J. R. Statist. Soc. B (2005)
67, *Part* 5, *pp.* 667–687

Statistical methods for regular monitoring data

Michael L. Stein

University of Chicago, USA

[Received September 2004. Final revision May 2005]

Summary. Meteorological and environmental data that are collected at regular time intervals on a fixed monitoring network can be usefully studied combining ideas from multiple time series and spatial statistics, particularly when there are little or no missing data. This work investigates methods for modelling such data and ways of approximating the associated likelihood functions. Models for processes on the sphere crossed with time are emphasized, especially models that are not fully symmetric in space–time. Two approaches to obtaining such models are described. The first is to consider a rotated version of fully symmetric models for which we have explicit expressions for the covariance function. The second is based on a representation of space–time covariance functions that is spectral in just the time domain and is shown to lead to natural partially nonparametric asymmetric models on the sphere crossed with time. Various models are applied to a data set of daily winds at 11 sites in Ireland over 18 years. Spectral and space–time domain diagnostic procedures are used to assess the quality of the fits. The spectral-in-time modelling approach is shown to yield a good fit to many properties of the data and can be applied in a routine fashion relative to finding elaborate parametric models that describe the space–time dependences of the data about as well.

Keywords: Coherence spectra; Multiple time series; Phase spectra; Random processes on the sphere; Space–time asymmetry; Spectral analysis

1. Introduction

It is common for meteorological and environmental data to be collected at regular intervals at fixed monitoring sites. Specifically, suppose that a process is observed at $(\mathbf{x}_j, t)$ for locations $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and times $t = 1, \ldots, T$. We shall call such data regular monitoring data and focus on this setting here. Although environmental data sets nearly always have missing observations, meteorological data sets are occasionally complete and, much more frequently, the fraction of missing observations is quite low. Thus, it is worthwhile to develop methods for the complete-data case with the idea that these can be combined with missing data methods (the EM algorithm, multiple imputations or Markov chain Monte Carlo methods) when the fraction of missing observations is not too high.

Regular monitoring data can be viewed as multiple time series, so it is tempting to want to use methods from this literature (Brillinger, 1981; Priestley, 1981; Reinsel, 1997) to analyse such data. However, the time series perspective does not readily deal with the problem of prediction at sites without a monitor, which is often of interest in meteorology and environmental monitoring. Furthermore, even if we were only interested in the process at the monitoring sites, when there are more than a few of them, we would generally want to make use of their spatial locations to model the dependences between the components of the time series, so that a purely time series perspective may not be best. Bennett (1979), Bras and Rodríguez-Iturbe (1985), Kyriakidis and

Address for correspondence: Michael L. Stein, Department of Statistics, University of Chicago, Chicago, IL 60637, USA.
E-mail: stein@galton.uchicago.edu

 1369–7412/05/67667

Journel (1999), de Luna and Genton (2005) and Sahu and Mardia (2005) considered various approaches to making use of spatial information in adapting multiple time series models to the space–time setting. Section 2 of this paper describes an approach to covariance modelling that is closely related to that of Cressie and Huang (1999) that is spectral in time but not in space and leads to models that naturally take account of spatial locations while automatically yielding valid covariance functions over all of space–time. The model is described in terms of a single spatial covariance function and three spectra: one describing the process at any single site, another describing the coherence spectra at two sites as a function of distance between them and a third describing the phase spectra between sites. The models are partially nonparametric in that the three spectra are not constrained to finite dimensional spaces but are still parsimonious relative to an unstructured multiple time series approach in which the spatial locations of the monitoring sites are ignored. For regular monitoring data, the resulting models can be naturally estimated in a relatively routine fashion by using series expansions for the three unknown spectra and a spectral approximation to the likelihood.

A common feature in space–time processes is a lack of what Gneiting (2002) called full symmetry in the covariance structure. Specifically, a space–time process Z has fully symmetric covariance structure if

$$\operatorname{cov}\{Z(\mathbf{x}, t), Z(\mathbf{y}, s)\} = \operatorname{cov}\{Z(\mathbf{x}, s), Z(\mathbf{y}, t)\}$$

for all $\mathbf{x}$, $\mathbf{y}$, s and t. The spectral-in-time approach provides a simple and interpretable way of capturing space–time asymmetries in covariance functions.

A further consideration which we may want to take into account when modelling a process on the Earth is the spherical nature of its surface. Specifically, it would be desirable to use models that are guaranteed to be valid over the whole sphere and all of time. Satisfying this constraint as well as allowing for space–time asymmetry is essential to developing space–time models that are suitable for all or a substantial fraction of the Earth. Section 3 describes two approaches to this problem. The first is to consider a rotation about the Earth's axis of a fully symmetric process and the second is to extend the spectral-in-time approach in Section 2 to processes on the sphere.

When regular monitoring data can be modelled as stationary in time, the resulting covariance matrix for the observations has what is known as a block Toeplitz structure, which simplifies the matrix computations that are needed for statistical analyses. For example, if the process is Gaussian, using the Levinson–Wiggins–Whittle–Robinson algorithm (Caines, 1988) or methods for computing Cholesky decompositions of block Toeplitz matrices, the exact likelihood function can be found in $O(n^3T^2)$ floating point operations rather than the $O(n^3T^3)$ operations that are needed if this structure is ignored. However, for sufficiently large T, this computation will still be infeasible and Section 4 describes a fast time domain approximation to the likelihood that is based on approximating the conditional density of each day given the past by conditioning on a fixed number of past days.

Section 5 presents several analyses of the Irish wind data which were first studied in Haslett and Raftery (1989). Two fairly elaborate parametric models (each with nine parameters) are fitted to these data by maximizing an approximate likelihood and the results are compared with that which is obtained by the spectral-in-time approach. The spectral-in-time approach is shown to provide a slightly superior fit to the data in terms of likelihood, agreement with empirical variograms and other diagnostic plots. Thus, this approach may be useful as a general tool for analysing regular monitoring data, although a lack of an explicit form for the space–time covariance function may be problematic when the goal is to predict the process at a large number of unobserved locations.

2. Models spectral in time

When we have long records of regular monitoring data at a modest number of scattered locations, it is convenient to consider models in the frequency domain in time but in the spatial domain in space. Specifically, instead of modelling the space–time covariance function $K(\mathbf{x}, t)$ directly, model the function $h(\omega, \mathbf{x})$ for which

$$K(\mathbf{x}, t) = \int_{\mathbb{R}} h(\omega, \mathbf{x}) \exp(\mathrm{i}t\omega)\, \mathrm{d}\omega.$$

A natural way to generate functions h for which K is a valid covariance function is to switch the roles of space and time in the approach of Cressie and Huang (1999) and to consider

$$h(\omega; \mathbf{x}) = S(\omega)\, \rho(\omega; \mathbf{x}),$$

where $S(\omega)$ is an even non-negative integrable function on $\mathbb{R}$ and, for each ω, $\rho(\omega; \mathbf{x})$ is a (possibly complex-valued) correlation function on $\mathbb{R}^d$. More specifically, the following form for ρ may prove useful in practice:

$$\rho(\omega; \mathbf{x}) = C\{|\mathbf{x}|\, \gamma(\omega)\} \exp\{\mathrm{i}\, \theta(\omega) \mathbf{u}'\mathbf{x}\} \tag{1}$$

with C a real-valued isotropic autocorrelation function on $\mathbb{R}^d$, γ a non-negative even function on $\mathbb{R}$, θ an odd function on $\mathbb{R}$ and $\mathbf{u} \in \mathbb{R}^d$ a unit vector. The following is nearly a special case of the generalized theorem 1 in Gneiting (2002), but here we do not require K to be continuous.

Theorem 1. For C a measurable real-valued isotropic autocorrelation function on $\mathbb{R}^d$, γ a non-negative even measurable function on $\mathbb{R}$, θ an odd measurable function on $\mathbb{R}$ and S an even non-negative integrable function on $\mathbb{R}$, the function

$$K(\mathbf{x}, t) = \int_{\mathbb{R}} S(\omega)\, C\{|\mathbf{x}|\, \gamma(\omega)\} \exp\{\mathrm{i}\, \theta(\omega) \mathbf{u}'\mathbf{x} + \mathrm{i}t\omega\}\, \mathrm{d}\omega \tag{2}$$

is a real-valued positive definite function on $\mathbb{R}^d \times \mathbb{R}$.

Proof. To prove that K is positive definite, it suffices to show that, for any set of locations $\mathbf{x}_1, \ldots, \mathbf{x}_N$ and any finite N, the matrix-valued function

$$\Phi(\omega) = (S(\omega)\, C\{|\mathbf{x}_j - \mathbf{x}_k| \gamma(\omega)\} \exp\{\mathrm{i}\, \theta(\omega)\, \mathbf{u}'(\mathbf{x}_j - \mathbf{x}_k)\})_{j,k=1}^N$$

is non-negative definite for all ω and is componentwise integrable over $\mathbb{R}$. Since S is integrable and $C\{|\mathbf{x}| \gamma(\omega)\} \exp\{\mathrm{i}\theta(\omega)\mathbf{u}'\mathbf{x}\}$ is bounded and measurable for all $\mathbf{x}$, the integrability of Φ is immediate. Furthermore, $\Phi(\omega)$ is a non-negative definite matrix for all ω, which follows from S non-negative, C positive definite and

$$(\exp\{\mathrm{i}\, \theta(\omega)\mathbf{u}'(\mathbf{x}_j - \mathbf{x}_k)\})_{j,k=1}^N = (\exp\{\mathrm{i}\, \theta(\omega)\mathbf{u}'\mathbf{x}_j\})_{j=1}^N ((\exp\{-\mathrm{i}\, \theta(\omega)\mathbf{u}'\mathbf{x}_j\})_{j=1}^N)'$$

non-negative definite, since a Hadamard (componentwise) product of non-negative definite matrices is non-negative definite. Finally, K real follows from the fact that $\Phi(-\omega) = \overline{\Phi(\omega)}$, where the complex conjugate is taken componentwise.

The functions S, C and θ all have natural interpretations. First, S is the temporal spectrum for the process at any one location. Furthermore, if C is non-negative, then $C\{|\mathbf{x}| \gamma(\omega)\}$ is the coherence at frequency ω between the time series at two locations separated by $\mathbf{x}$ and $\theta(\omega)\mathbf{u}'\mathbf{x}$ is the phase (Priestley, 1981). One potentially undesirable feature of equation (2) is that the purely

spatial covariance function is not generally isotropic unless $\theta(\omega) = 0$ for all ω, in which case the model is fully symmetric.

Stein (2005) considered a very simple translation-based approach to generating asymmetric models on $\mathbb{R}^d \times \mathbb{R}$. Specifically, if Z is a stationary process on $\mathbb{R}^d \times \mathbb{R}$ with covariance function $K(|\mathbf{x}|, t)$, then for any real ε and any unit vector $\mathbf{u}$ the process $Z(\mathbf{x} - \varepsilon t\mathbf{u}, t)$ has covariance function $K(|\mathbf{x} - \varepsilon t\mathbf{u}|, t)$. To see the relationship between this translation approach and equation (2), consider $\theta(\omega) = \varepsilon\omega$ in equation (2). For Z a stationary process on $\mathbb{R}^d \times \mathbb{R}$ with fully symmetric covariance function

$$K_0(\mathbf{x}, t) = \int_{\mathbb{R}} S(\omega)\, C\{|\mathbf{x}|\gamma(\omega)\} \exp(\mathrm{i}t\omega)\, \mathrm{d}\omega,$$

equation (2) yields $K(\mathbf{x}, t) = K_0(\mathbf{x}, t + \varepsilon\mathbf{u}'\mathbf{x})$, the covariance function of $Z(\mathbf{x}, t + \varepsilon\mathbf{u}'\mathbf{x})$. There is no obvious reason to prefer translating time via $Z(\mathbf{x}, t + \varepsilon\mathbf{u}'\mathbf{x})$ *versus* translating space via $Z(\mathbf{x} - \varepsilon t\mathbf{u}, t)$, but the fact that $\theta(\omega)$ need not be linear in equation (2) shows that theorem 1 allows a greater degree of flexibility in modelling asymmetries than translations of space or time.

The behaviour of $K(\mathbf{x}, 0)$ near $\mathbf{x} = \mathbf{0}$ determines the smoothness in space of the corresponding process (Stein, 1999). When C is continuous, its behaviour at the origin is generally closely related to the smoothness of $K(\mathbf{x}, 0)$ at $\mathbf{x} = \mathbf{0}$. For example, suppose that

$$C(r) = \sum_{j=0}^{m} \alpha_j r^{2j} + \beta r^{\nu} + o(r^{\nu})$$

as $r \downarrow 0$. If γ is bounded, then it is easy to show that

$$K(\mathbf{x}, 0) = \sum_{j=0}^{m} \alpha_j |\mathbf{x}|^{2j} \int_{\mathbb{R}} S(\omega)\, \gamma(\omega)^{2j} \cos\{\theta(\omega)\mathbf{u}'\mathbf{x}\}\, \mathrm{d}\omega$$
$$+ \beta|\mathbf{x}|^{\nu} \int_{\mathbb{R}} S(\omega)\, \gamma(\omega)^{\nu} \cos\{\theta(\omega)\mathbf{u}'\mathbf{x}\}\, \mathrm{d}\omega + o(|\mathbf{x}|^{\nu})$$

as $\mathbf{x} \to \mathbf{0}$. If θ is also bounded, then this implies that there is an analytic function A on $\mathbb{R}^d$ such that

$$K(\mathbf{x}, 0) = A(\mathbf{x}) + \beta|\mathbf{x}|^{\nu} \int_{\mathbb{R}} S(\omega)|\gamma(\omega)|^{\nu}\, \mathrm{d}\omega + o(|\mathbf{x}|^{\nu})$$

as $\mathbf{x} \to \mathbf{0}$. Thus, $K(\mathbf{x}, 0)$ has the same smoothness at the origin as C. If either γ or θ are unbounded, then $K(\mathbf{x}, 0)$ may be less smooth at the origin than C. Stein (2005) discusses the relevance of the smoothness of a space–time covariance function away from the origin to the modelling of processes in space–time. It is not easy in general to infer the smoothness of K away from the origin from the representation (2). However, when γ is constant and θ is identically 0, K is separable, in which case it is no smoother away from the origin than it is at the origin.

It is not necessary to assume that C in equation (2) is continuous. In particular, allowing C to have a jump at 0 makes it possible to include a spatial nugget effect (Gneiting, 2002), although the Irish wind data that are analysed in Section 5 appear to require a more elaborate form for the spatial nugget. In contrast with this spectral-in-time approach, a representation of $K(\mathbf{x}, t)$ that is spectral in both space and time will necessarily be continuous (if we exclude generalized functions).

A version of theorem 1 holds for processes in discrete time by changing the domain of ω from $\mathbb{R}$ to $(-\pi, \pi]$. This version of the theorem is especially useful in carrying out spectral analysis of regular monitoring data for which we naturally obtain estimates of the discrete time rather than the continuous time spectra.

It is worthwhile to compare equation (2) with an ordinary spectral analysis from the multiple time series perspective in which no use is made of the spatial information and no particular form is assumed for the spectra or cross-spectra of the process at the N monitoring sites. In equation (2), once the function C has been chosen, we need to specify only the three functions S, γ and θ. In contrast, an ordinary spectral analysis that made no use of the spatial information would require N univariate spectra and $\frac{1}{2}N(N-1)$ coherence and phase spectra. Furthermore, whereas equation (2) extends automatically to locations other than monitoring sites, the same cannot be said of an ordinary spectral analysis. Thus, using equation (2) as a model for space–time covariance functions in which C is specified parametrically and S, γ and θ are treated as smooth but unspecified functions of temporal frequency may provide a practically useful compromise between parametric and nonparametric approaches to space–time covariance modelling.

For processes in discrete time, a natural way to model the functions S, γ and θ on $(-\pi, \pi]$ is with trigonometric polynomials. For example, because γ must be even and non-negative, we could let

$$\log\{\gamma(\omega)\} = \sum_{k=0}^{K} a_k \cos(k\omega) \tag{3}$$

and, because θ must be odd,

$$\theta(\omega) = \sum_{k=1}^{K} b_k \sin(k\omega). \tag{4}$$

Now S must also be even and positive, so we could let

$$\log\{S(\omega)\} = \sum_{k=0}^{K} c_k \cos(k\omega),$$

but this may not be a good choice if the process exhibits long-range dependence in time, in which case we would want S to be unbounded in a neighbourhood of 0. In this case, it might be better to use

$$\log\{S(\omega)\} = -\beta \log(\sin|\tfrac{1}{2}\omega|) + \sum_{k=0}^{K} c_k \cos(k\omega), \tag{5}$$

where $\beta < 1$ ensures that S is integrable. We could of course replace $\Sigma_{k=0}^{K} c_k \cos(k\omega)$ in equation (5) by the log-spectral-density of an autoregressive AR(K) process or other standard time series model. For a given K, we can estimate the a_ks, b_ks, c_ks, β and any parameters of the function C by maximizing a spectral approximation to the likelihood. Specifically, writing $\mathbf{Z}_t$ for the vector of observations at time t, for $\omega_j = 2\pi j/T$ and

$$\mathbf{V}_j = (2\pi T)^{-1/2} \sum_{t=1}^{T} \mathbf{Z}_t \exp(\mathrm{i}\omega_j t),$$

assume that $\mathbf{V}_1, \ldots, \mathbf{V}_{\lfloor T/2 \rfloor}$ are independent complex multivariate normal random vectors with mean $\mathbf{0}$ and covariance matrix $\Phi(\omega_j)$. See Appendix A for further details. We could then choose K subjectively or by some model selection criterion such as the Akaike information criterion or Bayes information criterion. It is worth considering tapering and prewhitening as ways of improving the quality of approximation using this spectral approach, but we shall not pursue these possibilities here.

Although the spectral-in-time approach has several attractive features, the lack of an explicit expression for the covariance function may prove computationally challenging when we are interested in prediction at locations other than monitoring sites. The problem is that the calcu-

lation of the space–time covariances between any new site and the monitoring sites would in general require a numerical Fourier transform for every pair of an existing site and a prediction site. This may not be an obstacle if we must do these calculations for only a single fitted model, but it will become problematic if we are interested in a Bayesian predictive distribution at even a moderately large number of unmonitored sites, which would require calculating the space–time covariance function for many pairs of locations for many models. When modelling the discrete time spectra, a further disadvantage is that the resulting models do not extend easily to continuous time, unlike the intrinsically continuous time parametric models that are used in Section 5.

3. Models on spheres

Weber and Talkner (1993) noted the need for explicit classes of covariance functions on the sphere for applications in the atmospheric sciences. It is arguably at least as important to develop covariance function models for processes on the sphere that also vary in time. This section describes two simple ways of generating such models that include space–time asymmetry. Jun and Stein (2004) developed a more flexible approach to generating asymmetric models when the spatial domain is a sphere.

For a process on $\mathcal{S}_r^{d-1}$, the sphere of radius r in d dimensions, a natural requirement is for it to have an invariance property with respect to all rotations. A covariance function on the sphere is said to be homogeneous if it depends only on the angular distance between points on the sphere. Yadrenko (1983), Yaglom (1987), Weber and Talkner (1993), Gaspari and Cohn (1999) and Gneiting (1999) discussed properties and specific examples of homogeneous covariance functions on spheres. As noted by Yaglom (1987), page 389, an easy way to generate such homogeneous covariance functions is to consider the restriction of an isotropic covariance function on $\mathbb{R}^d$ to $\mathcal{S}_r^{d-1}$, i.e., if $K(|\mathbf{x}|)$ is the covariance function of the process Z on $\mathbb{R}^d$, then the restriction of Z to $\mathcal{S}_r^{d-1}$ is homogeneous. We consider only $d=3$ subsequently. Then, for $j=1,2$, if $\boldsymbol{\theta}_j=(L_j,l_j)$ are two points on $\mathcal{S}_r^2$ in terms of latitude and longitude and $A(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$ is the angle between them,

$$\operatorname{cov}\{Z(\boldsymbol{\theta}_1),Z(\boldsymbol{\theta}_2)\}=K[2r\sin\{A(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)/2\}].$$

It is not so clear what kinds of homogeneity for processes on $\mathcal{S}_r^2\times\mathbb{R}$ are most sensible. In particular, a process may be fundamentally different at different latitudes, so we may not want to assume homogeneity of a process with respect to latitude. In contrast, models that do not depend on the choice of time origin nor on what longitude becomes assigned 0 may be quite natural in many circumstances, so we shall restrict ourselves to models with this degree of homogeneity, i.e. we shall require that $\operatorname{cov}\{Z(L_1,l_1,s),Z(L_2,l_2,t)\}$ is of the form $Q(L_1,L_2,l_1-l_2,s-t)$ for some Q on $[-\frac{1}{2}\pi,\frac{1}{2}\pi]^2\times\mathbb{T}\times\mathbb{R}$, where $\mathbb{T}$ is the real line mod(2π).

In light of these considerations, define $\mathcal{D}$ to be the class of all real-valued functions Q on $[-\frac{1}{2}\pi,\frac{1}{2}\pi]^2\times\mathbb{T}\times\mathbb{R}$ that are symmetric and positive definite in the sense that $Q(L_1,L_2,l,t)=Q(L_2,L_1,-l,-t)$ and

$$\sum_{j,k=1}^{n}a_ja_kQ(L_j,L_k,l_j-l_k,t_j-t_k)\geqslant 0$$

for all real $a_1,\ldots,a_n$, $l_1,\ldots,l_n$ and $t_1,\ldots,t_n$, all $L_1,\ldots,L_n\in[-\frac{1}{2}\pi,\frac{1}{2}\pi]$ and all positive integers n. If $Q(L_1,L_2,l,t)=Q(L_1,L_2,l,-t)$ for all $(L_1,L_2,l,t)\in[-\frac{1}{2}\pi,\frac{1}{2}\pi]^2\times\mathbb{T}\times\mathbb{R}$, then we shall call Q fully symmetric.

Now consider a process Z on $\mathbb{R}^3 \times \mathbb{R}$ with covariance function $K(|\mathbf{x}|, |t|)$. The restriction of this process to $\mathcal{S}_r^2 \times \mathbb{R}$ has covariance function

$$\operatorname{cov}\{Z(\boldsymbol{\theta}, s), Z(\boldsymbol{\phi}, t)\} = K[2r \sin\{A(\boldsymbol{\theta}, \boldsymbol{\phi})/2\}, |s-t|],$$

which is in $\mathcal{D}$ and is fully symmetric. Indeed, this restriction of Z to $\mathcal{S}_r^2 \times \mathbb{R}$ is invariant to all rotations of the sphere and all translations in time.

We now consider two approaches to producing asymmetric covariance functions for processes on $\mathcal{S}_r^2 \times \mathbb{R}$ that are homogeneous in longitude and time. The first is to replace the translations in the translation-based approach that is described in Section 2 by rotations. Stein (2005) uses the idea of rotating a fully symmetric model but does not apply it in a way that guarantees that the resulting covariance function is in $\mathcal{D}$. In many cases, it will be natural to rotate about the north–south axis, but, in principle, any axis may be used. Denote by (L, l) the latitude and longitude of a point on the sphere with respect to this chosen axis. For definiteness, set $l = 0$ if $L = \pm\frac{1}{2}\pi$, although any other choice for longitude at the poles is equally valid. Now define $Y(L, l, t) = Z(L, l - \varepsilon t, t)$ and

$$d(L_1, L_2, l) = 2r\left\{\sin^2\left(\frac{L_1 - L_2}{2}\right) + \cos(L_1)\cos(L_2)\sin^2\left(\frac{l}{2}\right)\right\}^{1/2},$$

which gives the chordal distance between points $(L_1, 0)$ and (L_2, l) on $\mathcal{S}_r^2$. Then the covariance function $Q_Y \in \mathcal{D}$ for Y is

$$\begin{aligned} Q_Y(L_1, L_2, l, t) &= K(2r \sin[\tfrac{1}{2}A\{(L_1, l - \varepsilon t), (L_2, 0)\}], |t|) \\ &= K\{d(L_1, L_2, l - \varepsilon t), |t|\}. \end{aligned} \tag{6}$$

This covariance function has the property that, when $t = 0$, the resulting purely spatial covariance function is homogeneous on the sphere. Furthermore, if K is continuous, then Y is mean square continuous, even at the poles, which follows from the fact that $d(\frac{1}{2}\pi, L_n, l_n) \to 0$ if $L_n \to \frac{1}{2}\pi$ (or $-\frac{1}{2}\pi$) as $n \to \infty$, no matter how l_n varies.

The fact that the asymmetry parameter ε translates the longitude of Z by an amount that is proportional to time means, roughly, that the speed (measured in distance per unit time) of flow along a latitude decreases as we move polewards. This property may not always be desirable, especially if we are considering a region of $\mathcal{S}_r^2$ with large latitudinal variations. Another implication of equation (6) is that the temporal covariance function of the process at a single site generally varies with latitude when $\varepsilon \neq 0$, since

$$Q_Y(L, L, 0, t) = K\{2r \cos(L) \sin|\tfrac{1}{2}\varepsilon t|, |t|\}.$$

The second approach to generating asymmetric space–time covariance functions on the sphere is to adapt theorem 1 to the spherical setting. The following modification of theorem 1 holds.

Theorem 2. For C a measurable real-valued isotropic autocorrelation function on $\mathbb{R}^3$, γ a non-negative even measurable function on $\mathbb{R}$, θ an integer-valued odd measurable function on $\mathbb{R}$ and S an even non-negative integrable function on $\mathbb{R}$,

$$Q(L_1, L_2, l, t) = \int_{\mathbb{R}} S(\omega)\, C\{d(L_1, L_2, l)\gamma(\omega)\} \exp\{\mathrm{i}\,\theta(\omega)l + \mathrm{i}t\omega\}\, \mathrm{d}\omega \tag{7}$$

is in $\mathcal{D}$.

If we only want the existence of a process Z on $\mathcal{S}_r^2 \times \mathbb{R}$ for which

$$\operatorname{cov}\{Z(L_1, l_1, s), Z(L_2, l_2, t)\} = Q(L_1, L_2, l_1 - l_2, s - t),$$

then we can drop the requirement that θ be integer valued; the proof of this result is essentially identical to that of theorem 1. However, taking θ to be integer valued is necessary to ensure that Q depends on l_1 and l_2 only through $(l_1 - l_2) \bmod (2\pi)$, which is a requirement for Q to be in $\mathcal{D}$. Thus, although we do not need the integer-valued restriction to obtain a valid model, without it the model depends on an arbitrary choice of definition for longitude. For example, if we say that longitudes fall in the domain $(-\pi, \pi]$, then the model depends on where we place longitude 0. Furthermore, the process will generally be discontinuous along the 'International Date Line' (i.e. where the longitudes $-\pi$ and π meet). Thus, for global scale data, we shall generally want to require that θ be integer valued. This restriction implies that θ has jumps unless it is identically 0. For the Irish wind data, in which the stations cover only a narrow range of longitudes, results in Section 5 indicate that whether or not we enforce this constraint does not matter much.

Because S in equation (7) is the temporal spectrum for the process at any one location, the temporal covariance function is independent of location. This is in contrast with equation (6), for which the temporal covariance function generally depends on latitude. However, whereas models satisfying equation (6) have purely spatial covariances depending only on the angles between points, this property is generally false for models of the form (7) unless θ is identically 0. Thus, we can show that a model of the form (6) and a model of the form (7) must differ unless both are fully symmetric.

A somewhat troubling aspect of equation (7) is that, even when C is continuous, the resulting process is generally not mean square continuous at the poles unless θ is identically 0. Models for asymmetric processes on $\mathcal{S}_r^2 \times \mathbb{R}$ in Jun and Stein (2004) sometimes have a similar problem. Whether this feature is more of a practical problem than the property of equation (6) that speed of flows tends towards 0 at the poles is unclear. Both equation (6) and equation (7) should be used with caution when there are observations very near the poles.

4. Time domain likelihood approximation

Stein *et al.* (2004), following Vecchia (1988), considered approximating likelihoods for purely spatial data, and their approach can be applied in the space–time setting. The basic idea is simple, but effective implementation is sometimes not. For a vector of observations $\mathbf{Y}$ partitioned into subvectors $\mathbf{Y}_1, \ldots, \mathbf{Y}_b$, called prediction vectors, we always have

$$p(\mathbf{Y}; \boldsymbol{\theta}) = p(\mathbf{Y}_1; \boldsymbol{\theta}) \prod_{j=2}^{b} p(\mathbf{Y}_j | \mathbf{Y}_1, \ldots, \mathbf{Y}_{j-1}; \boldsymbol{\theta}), \tag{8}$$

where p indicates a generic marginal or conditional density. Then, instead of calculating exact conditional densities, replace $p(\mathbf{Y}_j | \mathbf{Y}_1, \ldots, \mathbf{Y}_{j-1}; \boldsymbol{\theta})$ by the conditional density of $\mathbf{Y}_j$ given just some subvector of $\mathbf{Y}_1, \ldots, \mathbf{Y}_{j-1}$, called the conditioning vector. Stein *et al.* (2004) showed how to extend this idea to approximating the restricted likelihood for Gaussian observations. I shall call estimates that are based on this approximation to the full or restricted likelihoods MVLEs, where the V is used in recognition of the work of Vecchia (1988). In the space–time context, it is natural to order the observations partially by using the time index and then to take each prediction vector $\mathbf{Y}_j$ to be those observations in a sequence of consecutive days.

With regular monitoring data, if we assume that the process is stationary in time, there are considerable efficiencies in picking the conditioning vectors to be all observations on a set of t consecutive days. Specifically, the covariance matrix for the observations on any set of consecutive days will be block Toeplitz: define $A_i = \operatorname{cov}(\mathbf{Z}_t, \mathbf{Z}'_{t+i})$ and, for $t > s$, $\mathbf{Z}_{s,t} = (\mathbf{Z}'_s, \ldots, \mathbf{Z}'_t)'$; then $B_t = \operatorname{cov}(\mathbf{Z}_{1,t}, \mathbf{Z}'_{1,t}) = (A_{j-k})_{j,k=1}^{t}$. The block Toeplitz structure of covariance matrices for

stationary vector-valued time series can be exploited to speed exact or approximate computations of the likelihood. Suppose, for simplicity, that Z has mean 0. Then, for $j>1$,

$$E(\mathbf{Z}_j|\mathbf{Z}_{1,j-1}) = C_{j-1}\mathbf{Z}_{1,j-1}$$

and

$$\operatorname{cov}(\mathbf{Z}_j, \mathbf{Z}_j'|\mathbf{Z}_{1,j-1}) = V_{j-1}$$

for fixed matrices C_{j-1} and V_{j-1}. The Levinson–Wiggins–Whittle–Robinson algorithm (Caines (1988), page 182) provides an efficient recursive procedure for calculating C_j and V_j from C_{j-1} and V_{j-1}. Furthermore, taking the prediction vectors to be the observations on a single day,

$$\begin{aligned}\log\{p_{\boldsymbol{\theta}}(\mathbf{Z}_{1,T})\} &= \log\{p_{\boldsymbol{\theta}}(\mathbf{Z}_1)\} + \sum_{j=2}^{T} \log\{p_{\boldsymbol{\theta}}(\mathbf{Z}_j|\mathbf{Z}_{1,j-1})\} \\ &= -\frac{NT}{2}\log(2\pi) - \frac{1}{2}\log\{\det(A_0)\} - \frac{1}{2}\sum_{j=2}^{T}[\log\{\det(V_j)\} \\ &\quad + (\mathbf{Z}_j - C_{j-1}\mathbf{Z}_{1,j-1})' V_j^{-1}(\mathbf{Z}_j - C_{j-1}\mathbf{Z}_{1,j-1})].\end{aligned}$$

Using the Levinson–Wiggins–Whittle–Robinson algorithm, this calculation requires $O(N^3T^2)$ operations and may be feasible for modest N and even fairly large T. However, even when this cannot be done, we may be able to approximate the likelihood well by

$$\log\{p_{\boldsymbol{\theta}}(\mathbf{Z}_{1,t})\} + \sum_{j=t+1}^{T} \log\{p_{\boldsymbol{\theta}}(\mathbf{Z}_j|\mathbf{Z}_{j-t,j-1})\}, \tag{9}$$

i.e. by conditioning on the previous t days when calculating the conditional distribution of $\mathbf{Z}_j$ given the past. Defining $P_t = -(C_t' I)'$, we then have

$$\begin{aligned}\sum_{j=t+1}^{T} \log\{p_{\boldsymbol{\theta}}(\mathbf{Z}_j|\mathbf{Z}_{j-t,j-1})\} &= -\frac{N(T-t)}{2}\log(2\pi) - \frac{T-t}{2}\log\{\det(V_t)\} \\ &\quad - \frac{1}{2}\sum_{j=t+1}^{T} \mathbf{Z}_{j-t,j}' P_t' V_t^{-1} P_t \mathbf{Z}_{j-t,j}.\end{aligned} \tag{10}$$

Furthermore, writing GG' and HH' for the Cholesky decompositions of $\Sigma_{j=t+1}^{T}\mathbf{Z}_{j-t,j}\mathbf{Z}_{j-t,j}'$ and V_t respectively,

$$\begin{aligned}\sum_{j=t+1}^{T} \mathbf{Z}_{j-t,j}' P_t' V_t^{-1} P_t \mathbf{Z}_{j-t,j} &= \operatorname{tr}\Big(P_t' V_t^{-1} P_t \sum_{j=t+1}^{T} \mathbf{Z}_{j-t,j}\mathbf{Z}_{j-t,j}'\Big) \\ &= \|H^{-1} P_t G\|^2,\end{aligned}$$

where $\|M\|^2$ for a matrix M means the sum of squares of its elements. G does not depend on $\boldsymbol{\theta}$ and can be calculated once and for all. Furthermore, although computing $\Sigma_{j=t+1}^{T}\mathbf{Z}_{j-t,j}\mathbf{Z}_{j-t,j}'$ looks like an $O(Tt^2N^2)$ calculation, it is easy to show that it can be done in $O(TtN^2)$ operations by using the fact that, except for minor 'edge corrections', the matrix is block Toeplitz.

By conditioning on the t most recent days, there are two reasons for the gains in efficiency over the general setting that was treated by Stein *et al.* (2004) in which the covariance matrices have no special structure that can be exploited. The first is that the resulting covariance matrices are block Toeplitz. The second, and often more important reason, is that the covariance matrices of the conditioning vectors are all the same, so the form of the conditional mean

and covariance matrix of $\mathbf{Z}_j$ given the conditioning vector $\mathbf{Z}_{j-t,j-1}$ needs to be calculated only once for any given $\boldsymbol{\theta}$. Stein *et al.* (2004) recommended conditioning vectors that were about twice as long as prediction vectors in the unstructured spatial setting that was considered there. In the setting here, the lack of dependence on j of the form of the conditional distributions leads to no advantage in choosing the prediction vectors to be more than a single day of observations.

In some circumstances, it may be worthwhile to abandon the exact block Toeplitz structure of the covariance matrices of the conditioning vectors while keeping at least most of these covariance matrices the same. In particular, if the process shows long-range dependence in time, adding a relatively small number of observations from the distant past to the conditioning vectors may improve the likelihood approximation substantially with little additional computational cost.

Other ways to approximate the likelihood are worth investigating. In particular, the block Toeplitz structure of B_T implies that we can use iterative methods to calculate the quadratic form $\mathbf{Z}_{1,T}' B_T^{-1} \mathbf{Z}_{1,T}$ in $O\{N^2 T \log(T)\}$ operations per iteration (Chan and Olkin, 1994). Although there does not appear to be an equally fast method for obtaining $\det(B_T)$, it may often be possible to obtain an excellent approximation to this determinant by calculating $\det(B_s)$ for $s = 1, \ldots, t$ for some moderate value of t and extrapolating the result to $s = T$ in some manner. We shall not pursue such possibilities here.

If the model that is under study yields covariance matrices with additional structure beyond being block Toeplitz, then we should certainly make use of it as in, for example, Haslett and Raftery (1989) and Ravishanker and Ray (2002). However, for irregularly sited monitors and the non-separable models that are considered here and in, for example, Cressie and Huang (1999) and Gneiting (2002), there does not appear to be any additional structure to exploit.

Another approach to approximating the likelihood is to work in the spectral domain and to assume that the vector-valued discrete Fourier transform of the observations viewed as a multiple time series is independent at each Fourier frequency. This approach is especially attractive when we have an explicit expression for the spectral density h of the multiple time series as, for example, in the discrete time version of theorem 1. When such an expression is not available, we could then approximate h by taking discrete Fourier transforms of the $N(N+1)/2$ cross-covariance functions or, perhaps better, compute the exact expected value of the vector-valued discrete Fourier transform of the observations at the Fourier frequencies. In either case, the computational efficiency of the spectral approach is not nearly as great as when h is known explicitly.

5. Application

We consider here the Irish wind data that were described in Haslett and Raftery (1989) and analysed in Gneiting (2002), de Luna and Genton (2005) and Stein (2005). This data set provides daily wind speeds over 18 years at 12 sites with no missing values and is available at Statlib, `http://lib.stat.cmu.edu/datasets/`. As in Haslett and Raftery (1989) and Stein (2005), we remove one of the sites, Rosslare, and analyse deseasonalized square-root wind speeds at the other 11 sites.

We consider three models: two parametric models of the rotation form (6) and a partially nonparametric model based on theorem 2. The first parametric model is

$$Q(L_1, L_2, l, t) = \frac{\phi_1 \mathbf{1}\{L_1 = L_2, l = 0\}}{(1 + \alpha_1^2 t^2)^{\nu}} + \frac{\phi_2}{(1 + \alpha_2^2 t^2)^{3\eta/2}} \exp\left\{ - \frac{\delta d(L_1, L_2, l - \varepsilon t)^{2\gamma}}{(1 + \alpha_2^2 t^2)^{3\eta/2}} \right\} \qquad (11)$$

for $\alpha_1, \alpha_2, \phi_1, \phi_2, \delta$ and ν positive, ε real and $\gamma, \eta \in (0, 1]$. Writing $\mathcal{M}_\nu$ for the Matérn covariance function with smoothness parameter ν (Stein, 1999), the second parametric model is

$$Q(L_1, L_2, l, t) = \frac{\phi_1 \mathbf{1}\{L_1 = L_2, l = 0\}\Gamma(\beta + |t|^\nu)}{\Gamma(\beta + |t|^\nu + \frac{3}{2})} + \frac{\phi_2 \mathcal{M}_{\mu+\delta|t|^\gamma}\{\alpha\, d(L_1, L_2, l - \varepsilon t)\}}{2^{\mu+\delta|t|^\gamma}\Gamma(\mu + \delta|t|^\gamma + \frac{3}{2})} \tag{12}$$

for $\alpha, \phi_1, \phi_2, \delta, \beta$ and μ positive, ε real and $\gamma, \nu \in (0, 2]$. The terms multiplying $\mathbf{1}\{L_1 = L_2, l = 0\}$ in models (11) and (12) are the spatial nugget effects and only contribute to the temporal covariance of a site with itself. The models that were applied to these data in Stein (2005) differ from models (11) and (12) in that Stein (2005) defined distances in a way that the resulting covariance functions were not provably positive definite on $\mathcal{S}_r^2 \times \mathbb{R}$, and they had a simpler parameterization for the spatial nugget effect. For convenience, we shall call equation (11) the Gneiting model and equation (12) the Matérn model. We shall assume that the mean of $Z(\mathbf{x}, t)$ depends only on $\mathbf{x}$. Define

$$\Delta Z(\mathbf{x}, t) = Z(\mathbf{x}, t + 1) - Z(\mathbf{x}, t),$$

so that ΔZ has mean 0 at every site. The likelihood of these first differences in time is the same as the restricted likelihood of the original observations. Models (11) and (12) were fitted via the MVLE method using equation (10) applied to ΔZ for $t = 180$. Stein (2005) used a cruder likelihood approximation for parameter estimation.

These parametric models have a degree of arbitrariness that is somewhat unsatisfying. Why use these models rather than any of the other space–time models that have been suggested in recent works such as Christakos (1992, 2000), Jones and Zhang (1997), Cressie and Huang (1999), Brown *et al.* (2000), de Iaco *et al.* (2001, 2002, 2003), Gneiting (2002), Hartfield and Gunst (2003), Ma (2003), Kolovos *et al.* (2004) and Stein (2005), or at least versions of these models adapted to $\mathcal{S}_r^2 \times \mathbb{R}$? The partially nonparametric spectral-in-time approach that was described in Section 2 requires far less experimentation and ingenuity on the part of the data analyst. As with models (11) and (12), allowing for a flexible spatial nugget term yields a better fit to the observations, so we consider

$$\begin{aligned} Q(L_1, L_2, l, t) = \int_{-\pi}^{\pi} & [S(\omega) C\{d(L_1, L_2, l)\gamma(\omega)\} \exp\{\mathrm{i}\,\theta(\omega) l\} \\ & + \phi(\omega)\, \mathbf{1}\{L_1 = L_2, l = 0\}] \exp(\mathrm{i}t\omega)\, \mathrm{d}\omega, \end{aligned} \tag{13}$$

where, in addition to the models for γ, θ and S in equations (3)–(5), we model

$$\log\{\phi(\omega)\} = -\beta' \log\{\sin|\tfrac{1}{2}\omega|\} + \sum_{k=0}^{K} d_k \cos(k\omega)$$

with $\beta' < 1$. In this formulation, θ and $C\{d(L_1, L_2, l_1 - l_2)\gamma(\omega)\}$ give the phase and coherence spectra for the spatially continuous part of the model, i.e. excluding ϕ. By using equation (4) for the phase, we are not constraining θ to be integer valued, so that, as pointed out in Section 3, the resulting model would not be tenable on a global scale. However, as we shall see, enforcing this constraint has little effect on the quality of the fit. We take $C(r) = \exp(-r^p)$ with $0 < p \leqslant 2$, so that, for a given K, equation (13) has $4K + 6$ parameters. We could also take $C(r) = \mathcal{M}_\nu(r)/\mathcal{M}_\nu(0)$, but, at least for smaller K, the fit is not improved, and the number of Bessel function evaluations that this choice requires leads to a substantial increase in computing time, at least when using the Bessel function program that is available in R. Using this model, the Akaike information criterion selects a value for K of 6 and the Bayes information criterion of 2. For $K = 2$, $\hat{\beta} = 0.208$, $\hat{\beta}' = 0.881$ and $\hat{p} = 1.286$. The fact that $\hat{\beta}' > \hat{\beta}$ by a substantial margin indicates that

the long-range dependence in the spatially continuous part in the model is considerably weaker than in the temporal variations that are independent between sites. A possible explanation for this finding is that local conditions around at least some of the sites were changing over long timescales in a way that was independent of what was happening at other sites.

We next consider various diagnostic plots for assessing the fits of the two parametric models and the spectral-in-time model. We begin by making the same comparisons between the fitted models and the empirical variograms as in Stein (2005). Sections 5.1 and 5.2 describe some further diagnostic plots. As in Stein (2005), we consider variograms of ΔZ to remove the effect of differences in mean wind speeds between sites. For $k \geqslant 0$, define

$$g(\mathbf{x}, \mathbf{y}; k) = \frac{1}{2(6573-k)} \sum_{j=1}^{6573-k} \{\Delta Z(\mathbf{x}, j+k) - \Delta Z(\mathbf{y}, j)\}^2 \tag{14}$$

for $k \geqslant 0$. Then $g(\mathbf{x}, \mathbf{y}; k) - g(\mathbf{y}, \mathbf{x}; k)$ provides a simple diagnostic for space–time asymmetry, since its expected value is 0 for a fully symmetric model. Fig. 1 shows some systematic discrepancies between the empirical variograms for ΔZ and the fitted parametric models, particularly in the strength of the asymmetry. Plotting symbols are used to distinguish results for the four most coastal sites from the other seven sites, since even on the square-root scale the coastal sites show somewhat greater variation (Stein, 2005). The plots show the results when $t = 180$ in the likelihood approximation, but nearly identical plots are obtained when using $t = 60$ or $t = 21$, so the discrepancies are not apparently due to any defect in the approximation.

Fig. 2 shows the same variograms by using the spectral-in-time approach with $K = 2$ and $K = 6$. The fit with $K = 2$ is fairly similar to that of the Gneiting model and the fit with $K = 6$ is slightly superior overall, doing a better job of fitting the temporal variogram in Fig. 2(b) and of fitting the asymmetry with 2-day lag in Fig. 2(f). Fig. 2(b) shows the exact lack of dependence on latitude of the temporal structure of the model, which, as can be seen in Fig. 1(b), is not the case for the rotation-based models of the form (6). In addition, Figs 2(c) and 2(e) show that these symmetrized statistics are more nearly a function of just intersite distance for the spectral-in-time models than the rotation-based models.

For purely spatial data, empirical variograms can be highly unreliable (Stein, 1999) and, hence, are of uncertain value as a diagnostic for model misspecification. However, for the Irish wind data, each variogram value is an average over 18 years of daily data and is reasonably reliable. For example, given sites $\mathbf{x}$ and $\mathbf{y}$, assume that the nine different estimates of $g(\mathbf{x}, \mathbf{y}; 2) - g(\mathbf{y}, \mathbf{x}; 2)$ based on breaking the 18 years of data into nine 2-year periods are independent and identically distributed normal. Calculating 95% confidence intervals for each pair of sites for which $\mathbf{x}$ is at least 2° east of $\mathbf{y}$, we find that the lower end points of the confidence intervals are positive in all except one case and, furthermore, are greater than the fitted values by using MVLE and the Matérn model in nine out of 15 cases. It is apparent that the asymmetry is stable across time and that the discrepancy in asymmetry between the fitted and empirical variograms cannot be blamed on the empirical variograms.

If the goal of the estimation procedure is to find a value for the parameters that visually mimics the empirical variogram summaries in Fig. 1, then we can in fact do better with either the Gneiting or Matérn model. Indeed, using either model and estimating the parameters by weighted least squares to the six empirical variogram summaries in Figs 1 or 2 yields fitted models that visually track the empirical variograms even better than the spectral-in-time fit with $K = 6$ (the results are not shown). However, these weighted least squares fits have much lower approximate likelihoods, particularly for the Matérn model, and also do poorly with some of the diagnostics that are described later in this section.

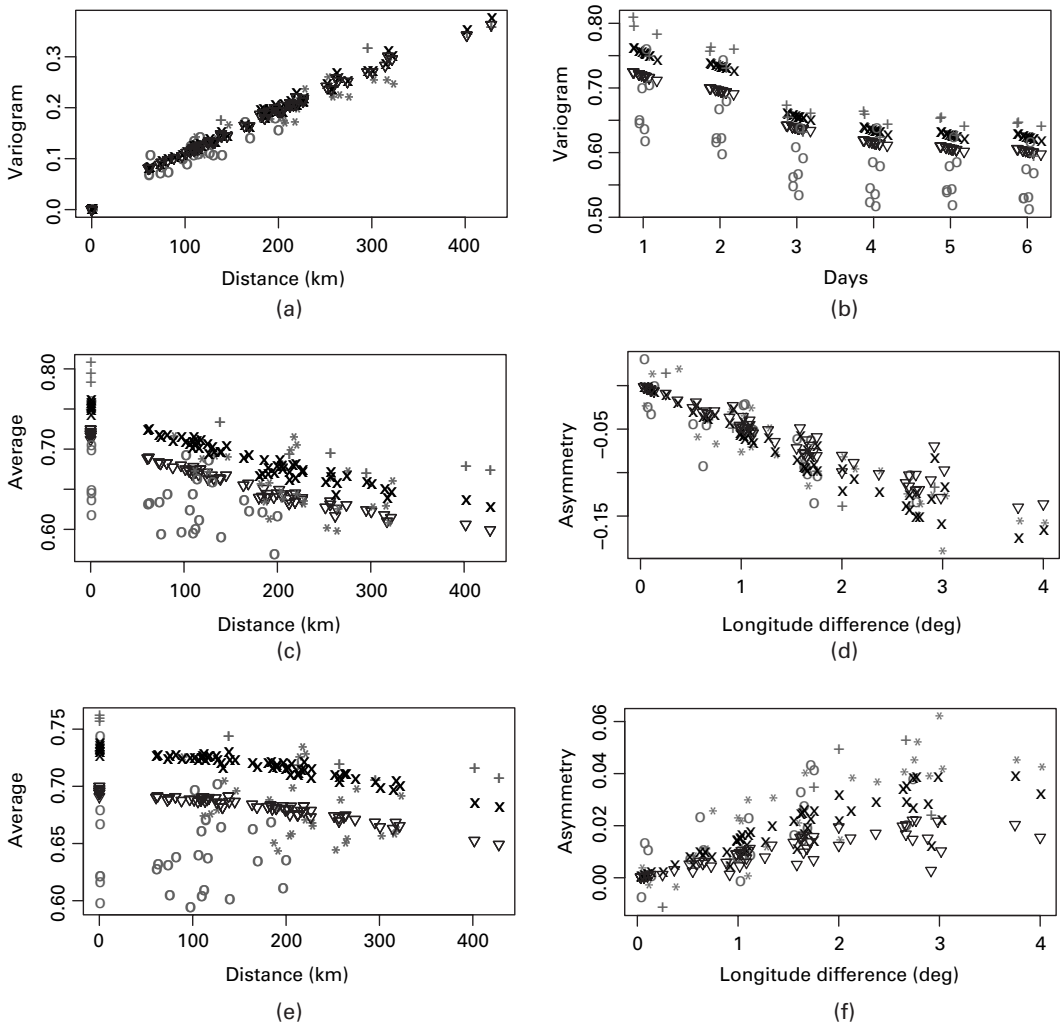

Fig. 1. Variograms for $\Delta Z(\mathbf{x}, t)$ for the Irish wind data and g defined by equation (14) (+, variogram value for two coastal sites, $*$, a coastal and an inland site; $\circ$, two inland sites; $\times$, fitted value under model (11); $\triangledown$, fitted value under model (12); both models fit by MVLE using $t = 180$ in equation (10)): (a) spatial variograms; (b) temporal variograms (horizontal offsets within each day are proportional to latitude, with more northerly sites to the right; for improved legibility, the temporal variogram at lag 0 is omitted, which is 0 for both empirical and fitted variograms); (c) $0.5\{g(\mathbf{x}, \mathbf{y}; 1) + g(\mathbf{y}, \mathbf{x}; 1)\}$; (d) $g(\mathbf{x}, \mathbf{y}; 1) - g(\mathbf{y}, \mathbf{x}; 1)$ (showing asymmetries for all pairs of sites for which $\mathbf{y}$ is east of $\mathbf{x}$); (e) $0.5\{g(\mathbf{x}, \mathbf{y}; 2) + g(\mathbf{y}, \mathbf{x}; 2)\}$; (f) $g(\mathbf{x}, \mathbf{y}; 2) - g(\mathbf{y}, \mathbf{x}; 2)$ (showing asymmetries for all pairs of sites for which $\mathbf{y}$ is east of $\mathbf{x}$)

Although the value of t in the likelihood approximation does not affect Fig. 1 much, it does have a substantial effect on the likelihoods of various estimates. Table 1 gives differences in approximate log-likelihood between eight different estimates and likelihood approximations based on conditioning on 21, 60 or 180 past days. The eight estimates are the MVLEs under the Gneiting and Matérn models based on conditioning on 21, 60 or 180 past days, the spectral-in-time model with $K = 2$ and a version of this model with the values of the estimated phase function $\hat{\theta}(\omega)$ rounded to the nearest integer for each ω. We see that, when conditioning on 21 or 60 days, MVLE under model (12) has a somewhat higher likelihood than MVLE under model

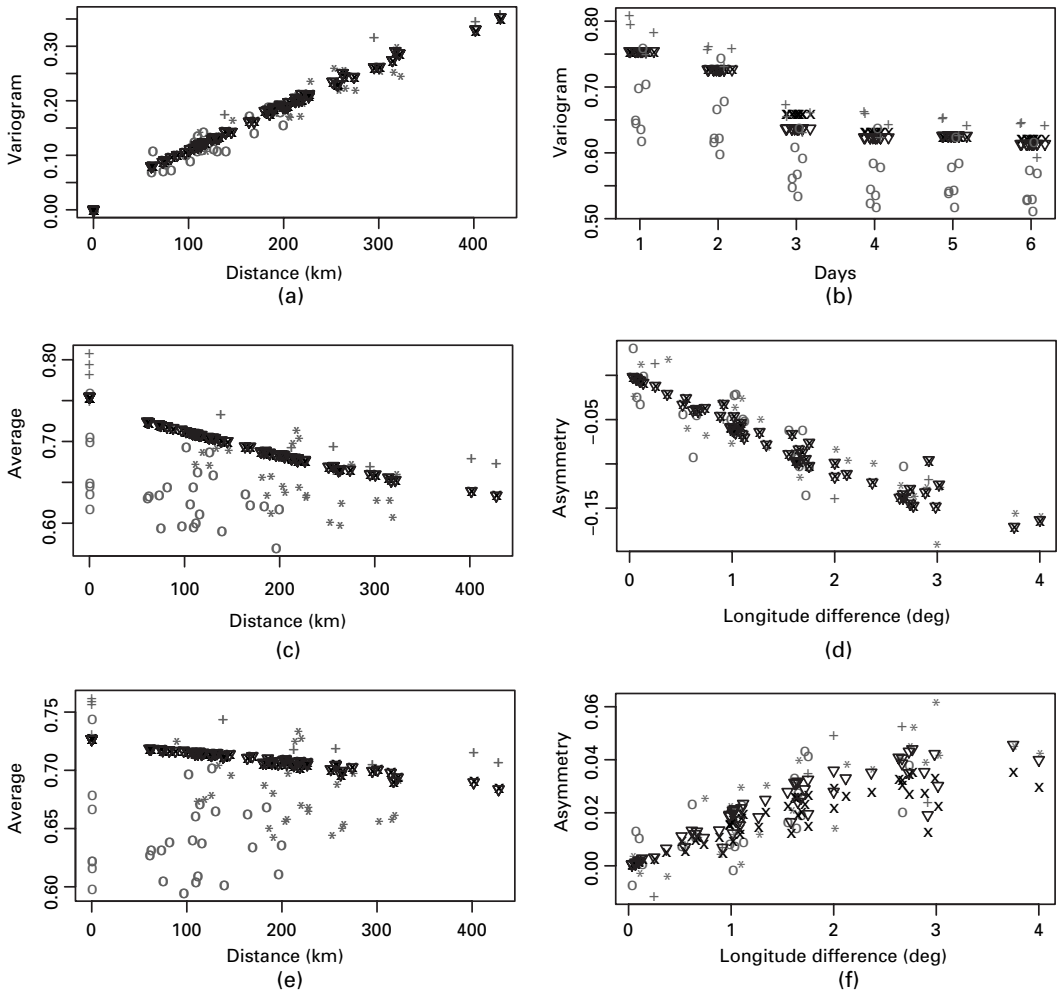

Fig. 2. Same empirical variograms as in Fig. 1: fitted variograms based on the spectral-in-time approach described in Section 5 (×, fitted value with $K = 2$; ▽, fitted value with $K = 6$)

(11), but the reverse is true when conditioning on 180 days. The approximate likelihoods for the spectral-in-time fits were computed by taking the discrete Fourier transform of the estimated spectra at each pair of sites to obtain an estimated space–time covariance matrix and then the corresponding approximate likelihoods evaluated by using equation (10). Table 1 shows that, for $t = 21, 60, 180$, the spectral-in-time method has larger approximate log-likelihood than either parametric model, and the advantage increases with t. Furthermore, forcing $\hat{\theta}$ to be integer valued has a rather small effect on the approximate log-likelihoods, reducing them by about 2 units for all values of t. Thus, at least for these data, requiring θ to be integer valued does not appear to be a serious restriction.

5.1. Spectral domain diagnostics

The variograms in Figs 1 and 2 consider only a limited range of measures of spatial–temporal variation in the data. Specifically, they are based on variances of pairwise differences of

Table 1. Comparison of approximate likelihoods for various estimates†

Criterion	*Estimate*							
	MA_{21}	MA_{60}	MA_{180}	GN_{21}	GN_{60}	GN_{180}	SP_2	SP_{2I}
Δl_{21}	0	−39	−127	−37	−60	−125	49	47
Δl_{60}	−81	0	−77	−69	−15	−63	122	120
Δl_{180}	−693	−184	0	−438	−88	17	203	201

†MA_t and GN_t are the MVLE models under results (12) and (11) when approximating the likelihood by conditioning on t past days. SP_K is the spectral-in-time model specified by equation (13) with upper limit K for the series expansions. The subscript I indicates that the function θ was rounded to take on only integer values. Δl_t gives the difference in approximate log-likelihood from MA_t, so positive numbers indicate a model with larger approximate log-likelihood than MA_t.

observations (in this case of ΔZ). This section and the next consider some other measures of variation based on variances of linear combinations of more than two observations that are particularly suited for studying regular monitoring data. Spectral domain diagnostics are the topic of this section and space–time domain diagnostics the topic of the next. These diagnostics compare empirical measures of variation with the values that are implied by fitted models. The fitted models that we consider here are the Matérn and Gneiting models and the spectral-in-time approach with $K = 2$.

For examining the purely temporal variation in the data, the marginal spectra at the 11 sites is a natural measure. Fig. 3 compares fitted with empirical marginal spectra for Z averaged over the 11 sites. All three fitted models track the marginal spectrum well, although the Matérn model appears to underestimate the power somewhat at the lowest frequencies and to overestimate the power slightly at middle frequencies.

Fig. 4 plots fitted and estimated phase spectra between Dublin and two sites to its west, Claremorris and Mullingar. The empirical phases were obtained by using the program `spec.pgram`

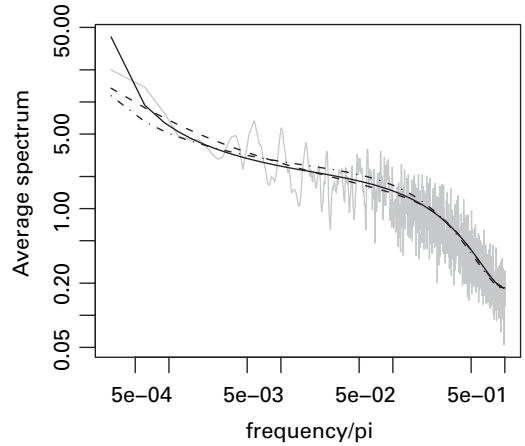

Fig. 3. Empirical and fitted average marginal spectra at the 11 sites: ———, empirical result (from R program spec.pgram with spans = 5); ———, MVLE result for the Gneiting model (11); - - - - - -, MVLE result for the Matérn model (12); -------, spectral-in-time approach described in Section 5 with $K = 2$

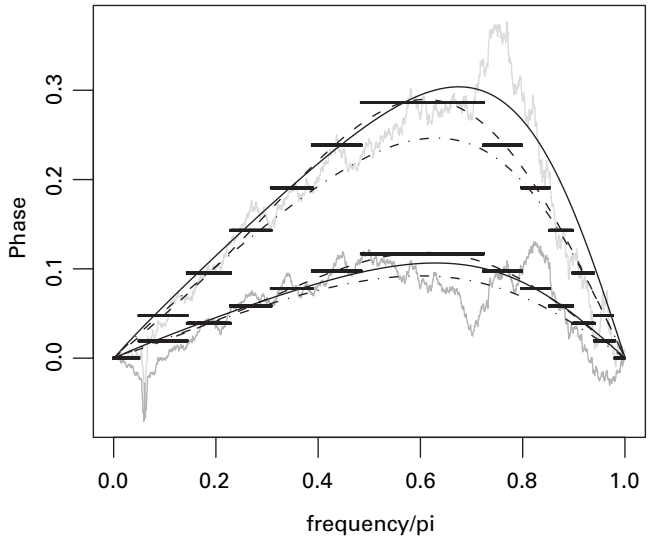

Fig. 4. Empirical and fitted phase spectra for Dublin and Claremorris (lower curves) and Dublin and Mullingar (upper curves): ———, empirical results (from R program spec.pgram with spans = 399) for Dublin and Claremorris; ———, empirical results for Dublin and Mullingar; ———, MVLE result for Gneiting model (11), · - · - ·, MVLE result for Matérn model (12); -------, spectral-in-time approach described in Section 5 with $K = 2$; — — —, spectral-in-time fit with $\hat{\theta}$-values rounded to the nearest integer

in R with the span set to 399 and no tapering. As expected, the phase is larger across the spectrum for the two sites that are farther apart in longitude. The Matérn model underestimates the phase, which makes sense since this fitted model clearly underestimates the asymmetries in Fig. 1. The other two models fit the phase spectra reasonably well. Fig. 4 also gives the fitted phase spectra under the spectral-in-time model with $\hat{\theta}$ rounded to integer values. Although the discontinuity in the fitted spectra is perhaps disconcerting, the agreement with the empirical phase is still fairly good.

Fig. 5 compares empirical and fitted coherence spectra for the two closest (Birr and Mullingar) and the two most distant sites (Valentia and Malin Head). The empirical coherences were obtained by using `spec.pgram` in R with the span set to 55 and no tapering (keeping in mind that `spec.pgram` provides squared coherences). The first 20 frequencies have been omitted because the empirical coherences have severe bias problems at these frequencies due to the long-range dependence. The Gneiting and Matérn fits have fairly similar coherence spectra. Both substantially underestimate the coherence for the two most distant sites at high frequencies and for the two nearest sites at low frequencies. The spectral-in-time fit appears to do better, if still not perfect, in both regards.

5.2. Space–time domain diagnostics

The space–time variograms in Figs 1 and 2 are examples of diagnostic plots in the space–time domain. Here we briefly describe some other space–time domain diagnostics that may be useful for regular monitoring data. Specifically, we consider plotting properties of the temporal covariance structure of various linear combinations of the sites. We shall normalize the sum of squared coefficients of the linear combination to 1. One obvious choice for a linear combination is the scaled sum over the sites. It is also valuable to consider differences across sites. Here, we look at four relatively close sites (Birr, Dublin, Mullingar and Clones) and take the linear combination

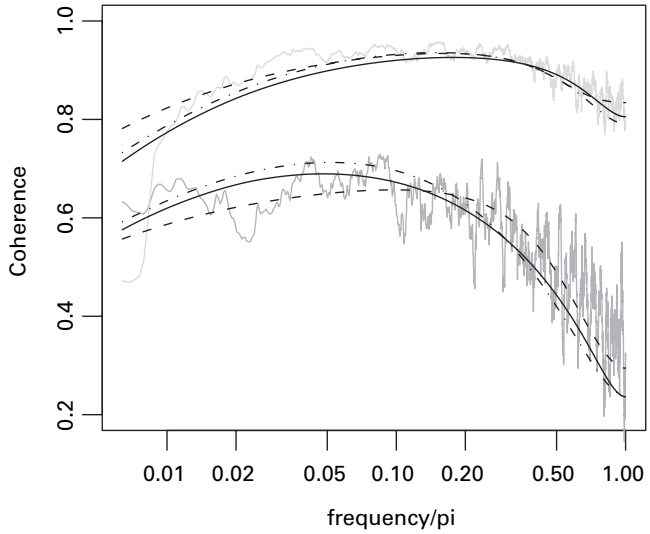

Fig. 5. Coherence spectra for the two nearest and two most distant sites (larger coherences correspond to nearest sites) (the first 20 frequencies have been omitted for improved legibility and because of difficulty in estimating coherence empirically at the lowest frequencies): ———, empirical spectrum (from R program spec.pgram with spans = 55) for the two most distant sites; ———, empirical spectrum for the two nearest sites; ———, MVLE result for Gneiting model (11); - · - · - ·, MVLE result for Matérn model (12); - - - - - -, spectral-in-time approach described in Section 5 with $K = 2$

of them that filters out a constant term, latitude and longitude, given by

$$0.40\,\text{Birr} + 0.15\,\text{Dublin} - 0.85\,\text{Mullingar} + 0.30\,\text{Clones}.$$

We could look at the temporal covariance function or temporal variogram of these linear combinations, but other possibilities may sometimes be more effective for highlighting the differences between models. For example, for $k > 0$ and $\mathbf{b} \in \mathbb{R}^n$, define

$$D_k(\mathbf{b}) = \frac{1}{k(T-2k+1)} \sum_{j=0}^{T-2k} \left\{ \sum_{l=1}^{k} (\mathbf{b}'\mathbf{Z}_{j+l} - \mathbf{b}'\mathbf{Z}_{j+k+l}) \right\}^2. \tag{15}$$

If there were no temporal dependence, then the expected value of $D_k(\mathbf{b})$ would not depend on k. As k increases, $D_k(\mathbf{b})$ puts a strong emphasis on long-term temporal correlations in $\mathbf{b}'\mathbf{Z}_j$.

We compare this statistic with what the three fitted models we are considering imply for its expected value. For the spatial average, Fig. 6 shows that the Matérn model slightly overestimates $D_k(\mathbf{b})$ throughout the range of k. The Gneiting and spectral-in-time models both do reasonably well, with perhaps a slight advantage for the spectral-in-time model. For the spatial difference filter, Fig. 7 shows that the two parametric fits give nearly identical results and are too large for all k, sometimes by up to a factor of 2. The spectral-in-time fit is clearly superior but it still is larger than the empirical values for all k displayed.

6. Discussion

Both the time domain and the spectral approximations to the likelihood that were used in the application assumed that we have regular monitoring data and that the process is stationary in time and Gaussian. Even with these strong assumptions, the results in Table 1 show that it is not easy to obtain accurate likelihood approximations when the process shows long-range

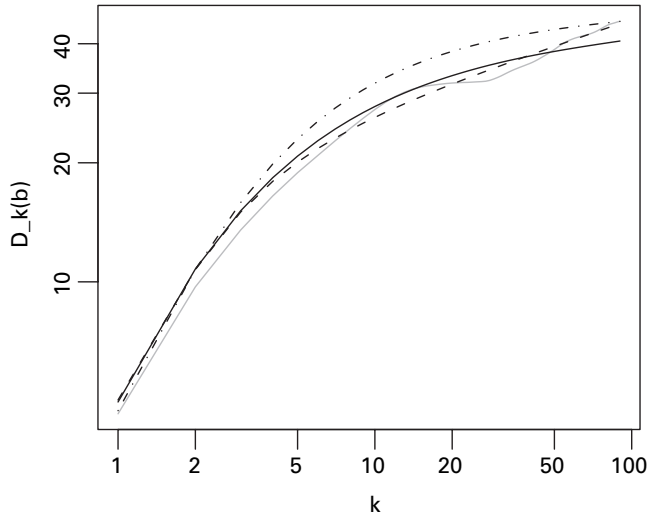

Fig. 6. Empirical (———) and fitted values for $D_k(\mathbf{b})$, defined by equation (15), where all components of **b** are equal: ———, MVLE result for Gneiting model (11); - · - · - ·, MVLE result for Matérn model (12); - - - - - - -, spectral-in-time approach described in Section 5 with $K = 2$

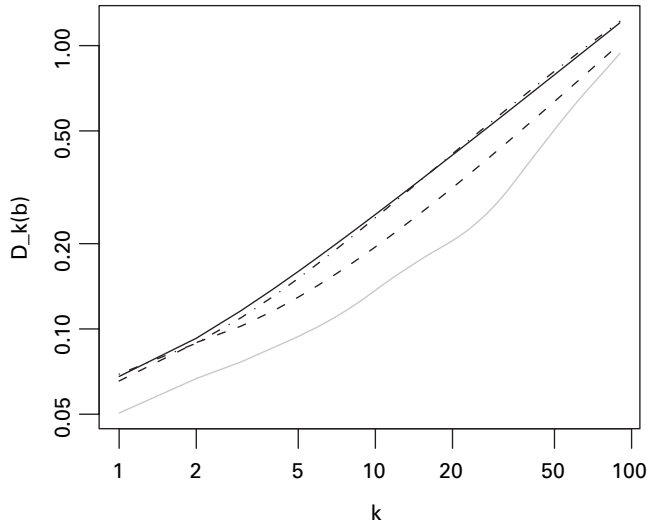

Fig. 7. Same as Fig. 6, but now **b** is a linear combination of four central Irish sites that filters out constants, latitude and longitude

dependence in time. If monitoring locations change frequently, then spectral approximations to the likelihood are no longer available, and, although we could still use equation (9), the efficiencies are no longer nearly as great as with regular monitoring data. If the monitoring network is fixed but there are missing observations, then some of the efficiencies of the regular monitoring data case can be recovered through the use of missing data algorithms. For example, Smith *et al.* (2003) showed how to apply the EM algorithm to such data in the case when the temporal effects can be captured in the mean function for the process and there is no temporal dependence in the observations. There is clearly scope for further development of missing data

methods to environmental monitoring data. In particular, it is not obvious how the various diagnostic plots that were used in Section 5 should be adapted when there are missing data.

The analysis of the Irish wind data demonstrates the feasibility of applying the partially non-parametric spectral-in-time approach to regular monitoring data. Nevertheless, the resulting fitted model shows some signs of misfit in the various diagnostic plots that were considered, although overall less than the two proposed parametric models. Furthermore, there is clearly scope for fitting these data better by allowing for spatial non-stationarity (de Luna and Genton, 2005; Stein, 2005).

For the Irish wind data, models on $\mathbb{R}^2 \times \mathbb{R}$ rather than on $\mathcal{S}_r^2 \times \mathbb{R}$ would probably suffice and it would be interesting to test out the models on $\mathcal{S}_r^2 \times \mathbb{R}$ that were presented here on data sets with greater spatial extent. Data sets that are collected via satellite would make obvious candidates, but these generally have large numbers of observations each day, for which the approximate likelihood methods that are used here would not be feasible even if there were no missing observations. The development of models that provide a realistic description of the space–time dependences in such processes and statistical methods for estimating these dependences present clear challenges to the statistical profession.

A fundamental issue that this work largely ignores is the relationship between the analysis and the objectives of a specific scientific study. Here we have taken as the goal accurately describing the spatial–temporal variations for regular monitoring data on a broad range of timescales using a model that extends naturally to a spherical spatial domain. To the extent that a model succeeds in this goal, it might be hoped that it could provide good results for many objectives. However, this might be too much to ask from one model, so there is certainly scope for fitting models for specific purposes. Thus, for example, if we are modelling space–time covariance structures as part of a data assimilation scheme (see, for example, Hamill *et al.* (2001) or Houtekamer and Mitchell (2001)), then the emphases will be largely on computational tractability and accurate modelling of dependence on shorter timescales. If, as in Gneiting *et al.* (2004) or Sahu and Mardia (2005), short-term forecasting is the objective, accurate modelling of long-range dependence is also unlikely to matter much. However, even when prediction is the goal, whether or not long-range dependence is important depends on what we want to predict. In Haslett and Raftery (1989), the goal was to estimate the average wind power that would be available over long time spans, in which case, modelling the long-range dependence accurately is critical.

Acknowledgements

Although the research that is described in this paper has been funded wholly or in part by the US Environmental Protection Agency through 'Science to achive results' co-operative agreement R-82940201-0 to the University of Chicago, it has not been subjected to the Agency's required peer and policy review and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred. The author thanks Ken Wilder for computing assistance and two referees for valuable comments leading to numerous improvements in this paper.

Appendix A

Theorem 4.4.1 in Brillinger (1981) shows that, under suitable regularity conditions, at the Fourier frequencies $\omega_l = 2\pi l/T$ for $l = 1, \ldots, \lfloor T/2 \rfloor$,

$$\mathbf{F}(\omega_l) = (2\pi T)^{-1/2} \sum_{t=1}^{T} \mathbf{X}_t \exp(\mathrm{i}\omega_l t)$$

are asymptotically independent complex Gaussian random vectors with mean 0 and covariance matrix $\Phi(\omega_l)$. Although this theorem assumes a summable covariance function, which is not so for the long-range dependence models that are used in Section 5, we shall nevertheless proceed. We can then approximate the log-likelihood by summing the log-likelihoods of the $\mathbf{F}(\omega_l)$s for $l = 1, \ldots, \lfloor T/2 \rfloor$. Because of the greater availability of software for real positive definite matrices, it is convenient to rewrite this result in terms of the $2n$-vector of real and imaginary parts of $\mathbf{F}(\omega_l)$. Write $\mathbf{G}(\omega_l)$ for the real-valued $2n$-vector that is made up of the alternating real and imaginary components of $\mathbf{F}(\omega_l)$. It then appears that we need to calculate the Cholesky decomposition of $\mathbf{G}(\omega_l)$ to obtain the approximate likelihood for frequency ω_l. However, when $\Phi(\omega)$ is of the form $(V_{jk} \exp\{\mathrm{i}(g_j - g_k)\})_{j,k=1}^n$ for V_{jk} and g_j real, which holds, for example, under equations (1) or (13), a nice simplification occurs.

Let $R(\theta)$ be the 2×2 matrix that rotates 2-vectors clockwise by the angle θ. Then the approximate covariance matrix of $\mathbf{G}(\omega)$ is given by $W = (V_{jk} R(g_j - g_k))_{j,k=1}^n$. Let $V = (V_{jk})_{j,k=1}^n$ and write $(V^{jk})_{j,k=1}^n$ for its inverse. Using the fact that

$$R(g_i - g_j)\, R(g_j - g_k) = R(g_i - g_k),$$

it is straightforward to show that the inverse of $(V_{jk}\, R(g_j - g_k))_{j,k=1}^n$ is $(V^{jk}\, R(g_j - g_k))_{j,k=1}^n$ Furthermore, it is possible to show that the determinant of $(V_{jk}\, R(g_j - g_k))_{j,k=1}^n$ is $\det(V)^2$. Writing $\mathbf{X}_j$ and $\mathbf{Y}_j$ for the vectors of real and imaginary parts of $\mathbf{F}(\omega_j)$, straightforward algebra yields that the approximate log-likelihood contribution from ω_j is (ignoring the additive constant)

$$\begin{aligned} &-\log\{\det(V)\} - \tfrac{1}{2}(\cos(g_j)\mathbf{X}_j - \sin(g_j)\mathbf{Y}_j)' V^{-1}(\cos(g_j)\mathbf{X}_j - \sin(g_j)\mathbf{Y}_j) \\ &\quad - \tfrac{1}{2}(\sin(g_j)\mathbf{X}_j + \cos(g_j)\mathbf{Y}_j)' V^{-1}(\sin(g_j)\mathbf{X}_j + \cos(g_j)\mathbf{Y}_j), \end{aligned}$$

which can be found readily from the Cholesky decomposition of the $n \times n$ real-valued matrix V.

References

Bennett, R. J. (1979) *Spatial Time Series: Analysis–Forecasting–Control*. London: Pion.

Bras, R. L. and Rodríguez-Iturbe, I. (1985) *Random Functions and Hydrology*. Reading: Addison-Wesley.

Brillinger, D. R. (1981) *Time Series: Data Analysis and Theory*, expanded edn. New York: McGraw-Hill.

Brown, P. E., Kåresen, K. F., Roberts, G. O. and Tonellato, S. (2000) Blur-generated non-separable space–time models. *J. R. Statist. Soc.* B, **62**, 847–860.

Caines, P. E. (1988) *Linear Stochastic Systems*. New York: Wiley.

Chan, T. and Olkin, J. (1994) Circulant preconditioners for Toeplitz-block matrices. *Numer. Alg.*, **6**, 89–101.

Christakos, G. (1992) *Random Field Models in Earth Sciences*. San Diego: Academic Press.

Christakos, G. (2000) *Modern Spatiotemporal Geostatistics*. New York: Oxford University Press.

Cressie, N. and Huang, H.-C. (1999) Classes of nonseparable, spatiotemporal stationary covariance functions. *J. Am. Statist. Ass.*, **94**, 1330–1340.

Gaspari, G. and Cohn, S. E. (1999) Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.*, **125**, 723–757.

Gneiting, T. (1999) Correlation functions for atmospheric data analysis. *Q. J. R. Meteorol. Soc.*, **125**, 2449–2464.

Gneiting, T. (2002) Nonseparable, stationary covariance functions for space-time data. *J. Am. Statist. Ass.*, **97**, 590–600.

Gneiting, T., Larson, K., Westrick, K., Genton, M. G. and Aldrich, E. (2004) Calibrated forecasting at the Stateline wind energy center: the regime-switching space-time (RST) method. Manuscript. (Available from `www.stat.washington.edu/www/research/reports/tr464.pdf`.)

Hamill, T. M., Whitaker, J. S. and Snyder, C. (2001) Distance-dependent filtering of background error covariance estimates in and ensemble Kalman filter. *Mnthly Weath. Rev.*, **129**, 2776–2790.

Hartfield, M. I. and Gunst, R. F. (2003) Identification of model components for a class of continuous spatio-temporal models. *J. Agric. Biol. Environ. Statist.*, **8**, 105–121.

Haslett, J. and Raftery, A. E. (1989) Space–time modelling with long-memory dependence: assessing Ireland's wind power resource (with discussion). *Appl. Statist.*, **38**, 1–50.

Houtekamer, P. L. and Mitchell, H. L. (2001) A sequential ensemble Kalman filter for atmospheric data assimilation. *Mnthly Weath. Rev.*, **129**, 123–137.

de Iaco, S., Myers, D. E. and Posa, D. (2001) Space-time analysis using a general product-sum model. *Statist. Probab. Lett.*, **52**, 21–28.

de Iaco, S., Myers, D. E. and Posa, D. (2002) Nonseparable space-time covariance models: some parametric families. *Math. Geol.*, **34**, 23–42.

de Iaco, S., Myers, D. E. and Posa, D. (2003) The linear coregionalization model and the product-sum space-time variogram. *Math. Geol.*, **35**, 25–38.

Jones, R. H. and Zhang, Y. (1997) Models for continuous stationary space-time processes. In *Modelling Longitudinal and Spatially Correlated Data* (eds T. G. Gregoire, D. R. Brillinger, P. J. Diggle, E. Russek-Cohen, W. G. Warren and R. D. Wolfinger), pp. 289–298. New York: Springer.

Jun, M. and Stein, M. L. (2004) An approach to producing space-time covariance functions on spheres. Manuscript. University of Chicago, Chicago. (Available from `www.stat.uchicago.edu/~cises/research/cises-tr18.pdf`.)

Kolovos, A., Christakos, G., Hristopulos, D. T. and Serre, M. L. (2004) Methods for generating non-separable spatiotemproal covariance models with potential environmental applications. *Adv. Wat. Resour.*, **27**, 815–830.

Kyriakidis, P. C. and Journel, A. G. (1999) Geostatistical space-time models: a review. *Math. Geol.*, **31**, 651–684.

de Luna, X. and Genton, M. G. (2005) Predictive spatio-temporal models for spatially sparse environmental data. *Statist. Sin.*, **15**, 547–568.

Ma, C. (2003) Families of spatio-temporal stationary covariance models. *J. Statist. Planng Inf.*, **116**, 489–501.

Priestley, M. B. (1981) *Spectral Analysis and Time Series*, vol. 2. London: Academic Press.

Ravishanker, N. and Ray, B. K. (2002) Bayesian prediction for vector ARFIMA processes. *Int. J. Forecast.*, **18**, 701–714.

Reinsel, G. C. (1997) *Elements of Multivariate Time Series Analysis*, 2nd edn. New York: Springer.

Sahu, S. K. and Mardia, K. V. (2005) A Bayesian kriged Kalman model for short-term forecasting of air pollution levels. *Appl. Statist.*, **54**, 223–244.

Smith, R. L., Kolenikov, S. and Cox, L. H. (2003) Spatiotemporal modeling of $PM_{2.5}$ data with missing values. *J. Geophys Res.*, **108**, 9004, doi:10.1029/2002JD002914.

Stein, M. L. (1999) *Interpolation of Spatial Data: Some Theory for Kriging*. New York: Springer.

Stein, M. L. (2005) Space-time covariance functions. *J. Am. Statist. Ass.*, **100**, 310–321.

Stein, M. L., Chi, Z. and Welty, L. J. (2004) Approximating likelihoods for large spatial data sets. *J. R. Statist. Soc.* B, **66**, 275–296.

Vecchia, A. V. (1988) Estimation and identification for continuous spatial processes. *J. R. Statist. Soc.* B, **50**, 297–312.

Weber, R. O. and Talkner, P. (1993) Some remarks on spatial correlation function models. *Mnthly Weath. Rev.*, **121**, 2611–2617.

Yadrenko, M. I. (1983) *Spectral Theory of Random Fields*. New York: Optimization Software.

Yaglom, A. M. (1987) *Correlation Theory of Stationary and Related Random Functions*, vol. I. New York: Springer.