Estimation of non-stationary spatial multivariate covariance structure

David J. Nott\(^1\) and William T.M. Dunsmuir\(^2\)

\(^1\) Department of Statistics, University of New South Wales, Sydney, NSW 2052, AUSTRALIA. (djn@maths.unsw.edu.au)

\(^2\) Division of Biostatistics, School of Public Health, A460, Mayo Building, 420 Delaware St SE, Minneapolis, MN 55455. (dunsmuir@biostat.umn.edu)

1 Introduction

For spatio-temporal modelling in meteorological and environmental applications assessment of spatial covariance structure from space-time data is an important task. In many problems the assumption of isotropic spatial covariance is often unrealistic and may not be necessary where we have spatial measurements replicated in time.

Modelling spatially non-stationary covariation behaviour for space-time data is a problem receiving an increasing amount of attention - see the reviews of Guttorp and Sampson (1994) and Sampson, Damian and Guttorp (2001). Existing methods include those based on empirical orthogonal functions, Bayesian hierarchical models, kernel smoothing and spatial deformation. We briefly review these here and refer the reader to Sampson, Damian and Guttorp (2001) for further details.

The empirical orthogonal functions method is described in Obled and Creutin (1986) but this method can have some problems in implementation and may not lead to parsimonious and interpretable representations of the observed covariance structure. See Guttorp and Sampson (1994) for discussion. Other related methods use basis function expansions to define non-stationary processes (Nychka, Royle and Wikle, 1999). Higdon (1988) develops a process convolution approach. Brown, Le and Zidek (1993) discuss a hierarchical Bayesian method of multivariate spatial interpolation which is able to incorporate information from spatial covariates. Loader and Switzer (1992) discuss an empirical Bayes technique for estimating a spatial covariance matrix at a collection of observing sites and suggest an extension of their estimate to locations where data are not available by adding sites sequentially. However the resulting estimated covariance matrix may not be invariant to the order in which the sites are considered and there may also be problems of positive definiteness. Oehlert (1993) introduces a kernel method but it can result in covariance estimates which are not necessarily non-negative definite.

Spatial deformation methods for assessing non-stationary spatial covariance structure are based on estimating a change of co-ordinates which will make the covariance function isotropic. This was proposed by Sampson (1986), and developed in Sampson and Guttorp (1992). A discussion of computational details is given by Meiring, Guttorp and Sampson (1997), and alternative approaches to implementation of the method are reviewed by
Sampson, Damian and Guttorp (2001). Recent attention has concentrated on a Bayesian implementation of this approach (Schmidt and O’Hagan, 2000, Damian, Sampson and Guttorp, 2001).

In this paper we discuss a method for estimating non-stationary spatial covariance structure developed in Nott and Dunsmuir (2001) and Nott et al (2000) that preserves a given spatial covariance matrix at observing stations and uses a stationary process to describe conditional behaviour given observing site values. A variant of the technique giving a localized description of the spatial covariance structure reproduces the covariance at observing sites while giving a description of conditional behaviour through a collection of stationary processes. The method is computationally attractive, produces a valid non-negative definite non-stationary covariance function and is readily extended to the assessment of covariance for multivariate processes. The technique is illustrated for data describing the two orthogonal components of Sydney winds. The method may be useful in applications to spatial prediction, network design and as a tool for visualization of spatial covariance structure.

2 The Method and its Application

2.1 Notation and description of the problem

Let

\[ Z = \{Z(s, w), s \in \mathbb{R}^d, w \in \mathbb{R}\} \]

be a space-time process (where \( s \) indexes space and \( w \) indexes time) and denote the spatial covariance function of \( Z \) as

\[ R_Z(s, t) = \text{Cov}(Z(s, w), Z(t, w)) \]

which does not depend on \( w \) if \( Z \) is temporally stationary. For time \( w_k \) let \( Z_k = (Z(s_1, w_k), ..., Z(s_n, w_k))^T, 1 \leq k \leq M \), and write \( z_k \) for the corresponding realization of \( Z_k \). Let \( \Gamma \) be the covariance matrix of \( z_k \). We requires throughout that the spatial realisations at different times are identically distributed, and that we can average over time to estimate spatial covariances. Our objective is to obtain an estimate of \( R_Z(s, t) \) which preserves the values \( \Gamma \) or non-negative definite estimates of it. For example, when there are no missing data the simple estimate

\[ \hat{\Gamma} = \frac{1}{M} \sum_{i=1}^{M} (z_i - \bar{z})(z_i - \bar{z})^T, \]

could be used. Other modifications to \( \hat{\Gamma} \) required to handle missing data and to incorporate shrinkage to a parametric model are described below.

2.2 Representation of a stationary spatial process

To describe our method consider a purely stationary, Gaussian random field \( \{W(s), s \in \mathbb{R}^d\} \) with zero mean and covariance function \( R(h), h \in \mathbb{R}^d \). For a designated collection of sites \( s_1, ..., s_n \), let \( W = (W(s_1), ..., W(s_n))^T \), let \( C \) be the covariance matrix of \( W \) and let
where $c(s) = (R(s - s_1), ... R(s - s_n))^T$ be the vector of cross-covariances between $W(s)$ and $W$. Note that

$$W(s) = c(s)^T C^{-1} W + \delta(s)$$

(1)

where $\delta(s)$ is a zero mean non-stationary random Gaussian field independent of $W$ with covariance function

$$R_\delta(s, t) = R(t - s) - c(s)^T C^{-1} c(t).$$

The function $c(s)^T C^{-1} w$ is the conditional mean of $W(s)$ given $W = w$ and, since the process $\delta(s)$ describes the residual variation in $W(s)$ after removing the conditional mean given $W$, it is zero at the sites $s_1, ..., s_n$.

We modify this representation to construct a non-stationary process with a spatial covariance function which we can use as a model for the non-stationary covariance function $R_Z(s, t)$ of $Z(s, w)$.

### 2.3 Construction of a non-stationary spatial process

Suppose that $J = \{s_1, ..., s_n\}$ represents a collection of monitored sites for the process $Z(s, w)$ and as before write $\Gamma$ for the spatial covariance matrix at sites in $J$. We construct a non-stationary spatial process which reproduces the covariance matrix $\Gamma$ at sites in $J$ by modifying the representation (1) by writing

$$W^*(s) = c(s)^T C^{-1} W^* + \delta(s)$$

where $W^* \sim N(0, \Gamma)$ independent of $\delta(s)$ and where $c(s)$ and $C$ are defined as before. This is a non-stationary process with covariance function

$$R_{W^*}(s, t) = R(t - s) + c(s)^T C^{-1}(\Gamma - C)C^{-1} c(t).$$

(2)

It is easy to check that for $(s, t) = (s_i, s_j)$ the value of the covariance function (2) is $\Gamma_{ij}$ since $c(s_i)$ and $c(s_j)$ are the $i$th and $j$th columns of $C$.

The function (2) is a valid non-negative definite covariance function, which reproduces the covariance matrix $\Gamma$ regardless of what stationary covariance function $R(h)$ is used in the construction. Note that if $s$ and $t$ are far away from monitored sites (2) is approximately $R(t - s)$ if $R(h) \rightarrow 0$ as $\|h\| \rightarrow \infty$.

### 2.4 Estimation

Let $\hat{\Gamma}$ be a non-negative definite estimate of $\Gamma$ based on temporally stationary observations on the field $Z$ at sites in $J$. We suggest fitting a parametric stationary covariance model $R(h) = R(h; \theta) \in \mathbb{R}^p$ to $\hat{\Gamma}$ to obtain an approximating stationary model and estimates $\hat{R}(t - s)$, $\hat{c}(s)$ and $\hat{C}$ for $R(t - s)$, $c(s)$ and $C$. We then estimate the non-stationary spatial covariance function $R_Z(s, t)$ of $Z(s, t)$ by

$$\hat{R}_Z(t - s) = \hat{R}(t - s) - \hat{c}(s)^T \hat{C}^{-1}(\hat{\Gamma} - \hat{C})\hat{C}^{-1} \hat{c}(t).$$

(3)

The approach used in fitting the stationary model is as follows. First, the empirical spatial dispersion matrix $D$ is standardized, so that $D$ corresponds to a variance one field. If the variance field is of interest, then it can be estimated separately. In practice we have found that if the variance nonstationarity is severe, this can have an undesirable effect on estimation of parameters describing the correlation structure. Write $\gamma(h; \theta) =$
for the spatial dispersion function of the stationary process used in our construction. See, for instance, Cressie (1993, pp. 85–86) for a list of possible parametric forms which could be adopted for the dispersion function. Various methods, such as ordinary or generalized least squares or maximum likelihood, could be used to estimate $\theta$.

Once a covariance function for $W$ has been determined, computation of (3) is straightforward. The computational cost in evaluating the covariance function is dominated by the need to solve two linear systems in $\hat{C}$. If (3) is to be evaluated for a large number of pairs of sites, which is usually the case, then an initial Cholesky decomposition of $\hat{C}$ could be obtained which enables the required solution of the linear systems to be done efficiently. The initial decomposition is feasible for up to several hundred sites, which makes the method computationally attractive relative to alternative approaches.

2.5 Connection with Loader and Switzer’s method.

The method of Loader and Switzer (1992) is used for estimating the covariance matrix at observing sites in the application described below. Loader and Switzer assume that the spatial replicates at different times are independent identically distributed normal random vectors with a known mean which is taken without loss of generality to be zero. Writing $\hat{\Gamma}$ for the empirical covariance matrix in this case,

$$\hat{\Gamma} = \frac{1}{M} \sum_{i=1}^{M} z_i z_i^T,$$

Loader and Switzer estimate the true spatial covariance matrix $\Gamma$ by

$$\hat{\Gamma}_{LS} = \lambda \hat{\Gamma} + (1 - \lambda)\hat{C},$$

where $0 < \lambda < 1$ is a shrinkage parameter, and $\hat{C}$ is a stationary covariance matrix obtained by fitting some parametric covariance function model. The shrinkage parameter $\lambda$ is chosen by an empirical Bayes argument: an inverse Wishart prior distribution is assumed for $\Gamma$, $\Gamma \sim w^{-1}((m - n - 1)\hat{C}, m)$. With this prior it can be shown that the posterior mean of $\Gamma$ is

$$\lambda \hat{\Gamma} + (1 - \lambda)\hat{C}$$

where $\lambda = M/(M + m - n - 1)$. The parameter $m$ can be estimated by maximum likelihood from which $\lambda$ is obtained. See Loader and Switzer (1992) for further details.

Note that if our method is applied with $\hat{\Gamma}_{LS}$ then our estimate of the cross-covariances between sites in $J$ and sites not in $J$ agrees with the proposal in Loader and Switzer. However when multiple sites are added sequentially using their suggestion the estimated covariance between pairs of sites not in $J$ may not be invariant to the order in which those sites are added.

2.6 Application to Sydney Wind Fields.

We illustrate the methods described above on some data on the 3pm east-west component of Sydney wind patterns during the 61 days of September and October. Our data consist of measurements of wind at a network of 45 monitoring stations in the Sydney...
area. Additional motivation for this analysis and a more complete description of the data network is provided in Nott and Dunsmuir (2001).

Since there are missing observations at some of the stations we obtained a positive definite estimate of the spatial covariance matrix using the EM algorithm (Little and Rubin, 1987) with an initial estimate of the covariance obtained by filling in missing values with site means. We then applied the technique of Loader and Switzer (1992) to shrink this raw estimate to a fitted stationary covariance matrix. We fitted an exponential covariance function model by estimating the variance as the average of the empirical variances across sites, giving an estimated variance of 58.04, and estimated the spatial dependence parameter $\theta$ by an ordinary least squares fit to the correlation matrix, giving an estimated $\theta$ of 1.115. After fitting the stationary model, the shrinkage parameter for Loader and Switzer’s technique was computed as $\lambda = 0.88$. The final covariance matrix was then standardized to obtain a correlation matrix.

The contour plots of Figure 1 show points which have equal estimated correlation with two different fixed sites. One station was chosen to be on the coast, location 33.85°S, 151.22°E, and the other to be in the western half of the region under study, location 33.74°S, 149.88°E. The contour plots for the coastal station show some anisotropy that is thought to reflect sea breeze effects, with stations the same distance from the coast influenced to a similar extent. For the western reference station, correlation seems to decay more slowly with distance and the shape of contours of high correlation can be interpreted in terms of the local topography with locations south of the reference site having a similar elevation.

A comparison of the above method with the Sampson-Guttorp approach is given in Nott and Dunsmuir (2001). Even after exclusion a number of sites which were spatially anomalous with respect to their neighbours (these sites appeared to have a large influence on the fitted spatial deformation) we experienced problems with “folding” of the deformation, and only a model which is very nearly stationary avoids the problem.

3 Extensions to the Method.

3.1 Spatially Adaptive Modification.

The method described above is appealing as a simple way of obtaining a valid non-negative definite covariance function respecting a given covariance matrix. We anticipate that it may be useful for visualizing covariance structure for large data sets. Of concern, however, is the lack of spatial adaptation in the description of the conditional behaviour in obtaining (3). It can easily be seen that simple kriging based on data at the observing sites and on the covariance estimate (3) is equivalent to simple kriging based on the stationary model used in constructing (3). Hence the use of the above non-stationary covariance estimate will not result in improved performance in simple statistical spatial prediction algorithms.

In Nott and Dunsmuir (2001) we consider an extension of our approach in which a collection of stationary processes is used to describe conditional behaviour. The extension is similar in spirit to a method of spatial prediction called moving window kriging due to Haas (1998) in which covariance non-stationarity is modelled implicitly by local fitting of stationary models in windows centered on prediction locations. When applied to a spatial covariance matrix, our approach automatically provides a valid spatial covariance function, so that modification of any estimated covariance matrices is not required (as it
3.2 Extension to Multivariate Processes.

Extension of our methods to assessment of covariance for multivariate processes is formally straightforward. Consider a multivariate spatio-temporal process

\[ Z = \{Z(s, w); s \in \mathbb{R}^d, w \in \mathbb{R}\} \]

where now \( Z(s, w) = (Z^1(s, w), \ldots, Z^K(s, w))^T \). As before, \( Z \) is assumed to be temporally ergodic. Define the \( K \times K \) matrix valued spatial covariance function of \( Z \) as

\[ R_Z(s, t) = [\text{Cov}(Z^i(s, w), Z^j(t, w))] \]

which does not depend on \( w \) by assumption. In a similar way to before we can estimate spatial covariances and cross-covariances among the components of \( Z \) at the monitored sites by averaging over time. Write \( \hat{\Gamma} \) for the estimated covariance matrix of

\[ (Z^1(s_1, w), \ldots, Z^K(s_1, w), \ldots, Z^1(s_n, w), \ldots, Z^K(s_n, w))^T. \]

We can describe the spatial covariance function of \( Z \) by honouring \( \hat{\Gamma} \) and describing conditional behaviour given monitored site values by a multivariate stationary process. Let \( \{W(s), s \in \mathbb{R}^d\} \) be a multivariate stationary process with \( K \) components, \( W(s) = (W^1(s), \ldots, W^K(s))^T \). Let \( R(h) \) be the \( K \times K \) matrix valued covariance function of \( W(s) \),

\[ R(h) = [\text{Cov}(W^i(s), W^j(s + h))]. \]

Let \( C \) be the covariance matrix of

\[ W = (W^1(s_1), \ldots, W^K(s_1), \ldots, W^1(s_n), \ldots, W^K(s_n))^T \]

and let \( C(s) \) be the \( nK \times K \) cross-covariance matrix between \( W \) and \( W(s) \). Then by similar reasoning to before we can show that a non-negative definite estimate of the spatial covariance \( R_Z(s, t) \) of \( Z \) honouring \( \hat{\Gamma} \) is

\[ \hat{R}(t - s) - \hat{C}(s)^T \hat{C}^{-1}(\hat{\Gamma} - \hat{C})\hat{C}^{-1}\hat{C}(t) \]

where \( \hat{R}(t - s), \hat{C}(s) \) and \( \hat{C} \) are obtained by fitting some appropriate parametric stationary model. The estimate above can be generalized in a similar way to before by using a collection of multivariate processes to describe conditional behaviour. Details and applications to the bivariate rectangular components of Sydney Harbour wind fields are given in Nott et al (2000).

References


Figure 1: Contours of equal estimated correlation with two different fixed sites. Location 33.85°S, 151.22°E (top) and location 33.74°S, 149.88°E (bottom). The marked sites show the locations of the monitored sites.