The University of Chicago
Center for Integrating Statistical and Environmental Science
www.stat.uchicago.edu/~cises

Chicago, Illinois USA

TECHNICAL REPORT NO. 4

SPACE-TIME COVARIANCE FUNCTIONS*

M. L. Stein

May 2003
Revised: November 2004

This research is funded by
U.S. EPA - Science To Achieve Results (STAR) Program
Grant # R-82940201-0

*Although the research described in this article has been funded wholly or in part by the United States Environmental Protection Agency through STAR Cooperative 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.

Space-Time Covariance Functions *

Michael L. Stein

ABSTRACT: This work considers a number of properties of space-time covariance functions and how these relate to the spatial-temporal interactions of the process. First, it examines how the smoothness away from the origin of a space-time covariance function affects, for example, temporal correlations of spatial differences. Models that are not smoother away from the origin than they are at the origin, such as separable models, have a kind of discontinuity to certain correlations that one might wish to avoid in some circumstances. Smoothness away from the origin of a covariance function is shown to follow from the corresponding spectral density possessing derivatives with finite moments. These results are used to obtain a parametric class of spectral densities whose corresponding space-time covariance functions are infinitely differentiable away from the origin and that allows for essentially arbitrary and possibly different degrees of smoothness for the process in space and time. Second, this work considers models that are asymmetric in space-time: the covariance between site $\mathbf{x}$ at time t and site $\mathbf{y}$ at time s is different than the covariance between site $\mathbf{x}$ at time s and site $\mathbf{y}$ at time t. A general approach is described for generating asymmetric models from symmetric ones by taking derivatives. Finally, the implications of a Markov assumption in time on space-time covariance functions for Gaussian processes are examined and an explicit characterization of all such continuous covariance functions given. Several of the new models described in this work are applied to wind data from Ireland.

Keywords: Spatial isotropy, Markov process, Spectral density, Matérn model, Random processes on the sphere

* Michael L. Stein is Professor, Department of Statistics, University of Chicago, Chicago, IL 60637 (email: stein@galton.uchicago.edu). The author gratefully acknowledges Steven Lalley for providing the key ideas to the proof of Proposition 5 and the reviewers for their detailed and thoughtful comments. Although the research described in this article has been funded wholly or in part by the United States Environmental Protection Agency through STAR cooperative 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.

1. INTRODUCTION

Stochastic models describing how processes vary across space and time are essential to the application of statistics to geophysical and environmental sciences. This work considers stationary models for the covariance functions of processes on $\mathbb{R}^d \times \mathbb{R}$, where d is the spatial dimension of the process. Specifically, letting $Z(\mathbf{x}, t)$ be the value of the random field at location $\mathbf{x}$ and time t, assume there exists a function K on $\mathbb{R}^d \times \mathbb{R}$, necessarily positive definite, such that $\text{cov}\{Z(\mathbf{x}, s), Z(\mathbf{y}, t)\} = K(\mathbf{x} - \mathbf{y}, s - t)$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ and all $s, t \in \mathbb{R}$. In addition, many of the models considered here are spatially isotropic in the sense that $K(\mathbf{x} - \mathbf{y}, 0)$ depends only on $|\mathbf{x} - \mathbf{y}|$. The restriction to stationary models is not meant to suggest that nonstationary models are unimportant. However, as with many recent efforts to develop nonstationary spatial models (Sampson and Guttorp 1992, Higdon 1998, Fuentes 2002, Fuentes and Smith 2001, Nott and Dunsmuir 2002, Clerc and Mallat 2003), stationary space-time models should play a central role as a building block for nonstationary models.

Christakos (1992, 2000), Jones and Zhang (1997), Cressie and Huang (1999), Brown, *et al.* (2000), de Iaco, Myers and Posa (2001, 2002, 2003), Gneiting (2002), Ma (2003) and Hartfield and Gunst (2003) describe some recent efforts to develop new classes of stationary space-time covariance functions on $\mathbb{R}^d \times \mathbb{R}$. Kyriakidis and Journel (1999) review recent work in the geostatistical literature on space-time modeling, including approaches based on a multiple time series perspective. There is considerable additional work on models in continuous space and discrete time; see, for example, Haslett and Raftery (1989) and Handcock and Wallis (1994). Before one can judge the utility of a class of models, one needs to have some understanding of what space-time covariances imply about the corresponding processes. Because it is often difficult to think about spatial and temporal variations simultaneously, it is tempting to focus on $K(\mathbf{0}, \cdot)$, how the covariances at a single place vary across time, and $K(\cdot, 0)$, how the covariances at a single time vary across space. If these were the only characteristics that mattered, then separable models, which are of the product form $K_1(\mathbf{x})K_2(t)$, would suffice, since K_1 positive definite on $\mathbb{R}^d$ and K_2 positive definite on $\mathbb{R}$ imply K_1K_2 is positive definite on $\mathbb{R}^d \times \mathbb{R}$. Section 2 shows that separable covariance functions generally imply that small changes in the locations of observations can lead to large changes in the correlations between certain linear combinations of observations. The source of this "discontinuity" can be traced to a lack of smoothness away from the origin of separable covariance functions or, more accurately, that they are not smoother away from the origin than at the origin. Furthermore, many of the nonseparable space-time covariance functions proposed in recent works have a similar lack of differentiability along certain axes and thus similar properties with their implied correlations.

One is thus led to seek space-time covariance functions that are smooth everywhere except possibly when $(\mathbf{x}, t) = (\mathbf{0}, 0)$. Another goal here is to find models that allow different degrees of smoothness across space than across time. For example, one may want the process to be mean square differentiable with respect to spatial coordinates but not with respect to time. Section 3 gives results showing that if the derivatives of a spectral density have certain moments, then the corresponding covariance function (i.e., the Fourier transform of the spectral density) possesses derivatives away from the origin. Section 4 uses these results to give a parametric class of spectral densities whose corresponding space-time covariance functions are infinitely differentiable away from the origin and for which one can achieve essentially any combination of degrees of smoothness in space and in time by appropriate choices of the parameters. Unfortunately, explicit expressions for the covariance functions are available only for some very limited special cases.

Gneiting (2002) calls a space-time covariance function K fully symmetric if $K(\mathbf{x}, t) = K(-\mathbf{x}, t) = K(\mathbf{x}, -t) = K(-\mathbf{x}, -t)$ and notes that such a property is inappropriate for processes, such as many atmospheric processes, for which there is a dominant flow direction over time. Section 5 shows how one can generate space-time covariance functions that are spatially isotropic but not fully symmetric in space-time by taking derivatives of spatially isotropic fully symmetric models.

One common method for restricting the classes of processes one considers is to assume some kind of Markov or autoregressive structure. Section 6 gives a characterization of the class of mean square continuous, stationary, space-time Gaussian processes Z that are Markov in time in the sense that the process at times $t > 0$ and the process at times $s < 0$ are conditionally independent given the process at time 0. Section 6 gives some specific examples of such covariance functions and shows that for any given $\mathbf{x}$, $Z(\mathbf{x}, \cdot)$ can possess long-range dependence in time. The fact that a space-time process that is Markov in time can be long-range dependent when observed at a single location across time is intriguing and may possibly help to explain the frequency with which geophysical time series exhibit long-range dependence.

Section 7 applies some of these models to 18 years of daily wind speeds at 11 sites in Ireland. These data were studied at length by Haslett and Raftery (1989), but not with an eye towards obtaining a model for the covariance function on $\mathbb{R}^2 \times \mathbb{R}$. Gneiting (2002) fitted one of his proposed models to these data via weighted least squares and demonstrated its superiority to a separable model. However, Gneiting (2002) found a clear lack of full symmetry in the data, a feature which he notes his chosen model does not capture. Using approximate likelihoods, Section 7 considers a number of models not possessing full symmetry and shows that they fit the asymmetries and other properties of the empirical space-time covariances reasonably well.

Proofs of all propositions are given in the Appendix.

2. COVARIANCE FUNCTIONS WITH RIDGES

This section shows that covariance functions that are not smoother along their axes than at the origin imply a kind of discontinuity in correlations of certain linear combinations of the random field. Stein (1999, p. 30 and p. 52) provides further results on covariance functions that are not smoother away from the origin than at the origin.

For a nonnegative integer m, suppose $Z(t,s)$ is m times mean square differentiable in its first coordinate and write $Z_m(t,s)$ for this mth mean square derivative. The covariance function of Z_m is given by $K_m(t,s) = (-1)^m \partial^{2m} K(t,s)/\partial t^{2m}$. Define $\rho_\epsilon^m(t,s) = \mathrm{corr}\{Z_m(\epsilon,0) - Z_m(0,0), Z_m(t+\epsilon,s) - Z_m(t,s)\}$ and let $\rho^m(t,s)$ be its limit as $\epsilon \downarrow 0$, assuming the limit exists.

Proposition 1. *Suppose K_m is a continuous function on $\mathbb{R}^2$, $0 < \alpha_1 < \cdots < \alpha_p < 2$, $C_1, \ldots, C_p$ are even functions on $\mathbb{R}$ with $C_1(0) \neq 0$ such that*

$$K_m(t,s) = K_m(0,s) + \sum_{j=1}^{p} C_j(s)|t|^{\alpha_j} + R_s(t),$$

where, for any given s, $R_s(t) = O(t^2)$ as $t \to 0$, and $R_s(\cdot)$ has a bounded second derivative. Then

$$\sup_{t\in\mathbb{R}} \lim_{\epsilon\downarrow 0} \left| \frac{C_1(s)\{|t+\epsilon|^{\alpha_1} - 2|t|^{\alpha_1} + |t-\epsilon|^{\alpha_1}\}}{2C_1(0)\epsilon^{\alpha_1}} - \rho_\epsilon^m(t,s) \right| = 0 \tag{1}$$

and $\rho^m(t,s)$ exists for all (t,s) with

$$\rho^m(t,s) = \begin{cases} C_1(s)/C_1(0), & t = 0 \\ 0, & t \neq 0. \end{cases} \tag{2}$$

Proposition 1 has direct implications for the properties of optimal linear predictors. To simplify matters, assume $m = 0$ and consider predicting $Z(\epsilon, 0)$ based on $Z(0,0), Z(0,s)$ and $Z(\epsilon, s)$ when the mean is an unknown constant. It is possible to show that as $\epsilon \downarrow 0$, the best linear unbiased predictor, or ordinary kriging predictor, of $Z(\epsilon, 0)$ has mean squared error $2\{C_1(0) - \frac{C_1(s)^2}{C_1(0)}\}\epsilon^{\alpha_1} + o(\epsilon^{\alpha_1})$, whereas the predictor $Z(0,0)$ has mean squared error $2C_1(0)\epsilon^{\alpha_1} + o(\epsilon^{\alpha_1})$. We have that $Z(0,0)$ is an asymptotically optimal predictor of $Z(\epsilon, 0)$ if and only if $C_1(s) = 0$. Thus, the lack of continuity of $\rho^0(t,s)$ along the s axis leads to best linear unbiased predictors that, even though there is an observation $Z(0,0)$ whose correlation with the predictand $Z(\epsilon, 0)$ tends to 1 as $\epsilon \downarrow 0$, depend nontrivially on observations whose correlations with the predictand are bounded away from 1 as $\epsilon \downarrow 0$.

Many separable covariance functions $K(t,s) = K_1(t)K_2(s)$ satisfy the conditions of Proposition 1 with $C_1(s) \neq 0$ for all s, since whatever lack of smoothness $K(t,0)$ has for t near 0 will automatically be preserved in $K(t,s)$ for t near 0 unless $K_2(s) = 0$. Thus, separable covariance functions tend to have "ridges" along their axes. More specifically, if $K_1^{(2m)}(t) = C_0 + \sum_{j=1}^p C_j|t|^{\alpha_j} + R(t)$, where $R(t) = O(t^2)$ and has a bounded second derivative, then the conditions of Proposition 1 are satisfied. Furthermore, (2) holds under the weaker condition that R has a bounded second derivative in some neighborhood of the origin. Some covariance functions have logarithmic terms in their expansions at the origin (e.g., a term like $t^2 \log(|t|)$) for which it is possible to obtain an extension of Proposition 1. The only other commonly used continuous covariance functions K_1 for which $K(t,s) = K_1(t)K_2(s)$ would not satisfy the conditions of Proposition 1 are covariance functions that are analytic, such as e^{-t^2}. Stein (1999, pp. 30 and 69) argues that such covariance functions are physically unrealistic because they imply implausibly smooth processes. For example, if $K(t,0)$ is analytic in t, then, for any $\epsilon > 0$, based on observing $Z(t,s)$ for $t \in [0,\epsilon]$, it is possible to predict $Z(t,s)$ without error for all t.

Cressie and Huang (1999), de Iaco, Myers and Posa (2001, 2002) and Gneiting (2002) describe approaches of generating nonseparable space-time covariance functions. All of the examples in Cressie and Huang (1999) are analytic along either the spatial or temporal coordinates, although their general approach can yield covariance functions without this property. De Iaco, Myers and Posa (2001) consider a sum of product form models with functions of just time and just space, which share any ridges possessed by any one of the summands. De Iaco, Myers and Posa (2002) consider integrated versions of these models, which do not necessarily have ridges, but the specific examples they provide have the same smoothness along their axes as they do at the origin (and thus, for example, have a ridge along the time axis unless they are analytic in time). Gneiting (2002) shows that all functions of the form $K(\mathbf{x}, t) = \sigma^2 \psi(t^2)^{-d/2} \varphi\big(|\mathbf{x}|^2/\psi(t^2)\big)$, where φ is completely monotone on $[0, \infty)$ and ψ is positive and has a completely monotone derivative on $[0, \infty)$ are positive definite on $\mathbb{R}^d \times \mathbb{R}$. For these models, whatever lack of smoothness $K(\mathbf{x}, 0)$ has for $\mathbf{x}$ near $\mathbf{0}$ will be shared by $K(\mathbf{x}, t)$ for $t \neq 0$ and $\mathbf{x}$ near 0, since fixing $t \neq 0$, $K(\mathbf{x}, t)$ is just a rescaling of $K(\mathbf{x}, 0)$. Moreover, whatever lack of smoothness $K(\mathbf{0}, t) = \sigma^2 \psi(t^2)^{-d/2} \varphi(0)$ has for t near 0 will also hold for fixed $\mathbf{x} \neq \mathbf{0}$ and t near 0 in $K(\mathbf{x}, t)$ unless $\varphi\big(|\mathbf{x}|^2/\psi(t^2)\big)$ just happens to cancel the lack of smoothness in $\psi(t^2)^{d/2}$. All of the examples of completely monotone functions in Gneiting (2002) are strictly monotone, so at least for these φ, there cannot be such a cancellation for more than one value of $|\mathbf{x}|$. Thus, it would appear that the nonseparable covariance functions proposed in Gneiting (2002) cannot be smoother along their axes than at the origin.

For erfc the complementary error function, Ma (2003, (5.3)) shows that

$$K(\mathbf{x}, t) = e^{-\alpha|t|} \operatorname{erfc}\left(\gamma(\mathbf{x})^{1/2} - \frac{\alpha|t|}{2\gamma(\mathbf{x})^{1/2}}\right) + e^{\alpha|t|} \operatorname{erfc}\left(\gamma(\mathbf{x})^{1/2} + \frac{\alpha|t|}{2\gamma(\mathbf{x})^{1/2}}\right) \qquad (3)$$

is positive definite on $\mathbb{R}^d \times \mathbb{R}$ for γ any valid continuous variogram on $\mathbb{R}^d$. Heine (1955) obtained the special case of (3) when $d = 1$ and $\gamma(x) \propto |x|$. If γ is infinitely differentiable away from the origin, then under mild additional conditions on γ, it is possible to show that (3) is as well (see Appendix). However, $K(\mathbf{0}, t) = 2e^{-\alpha|t|}$ for any γ, so that this class of models is limited in its possible purely temporal covariance functions.

3. SMOOTHNESS OF SPECTRAL DENSITIES AND COVARIANCE FUNCTIONS

This section explores the relationship between the existence of moments for derivatives of a spectral density and derivatives of the corresponding covariance function at locations other than the origin. Suppose f is the spectral density for a real-valued weakly stationary process, so that f is nonnegative, even and integrable on $\mathbb{R}^d$. Then $K(\mathbf{x}) = \int_{\mathbb{R}^d} e^{i\mathbf{w}'\mathbf{x}} f(\mathbf{w}) \mathbf{dw}$ is the corresponding continuous covariance function. Write $\mathbf{w} = (w_1, \dots, w_d)'$, $\mathbf{x} = (x_1, \dots, x_d)'$ and for a d-tuple $\mathbf{m}$ with nonnegative integer components, set $m = m_1 + \cdots + m_d$. Let $D^{\mathbf{m}}$ denote the differential operator $\partial^m / \partial x_1^{m_1} \cdots \partial x_d^{m_d}$. For vectors $\mathbf{a}$ and $\mathbf{b}$ of length d, define $\mathbf{a}^{\mathbf{b}} = \prod_{i=1}^d a_i^{b_i}$ and say $\mathbf{a} \leq \mathbf{b}$ if $a_i \leq b_i$ for $1 \leq i \leq d$. It is well-known (Stein and Weiss (1971), p. 5) that if $\mathbf{w}^{\mathbf{m}} f(\mathbf{w})$ is integrable, then $D^{\mathbf{m}} K$ exists and $D^{\mathbf{m}} K(\mathbf{x}) = \int_{\mathbb{R}^d} (i\mathbf{w})^{\mathbf{m}} e^{i\mathbf{w}'\mathbf{x}} f(\mathbf{w}) \mathbf{dw}$. We are interested here in proving $D^{\mathbf{m}} K(\mathbf{x})$ exists for $\mathbf{x} \neq \mathbf{0}$ when $\mathbf{w}^{\mathbf{m}} f(\mathbf{w})$ is not integrable. Define $(k)_j = k(k+1) \cdots (k+j-1)$ and $(k)_0 = 1$.

Proposition 2. *Suppose $D^{\mathbf{k}}f$ exists, $D^{\mathbf{q}}f$ is integrable for all $\mathbf{q} \le \mathbf{k}$ and $\mathbf{w}^{\mathbf{m}}D^{\mathbf{k}}f(\mathbf{w})$ is integrable. If $x_j \neq 0$ for every j such that $k_j > 0$ then $D^{\mathbf{m}}K(\mathbf{x})$ exists and is given by*

$$D^{\mathbf{m}}K(\mathbf{x}) = \sum_{\mathbf{p}+\mathbf{q}=\mathbf{m}} \prod_{j=1}^{d} \binom{m_j}{p_j} i^{k_j}(-1)^{p_j}(k_j)_{p_j} x_j^{-k_j-p_j} \int_{\mathbb{R}^d} (i\mathbf{w})^{\mathbf{q}} \left\{ D^{\mathbf{k}} f(\mathbf{w}) \right\} e^{i\mathbf{w}'\mathbf{x}} \mathbf{dw}.$$

Note that if $k_j = 0$, $(k_j)_{p_j} = 0$ for $p_j > 0$, in which case, take $(k_j)_{p_j} x_j^{-k_j-p_j}$ to be 0 even if $x_j = 0$.

Corollary 1. *If the conditions on f in Proposition 2 hold only on some set $|\mathbf{w}| > R$ for some finite R and $x_j \neq 0$ for every j such that $k_j > 0$, then $D^{\mathbf{m}}K(\mathbf{x})$ exists.*

Let us apply Proposition 2 to the spectral density $f(w_1, w_2) = (1+w_1^2)^{-1}(1+w_1^2+w_2^2)^{-1}$. It is possible to show that for any given $\mathbf{m}$, choosing k sufficiently large makes $\mathbf{w}^{\mathbf{m}}D^{(0,k)}f(\mathbf{w})$ integrable. Thus, f is infinitely differentiable for all $\mathbf{x}$ with $x_2 \neq 0$. Furthermore, $w_1^k D^{(k,0)} f(\mathbf{w})$ is integrable for all positive integers k, so that $D^{(k,0)}K(x_1, 0)$ exists for all $x_1 \neq 0$. However, $w_2 D^{(k,0)} f(\mathbf{w})$ is not integrable for any positive integer k, so Proposition 2 tells us nothing about the differentiability of K with respect to x_2 along the x_1 axis. In fact,

$$K(x_1, x_2) = 2\pi \int_0^{\infty} \frac{\cos(x_1 w_1) e^{-|x_2|(1+w_1^2)^{1/2}}}{(1 + w_1^2)^{3/2}} dw_1,$$

from which one can show that for any x_1, $K(x_1, x_2) - K(x_1, 0) = -\pi^2 |x_2| e^{-|x_1|} + o(|x_2|)$ as $x_2 \to 0$, so that indeed K is not differentiable with respect to x_2 anywhere on the x_1 axis.

The next proposition gives simple conditions on f under which $D^{\mathbf{m}}K(\mathbf{x})$ exists for all $\mathbf{x} \neq \mathbf{0}$.

Proposition 3. *For $j = 1, \ldots, d$, suppose $(\partial^{\ell}/\partial w_j^{\ell}) f(\mathbf{w})$ exists and is integrable for $\ell \le k$ and $|\mathbf{w}|^n (\partial^k/\partial w_j^k) f(\mathbf{w})$ is integrable. Then for all $\mathbf{x} \neq \mathbf{0}$, $D^{\mathbf{m}}K(\mathbf{x})$ exists for all $\mathbf{m}$ such that $m_1 + \cdots + m_d \le n$. Furthermore, if $x_j \neq 0$,*

$$D^{\mathbf{m}}K(\mathbf{x}) = \sum_{p=0}^{m_j} \binom{m_j}{p} i^k (-1)^p (k)_p x_j^{-k-p} \int_{\mathbb{R}^d} \frac{(i\mathbf{w})^{\mathbf{m}}}{(iw_j)^p} \left\{ \frac{\partial^k}{\partial w_j^k} f(\mathbf{w}) \right\} e^{i\mathbf{w}'\mathbf{x}} \mathbf{dw}.$$

This follows immediately from Proposition 2. A corollary similar to Corollary 1 holds here as well.

The conditions in Proposition 3 may not be easy to verify in practice, so the next result gives a general class of spectral densities for which these conditions always hold. Write $\mathbf{w}' = (\mathbf{w}_1', \mathbf{w}_2')$ where $\mathbf{w}_j \in \mathbb{R}^{d_j}$ for $j = 1, 2$. Although d_2 is always 1 for space-time models, Proposition 4 and certain results in the next section take on a more symmetric and transparent form when treating the more general case considered here.

PROPOSITION 4. *Suppose* $f(\mathbf{w}) = \{P_1(|\mathbf{w}_1|^2) + P_2(|\mathbf{w}_2|^2)\}^{-\nu}$, *where* $\nu > 0$, f *is bounded and* P_1 *and* P_2 *are nonnegative polynomials on* $[0, \infty)$ *of positive degree* α_1 *and* α_2, *respectively. Then* f *is integrable if and only if* $d_1/(\alpha_1\nu) + d_2/(\alpha_2\nu) < 2$ *and, if so, its Fourier transform* $K(\mathbf{x})$ *possesses all partial derivatives of all orders for all* $\mathbf{x} \neq \mathbf{0}$.

4. A NEW CLASS OF SPACE-TIME COVARIANCE FUNCTIONS

One rich class of spectral densities that satisfies the conditions of Proposition 4 is

$$f(\mathbf{w}) = \left\{c_1(a_1^2 + |\mathbf{w}_1|^2)^{\alpha_1} + c_2(a_2^2 + |\mathbf{w}_2|^2)^{\alpha_2}\right\}^{-\nu} \tag{4}$$

for c_1 and c_2 positive, $a_1^2 + a_2^2 > 0$, α_1 and α_2 positive integers and $d_1/(\alpha_1\nu) + d_2/(\alpha_2\nu) < 2$. Jones and Zhang (1997) consider $d_1 = 2$ and $\alpha_2 = \nu = d_2 = 1$. If α_1 and α_2 are positive but not both integers, then (4) is still a valid spectral density whenever $d_1/(\alpha_1\nu) + d_2/(\alpha_2\nu) < 2$, but Proposition 4 no longer applies. The Fourier transform $K(\mathbf{x}_1, \mathbf{x}_2)$ of (4) depends only on $|\mathbf{x}_1|$ and $|\mathbf{x}_2|$. To see how K behaves when either $|\mathbf{x}_1|$ or $|\mathbf{x}_2|$ are 0, consider

$$K_1(\mathbf{x}_1) = K(\mathbf{x}_1, \mathbf{0}) = \int_{\mathbb{R}^{d_1}} e^{i\mathbf{w}_1'\mathbf{x}_1} \left\{\int_{\mathbb{R}^{d_2}} f(\mathbf{w}_1, \mathbf{w}_2)\mathbf{dw}_2\right\} \mathbf{dw}_1,$$

so that $f_1(\mathbf{w}_1) = \int_{\mathbb{R}^{d_2}} f(\mathbf{w}_1, \mathbf{w}_2)\mathbf{dw}_2$ is the spectral density for the covariance function $K_1(\mathbf{x}_1)$. The behavior of $K_1(\mathbf{x}_1)$ for $\mathbf{x}_1$ near $\mathbf{0}$ depends on the behavior of $f_1(\mathbf{w}_1)$ for large $|\mathbf{w}_1|$. Letting $r_j = |\mathbf{w}_j|$ for $j = 1, 2$, as $r_1 \to \infty$, it is possible to show

$$\begin{aligned} f_1(\mathbf{w}_1) &= \int_{\mathbb{R}^{d_2}} \left\{c_1(a_1^2 + |\mathbf{w}_1|^2)^{\alpha_1} + c_2(a_2^2 + |\mathbf{w}_2|^2)^{\alpha_2}\right\}^{-\nu} \mathbf{dw}_2 \\ &= \frac{2\pi^{d_2/2}}{\Gamma(d_2/2)} \int_0^\infty \left\{c_1(a_1^2 + r_1^2)^{\alpha_1} + c_2(a_2^2 + r_2^2)^{\alpha_2}\right\}^{-\nu} r_2^{d_2-1} dr_2 \\ &\sim \frac{2\pi^{d_2/2}}{\Gamma(d_2/2)} \int_0^\infty \left\{c_1 r_1^{2\alpha_1} + c_2 r_2^{2\alpha_2}\right\}^{-\nu} r_2^{d_2-1} dr_2 \\ &= \frac{\pi^{d_2/2}}{\Gamma(d_2/2)\alpha_2 c_1^\nu} \left(\frac{c_1}{c_2}\right)^{d_2/(2\alpha_2)} \frac{\Gamma\left(\frac{d_2}{2\alpha_2}\right)\Gamma\left(\nu - \frac{d_2}{2\alpha_2}\right)}{\Gamma(\nu) r_1^{\alpha_1(2\nu - d_2/\alpha_2)}} \end{aligned} \tag{5}$$

using Gradshteyn and Ryzhik (2000, 3.241.4). This result does not require α_1 or α_2 to be integers. A similar result holds with the roles of $\mathbf{w}_1$ and $\mathbf{w}_2$ switched. A desirable feature of the class of models (4) with α_1 and α_2 positive reals is that for all $\gamma_j > 0$ and $\beta_j > 0$, one can show there is an element in the class such that $f_j(\mathbf{w}_j) \sim \beta_j r_j^{-\gamma_j - d_j}$ for $j = 1, 2$. The coefficient γ_j controls the smoothness of the process in $\mathbf{x}_j$, so that, for example, Z is m times mean square differentiable in each component of $\mathbf{x}_j$ if and only if $\gamma_j > 2m$. Thus, roughly speaking, one can separately allow for any degree of smoothness of the process along its first d_1 dimensions and any different degree of smoothness along its last d_2 dimensions. However, for Proposition 4 to apply, we need α_1 and α_2 to be integers. Whenever γ_1/γ_2 is rational, it is possible to choose α_1 and α_2 integer-valued. Specifically, suppose for $\zeta > 0$ and positive integers p_1 and p_2, $\gamma_j = \zeta p_j$ for $j = 1, 2$. For any positive integer q, setting $\alpha_1 = p_1 q$, $\alpha_2 = p_2 q$, $\nu = (\zeta + d_1/p_1 + d_2/p_2)/(2q)$ and choosing c_1 and c_2 appropriately yields $f_j(\mathbf{w}_j) \sim \beta_j r_j^{-\gamma_j - d_j}$ for $j = 1, 2$. A simple calculation shows $d_1/(\alpha_1 \nu) + d_2/(\alpha_2 \nu) < 2$, so the resulting f is integrable.

There are only a few special cases of spectral densities of the form (4) for which explicit expressions for K are available. Consider first $\alpha_1 = \alpha_2 = 1$, in which case there is no loss of generality in taking $a_2 = 0$. Setting $\beta = (c_1/c_2)^{1/2}$, $K(\mathbf{x}_1, \mathbf{x}_2)$ is proportional to $\mathcal{M}_{\nu-(d_1+d_2)/2}\big(a_1(|\mathbf{x}_1|^2 + \beta^2|\mathbf{x}_2|^2)^{1/2}\big)$, where $\mathcal{M}_\nu(r) = r^\nu \mathcal{K}_\nu(r)$ and $\mathcal{K}_\nu$ is a modified Bessel function of order ν. For ϕ, θ and ν all positive, every function of the form $\phi\mathcal{M}_\nu(\theta r)$ is an isotropic covariance function in any number of dimensions. This class of covariance functions is called the Matérn class (Handcock and Stein 1993) in honor of Matérn's pioneering work in spatial statistics (Matérn 1960). The parameter ν controls the smoothness of the corresponding process, which is m times mean square differentiable in any direction if and only if $\nu > m$ (Stein (1999), p. 31).

Equation (4) can sometimes be transformed analytically when $\alpha_1 \neq \alpha_2$ if α_1, α_2, and ν are all integers. For example, a covariance function of the form (3) with $\gamma(\mathbf{x}) = |\mathbf{x}|$ is obtained when $d_1 = d_2 = 1$, $\alpha_1 = 1$, $\alpha_2 = 2$, $a_2 = 0$ and $\nu = 1$. It is possible to extend this result to other positive integer values for ν. Rather than pursuing models with $d_1 = 1$ further, let us instead consider $d_1 = 3$, $d_2 = 1$, $\alpha_1 = 2$, $\alpha_2 = 1$, $a_2 = 0$ and $\nu > 1$ an integer, for which it is also possible to obtain an explicit expression for K. For $d_1 = 3$, one cannot take $\nu = 1$, because the resulting f is not integrable. To find, for example, the Fourier transform of $\big\{c^2(a^2 + |\mathbf{w}|^2)^2 + v^2\big\}^{-2}$, first note that

$$\int_{\mathbb{R}} \frac{e^{ivt}\, dv}{\big\{c^2(a^2 + |\mathbf{w}|^2)^2 + v^2\big\}^2} = \frac{\pi e^{-c^2(a^2+|\mathbf{w}|^2)|t|}}{2c^6(a^2 + |\mathbf{w}|^2)^3}\{1 + c^2(a^2 + |\mathbf{w}|^2)|t|\},$$

so letting $r = |\mathbf{x}|$ (Yaglom 1987),

$$
\begin{aligned}
K(\mathbf{x},t) &= 2\pi^2 \int_0^\infty \frac{e^{-c^2(a^2+k^2)|t|}}{c^6(a^2+k^2)^3}\{1+c^2(a^2+k^2)|t|\}\frac{\sin(kr)}{kr}k^2 dk \\
&= \frac{2\pi^2 e^{-c^2a^2|t|}}{rc^6}\left\{\int_0^\infty \frac{e^{-c^2k^2|t|}}{(a^2+k^2)^3}k\sin(kr)dk + c^2|t|\int_0^\infty \frac{e^{-c^2k^2|t|}}{(a^2+k^2)^2}k\sin(kr)dk\right\}.
\end{aligned}
$$

Using integration by parts and 3.954 in Gradshteyn and Ryzhik (2000) on both of these integrals yields

$$
\begin{aligned}
K(\mathbf{x},t) &= \frac{\pi^2}{16c^6}e^{ar}\,\mathrm{erfc}\left(ca|t|^{1/2} + \frac{r}{2c|t|^{1/2}}\right)\left(\frac{1}{a^3} - \frac{r}{a^2} + \frac{4c^4t^2}{r}\right) \\
&\quad + \frac{\pi^2}{16c^6}e^{-ar}\,\mathrm{erfc}\left(ca|t|^{1/2} - \frac{r}{2c|t|^{1/2}}\right)\left(\frac{1}{a^3} + \frac{r}{a^2} - \frac{4c^4t^2}{r}\right) \\
&\quad + \frac{\pi^{3/2}|t|^{1/2}}{4c^5a^2}\exp\left(-c^2a^2|t| - \frac{r^2}{4c^2|t|}\right). \qquad (6)
\end{aligned}
$$

For $\mathbf{x} = \mathbf{0}$ or $t = 0$, define K by continuity, so that $K(\mathbf{x},0) = \frac{\pi^2}{8c^6a^3}e^{-ar}(1+ar)$, which implies the corresponding process is exactly once mean square differentiable in any spatial direction (Stein (1999), p. 28). Furthermore,

$$
K(\mathbf{0},t) = \frac{\pi^2}{8c^6}\left(\frac{1}{a^3} + 4c^4t^2\right)\mathrm{erfc}\left(ca|t|^{1/2}\right) + \frac{\pi^{3/2}e^{-c^2a^2|t|}}{c^6}\left(\frac{c|t|^{1/2}}{4a^2} - \frac{c^3|t|^{3/2}}{2}\right)
$$

and applying Taylor series, it follows that $K(\mathbf{0},t) = \frac{\pi^2}{8c^6a^3} - \frac{2\pi^{3/2}}{3c^3}|t|^{3/2} + O(t^2)$ as $t \to 0$. Thus, the process is not mean square differentiable in time (Stein (1999), p. 27). Finally, although it is not obvious from (6), Proposition 4 implies that this function is infinitely differentiable away from the origin.

If $\alpha_2 = 1$ in (4), then the Fourier transform over $\mathbf{w}_2$ can be obtained analytically. Specifically, defining $\theta(t) = \{c_1c_2^{-1}(a_1^2+t)^{\alpha_1} + a_2^2\}^{1/2}$ for $t \geq 0$,

$$
\int_{\mathbb{R}^{d_2}} \left\{c_1(a_1^2+|\mathbf{w}_1|^2)^{\alpha_1} + c_2(a_2^2+|\mathbf{w}_2|^2)\right\}^{-\nu} e^{i\mathbf{w}_2'\mathbf{x}_2}\mathbf{dw}_2 = \frac{\pi^{d_2/2}\mathcal{M}_{\nu-d_2/2}\big(\theta(|\mathbf{w}_1|^2)|\mathbf{x}_2|\big)}{2^{\nu-d_2/2-1}c_2^\nu\Gamma(\nu)\theta(|\mathbf{w}_1|^2)^{2\nu-d_2}}.
$$

Thus, $K(\mathbf{x})$ can be computed by numerically carrying out a one-dimensional Bessel transform. If d_1 is odd, then this Bessel transform reduces to one-dimensional Fourier transforms, which, for every value of $|\mathbf{x}_2|$ of interest, can be approximated quickly for a large number of t values using the fast Fourier transform. However, the fast Fourier transform is much less efficient for computing

K at a modest set of $(\mathbf{x}, t)$ values for many sets of parameter values, which is what is needed for statistical inference. Given the strong analyticity properties of spectral densities of the form (4), one can hope that quickly converging series expansions for their Fourier transforms can be found.

5. MODELS LACKING FULL SYMMETRY

This section describes a simple and general approach to deriving space-time covariance functions that are spatially isotropic but not fully symmetric. Suppose $\mathbf{z} \in \mathbb{R}^d$ has length 1, scalars a, c_1 and c_2 are nonnegative, $b^2 \leq 4c_1c_2$ and $f(\mathbf{w}, v) = \{a + b(\mathbf{w}'\mathbf{z})v + c_1|\mathbf{w}|^2 + c_2v^2\}\Psi(|\mathbf{w}|, |v|)$, where $(1 + |\mathbf{w}|^2 + v^2)\Psi((|\mathbf{w}|, |v|) \geq 0$ is integrable over $\mathbb{R}^{d+1}$. These conditions imply that f and Ψ are nonnegative and integrable. Furthermore, $f_1(\mathbf{w}) = \int_{\mathbb{R}} f(\mathbf{w}, v)dv$ depends on $\mathbf{w}$ only through $|\mathbf{w}|$, so that K is spatially isotropic. Now $K_0(\mathbf{x}, t)$, the Fourier transform of Ψ, only depends on $(\mathbf{x}, t)$ through $(|\mathbf{x}|, t)$, so define $\overline{K}_0$ on $\mathbb{R}^+ \times \mathbb{R}$ by $\overline{K}_0(|\mathbf{x}|, t) = K_0(\mathbf{x}, t)$. Writing $\overline{K}_0^{(m_1,m_2)}$ for $D^{(m_1,m_2)}\overline{K}_0$, we get (see Appendix)

$$
\begin{aligned}
K(\mathbf{x}, t) = {} & a\overline{K}_0(|\mathbf{x}|, t) - b\frac{\mathbf{x}'\mathbf{z}}{|\mathbf{x}|}\overline{K}_0^{(1,1)}(|\mathbf{x}|, t) - c_1\overline{K}_0^{(2,0)}(|\mathbf{x}|, t) \\
& - c_1\frac{d-1}{|\mathbf{x}|}\overline{K}_0^{(1,0)}(|\mathbf{x}|, t) - c_2\overline{K}_0^{(0,2)}(|\mathbf{x}|, t).
\end{aligned} \tag{7}
$$

In practice, whenever an analytic expression for $\overline{K}_0$ is available, there is commonly an analytic expression available for its derivatives. Furthermore, if K_0 is infinitely differentiable everywhere but at the origin, then so is K.

Let us consider a specific example for which all of the calculations can be done analytically. For β_1 and β_2 positive, consider $\overline{K}_0(r, t) = \mathcal{M}_{\nu+1}\left((\beta_1^2r^2 + \beta_2^2t^2)^{1/2}\right)$ with $\nu > 0$, for which the corresponding K_0 is twice differentiable. Letting $y = (\beta_1^2r^2 + \beta_2^2t^2)^{1/2}$ yields (Abramowitz and Stegun (1965), 9.6.26) $\overline{K}_0^{(1,0)}(r, t) = -\beta_1^2r\mathcal{M}_\nu(y)$, $\overline{K}_0^{(1,1)}(r, t) = \beta_1^2\beta_2^2rt\mathcal{M}_{\nu-1}(y)$, $\overline{K}_0^{(2,0)}(r, t) = -\beta_1^2\mathcal{M}_\nu(y) + \beta_1^4r^2\mathcal{M}_{\nu-1}(y)$ and $\overline{K}_0^{(0,2)}(r, t) = -\beta_2^2\mathcal{M}_\nu(y) + \beta_2^4t^2\mathcal{M}_{\nu-1}(y)$. Using $\mathcal{M}_{\nu+1}(y) = y^2\mathcal{M}_{\nu-1}(y) + 2\nu\mathcal{M}_\nu(y)$,

$$
\begin{aligned}
K(\mathbf{x}, t) = {} & \left\{(a - c_1\beta_1^2)\beta_1^2r^2 + (a - c_2\beta_2^2)\beta_2^2t^2 - b\beta_1^2\beta_2^2(\mathbf{x}'\mathbf{z})t\right\}\mathcal{M}_{\nu-1}(y) \\
& + \left(2a\nu + c_2\beta_2^2 + c_1d\beta_1^2\right)\mathcal{M}_\nu(y).
\end{aligned}
$$

If $c_1 = a/\beta_1^2$ and $c_2 = a/\beta_2^2$, then $K(\mathbf{x}, t) = a(2\nu + d + 1)\mathcal{M}_\nu(y) - b\beta_1^2\beta_2^2(\mathbf{x}'\mathbf{z})t\mathcal{M}_{\nu-1}(y)$. Defining $\tau = b\beta_1\beta_2/(2a)$ gives the alternative form

$$K(\mathbf{x}, t) = a\big\{(2\nu + d + 1)\mathcal{M}_\nu(y) - 2\tau(\beta_1\mathbf{x}'\mathbf{z})\beta_2 t\mathcal{M}_{\nu-1}(y)\big\}, \tag{8}$$

where $0 \leq \tau \leq 1$ guarantees that K is positive definite.

Setting $d = 2$, $a = \beta_1 = \beta_2 = 1$, $\mathbf{z} = (1, 0)', \nu = 0.5$ and $\tau = 1$ in (8), Figure 1 plots $K((x_1, 0), t)/K(\mathbf{0}, 0)$. The lack of full symmetry is apparent. On the other hand, $K((0, x_2), t) = K((0, -x_2), t)$ for all x_2 and t. Other directions for $\mathbf{x}$ yield a linear combination of these two cases.

Models of the form (8) have equal degrees of smoothness in space and in time. Thus, we can compare these models to what one would get by setting $s = \big(|\mathbf{x}|^2 + 2t\mathbf{b}'\mathbf{x} + c^2t^2\big)^{1/2}$ for $|\mathbf{b}| \leq c$ and letting $K(\mathbf{x}, t) = C(s)$ for C an isotropic covariance function in $d + 1$ dimensions. This approach always yields elliptical contours for the covariance function, in contrast to what Figure 1 shows. In addition, note that K in Figure 1 becomes slightly negative for x_1 and t large and positive but remains positive when $x_1 < 0$ and $t > 0$, which could not happen for a covariance function with elliptical contours. Furthermore, the general approach outlined in this section can be used to obtain asymmetric covariance functions that have different smoothness across space than across time, which an affine coordinate transformation of an isotropic model cannot achieve.

6. MARKOV MODELS

One commonly used principle for restricting the class of stochastic processes to consider is to require some kind of Markov property. Although there are notions of Markovian behavior for spatial processes, the Markov property is more frequently used to describe dependence structure in time, in which case it has the interpretation that the future and the past are conditionally independent given the present. Note that we are only assuming the conditional independence holds if Z is observed throughout $\mathbb{R}^d$ at a given time t. If Z is observed on some $A \subset R^d$ at time t, we will generally not have conditional independence of the future and past even on A. So far, this work has only considered the first two moments of space-time processes, but by adding the assumption that the process is Gaussian, then the process is Markov if and only if for every $t_0 \in \mathbb{R}$, $t > t_0$ and $\mathbf{x} \in \mathbb{R}^d$, the best linear predictor of $Z(\mathbf{x}, t)$ in terms of $Z(\cdot, t_0)$ is the same as the best linear predictor of $Z(\mathbf{x}, t)$ in terms of $\{Z(\mathbf{y}, s) : \mathbf{y} \in \mathbb{R}^d, s \leq t_0\}$. If the process Z is stationary in space-time, then it suffices to verify this property for $\mathbf{x} = \mathbf{0}$ and $t_0 = 0$. The following result characterizes all continuous space-time covariance functions satisfying this condition.

Proposition 5. *Let Z be a stationary, mean square continuous real-valued space-time process with covariance function K. Then the best linear predictor of $Z(\mathbf{0}, t)$ in terms of $Z(\cdot, 0)$ is the same as the best linear predictor of $Z(\mathbf{0}, t)$ in terms of $\{Z(\mathbf{y}, s) : \mathbf{y} \in \mathbb{R}^d, s \leq 0\}$ for all $t > 0$ if and only if K is of the form*

$$K(\mathbf{x}, t) = \int_{\mathbb{R}^d} \exp\{i\mathbf{w}'\mathbf{x} - |t|\beta(\mathbf{w}) - it\phi(\mathbf{w})\} F(\mathbf{dw}), \tag{9}$$

where β is an even nonnegative Borel-measurable function, ϕ is an odd Borel-measurable function and F is a positive, finite symmetric measure on $\mathbb{R}^d$.

Note that $K(\mathbf{x}, 0) = \int_{\mathbb{R}^d} e^{i\mathbf{w}'\mathbf{x}} F(\mathbf{dw})$, so that F is the spectral distribution for the spatial variation of Z. Furthermore, K is spatially isotropic if and only if the measure F depends on $\mathbf{w}$ only through $|\mathbf{w}|$. The function β can be viewed as a damping factor in time for each spatial frequency $\mathbf{w}$ and the function ϕ a phase modulation for each frequency. Note that $\beta(\mathbf{w})$ may equal 0 for certain frequencies, which means that those frequencies are not damped at all in time.

If $\beta(\mathbf{w}) > 0$ almost everywhere with respect to F, one can rewrite (9) as

$$K(\mathbf{x}, t) = \frac{1}{\pi} \int_{\mathbb{R}^{d+1}} \frac{\beta(\mathbf{w}) e^{ivt + i\mathbf{w}'\mathbf{x}}}{\beta(\mathbf{w})^2 + \{v + \phi(\mathbf{w})\}^2}\, dv\, F(\mathbf{dw}). \tag{10}$$

If $1/\beta(\mathbf{w})$ is integrable over $\mathbb{R}^d$, then taking $F(\mathbf{dw}) = \mathbf{dw}/\beta(\mathbf{w})$ yields $[\beta(\mathbf{w})^2 + \{v + \phi(\mathbf{w})\}^2]^{-1}$ as the spectral density of a stationary space-time Gaussian process that is Markov in time. Thus, spectral densities of the form (4) with $d_2 = 1$ correspond to Gaussian Markov processes if and only if $\nu = \alpha_2 = 1$.

The model given by Brown, *et al.* (2000) is the special case of (10) when $\phi(\mathbf{w}) = \mathbf{w}'\mathbf{u}$, $\beta(\mathbf{w}) = \lambda + \frac{1}{2}\mathbf{w}'\mathbf{\Sigma}\mathbf{w}$ and $F(\mathbf{dw})/\mathbf{dw} = \exp\big(-\frac{1}{2}\mathbf{w}'\mathbf{\Sigma}\mathbf{w}\big)/\big(\lambda + \frac{1}{2}\mathbf{w}'\mathbf{\Sigma}\mathbf{w}\big)$ for some $\lambda \geq 0$, a vector $\mathbf{u}$ and positive definite matrix $\mathbf{\Sigma}$. All such processes are analytic in space, which may limit their utility.

A class of models for which (9) can be evaluated analytically is $\beta(\mathbf{w}) = \zeta \log(1 + \alpha^{-2}|\mathbf{w}|^2)$, $F(\mathbf{dw})/\mathbf{dw} = (1 + \alpha^{-2}|\mathbf{w}|^2)^{-(\nu+d)/2}$ and $\phi(\mathbf{w}) = \epsilon\mathbf{w}'\mathbf{z}$ for some unit vector $\mathbf{z}$ and positive α, ν and ζ. In this case,

$$K(\mathbf{x}, t) = \int_{\mathbb{R}^d} \frac{e^{i\mathbf{w}'(\mathbf{x} - \epsilon t\mathbf{z})}}{(1 + \alpha^{-2}|\mathbf{w}|^2)^{\nu + \zeta|t| + d/2}} \mathbf{dw}$$

$$= \frac{\pi^{d/2}\alpha^d}{2^{\nu+\zeta|t|-1}\Gamma\big(\nu+\zeta|t|+\frac{1}{2}d\big)}\mathcal{M}_{\nu+\zeta|t|}(\alpha|\mathbf{x}-\epsilon t\mathbf{z}|). \tag{11}$$

Taking $t=0$ essentially recovers the Matérn model for the spatial variation: $K(\mathbf{x},0)\propto\mathcal{M}_\nu(\alpha|\mathbf{x}|)$, so that one can get any possible degree of differentiability in space. When $\mathbf{x}=\mathbf{0}$, $d=2$ and $\epsilon=0$, $K(\mathbf{0},t)=\pi\alpha^2/(\nu+\zeta|t|)$. For general d, using Stirling's formula, it is possible to show that as $|t|\to\infty$, $K(\mathbf{0},t)\sim\{\pi\alpha^2/(\zeta|t|)\}^{d/2}$. For $d=1$ or 2 and $\epsilon=0$, $K(\mathbf{0},t)$ is not integrable in t, so the process at a single location exhibits long-range dependence.

That a Markov process can exhibit long-range dependence along one of its margins may be unexpected; certainly, this cannot happen for a finite-dimensional ergodic Markov process. The source of the long-range dependence for this process is that the damping factor, $\beta(\mathbf{w})$, is near 0 for $|\mathbf{w}|$ near 0, so that the low frequency spatial variations decay very slowly. Geophysical time series often exhibit long-range dependence and perhaps this can be explained by the fact that one generally analyzes low-dimensional margins of high or infinite-dimensional systems that are actually Markov but for which the damping of broad scale spatial features of the system is sufficiently weak to produce slowly decaying correlations in time.

If $\epsilon\neq 0$ in (11), then the model is not fully symmetric. Furthermore, from Abramowitz and Stegun (1965, 9.7.8), $K(\mathbf{0},t)$ decays exponentially as $t\to\infty$. Instead, $K(\epsilon t\mathbf{z},t)\sim\{\pi\alpha^2/(\zeta|t|)\}^{d/2}$ as $t\to\infty$, so that the spatial-temporal correlations decay algebraically along this line in space-time.

Let us next consider the smoothness away from the origin of (11), setting $\epsilon=0$ for simplicity. First of all, K is infinitely differentiable at $(\mathbf{x},t)$ if $t|\mathbf{x}|\neq 0$. Because the smoothness of $\mathcal{M}_{\nu+\zeta|t|}(\alpha|\mathbf{x}|)$ in $\mathbf{x}$ as $\mathbf{x}\to\mathbf{0}$ increases with $|t|$ (Stein (1999), p. 31), K is smoother in $\mathbf{x}$ along the spatial axes than it is at the origin. However, it is possible to show that for any given $\mathbf{x}$, $K(\mathbf{x},t)=K(\mathbf{x},0)-C_{\mathbf{x}}|t|+o(|t|)$ as $t\to 0$ with $C_{\mathbf{x}}\neq 0$, so K is no smoother along the time axis than at the origin.

Similar to (3), for a valid variogram γ on $\mathbb{R}$, it is possible to show that if $|t|$ is replaced by $\gamma(t)$ on the right-hand side of (9), the resulting model is still positive definite (although no longer Markov unless $\gamma(t)\propto|t|$). The proof is similar to that of Theorem 2 in Ma (2003). Applying this extension to (11) yields

$$K(\mathbf{x},t)=\frac{\pi^{d/2}\alpha^d}{2^{\nu+\gamma(t)-1}\Gamma\big(\nu+\gamma(t)+\frac{1}{2}d\big)}\mathcal{M}_{\nu+\gamma(t)}(\alpha|\mathbf{x}-\epsilon t\mathbf{z}|) \tag{12}$$

as an explicit class of space-time covariance functions that can achieve any degree of differentiability in space and in time.

7. IRISH WIND DATA

This section applies some of the models described here to the Irish wind data studied in Haslett and Raftery (1989), Gneiting (2002) and de Luna and Genton (2005). The daily average wind speeds from 1961–1978 at 12 sites and the intersite distances are available at Statlib, http://lib.stat.cmu.edu/datasets/. Tilmann Gneiting kindly provided the latitudes and longitudes of the sites. As in Haslett and Raftery (1989), the analyses here are on the square roots of the wind speeds because this transformation makes the data nearly Gaussian. Similar to Haslett and Raftery (1989) and Gneiting (2002), the data were deseasonalized by regressing the wind speeds averaged across sites on a small number of annual harmonics and one of the sites, Rosslare, was removed because of some clear nonstationarities that result when it is included. Denote by Z the deseasonalized square root wind speeds. The mean of Z clearly varies with site, so $EZ(\mathbf{x}_i, t)$ is modeled as an unknown constant m_i. To have a complete model for the wind speed process over Ireland, one would need a spatial model for $EZ(\mathbf{x}, t)$ as a function of $\mathbf{x}$. The m_is are larger at the coastal sites than inland, and any sensible model for $EZ(\mathbf{x}, t)$ would need to incorporate this information. By analyzing only differences in time of the observations, the problem of modeling the m_is is avoided here.

One way to compare the models would be through the maxima of their likelihoods, but it is difficult to calculate likelihoods exactly for all 72,314 $= 11 \times 6574$ observations. Write $\Delta Z(\mathbf{x}, t) = Z(\mathbf{x}, t+1) - Z(\mathbf{x}, t)$ and set $\mathbf{D}_t = (\Delta Z(\mathbf{x}_1, t), \cdots, \Delta Z(\mathbf{x}_{11}, t))$. For $j = 1, \ldots 313$, let $\mathbf{Y}_j = (\mathbf{D}_{(j-1)t+1}, \cdots, \mathbf{D}_{(j-1)t+21})$. Since $6574 = 21 \times 313 + 1$, $\mathbf{Y} = (\mathbf{Y}_1, \cdots, \mathbf{Y}_{313})$ is the vector of all first differences in time of the data and $E\mathbf{Y} = \mathbf{0}$. Indexing the model for the covariance function for Z by $\boldsymbol{\theta}$, the likelihood for $\boldsymbol{\theta}$ in terms of $\mathbf{Y}$ is the restricted likelihood of the full dataset and its maximizer is called the restricted maximum likelihood estimator (Christensen 1996). A simple approximation to the restricted loglikelihood is given by $\sum_{j=1}^{313} \log p(\mathbf{Y}_j \mid \boldsymbol{\theta})$, which is a special case of an approximation studied in Stein, Chi and Welty (2004) that extends an approach due to Vecchia (1988) to the restricted likelihood setting. This approximation ignores the dependence between the $\mathbf{Y}_j$s, but does yield an unbiased estimating equation for $\boldsymbol{\theta}$ (Stein, Chi and Welty 2004).

It is natural to compare any fitted covariance function to empirical measures of space-time variations. The empirical space-time variogram for Z would include the effect of the m_is. To avoid this problem, Figure 2 plots summaries of the space-time variogram for ΔZ, which is invariant to changes in the m_is. Figure 2a shows the strong dependence on distance of the spatial variogram of ΔZ. Figure 2b shows that the temporal variogram of ΔZ decreases from lag 1 to lag 6. The temporal variogram of the undifferenced data (not shown) is increasing and concave in k and this concavity leads to the temporal variogram of ΔZ being above its sill for k small and positive. Alternatively, one could say that taking first differences in time slightly overdifferences the data, inducing negative correlations between $\Delta Z(\mathbf{x}, t+k)$ and $\Delta Z(\mathbf{x}, t)$ for k small and positive.

Define the empirical space-time variogram of ΔZ by

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

for $k \geq 0$. Figures 2d and f demonstrate the clear asymmetry in $g(\mathbf{x}, \mathbf{y}; k)$ for $k = 1$ or 2, which appears to be largely a function of differences in longitudes between sites. If ΔZ is stationary in space-time with covariance function K, then $E\{g(\mathbf{x}, \mathbf{y}; k) - g(\mathbf{y}, \mathbf{x}; k)\} = K(\mathbf{y}-\mathbf{x}, k) - K(\mathbf{x}-\mathbf{y}, k)$. Thus, the increase in $g(\mathbf{x}, \mathbf{y}; k) - g(\mathbf{y}, \mathbf{x}; k)$ with longitude in Figure 2d shows that if $\mathbf{y}$ is east of $\mathbf{x}$, then the covariance between ΔZ at $\mathbf{x}$ today and $\mathbf{y}$ tomorrow is generally larger than the covariance between ΔZ at $\mathbf{y}$ today and $\mathbf{x}$ tomorrow, with the difference being approximately proportional to the difference in longitudes between the sites. Figure 2f shows that this pattern is reversed for $k = 2$: for $\mathbf{y}$ east of $\mathbf{x}$, $\Delta Z(\mathbf{y}, t+2)$ generally has weaker covariance with $\Delta Z(\mathbf{x}, t)$ than $\Delta Z(\mathbf{y}, t)$ with $\Delta Z(\mathbf{x}, t+2)$. This reversal in asymmetry does not have a simple explanation, but it is at least partially a consequence of considering the space-time variogram of ΔZ rather than Z itself. More specifically, if one looks at empirical estimates of $\mathrm{cov}\{Z(\mathbf{x}, t+k), Z(\mathbf{y}, t)\} - \mathrm{cov}\{Z(\mathbf{y}, t+k), Z(\mathbf{x}, t)\}$ for $\mathbf{y}$ east of $\mathbf{x}$, then one gets the expected result of a generally decreasing relationship as the difference in longitude increases for both $k = 1, 2$ and even $k = 3$. However, despite the perhaps greater difficulty of interpreting the space-time variogram of ΔZ rather than Z, the fact that differencing automatically removes any constant spatial effect makes it attractive.

Another feature that any good model should capture is an apparent discontinuity in $K(\mathbf{x}, t)$ at $\mathbf{x} = \mathbf{0}$ for all t, which can be seen if one looks carefully at Figure 2c or at Figure 5 in Gneiting (2002). Gneiting (2002) calls this phenomenon a spatial nugget effect. Because the 11 stations are fairly evenly spread throughout Ireland, one cannot say whether this apparent discontinuity would be present for two sites within, say, a few kilometers of each other.

A simple way to introduce asymmetry into a model of the form $K(|\mathbf{x}|, t)$ is to consider, for some unit vector $\mathbf{z}$, the model $K(|\mathbf{x}-\epsilon t\mathbf{z}|, t)$, which is positive definite on $\mathbb{R}^d \times \mathbb{R}$ whenever $K(|\mathbf{x}|, t)$ is. Since the asymmetry appears to be almost entirely a function of differences in longitude, letting x_1 indicate latitude and x_2 longitude, we will set $\mathbf{z} = \mathbf{u}_2 = (0, 1)'$ rather than trying to estimate it in the various models.

A general issue that must be addressed is how to interpret $|\mathbf{x}-\epsilon t\mathbf{u}_2|$ when the spatial locations are points on the sphere $\mathcal{S}$ and not the plane. The simplest solution is just to map the relevant portion of $\mathcal{S}$ to $\mathbb{R}^2$. For example, if $\overline{L}$ is the mean latitude of the 11 sites, one could use the mapping from the sphere to the plane of $(x_1, x_2) \mapsto (x_1, x_2 \cos \overline{L})$. This suggests defining

$$\text{cov}\{Z(\mathbf{x}, s), Z(\mathbf{y}, t)\} = K\big(r\left[(x_1 - y_1)^2 + \{(x_2 - y_2)\cos \overline{L} - \epsilon(s-t)\}^2\right]^{1/2}, s-t\big), \tag{13}$$

where r is the conversion factor from a degree latitude to kilometers, roughly 111. If K is a valid covariance function for the process on $\mathbb{R}^2 \times \mathbb{R}$, then the resulting covariance function for the process on $\mathcal{S} \times \mathbb{R}$ is valid as well. However, the ratio of the arclength of a degree longitude of the southernmost to the northernmost site is 1.088, so this transformation entails a nontrivial distortion of distances. Transformations from $\mathcal{S}$ to $\mathbb{R}^2$ that distort distances less would require that lines of longitude not be mapped into vertical lines, which would also be undesirable.

An approach that does not distort distances is to define the function $G(\mathbf{x}, \mathbf{y}) = r \cos^{-1}\big\{\sin x_1 \sin y_1 + \cos x_1 \cos y_1 \cos(x_2 - y_2)\big\}$, the great circle distance between two points on the Earth assuming it is a perfect sphere, and set

$$\text{cov}\{Z(\mathbf{x}, s), Z(\mathbf{y}, t)\} = K\big(G(\mathbf{x}, \mathbf{y} + \epsilon(s-t)\mathbf{u}_2), s-t\big). \tag{14}$$

However, K positive definite on $\mathbb{R}^2 \times \mathbb{R}$ no longer guarantees that the resulting model is positive definite. The approximate likelihood approach used here at least guarantees that the fitted models are positive definite at the 11 sites for 21 days of first differences. Note that this definition leads to the temporal autocovariance function of $Z(\mathbf{x}, \cdot)$ depending on latitude.

Of course, the best solution would be to develop directly valid covariance function models for $\mathcal{S} \times \mathbb{R}$, but that would be beyond the scope of this paper. An easy way to generate valid models for processes on $\mathcal{S}$ is to consider restrictions of models on $\mathbb{R}^3$ to the sphere (Yaglom 1987, Gneiting 1999). This idea can be easily extended to yield fully symmetric models on $\mathcal{S} \times \mathbb{R}$, but it is not so clear what might be good asymmetric models on $\mathcal{S} \times \mathbb{R}$.

Extensive experimentation with various models led to the following two best fitting models in terms of the approximate restricted likelihood. The first is an extension of (4) and (21) in Gneiting (2002) that allows for space-time asymmetry and a more flexible spatial nugget effect than in (21) in Gneiting (2002):

$$\mathrm{cov}\{Z(\mathbf{x}, s), Z(\mathbf{y}, t)\} = \frac{\phi}{(1 + a|s-t|^2)^{\alpha}} \exp\left\{ -\frac{cG(\mathbf{x}, \mathbf{y} + \epsilon(s-t)\mathbf{u}_2)^{2\gamma}}{(1 + a|s-t|^2)^{\alpha\gamma}} \right\} + \frac{\delta 1\{\mathbf{x} = \mathbf{y}\}}{(1 + a'|s-t|^2)^{\alpha'}}. \quad (15)$$

The parameter domain is all real ϵ, all nonnegative a, a', c, ϕ, α' and δ and all α' and γ in $[0, 1]$. This nine-parameter models sets the parameter β in (21) of Gneiting (2002) to its maximum possible value, 1, since this boundary value maximizes the approximate likelihood. The other best fitting model is an eight-parameter model based on (12) with $\gamma(t) = \zeta|t|^{\gamma}$ that includes a two-parameter spatial nugget term suggested by the form of (12) when $\mathbf{x} = \mathbf{0}$ and $\epsilon = 0$:

$$\mathrm{cov}\{Z(\mathbf{x}, s), Z(\mathbf{y}, t)\} = \frac{\phi \mathcal{M}_{\nu+\zeta|s-t|^{\gamma}}(\alpha G(\mathbf{x}, \mathbf{y} + \epsilon(s-t)\mathbf{u}_2))}{2^{\nu+\zeta|s-t|^{\gamma}} \Gamma\big(\nu + \zeta|s-t|^{\gamma} + 1\big)} + \frac{\delta 1\{\mathbf{x} = \mathbf{y}\}}{\nu' + |s-t|^{\gamma}} \quad (16)$$

with ϵ real, ϕ, ζ, α and δ nonnegative, ν and ν' positive and $\gamma \in (0, 2)$. The approximate likelihoods for these and many other models were maximized using the routine nlm in R. These two models (or minor variants of them) yielded higher likelihoods than similar extensions of (3), (8) or of separable models, and the maximized approximate loglikelihood for (16) is larger than that for (15) by 15.3. The parameter estimates for (15) are $(\widehat{\phi}, \widehat{a}, \widehat{\alpha}, \widehat{c}, \widehat{\gamma}, \widehat{\epsilon}, \widehat{\delta}, \widehat{a'}, \widehat{\alpha'}) = (0.581, 0.876, 0.817, 6.94 \times 10^{-5}, 0.773, -0.0661, 0.138, 4.47, 0.479)$ and for (16) are $(\widehat{\phi}, \widehat{\nu}, \widehat{\zeta}, \widehat{\gamma}, \widehat{\alpha}, \widehat{\epsilon}, \widehat{\delta}, \widehat{\nu'}) = (1.25, 0.921, 0.499, 1.43, 0.00185, -0.0584, 0.0199, 0.689)$. Given the approximate nature of the likelihood and the considerable amount of informal model selection involved in arriving at these two models, it would be imprudent to read too much into the moderate advantage of (16) over (15). Figure 2 shows fitted variograms for ΔZ for each of these two models. Both models fit the spatial variogram well. Model (16) fits the temporal variogram and the average of the time-lagged spatial variograms better than (15), but (15) fits the asymmetries better. The models fitted in Haslett and Raftery (1989) and Gneiting (2002) are fully symmetric and hence produce fitted values in Figures 2d and f that are identically 0. Figures 2b, c and e show greater variation at and between coastal locations than inland locations, a feature that none of the models considered here can fit. Allowing different variances at different sites can improve the fit considerably. For example, by extending (16) to have two overall scale parameters, one for the four most coastal sites (Roche's Point, Valentia, Belmullet and Malin Head) and another for the other seven, the maximized approximate loglikelihood increases by 145.7 units and (results not shown) produces substantially better agreement between the observed and fitted variograms in Figures 2b, c and e. The vector autoregressive models used by de Luna and Genton (2005) allow for asymmetry in space-time and spatial nonstationarity, but these models do not naturally extend to locations other than the observation sites.

If instead of (14), one uses the flat Earth approximation (13) in (15) and (16), the maximized approximate loglikelihood decreases by 18.8 and 19.7 units, respectively. Thus, the difference in loglikelihoods between (15) and (16) is nearly unchanged by how we define distances on the Earth's surface, but both models fit moderately better when using the approach that does not distort great circle distances.

8. DISCUSSION

A number of recent papers have derived new space-time covariance functions and this work adds some further ones. However, to develop appropriate models for specific spatial-temporal processes, one needs more than just a laundry list of potential models. Three additional critical aspects of model-building are theoretical frameworks for thinking about the differences in models, methods for fitting and comparing models, and diagnostics for assessing the adequacy of models. These issues are particularly critical with regards to the nature of spatial-temporal interactions of processes. Sections 2–6 consider some theoretical bases for describing spatial-temporal interactions implied by various covariance functions, including smoothness away from the origin, space-time asymmetry and Markov properties, although this work clearly only scratches the surface of this problem.

For fitting and comparing models, it is natural to use likelihood-based (either Bayesian or frequentist) methods where possible. However, even for Gaussian processes, exact likelihoods are often impossible to compute for large space-time datasets, creating the need for methods of approximating likelihoods and for developing inferential procedures that take account of the fact that the exact likelihood is not available. For the Irish wind data, the lack of missing values and stationarity in time imply that the modeled covariance matrices for any set of consecutive days have a block

Toëplitz structure. The SLICOT Library (http://www.win.tue.nl/niconet/NIC2/slicot.html) provides Fortran 77 routines for efficiently computing, for example, the Cholesky decompositions of block Toëplitz matrices needed for the approximate likelihood calculations reported on in Section 7.

For assessing model adequacy, spatial-temporal datasets of long duration provide opportunities not available for purely spatial data. In the purely spatial context, Stein (1999) argues that empirical variograms can be highly unreliable and hence comparing empirical variograms and maximum likelihood estimates of the variogram may not be a good way of assessing model adequacy. However, for the Irish wind data, the empirical variogram estimates are averages over 18 years of daily data and are quite accurate, which can be seen, for example, by comparing estimates based on the first 9 and last 9 years of data. Thus, comparison of the fitted and empirical variograms can be a powerful diagnostic for space-time covariance functions when one has a fixed monitoring network of long duration. In particular, I believe the disagreements between the fitted models and empirical variogram estimates in Figure 2 are a sign of model misspecification. When making comparisons between empirical and fitted variograms for spatially sparse monitoring networks, it is critical to remove any purely spatial component of the variation, since these effects will be poorly estimated no matter how long the data record. This problem was solved here by taking first differenes in time at each monitoring site. However, one would need to develop a separate model for this purely spatial component if one wanted to predict winds at a site without a monitor.

The possibility of space-time asymmetries in covariance functions leads to important and fascinating challenges in modeling and diagnostics. Figures 2d and f, which plot empirical and fitted asymmetries as a function of the difference in longitude between sites, proved useful for the Irish wind data, but one could not expect the difference in latitude between sites to play so small a rule in the asymmetry for a region much larger than Ireland. Thus, there is a need for visualization methods for variograms that take account of how they jointly depend on differences in latitude, longitude and time. Finally, there is clear scope for models and methods that reflect the Earth's spherical shape and its rotation about a fixed axis.

APPENDIX: PROOFS

Proof of Proposition 1. The conditions on K_m imply $\mathrm{var}\{Z_m(0,\epsilon) - Z_m(0,0)\} \sim -2C_1(0)\epsilon^{\alpha_1}$ as $\epsilon \downarrow 0$. Now

$$\mathrm{var}\{Z_m(0,\epsilon) - Z_m(0,0)\}\rho_\epsilon^m(t,s) = \sum_{j=1}^{p} C_j(s)\{|t+\epsilon|^{\alpha_j} - 2|t|^{\alpha_j} + |t-\epsilon|^{\alpha_j}\}$$
$$+ R_s(t+\epsilon) - 2R_s(t) + R_s(t-\epsilon).$$

For two functions f and g on a domain A, write $f \ll g$ if there exists $C < \infty$ such that $|f(x)| \le Cg(x)$ for all $x \in A$. Then for fixed s, $R_s(t+\epsilon) - 2R_s(t) + R_s(t-\epsilon) \ll \epsilon^2$ for $t \in \mathbb{R}$ and $\epsilon > 0$, and if $p = 1$, (1) follows immediately. If $p > 1$, then for $|t| \le 2\epsilon$, $|t+\epsilon|^{\alpha_j} - 2|t|^{\alpha_j} + |t-\epsilon|^{\alpha_j} \ll \epsilon^{\alpha_j}$, and for $|t| > 2\epsilon$,

$$|t+\epsilon|^{\alpha_j} - 2|t|^{\alpha_j} + |t-\epsilon|^{\alpha_j} = |t|^{\alpha_j}\left\{\left(1 - \frac{\epsilon}{t}\right)^{\alpha_j} - 2 + \left(1 + \frac{\epsilon}{t}\right)^{\alpha_j}\right\} \ll |t|^{\alpha_j}\left(\frac{\epsilon}{|t|}\right)^{\alpha_j}$$

and (1) follows. Since, as $\epsilon \downarrow 0$, $\mathrm{var}\{Z_m(0,\epsilon) - Z_m(0,0)\} \sim 2C_1(0)\epsilon^{\alpha_1}$ and $\{|t+\epsilon|^{\alpha_1} - 2|t|^{\alpha_1} + |t-\epsilon|^{\alpha_1}\}/(2\epsilon^{\alpha_1}) \to 1$ if $t = 0$ and to 0 otherwise, (2) follows.

Differentiability of (3). Assume $\gamma(\mathbf{x}) \neq 0$ for $\mathbf{x} \neq \mathbf{0}$, which only excludes processes that are exactly periodic in some direction. First consider the differentiability in t of (3). From (5.1) in Ma (2003), the covariance function in (3) equals $\frac{\pi}{2}\int_0^\infty \cos(\alpha t w)\exp\{-(1+w^2)\gamma(\mathbf{x})\}(1+w^2)^{-1}dw$. For all $\mathbf{x} \neq \mathbf{0}$, this integral is the cosine transform of a function with finite moments of all orders, so it is infinitely differentiable with respect to t. Since $K(\mathbf{0}, t) = 2e^{-\alpha|t|}$, K is infinitely differentiable in t for $\mathbf{x} = \mathbf{0}$ and $t \neq 0$. Next consider $D^{(\mathbf{m},k)}K(\mathbf{x}, t)$ for $\mathbf{m} \neq \mathbf{0}$. For $\mathbf{x} \neq \mathbf{0}$,

$$\frac{\partial}{\partial x_j}K(\mathbf{x}, t) = -\frac{2}{\pi^{1/2}}\exp\left\{\gamma(\mathbf{x}) - \frac{\alpha^2 t^2}{4\gamma(\mathbf{x})}\right\}\gamma(\mathbf{x})^{-1/2}\frac{\partial}{\partial x_j}\gamma(\mathbf{x}).$$

Thus, for $\mathbf{x} \neq \mathbf{0}$, $D^{(\mathbf{m},k)}K(\mathbf{x}, t)$ exists and is of the form

$$\exp\left\{\gamma(\mathbf{x}) - \frac{\alpha^2 t^2}{4\gamma(\mathbf{x})}\right\}\sum_j \lambda_j t^{\beta_j}\gamma(\mathbf{x})^{-\alpha_j - 1/2}\prod_\ell \{D^{(\mathbf{k}_{j\ell})}\gamma(\mathbf{x})\}^{\delta_{j\ell}},$$

where the sum and product are both over a finite number of terms and the α_js, β_js and $\delta_{j\ell}$s are nonnegative integers. Finally, consider $\mathbf{x} = \mathbf{0}$ and $t \neq 0$. Because $\exp\left(-\frac{\alpha^2 t^2}{4\gamma(\mathbf{x})}\right)$ tends to 0 very quickly as $\gamma(\mathbf{x}) \downarrow 0$, $D^{(\mathbf{m},k)}K(\mathbf{0}, t) = 0$ for all $\mathbf{m} \neq \mathbf{0}$ under mild conditions on γ. For example, the following conditions suffice: there exist positive C and β such that $\gamma(\mathbf{x}) > C|\mathbf{x}|^\beta$ in some neighborhood of the origin and, for any given $\mathbf{n} \neq \mathbf{0}$ with nonnegative integer components, there exist constants B and τ and positive R (possibly depending on $\mathbf{n}$), such that $|D^{(\mathbf{n})}\gamma(\mathbf{x})| < B|\mathbf{x}|^\tau$ for $0 < |\mathbf{x}| < R$.

Proof of Proposition 2. Suppose for an integrable function g on $\mathbb{R}^d$, $(\partial/\partial w_j)g(\mathbf{w})$ exists and is integrable. If $x_j \neq 0$, then for almost every $(w_1, \cdots, w_{j-1}, w_{j+1}, \cdots, w_d)'$,

$$\int_{\mathbb{R}} g(\mathbf{w}) e^{i w_j x_j} dw_j = -\frac{1}{i x_j} \int_{\mathbb{R}} \left\{ \frac{\partial}{\partial w_j} g(\mathbf{w}) \right\} e^{i w_j x_j} dw_j.$$

Thus,

$$\int_{\mathbb{R}^d} g(\mathbf{w}) e^{i\mathbf{w}'\mathbf{x}} \mathbf{dw} = -\frac{1}{i x_j} \int_{\mathbb{R}^d} \left\{ \frac{\partial}{\partial w_j} g(\mathbf{w}) \right\} e^{i\mathbf{w}'\mathbf{x}} \mathbf{dw}.$$

Applying this result repeatedly and using $D^{\mathbf{q}} f$ integrable for all $\mathbf{q} \leq \mathbf{k}$ gives

$$\int_{\mathbb{R}^d} f(\mathbf{w}) e^{i\mathbf{w}'\mathbf{x}} \mathbf{dw} = \frac{1}{(-i\mathbf{x})^{\mathbf{k}}} \int_{\mathbb{R}^d} \{D^{\mathbf{k}} f(\mathbf{w})\} e^{i\mathbf{w}'\mathbf{x}} \mathbf{dw}.$$

Since $\mathbf{w}^{\mathbf{m}} D^{\mathbf{k}} f(\mathbf{w})$ is integrable, for all $\mathbf{j} \leq \mathbf{m}$,

$$D^{\mathbf{j}} \int_{\mathbb{R}^d} \{D^{\mathbf{k}} f(\mathbf{w})\} e^{i\mathbf{w}'\mathbf{x}} \mathbf{dw} = \int_{\mathbb{R}^d} (i\mathbf{w})^{\mathbf{j}} \{D^{\mathbf{k}} f(\mathbf{w})\} e^{i\mathbf{w}'\mathbf{x}} \mathbf{dw}$$

(Stein and Weiss 1971, p. 5). Proposition 2 follows by repeated application of the product rule for differentiation and elementary counting arguments.

Proof of Corollary 1. Let h be an infinitely differentiable function on $[0, \infty)$ such that $h(t) = 0$ on $[0, R]$ and $h(t) = 1$ on $[R + 1, \infty)$. Then Proposition 2 applies to $h(|\mathbf{w}|)f(\mathbf{w})$ and $\{1 - h(|\mathbf{w}|)\}f(\mathbf{w})$ has bounded support, so its Fourier transform is infinitely differentiable and, hence, the Fourier transform of f has the required derivatives.

Proof of Proposition 4. The integrability of f holds if and only if

$$\int_0^\infty \int_0^\infty (1 + r_1^{2\alpha_1} + r_2^{2\alpha_2})^{-\nu} r_1^{d_1 - 1} r_2^{d_2 - 1} dr_1\, dr_2 < \infty,$$

which one can verify holds for α_1, α_2 positive if and only if $d_1/(\alpha_1\nu) + d_2/(\alpha_2\nu) < 2$.

Write $\mathbf{w}_j = (w_{j1}, \ldots, w_{jd_j})'$ for $j = 1, 2$, and, without loss of generality, we will show $|\mathbf{w}|^m (\partial^k/\partial w_{11}^k) f(\mathbf{w})$ is integrable for k sufficiently large. Using the chain rule for higher derivatives of composite functions (Gradshteyn and Ryzhik 2000, 0.430.2), for $k > 2\alpha_1$, $(\partial^k/\partial w_{11}^k)f(\mathbf{w})$ is a linear combination of terms like

$$\{P_1(|\mathbf{w}_1|^2) + P_2(|\mathbf{w}_2|^2)\}^{-k-\nu+a_0} \prod_{\ell=1}^{2\alpha_1} \left\{ \frac{\partial^\ell}{\partial w_{11}^\ell} P_1(|\mathbf{w}_1|^2) \right\}^{a_\ell} \tag{17}$$

for which $\sum_{j=0}^{2\alpha_1} a_j = \sum_{j=0}^{2\alpha_1} j a_j = k$ with $a_0, \ldots, a_{2\alpha_1}$ nonnegative integers. We do not have to consider $\ell > 2\alpha_1$ in (17) because then $(\partial^\ell/\partial w_{11}^\ell)P_1(|\mathbf{w}_1|^2) = 0$. To obtain $\sum_{j=0}^{2\alpha_1} j a_j = k$, we must have $a_0 \le k\{1 - 1/(2\alpha_1)\}$, or $k - a_0 \ge k/(2\alpha_1)$. Now, for a given m, (17) times $|\mathbf{w}|^m$ will be integrable if $|\mathbf{w}|^m(1+|\mathbf{w}_1|)^{k(2\alpha_1-1)-2a_0\alpha_1}(1+|\mathbf{w}_1|^{2\alpha_1}+|\mathbf{w}_2|^{2\alpha_2})^{a_0-k-\nu}$ is integrable, and if $k > m - 2\alpha_1\nu + d_1$,

$$\begin{aligned}
&\int_{\mathbb{R}^{d_1+d_2}} |\mathbf{w}|^m(1+|\mathbf{w}_1|)^{k(2\alpha_1-1)-2a_0\alpha_1}(1+|\mathbf{w}_1|^{2\alpha_1}+|\mathbf{w}_2|^{2\alpha_2})^{a_0-k-\nu}\mathbf{dw} \\
&\ll 1 + \int_1^\infty \int_1^\infty \frac{(r_1^m + r_2^m) r_1^{k(2\alpha_1-1)-2a_0\alpha_1}}{r_1^{2\alpha_1(k+\nu-a_0)} + r_2^{2\alpha_2(k+\nu-a_0)}} r_1^{d_1-1} r_2^{d_2-1} dr_1 dr_2 \\
&\ll 1 + \int_1^\infty \left\{ \int_1^{r_2^{\alpha_2/\alpha_1}} \frac{(r_1^m + r_2^m) r_1^{k(2\alpha_1-1)-2a_0\alpha_1}}{r_2^{2\alpha_2(k+\nu-a_0)}} r_1^{d_1-1} r_2^{d_2-1} dr_1 \right. \\
&\qquad \left. + \int_{r_2^{\alpha_2/\alpha_1}}^\infty \frac{(r_1^m + r_2^m) r_1^{k(2\alpha_1-1)-2a_0\alpha_1}}{r_1^{2\alpha_1(k+\nu-a_0)}} r_1^{d_1-1} r_2^{d_2-1} dr_1 \right\} dr_2 \\
&\ll 1 + \int_1^\infty r_2^{d_2-1-2\alpha_2\nu} \left\{ r_2^{m-2\alpha_2(k-a_0)} + r_2^{m+(\alpha_2/\alpha_1)(d_1-k)} + r_2^{(\alpha_2/\alpha_1)(m-k+d_1)} \right\} dr_2,
\end{aligned}$$

which, using $k - a_0 \ge k/(2\alpha_1)$, is finite for k sufficiently large. Proposition 4 follows.

Proof of (7). Using Theorem 17 in Stein and Weiss (1971),

$$K(\mathbf{x},t) = aK_0(\mathbf{x},t) - b\sum_{j=1}^d z_j \frac{\partial^2}{\partial x_j\, \partial t} K_0(\mathbf{x},t) - c_1 \sum_{j=1}^d \frac{\partial^2}{\partial x_j^2} K_0(\mathbf{x},t) - c_2 \frac{\partial^2}{\partial t^2} K_0(\mathbf{x},t).$$

Next,

$$\begin{aligned}
\sum_{j=1}^d \frac{\partial^2}{\partial x_j^2} K_0(\mathbf{x},t) &= \sum_{j=1}^d \left[\overline{K}_0^{(2,0)}(|\mathbf{x}|,t) \frac{x_j^2}{|\mathbf{x}|^2} + \overline{K}_0^{(1,0)}(|\mathbf{x}|,t) \left\{ \frac{|\mathbf{x}|^2 - x_j^2}{|\mathbf{x}|^3} \right\} \right] \\
&= \overline{K}_0^{(2,0)}(|\mathbf{x}|,t) + \frac{d-1}{|\mathbf{x}|} \overline{K}_0^{(1,0)}(|\mathbf{x}|,t)
\end{aligned}$$

and

$$\sum_{j=1}^{d} z_j \frac{\partial^2}{\partial x_j\, \partial t} K_0(\mathbf{x}, t) = \frac{\mathbf{x}'\mathbf{z}}{|\mathbf{x}|} \overline{K}_0^{(1,1)}(|\mathbf{x}|, t),$$

where the conditions on Ψ guarantee the existence of the relevant derivatives of K_0. Equation (7) follows.

Proof of Proposition 5. Let us first prove that Z Markov implies its covariance function is of the form (11). Write $K(\mathbf{x}, t) = \int e^{i\mathbf{x}'\mathbf{w}+ivt} G(\mathbf{dw}, dv)$ for the spectral representation of K. Because G is a positive finite measure, one can define the measure F given by $F(\mathbf{dw}) = \int_v G(\mathbf{dw}, dv)$. By the Radon-Nikodym Theorem, $G(\mathbf{dw}, dv) = H_{\mathbf{w}}(dv) F(\mathbf{dw})$, where $H_{\mathbf{w}}(\cdot)$ is a positive, finite measure for F-a.e. $\mathbf{w}$. Set $\eta_{\mathbf{w}}(t) = \int e^{ivt} H_{\mathbf{w}}(dv)$, which is well-defined for F-a.e. $\mathbf{w}$.

Let $\mathcal{L}^2(G)$ be the closed real linear hull of $e^{i\mathbf{x}'\mathbf{w}+ivt}$ with respect to G and $\mathcal{H}(G)$ the closed real linear hull of $Z(\mathbf{x}, t)$ with respect to G. Then $e^{i\mathbf{x}'\mathbf{w}+ivt} \longleftrightarrow Z(\mathbf{x}, t)$ defines a Hilbert space isometry between $\mathcal{L}^2(G)$ and $\mathcal{H}(G)$. Let $\mathcal{P}_t$ be the orthogonal projection operator in $\mathcal{L}^2(G)$ onto the closed real linear hull of $e^{i\mathbf{x}'\mathbf{w}+ivs}$ for $s \le t$, $\mathbf{x} \in \mathbb{R}^d$. It suffices to show that for all $\mathbf{x} \in \mathbb{R}^d$ and all s, t nonnegative, $\mathrm{cov}[Z(\mathbf{0}, t) - E\{Z(\mathbf{0}, t) \mid Z(\cdot, 0)\}, Z(-\mathbf{x}, -s)] = 0$, or in terms of elements of $\mathcal{L}^2(G)$, $\int e^{i\mathbf{x}'\mathbf{w}+ivs}\big(e^{ivt} - \mathcal{P}_0 e^{ivt}\big) G(\mathbf{dw}, dv) = 0$. The Markov property implies $\mathcal{P}_0 e^{ivt}$ is in the closed real linear hull of $e^{i\mathbf{x}'\mathbf{w}}$ (i.e., it does not depend on v), so define $A(t; \mathbf{w}) = \mathcal{P}_0 e^{ivt}$. Now, $\lim_{s\to 0} A(t+s; \mathbf{w}) = A(t; \mathbf{w})$ in $\mathcal{L}^2(F)$ because $Z(\mathbf{x}, t)$ is mean square continuous by hypothesis. Thus, $A(t; \mathbf{w})$ is jointly measurable in $t, \mathbf{w}$. Next, for s, t nonnegative, using the stationarity of Z,

$$\begin{aligned} A(t+s; \mathbf{w}) &= \mathcal{P}_0 e^{iv(t+s)} = \mathcal{P}_0 \mathcal{P}_s e^{iv(t+s)} \\ &= \mathcal{P}_0(e^{ivs} A(t; \mathbf{w})) = A(s; \mathbf{w}) A(t; \mathbf{w}) \end{aligned}$$

and it follows that $A(t; \mathbf{w}) = e^{-\alpha(\mathbf{w})t}$ for all $t \ge 0$ for some measurable function α with real and imaginary parts β and ϕ, respectively. Since $\overline{A(t; -\mathbf{w})} = A(t; \mathbf{w})$, β must be even and ϕ odd. Next, K even implies $A(-t; -\mathbf{w}) = A(t; \mathbf{w})$, so for $t \le 0$, $A(t; \mathbf{w}) = \overline{A(-t; \mathbf{w})} = e^{t\beta(\mathbf{w}) - it\phi(\mathbf{w})}$ and it follows that K is of the form (9) with β even and ϕ odd. Finally, to see that β must be nonnegative, note that if $\beta(\mathbf{w}) < 0$, $\limsup_{t\to\infty} \left|\int_v e^{itv} H_{\mathbf{w}}(dv)\right| = \limsup_{t\to\infty} e^{-|t|\beta(\mathbf{w})} = \infty$, which can only happen for a set of $\mathbf{w}$ with measure 0 under F, so that one can assume β nonnegative everywhere without loss of generality.

To prove the converse, if K is of the form (9), then one needs to show $\mathcal{P}_0 e^{ivt} \in \mathcal{L}^2(F)$. Since $K(\mathbf{x}, t) = \int e^{i\mathbf{w}'\mathbf{x}} \eta_{\mathbf{w}}(t) F(\mathbf{dw})$, it follows from (9) that $\eta_{\mathbf{w}}(t) = e^{-|t|\beta(\mathbf{w}) - it\phi(\mathbf{w})}$ F-a.e. Then for s, t

nonnegative,

$$\begin{aligned}
\int \left(e^{ivt} - e^{-|t|\beta(\mathbf{w})-it\phi(\mathbf{w})}\right) e^{i\mathbf{w}'\mathbf{x}+ivs} G(\mathbf{dw}, dv) \\
= K(\mathbf{x}, s+t) - \int e^{-|t|\beta(\mathbf{w})-it\phi(\mathbf{w})+i\mathbf{w}'\mathbf{x}} \eta_{\mathbf{w}}(s) F(\mathbf{dw}) \\
= K(\mathbf{x}, s+t) - \int e^{-|s+t|\beta(\mathbf{w})-i(s+t)\phi(\mathbf{w})+i\mathbf{w}'\mathbf{x}} F(\mathbf{dw}),
\end{aligned}$$

which is 0 by (9). Thus, $\mathcal{P}_0 e^{ivt} = \eta_{\mathbf{w}}(t)$, which, for β even and nonnegative and ϕ odd, is in $\mathcal{L}^2(F)$ as required.

REFERENCES

Abramowitz, M., and Stegun, I. (1965), *Handbook of Mathematical Functions* (ninth ed.), New York: Dover.

Brown, P. E., Karesen, K. F., Roberts, G. O., and Tonellato, S. (2000), "Blur-Generated Non-Separable Space-Time Models," *Journal of the Royal Statistical Society*, Ser. B, 62, 847–860.

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

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

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

Christensen, R. (1996), *Plane Answers to Complex Questions: The Theory of Linear Models* (second ed.), New York: Springer.

Clerc, M., and Mallat, S. (2003), "Estimating Deformations of Stationary Processes," *Annals of Statistics*, 31, 1772–1821.

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

Cressie, N., and Huang, H.-C. (1999), "Classes of Nonseparable, Spatiotemporal Stationary Covariance Functions," *Journal of the American Statistical Association*, 94, 1330–1340.

de Iaco, S., Myers, D. E., and Posa, D. (2001), "Space-Time Analysis Using a General Product-Sum Model," *Statistics & Probability Letters*, 52, 21–28.

de Iaco, S., Myers, D. E., and Posa, D. (2002), "Nonseparable Space-Time Covariance Models: Some Parametric Families," *Mathematical Geology*, 34, 23–42.

de Iaco, S., Myers, D. E., and Posa, D. (2003), "The Linear Coregionalization Model and the Product-Sum Space-Time Variogram," *Mathematical Geology*, 35, 25–38.

de Luna, X., and Genton, M. G. (2005), "Predictive Spatio-Temporal Models for Spatially Sparse Environmental Data," in press, *Statistica Sinica*.

Fuentes, M. (2002), "Spectral Methods for Nonstationary Spatial Processes," *Biometrika*, 89, 197–210.

Fuentes, M., and Smith, R. (2001), "A New Class of Nonstationary Spatial Models," unpublished manuscript, www4.stat.ncsu.edu/~fuentes/sdf.ps.

Gneiting, T. (1999), "Correlation functions for atmospheric data analysis," *Quarterly Journal of the Royal Meteorological Society*, 125, 2449–2464.

Gneiting, T. (2002), "Nonseparable, Stationary Covariance Functions for Space-Time Data," *Journal of the American Statistical Association*, 97, 590–600.

Gradshteyn, I. S., and Ryzhik, I. M. (2000), *Table of Integrals, Series, and Products* (sixth ed.), San Diego: Academic Press.

Handcock, M. S., and Stein, M. L. (1993), "A Bayesian Analysis of Kriging," *Technometrics*, 35, 403–410.

Handcock, M. S., and Wallis, J. R. (1994), "An Approach to Statistical Spatial-Temporal Modeling of Meteorological Fields" (with comments), *Journal of the American Statistical Association*, 89, 368–390.

Hartfield, M. I., and Gunst, R. F. (2003), "Identification of Model Components for a Class of Continuous Spatiotemporal Models," *Journal of Agricultural, Biological, and Environmental Statistics*, 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), *Applied Statistics*, 38, 1–50.

Heine, V. (1955), "Models for Two-Dimensional Stationary Stochastic Processes," *Biometrika*, 42, 170–178.

Higdon, D. (1998), "A Process-Convolution Approach to Modeling Temperatures in the North-Atlantic Ocean," *Environmental and Ecological Statistics*, 5, 173–190.

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), New York: Springer-Verlag, pp. 289–298.

Kyriakidis, P. C., and Journel, A. G. (1999), "Geostatistical Space-Time Models: A Review," *Mathematical Geology*, 31, 651–684.

Ma, C. (2003), "Families of Spatio-Temporal Stationary Covariance Models," *Journal of Statistical Planning and Inference*, 116, 489–501.

Matérn, B. (1960), *Spatial Variation*, Meddelanden från Statens Skogsforskningsinstitut, (Vol. 49, No. 5), Stockholm: Almaenna Foerlaget. Second edition (1986), Berlin: Springer-Verlag.

Nott, D. J., and Dunsmuir, W. T. M. (2002), "Estimation of Nonstationary Spatial Covariance Structure," *Biometrika*, 89, 819–829.

Sampson, P. D., and Guttorp, P. (1992), "Nonparametric Estimation of Nonstationary Spatial Covariance Structure," *Journal of the American Statistical Association*, 87, 108–119.

Stein, E. M., and Weiss, G. (1971), *Introduction to Fourier Analysis on Euclidean Spaces*, Princeton, NJ: Princeton University Press.

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

Stein, M. L. (2002), "The Screening Effect in Kriging," *Annals of Statistics*, 30, 298–323.

Stein, M. L., Chi, Z. and Welty, L. J. (2004), "Approximating Likelihoods for Large Spatial Datasets," in press, *Journal of the Royal Statistical Society*, Ser. B.

Vecchia, A. V. (1988), "Estimation and Identification for Continuous Spatial Processes," *Journal of the Royal Statistical Society*, Ser. B, 50, 297–312.

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

Figure captions

Figure 1. Contour plot for correlation function $\{4\mathcal{M}_{1/2}\big((x_1^2+x_2^2+t^2)^{1/2}\big)-2x_1t\mathcal{M}_{-1/2}\big((x_1^2+x_2^2+t^2)^{1/2}\big)\}/\{4\mathcal{M}_{1/2}(0)\}$ for $x_2=0$.

Figure 2. Variograms for $\Delta Z(\mathbf{x}, t+1)$ for Irish wind data. In each figure, + indicates a variogram value for coastal sites, $*$ indicates a coastal and an inland site and $\circ$ indicates two inland sites. The $\times$ indicates a fitted value under model (15) and ∇ indicates (16). In Figure b, horizontal offsets within each day are proportional to latitude, with more northerly sites to the right. For improved legibility, Figure b omits the temporal variogram at lag 0, which is necessarily 0 for both empirical and fitted variograms. Figures d and f plot asymmetries for all pairs of sites for which $\mathbf{y}$ is east of $\mathbf{x}$.

Figure 1

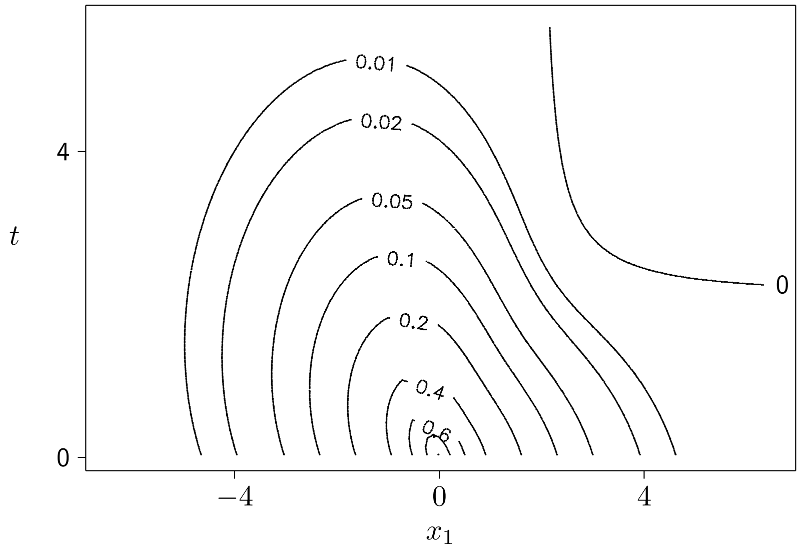
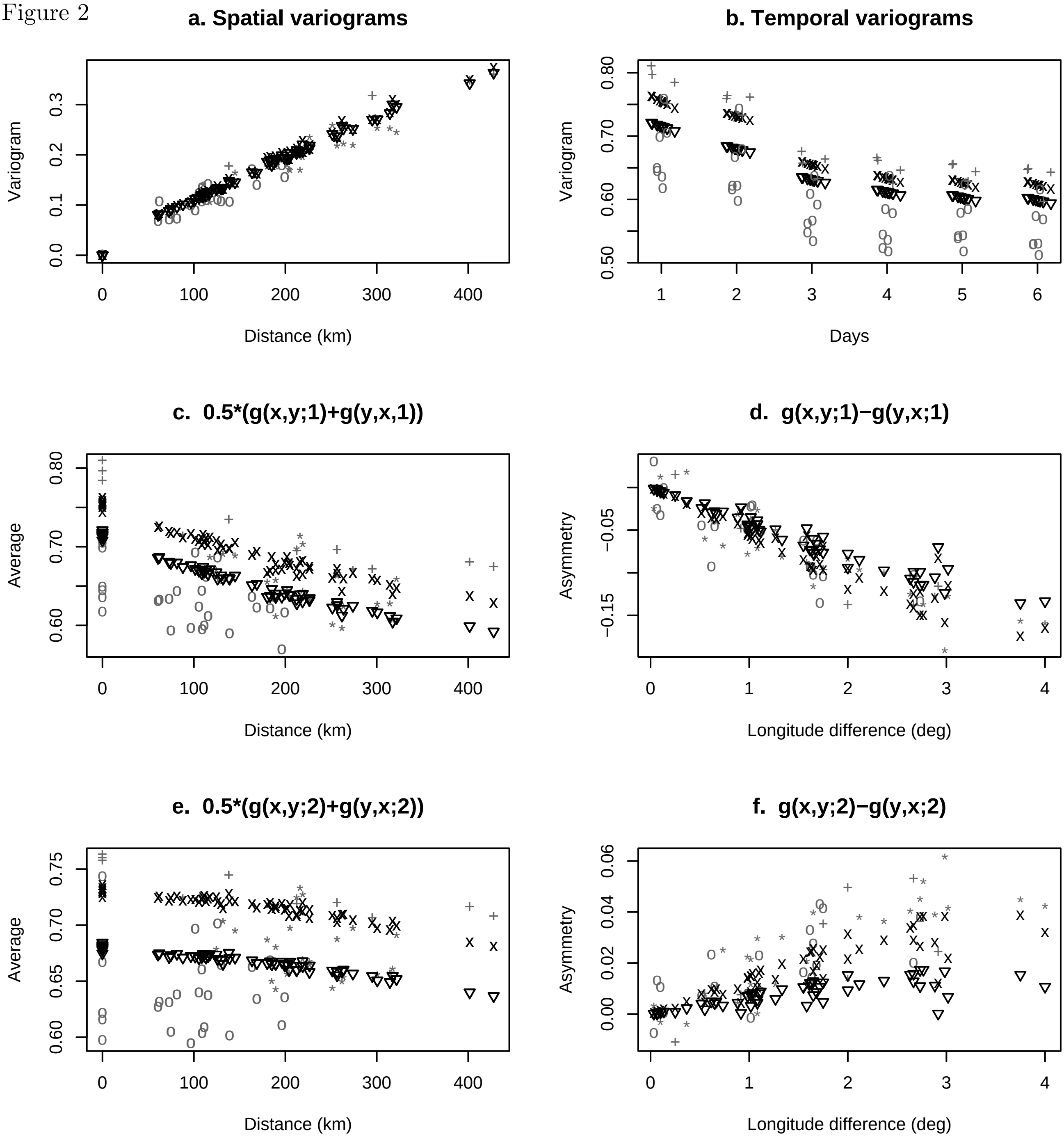


Figure 2


a. Spatial variograms
Variogram
Distance (km)
b. Temporal variograms
Variogram
Days
c. 0.5*(g(x,y;1)+g(y,x,1))
Average
Distance (km)
d. g(x,y;1)−g(y,x;1)
Asymmetry
Longitude difference (deg)
e. 0.5*(g(x,y;2)+g(y,x;2))
Average
Distance (km)
f. g(x,y;2)−g(y,x;2)
Asymmetry
Longitude difference (deg)