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

CISES

Chicago, Illinois USA

TECHNICAL REPORT NO. 4

SPACE-TIME COVARIANCE FUNCTIONS*

M. Stein
May 2003

*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.

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

Space-Time Covariance Functions *

Michael L. Stein

ABSTRACT: A good model for the covariance function of a stationary process in space and time should accurately describe the variances and correlations of all linear combinations of the process. In particular, it does not suffice to find a model that describes the purely temporal covariances and the purely spatial covariances accurately. Rather, it is critical to capture the spatial-temporal interactions as well. This work considers a number of properties of spatial-temporal 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 spatial-temporal 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 might be undesirable in some circumstances. A class of spectral densities is given whose corresponding spatial-temporal covariance functions are infinitely differentiable away from the origin and that allows for 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 correlation between site $\mathbf{x}$ at time t and site $\mathbf{y}$ at time s is different than the correlation 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 spatial-temporal covariance functions are examined and an explicit characterization of all such continuous covariance functions given. Some explicit examples of Markov models are obtained. 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

---

* 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 7. 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. Although data for geophysical processes are often taken at fixed locations and times, the processes themselves usually are defined on continuous index sets in space and time, and the interest here is in modeling the processes themselves and not any particular dataset. Thus, this work addresses models for processes on $\mathbb{R}^d \times \mathbb{R}$, where d is the spatial dimension of the process, generally 1, 2 or 3. Let $Z(\mathbf{x}, t)$ be the value of the random field at location $\mathbf{x}$ and time t. A reasonable starting point is to develop models that describe the first two moments of Z. This work focuses on stationary models; i.e., assume there exists a function K on $\mathbb{R}^d \times \mathbb{R}$ 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 $s, t \in \mathbb{R}$. Of course, K cannot be just any function; it must be positive definite so that variances of all linear combinations of Z are nonnegative. 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), Gneiting (2002), Ma (2003) and Hartfield and Gunst (2003) describe some recent efforts to develop new classes of stationary space-time covariance functions on $\mathbb{R}^d \times \mathbb{R}$. 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 $\mathbb{R}^d \times \mathbb{R}$. Allowing the spatial covariances and the temporal covariances to define the spatial-temporal covariances is a severe restriction, however. Expanding on considerations in Stein (1999, Section 2.11), 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. For example, the space-time covariance function $\exp(-|\mathbf{x}| - |t|)$ is not differentiable at the origin, but is also not differentiable in t for $t = 0$ and any $\mathbf{x}$, nor is it differentiable in the components of $\mathbf{x}$ for $\mathbf{x} = \mathbf{0}$ and any t. 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 $\mathbf{x} = \mathbf{0}$ and $t = 0$. Another goal here is to find models that allow different degrees of smoothness across space than across time. For example, one may want to allow the process to be mean square differentiable with respect to spatial coordinates but not with respect to time. It is difficult to write down directly space-time covariance functions that possess different amounts of smoothness in space than in time and are much smoother away from the origin than at the origin. As in many problems with covariance functions, it is helpful to work in the spectral domain. In particular, 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. Denoting the temporal frequency by v and the spatial frequency by $\mathbf{w}$, these results can be used to show that for $c_1, c_2, \alpha_1, \alpha_2, \nu$, $a_1^2 + a_2^2$ positive and $d/\alpha_1 + 1/\alpha_2 < 2\nu$, the Fourier transform of

$$f(\mathbf{w}, v) = \left\{c_1(a_1^2 + |\mathbf{w}|^2)^{\alpha_1} + c_2(a_2^2 + v^2)^{\alpha_2}\right\}^{-\nu} \tag{1}$$

is positive definite and infinitely differentiable away from the origin. Furthermore, by appropriately choosing α_1, α_2 and ν, one can achieve any combination of degrees of smoothness in space and in time.

There are other properties besides smoothness away from the origin to consider in developing models for space-time covariance functions. One desirable practical property is to have an explicit expression for the space-time covariance function or, if that is not available, at least a fast and accurate algorithm for computing it numerically. Explicit expressions for the Fourier transforms of the spectral densities in (1) are available only in some very limited special cases. If α_1 or α_2

equals 1, then the calculation of $K(\mathbf{x}, t)$ can be reduced to one-dimensional integral transforms, which can be computed numerically fairly readily.

Another property that one may often want for space-time models is isotropy in space. It is not completely obvious how to define spatial isotropy in the space-time context. If $K(\mathbf{x}, 0)$ only depends on $|\mathbf{x}|$, let us say that K is spatially isotropic. If for every t, $K(\mathbf{x}, t)$ depends on $\mathbf{x}$ only through $|\mathbf{x}|$, then because every real, positive definite function must be even, $K(\mathbf{x}, t)$ depends only on $(|\mathbf{x}|, |t|)$. In this case, since $K(\mathbf{x}, t) = K(\mathbf{H}\mathbf{x}, t)$ for every orthogonal matrix $\mathbf{H}$, let us call K directionally invariant.

All spectral densities of the form (1) depend on $(\mathbf{w}, v)$ only through $(|\mathbf{w}|, |v|)$. Consequently, the corresponding covariance functions are all directionally invariant. Directional invariance implies full symmetry in space-time in the sense that $K(\mathbf{x}, t) = K(-\mathbf{x}, t) = K(\mathbf{x}, -t) = K(-\mathbf{x}, -t)$ (Gneiting 2002). Such an assumption is innapropriate for processes in which there is a dominant flow direction over time. Directional flows are common for atmospheric processes that tend to track dominant wind patterns. Section 5 shows how one can generate spatially isotropic models that are not fully symmetric by taking derivatives of directionally invariant 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. From this characterization, one can see that the requirement of stationarity in space-time restricts the temporal dynamics to a damping term and a phase modulation for each spatial frequency. Section 6 gives some specific examples of such covariance functions and investigates their properties in terms of the nature of this damping and phase modulation. In particular, if the damping is sufficiently weak at some frequencies, then 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 12 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) noted a clear lack of full symmetry in the data, a feature which his model did not capture. Using approximate likelihoods, Section 7 considers a number of models not possessing full symmetry and shows that they fit the data substantially better than the model Gneiting (2002) used.

Proofs of all propositions are given in the Appendix.

2. PROPERTIES OF COVARIANCE FUNCTIONS THAT ARE NOT SMOOTH AWAY FROM THE ORIGIN

Stein (1999, p. 30 and p. 52) criticizes covariance functions that are not sufficiently smooth away from the origin. This section revisits some of the examples in Stein (1999) and shows how covariance functions that are not sufficiently smooth away from the origin can imply a kind of discontinuity in correlations of certain linear combinations of the random field.

Let us start by considering the triangular covariance function, $K(t) = (1 - |t|)^+$, which is positive definite on $\mathbb{R}$. This function is not differentiable at $t = \pm 1$. Suppose Z is a stationary process on $\mathbb{R}$ with covariance function $(1-|t|)^+$. Define $\rho_\epsilon(t) = \text{corr}\{Z(\epsilon) - Z(0), Z(t+\epsilon) - Z(t)\}$. If $0 < \epsilon < 1/2$, then for $|t| < \epsilon$, $\rho_\epsilon(t) = 1 - |t|/\epsilon$, for $||t| - 1| < \epsilon$, $\rho_\epsilon(t) = (||t| - 1|/\epsilon - 1)/2$ and for other t, $\rho_\epsilon(t) = 0$. Thus, for all t, $\rho(t) = \lim_{\epsilon \to 0} \rho_\epsilon(t)$ exists, $\rho(0) = 1$, $\rho(\pm 1) = -1/2$ and $\rho(t) = 0$ otherwise. The discontinuity in ρ at 0 is not troubling and is, indeed, inevitable for any model for which $K'(0^+) < 0$. The discontinuity at ± 1 is a concern and is a direct result of the lack of differentiability of K at ± 1. Indeed, if a covariance function K is continuous, has a continuous and bounded derivative on $(0, \infty)$ and $K'(0^+) < 0$, then it is possible to show that $\rho(t)$ exists for all t and is 0 for all $t \neq 0$.

Next consider a random field Z on $\mathbb{R}^2$ with covariance function $K(t, s) = e^{-|t|-|s|}$. This separable covariance function is criticized in Stein (1999, Section 2.11) for the severe dependence of variances of certain linear combinations of the random field on the choice of axes. In fact, the problem with this covariance function is very similar to the problem with the triangular covariance function. Define $\rho_\epsilon(t, s) = \text{corr}\{Z(\epsilon, 0) - Z(0, 0), Z(t+\epsilon, s) - Z(t, s)\}$. For every s, $\lim_{\epsilon \to 0} \rho_\epsilon(t, s) = \rho(t, s)$ exists, $\rho(t, s) = 0$ whenever $t \neq 0$ and $\rho(0, s) = e^{-|s|}$. Thus, for all s, ρ is not continuous at $(0, s)$. Again, the discontinuity is only troubling for $s \neq 0$ and can be traced to the lack of smoothness of K at $(0, s)$. Specifically, if K is continuous, $\{\partial K(t,0)/\partial t\}|_{t=0^+} \in (-\infty, 0)$ and K has continuous and bounded first partial derivatives everywhere except at the origin, then $\rho(t,s) = 0$ everywhere except the origin. When s and t are both axes in space, such a dramatic dependence of correlations on the locations of the observations would be hard to justify for natural processes. When t represents time and s space, a ridge in the covariance function along the time and/or space axis may or may not be appropriate. What is certainly the case is that models with such ridges are different in an important way from models not having these ridges and that the decision on whether to use a model with ridges should be made consciously and in light of the evidence in the data.

This discontinuity of limiting correlations occurs for a broad class of covariance functions with a "ridge" along one of its axes. 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*

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

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

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

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

$$\rho^m(t,s) = \begin{cases} C_1(s)/C_1(0), & t = 0 \\ 0, & t \neq 0. \end{cases} \tag{3}$$

□

Many separable covariance functions $K(t,s) = K_1(t)K_2(s)$ satisfy the conditions of Proposition 1, 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, (3) 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) argues that such covariance functions are physically unrealistic because they imply implausibly smooth processes.

Cressie and Huang (1999) and Gneiting (2002) describe approaches of generating nonseparable space-time covariance functions. However, all of the examples in Cressie and Huang (1999) are analytic along either the spatial or temporal coordinates. Gneiting (2002) shows that all functions of the form

$$K(\mathbf{x},t) = \frac{\sigma^2}{\psi(t^2)^{d/2}}\varphi\left(\frac{|\mathbf{x}|^2}{\psi(t^2)}\right),$$

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 $\mathbb{R}^d \times \mathbb{R}$. Gneiting (2002) gives examples of possible φ and ψ such that K is not analytic in $\mathbf{x}$ or t. However, in these cases, 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) do not remove one of the potentially undesirable features of separable covariance functions.

For $d = 1$, Heine (1955) gives the following covariance function that is infinitely differentiable away from the origin and for which the process is smoother in space than in time:

$$K(x,t) = e^{-\alpha|x|}\operatorname{erfc}\left(\beta|t|^{1/2} - \frac{\alpha|x|}{2\beta|t|^{1/2}}\right) + e^{\alpha|x|}\operatorname{erfc}\left(\beta|t|^{1/2} + \frac{\alpha|x|}{2\beta|t|^{1/2}}\right), \qquad (4)$$

where erfc is the complementary error function and α and β are positive constants. In fact, (4) is the Fourier transform of (1) when $d = 1$, $a_2 = 0$, $\alpha_1 = 2$, $\alpha_2 = \nu = 1$ and a_1, c_1 and c_2 are chosen appropriately. Thus, it follows from Proposition 6 in Section 3 that (4) is infinitely differentiable away from the origin. Ma (2003, (5.3)) generalizes (4) by showing that

$$K(x,t) = e^{-\alpha|x|}\operatorname{erfc}\left(\gamma(t)^{1/2} - \frac{\alpha|x|}{2\gamma(t)^{1/2}}\right) + e^{\alpha|x|}\operatorname{erfc}\left(\gamma(t)^{1/2} + \frac{\alpha|x|}{2\gamma(t)^{1/2}}\right) \tag{5}$$

is positive definite for γ any valid variogram on $\mathbb{R}$. If γ is infinitely differentiable away from the origin, then (5) is as well. However, $K(x,0) = 2e^{-\alpha|x|}$ for any γ, so that this class of models is limited in its possible purely spatial covariance functions.

Another way to quantify the differences between covariance functions with ridges and those without is to consider the Kullback-Liebler divergence between models for given finite sets of observations. Suppose Z is a mean 0, stationary Gaussian random field on $\mathbb{R}^2$ and is observed at (j,k) for $j,k \in 1,\ldots,20$. Suppose further that the true covariance function for Z is $\exp\big\{-(t^2 + s^2)^{1/2}/10\big\}$, an isotropic exponential model, but one fits the separable model $\theta_1 \exp\big\{-(|t|+|s|)/\theta_2\big\}$ to the data. If $\theta_1 = 1$ and $\theta_2 = 10$, then the covariances of the two models are identical when either t or s is 0. The Kullback-Liebler divergence from a density f to a density g is $KL(f,g) = \int f \log(f/g)$, that is, the expected log likelihood of f relative to g when f is true. This divergence is 0 if $g = f$ and is positive otherwise. Setting g equal to the joint density of the observations under the separable covariance with $\theta_1 = 1$ and $\theta_2 = 10$ yields $KL(f,g) = 672.23$, or well over one unit of divergence per observation, an enormous value. Setting $\theta_2 = 10$ and choosing θ_1 to minimize the divergence yields $\theta_1 = 5.443$ and a divergence of 122.52. Selecting both θ_1 and θ_2 to minimize the divergence gives $(\theta_1,\theta_2) = (0.4562, 2.426)$ and a divergence of 56.34. Thus, when fitting the separable model to data from the isotropic model, one would expect the maximum likelihood estimates of (θ_1,θ_2) to be near $(0.4562, 2.426)$ (or at least much nearer to these values than to $(1,10)$).

To illustrate some of these problems, consider a simulated realization of these 400 observations from a Gaussian random field with mean 0 and isotropic covariance function $\exp\big\{-(t^2+s^2)^{1/2}/10\big\}$. Figure 1 plots the empirical variogram along each axis at distances $1,2,\ldots,10$ and along the $45°$ and $135°$ lines at distances $2^{1/2}j$ for $j = 1,\ldots,7$. Even though the truth is isotropic, the empirical variogram perhaps superficially suggests some anisotropy. When fitting the correct model

$\theta_1 \exp\{-(t^2+s^2)^{1/2}/\theta_2\}$, the maximum likelihood estimates are $\widehat{\theta}_1 = 1.725$ and $\widehat{\theta}_2 = 17.5$. Although these estimates are not very accurate, $\widehat{\theta}_1/\widehat{\theta}_2 = 0.0986$, which is very close to the true value for θ_1/θ_2 of 0.1, as one should expect since θ_1/θ_2 controls the local behavior of Z (Stein (1999), Chapter 6). Figure 1 shows that the true variogram and the MLE (maximum likelihood estimate) under the correct model are nearly identical at shorter distances. Furthermore, both of these curves track the empirical variograms reasonably well at the shorter distances. Fitting the separable model $\theta_1 \exp\{-(|t|+|s|)/\theta_2\}$ to these data via maximum likelihood gives $\widehat{\theta}_1 = 0.656241$ and $\widehat{\theta}_2 = 2.93$, which are not all that far from the values 0.4562 and 2.426 that minimize the Kullback-Liebler divergence. Figure 1 shows that this fitted model badly disagrees with both the empirical variogram and the truth at the shorter distances. It is certainly possible to pick θ_1 and θ_2 so that the fitted model and the empirical variogram agree reasonably well visually. For example, Figure 1 plots a fitted separable model with estimates $\widehat{\theta}_1 = 1.725$ and $\widehat{\theta}_2 = 20.8$, which tracks the the empirical variogram much better than does the MLE.

One conclusion to draw from this example is that MLEs can produce bizarre-looking estimates when the model is seriously flawed. However, it would be a serious mistake to conclude that visually agreeable fits to the empirical variogram are a safer approach to parameter estimation when one is concerned about model misspecification. Despite the appearance of Figure 1, the visually fit separable model is arguably a worse fit to the data than the MLE. The problem is that the empirical variogram only provides information about some aspects of the data. Consider $\mathrm{var}\{Z(0,0)-Z(0,1)-Z(1,0)+Z(1,1)\}$, which equals $4K(0,0)-8K(1,0)+4K(1,1)$ under either the isotropic or separable models. The true value for this quantity is 0.2338, the estimated value under the MLE for the isotropic model is 0.2308 and the estimated value under the MLE for the separable model is 0.2195. In contrast, the estimated value under the visual fit for the separable model is 0.01518, which is off by more than an order of magnitude. We see that considering variances of linear combinations of more than two observations can be helpful in identifying problems with models for 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. Let us first consider the problem in one dimension. Suppose f is the spectral density for a real-valued weakly stationary process, so that f is nonnegative, even and integrable on $\mathbb{R}$. Then $K(x) = \int_{\mathbb{R}} e^{iwx} f(w)dw$ is the corresponding continuous covariance function. It is well-known that if f has a finite mth moment, then K is m times differentiable and $K^{(m)}(x) = \int_{\mathbb{R}} (iw)^m e^{iwx} f(w)dw$. If m is even, then the condition that f has a finite mth moment is also necessary for K to have an mth derivative at 0. For example, if $f(w) = (1 + w^2)^{-1}$, then f does not have a first moment, so that one cannot conclude that K has a derivative at 0. In fact, because f is even, it is possible to show (Feller 1971, p. 565) that $x \int_x^\infty f(w)dw \to 0$ as $x \to \infty$ is necessary and sufficient for $K'(0)$ to exist. When $f(w) = (1 + w^2)^{-1}$, $x \int_x^\infty f(w)dw \to 1$ as $x \to \infty$, so that $K'(0)$ does not exist. Indeed, $K(x) = \pi e^{-|x|}$, which is not differentiable at 0.

Of course, $\pi e^{-|x|}$ is infinitely differentiable everywhere except the origin and the goal here is to find verifiable conditions on the spectral density directly that would imply that K has derivatives away from the origin. If f is differentiable with f' integrable and $f(w) \to 0$ as $|w| \to \infty$, then for $x \neq 0$, integrating by parts, $K(x) = ix^{-1} \int_{\mathbb{R}} e^{iwx} f'(w)dw$. If, in addition, $wf'(w)$ is integrable, then for $x \neq 0$,

$$K'(x) = -\frac{i}{x^2} \int_{\mathbb{R}} e^{iwx} f'(w)dw - \frac{1}{x} \int_{\mathbb{R}} we^{iwx} f'(w)dw. \tag{6}$$

When $f(w) = (1 + w^2)^{-1}$, $f'(w) = -2w/(1 + w^2)^2$, so that $wf'(w)$ is integrable. One can then use (6) to recover $K'(x) = -\pi e^{-x}$ for $x > 0$ and $K'(x) = \pi e^x$ for $x < 0$. However, we would not generally use (6) to calculate $K'(x)$, since whenever it is possible to calculate the integrals in (6) explicitly, it would usually be easier to calculate K itself and differentiate the result. Consider instead $f(w) = \{1 + (1 + w^4)^{1/2}\}^{-1}$, for which there does not appear to be an explicit expression for the corresponding covariance function K. In this case, $x \int_x^\infty f(w)dw \to 1$ as $x \to \infty$, so K is not differentiable at 0. Nevertheless, it is easy to verify that $wf'(w)$ is integrable and hence $K'(x)$ exists for all $x \neq 0$.

If f has higher derivatives possessing moments, it is possible to express derivatives of K for $x \neq 0$ in terms of these derivatives of f. Define $(k)_j = k(k + 1) \cdots (k + j - 1)$ and $(k)_0 = 1$.

Proposition 2. *If f is k times differentiable and $f^{(k)}(w)$ and $w^m f^{(k)}(w)$ are integrable, then $K^{(m)}(x)$ exists for all $x \neq 0$ and is given by*

$$K^{(m)}(x) = i^k \sum_{j=0}^{m} \binom{m}{j} (-1)^j (k)_j x^{-k-j} \int_{\mathbb{R}} (iw)^{m-j} f^{(k)}(w) e^{iwx} dw. \tag{7}$$

□

If f itself does not satisfy the conditions of Proposition 2, it may still be possible to show its Fourier transform possesses derivatives away from the origin by writing f as the sum of a function satisfying the conditions of Proposition 2 and another function possessing finite moments.

PROPOSITION 3. *Suppose $f = f_1 + f_2$, where f_1 is k times differentiable and $f^{(k)}(w)$, $w^m f_1^{(k)}(w)$ and $w^m f_2(w)$ are integrable. Then K is m times differentiable for all $x \neq 0$ and*

$$K^{(m)}(x) = i^k \sum_{j=0}^{m} \binom{m}{j} (-1)^j (k)_j x^{-k-j} \int_{\mathbb{R}} (iw)^{m-j} f_1^{(k)}(w) e^{iwx} dw + \int_{\mathbb{R}} (iw)^m f_2(w) e^{iwx} dw.$$

□

For example, $f(w) = (1 + |w|)^{-2}$ is not differentiable at 0, but defining $f_1(w) = (1 + w^2)^{-1}$ and $f_2 = f - f_1$, Proposition 3 proves that the corresponding covariance function is differentiable for $x \neq 0$. Indeed, using Gradshteyn and Ryzhik (2000, 3.722) and integration by parts, one can show that the Fourier transform of $(1 + |w|)^{-2}$ is infinitely differentiable away from the origin.

If f does not satisfy the conditions of Propositions 2 or 3, the corresponding K may or may not have derivatives away from the origin. Consider the spectral density $f(w) = (1-\cos\, w)/(\pi w^2)$. This integrable function does not have a finite first moment. Furthermore, $x \int_x^\infty f(w)dw$ does not tend to 0 as $x \to \infty$, so the corresponding covariance function K does not have a derivative at 0. Now f is infinitely differentiable, but $wf^{(k)}(w)$ is not integrable for any k, which follows from $f^{(k)}(w) = \cos^{(k)} w/(\pi w^2) + O(|w|^{-3})$ as $|w| \to \infty$. Thus, the sufficient condition in Proposition 2 for showing K is differentiable for all $x \neq 0$ is not satisfied for any k. In fact, $K(x) = (1 - |x|)^+$, which is not differentiable at $x = \pm 1$.

Let us next consider an example in which neither Propositions 2 nor 3 apply, but K is still infinitely differentiable away from the origin. If $f(w) = \sin(w^2)/w^2 + 2/(1 + w^2)$, then f is nonnegative and integrable and $K(x) = \pi \left\{ 2e^{-|x|} + S(\frac{1}{2}|x|) - C(\frac{1}{2}|x|) + \pi^{1/2} \sin\left(\frac{x^2+\pi}{4}\right) \right\}$, where C and S are the Fresnel integrals $C(x) = (2/\pi)^{1/2} \int_0^x \cos(t^2)dt$ and $S(x) = (2/\pi)^{1/2} \int_0^x \sin(t^2)dt$, (Bateman Manuscript Project 1954). It is straightforward to verify that K is not differentiable at 0 but is infinitely differentiable for all $x \neq 0$. Furthermore, Proposition 2 clearly does not apply for any m positive and it is possible to show that neither does Proposition 3, since any f_1 that follows the oscillations of f sufficiently well to make f_2 have any finite moments cannot have any derivatives with finite moments.

Now let us consider the situation in $\mathbb{R}^d$. Write $\mathbf{w} = (w_1, \ldots, w_d)'$, $\mathbf{x} = (x_1, \ldots, x_d)'$ 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 $1 \leq i \leq d$.

Proposition 4. *Suppose $D^{\mathbf{k}} f$ exists, $D^{\mathbf{q}} f$ is integrable for all $\mathbf{q} \leq \mathbf{k}$ 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*

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

$\square$

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 $x_j = 0$.

Let us apply this result to the spectral density $f(w_1, w_2) = \pi^{-2}(1+w_1^2)^{-1}(1+w_2^2)^{-1}$. Neither $w_1 f(w_1, w_2)$ nor $w_2 f(w_1, w_2)$ are integrable, but it is easy to show that $D^{\mathbf{k}} f(\mathbf{w})$ and $\mathbf{w}^{\mathbf{k}} D^{\mathbf{k}} f(\mathbf{w})$ are integrable for all $\mathbf{k} \geq \mathbf{0}$. It follows that for all $\mathbf{x}$ with $x_1 x_2 \neq 0$, $D^{\mathbf{m}} K(\mathbf{x})$ exists for all $\mathbf{m} \geq \mathbf{0}$ and if $t \neq 0$, $D^{(m,0)} K(t, 0)$ and $D^{(0,m)} K(0, t)$ exist for all m. However, neither $w_2 D^{(m,0)} f(\mathbf{w})$ nor $w_1 D^{(0,m)} f(\mathbf{w})$ are integrable for any $m \geq 0$. Thus, Proposition 4 says nothing about the existence of $D^{(0,1)} K(t, 0)$ or $D^{(1,0)} K(0, t)$ for $t \neq 0$. In fact, $K(\mathbf{x}) = e^{-|x_1| - |x_2|}$, so that $D^{(0,1)} K(t, 0)$ and $D^{(1,0)} K(0, t)$ do not exist for $t \neq 0$.

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 5. *For $j = 1, \ldots, d$, suppose $(\partial^\ell / \partial w_j^\ell) f(\mathbf{w})$ exists and is integrable for $\ell \leq k$ 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 $m_1 + \cdots + m_d \leq n$. Furthermore, if $x_j \neq 0$,*

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

$\square$

This follows immediately from Proposition 4.

The conditions in Proposition 5 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. For two functions f and g on a domain D, write $f \ll g$ if there exists $C < \infty$ such that $|f(x)| \le Cg(x)$ for all $x \in D$. 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 6 and certain results in the next section take on a more symmetric and transparent form when treating the more general case considered here.

PROPOSITION 6. *Suppose* $f(\mathbf{w}) = \{g_1(|\mathbf{w}_1|^2) + g_2(|\mathbf{w}_2|^2)\}^{-\nu}$, *where* $\nu > 0$, f *is bounded,* g_1 *and* g_2 *are infinitely differentiable functions on* $[0, \infty)$ *and there exist* α_1, α_2 *positive such that* $g_j^{(k)}(t) \ll (1+t)^{\alpha_j - k}$ *for* $j = 1, 2$ *and all nonnegative integers* k. *A necessary and sufficient condition for* f *to be integrable is* $d_1/(\alpha_1\nu) + d_2/(\alpha_2\nu) < 2$. *If* f *is integrable, its Fourier transform* $K(\mathbf{x})$ *possesses all partial derivatives of all orders for all* $\mathbf{x} \ne \mathbf{0}$. □

Many algebraic functions satisfy the required conditions on g_j. For example, if P is a strictly positive polynomial of order p on $[0, \infty)$ and $\gamma > 0$, then $(d^k/dt^k)P(t)^\gamma \ll (1+t)^{p\gamma - k}$ on $[0, \infty)$ for all k.

4. A NEW CLASS OF SPACE-TIME COVARIANCE FUNCTIONS

One rich class of spectral densities that satisfies the conditions of Proposition 6 is

$$f(\mathbf{w}) = \left\{c_1(a_1^2 + |\mathbf{w}_1|^2)^{\alpha_1} + c_2(a_2^2 + |\mathbf{w}_2|^2)^{\alpha_2}\right\}^{-\nu} \tag{8}$$

for c_1, c_2, a_1, a_2, α_1, α_2 positive and $d_1/(\alpha_1\nu) + d_2/(\alpha_2\nu) < 2$. Note that (8) is just (1) when $d_2 = 1$. Jones and Zhang (1997) consider $d_1 = 2$ and $\alpha_2 = \nu = d_2 = 1$. The Fourier transform $K(\mathbf{x}_1, \mathbf{x}_2)$ of (8) 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

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

so that $f_1(\mathbf{w}_1) = \int_{\mathbb{R}^{d_2}} f(\mathbf{w}_1, \mathbf{w}_2)\mathbf{d}\mathbf{w}_2$ 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

$$\int_{\mathbb{R}^{d_2}} \left\{c_1(a_1^2 + |\mathbf{w}_1|^2)^{\alpha_1} + c_2(a_2^2 + |\mathbf{w}_2|^2)^{\alpha_2}\right\}^{-\nu} \mathbf{d}\mathbf{w}_2$$

$$
\begin{aligned}
&= \frac{2\pi^{d_2/2}}{\Gamma(d_2/2)} \int_0^\infty \left\{ c_1(a_1^2 + r_1^2)^{\alpha_1} + c_2(a_2^2 + r_2^2)^{\alpha_2} \right\}^{-\nu} r_2^{d_2-1} dr_2 \\
&\sim \frac{2\pi^{d_2/2}}{\Gamma(d_2/2)} \int_0^\infty \left\{ c_1 r_1^{2\alpha_1} + c_2 r_2^{2\alpha_2} \right\}^{-\nu} r_2^{d_2-1} dr_2 \\
&= \frac{\pi^{d_2/2}}{\Gamma(d_2/2)\alpha_2 c_1^{\nu+1}} \left(\frac{c_1}{c_2} \right)^{d_2/(2\alpha_2)} \frac{\Gamma\left(\frac{d_2}{2\alpha_2}\right)\Gamma\left(\nu - \frac{d_2}{2\alpha_2}\right)}{\Gamma(\nu)} r_1^{-\alpha_1(2\nu - d_2/\alpha_2)}
\end{aligned} \tag{9}
$$

using Gradshteyn and Ryzhik (2000, 3.241.4). A similar result holds with the roles of $\mathbf{w}_1$ and $\mathbf{w}_2$ switched. A desirable feature for a class of models for f is that for all $\gamma_j > 0$ and $\beta_j > 0$, there is an element in the class such that $f_j(\mathbf{w}_j) \sim \beta_j r_j^{-\gamma_j - d_j}$ is achievable for $j = 1, 2$. Because γ_j is directly related to the smoothness of the process in $\mathbf{x}_j$ (Stein 1999), roughly speaking, one can then 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. This can in fact be done with any one of ν, α_1 or α_2 fixed in the model (8). For example, if ν is fixed, by taking

$$
\alpha_1 = \frac{\gamma_1 + d_1 + d_2\gamma_1/\gamma_2}{2\nu}, \qquad \alpha_2 = \frac{\gamma_2 + d_2 + d_1\gamma_2/\gamma_1}{2\nu},
$$

and c_1 and c_2 appropriately, $f_j(\mathbf{w}_j) \sim \beta_j r_j^{-\gamma_j - d_j}$ as $r_j \to \infty$ for $j = 1, 2$. To check that the resulting f is in fact integrable, note that

$$
\begin{aligned}
\frac{d_1}{\nu\alpha_1} + \frac{d_2}{\nu\alpha_2} &= \frac{2d_1}{\phi_1 + d_1 + d_2\phi_1/\phi_2} + \frac{2d_2}{\phi_2 + d_2 + d_1\phi_2/\phi_1} \\
&< \frac{2d_1}{d_1 + d_2\phi_1/\phi_2} + \frac{2d_2}{d_2 + d_1\phi_2/\phi_1} = 2.
\end{aligned}
$$

Next, if $\alpha_2 > 0$ is fixed, by taking

$$
\alpha_1 = \frac{\alpha_2(\phi_1 + d_1 - 1)}{\phi_2}, \qquad \nu = \frac{(\phi_1 + d_1)(\phi_2 + d_2) - d_2}{2\alpha_2(\phi_1 + d_1 - 1)}
$$

and c_1 and c_2 appropriately, $f_j(\mathbf{w}_j) \sim \beta_j r_j^{-\gamma_j - d_j}$ as $r_j \to \infty$ for $j = 1, 2$. The resulting f is integrable since

$$
\frac{d_1}{\nu\alpha_1} + \frac{d_2}{\nu\alpha_2} = 2\frac{(\phi_1 + d_1)(\phi_2 + d_2) - \phi_1\phi_2 - d_2}{(\phi_1 + d_1)(\phi_2 + d_2) - d_2} < 2.
$$

A similar result holds if $\alpha_1 > 0$ is fixed.

Thus, in terms of controlling the high frequency behavior of the resulting space-time process as just a spatial or just a temporal process, one might as well set one of ν, α_1 or α_2 to a fixed value.

However, if f and $\widetilde{f}$ with parameter values $(\nu, \alpha_1, \alpha_2, c_1, c_2, a_1, a_2)$ and $(\widetilde{\nu}_1, \widetilde{\alpha}_1, \widetilde{\alpha}_2, \widetilde{c}_1, \widetilde{c}_2, \widetilde{a}_1, \widetilde{a}_2)$ respectively give the same values for $(\beta_1, \beta_2, \gamma_1, \gamma_2)$, then the two spectral densities are not asymptotically the same as $|\mathbf{w}| \to \infty$. For example, if $d_1 = d_2 = 1$, $f(w_1, w_2) = 2^{1/2}(2 + w_1^2 + w_2^2)^{-2}$ and $\widetilde{f}(w_1, w_2) = \left\{(1 + w_1^2)^2 + (1 + w_2^2)^2\right\}^{-1}$, then both models give $\gamma_1 = \gamma_2 = 2$ and $\beta_1 = \beta_2 = 2^{-1/2}\pi$, but $f(w_1, w_2)/\widetilde{f}(w_1, w_2)$ does not tend to a constant as $|\mathbf{w}| \to \infty$. However, $f(w_1, w_2)/\widetilde{f}(w_1, w_2)$ is bounded away from 0 and ∞, so that while the two models are different at high frequencies, they are arguably not very different. In particular, if $0 < L \le f(\mathbf{w})/\widetilde{f}(\mathbf{w}) \le U < \infty$ for all $\mathbf{w}$, then the ratio of variances assigned to any linear combination of values of the random field by the spectral densities f and $\widetilde{f}$ will be in $[L, U]$.

Thus, if one is hoping to reduce the number of parameters in (8) from seven to some lower number, fixing one of α_1, α_2 or ν is a reasonable place to start. Suppose from now on that $\alpha_2 = 1$. To further reduce the number of parameters, a_1 and a_2, which are range parameters of a sort, would be a place to look. If either a_1 or a_2 are 0 but not both, then the resulting f is still integrable. Furthermore, if $a_2 = 0$, by a minor modification of the argument in Proposition 6, the resulting covariance function is infinitely differentiable away from the origin. One could set $a_1 = a_2 = 0$, in which case f is not integrable, but it does correspond to the spectral "density" for an intrinsic space-time random function (Christakos 1992) and leaves one with a 4-parameter model for spatial-temporal variations with great flexibility in the local purely spatial and purely temporal variations.

There are only a few special cases of spectral densities of the form (8) 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).

Equation (8) can sometimes be transformed analytically when $\alpha_1 \neq \alpha_2$ if α_1, α_2, and ν are all integers. For example, as noted earlier, the covariance function in (4) is obtained when $d_1 = d_2 = 1$, $\alpha_1 = 2$, $\alpha_2 = 1$, $a_2 = 0$ 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 $\left\{c^2(a^2+|\mathbf{w}|^2)^2+v^2\right\}^{-2}$, first note that

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

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

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

Integrating by parts and using Gradshteyn and Ryzhik (2000, 3.954),

$$\begin{aligned}
&\int_0^\infty \frac{e^{-c^2k^2|t|}}{(a^2+k^2)^2}k\sin(kr)dk \\
&= \frac{1}{2}\int_0^\infty \frac{e^{-c^2k^2|t|}}{a^2+k^2}\{-2c^2k|t|\sin(kr)+r\cos(kr)\}dk \\
&= \frac{\pi}{4}c^2|t|e^{c^2a^2|t|}\left\{e^{ar}\operatorname{erfc}\left(ac|t|^{1/2}+\frac{r}{2c|t|^{1/2}}\right) - e^{-ar}\operatorname{erfc}\left(ac|t|^{1/2}-\frac{r}{2c|t|^{1/2}}\right)\right\} \\
&\quad + \frac{\pi r}{8a}e^{c^2a^2|t|}\left\{e^{ar}\operatorname{erfc}\left(ac|t|^{1/2}+\frac{r}{2c|t|^{1/2}}\right) + e^{-ar}\operatorname{erfc}\left(ac|t|^{1/2}-\frac{r}{2c|t|^{1/2}}\right)\right\}.
\end{aligned}$$

Similarly evaluating $\int_0^\infty e^{-c^2k^2|t|}(a^2+k^2)^{-3}k\sin(kr)dk$ yields

$$\begin{aligned}
K(\mathbf{x},t) = &\frac{\pi^2}{16c^6}e^{ar}\operatorname{erfc}\left(ac|t|^{1/2}+\frac{r}{2c|t|^{1/2}}\right)\left(\frac{1}{a^3}-\frac{r}{a^2}+\frac{4c^4t^2}{r}\right) \\
&+ \frac{\pi^2}{16c^6}e^{-ar}\operatorname{erfc}\left(ac|t|^{1/2}-\frac{r}{2c|t|^{1/2}}\right)\left(\frac{1}{a^3}+\frac{r}{a^2}-\frac{4c^4t^2}{r}\right) \\
&+ \frac{\pi^{3/2}|t|^{1/2}}{4c^5a^2}\exp\left(-c^2a^2|t|-\frac{r^2}{4c^2|t|}\right).
\end{aligned} \tag{10}$$

For $\mathbf{x} = \mathbf{0}$ or $t = 0$, define K by continuity, so that $K(\mathbf{x},0) = \frac{\pi^2}{8c^6a^3}e^{-ar}(1+ar)$ and

$$K(\mathbf{0},t) = \frac{\pi^2}{8c^6}\left(\frac{1}{a^3}+4c^4t^2\right)\operatorname{erfc}\left(ac|t|^{1/2}\right) + \frac{\pi^{3/2}e^{-c^2a^2|t|}}{c^6}\left(\frac{c|t|^{1/2}}{4a^2}-\frac{c^3|t|^{3/2}}{2}\right). \tag{11}$$

From (9), $f_1(\mathbf{w}_1) \sim C_1|\mathbf{w}_1|^{-6}$ as $|\mathbf{w}_1| \to \infty$ for a positive constant C_1, and it follows (Bingham, Goldie and Teugels 1987) that $K(\mathbf{x},0)$ must be of the form $b_0 - b_1|\mathbf{x}|^2 + b_2|\mathbf{x}|^3 + o(|\mathbf{x}|^3)$ as $\mathbf{x} \to \mathbf{0}$, with $b_2 \neq 0$. In fact, $K(\mathbf{x}, 0) = \frac{\pi^2}{8c^6}\big(\frac{1}{a^3} - \frac{1}{2a}|\mathbf{x}|^2 + \frac{1}{3}|\mathbf{x}|^3\big) + O(|\mathbf{x}|^4)$ as $\mathbf{x} \to \mathbf{0}$. From either the spectral or spatial domain, one sees that the process is exactly once mean square differentiable in any spatial direction. From (9), $f_2(v) \sim C_2|v|^{-5/2}$ as $v \to \infty$ for some positive C_2, which implies that $K(\mathbf{0}, t) = b_0 - b_1|t|^{3/2} + o(|t|^{3/2})$ as $t \to 0$ for some $b_1 \neq 0$. Indeed, applying Taylor series to (11), it is possible to show that as $t \to 0$, $K(\mathbf{0}, t) = \frac{\pi^2}{8c^6a^3} - \frac{2\pi^{3/2}}{3c^3}|t|^{3/2} + O(t^2)$. Thus, the process is not mean square differentiable in time. Finally, although it is not obvious from (10), Proposition 6 implies that this function is infinitely differentiable away from the origin.

If $\alpha_2 = 1$ in (8), 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 can be approximated quickly for a large number of x_1 values using the fast Fourier transform. A separate transform needs to be done for every value of $|\mathbf{x}_2|$ of interest, but this should still be feasible in many circumstances.

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. If K has spectral density $f(\mathbf{w}, v)$, then K is spatially isotropic if and only if $f_1(\mathbf{w}) = \int_{\mathbb{R}} f(\mathbf{w}, v)\, dv$ depends (almost everywhere) on $\mathbf{w}$ only through $|\mathbf{w}|$. Consider a nonnegative, integrable function of the form

$$f(\mathbf{w}, v) = g_1(\mathbf{w})g_2(v)h_1(|\mathbf{w}|, |v|) + h_2(|\mathbf{w}|, |v|), \tag{12}$$

with g_1 and g_2 odd functions. A density of this form is an even function of $(\mathbf{w}, v)$ and hence is the spectral density of a real-valued stationary space-time random field. Furthermore, $f_1(\mathbf{w}) = \int_{\mathbb{R}} h_2(|\mathbf{w}|, |v|)\, dv$, so the corresponding random field is spatially isotropic. Note that the inclusion of h_2 is essential in (12), since the only way $g_1(\mathbf{w})g_2(v)h_1(|\mathbf{w}|, |v|)$ can be nonnegative everywhere is to be identically 0. For a particular subclass of spectral densities of the form (12), it is commonly possible to obtain an explicit expression for the covariance function $K(\mathbf{x}, t)$. Specifically, 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

$$f(\mathbf{w}, v) = \big\{a + b(\mathbf{w}'\mathbf{z})v + c_1|\mathbf{w}|^2 + c_2v^2\big\}\Psi(|\mathbf{w}|, |v|), \tag{13}$$

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 is nonnegative and integrable and, since it is of the form (12), the resulting K is spatially isotropic. Furthermore, if $K_0(\mathbf{x}, t)$ is the Fourier transform of Ψ, then

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

Now $K_0(\mathbf{x}, t)$ only depends on $(\mathbf{x}, t)$ through $(|\mathbf{x}|, t)$, so define the function $\overline{K}_0$ 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 $D^{(m_1,m_2)}\overline{K}_0$,

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

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. Putting these results together yields

$$\begin{aligned} K(\mathbf{x}, t) = {} & a\overline{K}_0(|\mathbf{x}|, t) - b\frac{\mathbf{x}'\mathbf{z}}{|\mathbf{x}|} \overline{K}_0^{(1,1)}(|\mathbf{x}|, t) - c_1\overline{K}_0^{(2,0)}(|\mathbf{x}|, t) \\ & - c_1 \frac{d-1}{|\mathbf{x}|} \overline{K}_0^{(1,0)}(|\mathbf{x}|, t) - c_2\overline{K}_0^{(0,2)}(|\mathbf{x}|, t). \end{aligned} \tag{14}$$

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 $\overline{K}_0(r, t) = \mathcal{M}_{\nu+1}\big((\beta_1^2 r^2 + \beta_2^2 t^2)^{1/2}\big)$ with $\nu > 0$, for which the corresponding K_0 is twice differentiable. Letting $y = (\beta_1^2 r^2 + \beta_2^2 t^2)^{1/2}$ yields (Abramowitz and Stegun (1965), 9.6.26) $\overline{K}_0^{(1,0)}(r,t) = -\beta_1^2 r \mathcal{M}_\nu(y)$, $\overline{K}_0^{(1,1)}(r,t) = \beta_1^2\beta_2^2 rt\mathcal{M}_{\nu-1}(y)$, $\overline{K}_0^{(2,0)}(r,t) = -\beta_1^2\mathcal{M}_\nu(y) + \beta_1^4 r^2 \mathcal{M}_{\nu-1}(y)$ and $\overline{K}_0^{(0,2)}(r,t) = -\beta_2^2\mathcal{M}_\nu(y) + \beta_2^4 t^2\mathcal{M}_{\nu-1}(y)$. Using $\mathcal{M}_{\nu+1}(y) = y^2\mathcal{M}_{\nu-1}(y) + 2\nu\mathcal{M}_\nu(y)$,

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

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

$$K(\mathbf{x},t) = a\big\{(2\nu + d + 1)\mathcal{M}_\nu(y) - 2\tau(\beta_1\mathbf{x}'\mathbf{z})\beta_2 t\mathcal{M}_{\nu-1}(y)\big\}, \tag{15}$$

where $0 \le \tau \le 1$ 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 (15), Figure 2 plots $K((x_1,0),t)/K(\mathbf{0},0)$. Note that along the x_1 axis, $\mathbf{x}$ and $\mathbf{z}$ are parallel. The lack of full symmetry is apparent. On the other hand, $\mathbf{z}$ is perpendicular to $(0,x_2)$ and $K((0,x_2),t) = K((0,-x_2),t)$ for all x_2 and t. Other angles between $\mathbf{x}$ and $\mathbf{z}$ yield a convex combination of the parallel and perpendicular cases.

Models of the form (15) 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 $|\mathbf{b}| \le c$ and letting $K(\mathbf{x},t) = C(s)$ for $C \in \mathcal{D}_{d+1}$. This approach always yields elliptical contours for the covariance function, in contrast to what Figure 2 shows. Whether covariance functions of the form (15) will prove more useful than, say, covariance functions like $\phi\mathcal{M}_\nu(\beta s)$ remains to be seen. However, the general approach outlined in this section can also be applied to covariance functions that have different smoothness across space than across time.

6. MARKOV MODELS

One commonly used principle for restricting the class of stochastic processes one wishes 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. So far, this work has only considered the first two moments of spatial-temporal processes, but by adding the assumption that the process is Gaussian, then the question of a process being Markov reduces to whether 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 $\{Z(\mathbf{y}, s) : \mathbf{y} \in \mathbb{R}^d, s \le t_0\}$. 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 spatial-temporal covariance functions satisfying this condition.

Proposition 7. *A stationary, mean square continuous real-valued space-time process Z has the property that, for all $t > 0$, 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 $\{Z(\mathbf{y}, s) : \mathbf{y} \in \mathbb{R}^d, s \le 0\}$ if and only if its covariance function K is of the form*

$$K(\mathbf{x}, t) = \int_{\mathbb{R}^d} \exp\{i\mathbf{w}'\mathbf{x} - |t|\beta(\mathbf{w}) - it\phi(\mathbf{w})\} F(\mathbf{dw}), \tag{16}$$

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, setting $t = 0$ in (16) shows that K is spatially isotropic if and only if the measure F depends on $\mathbf{w}$ only through $|\mathbf{w}|$. The function β gives 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 (16) as

$$K(\mathbf{x}, t) = \frac{1}{\pi} \int_{\mathbb{R}^{d+1}} e^{ivt + i\mathbf{w}'\mathbf{x}} \frac{\beta(\mathbf{w})}{\beta(\mathbf{w})^2 + \{v + \phi(\mathbf{w})\}^2} \, dv \, F(\mathbf{dw}). \tag{17}$$

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 (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 (17) when $\phi(\mathbf{w}) = \mathbf{w}'\mathbf{u}$, $\beta(\mathbf{w}) = \lambda + \frac{1}{2}\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}$ and $F(\mathbf{dw})/\mathbf{dw} = \exp\left(-\frac{1}{2}\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}\right)/\left(\lambda + \frac{1}{2}\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}\right)$ for some $\lambda \ge 0$, a vector $\mathbf{u}$ and positive definite matrix $\boldsymbol{\Sigma}$. All such processes are analytic in space, which may limit their practical utility.

A class of models for which (16) 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,

$$\begin{aligned} K(\mathbf{x}, t) &= \int_{\mathbb{R}^d} \frac{e^{i\mathbf{w}'(\mathbf{x}-\epsilon t\mathbf{z})}}{(1 + \alpha^{-2}|\mathbf{w}|^2)^{\nu+\zeta|t|+d/2}} \mathbf{dw} \\ &= \frac{\pi^{d/2}\alpha^d}{2^{\nu+\zeta|t|-1}\Gamma\left(\nu + \zeta|t| + \frac{1}{2}d\right)} \mathcal{M}_{\nu+\zeta|t|}(\alpha|\mathbf{x} - \epsilon t\mathbf{z}|). \end{aligned} \tag{18}$$

Taking $t = 0$ essentially recovers the Matérn model for the spatial variation:

$$K(\mathbf{x}, 0) = \frac{\pi^{d/2}\alpha^d}{2^{\nu-1}\Gamma\left(\nu + \frac{1}{2}d\right)} \mathcal{M}_\nu(\alpha|\mathbf{x}|),$$

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.

If $\epsilon \neq 0$ in (18), 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.

That a Markov process can exhibit long-range dependence along one of its margins is perhaps 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.

Let us next consider the smoothness away from the origin of (18), setting $\epsilon = 0$ for simplicity. First of all, K is infinitely differentiable at $(\mathbf{x}, t)$ if $t|\mathbf{x}| \neq 0$. If $m < \nu < m + 1$ for a nonnegative integer m, then $\mathcal{M}_\nu(t) = \sum_{j=0}^m C_j t^{2j} + C_\nu |t|^{2\nu} + o(|t|^{2\nu})$ as $t \to 0$, so as t increases, the smoothness of $K(\mathbf{x}, t)$ in $\mathbf{x}$ at $\mathbf{x} = \mathbf{0}$ increases with t, although K is not infinitely differentiable in $\mathbf{x}$ at $(\mathbf{0}, t)$ for any t. 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 has a ridge along the line $t = 0$. Thus, the models in this class are smoother along the space axes than they are at the origin but not smoother along the time axis. Whether such a property will be of practical value remains to be seen, but it is a difference between these models, models that are smooth everywhere but at the origin, and models that have ridges along both the space and time axes.

Let us consider one further special case of (16) for which K can be calculated analytically. For $d = 1$, $\phi(w) = 0$, $\beta(w) = cw^2$ with $c > 0$ and $F(dw)/dw = (a^2 + w^2)^{-1}$, using Gradshteyn and Ryzhik (2000, 3.954.2),

$$\frac{1}{\pi} \int_{\mathbb{R}^2} \frac{cw^2}{c^2 w^4 + v^2} \cdot \frac{1}{a^2 + w^2} e^{i(vt+wx)} dv\, dw$$
$$= \frac{\pi e^{a^2 c|t|}}{2a} \left[e^{-ax} \operatorname{erfc} \left\{ a(c|t|)^{1/2} - \frac{x}{2(c|t|)^{1/2}} \right\} + e^{ax} \operatorname{erfc} \left\{ a(c|t|)^{1/2} + \frac{x}{2(c|t|)^{1/2}} \right\} \right]. \tag{19}$$

One immediately sees that $K(x,t)$ is infinitely differentiable if $t \neq 0$. It is possible to show that $K(0,t) = \pi a^{-1} - 2\pi^{1/2}(c|t|)^{1/2} + O(|t|)$ as $t \to 0$ and for fixed $x \neq 0$, $K(x,t) = \pi a^{-1}(1 + a^2 c|t|)e^{-a|x|} + O(t^2)$ as $t \to 0$. Thus, $K(x,t)$ does not have a derivative in t when $t = 0$ for any x, although it is somewhat smoother away from the origin than at the origin because of the $O(|t|^{1/2})$ term in the expansion for $K(0,t)$. More interestingly, although $K(0,t)$ is necessarily maximized at $t = 0$, for every $x \neq 0$, $K(x,t)$ viewed as a function of t achieves a local minimum at $t = 0$. It is perhaps surprising that there exists a positive definite function possessing such a property.

Similar to (5), 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 (16), the resulting model is still positive definite (although no longer Markov unless $\gamma(|t|)$ is a multiple of $|t|$). To see this, note that, as in Ma (2003), it holds when ϕ is identically 0. Setting $u = v - \phi(\mathbf{w})$ in (17) yields that the resulting K is the Fourier transform of a positive, finite, symmetric measure. Thus, in (18), one can replace each appearance of $\zeta|t|$ by $\gamma(|t|)$ and obtain an explicit class of space-time covariance functions with 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) and Gneiting (2002). 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/. The models lacking full symmetry require the actual locations of the sites, which were not available from the website, so the latitudes and longitudes used here were reconstructed from an atlas, yielding intersite distances differing slightly from those in Haslett and Raftery (1989). 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. One can undoubtedly fit the data better by including any number of nonstationarities. Haslett and Raftery (1989) show that excluding one of the sites, Rosslare, leads to a much better fit of an isotropic model to the spatial variations in wind. However, there are other apparent nonstationarities in the data, so removing Rosslare does not solve the problem. For example, by looking at the temporal periodogram of the differences between certain sites (e.g., Birr and Dublin), it is apparent that the seasonal cycle is not identical at all sites. Furthermore, a CUSUM plot of the deseasonalized winds at Clones shows a clear shift in the mean at that site towards the end of 1967. Rather than seek to find and fit all such nonstationarities, the goal here is to find a good fit among stationary models for all of the data including Rosslare. The one nonstationarity taken account of is that the mean wind speed can vary spatially. Specifically, the mean wind speed at site $\mathbf{x}_i$ 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 this average wind speed as a function of location. The mean winds are clearly stronger at the coastal sites than inland, and any sensible model for the mean winds 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 78,888 = 12 × 6574 observations. Letting $t = 1, \ldots, 6574$ indicate the day of the study period, set $\mathbf{Z}_t = (Z(\mathbf{x}_1, t), \cdots, Z(\mathbf{x}_{12}, t))$. For $j = 1, \ldots 939$, let $\mathbf{Y}_j = (\mathbf{Z}_{(j-1)t+2} - \mathbf{Z}_{(j-1)t+1}, \cdots, \mathbf{Z}_{(j-1)t+8} - \mathbf{Z}_{(j-1)t+7})$, a vector of 7 days of first differences in time of the $\mathbf{Z}_t$s. Since $6574 = 7 \times 939 + 1$, $\mathbf{Y} = (\mathbf{Y}_1, \cdots, \mathbf{Y}_{939})$ 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 the maximizer of this likelihood is called the restricted maximum likelihood estimator (Christensen 1996). A simple approximation to the restricted loglikelihood is given by $\sum_{j=1}^{939} \log p(\mathbf{Y}_j \mid \boldsymbol{\theta})$, which is a special case of an approximation studied in Stein, Chi and Welty (2003) 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 it at least yields an unbiased estimating equation for $\boldsymbol{\theta}$ (Stein, Chi and Welty 2003).

Using this approximate likelihood, one can compare an extension of the model proposed in Gneiting (2002) to some of the models proposed here. Combining (4) and (21) in Gneiting (2002) yields that

$$K(\mathbf{x},t) = \frac{\phi(1+\delta 1\{\mathbf{x}=\mathbf{0}\})}{(1+a|t|^{2\alpha})^{\beta d/2}} \exp\left\{-\frac{c|\mathbf{x}|^{2\gamma}}{(1+a|t|^{2\alpha})^{\beta\gamma}}\right\} \tag{20}$$

is positive definite on $\mathbb{R}^d \times \mathbb{R}$ for all nonnegative a, c, ϕ and δ and all α, β and γ in $[0,1]$. Gneiting (2002) fits this model to the empirical space-time covariances assuming $\gamma = \frac{1}{2}$. The presence of the term $\delta 1\{\mathbf{x}=\mathbf{0}\}$ is crucial to obtaining a good fit to the empirical covariances because of what appears to be a discontinuity in $K(\mathbf{x},t)$ at $\mathbf{x}=\mathbf{0}$ for all t, which can be seen if one looks carefully at Figure 5 in Gneiting (2002). Because the 12 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. In any case, the inclusion of such a term, which adds only one parameter to the model, helps considerably with the fits of the models developed here to these data.

Gneiting (2002) notes that (20) is fully symmetric and that empirical space-time covariances of the data are not. Figure 3 plots $\operatorname{corr}\{(Z(\mathbf{x}_i,t+1), Z(\mathbf{x}_j,t)\} - \operatorname{corr}\{(Z(\mathbf{x}_i,t), Z(\mathbf{x}_j,t+1)\}$ versus the difference in longitudes between $\mathbf{x}_i$ and $\mathbf{x}_j$, showing that the asymmetry in the correlations is strongly related to longitude. Gneiting (2002) discusses how one might extend his models to include asymmetries, but the method he proposes does not obviously yield explicit expressions for the covariance function. One simple extension of (20) that allows asymmetry is

$$K(\mathbf{x},t) = \frac{\phi(1+\delta 1\{\mathbf{x}=\mathbf{0}\})}{(1+a|t|^{2\alpha})^{\beta d/2}} \exp\left\{-\frac{c|\mathbf{x}-\epsilon t\mathbf{v}|^{2\gamma}}{(1+a|t|^{2\alpha})^{\beta\gamma}}\right\} \tag{21}$$

for some unit vector $\mathbf{v} \in \mathbb{R}^2$. In light of Figure 3 and other exploratory analyses, if the first component of $\mathbf{x}$ is longitude, it is plausible to set $\mathbf{v} = (1,0)'$ rather than trying to estimate $\mathbf{v}$.

This model will be compared to three proposed here extended to include the δ parameter from (20). The first is the Markov model in (18):

$$K(\mathbf{x}, t) = \frac{\pi^{d/2}\phi\alpha^d(1 + \delta 1\{\mathbf{x} = \mathbf{0}\})}{2^{\nu+\zeta|t|-1}\Gamma\big(\nu + \zeta|t| + \frac{1}{2}d\big)}\mathcal{M}_{\nu+\zeta|t|}(\alpha|\mathbf{x} - \epsilon t(1, 0)'|). \tag{22}$$

For $\delta > 0$, this model is no longer Markov. The second is

$$\phi(1 + \delta 1\{\mathbf{x} = \mathbf{0}\})\mathcal{M}_\nu\big(\beta\big(|\mathbf{x}|^2 + 2bx_1t + c^2t^2\big)^{1/2}\big). \tag{23}$$

The third is adapted from (15):

$$\phi(1 + \delta 1\{\mathbf{x} = \mathbf{0}\})\big\{(2\nu + 3)\mathcal{M}_\nu(y) - \tau\beta_1\beta_2 x_1 t\mathcal{M}_{\nu-1}(y)\big\}, \tag{24}$$

where $y = \big(\beta_1^2|\mathbf{x}|^2 + \beta_2^2t^2\big)^{1/2}$. The approximate likelihoods for each of the models (20)–(24) were maximized using the routine nlm in R. The differences between these maximized approximate likelihoods and the maximized approximate likelihood for the model given by (20) with $\gamma = \frac{1}{2}$ (effectively the parametric form fit by Gneiting (2002)) are 22.9, 140.8, 126.7, 186.5, 204.2, respectively for models (20)–(24). The number of parameters in each of these models is 6, except for (20), which has 7, and (21), which has 8. The approximate likelihoods suggest that model (24) fits the data best. The estimated model parameters are $(\widehat{\phi}, \widehat{\nu}, \widehat{\beta}_1, \widehat{\beta}_2, \widehat{\tau}, \widehat{\delta}) = (9.34, 0.768, 0.00327, 1.11, 1, 0.768)$. Note that $\widehat{\tau} = 1$ is on the boundary of the permissible parameter space. With such a large dataset, one could undoubtedly find stationary models that fit the data better and nonstationary models that fit better still.

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

$$\begin{aligned}\rho_\epsilon^m(t, s) = &\sum_{j=1}^{p} C_j(s)\{|t + \epsilon|^{\alpha_j} - 2|t|^{\alpha_j} + |t - \epsilon|^{\alpha_j}\}\\ &+ R_s(t + \epsilon) - 2R_s(t) + R_s(t - \epsilon).\end{aligned}$$

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 $p = 1$, (2) follows immediately. If $p > 1$, then for $|t| \leq 2\epsilon$, $|t+\epsilon|^{\alpha_j} - 2|t|^{\alpha_j} + |t-\epsilon|^{\alpha_j} \ll \epsilon^{\alpha_j}$, 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 (2) follows. Since $\{|t+\epsilon|^{\alpha_1} - 2|t|^{\alpha_1} + |t-\epsilon|^{\alpha_1}\}/(2\epsilon^{\alpha_1}) \to 1$ if $t = 0$ and to 0 otherwise, (3) follows. $\square$

Proof of Proposition 2. Adams (1978, Theorem 4.13) implies $f^{(j)}$ is integrable for $j = 1, \ldots, k-1$. Thus, for $j \leq k-1$,

$$\lim_{w,w' \to \infty} \left\{f^{(j)}(w') - f^{(j)}(w)\right\} = \int_{\mathbb{R}} f^{(j+1)}(\tau)d\tau,$$

so that $\lim_{w \to \infty} f^{(j)}(w)$ exists. Since $f^{(j)}$ is integrable for $j \leq k-1$, one has $f^{(j)}(w) \to 0$ as $w \to \infty$. Since f is even, one also has $f^{(j)}(w) \to 0$ as $w \to -\infty$ for $j \leq k-1$. Thus, for $x \neq 0$, integrating by parts k times yields $K(x) = (i/x)^k \int_{\mathbb{R}} f^{(k)}(w)e^{iwx}dw$. Since $f^{(k)}$ has a finite mth moment, $(\partial^j/\partial x^j)\int_{\mathbb{R}} f^{(k)}(w)e^{iwx}dw = \int_{\mathbb{R}}(iw)^j f^{(k)}(w)e^{iwx}dw$ for $j \leq m$ and (7) follows. $\square$

Proof of Proposition 4. 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 $(w_1, \cdots, w_{j-1}, w_{j+1}, \cdots, w_d)'$,

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

Thus,

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

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 4 follows by repeated application of the product rule for differentiation and elementary counting arguments. $\square$

Proof of Proposition 6. The integrability of f holds if and only if

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

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 let $\lfloor x \rfloor$ be the greatest integer less than or equal to x. Then for appropriate constants $c_{k\ell}$,

$$\frac{\partial^k}{\partial w_{j1}^k} g_j(|\mathbf{w}_j|^2) = \sum_{\ell=0}^{\lfloor k/2 \rfloor} c_{k\ell}\, g_j^{(k-\ell)}(|\mathbf{w}_j|^2) w_{j1}^{k-2\ell},$$

and elementary calculations yield $(\partial^k/\partial w_{j1}^k) g_j(|\mathbf{w}_j|^2) \ll 1 + |\mathbf{w}_j|^{2\alpha_j - k}$. For an infinitely differentiable function h on $[0, \infty)$ and $\nu > 0$, $(d^k/dt^k)h(t)^{-\nu}$ is a linear combination of terms like $h(t)^{-k-\nu} \prod_{j=0}^{k} \{h^{(j)}(t)\}^{a_j}$ for which $\sum_{j=0}^{k} a_j = \sum_{j=0}^{k} j a_j = k$ with $a_1, \ldots, a_k$ nonnegative integers. It follows that $|\mathbf{w}|^m (\partial^k/\partial w_{j1}^k) f(\mathbf{w})$ is integrable if $|\mathbf{w}|^m (1 + |\mathbf{w}_1|)^{k(2\alpha_1 - 1)} (1 + |\mathbf{w}_1|^{\alpha_1} + |\mathbf{w}_2|^{\alpha_2})^{-k-\nu}$ is integrable, and

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

which, for any given m, can be made finite by taking k sufficiently large. The same argument applies when taking the derivative of k with respect to any component of $\mathbf{w}$, so Proposition 6 follows. □

Proof of Proposition 7. Let us first prove that Z Markov implies its covariance function is of the form (18). 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 $F(\mathbf{dw}) = \int_v G(\mathbf{dw}, dv)$.

Then $G(\mathbf{dw}, dv) = H_{\mathbf{w}}(dv)F(\mathbf{dw})$, 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 $s \leq t$, $\mathbf{x} \in \mathbb{R}^d$. It suffices to show that for all $\mathbf{x} \in \mathbb{R}^d$ and all s, t nonnegative, $\operatorname{cov}\{Z(\mathbf{0}, t), Z(-\mathbf{x}, -s)\} = \operatorname{cov}[E\{Z(\mathbf{0}, t) \mid Z(\cdot, 0)\}, Z(-\mathbf{x}, -s)]$, or in terms of elements of $\mathcal{L}^2(G)$, $\int e^{i\mathbf{x}'\mathbf{w}+iv(s+t)} G(\mathbf{dw}, dv) = \int (\mathcal{P}_0 e^{ivt}) e^{i\mathbf{x}'\mathbf{w}+ivs} G(\mathbf{dw}, dv)$. The Markov property implies $\mathcal{P}_0 e^{ivt}$ is in the closed real linear hull of $e^{i\mathbf{x}'\mathbf{w}}$, 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,

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

and it follows that $A(t; \mathbf{w}) = e^{-\alpha(\mathbf{w})t}$ for all $t \geq 0$ 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 $t \leq 0$, $A(t; \mathbf{w}) = \overline{A(-t; \mathbf{w})} = e^{t\beta(\mathbf{w}) - it\phi(\mathbf{w})}$ and it follows that K is of the form (18) 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 (18), 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 that $\eta_{\mathbf{w}}(t) = e^{-|t|\beta(\mathbf{w}) - it\phi(\mathbf{w})}$ F-a.e. Then for s, t nonnegative,

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

$$= 0,$$

so that $\mathcal{P}_0 e^{ivt} = \eta_{\mathbf{w}}(t)$, 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.

Adams, R. A. (1978), *Sobolev Spaces*, San Diego: Academic Press.

Bateman Manuscript Project (1954), *Tables of Integral Transforms*, ed. A. Erdélyi, research associates W. Magnus, F. Oberhettinger, F. G. Tricomi, New York: McGraw-Hill.

Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1987), *Regular Variation*, New York: Cambridge University Press.

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.

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," in press, *Annals of Statistics*.

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

Feller, W. (1971), *An Introduction to Probability Theory and Its Applications*, (vol. II), New York: Wiley.

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, http://www4.stat.ncsu.edu/~fuentes/sdf.ps.

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.

Ma, C. (2003), "Families of Spatio-Temporal Stationary Covariance Models," in press, *Journal of Statistical Planning and Inference*.

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 Dusnmuir, 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., Chi, Z. and Welty, L. J. (2003), "Approximating Likelihoods for Large Spatial Datasets," Technical Report 532, University of Chicago, Dept. of Statistics.

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. Empirical and fitted variograms for simulated data. Numbers indicate empirical variograms along one of four indicated directions. Solid curve is true variogram. Dashed curve is MLE under correct isotropic model, $K(x, y) = \theta_1 \exp\big\{-(x^2 + y^2)^{1/2}/\theta_2\big\}$. The two dotted curves show the MLE under the separable model, $K(x, y) = \theta_1 \exp\big\{-(|x| + |y|)/\theta_2\big\}$. The lower dotted curve shows the fitted curve for directions 0° and 90° and the upper curve for 45° and 135°. The dotted-dashed curves give curves under the separable model fitted by eye to the empirical variogram; the lower curve corresponding to the directions 0° and 90° and the upper curve to 45° and 135°.

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

Figure 3. Plot of $\mathrm{corr}\{Z(\mathbf{x}_i, t+1), Z(\mathbf{x}_j, t)\} - \mathrm{corr}\{Z(\mathbf{x}_i, t), Z(\mathbf{x}_j, t+1)\}$ for Irish wind data for all pairs of sites for which $\mathbf{x}_i$ is east of $\mathbf{x}_j$.

Figure 1

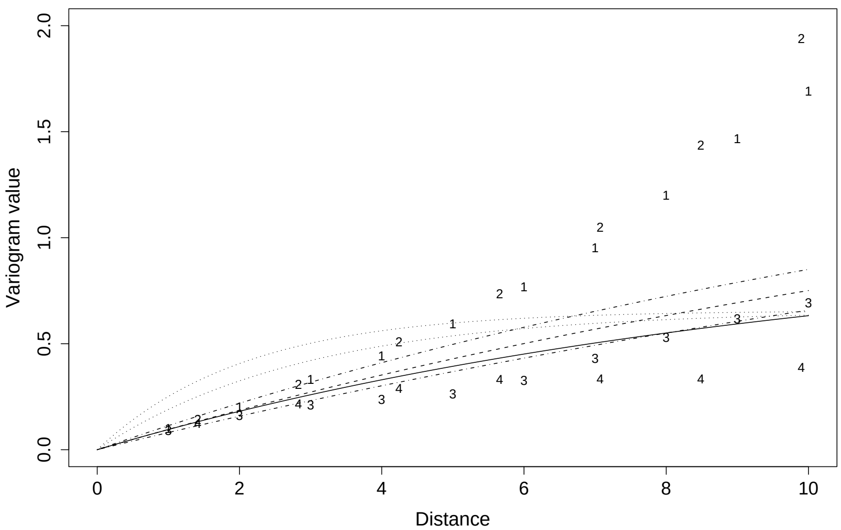


Figure 2

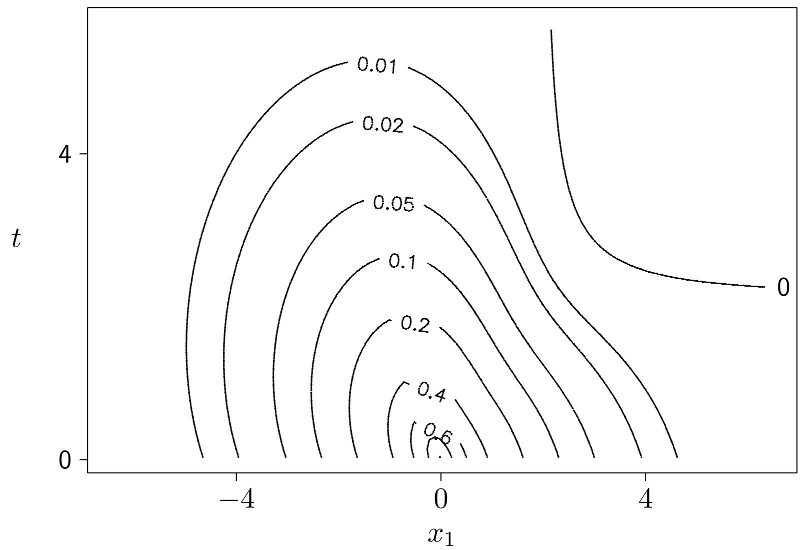
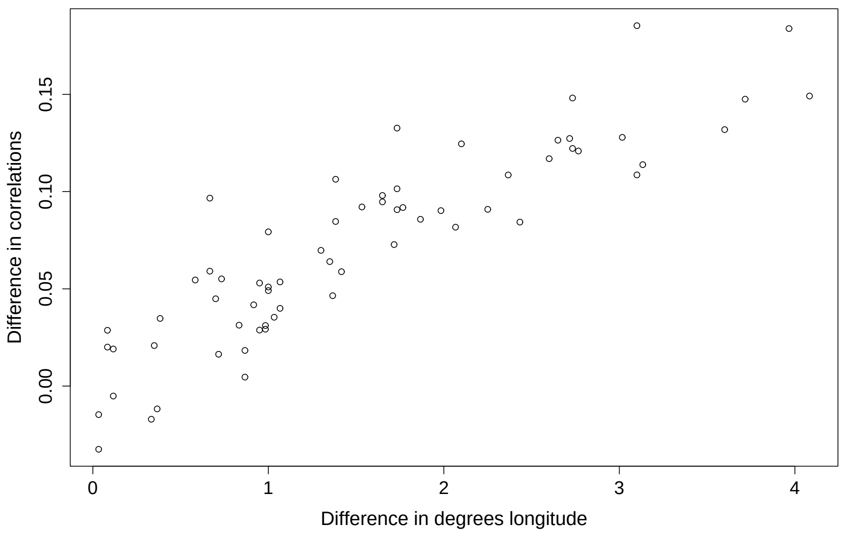


Figure 3


Difference in correlations
0.00
0.05
0.10
0.15
0
1
2
3
4
Difference in degrees longitude