J. R. Statist. Soc. B (2003)
65, *Part* 3, *pp.* 743–758

Bayesian inference for non-stationary spatial covariance structure via spatial deformations

Alexandra M. Schmidt

Federal University of Rio de Janeiro, Brazil

and Anthony O'Hagan

University of Sheffield, UK

[Received May 2001. Revised January 2003]

Summary. In geostatistics it is common practice to assume that the underlying spatial process is stationary and isotropic, i.e. the spatial distribution is unchanged when the origin of the index set is translated and under rotation about the origin. However, in environmental problems, such assumptions are not realistic since local influences in the correlation structure of the spatial process may be found in the data. The paper proposes a Bayesian model to address the anisotropy problem. Following Sampson and Guttorp, we define the correlation function of the spatial process by reference to a latent space, denoted by D, where stationarity and isotropy hold. The space where the gauged monitoring sites lie is denoted by G. We adopt a Bayesian approach in which the mapping between G and D is represented by an unknown function $\mathbf{d}(\cdot)$. A Gaussian process prior distribution is defined for $\mathbf{d}(\cdot)$. Unlike the Sampson–Guttorp approach, the mapping of both gauged and ungauged sites is handled in a single framework, and predictive inferences take explicit account of uncertainty in the mapping. Markov chain Monte Carlo methods are used to obtain samples from the posterior distributions. Two examples are discussed: a simulated data set and the solar radiation data set that also was analysed by Sampson and Guttorp.

Keywords: Anisotropy; Augmented covariance matrix; Environmental monitoring; Gaussian process; Spatial deformation

1. Introduction

Suppose that we have a network of n monitoring stations over a region of interest, G. The locations where the n monitoring sites lie are known in the literature as gauged locations. Any other location is an ungauged location. In environmental problems it is common that at each gauged location we have repeated measurements of some random variable Y. Specifically, suppose that we observe $Y(\mathbf{x}_i, t)$ for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, where $Y(\mathbf{x}, t)$ is a spatiotemporal process defined for $\mathbf{x} \in G$ and arbitrary time t. On the basis of the information obtained from the gauged locations the interest is in predicting $Y(\mathbf{x}, t)$ at any ungauged location $\mathbf{x}$ and time t of interest. This is a typical problem in environmental research. One usually has to describe the spatial structure of the process being considered, the mean structure and high order stochastic structures such as the spatial covariance $\mathbf{\Sigma}$. More specifically, in such applications, it is often important to describe the covariance structure between a set of gauged and ungauged locations in G.

Address for correspondence: Alexandra M. Schmidt, Departamento de Métodos Estatísticos, Instituto de Matemática, Universidade Federal do Rio de Janeiro, CP 68530 Rio de Janeiro, CEP 21.945-970, Brazil.
E-mail: alex@im.ufrj.br

 1369–7412/03/65743

In geostatistics the interpolation of the spatial process to ungauged locations of interest is known as *kriging* (Cressie, 1993). The covariance structure of the process is of great importance when using kriging. This method usually supposes that the stochastic process is stationary, which means that the distribution is unchanged when the origin of the index set is translated, and isotropic, i.e. the process is invariant under rotation about the origin. When a spatial process is stationary and isotropic it is called *homogeneous*. As Sampson and Guttorp (1992) pointed out, in the analysis of most spatiotemporal processes underlying environmental studies, there is little reason to expect spatial covariance structures to be stationary over the spatial scales of interest because there may be local influences in the correlation structure of the spatial random process. Guttorp *et al*. (1994) said that in analysing a variety of environmental data sets they had rarely found the assumption of homogeneity of the spatial process to be reasonable. Sampson and Guttorp (1992) pioneered an approach to modelling non-stationarity and anisotropy involving a latent space called D-space, where D stands for dispersion. The main idea is based on a non-linear transformation of the G- (geographical) space, into D-space, within which the spatial structure is stationary and isotropic. To obtain the locations of the sampled sites in D-space, Sampson and Guttorp (1992) made use of multidimensional scaling. Once the locations have been obtained in D-space, the next step is to estimate a realization of the spatial random process at an ungauged site, and Sampson and Guttorp interpolated the mapping of the gauged sites by thin plate splines. This approach has been applied to various problems (Guttorp *et al*. (1992, 1994), etc.) and is an important reference in this field. However, the use of multidimensional scaling and thin plate splines seems quite arbitrary. Another criticism of this approach is that once the points in D-space have been obtained they are fixed and the interpolation for ungauged locations is made on the basis of these fixed locations without taking into account uncertainty about the mapping. In the approach that is proposed here each location has a prior distribution, and its posterior distribution can be obtained by following the Bayes paradigm.

Two more difficulties with the approach of Sampson and Guttorp (1992) may be mentioned here. First, there can be many quite different mappings of the gauged locations into D-space that are almost equally good on any criterion of fit. This problem is acknowledged in multidimensional scaling but will also represent a potential difficulty for other related methods. Second, this mapping itself, or its interpolation or smoothing to allow prediction of ungauged sites, may not be bijective, i.e. the mapping may produce a surface that folds back on itself. Substantial folding of this type is often thought to be undesirable and/or implausible for environmental data.

Monestiez and Switzer (1991) also discussed the covariance structure for a non-stationary random field by using a transformation of the random field in $\mathbb{R}^2$ to a stationary isotropic field in a low dimensional space $\mathbb{R}^p$, $p = 2, \ldots, n-1$. They performed a simultaneous estimation of the isotropic covariance model in the rescaled space together with the multidimensional scaling of observed station locations, by an alternating least squares algorithm. Mardia and Goodall (1993) took a likelihood-based approach to the transformed isotropic covariance model introduced by Sampson and Guttorp (1992). They claimed that in many situations the maximum likelihood estimator yields estimates that are close to those obtained by a non-linear least squares approach. Actually they proposed an algorithm that is similar to that proposed by Monestiez and Switzer (1991), the difference being that they performed a unified minimization of the set of parameters in the model. Smith (1996) considered an alternative approach to that of Sampson and Guttorp, also based on the likelihood method but with the elements of the covariance function represented by a linear combination of radial basis functions. In terms of the thin plate splines, unlike Sampson and Guttorp, Smith (1996) did not make use of the smoothing parameter. He claimed that instead one can restrict the representation of the thin plate spline to a subset

of radial basis functions. However, it can be difficult to choose how many basis functions to use, and at which observations they should be centred. Problems of multiple solutions (multiple local maxima of the likelihood function) and folding arise in this approach also. Iovleff and Perrin (1999) also made use of spatial deformation to tackle spatial heterogeneity. However, they guaranteed bijectiveness of the function $\mathbf{d}(\cdot)$ by constructing a Delaunay triangulation for the geographical sites.

Loader and Switzer (1992) also discussed the estimation of non-homogeneous spatial processes. They made use of empirical Bayes techniques, so not taking into account the variability of the hyperparameters. They briefly discussed the idea of a latent space D wherein isotropy holds. Their covariance is not completely specified for all the space $\mathbb{R}^2 \times \mathbb{R}^2$; the model cannot be extended to describe the covariance between any two new locations that are not in the set $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Le and Zidek (1992) proposed a Bayesian alternative to kriging in which the uncertainty about the covariance matrix $\boldsymbol{\Sigma}$ is reflected. They defined an inverse Wishart prior for $\boldsymbol{\Sigma}$ with scale matrix $\boldsymbol{\Psi}$. Therefore, $\boldsymbol{\Sigma}$ does not necessarily take an isotropic form *a priori*. They used a partition for $\boldsymbol{\Sigma}$ and $\boldsymbol{\Psi}$ such that interpolation for ungauged locations is obtained through the gauged sites. They mentioned that the assumption of $\boldsymbol{\Sigma}$ being isotropic seems strong for environmetric problems. One suggestion is to use the nonparametric approach of Sampson and Guttorp (1992) to estimate $\boldsymbol{\Psi}$, and an example of its application can be seen in Sun *et al.* (1998). Higdon *et al.* (1999) proposed an alternative model for heterogeneity in the spatial covariance function, which is based on a moving average specification of a Gaussian process. They used a Bayesian hierarchical modelling structure so that uncertainty about all quantities involved in the model can be obtained.

This paper proposes a fully Bayesian model for the estimation of the spatial covariance function of processes that are used for environmental problems. Following the idea of Sampson and Guttorp (1992) we make use of a latent space D but map the locations into the latent space through a Gaussian process prior. Having derived the posterior distribution, predicted values and associated measures of uncertainty can be obtained at any location of the region being observed. Damian *et al.* (2001) have independently proposed a semiparametric Bayesian approach for the spatial deformation model. For a comparison of the two models see Schmidt (2001).

This paper is organized as follows. The next section presents the model and discusses the role of each of its parameters. Section 3.1 discusses the computational issues that are involved in obtaining the posterior distribution of the parameters of the model. Section 3.2 presents the predictive distributions that are obtained according to the model specified in Section 2. Section 4 presents the analysis of two examples. The first is a simulated data set with $n = 6$ sites. The other is the solar radiation data set that was analysed by Sampson and Guttorp (1992). Section 5 concludes the paper and discusses some future work.

2. The model

Following Sampson and Guttorp (1992), we consider a latent space D in which stationarity and isotropy hold. Here we discuss the $\mathbf{d}(\cdot)$ process, which is the function that maps the geographical locations into D-space. We consider only mappings from $\mathbb{R}^2$ to $\mathbb{R}^2$, but mappings to other dimensions can be developed as a generalization of the theory that is presented here.

Our observations comprise $Y_{it} = Y(\mathbf{x}_i, t)$ for $i = 1, \ldots, n$ and $t = 1, \ldots, T$. Let $\mathbf{Y}_t = (Y_{1t}, Y_{2t}, \ldots, Y_{nt})'$ for $t = 1, \ldots, T$. We suppose that $\mathbf{Y}_1, \ldots, \mathbf{Y}_T$ are independent and identically distributed with density $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the n-dimensional multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. We do not consider the temporal structure of the spatiotemporal process $Y(\mathbf{x}, t)$. In common with other researchers, we focus on the spatial structure (and

in particular on modelling non-homogeneity) and suppose that temporal dependence has been removed by preprocessing. For instance, we might typically adopt the same temporal model for each of the gauged sites, fit the parameters of that model at each location and compute the residuals at each time point. Effectively, we suppose that our observations Y_{it} are these residuals, or the results of some equivalent removal of temporal structure, and we ignore any correlations that are induced in these residuals by the fitting process. We shall concentrate on the spatial structure that is left in the data, and we shall refer to it as a spatial process only, although we retain the notation $Y(\mathbf{x}, t)$. Data in environmental statistics rarely follow a normal distribution, and where appropriate we further assume that some suitable transformation has previously been carried out.

Thus, the data yield an $n \times n$ sample covariance matrix $\mathbf{S}$, obtained from data at n spatial locations $\mathbf{x}_1, \ldots, \mathbf{x}_n$ over T time points. Our primary interest is in $\boldsymbol{\Sigma}$. After integrating out $\boldsymbol{\mu}$, using a uniform prior, the likelihood for $\boldsymbol{\Sigma}$ has a Wishart form:

$$f(\mathbf{S} \mid \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(T-1)/2} \exp\left\{ -\frac{T}{2} \operatorname{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) \right\}. \tag{1}$$

Each element of $\boldsymbol{\Sigma}$ is modelled through

$$\operatorname{cov}\{Y(\mathbf{x}_i, t), Y(\mathbf{x}_j, t)\} = \sqrt{\{v(\mathbf{x}_i) v(\mathbf{x}_j)\}}\, c_d(\mathbf{x}_i, \mathbf{x}_j), \tag{2}$$

where for all t

$$v(\mathbf{x}) = \operatorname{var}\{Y(\mathbf{x}, t)\},$$
$$c_d(\mathbf{x}, \mathbf{x}') = \operatorname{corr}\{Y(\mathbf{x}, t), Y(\mathbf{x}', t)\},$$

with $\operatorname{corr}(\cdot, \cdot)$ denoting correlation. For simplicity we model the variances as exchangeable. *A priori*, the variances $v(\mathbf{x})$ have inverse gamma distributions with mean τ^2 and known degrees of freedom f, i.e.

$$v(\mathbf{x}) \mid \tau^2, f \sim \mathrm{IG}\{\tau^2(f-2), f\}, \qquad \mathbf{x} \in G, \tag{3}$$
$$\pi(\tau^2) \propto \tau^{-2},$$

giving τ^2 a non-informative prior distribution.

2.1. Mapping the spatial correlation

The main feature of the model is in the correlation structure $c_d(\cdot, \cdot)$ of the spatial random process, defined by reference to a latent space D. In geostatistics it is usual to represent spatial variability through the variogram rather than the covariance or correlation function. This is defined as the variance of the difference of the process at two locations, i.e. $D^2(\mathbf{x}, \mathbf{x}') = \operatorname{var}\{Y(\mathbf{x}, t) - Y(\mathbf{x}', t)\}$. Sampson and Guttorp (1992) asserted that nearly all variogram models that are common in geostatistical practice assuming stationarity can be expressed as

$$D^2(\mathbf{x}, \mathbf{x}') = f^*(||\mathbf{d}(\mathbf{x}) - \mathbf{d}(\mathbf{x}')||), \tag{4}$$

where $\mathbf{d}(\cdot)$ is a linear transformation in the case of elliptical anisotropy and the identity function, or any orthonormal transformation, in the case of an isotropic covariance structure, and where f^* is an appropriately chosen monotone function. $||\cdot||$ denotes a fixed distance measure or norm on D, and Euclidean distance is assumed. The model of Sampson and Guttorp (1992) accommodates non-stationary structure by permitting $\mathbf{d}(\cdot)$ to be non-linear. Their method for estimating $\mathbf{d}(\cdot)$ and f^* was briefly described in Section 1; for more details see Sampson and Guttorp (1992). Here we propose the modelling of the spatial covariances as shown in

equation (2). The mapping of the gauged sites occurs through the correlation function defined as

$$c_d(\mathbf{x}_i, \mathbf{x}_j) = g(||\mathbf{d}(\mathbf{x}_i) - \mathbf{d}(\mathbf{x}_j)||), \tag{5}$$

where $g(\cdot)$ is a monotone function. Similarly to Sampson and Guttorp's approach, $g(\cdot)$ is defined on the basis of the general representation for a two-dimensional isotropic covariance function (Cressie (1993), page 85), as a mixture of K Gaussian correlation functions. Specifically,

$$g(h) = \sum_{k=1}^{K} a_k \exp(-b_k h^2). \tag{6}$$

For parsimony, the number of components K should be as small as possible consistent with the data, and the roughness parameter constants b_k and the coefficients a_k are unknown with the a_ks subject to $\Sigma_{k=1}^{K} a_k = 1$. The b_ks satisfy $b_1 > \ldots > b_K$ to avoid identifiability problems and $a_k > 0$ and $b_k > 0$, $k = 1, \ldots, K$. A nugget effect can be easily introduced by letting b_1 in equation (6) tend to ∞.

2.2. The $\mathbf{d}(\cdot)$ process

The correlation function depends on the distance between the sites in D-space, and $\mathbf{d}(\cdot)$ is the function that maps locations in G-space into D-space. Here we assign a Gaussian process (GP) (see for example O'Hagan (1994)) prior to the function $\mathbf{d}(\cdot)$:

$$\mathbf{d}(\cdot) \,|\, \mathbf{m}(\cdot), \sigma_d^2, R_d(\cdot,\cdot) \sim \mathrm{GP}\{\mathbf{m}(\cdot), \sigma_d^2 R_d(\cdot,\cdot)\}, \tag{7}$$

where $\mathbf{m}(\cdot)$ is the prior mean function. The covariance function is a 2×2 covariance matrix σ_d^2 multiplied by the scalar function $R_d(\cdot,\cdot)$, which measures the prior correlation between the sites, such that $R_d(\mathbf{x}, \mathbf{x}) = 1$.

In particular, let $\mathbf{D} = (\mathbf{d}_1, \ldots, \mathbf{d}_n)$, where $\mathbf{d}_i = \mathbf{d}(\mathbf{x}_i)$, be the $2 \times n$ matrix of D-space coordinates of the gauged sites, let $\mathbf{m} = (\mathbf{m}(\mathbf{x}_1), \ldots, \mathbf{m}(\mathbf{x}_n))$ and let $\mathbf{R}_d$ be the $n \times n$ matrix with elements $R_d(\mathbf{x}_i, \mathbf{x}_j)$. Then the prior for $\mathbf{D}$ is the matrix normal distribution $(\mathbf{D} \,|\, \mathbf{m}, \sigma_d^2, \mathbf{R}_d) \sim N_{(2\times n)}(\mathbf{m}, \sigma_d^2, \mathbf{R}_d)$. If we consider the vector that is obtained by stacking the columns of $\mathbf{D}$, then we have a multivariate normal distribution of dimension $2n$, whose covariance matrix is given by the Kronecker product between σ_d^2 and $\mathbf{R}_d$.

The GP formulation for $\mathbf{d}(\cdot)$ tends to eliminate the kind of non-injective mappings that were noted by Sampson and Guttorp (1992). It will also tend to reduce bimodality problems, referred to particularly by Smith (1996).

Although $\mathbf{m}(\cdot)$ might in principle take any form reflecting prior beliefs about the mapping from G- to D-space, it is natural to set $\mathbf{m}(\mathbf{x}) = \mathbf{x}$ when there is no prior information about how D would differ from G. This also helps to control overfitting. Although we do not observe the $\mathbf{d}(\cdot)$ process, the likelihood brings information about the distances between the sites in D-space. Therefore the prior of $\mathbf{D}$ pulls the posterior towards the G-locations, whereas the likelihood gives higher posterior probability to configurations in D-space for which the implied correlations given by equation (5) approximate the sample correlations given by $\mathbf{S}$. The covariance structure of the GP is of great importance in describing the behaviour of the function $\mathbf{d}(\cdot)$. It controls the degree of distortion of the mapping *a priori*. The following subsections discuss each of its components.

2.2.1. *Specification of $R_d(\cdot,\cdot)$*

The correlation matrix $R_d(\cdot,\cdot)$ gives information about local distortions among the sites in D-space and controls the degree of smoothness of the GP. In particular, the $n \times n$ matrix $\mathbf{R}_d$

describes the prior correlation structure of the gauged locations in D-space. This is based on the correlation between the sites in G-space. It is reasonable to expect that neighbouring sites in G-space will tend to stay close together when mapped into D and so their locations in D-space should be highly correlated. Elements of $R_d(\cdot,\cdot)$ are accordingly modelled via a Gaussian correlation function,

$$R_d(\mathbf{x},\mathbf{x}') = \exp(-b_d||\mathbf{x}-\mathbf{x}'||^2),$$

where b_d controls the prior shape of the configuration of the gauged locations in D-space. One suggestion is to fix b_d equal to $1/2a$, where a is a typical squared distance between gauged locations in G-space. This choice means that gauged locations that are relatively close together in G-space will tend to stay close together in D-space, but that points further apart in G-space are less well correlated *a priori*, and so may have their distances more distorted by the mapping.

2.2.2. *Specification of* $\boldsymbol{\sigma}_d^2$

The prior covariance structure of the co-ordinate system in D-space is also controlled by $\boldsymbol{\sigma}_d^2 = \text{var}\{\mathbf{d}(\mathbf{x})\}$. Whereas $R_d(\cdot,\cdot)$ controls the nature of the distortion which is expected *a priori* in mapping from G to D, by determining which locations are strongly correlated, $\boldsymbol{\sigma}_d^2$ controls the amount of that distortion. The larger the elements of $\boldsymbol{\sigma}_d^2$, the more the prior distribution allows individual gauged locations to move.

It can be shown that this parameter is unidentifiable in the sense of Dawid (1979). The sample covariance matrix $\mathbf{S}$ brings information about the distances in D-space; therefore it gives information about the size and shape of the configuration of points in D-space. It can be proved (see Appendix A) that $\mathbf{S}$ will at most give information about the eigenvalues of $\boldsymbol{\sigma}_d^2$. For this reason we model $\boldsymbol{\sigma}_d^2$ as a diagonal matrix. We suppose that the elements of the main diagonal are independent and we assign an inverse gamma prior with β_i degrees of freedom and scale parameter α_i to diagonal element $\boldsymbol{\sigma}_{d_{ii}}^2$, $i = 1, 2$.

2.3. Posterior distribution

By defining the prior distribution of the parameters in the correlation function in equation (6) and combining this information with the likelihood in distribution (1), the posterior distribution of the whole set of parameters can be obtained by using Bayes's theorem.

Considering equation (6), suppose that, conditional on K, a_j and b_l are independent for every $j \neq l$ and $j, l = 1, \ldots, K$ with prior density given by

$$\pi(a_1,\ldots,a_K,b_1,\ldots,b_K|K) \propto \prod_{k=1}^{K} \pi_k(b_k) \qquad \text{with } \sum_{k=1}^{K} a_k = 1 \text{ and } b_1 > \ldots > b_K, \tag{8}$$

where $\Pi_{k=1}^{K}\pi_k(b_k)$ is the kernel of the prior joint density of $\mathbf{b}' = (b_1,\ldots,b_K)$ and $\pi_k(b_k)$ is the kernel of the log-normal density whose associated normal has mean μ_b and variance σ_b^2, $k = 1,\ldots,K$. The a_ks have a uniform prior distribution over the $(K-1)$-simplex. The complete set of parameters is

$$\boldsymbol{\theta} = \{(v_1, v_2, \ldots, v_n), (\mathbf{d}_1,\ldots,\mathbf{d}_n), (a_1,\ldots,a_K,\mathbf{b}), (\tau^2, \boldsymbol{\sigma}_d^2)\}.$$

According to Bayes's theorem, the posterior for $\boldsymbol{\theta}$ is proportional to the prior times the likelihood, which is easy to obtain following the specifications that have so far been presented,

$$\pi(\boldsymbol{\theta}|\mathbf{S}) \propto |\boldsymbol{\Sigma}|^{-(T-1)/2} \exp\left\{-\frac{T}{2}\,\mathrm{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1})\right\} \left[\prod_{i=1}^{n} v_i^{-(f+2)/2} \exp\left\{-\frac{(f-2)\tau^2}{2v_i}\right\}\right]$$
$$\times\, \tau^{(nf-2)/2}\left(\prod_{k=1}^{K}\frac{1}{b_k}\exp\left[\frac{-\{\log(b_k)-\mu_b\}^2}{2\sigma_b^2}\right]\right)$$
$$\times\, |\boldsymbol{\sigma}_d^2|^{-n/2}|\mathbf{R}_d|^{-1}\exp\left\{-\frac{1}{2}\mathrm{tr}(\mathbf{D}-\mathbf{m})'\boldsymbol{\sigma}_d^{-2}(\mathbf{D}-\mathbf{m})\mathbf{R}_d^{-1}\right\}$$
$$\times\, \sigma_{d_{11}}^{-(\beta_1+2)/2}\exp\left(-\frac{\alpha_1}{2\sigma_{d_{11}}^2}\right)\sigma_{d_{22}}^{-(\beta_2+2)/2}\exp\left(-\frac{\alpha_2}{2\sigma_{d_{22}}^2}\right). \tag{9}$$

3. Implementation and predictive inference

3.1. Markov chain Monte Carlo algorithm

Analytical summarization of expression (9) seems infeasible and we resort to Markov chain Monte Carlo simulation (Gamerman, 1997).

The model that was introduced in the previous section has a complex hierarchical structure. It is important to recognize that without constraints there is an inherent unidentifiability due to the way that $\mathbf{d}(\cdot)$ appears in the $g(\cdot)$ function. First, any transformation of co-ordinates $\mathbf{d}(\mathbf{x}_i)$ in D which leaves distances unchanged (translation and/or rotation) is observationally equivalent. Second, because $g(\cdot)$ has unspecified roughness parameters b_k, any transformation which multiplies all distances in D-space by a constant is also unidentifiable. Since the GP for $\mathbf{d}(\cdot)$ acts as a proper prior distribution, this unidentifiability in the data causes no fundamental problems. It means that the posterior for $\mathbf{d}(\cdot)$ will always be proportional to the prior for any of these observationally equivalent configurations, but the posterior will be proper and the computational methods that are considered later will work. However, it is important to acknowledge this to design an efficient algorithm. The algorithm to obtain samples from the posterior in expression (9) is a Gibbs sampler. Following the full conditionals of each of the parameters we

(a) sample v_i by using adaptive rejection sampling (Gilks and Wild, 1992) as its full conditional is log-concave when expressed in terms of $v_i^{-1/2}$,
(b) sample elements of $\mathbf{D}$ by Metropolis–Hastings steps,
(c) sample $a_1, \ldots, a_K$,
(d) also sample $b_2, \ldots, b_K$ by Metropolis–Hastings steps,
(e) sample τ^2 from its full conditional gamma distribution and
(f) sample $\sigma_{d_{11}}^2$ and $\sigma_{d_{22}}^2$ from their respective inverse gamma distributions.

The two most complex steps are to update $\mathbf{D}$ and $b_2, \ldots, b_K$. As the likelihood brings information about the distances between the sites in D-space and the prior gives information about their locations in D, it is difficult to ensure that the chain mixes well. Below we concentrate on the description of how to obtain samples of $\mathbf{D}$ and $b_2, \ldots, b_K$.

Following the posterior in expression (9) the full conditional for $\mathbf{d}_i$ is given by

$$\pi^*(\mathbf{d}_i) \propto |\boldsymbol{\Sigma}|^{-(T-1)/2}\exp\left\{-\frac{T}{2}\,\mathrm{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1})\right\}\exp\left\{-\frac{1}{2}\,\mathrm{tr}(\mathbf{D}-\mathbf{m})'\boldsymbol{\sigma}_d^{-2}(\mathbf{D}-\mathbf{m})\mathbf{R}_d^{-1}\right\}.$$

The Metropolis–Hastings proposal that we use is based on the principal components of the sample covariance matrix $\mathbf{S}$, and instead of sampling locations we sample directions. If we sample the locations separately, one at a time, then those which are highly correlated tend not to move much. In sampling the directions we overcome this problem because the principal components of $\mathbf{S}$ indicate how the locations are correlated. Making use of them ensures that sites

which are highly correlated, those with small values of the principal components, tend to move together along the D-space. The full conditional posterior distribution of $\mathbf{D}$ is the combination between the prior multivariate normal distribution and the likelihood, which is invariant under rotation, location and scale changes of the configuration of points. By reparameterizing the $\mathbf{d}$s in terms of the principal components of $\mathbf{S}$, we are sampling in the region of the D-space where the likelihood gives more support. In practice, we define $\mathbf{E} = \mathbf{A}^{-1}\mathbf{D}$, with $\mathbf{A}$ the $n \times n$ matrix whose ith column is the ith principal component of $\mathbf{S}$. Then we use a Metropolis random-walk step to sample $\mathbf{E}_i$, $i = 1, \ldots, n$. Although movement is proposed only to the ith component, as $\mathbf{D} = \mathbf{AE}$, all the locations move accordingly. The variances of the proposal distributions are proportional to the corresponding eigenvalues of $\mathbf{S}$.

To sample $b_2, \ldots, b_K$ we first suppose that the correlation function (6) has only two components ($K = 2$), one of them a nugget effect. Therefore we have only one roughness parameter, b_2, in the correlation function. The parameter b_2 brings information about the size of the configuration of the locations in D-space. The full conditional distribution of b_2 is tightly concentrated, allowing little movement in the Markov chain. To improve mixing, our algorithm is based on moving $\mathbf{D} = (\mathbf{d}_1, \ldots, \mathbf{d}_n)$ and b_2 together, but making proposals only to b_2 and obtaining the proposal for $\mathbf{D}$ such that the distances between the proposed points in D-space keep the current correlations the same. Thus, a proposal is made for $\log(b_2)$ from a normal distribution with mean equal to the logarithm of its current value and fixed variance U_b. The proposed value for $\mathbf{d}_i$ is

$$\mathbf{d}^{\text{prop}}(\mathbf{x}_i) = \sqrt{\left(\frac{b_2^{\text{cur}}}{b_2^{\text{prop}}}\right)}\{\mathbf{d}^{\text{cur}}(\mathbf{x}_i) - \bar{\mathbf{x}}\} + \bar{\mathbf{x}}, \tag{10}$$

where $\bar{\mathbf{x}}$ is the mean point of the configuration of points in G-space, cur stands for the current value and prop stands for the proposed value of a parameter. The proposed $\mathbf{D}$ in equation (10) satisfies the equality

$$b_2^{\text{cur}}||\mathbf{d}^{\text{cur}}(\mathbf{x}_i) - \mathbf{d}^{\text{cur}}(\mathbf{x}_j)||^2 = b_2^{\text{prop}}||\mathbf{d}^{\text{prop}}(\mathbf{x}_i) - \mathbf{d}^{\text{prop}}(\mathbf{x}_j)||^2, \tag{11}$$

so the likelihood is unchanged. The movement is accepted with probability

$$\alpha = \min\{1, \pi(\mathbf{D}^{\text{prop}})\,\pi(b_2^{\text{prop}})/\pi(\mathbf{D}^{\text{cur}})\,\pi(b_2^{\text{cur}})\}.$$

When there are more than two components in the correlation function, we still make proposals for $\mathbf{d}_1, \ldots, \mathbf{d}_n$ and $b_2, \ldots, b_K$ together but now the equality in expression (11) cannot hold for all K exponential terms in the correlation function $g(\cdot)$. To obtain an effect similar to that of equation (11) we take $\exp(x) \approx 1 + x$. Then the analogous equality to expression (11) is

$$\begin{aligned}(a_2 b_2^{\text{cur}} + a_3 b_3^{\text{cur}} + \ldots + a_K b_K^{\text{cur}})||\mathbf{d}^{\text{cur}}(\mathbf{x}_i) - \mathbf{d}^{\text{cur}}(\mathbf{x}_j)||^2 \\ = (a_2 b_2^{\text{prop}} + a_3 b_3^{\text{prop}} + \ldots + a_K b_K^{\text{prop}})||\mathbf{d}^{\text{prop}}(\mathbf{x}_i) - \mathbf{d}^{\text{prop}}(\mathbf{x}_j)||^2.\end{aligned} \tag{12}$$

Now the proposal for $\phi = (b_2, \ldots, b_K, \mathbf{D})$ is obtained by sampling each $\log(b_k)$ from a normal distribution with mean equal to the logarithm of its current value, and the proposal for $\mathbf{D}$ is obtained by using

$$\mathbf{d}^{\text{prop}}(\mathbf{x}_i) = \sqrt{\left(\frac{a_2 b_2^{\text{cur}} + \ldots + a_K b_K^{\text{cur}}}{a_2 b_2^{\text{prop}} + \ldots + a_k b_K^{\text{prop}}}\right)}\{\mathbf{d}^{\text{cur}}(\mathbf{x}_i) - \bar{\mathbf{x}}\} + \bar{\mathbf{x}}.$$

For a general K the movement from ϕ^{cur} to ϕ^{prop} is accepted with probability

$$\alpha = \min\left\{\frac{f(\mathbf{S}|\phi^{\text{prop}})\,\pi(\mathbf{D}^{\text{prop}})\,\pi(\mathbf{b}^{\text{prop}})}{f(\mathbf{S}|\phi^{\text{cur}})\,\pi(\mathbf{D}^{\text{cur}})\,\pi(\mathbf{b}^{\text{cur}})}, 1\right\}.$$

3.2. Predictive inference

This subsection obtains the predictive distribution for the simple case of one ungauged location. Assume that we wish to predict $Y^*_{n+1,t} = Y^*$ at an ungauged location $\mathbf{x}_{n+1}$, at some time $t \in \{1, \ldots, T\}$, conditioned on the observations $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_T)$, of the n gauged sites. Now the parameter vector $\boldsymbol{\theta}$ is augmented to $\boldsymbol{\theta}^* = (\boldsymbol{\theta}, v^*_{n+1}, \mathbf{d}^*_{n+1})$, where v^*_{n+1} is the variance of the process at this ungauged location and $\mathbf{d}^*_{n+1}$ is its position in D-space. The predictive distribution of $Y^* \mid \mathbf{Y}$ is now given by

$$p(Y^*|\mathbf{Y}) = \int_{\boldsymbol{\theta}^*} p(Y^* \mid \mathbf{Y}, \boldsymbol{\theta}^*)\, p(\boldsymbol{\theta}, v^*_{n+1}, \mathbf{d}^*_{n+1} \mid \mathbf{Y})\, \mathrm{d}\boldsymbol{\theta}\, \mathrm{d}v^*_{n+1}\, \mathrm{d}\mathbf{d}^*_{n+1}. \tag{13}$$

Using basic results from multivariate normal theory we have

$$(Y^* \mid \mathbf{Y}, \boldsymbol{\theta}^*) \sim N(\mu^* + \boldsymbol{\psi}'\boldsymbol{\Sigma}^{-1}\mathbf{Y}; v^*_{n+1} - \boldsymbol{\psi}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\psi}), \tag{14}$$

where μ^* is the mean of the process at $\mathbf{x}_{n+1}$ and $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_n)'$ with

$$\psi_i = \sqrt{(v^*_{n+1} v_i)} \sum_{k=1}^{K} a_k \exp\{-b_k ||\mathbf{d}^*_{n+1} - \mathbf{d}(\mathbf{x}_i)||^2\}.$$

We obtain $p(\boldsymbol{\theta}, v^*_{n+1}, \mathbf{d}^*_{n+1} \mid \mathbf{Y})$ in equation (13), via

$$p(\boldsymbol{\theta}, v^*_{n+1}, \mathbf{d}^*_{n+1}|\mathbf{Y}) = p(v^*_{n+1}, \mathbf{d}^*_{n+1}|\boldsymbol{\theta}, \mathbf{Y})\, p(\boldsymbol{\theta}|\mathbf{Y}). \tag{15}$$

As v^*_{n+1} and $\mathbf{d}^*_{n+1}$ do not figure in the likelihood it follows that $p(v^*_{n+1}, \mathbf{d}^*_{n+1} \mid \boldsymbol{\theta}, \mathbf{Y}) = p(v^*_{n+1}, \mathbf{d}^*_{n+1} \mid \boldsymbol{\theta})$. Therefore $(v^*_{n+1} \mid \boldsymbol{\theta}, \mathbf{S}) \sim \text{IG}\{\tau^2(f-2); f\}$ and the posterior distribution of $\mathbf{d}^*_{n+1}$ follows from its prior conditioned on $\boldsymbol{\theta}$ together with results for partitions of the multivariate normal distribution, i.e.

$$(\mathbf{d}^*_{n+1}|\boldsymbol{\theta}, \mathbf{S}) \sim N_2(\mathbf{m}_{d^*}; \boldsymbol{\Sigma}_{d^*}) \tag{16}$$

where

$$\mathbf{m}_{d^*} = \mathbf{m}^* + (\mathbf{I}_2 \otimes \mathbf{R}^{*\prime}_d \mathbf{R}_d^{-1})(\mathbf{D} - \mathbf{m}),$$

$$\boldsymbol{\Sigma}_{d^*} = \sigma^2_d (1 - \mathbf{R}^{*\prime}_d \mathbf{R}_d^{-1} \mathbf{R}^*_d),$$

$\mathbf{R}^{*\prime}_d = (R_{d_1}, \ldots, R_{d_n})$ and $R_{d_i} = \exp(-b_d ||\mathbf{x}^*_{n+1} - \mathbf{x}_i||^2)$, $i = 1, 2, \ldots, n$. If the prior mean of $\mathbf{D}$ is the identity then $\mathbf{m}^*$ is the vector containing the G-space co-ordinates of the ungauged location.

The integration in equation (13) is analytically intractable but can be computed by Monte Carlo methods. A simpler question is how to obtain the augmented covariance matrix for both gauged and ungauged locations. As mentioned previously, we can use the Markov chain Monte Carlo algorithm to obtain a sample of size L from the posterior distribution of the parameter vector $\boldsymbol{\theta}$ and then for each $l = 1, \ldots, L$

(a) sample v^*_{n+1} from $(v^*_{n+1}|\tau^{2^l}, \mathbf{S}) \sim \text{IG}\{\tau^{2^l}(f-2), f\}$,
(b) sample $\mathbf{d}^*_{n+1}$ from $(\mathbf{d}^*_{n+1}|\mathbf{D}^l, \mathbf{S}) \sim N_2(\mathbf{m}_{d^*}, \boldsymbol{\Sigma}_{d^*})$, where $\mathbf{m}_{d^*}$ and $\boldsymbol{\Sigma}_{d^*}$ are the same as in distribution (16), and

(c) compute

$$\Sigma_{ij}^{l} = \sqrt{\{v^{l}(\mathbf{x}_i)\, v^{l}(\mathbf{x}_j)\}} \sum_{k=1}^{K} a_k^{l} \exp\{-b_k^{l} ||\mathbf{d}(\mathbf{x}_i)^{l} - \mathbf{d}(\mathbf{x}_j)^{l}||^2\}, \qquad \text{for } i, j = 1, \ldots, n+1;$$

$\boldsymbol{\Sigma}^1, \ldots, \boldsymbol{\Sigma}^L$ is then a sample from the posterior distribution of the larger covariance matrix.

4. Examples

4.1. Simulated data set

We first fit the model to a simulated data set to check the behaviour of the posterior inferences about parameters of the model when we know their true values.

Initially we established the co-ordinates of $n = 6$ sites in G-space; the values of τ^2, f, b_2, b_d and $\boldsymbol{\sigma}_d^2$ were fixed and after generating the locations in D-space from the prior in distribution (7), and the variance at each gauged location from the prior in distribution (3), a 'true' covariance matrix $\boldsymbol{\Sigma}$ was obtained with elements given by equation (2). The variables were fixed at the values $\tau^2 = 2, f = 12, b_2 = 0.25, a_1 = 0.1, a_2 = 0.9, \sigma_{d_{11}}^2 = 0.25, \sigma_{d_{22}}^2 = 0.375$ and $b_d = 0.5$.

As this data set is being simulated straight from the model, it is not necessary to remove the temporal effect from the data at each gauged site. Since the objective here is to test the model in ideal conditions, the sample covariance matrix that was used was the true simulated matrix but we suppose that it arose from a sample of $T = 200$ observations. The next step is to assign prior distributions for the parameters in the model. The prior for b_2 is assumed to be log-normal with mean 0.25 and variance such that the prior probability of b_2 being greater than 2 is 0.01. The parameters τ^2, a_1 and a_2 have non-informative priors.

As discussed in Section 2, the prior information about $\mathbf{d}(\cdot)$ is assigned through $\mathbf{m}(\cdot)$, b_d and $\boldsymbol{\sigma}_d^2$. We are using the geographical location $\mathbf{x}$ as the prior mean of $\mathbf{d}(\mathbf{x})$. Following a suggestion by one of the referees we ran several different models to check the effect of the priors for b_d and $\boldsymbol{\sigma}_d^2$ on the posterior of $\mathbf{d}(\cdot)$. Two results are shown here:

(a) for $b_d = 0.25$ and prior means for $\sigma_{d_{11}}^2$ and $\sigma_{d_{22}}^2$ equal to 0.5 and 0.75 respectively and
(b) for $b_d = 1.0$ and prior means for $\sigma_{d_{11}}^2$ and $\sigma_{d_{22}}^2$ equal to 1.0 and 1.5 respectively.

These models were run for 90000 iterations with a burn-in of 20000 iterations and samples were stored after every 35th iteration. The trace plots of all the parameters suggest that convergence has been reached.

In shape analysis, Procrustes superimposition (Dryden and Mardia, 1998) is a useful tool to compare the shape between two different configurations of n points, as it minimizes the distance in shape between them. Fig. 1 shows the Procrustes superimposition of the posterior mean of $\mathbf{D}$ onto the original configuration in G-space. G_i marks the location of site i in G-space, whereas s_i marks the Procrustes superimpostion of the posterior mean of its location in D-space. Note that the shapes of both configurations in D-space do not differ much. We use this transformation to display other aspects of the mapping $\mathbf{d}(\cdot)$ in subsequent figures. For instance, Fig. 2 shows the posterior mean for both models of the mapping of a grid with 200 ungauged sites in G-space. This grid was computed by drawing a square grid over the G-space and, at each point where two lines on the grid meet, the corresponding position in D-space was computed following equation (16) and Procrustes transformation. As can be observed, the mapping of the locations into D-space is quite smooth. There is shrinkage of the configuration in the north-east direction on the bottom left-hand side of the grid. This shrinkage is smoother in Fig. 2(a) than in Fig. 2(b), in accordance with the prior information about the deformation. Fig. 3 exhibits the

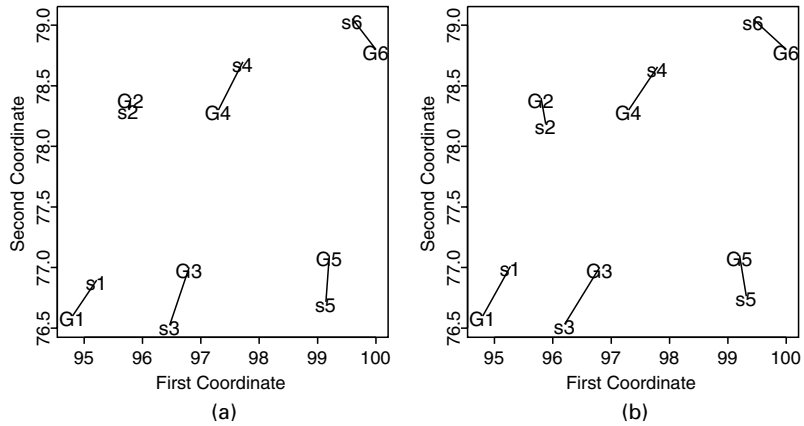

Fig. 1. Procrustes superimpositions of the mean configurations in $D(s_i)$ onto the original locations in G-space (G_i): (a) $b_d = 0.25$, $E(\sigma^2_{d_{11}}) = 0.5$ and $E(\sigma^2_{d_{22}}) = 0.75$; (b) $b_d = 1.0$, $E(\sigma^2_{d_{11}}) = 1.0$ and $E(\sigma^2_{d_{22}}) = 1.5$

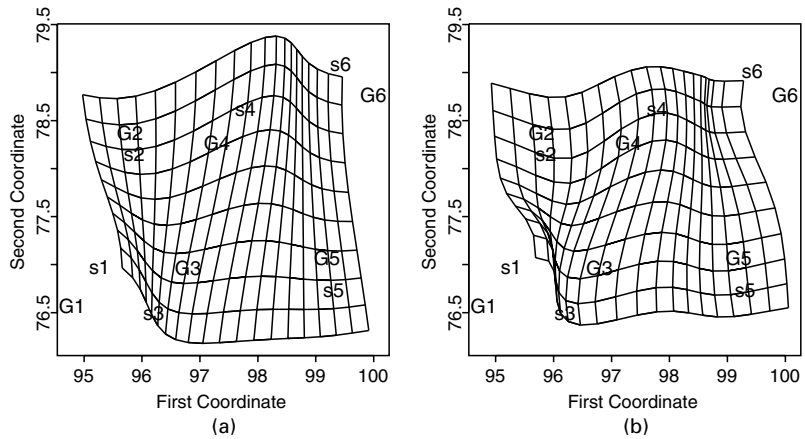

Fig. 2. Mapping of a rectangular grid of 200 points from G-space into D-space: (a) $b_d = 0.25$, $E(\sigma^2_{d_{11}}) = 0.5$ and $E(\sigma^2_{d_{22}}) = 0.75$; (b) $b_d = 1.0$, $E(\sigma^2_{d_{11}}) = 1.0$ and $E(\sigma^2_{d_{22}}) = 1.5$

estimated correlation function and pointwise 95% highest posterior density interval under each prior distribution. The dots in the plot represent the observed correlations. In both cases we obtain an isotropic correlation function. Fig. 4 illustrates the posterior uncertainty in each $\mathbf{d}_i$ by plotting the principal axes of ellipses centred at the posterior mean s_i, together with the original geographical location G_i and the 'true' location in D-space (shown as D_i). Again, this is shown for the two alternative prior distributions. We see that the effect of varying the parameters b_d and the prior distribution of $\boldsymbol{\sigma}^2_d$ has a modest but potentially important effect on posterior inferences. Prior (a) expresses a belief that the mapping $\mathbf{d}(\cdot)$ will be closer to its prior mean (the identity function) than prior (b), and in particular that the shape of the configuration in D-space should be close to its prior G-space expectation. The graphs are consistent with these differences in prior specification. There is relatively little information in the data about the mapping $\mathbf{d}(\cdot)$, and our Bayesian approach recognizes and quantifies the consequent posterior uncertainty in $\mathbf{d}(\cdot)$, and the fact that inferences will be influenced by prior information. The original approach

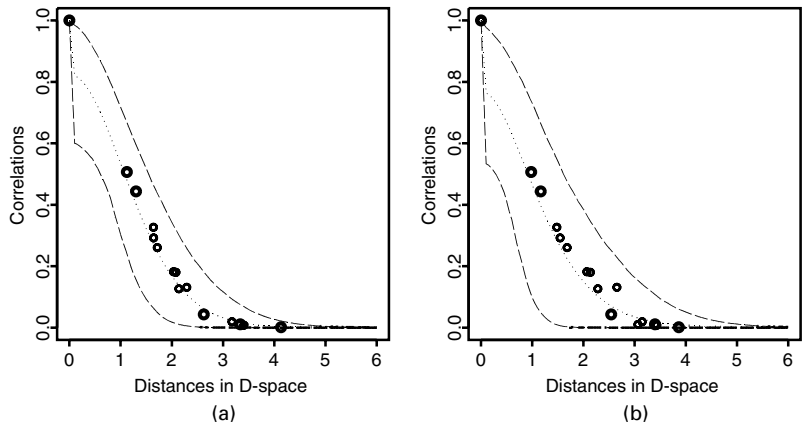

Fig. 3. Posterior mean (·····) with the associated 95% highest posterior density interval (– – –): (a) $b_d = 0.25$, $E(\sigma^2_{d_{11}}) = 0.5$ and $E(\sigma^2_{d_{22}}) = 0.75$; (b) $b_d = 1.0$, $E(\sigma^2_{d_{11}}) = 1.0$ and $E(\sigma^2_{d_{22}}) = 1.5$

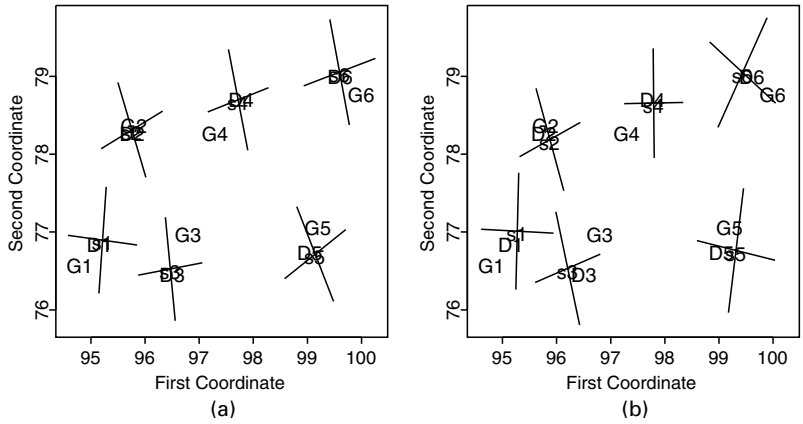

Fig. 4. Locations of the posterior mean of the sites in D-space together with their associated uncertainty: (a) $b_d = 0.25$, $E(\sigma^2_{d_{11}}) = 0.5$ and $E(\sigma^2_{d_{22}}) = 0.75$; (b) $b_d = 1.0$, $E(\sigma^2_{d_{11}}) = 1.0$ and $E(\sigma^2_{d_{22}}) = 1.5$

of Sampson and Guttorp does not represent uncertainty in $\mathbf{d}(\cdot)$ at all and so may be seen as giving a misleading impression of precision in its estimates.

4.2. Solar radiation data set

Here we apply the model to the data set that was analysed by Sampson and Guttorp (1992), i.e. the solar radiation data set collected in south-western British Columbia, Canada. As shown in Fig. 3 of Sampson and Guttorp (1992), the network consists of $n = 12$ monitoring stations and we analyse only the spring–summer data set, comprising observations from March 22nd to September 20th, for 4 years, from 1980 until 1983. Thus $T = 732$. We used the seasonal sample covariance matrices after removing the temporal effect from the original data, as analysed by Sampson and Guttorp.

Fig. 5(a) shows the observed spatial correlations *versus* the geographical distances. Anisotropy in the data is clear from this plot. Many sites are separated by roughly the same distance, yet the correlations between them are different. It is difficult to believe that any of the usual

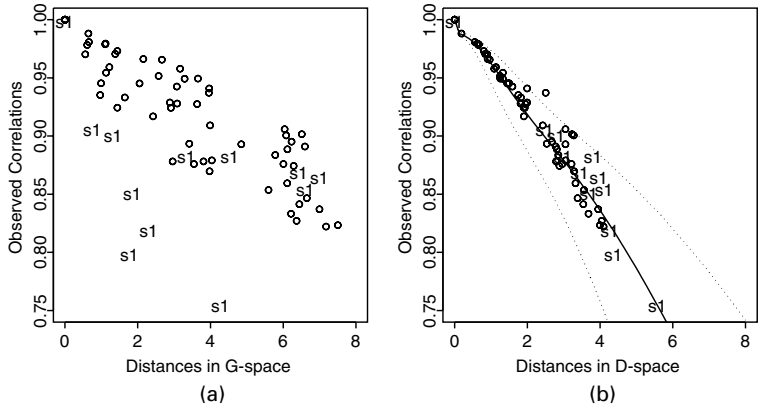

Fig. 5. (a) Geographical distances *versus* the observed correlations of the solar radiation data set and (b) estimated posterior mean (———) and 95% highest posterior density interval (· · · · · · ·) of the correlation function for the solar radiation data set

isotropic correlation functions would fit these data. As pointed out by Sampson and Guttorp (1992), site 1 (Grouse Mountain station) is at an elevation of 1128 m, whereas all the other stations lie below 130 m altitude; this might be one of the sources of the anisotropy in this data set.

From Fig. 5(a) we see that the correlations vary in the range $(0.75, 1)$, which indicates that the roughness parameters $b_2, \ldots, b_K$ of the correlation function $g(\cdot)$ in equation (6) should be small. Therefore the prior of b_i is a log-normal distribution with mean 0.1 and variance such that the probability that b_i is greater than 2 is 0.01, $i = 1, 2$. The prior degrees of freedom of the variances f were set equal to 10, as we do not have strong prior belief about the variances of the locations. For this example we set b_d, the roughness parameter of the prior correlation function of the locations in D-space, to 1.7. As previously discussed, $\sigma^2_{d_{ii}}$, $i = 1, 2$, controls how far the sites are expected to move *a priori*. The prior expectation of $\sigma^2_{d_{ii}}$ was set to 0.5. Initially we attempted to fit the model with $K = 2$ components in the correlation function but the estimated correlation function did not fit the observed correlations well. We therefore considered $K = 3$. Once convergence was considered to be reached we stored a sample of size 2000 of the set of parameters. Fig. 5(b) presents the posterior mean of the correlation function with its 95% highest posterior density interval. It is clear that the model fitted succeeds in making the observed correlations isotropic.

Fig. 6(a) shows the Procrustes superimposition of the posterior mean of the gauged sites in D onto the original configuration in G-space. Site 1 moves further north indicating that it is less correlated with the others than the original configuration suggests. Note that, when considering the whole configuration in D, we obtain an east–west shrinkage but a north–south stretching. This agrees with the results obtained by Sampson and Guttorp (1992). Fig. 6(b) shows the result of the mapping of a grid with 200 ungauged points in G-space. This mapping results in some folding over itself, close to sites 7, 8 and 9 and again in the area around sites 2, 3 and 11. Probably, one of the reasons for this folding is the difference in altitude between site 1 and the other sites. The posterior mean of a_1 is 0.012. Since $2a_1$ times the variance of Y at each location represents the nugget effect in the variogram, this agrees with Sampson and Guttorp's finding of a nugget effect in the range 2–4%, which they say is consistent with the known magnitude of the measurement error in these data.

To demonstrate the ability to predict correlations with ungauged locations and also to provide a kind of cross-validation check on our model, we ran the chain for 11 sites and dropped site 6

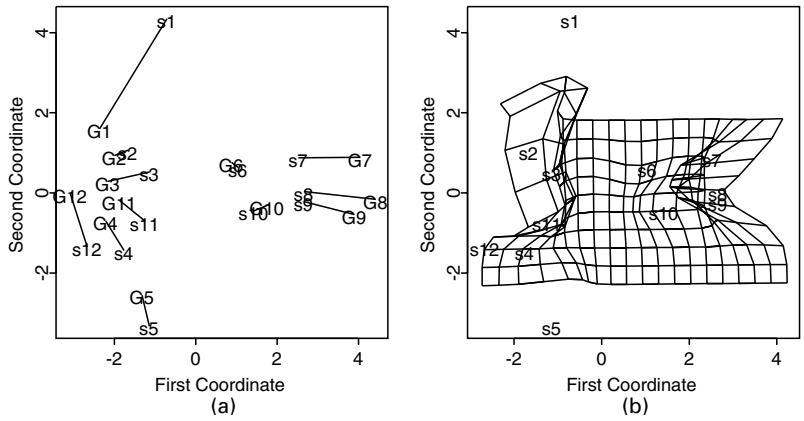

Fig. 6. (a) Procrustes superimposition of the posterior mean of the locations in D-space (s_i) onto the original geographical locations in G-space (G_i) and (b) mapping of a rectangular grid of 200 points from G-space into D-space

Table 1. Prediction of the covariances between site 6 and the others when site 6 was dropped from the solar radiation data set

Site	*Results for the following quantiles:*			*Observed*
	25%	*50%*	*75%*	
1	44.70	53.11	63.07	54.57
2	42.49	49.91	59.71	52.63
3	41.73	48.84	58.35	52.44
4	35.44	41.35	49.40	44.34
5	32.79	38.54	45.87	41.55
6	39.37	53.34	74.81	56.10
7	43.04	50.50	60.03	54.07
8	40.12	47.02	56.14	50.24
9	40.67	47.56	56.78	51.07
10	41.35	48.48	57.63	52.40
11	39.11	45.72	54.44	49.30
12	36.47	42.76	51.05	46.12

from the analysis, and then predicted the augmented covariance matrix following the algorithm shown at the end of Section 3.2. The posterior distribution that was obtained was similar to the case when all the $n = 12$ gauged sites were considered. Table 1 presents the predicted and the observed covariances between site 6 and the remaining 11 sites. The uncertainty in predicting the variance of site 6 is quite high, because the posterior distribution of the variance conditioned on τ^2 is the same as its prior distribution, as shown in Section 3.2. Observe that there are many other covariances which are well estimated by the respective posterior median.

5. Conclusions

We have presented a Bayesian model for the spatial deformation approach that was introduced by Sampson and Guttorp (1992). Observations comprise a spatial covariance matrix obtained

over T time points at n gauged locations. The idea is to use the observed correlations between gauged sites to model the spatial correlation structure of the process of interest. The main feature of the model is that it does not require the usual assumptions of stationarity and isotropy of the spatial covariance structure. As in Sampson and Guttorp (1992), we map the monitoring sites from their geographical locations G to a new space D, wherein isotropy holds. This was done by assigning a GP prior to the function that maps the locations from G-space into D-space. Unlike Sampson and Guttorp (1992), we formulate the whole inference in a single framework, i.e. through the posterior distribution of the parameters involved in the model. Inferences are computed by Markov chain Monte Carlo sampling, and Section 3.1 presents some key features of the algorithm that were needed to improve mixing in this complex application. Prediction of the process $Y(\mathbf{x}, t)$, with its associated uncertainty, at an ungauged location $\mathbf{x}$, is straightforward.

In Section 4 we applied the model to two different data sets, first to a simulated data set with only $n = 6$ gauged sites and $K = 2$ and then to a real data set analysed by Sampson and Guttorp. The model seems to work well in modelling the anisotropy in the data in both applications.

The following natural extensions of this model will form the basis of future research. Here we have not allowed for a prior spatial structure of the variances of the process. Instead we modelled them as exchangeable, *a priori*. One suggestion is to model the variances as a log-normal spatial process to ensure positivity. It is also known that a more general class of models includes mappings from $\mathbb{R}^2 \to \mathbb{R}^p$, $p = 2, \ldots, n-1$. Furthermore, when $p > 2$ we can include covariate information by using covariates as the other axes of the D-space. Finally a natural step towards the generalization of this model would be the joint modelling of space and time.

Acknowledgements

This research benefited from discussions with Professor Jim Zidek and colleagues. Alexandra M. Schmidt is grateful to the Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil, for financial support during her doctoral studies. The authors thank Professor Paul Sampson for making available the solar radiation data set. We are also grateful to Dr Huiling Le for proving the result in Appendix A and for some useful discussions about shape analysis. The authors thank two referees and the Joint Editor for valuable comments.

Appendix A: Information about σ_d^2 in the likelihood

Here we prove that $\mathbf{S}$, the sample covariance matrix, brings information at most about the eigenvalues of $\boldsymbol{\sigma}_d^2$. First suppose that we could learn the value of $\boldsymbol{\Sigma}$.

Write $\boldsymbol{\sigma}_d^2 = T\Lambda T'$, where $T \in \mathrm{SO}(2)$, the set of all 2×2 rotation matrices, which is known as the special orthogonal group, and Λ is the diagonal matrix whose diagonal elements are the eigenvalues of $\boldsymbol{\sigma}_d^2$. And consider $\tilde{\mathbf{D}} \sim N_{2n}(\mathbf{m}, \Lambda \otimes \mathbf{R}_d)$. If we rotate $\tilde{\mathbf{D}}$ by T and since $\boldsymbol{\sigma}_d^2 \otimes \mathbf{R}_d = \mathrm{diag}(T, \ldots, T)(\Lambda \otimes \mathbf{R}_d)\ \mathrm{diag}(T', \ldots, T')$, we have

$$T\tilde{\mathbf{D}} \sim N_{np}(T\mathbf{m}, \boldsymbol{\sigma}_d^2 \otimes \mathbf{R}_d).$$

However, $T\tilde{\mathbf{D}}$ differs from $\mathbf{D}$ only by a fixed rotation and hence the shape of $\tilde{\mathbf{D}}$, which is the shape of $T\tilde{\mathbf{D}}$, has the same distribution as that of the shape of $\hat{\mathbf{D}}$, where

$$\hat{\mathbf{D}} \sim N_{np}(T\mathbf{m}, \boldsymbol{\sigma}_d^2 \otimes \mathbf{R}_d).$$

However, as $T\mathbf{m}$ has the same shape as $\mathbf{m}$, the shape of $\hat{\mathbf{D}}$ has the same distribution as that of the shape of $\mathbf{D}$. Thus, the shape of $\tilde{\mathbf{D}}$ has the same distribution as that of the shape of $\mathbf{D}$. So, the information on $\boldsymbol{\Sigma}$, which gives the information on the size and shape of $\mathbf{D}$, will at most give information on the eigenvalues of $\boldsymbol{\sigma}_d^2$. Since even knowing $\boldsymbol{\Sigma}$ would only provide information about the eigenvalues of $\boldsymbol{\sigma}_d^2$, observing $\mathbf{S}$ also gives information at most about these eigenvalues.

References

Cressie, N. A. C. (1993) *Statistics for Spatial Data*, revised edn. New York: Wiley.

Damian, D., Sampson, P. D. and Guttorp, P. (2001) Bayesian estimation of semiparametric non-stationary spatial covariance structures. *Environmetrics*, **12**, 161–178.

Dawid, A. P. (1979) Conditional independence in statistical theory (with discussion). *J. R. Statist. Soc.* B, **41**, 1–31.

Dryden, I. L. and Mardia, K. V. (1998) *Statistical Shape Analysis*. Chichester: Wiley.

Gamerman, D. (1997) *Markov Chain Monte Carlo—Stochastic Simulation for Bayesian Inference*. London: Chapman and Hall.

Gilks, W. R. and Wild, P. (1992) Adaptive rejection sampling for Gibbs sampling. *Appl. Statist.*, **41**, 337–348.

Guttorp, P., Meiring, W. and Sampson, P. D. (1994) A space-time analysis of ground-level ozone data. *Environmetrics*, **5**, 241–254.

Guttorp, P., Sampson, P. D. and Newman, K. (1992) Nonparametric estimation of spatial covariance with application to monitoring network evaluation. In *Statistics in the Environmental & Earth Sciences* (eds A. Walden and P. Guttorp), pp. 39–51. London: Arnold.

Higdon, D., Swall, J. and Kern, J. (1999) Non-stationary spatial modeling. In *Bayesian Statistics 6* (eds J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), pp. 761–768. Oxford: Clarendon.

Iovleff, G. and Perrin, O. (1999) Estimating non-stationary spatial structure using simulated annealing. *Technical Report*. Groupe de Recherche en Economie Mathematique et Quantitative, University of Toulouse I, Toulouse.

Le, N. D. and Zidek, J. V. (1992) Interpolation with uncertain spatial covariances: a Bayesian alternative to kriging. *J. Multiv. Anal.*, **43**, 351–374.

Loader, C. and Switzer, P. (1992) Spatial covariance estimation for monitoring data. In *Statistics in the Environmental & Earth Sciences* (eds A. Walden and P. Guttorp), pp. 52–70.

Mardia, K. V. and Goodall, C. R. (1993) Spatial-temporal analysis of multivariate environmental monitoring data. In *Multivariate Environmental Statistics* (eds G. P. Patil and C. R. Rao), pp. 347–386. Amsterdam: Elsevier.

Monestiez, P. and Switzer, P. (1991) Semiparametric estimation of nonstationary spatial covariance models by metric multidimensional scaling. *Technical Report*. Department of Statistics, Stanford University, Stanford.

O'Hagan, A. (1994) *Kendall's Advanced Theory of Statistics*, vol. 2B, *Bayesian Inference*. Chichester: Wiley.

Sampson, P. D. and Guttorp P. (1992) Nonparametric estimation of nonstationary spatial covariance structure. *J. Am. Statist. Ass.*, **87**, 108–119.

Schmidt, A. M. (2001) Bayesian spatial interpolation of environmental monitoring stations. *PhD Thesis*. Department of Probability and Statistics, University of Sheffield, Sheffield.

Smith, R. L. (1996) Estimating nonstationary spatial correlations. *Technical Report*. Cambridge University, Cambridge.

Sun, W., Le, N. D., Zidek, J. V. and Burnett, R. (1998) Assessment of a Bayesian multivariate interpolation approach for health impact studies. *Environmetrics*, **9**, 565–586.