Half-Spectral Space-Time Covariance Models

Michael T. Horrell & Michael L. Stein

*University of Chicago
Department of Statistics*

Abstract

We develop a new class of space-time Gaussian process models by specifying covariance functions using what we call a half-spectral representation. We establish general properties of half-spectral models and appeal to results on a screening effect analysis to claim more natural space-time interaction properties in the class of models we develop. To test this claim and to more generally test this class of models, we fit models we develop in this paper to a wind power dataset. We show our models fit these data better than other common separable and non-separable space-time models.

Keywords: Space-Time Processes, Spectral Density, Fourier Transform, Covariance Function

1. Introduction

Continuous natural phenomena in space or space-time are often modeled statistically as Gaussian processes. Generally, a Gaussian process model requires specification of a mean structure and a covariance or dependence structure. Since covariance models must be positive definite, mathematically valid
5 dependence structures can be quite difficult to define. Therefore, a primary focus in defining new Gaussian process models is on defining valid dependence structures. In this paper, we develop properties of a specific set of stationary space-time dependence structures using what we call a half-spectral approach.
10 We present a new classes of models built using this half-spectral approach and further develop properties of this class. In particular, our theoretical exploration

of these models focuses on space-time interaction properties, and we show by way of an application that careful consideration of space-time interaction can lead to better fitting models.

15 Half-spectral approaches to defining dependence structures have been studied previously by several authors ([1], [2], [3]). What distinguishes half-spectral approaches from other modeling methods is that valid dependence models can be defined using both spectral and non-spectral elements. Notably for this paper, half-spectral approaches can also enable direct modeling of space-time
20 interaction. Details and definitions of half-spectra and their formulations follow in Section 2.1.

The investigation in this paper is distinguished from other articles by our focus on developing the space-time interaction properties of the models we present. We believe this paper is novel in this regard. Commonly, models that may be
25 factored into purely spatial and purely temporal components—often referred to as separable models—are avoided [4]. Similarly, models with non-differentiable ridges or kinks in the covariance function have been shown to have undesirable properties ([5], [6]). In this paper we give positive and specific recommendations in terms of space-time interaction, and we build a class of space-time models
30 that meets these recommendations.

The space-time interaction behavior we seek is motivated by screening effect considerations [7]. A screening effect is an intuitively desirable property of some covariance models that implies that information at a target point in a random field is asymptotically completely determined by other points in the target
35 point’s local neighborhood. An overview of the screening effect and our complete analysis are in Section 2.2 and Section 3. To summarize, we develop our models by following a condition developed in [8] that guarantees a screening effect for non-differentiable models. Though this condition guarantees a screening effect in only some models, [8] also suggests it may lead to some generally desirable
40 model behavior as it will limit irregularities in the high-frequency components of a model.

The class of models we develop in this paper is a set of stationary non-

separable space-time dependence models. This class allows for nearly arbitrary marginal temporal dependence modeling while at the same time is Matérn-like marginally in space. In this way, arbitrarily and independently, different smoothnesses may be specified in time and space in this class. This class of models furthermore satisfies the screening effect condition in [8] under mild conditions on the marginal temporal spectrum.

Outside of half-spectral approaches, there are a host of techniques that can be used to produce valid non-separable space-time and other multidimensional covariance models. Convolution and mixtures methods provide a wide range of solutions [[9], [10], [11], [12], [13]]. Full spectral methods also give tremendous flexibility to produce stationary [[12], [5], [14]] and non-stationary models [15]. [16] approximates non-separable models using separable covariances. Also, expanding on the popular approach in [3], [17] provides characterization theorems and necessary conditions to produce space-time models with compact support, and [18] creates space-time models focusing on spatial anisotropy.

This paper is structured as follows. Section 2 introduces the theory and definitions of the half-spectral approach and describes in more detail the screening effect condition in [8]. Section 3 develops theory of certain half-spectral models with respect to the condition in [8]. Basic properties of the half-spectral models we develop are given, and we develop two necessary conditions for models to meet the screening condition. Section 4 presents the new class of models that meets the restrictions developed in Section 3. In Section 5, we fit examples of models from this class to the frequently used Irish wind dataset, and we compare these models to other separable and non-separable models.

2. Half-spectrum definitions and the screening effect condition

2.1. Half-Spectra

All continuous stationary covariance functions, K , for a real valued process on $\mathbb{R}^m$ can be written in the form $K(x) = \int \exp(i\xi^T x) F(d\xi)$ for F a symmetric

finite positive measure on $\mathbb{R}^m$ and $\xi \in \mathbb{R}^m$. If f exists such that $F(d\xi) = f(\xi)d\xi$, f is called the spectral density.

For a stationary, continuous space-time covariance model $K(s, t)$ with $s \in \mathbb{R}^d$ and $t \in \mathbb{R}$ with integrable spectral density, $g(\lambda, \omega)$, where $\lambda \in \mathbb{R}^d$ and $\omega \in \mathbb{R}$, we can represent K with g via Fourier transform,

$$K(s, t) = \frac{1}{(2\pi)^{d+1}} \int_{\mathbb{R}} \int_{\mathbb{R}^d} g(\lambda, \omega) \exp(is^T \lambda + it\omega) d\lambda d\omega. \quad (1)$$

Performing one of the two integration steps in (1) with respect to either λ or ω gives an expression we call either a half-spectral representation in space (if
75 integrating over ω) or time (if integrating over λ).

The specific half-spectra we study are half-spectra in time of the form $f(\omega)\mathbb{C}(s\delta(\omega)) \exp(i\theta(\omega)\phi^T s)$ defined such that

$$K(s, t) = \int_{\mathbb{R}} f(\omega)\mathbb{C}(s\delta(\omega)) \exp(i\theta(\omega)\phi^T s) \exp(it\omega) d\omega. \quad (2)$$

Here f is the spectral density of $K(0, t)$, the temporal covariance function at a single spatial site; $\mathbb{C}$ is any valid correlation function in $\mathbb{R}^d$ with a spectral density; and δ is an even non-negative function that determines the space-time interaction properties of the model. The odd, real function, $\theta(\cdot)$, and the unit
80 vector, $\phi \in \mathbb{R}^d$, determine space-time asymmetry in K by specifying a phase relationship. By Theorem 1 in [1], we know K thus defined is a valid covariance model.

Models not exhibiting any space-time asymmetry are called fully-symmetric [3]. We primarily consider the space-time symmetric case with $\theta(\omega) = 0$; hence,
85 we focus our study on half-spectra of the form $\psi(\omega, s) = f(\omega)\mathbb{C}(s\delta(\omega))$. However, as we show in Section 3, some important modeling properties of K will hold, independent of θ or ϕ . Restriction of our analysis to models defined by $\psi(\omega, s)$ should therefore not be seen as a simplification.

The space-time interaction properties of these models are determined by δ ,
90 which must be even and non-negative. However, when δ is constant, $K(s, t)$ becomes separable. In Section 3, we give specific recommendations on the form δ should take. Our recommendations depend on the screening effect analysis in

[8].

2.2. The screening effect condition

95 Screening effects generally refer to specific interpolation properties of a model and a random field. Often it is desired that closer observations to an interpolation site should best indicate what value the random field should take at that site. Or in other words, the use of far-off observations for interpolation purposes should be “screened out” by the close observations. This broad concept taken
100 as a whole is what is referred to as the screening effect [7].

[8] formalizes a version of this concept and builds a condition that guarantees a screening effect in some models. For a space-time spectral density, g , this condition can be written as

$$\lim_{\|(\lambda, \omega)\| \rightarrow \infty} \sup_{\|(u, v)\| < R} \left| \frac{g(\lambda + v, \omega + u)}{g(\lambda, \omega)} - 1 \right| = 0 \quad (3)$$

for all finite R . The condition in (3) essentially requires g to flatten out in a relative sense at high values of λ and/or ω .

The theorems in [8] are not completely general, but they, together with further results in [19], suggest that a space or space-time model with a spectral
105 representation that does not follow the condition in (3) may have undesirable interpolation properties. We therefore use this condition to guide development of our classes of half-spectral models and specifically the space-time interaction function, δ .

As further confirmation that the condition in (3) may describe natural mod-
110 els, a number of physically derived Gaussian process models will satisfy (3). For example, models defined using stochastic partial differential equations will often meet (3) when white noise forcing terms lead to processes with integrable spectral densities. In particular, Gaussian process solutions to stochastic versions of convection-diffusion equations will satisfy (3) [20].

115 Several classical models do not meet the condition in (3). Separable models in particular do not follow (3). In one or more dimensions, m , the squared exponential covariance, $K(x) = \exp(-\|x\|^2)$, $x \in \mathbb{R}^m$, which produces analytic

realizations is also excluded since its spectral density—which is proportional to $\exp(-\|\xi\|^2/4)$, $\xi \in \mathbb{R}^m$ —goes to zero too quickly as $\|\xi\| \rightarrow \infty$ to satisfy (3).
 120 The triangular covariance function $K(x) = \max(1 - |x|, 0)$, $x \in \mathbb{R}$ which has documented undesirable interpolation properties [6] is excluded as well since its spectral density, $2(1 - \cos(\xi))/\xi^2$, $\xi \in \mathbb{R}$, oscillates too violently to flatten out in the relative sense required by (3).

3. Restrictions on half-spectral forms

The half-spectral form in (2) in general does not meet the condition in (3). Let h be the spectral density of the spatial covariance function $\mathbb{C}$. The full spectral representation of K is

$$K(s, t) = \int_{\mathbb{R}} \int_{\mathbb{R}^d} f(\omega) \delta(\omega)^{-d} h\left(\frac{\lambda - \theta(\omega)\phi}{\delta(\omega)}\right) \exp(is^T \lambda + it\omega) d\lambda d\omega. \quad (4)$$

125 Let $g(\lambda, \omega)$ be the fully-symmetric spectrum, $g(\lambda, \omega) = f(\omega) \delta(\omega)^{-d} h(\lambda/\delta(\omega))$; therefore, we can write the general spectrum of K as $g(\lambda - \theta(\omega)\phi, \omega)$. To show when $g(\lambda, \omega)$ or $g(\lambda - \theta(\omega)\phi, \omega)$ satisfy (3), we first establish the result in Theorem 1: $g(\lambda, \omega)$ satisfies (3) if and only if $g(\lambda - \theta(\omega)\phi, \omega)$ satisfies (3) under a mild regularity condition requiring $\theta(\omega)$ to be locally bounded. This result
 130 permits the theoretical focus of this section to center on fully-symmetric models without a practical reduction of the generality of our results. With Theorem 1 established, the remaining theory in this section gives necessary conditions and restrictions on a fully-symmetric model (defined only by f , $\mathbb{C}$ and δ) that must be met in order for g to satisfy (3). Proofs of Theorem 1 and the Restrictions
 135 presented in this section are in the Supplemental Material.

Theorem 1. *Let $g(\lambda, \omega)$ be the full spectrum of a space-time covariance function where λ is the spatial wavenumber and ω is the temporal frequency. Let $\theta(\cdot)$ be an odd function, and let ϕ be a unit vector in $\mathbb{R}^d$. The following two statements hold:*

- 140 (i) *Let $\theta(\cdot)$ be locally bounded. The full spectrum $g(\lambda, \omega)$ satisfies (3) if and only if $g(\lambda - \theta(\omega)\phi, \omega)$ satisfies (3).*

(ii) Suppose $g(\lambda, \omega)$ is strictly positive and satisfies (3). If $g(\lambda - \theta(\omega)\phi, \omega)$ satisfies (3), then $\theta(\omega)$ must be locally bounded.

With Theorem 1 in mind, we focus on fully-symmetric half-spectra $\psi(\omega, s)$ with full spectra $g(\lambda, \omega)$. Since $g(\lambda, \omega) = f(\omega)\delta(\omega)^{-d}h(\lambda/\delta(\omega))$, we can more simply write g in terms of a marginal function in ω and a joint function in λ and ω . Let $g(\lambda, \omega) = p(\omega)q(\lambda, \omega)$. The following restrictions establish necessary conditions on p and q that must be met in order for g to satisfy the condition in (3).

Restriction 1. Let the spectral representation of covariance function K have the form $g(\lambda, \omega) = p(\omega)q(\lambda, \omega)$. The following two statements hold for spectral densities with this parameterization.

- (i) If g and q satisfy the condition in (3), then p must be constant in ω .
- (ii) If g satisfies the condition in (3), then for every point ω_0 such that $p(\omega_0)$ is finite and positive, the marginal spectrum $g_{\omega_0}^*(\lambda) = g(\lambda, \omega_0)$ must meet the condition in (3).

The implication of Restriction 1 is that the space-time interaction function δ cannot be determined completely independently of $\mathbb{C}$ and f if we wish to use models that follow the condition in (3). Of note is the fact that this restriction applies to several models proposed in [2], [3] and [1].

One may define the spatial interaction function as $\delta(\omega) = f(\omega)^{1/d}$ to escape the conditions of Restriction 1; however, $\delta(\omega)$ defined in this way decreases to zero for increasingly large ω since f must be integrable. However, δ decreasing to zero or having finite limit as ω approaches infinity will not generally lead to models satisfying the condition in (3).

Restriction 2. Let $g(\lambda, \omega) = p(\omega)q(\lambda/\delta(\omega))$, where $p(\omega)$ is positive and finite for all sufficiently large $|\omega|$, q is a non-negative, continuous, integrable function that is not identically zero, and $\delta(\omega)$ is even with a well-defined limit as $|\omega| \rightarrow \infty$. If g satisfies (3), then $\lim_{|\omega| \rightarrow \infty} \delta(\omega) = \infty$.

170 Restrictions 1 and 2 confine the form of δ . Implicitly, the form of h is also restricted. For half-spectral models satisfying (3), if δ must increase to infinity, $f(\omega)\delta(\omega)^{-d}$ cannot be constant, implying $h(\lambda/\delta(\omega))$ itself cannot satisfy (3). On the other hand, for fixed ω , $h(\lambda/\delta(\omega))$ must satisfy (3) in λ by part (ii) of Restriction 1. One way forward is to choose h and δ such that $h(\lambda/\delta(\omega))$ admits
175 the representation, $h(\lambda/\delta(\omega)) = H(\lambda, \omega)\delta(\omega)^d/f(\omega)$, where H satisfies (3). A half-spectral model thus specified will satisfy (3). Finally, by Theorem 1, with any locally bounded θ , $h((\lambda - \theta(\omega)\phi)/\delta(\omega))$ specified in this way will also lead to models satisfying (3).

4. A new class of models

180 We build a new class of fully-symmetric models; thus, $\theta(\omega) = 0$. Following Theorem 1, if $\theta(\omega)$ is non-zero, models with space-time asymmetry can be defined that will satisfy (3) if θ is locally bounded.

Let $h(\lambda) = \rho(\alpha^2 + \|\lambda\|^2)^{-(\nu+d/2)}$, the Matérn spectral density with smoothness $\nu > 0$, $\alpha > 0$ an inverse range parameter and $\rho > 0$, a variance parameter. For a given f , to ensure Restrictions 1 and 2 are not violated, choose $\delta(\omega) = f(\omega)^{-\frac{1}{2\nu}}$. Plugging into the space-time spectrum in (4), a new set of spectral densities can be written

$$g(\lambda, \omega) = \rho \left(\alpha^2 f(\omega)^{-1/\nu} + \|\lambda\|^2 \right)^{-(\nu+d/2)}. \quad (5)$$

If f , itself a spectral density, satisfies $f \sim c_f \omega^{-k}$ for some $k > 1$ and c_f a positive constant depending on f , then (5) describes a class of models that satisfies
185 (3). We write covariance functions with this half-spectrum K_f to indicate the temporal spectrum f determines a specific model in this class.

We use a Matérn spectral density in space to form this class of models for three reasons. First, it is a natural model to analyze since spatial modeling is often carried out using a Matérn covariances. Second, its form leads to easy
190 comparison with the conditions of Restrictions 1 and 2. Third, the Matérn spectrum gives (5) some desirable Matérn-like properties. For example, the pa-

parameter ν , is directly related to the spatial smoothness of the resulting Gaussian process. Example 1 below shows one application of this.

The half-spectral representation of (5) is given simply by plugging the Matérn covariance into the relevant portion of (2).

$$\psi(\omega, s) = f(\omega) \frac{\pi^{d/2} \rho}{2^{\nu-1} \Gamma(\nu + d/2) \alpha^{2\nu}} (\alpha \|s\| f(\omega)^{-\frac{1}{2\nu}})^\nu \mathcal{K}_\nu \left(\alpha \|s\| f(\omega)^{-\frac{1}{2\nu}} \right) \quad (6)$$

where $\mathcal{K}_\nu$ is the modified Bessel function of the second kind with argument ν [21]. If we let ρ absorb terms not depending on ω in (6) and write $\mathcal{M}_\nu(s) = \|s\|^\nu \mathcal{K}_\nu(\|s\|)$, (6) can be simplified to

$$\psi(\omega, s) = \rho f(\omega) \mathcal{M}_\nu \left(\alpha \|s\| f(\omega)^{-\frac{1}{2\nu}} \right). \quad (7)$$

Mean-square differentiable properties of this class of models are readily found. Following arguments in [5] and [22, p. 5], if f integrable and $f \sim c_f |\omega|^{-k}$ as ω approaches infinity, then a Gaussian space-time process governed by $K(s, t)$ will be m times mean-square differentiable in time if $(k - 1)/2 > m$. Similarly, mirroring the argument in [5, p. 313], a Gaussian space-time process governed by $K(s, t)$ will be m times mean-square differentiable in space if and only if $\nu(k - 1)/k > m$.

Therefore, with these models, smoothnesses in space and time can be defined via ν and the limiting behavior of f determined by k . We call $\nu(k - 1)/k$ and $(k - 1)/2$ the effective smoothnesses of the purely spatial and purely temporal processes respectively. Since $\nu > 0$ and $k > 1$, arbitrary effective smoothnesses in space and time can be specified using models in this class.

Example 1 (Matérn in Time)

Let $f(\omega) = (\beta^2 + \omega^2)^{-(\kappa+1/2)}$, a one dimensional Matérn spectral density, and let $\nu' > -1/2$. Plugging f into (5) and (7) and using $\nu' + 1/2$ in place of ν , we get the spectral and half-spectral forms.

$$\begin{aligned} g(\lambda, \omega) &= \rho \left(\alpha^2 (\beta^2 + \omega^2)^{(\kappa+1/2)/(\nu'+1/2)} + \lambda^2 \right)^{-(\nu'+(d+1)/2)}; \\ \psi(\omega, s) &= \rho (\beta^2 + \omega^2)^{-(\kappa+1/2)} \mathcal{M}_{\nu'+1/2} \left(\alpha \|s\| (\beta^2 + \omega^2)^{1/2 \frac{\kappa+1/2}{\nu'+1/2}} \right). \end{aligned} \quad (8)$$

Figure 1: Contour and marginal correlation plots for the Matérn in Time model in Example 1. Effective smoothness in space is 0.5. Effective smoothness in time is 2. Parameter values are: $\beta = 1$, $\alpha = 1/4$, $\nu' = 1/8$, $\kappa = 2$, $\rho = \beta^{2\kappa}\Gamma(\kappa + 1/2)/(2\nu'^{-1/2}\Gamma(\kappa)\Gamma(\nu' + 1/2)\sqrt{\pi}) \approx 0.678$

Here $k = 2\kappa + 1$; hence this process has effective spatial smoothness of $\kappa \frac{\nu'+1/2}{\kappa+1/2}$ and effective temporal smoothness of κ . This class extends the Matérn class. If $\kappa = \nu'$, K_f reduces to a $d + 1$ dimensional Matérn model. A numerically
215 calculated contour plot for $\nu' = 1/8$, $\kappa = 2$ is in Figure 1(a). Marginal plots of $K_f(0, s)$ and $K_f(t, 0)$ are in Figure 1(b).

Of note in Figure 1(a) is the fact that a “dimple”-type property is present in this model [23]. [23] gives a definition of a dimple; however, a detail of the contour plot in Figure 1 (see Supplemental Material) shows that K_f in the
220 example in Figure 1 does not precisely meet their definition of a dimple. We remark more generally on the dimple shape formed by the non-convex shape of the contour lines of K_f . In our setting, we say a dimple exists if there exist lags s and t such that $K(s, t) > K(0, t)$ and $K(-s, t) > K(0, t)$.

Empirically, the dimple in Figure 1(a) appears to be caused by the large
225 difference in κ and ν' and $\nu' < \kappa$. For κ and ν' closer together or $\nu' > \kappa$, numerical experiments suggest this dimple either disappears or gets pushed further from the origin. Note finally that this model is a special case of a class of models considered by [5], but the half-spectral form of this model is not given in [5] nor is the connection of this model to Matérn models mentioned therein.

230 5. Application

We apply examples from the class of models presented in Section 4 to the Irish wind dataset first analyzed by [24]. We compare fits from models in either class to similarly parameterized separable and non-separable models. In making these comparisons, we focus on space-time interaction properties of the models we fit; hence, when possible, marginal spatial and temporal covariances and smoothnesses are fixed at common values across the models we fit. We focus on space-time interaction to show the models we develop may be natural first choices to consider when analyzing space-time data.

The Irish wind dataset is a record of daily average wind speeds collected at 12 sites in Ireland from 1961 to 1978. This dataset has been analyzed several times previously by [24], [3], [25] and [1] among others. Like these authors, we deseasonalize the data by regressing wind-speed on a few (four) annual harmonics, and we discard data from the Rosslare site due to nonstationarities. We also take square-roots to obtain more marginally Gaussian data, and we subtract station means of these deseasonalized and transformed data. Finally, we divide each station's time series by its respective sample standard deviation. The resulting dataset contains $6,574 \times 11 = 72,314$ observations with none missing.

There are subtleties in these data. Most notably these data have been documented as having longer memory in time than covariance models with exponential tails can produce [24]. Therefore, to make best comparisons within groups of models, we separate comparisons based on the temporal tail behavior (exponential tails versus algebraic tails) of the covariance models we consider.

The wind process also has properties that are not fully-symmetric [3]. The primary model comparisons we make in this section involve fully-symmetric models because these provide the best comparisons to established models. However, at the end of the application we fit a model that does not exhibit full-symmetry by using a simple model for the phase shift present in (2).

Since these data are collected at regular intervals through time but at dif-

260 ferent spatial locations, these data are an example of regular monitoring data
[1]. In-depth details on computation and fitting models to these data can be
found in the Supplemental Material. Generally, we fit models using two meth-
ods. Primarily we maximize a Whittle likelihood to find parameter estimates
since this method can be used to directly fit half-spectral forms [1]. When we
265 compare half-spectral models to models that do not have explicit half-spectral
forms, however, we instead compare true likelihoods by numerically calculating
(2) for the half-spectral models.

5.1. Results

5.1.1. Models with Exponentially Decaying Temporal Tails

270 We fit the model in Example 1 to these data. For comparison, we fit a Matérn
model as well as a separable model with exponential marginal covariances. We
also fit a model in [2] that was found to best fit a different dataset of wind speeds.
The half-spectral form of the model in [2] is $\rho(\beta^2 + \omega^2)^{-1} \exp(-\alpha^2 \|s\|^2)$. Note
 δ is constant in this model due to its separability. [2] uses variograms to fit the
275 model. In the model variogram they include a term related to $\|s\|$. Since we
subtract site means prior to analysis of the Irish wind data, we do not need to
use an analog of this term in our models.

The only modification we make to all four exponential temporal tail models
is to add a spatial nugget similar to [3] and [1]. The half-spectral representations
of our models therefore take the general form

$$f(\omega) [\mathbb{C}(s\delta(\omega)) + \eta^2 \mathbb{I}_{\{s=0\}}] \quad (9)$$

where $\mathbb{I}_{\{ \cdot \}}$ is an indicator function and $\eta^2 \geq 0$.

For models with flexibility in determining smoothness in time (Example 1
280 and the Matérn model), temporal smoothness parameters, κ , were fixed at $1/2$
to allow for better comparison of all models. Note for the Matérn model, this
also fixes the spatial smoothness. To evaluate the assumption that $\kappa = 1/2$, we
fit a univariate Matérn model to a subset of the time series at each station, and
we found $\kappa = 1/2$ to be a reasonable estimate of temporal smoothness.

Table 1: Comparison of log-likelihoods of short memory in time models fitted to the Irish wind dataset.

Whittle Log-Likelihoods of Models with Exponential Temporal Tails			
	Log-Likelihood	Diff. From Ex. 1	# Param. Fit
Example 1 ($\kappa = 1/2$)	20,318	0	5
Matérn (Example 1, $\kappa = \nu' = 1/2$)	20,199	119	4
Separable Exponential	18,703	1,615	4
[2]	18,378	1,940	4

285 A comparison of the multivariate Whittle likelihoods is in Table 1. Here, the model from Example 1 is shown to provide the best fit for these data. That Example 1 outperforms a Matérn model is expected since Example 1 extends the Matérn class. The added flexibility in specifying the spatial smoothness appears to add a small amount to the log-likelihood in this case. With Example
290 1, the effective smoothness in space is estimated to be 0.4; hence a modest improvement in log-likelihood is not surprising. For processes estimated to have drastically different smoothnesses in space and time, we may expect a larger improvement in using Example 1 over a Matérn model.

A look at the remaining fits shows smoothness and space-time interaction
295 matter in terms of model fitting. For example, the Matérn model compared to the separable exponential model shows an increase in 1,496 of the log-likelihood. Both models have the same number of fitted parameters and the same smoothnesses in space and time; yet, the Matérn appears able to capture the space-time interaction of the Irish wind process better. Comparing the separable
300 exponential to the model in [2], we see modeling variations in space as squared exponential is perhaps unreasonable.

5.1.2. Models with Algebraic Temporal Tails

[3] uses the following covariance model, denoted by G , to analyze these data:

$$G(s, t) = \rho(\beta|t|^\ell + 1)^{-1} \exp\left(-\frac{\alpha\|s\|}{(\beta|t|^\ell + 1)^{\gamma/2}}\right) + \eta^2(\beta|t|^\ell + 1)^{-1}\mathbb{I}_{\{s=0\}} \quad (10)$$

where $\beta > 0$, $\ell \in [0, 2]$ and $\gamma \in [0, 1]$. The last term in (10) is the spatial nugget as also seen in (9). The temporal marginal covariance $G(0, t)$ has algebraic tail behavior. When $\ell \leq 1$, G is a classical long memory model in time. In the case where $\ell > 1$, $G(s, t)$ is not formally long memory in time since $G(0, t)$ will be integrable; however, the algebraic decay in $G(0, t)$ for large t is much slower than the exponential decay present in the models in Table 1 and fits the Irish wind data better than exponentially decaying models in Section 5.2.1.

To fit G using a multivariate Whittle likelihood method, the half-spectrum of G is needed, but we do not know of a general closed form half-spectral representation of G currently. However, specific separable half-spectral forms of G can be found. For $\gamma = 0$, G is separable and exponential in space. For $\ell = 1$, a closed form for the spectral density of $G(0, t)$ is readily calculated (derivation in the Supplemental material); hence, the half-spectrum of a separable G that is long-memory in time can be specified as long as $\ell = 1$ and $\gamma = 0$ are fixed. Note that [3] estimated $\kappa = 1.544$. If we instead let $\ell = 2$, a closed form spectrum for $G(0, t)$ can also be found, but it is exponential and thus $G(0, t)$ for this model will not meet the condition in (3). With $\ell = 1$, $G(0, t)$ meets the condition in (3) since the spectrum of $G(0, t)$ denoted f_G decays proportional to ω^{-2} (details in the Supplemental Material).

Since f_G decays asymptotically proportional to ω^{-2} , we can plug f_G into the class in (7) to obtain K_{f_G} , a model that satisfies the condition in (3). Fitting both K_{f_G} with a spatial nugget and G with $\ell = 1$ and $\gamma = 0$ by their respective half-spectra, we obtain the results in Table 2. We fix $\nu = 1$ for K_{f_G} to achieve the same smoothness in space for K_{f_G} as exists for G with $\gamma = 0$ and $\ell = 1$.

Table 2 shows again that compared to similar non-separable models, a separable model will not adequately fit these data. This exercise shows further that the class of models presented in Section 4.1 is quite flexible and can be adapted straightforwardly to many different scenarios. Comparing the log-likelihoods of

Table 2: Comparison of log-likelihoods of longer memory models fitted to the Irish wind dataset.

Whittle Log-Likelihoods of Models with Algebraic Temporal Tails				
		Log-Likelihood	Difference From K_{f_G}	# Parameters Fit
K_{f_G}	$(\nu = 1)$	21,655	0	4
G	$(\gamma = 0, \ell = 1)$	21,245	410	4

the short memory models in Table 1 to the longer memory models in Table 2, we find that G with $\gamma = 0$, a separable model, to better fit these data than the non-separable models in Table 1. The difference between G in Table 2 and Example 1 from Table 1 is 927 Whittle log-likelihood units. This is not surprising given that long memory properties of these data are apparent in looking at marginal correlation plots [24].

As a final comparison, we fit K_{f_G} and a non-separable version of G by maximizing space-time likelihoods (as opposed to the frequency domain Whittle likelihoods). The size of this dataset (72,314 observations) is quite difficult for space-time likelihood fitting of these models without making use of computational shortcuts. Specifically, for these data, calculating a log-determinant and a solve involving a $72,314 \times 72,314$ covariance matrix is required to compute the space-time likelihood under the multivariate Gaussian assumption; however, even storing a matrix of this size in RAM will not be possible on most machines as it requires over 40GB of memory using standard precisions. To perform this calculation, we make use of the regular monitoring structure of these data and the temporal stationarity of these models. Sorting the data first by time and second by location leads to a Block-Toeplitz covariance matrix with blocks of size 11×11 . Using a Block-Toeplitz Levinson-type algorithm, one can compute the log-determinant and solve terms simultaneously with greatly reduced computational burden [26].

The covariance model G is given in its space-time representation in (10). The space-time covariance function K_{f_G} must be computed numerically from

Table 3: Comparison of log-likelihoods of longer memory models fitted to the Irish wind dataset. In contrast to Table 2, G is a non-separable model.

Space-Time Log-Likelihoods of Models with Algebraic Temporal Tails				
		Log-Likelihood	Difference From K_{f_G}	# Parameters Fit
K_{f_G}	$(\nu = 1)$	21,471	0	4
G	$(\ell = 1)$	21,327	144	5

the half-spectrum specified by (6). We approximate the covariance function
355 K_{f_G} via a finite approximation of the integral in (2). Details and computation times can be found in the Supplemental Material.

To ensure G and K_{f_G} are most comparable, we fix $\nu = 1$ for K_{f_G} and $\ell = 1$ for G . With these parameter values, both models have the same smoothnesses in space and (up to range and scale parameters) identical marginal temporal
360 covariance models. For G , the separability parameter γ is estimated from the data. Results from fitting these models are in Table 3.

Model K_{f_G} fits the data better than G while simultaneously requiring fewer estimated parameters. The separability parameter γ is estimated to be 0.5; hence G is non-separable. The marginal spatial and marginal temporal covari-
365 ances in both models are quite similar; thus the slight edge K_{f_G} appears to have over G in this case may be due to the ability of K_{f_G} to better capture space-time interactions in these data. Comparing Table 3 to Table 2, we see a difference in likelihoods for K_{f_G} . Due to the approximations made in calculating both Whittle-type and space-time likelihoods, some discrepancy is expected.

370 5.1.3. A Space-time Asymmetric Model

Starting with K_{f_G} , we generate a space-time asymmetric model, $\tilde{K}_{f_G}$, by picking $\theta(\omega)\phi$ to be non-zero. We use a simple linear translation model for the phase shift, giving $\theta(\omega) = r\omega$, where $r \in \mathbb{R}$. The spatial direction of this shift is determined by ϕ . Since $\theta(\omega)$ is locally bounded, $\tilde{K}_{f_G}$ will satisfy (3). [3]
375 remarks on the east-west asymmetry in these data; hence, for simplicity, we fix $\phi \in \mathbb{R}^3$ to be a unit vector pointing in a direction consistent with an east-to-west

direction over Ireland. In principle, ϕ can be estimated from the data. Lastly, for best comparison to K_{f_G} in Table 2, ν is again fixed at 1 for $\tilde{K}_{f_G}$.

Using Whittle likelihood fitting, $\tilde{K}_{f_G}$ has a log-likelihood of 21,922. Comparing to K_{f_G} in Table 2, we see a marked advantage to taking into account space-time asymmetry with these data. The increase in Whittle log-likelihood is 267 units. The value r was estimated to be positive; hence, $\tilde{K}_{f_G}$ is consistent with the simple measures of asymmetry used in [3].

Lastly, it is important to note that in the context of modeling a process on the entire globe, this phase shift model will not be entirely sensible. The vector $r\omega\phi$ indicates a linear phase shift through 3-dimensional Euclidean space. Our data are on the globe however; therefore, an east-west phase shift is not everywhere accurately described by the single vector ϕ . For example, while ϕ is an east-to-west vector over Ireland, it is a west-to-east vector on the opposite side of the globe (for example, in Japan). A better way to model phase shifts on the globe in the context of half-spectral models is given in [1], wherein phase shifts are akin to rotations of the globe. In this type of phase shift model, phases depend on differences in longitude and/or latitude and not a lag vector in $\mathbb{R}^3$. Unfortunately, rotation based phase models do not fit within the theory presented in Section 3. Therefore, we have opted to use the locally accurate phase shift described by $r\omega\phi$. Since Ireland covers only a small range of longitudes, using this simplification should not affect results in any substantial way.

6. Discussion

We developed a new class of continuous space-time dependency models via half-spectra defined by the form $\psi(\omega, s) \exp(i\theta(\omega)\phi)$. This class of models allows for substantial flexibility in modeling the marginal processes defined by $K(s, 0)$ and $K(0, t)$. Moreover, this class of models is non-separable. These models are further theoretically validated by meeting both restrictions we developed in Section 3 to guide our model building. These restrictions were developed through consideration of a natural condition posed in [8]. Our model building

focused on fully-symmetric models defined by $\psi(\omega, s)$; however, Theorem 1 in Section 3 established that inducing space-time asymmetry in these models using a non-zero $\theta(\omega)$ can be done naturally and easily.

We compared fully-symmetric examples from the new classes of models in this paper to separable models and to a non-separable model developed in [3] by fitting these models to the Irish wind dataset in Section 5. Models from the class of half-spectra defined in (6) were shown to fit the Irish wind data better than all other models we examine in this paper in terms of likelihood. [24] noted these data exhibit long-memory properties in time, and in Section 4, we showed the flexibility of the first class of models by fitting a model that is long-memory in time that also meets the condition (3). We also fit an adaptation of this model using a non-zero $\theta(\omega)$ to illustrate the space-time asymmetry present in these data. Half-spectral models of the form in (7) therefore appear to be quite flexible and may be a good starting point for many model building purposes.

The half-spectral forms present in this paper serve as one step toward a more coherent melding of spatial and temporal methods in statistics. A primary avenue of future research may be to develop computationally efficient fitting procedures for half-spectral models on data that do not have the nice structure of non-missing regular monitoring data. Adaptation of composite likelihood methods [[27], [28], [29], [30]] may serve as one way forward as these methods allow for considerations of only subsets of the full dataset.

7. Acknowledgment

The first author was supported by Department of Education award P200A100171 and STATMOS, an NSF-funded Network (NSF-DMS awards 1106862, 1106974 and 1107046). The second author was supported by the U.S. Department of Energy Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics program under Award Number DE-SC0002557.

References

- [1] M. Stein, Statistical methods for regular monitoring data, J. R. Statist. Soc. Ser. B 67 (2005) 667–687.
- [2] N. Cressie, H.-C. Huang, Classes of nonseparable, spatio-temporal stationary covariance functions, J. Amer. Statist. Assoc. (1999) 1330–1340.
- [3] T. Gneiting, Nonseparable, stationary covariance functions for spacetime data, J. Amer. Statist. Assoc. 97.
- [4] P. Kyriakidis, A. Journel, Geostatistical space-time models: a review, Mathematical Geology 31 (1999) 651–684.
- [5] M. Stein, Space-time covariance functions, J. Amer. Statist. Assoc. 100 (2005) 310–321.
- [6] M. Stein, Interpolation of Spatial Data: Some Theory for Kriging, Springer, 1999.
- [7] J. Chilès, P. Delfiner, Geostatistics: Modeling Spatial Uncertainty, Wiley, 1999.
- [8] M. Stein, 2010 Rietz lecture: When does the screening effect hold? Ann. Stat. 39 (2011) 2795–2819.
- [9] P. Brown, K. Kåresen, G. Roberts, S. Tonellato, Blur-generated nonseparable space-time models, J. R. Statist. Soc. Ser. B 62 (2000) 847–860.
- [10] C. Ma, Families of spatio-temporal stationary covariance models, J. Stat. Plan. and Infer. 116 (2003) 489–501.
- [11] E. Porcu, J. Mateu, G. Christakos, Quasi-arithmetic means of covariance functions with potential applications to space-time data, Journal of Multivariate Analysis 100 (2010) 1803–1844.

- [12] A. Kolovos, G. Christakos, D. Hristopulos, M. Serre, Methods for generating non-separable spatiotemporal covariance models with potential environmental applications, *Advances in Water Resources* 27 (2004) 815–830.
- 460 [13] T. Fonseca, M. Steel, A general class of nonseparable space-time covariance models, *Environmetrics* 22 (2011) 224–242.
- [14] M. Fuentes, L. Chen, J. Davis, A class of nonseparable and nonstationary spatial temporal covariance functions, *Environmetrics* 19 (2007) 487–507.
- [15] E. Porcu, P. Gregori, J. Mateu, Archimedean spectral densities for nonstationary space-time geostatistics, *Statistica Sinica* 19 (2009) 273–286.
- 465 [16] M. Genton, Separable approximations of space-time covariance matrices, *Environmetrics* 18 (2007) 681–695.
- [17] V. Zastavnyi, E. Porcu, Characterization theorems for the gneiting class of space-time covariance functions, *Bernoulli* 17 (2011) 456–465.
- 470 [18] E. Porcu, P. Gregori, J. Mateu, Nonseparable stationary anisotropic space-time covariance functions, *Stochastic Environmental Research and Risk Assessment* 21 (2006) 113–122.
- [19] M. Stein, When does the screening effect not hold? *Spatial Statistics* 11 (2014) 65–80.
- 475 [20] F. Lindgren, H. Rue, J. Lindström, An explicit link between gaussian fields and gaussian markov random fields: the spde approach, *J. R. Statist. Soc. B* 73 (2011) 423–498.
- [21] F. Olver, D. Lozier, R. Boisvert, C. Clark, *NIST Handbook of Mathematical Functions*, National Institute of Standards and Technology and Cambridge University Press, 2010.
- 480 [22] E. Stein, G. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces*, Princeton University Press, 1971.

- [23] J. Kent, M. Mohammadzadeh, A. Mosammam, The dimple in Gneiting's spatial-temporal covariance model, *Biometrika* 98 (2011) 489–494.
- 485 [24] J. Haslett, A. Raftery, Space-time modeling with long-memory dependence: Assessing ireland's wind power resource, *Appl. Stat.* 38 (1989) 1–50.
- [25] X. de Luna, M. Genton, Predictive spatio-temporal models for spatially sparse environmental data, *Statistica Sinica* 15 (2005) 547–568.
- [26] H. Akaike, Block toeplitz matrix inversion, *SIAM J. Appl. Math.* 24 (1973) 234–241.
- 490 [27] B. Lindsay, Composite likelihood methods, *Contemporary Mathematics* 80 (1988) 220–239.
- [28] A. V. Vecchia, Estimation and model identification for continuous spatial processes, *J. R. Statist. Soc. B* 50 (1988) 297–312.
- 495 [29] M. Stein, Z. Chi, L. Welty, Approximating likelihoods for large spatial data sets, *J. R. Statist. Soc. B* 66 (2004) 275–296.
- [30] M. Bevilacqua, C. Gaetan, J. Mateu, E. Porcu, Estimating space and space-time covariance functions for large data sets: A weighted composite likelihood approach, *J. Amer. Statist. Assoc.* 107 (2013) 268–280.