Spatio-Temporal Modelling using Integro-Difference Equations with Bivariate Stable Kernels

Robert Richardson

Brigham Young University, Provo, Utah, U.S.A.

Athanasios Kottas and Bruno Sansó

University of California, Santa Cruz, California, U.S.A.

Abstract

An integro-difference equation can be represented as a hierarchical spatio-temporal dynamic model using appropriate parameterizations. The dynamics of the process defined by an integro-difference equation depends on the choice of a bivariate kernel distribution, where more flexible shapes generally result in more flexible models. Under a Bayesian modelling framework, we consider the use of the stable family of distributions for the kernel, as they are infinitely divisible and offer a variety of tail behaviors, orientations and skewness. Many of the attributes of the bivariate stable distribution are controlled by a measure, which we model using a flexible Bernstein polynomial basis prior. The method is the first attempt to incorporate non-Gaussian kernels in a two-dimensional integro-difference equation model and will be shown to improve prediction over the Gaussian kernel model for a data set of Pacific sea surface temperatures.

Keywords: Bernstein polynomials, Elliptically contoured stable distributions, Fourier series, Markov chain Monte Carlo, Semiparametric Bayesian modelling

1 Introduction

Spatio-temporal data analysis requires the specification of spatial and temporal dynamics. In separable models, the space and time components are modeled independently. However, spatio-temporal

processes that occur naturally in fields such as climate, ecology, and epidemiology often exhibit an interaction between the spatial and temporal dynamics. Learning the nature of these space-time interactions is a key component of spatio-temporal modeling. Hierarchical dynamic spatio-temporal models (Wikle & Hooten, 2010) provide a framework that can capture a wide variety of non-separable model parameterizations. We explore a particular parameterization of a hierarchical dynamic spatio-temporal model that offers a great deal of flexibility in capturing complicated physical characteristics, while providing computational tractability.

Denoting by $X_t(s)$ the process at time t and spatial location s, an integro-difference equation (IDE) model can be written as

$$X_{t+1}(s) = \int k(u - s \mid \theta)\, X_t(u)\, \mathrm{d}u \; + \; \varepsilon_t(s) \tag{1}$$

where $\varepsilon_t(s)$ is a zero mean error process which may be spatially colored, and k is a bivariate redistribution kernel with parameter set θ. IDE models have been used in space-time modeling since Wikle & Cressie (1999). The use of Fourier basis functions and the extension to IDE kernels with spatially varying parameters is studied in Wikle (2002) and Xu et al. (2005). Moreover, IDE models have been used for non-linear space-time systems (Wikle & Holan, 2011). Modelling choices and estimation procedures have also been explored in the engineering literature (e.g., Scerri et al., 2009).

The concept of an integro-difference equation is that the process at time t is propagated in time by re-weighting it using kernel k, acting in space. This model provides a simple, yet effective and interpretable way to express the relationship between the space and time components, using linear dynamics. As the kernel shifts or expands, the nature of the dependence between the process at a specific location and the process at previous time points changes accordingly. In fact, the first two moments of the kernel determine, respectively, the advection and the diffusion of the process (Brown et al., 2000; Storvik et al., 2002). Richardson et al. (2017) show that IDEs with non-Gaussian kernels allow for the description of higher order properties in the spatio-temporal process. Thus, although complex spatial fields will often exhibit non-linear dynamics, non-Gaussian kernels can help explain some of the complexity. This feature can be enhanced by making the kernel parameters vary across the spatial region. The introduction of spatial non-stationarity in the model is then used to capture the physical properties of the field of interest, through the characteristics of the kernels (Lemos &

Sansó, 2009). This is the approach we follow here. In contrast, Wikle & Holan (2011) and Cressie & Wikle (2011) detail methods that account explicitly for non-linearity. Such approaches offer powerful modelling tools, but shoulder the burden of increased computational complexity and run time.

In essentially all applications of IDE models, Gaussian kernels are used, owing to the convenience in their specification and computation. Richardson et al. (2017, 2018) give strong evidence that inference and prediction may be improved by using non-Gaussian kernels, but due to computational difficulties of extending the methods to two dimensions, the results are limited to one dimension.

To apply flexible IDE modeling to spatio-temporal data, we propose a bivariate stable kernel due to its ability to capture varying tail behavior and levels of skewness. In addition, stable kernels are infinitely divisible. This property is important to guarantee that the evolution equation of the process remains valid when the time scale is changed (Brown et al., 2000). An important feature of the bivariate stable distribution is that it is defined through a measure Γ which dictates key distributional properties, such as skewness, spread, and orientation. By flexibly modeling this measure we can achieve a wide variety of kernel shapes in two dimensions, resulting in general IDE models. To our knowledge, this is the first attempt at comparing Gaussian versus non-Gaussian kernels in IDEs for two-dimensional space. This work also offers a novel Bayesian semiparametric modelling approach for the bivariate stable distribution. Finally, while the flexibility afforded by the proposed model is greater than that of Gaussian kernel models, the computational burden is not significantly larger.

In addition to general bivariate stable kernels, we explore IDE models with bivariate elliptically contoured stable (ECS) kernels (Nolan, 2013). The family of ECS distributions includes the Gaussian as a special case, as well as distributions with polynomial tails. More specifically, the Gaussian family is a subset of ECS distributions, which in turn are contained in the stable distributions family. While the ECS family does not have the same flexibility as general stable kernels, IDE models with ECS kernel distributions are essentially as simple to implement as Gaussian kernel IDE models and can directly address whether heavy tails or skewness are affecting differences in model fit.

A data set that will benefit from a stable kernel compared to Gaussian kernels is one that has long-range lagged dependence or has a highly complicated system of lagged dependence that can not be accurately represented by simpler kernels. Pacific sea surface temperatures are an example of a system that displays complicated dynamics, often described through non-linear models (Cressie & Wikle, 2011; Richardson, 2017). We will show that these dynamics can be better explained by

flexibly modeling the dependence structure from one time point to the next through non-Gaussian IDE models. IDE models have been applied to sea surface temperature data (Wikle & Holan, 2011), although the specific nature of the kernel distribution was not the focus of that application.

Section 2 presents the general modelling approach, with emphasis on properties of stable distributions relevant to IDE modelling. A flexible, yet parsimonious model for the spectral measure Γ that characterizes the stable distribution is developed in Section 3. The posterior simulation method is discussed in Section 4. Section 5 details the analysis of sea surface temperature data, and demonstrates how bivariate stable and ECS kernels improve prediction over Gaussian kernel IDE models. Finally, Section 6 concludes with a summary.

2 Stable kernel IDE models

The usual approach to fitting IDE models is to convert Equation (1) into a dynamic linear model, conditional on the kernel and error process parameters. This is achieved by using a set of orthogonal basis functions to represent both the kernel and the spatio-temporal process (e.g., Wikle, 2002). Of particular interest is the Fourier basis, as the corresponding coefficients can be obtained from the characteristic function of the kernel. This is the approach that we follow to obtain a model for IDE kernels with flexible skewness and tail behavior based on the family of stable distributions.

More specifically, for a given set of basis functions, $\psi_i(s)$, the process is written as $X_t(s) = \sum_{i=1}^{\infty} a_i(t)\psi_i(s)$ and the kernel as $k(u-s \mid \theta_s) = \sum_{j=1}^{\infty} b_j(s,\theta_s)\psi_j(u)$, considering IDE kernels with spatially varying parameters. Then, any orthonormal basis leads to the process representation at time $t+1$ as $X_{t+1}(s) = \sum_{i=1}^{\infty} a_i(t)b_i(s,\theta_s)$. The process coefficients, $a_i(t)$, must be estimated, whereas, given parameters θ_s, the kernel coefficients, $b_j(s,\theta_s)$, are determined by the choice of kernel distribution.

Consider a spatial data vector $Y_t = (Y_t(s_{t,1}), ..., Y_t(s_{t,n_t}))'$ where the number of data points and the locations of observations may change over time. We will utilize process convolutions to represent the spatially varying IDE kernel parameters θ_s (see Sections 3 and 5.2). Denote by φ the parameter vector of the model for θ_s. Then, the hierarchical model for the data can be expressed as

$$Y_t \mid a_t, \sigma^2 \sim \mathrm{N}(\Psi_t a_t, \sigma^2 I_{n_t}), \quad t = 1, ..., T \tag{2}$$

$$a_t \mid a_{t-1}, \tau^2, \varphi \sim \mathrm{N}(G_t\, B_{\varphi,t}\, a_{t-1}, \tau^2 G_t V_t G_t'), \quad t = 1, ..., T \tag{3}$$

where the i-th row of Ψ_t contains the Fourier basis functions evaluated at the i-th spatial location at time t, a_t contains the stochastic basis function coefficients for the process (which will be estimated), $G_t = (\Psi_t'\Psi_t)^{-1}\Psi_t'$, and the i-th row of $B_{\varphi,t}$ contains the coefficients for the IDE kernel basis expansion (which are specified given φ). The observational variance $\sigma^2 I_{n_t}$ is a scaled identity. The matrix V_t is a spatial covariance that is transformed spectrally by G_t and scaled by τ^2. More details on how this formulation is derived for a general kernel is found in Wikle (2002) and Xu et al. (2005).

We develop spatio-temporal IDE models based on general stable and ECS distributions for the kernel $k(\cdot \mid \theta_s)$. In Section 2.1, we establish Gaussian kernel IDE models as a special case of our modelling framework, and study the benefits of the general skewness and tail behavior of stable distributions. Section 2.2 provides the Fourier basis representation for general stable and ECS kernels.

2.1 Model properties

A random vector $X \in \mathbb{R}^d$ belongs to the stable family of distributions if and only if for any constants $a > 0$ and $b > 0$ and independent copies $X^{(1)}$ and $X^{(2)}$ of X, there exists a constant $c > 0$ and vector $D \in \mathbb{R}^d$ such that $aX^{(1)} + bX^{(2)}$ and $cX + D$ are equal in distribution (Samorodnitsky & Taqqu, 1997). Many properties stem from this definition, including infinite divisibility. IDEs with infinitely divisible kernels can be connected to well-studied partial differential equation systems, and thus an infinitely divisible kernel is an attractive choice. As it comprises distributions without finite moments, the stable family has been applied to model heavy-tailed distributions (Panorska, 1996; Nolan, 2014).

Lacking a closed-form expression for its density, the stable distribution is typically defined through its characteristic function. For $\alpha \neq 1$, and for $d \geq 2$, the d-dimensional stable characteristic function:

$$g(t \mid \alpha, \mu, \Gamma) = \exp\left\{ it'\mu - \int_{v \in S_d} |t'v|^\alpha \left[1 - i \ \text{sign}(t'v) \tan(\pi\alpha/2)\right] \mathrm{d}\Gamma(v) \right\}, \tag{4}$$

where $\mu \in \mathbb{R}^d$ is a location parameter vector, $\alpha \in (0, 2]$ controls tail behavior, and Γ is a finite measure that controls characteristics of the distribution such as skewness, orientation, and spread. (Depending on the context, we use Γ to denote either the finite measure or its distribution function.) The set S_d is the unit sphere in $\mathbb{R}^d$, that is, the unit circle, S_2, in two dimensions. In this case, a change of variables can be made to $v = (\cos(z), \sin(z))'$ where the integral is now over $z \in [0, 2\pi]$. The derivative of the distribution Γ, which we will denote as $\gamma = \mathrm{d}\Gamma$, must be non-negative, but can take on a variety of

shapes. The skewness of the bivariate stable distribution changes when $\gamma(z)$ and $\gamma(z+\pi)$ are more disparate, and its orientation rotates with a shift in γ. The spread of the distribution changes with the scale of the measure; a measure $\gamma(z)$ with larger values for all z yields larger spread.

We notice that Equation (4) is defined in the literature for $\alpha \in (0, 2)$ (Samorodnitsky & Taqqu, 1997). Press (1972) indicates that the $\alpha = 2$ case corresponds to a Gaussian distribution for a different form of the characteristic function. When considering Equation (4), the case $\alpha = 2$ is explicitly excluded. However, for modelling purposes, $\alpha = 2$ is of particular interest, as it determines the change from polynomial to exponential tails. Thus, estimating α yields useful qualitative information about the IDE kernel. This motivates the following result, which, to our knowledge, is novel.

Lemma 1. *The expression given in Equation* (4) *with* $\alpha = 2$ *is the characteristic function for a Gaussian distribution.*

Proof. Let X be a random vector with characteristic function given by Equation (4) with $\alpha = 2$. Then, for any vector $b = (b_1, ..., b_d)'$, the characteristic function of the linear combination $Y = b'X$ can be written as $\exp\left\{it(b'\mu) - t^2 \int_{v \in S_d} (b'v)^2 \mathrm{d}\Gamma(v)\right\}$, with scalar argument t. This is the characteristic function for a univariate Gaussian distribution with mean $b'\mu$ and variance $2\int_{v \in S_d} (b'v)^2 \mathrm{d}\Gamma(v)$. Since all linear combinations of X are Gaussian, X is a Gaussian random vector. □

The lemma implies that our stable kernel IDE model includes the Gaussian kernel model as a special case. The validity of a Gaussian kernel IDE model can thus be evaluated through the prior to posterior updating. This is illustrated in the data analysis of Section 4 where we consider a discrete prior for α that includes the value of $\alpha = 2$ in its support.

To explore the impact of using a stable kernel in an IDE model, we present in Figure 1 a noise free propagation of a spatial surface from time 0 to time 10. Without random noise added, the IDE process averages nearby values leading to a smooth surface. The left column shows the IDE kernel shapes that differ only in parameter $\alpha = 1.25$, 1.75, and 2, the latter corresponding to a Gaussian kernel. Even though global features are similar, the stable kernels allow for more local variability in the propagated spatial surfaces. Richardson et al. (2017) provide similar results in one dimension.

In Richardson et al. (2017) it is shown that the degree of skewness of an IDE kernel can be connected to the physical characteristic of the dispersion for the process. To understand their effect on IDE space-time modelling, we study skewness and tail properties of stable distributions, including

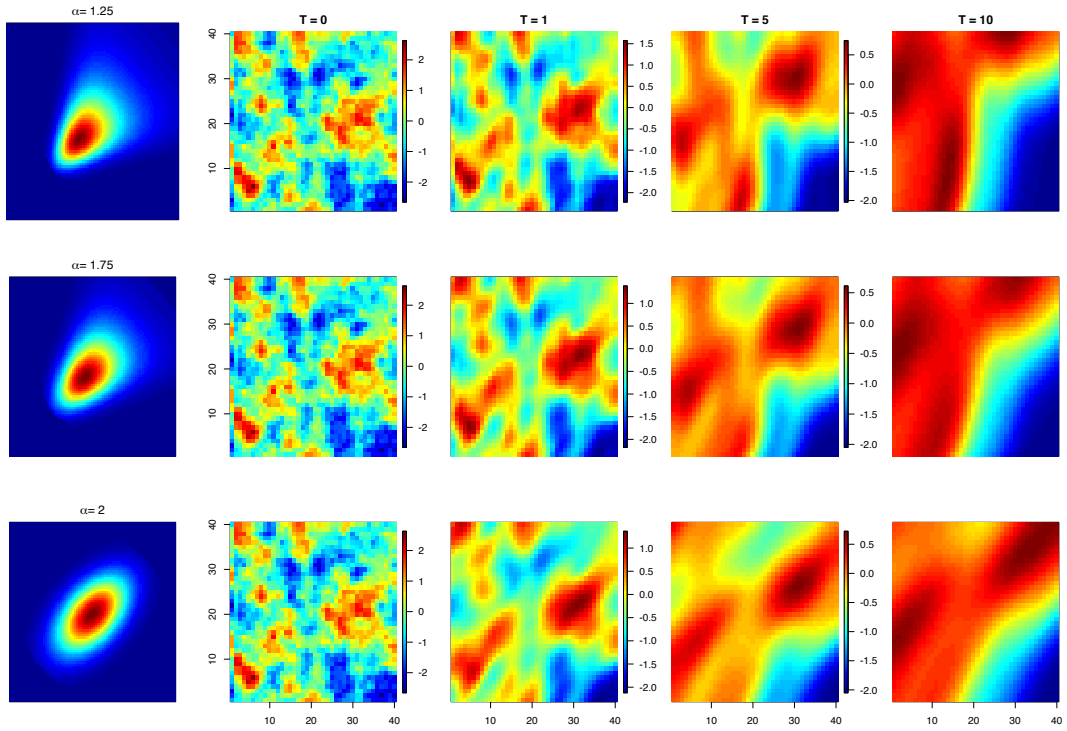


Figure 1: The propagation of a spatial surface with different IDE kernels is shown without noise added to the model. The kernel shape is shown in the left most column. We use the same initial iterate for all three kernels and the spatial surfaces for times 1, 5, and 10 are plotted in subsequent columns.

two sub-families, ECS and elliptically symmetric stable (ESS) distributions.

ECS distributions are a simplification of stable laws (Nolan, 2013) that have characteristic function:

$$\exp\{it'\mu - (t'\Sigma t)^{\alpha/2}\}, \tag{5}$$

where Σ is a positive definite matrix, μ a location parameter vector, and $\alpha \in (0, 2]$. Thus, the Gaussian family is a subset of ECS distributions corresponding to $\alpha = 2$. Equation 2.5.8 in Samorodnitsky & Taqqu (1997) provides a specific form for Γ that yields ECS laws.

A stable distribution is called symmetric α-stable if there exists μ such that $-X + \mu$ and $X + \mu$ have the same distribution. This property is equivalent to that of elliptical symmetry. The following lemma provides an alternative definition for elliptical symmetry.

Lemma 2. *The condition $\gamma(z) = \gamma(z + \pi)$, for $z \in [0, \pi]$, is necessary and sufficient for a bivariate stable distribution to be elliptically symmetric.*

Proof. Without loss of generality, assume $\mu = (0,0)'$. Using a Fourier representation of the density, $f(x) = (2\pi)^{-2} \sum_{t\in\mathbb{Z}^2} g(t) \exp(it'x)$, where $g(t)$ is given by (4) with $\mu = (0,0)'$. Hence, $f(x) - f(-x)$ can be written as $(2\pi)^{-2} \sum_{t\in\mathbb{Z}^2} \{g(t) - g(-t)\} \exp(it'x)$.

First, assume the distribution is elliptically symmetric, which implies $f(x) - f(-x) = 0$. Then, Fourier and inverse Fourier transform properties yield $g(t) - g(-t) = 0$, for all $t \in \mathbb{Z}^2$. By collecting terms and dividing constants, we obtain $\int_{z=0}^{2\pi} \text{sign}(t'v)|t'v|^{\alpha} \text{d}\Gamma(z) = 0$. The integral can be split and written as $\int_{z=0}^{\pi} |t'v|^{\alpha} \text{d}\Gamma(z) - \int_{z=0}^{\pi} |t'v|^{\alpha} \text{d}\Gamma(z+\pi) = 0$, for all $t \in \mathbb{Z}^2$, which holds true only when $\gamma(z) = \gamma(z+\pi)$, for $z \in [0,\pi]$. Now, assume that $\gamma(z) = \gamma(z+\pi)$, for $z \in [0,\pi]$. By working backwards in the above argument, this implies $\int_{z=0}^{2\pi} \text{sign}(t'v)|t'v|^{\alpha} \text{d}\Gamma(z) = 0$. The characteristic function then becomes $\exp\left\{\int_{z=0}^{2\pi} |t'v|^{\alpha} \text{d}\Gamma(z)\right\}$. Therefore, $g(t) - g(-t) = 0$, for $t \in \mathbb{Z}^2$, and thus $f(x) - f(-x) = 0$, resulting in the property of elliptical symmetry. □

Lemma 2 characterizes ESS distributions, and formalizes the previous statement that varying degrees of skewness in bivariate stable distributions can be achieved with more or less disparate values of $\gamma(z)$ and $\gamma(z+\pi)$. Note that ECS distributions form a strict subset of the ESS family of distributions; see, e.g., Example 2.3.5 in Samorodnitsky & Taqqu (1997).

As skewness is a key difference between stable and Gaussian kernel distributions, it is important to understand what is gained from it within the IDE modelling framework. Wikle (2002) indicates that any kernel with spatially varying parameters produces an asymmetric propagation of the spatio-temporal process over time, although the exact nature of this asymmetry has not been shown. This is clarified in the following discussion which also addresses the impact of stable IDE kernels.

The IDE kernel, $k(u - s \mid \theta)$, in Equation (1) can be interpreted as a weighting function that maps the process at location u and time t to the process at location s and time $t+1$. For instance, note that for location s' at time t, $k(s' - s \mid \theta)$ is (approximately) the derivative of the conditional expectation $\partial \text{E}(X_{t+1}(s) \mid X_t = x_t)/\partial x_t(s')$, where $x_t = \{x_t(u) : u \in \mathcal{D}\}$ is a realization of X_t over the spatial domain $\mathcal{D}$. Hence, a higher kernel value increases the effect of process values at time t to process values at time $t+1$.

For a Gaussian kernel distribution with mean vector μ and covariance matrix Σ that are not spatially varying, $k(s' - s \mid \mu, \Sigma) \propto |\Sigma|^{-1/2} \exp\{-0.5(s' - s - \mu)' \Sigma^{-1} (s' - s - \mu)\}$. The Gaussian kernel is restricted by two forms of elliptical symmetry in terms of the process evolution over time:

(A) For a given location s at time $t+1$, the kernel contribution of s' at time t is equivalent for all locations $s' = s + \mu + \omega$ such that $\omega' \Sigma^{-1} \omega$ has the same value.

(B) For a given location s' at time t, the kernel contribution to the process at location s and time $t+1$ is equivalent for all locations $s = s' - \mu + \omega$ where $\omega' \Sigma^{-1} \omega$ has the same value.

In other words: (A) every location at time $t+1$ is equally influenced by sets of elliptical contours at time t; (B) every location at time t equally influences sets of elliptical contours at time $t+1$. These restrictions follow from the elliptical symmetry of the Gaussian distribution probability contours.

When the Gaussian kernel parameters are spatially varying, as in $k(u - s \mid \mu(s), \Sigma(s))$, then symmetry restriction (B) is relaxed, since the kernel contribution to the process at time $t+1$ depends on s. However, restriction (A) still holds. Alternatively, the kernel can be written as $k(u-s \mid \mu(u), \Sigma(u))$. In this case, symmetry restriction (A) is no longer valid, but (B) is. The spatially varying Gaussian kernel is unable to remove both types of elliptical symmetry restrictions. In contrast, IDE models with spatially varying stable kernels can break both types of restrictions as the contours of the bivariate stable are not limited to ellipses. The ECS and ESS kernels have similar restrictions as the Gaussian kernel with respect to asymmetric propagation of the spatio-temporal process, but allow for more flexibility through their more general tail behavior.

We explore IDE models with ECS kernels and general stable kernels. A key feature of the modelling framework is the nested structure of IDE models that can be used to capture and assess the effect of kernel skewness by comparing the full bivariate stable with the ECS kernel, and the effect of tail width by comparing the ECS with the Gaussian kernel. While the ESS family is not modelled directly, Lemma 2 is used in Section 5 to assess the symmetry of the full stable distribution. The general stable kernel provides a comprehensive IDE model formulation, with the ESS kernel, ECS kernel, and Gaussian kernel IDE models getting respectively more restrictive in a direct nested structure.

2.2 Fourier series representation for bivariate stable distributions

Let $g(t)$, where $t = (t_1, t_2)'$, be the characteristic function of a bivariate density f, such that $f(x) = (2\pi)^{-2} \sum_{t \in \mathbb{Z}^2} g(t) \exp(it'x)$. When $g(t) = \exp\{a(t) + ib(t)\}$, where $a(t)$ and $b(t)$ are real-valued scalar functions, the real basis functions are $\cos(t'x)$ and $\sin(t'x)$ and the coefficients of the expansion are, respectively, $\exp\{a(t)\}\cos\{b(t)\}$ and $\exp\{a(t)\}\sin\{b(t)\}$. For identifiability, we can combine

Table 1: Basis functions and coefficients for the real-valued Fourier basis expansion of the bivariate stable distribution and the elliptically contoured stable (ECS) distribution, where $t \in \mathbb{Z}^2$.

Basis Function	General bivariate stable coefficients	ECS coefficients
$(2\pi)^{-1}\cos(t'x)$	$\pi^{-1}\exp\{\int_{z=0}^{2\pi}\lvert t'v\rvert^{\alpha}\mathrm{d}\Gamma(z)\}\cos\{t'\mu - \int_{z=0}^{2\pi}\mathrm{sign}(t'v)\lvert t'v\rvert^{\alpha}\tan(\pi\alpha/2)\mathrm{d}\Gamma(z)\}$	$\pi^{-1}\exp(t'\Sigma t)^{\alpha}\cos(t'\mu)$
$(2\pi)^{-1}\sin(t'x)$	$\pi^{-1}\exp\{\int_{z=0}^{2\pi}\lvert t'v\rvert^{\alpha}\mathrm{d}\Gamma(z)\}\sin\{t'\mu - \int_{z=0}^{2\pi}\mathrm{sign}(t'v)\lvert t'v\rvert^{\alpha}\tan(\pi\alpha/2)\mathrm{d}\Gamma(z)\}$	$\pi^{-1}\exp(t'\Sigma t)^{\alpha}\sin(t'\mu)$

certain basis functions such that only the indices where $t_1 \geq 1$ for all t_2, or for $t_2 \geq 1$ when $t_1 = 0$, are included in the basis function set. The resulting coefficients are $\exp\{a(t)\}\cos\{b(t)\} + \exp\{a(-t)\}\cos\{b(-t)\}$ and $\exp\{a(t)\}\sin\{b(t)\} - \exp\{a(-t)\}\sin\{b(-t)\}$. When $t_1 = t_2 = 0$, the basis coefficient is $\exp\{a(0)\}\cos\{b(0)\}$.

The characteristic function in Equation (4) can be written for the bivariate stable distribution as $\exp\{a(t) + ib(t)\}$, where $a(t) = \int_{z=0}^{2\pi}\lvert t'v\rvert^{\alpha}\mathrm{d}\Gamma(z)$, and $b(t) = t'\mu - \int_{z=0}^{2\pi}\mathrm{sign}(t'v)\lvert t'v\rvert^{\alpha}\tan(\pi\alpha/2)\mathrm{d}\Gamma(z)$, with $v = (\cos(z), \sin(z))'$. Table 1 includes the Fourier basis functions and coefficients for the bivariate stable distribution. These basis coefficients are defined for $t_1 \geq 1$ for all t_2, and for $t_2 \geq 1$ when $t_1 = 0$. The basis coefficient when $t_1 = t_2 = 0$ is $(2\pi)^{-1}$, and only the cosine basis function are included. To obtain the coefficients in the spatially varying case, parameters μ and Γ are replaced with $s + \mu_s$ and Γ_s. Table 1 includes also the basis coefficients for bivariate ECS distributions with characteristic function in Equation (5). In the spatially varying case, μ and Σ are extended to μ_s and Σ_s.

3 Semiparametric model for the stable kernel distribution

As discussed in Section 2, the stable kernel distribution enters the model in Equations (2) and (3) through its characteristic function, thus overcoming the lack of a closed-form expression for the density. Its definition in terms of parameter vector (α, μ) and measure Γ provides a natural setting for a Bayesian semiparametric model specification. In particular, we develop a model for $\gamma = \mathrm{d}\Gamma$ building on ideas from Bernstein polynomial priors for densities with compact support (Petrone, 1999). In light of the ensuing extension to spatially dependent IDE kernel parameters, a key consideration for the model structure is to balance model flexibility and computational feasibility for inference.

Since Γ is a finite measure on $[0, 2\pi]$, $\gamma = \mathrm{d}\Gamma$ is a probability density function on $[0, 2\pi]$ up to normalizing parameter $c > 0$. Denote by $\mathrm{be}(\cdot \mid a, b)$ the density of the Beta distribution with mean

$a/(a+b)$. We model γ using a weighted combination of Beta densities with different shapes,

$$c^{-1}\,\gamma(z) = (2\pi)^{-1} \sum_{m=1}^{M} w_m \,\mathrm{be}(z/2\pi \mid m, M-m+1), \quad z \in [0, 2\pi]. \tag{6}$$

The mixture weights are defined through a distribution function F on $[0, 2\pi]$, more specifically, $w_m = F(2\pi m/M) - F(2\pi(m-1)/M)$, for $m = 1, ..., M$.

To motivate this model structure, note that, as $M \to \infty$, the right-hand-side of Equation (6) converges uniformly to the density of F, assuming such density is continuous. Hence, in our context, F plays the role of the distribution function Γ/c, which corresponds to the density function γ/c.

Therefore, key to model flexibility is the capacity of F to admit general shapes, which allows the effective selection of the Beta mixture components that contribute most to the construction of γ. This motivates using a nonparametric prior for F that supports general discrete distributions on $[0, 2\pi]$. We assign a geometric weights prior (Mena et al., 2011) to F, which implies that the corresponding distribution is (almost surely) of the form $\sum_{j=1}^{\infty} q(1-q)^{j-1}\,\delta_{x_j}$. This is a discrete distribution with random atoms, x_j, and weights, $q(1-q)^{j-1}$, defined through a single random variable q. The x_j are independent and identically distributed from a base distribution on $[0, 2\pi]$, and q follows a distribution on $(0, 1)$, independently of the x_j. This model structure is more economical in the number of parameters than, for instance, the stick-breaking weights of the Dirichlet process, enabling the extension to a spatially varying kernel distribution.

We use a uniform base distribution for the x_j, and fix M to a sufficiently large value. In practice, we recommend prior sensitivity analysis for the value of M. Moreover, the countable representation for F is truncated at a finite number of elements, J, defining the last weight to be 1 minus the sum of the previous weights. Since $\sum_{j=J+1}^{\infty} q(1-q)^{j-1} = (1-q)^J$, we can specify J by, for instance, setting $(1 - \mathrm{E}(q))^J$ equal to some small number and solving for J.

Note that w_m arises by summing up the geometric weights $q(1-q)^{j-1}$ for the x_j that lie in interval $(2\pi(m-1)/M, 2\pi m/M)$. Since the x_j are uniformly distributed on $[0, 2\pi]$, we can associate a Multinomial$(1, (1/M, ..., 1/M))$ vector, $(Z_{j1}, ..., Z_{jM})$, with each x_j, and express the mixture weights as $w_m = \sum_{j=1}^{J} q(1-q)^{j-1} Z_{jm}$, for $m = 1, ..., M$. This replaces a continuous latent variable with a discrete variable, which facilitates estimation. By construction, only one element of $(Z_{j1}, ..., Z_{jM})$ is 1, the rest being 0, and thus the dimensionality of the new parameter set remains effectively the same.

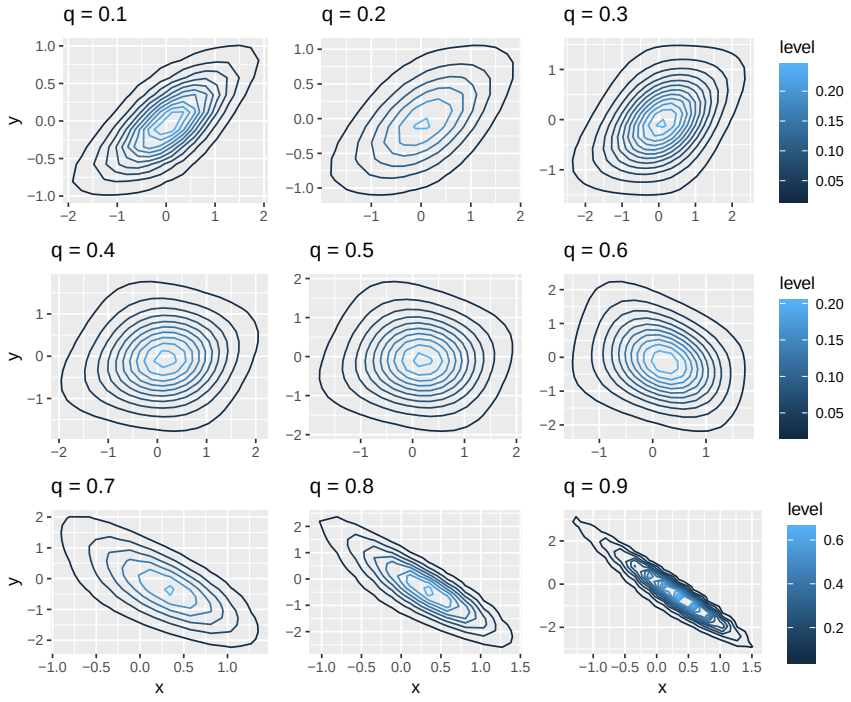


Figure 2: Bivariate stable densities under the model of Section 3 for different values of the geometric weight, q, with $\mu = (0,0)'$, $c = 2\pi$, $\alpha = 1.5$, $M = 40$ and $J = 40$ for each plot. The first 10 atoms of the geometric weights prior are $(2, 5.14, 3.6, 0.4, 3.4, 0.6, 3.5, 0.5, 3.5, 0.5)$. The other atoms were randomly drawn from the uniform distribution on $[0, 2\pi]$, but are the same for all 9 plots.

Therefore, measure Γ is defined in terms of parameters c, q, and $\{Z_{jm} : j = 1, ..., J; m = 1, ..., M\}$, and the stable IDE kernel specification is completed with location parameter $\mu = (\mu_1, \mu_2)$ and tail parameter α. To construct a non-stationary spatio-temporal process, we elaborate on the model such that parameters μ, c, and q are spatially varying. To illustrate the flexibility of the model and the effect of spatially varying q, Figure 2 shows how the bivariate stable density changes with q. Even though the atoms of the geometric weights prior are the same, the density shape can drastically change in both skewness and orientation by only varying q. Also, Figure 2 shows that the kernel changes smoothly with q, implying that kernels at nearby locations and having similar values for q, will be similar in shape. This smooth transition of kernel shape is desirable for IDE spatio-temporal models.

The spatially dependent parameters are modelled via kernel process convolutions (Higdon, 1998) for $\mu_1(s)$, $\mu_2(s)$, $\log(c(s))$, and $\Phi^{-1}(q(s))$, where Φ is the standard normal distribution function. For instance, for $\mu_1(s)$, consider a grid of knots $u_1, ..., u_Q$ and latent variables $(\zeta_{\mu_1}(u_1), ..., \zeta_{\mu_1}(u_Q))$ which are independent $\mathrm{N}(0, \sigma_{\mu_1})$. Then, $\mu_1(s) = \mu_{\mu_1} + \sum_{i=1}^{Q} k_\zeta(u_i, s)\zeta_{\mu_1}(u_i)$, where $k_\zeta(u_i, s)$ is the

convolution kernel, e.g., the Gaussian or Matern function. The process $\mu_2(s)$ has a similar construction. Analogously, $\log(c(s)) = \mu_c + \sum_{i=1}^{Q} k_\zeta(u_i, s)\zeta_c(u_i)$, where $\zeta_c(u_i) \sim \mathrm{N}(0, \sigma_c)$, and $\Phi^{-1}(q(s)) = \mu_q + \sum_{i=1}^{Q} k_\zeta(u_i, s)\zeta_q(u_i)$, where $\zeta_q(u_i) \sim \mathrm{N}(0, \sigma_q)$. Using process convolutions reduces the parameter space as Q is typically set to be much smaller than the number of locations, but they may potentially ignore some short range variability. The parameter processes, however, should vary smoothly in space, so a process convolution approximation is arguably an appropriate modelling strategy.

The parameter α is kept constant over the spatial region, which aids in estimation and also allows the spatially varying Gaussian kernel to be retained as a special case.

In Section 5, model flexibility is explored through the analysis of sea surface temperature data, including comparison with IDE models based on spatially varying ECS and Gaussian kernels. The Gaussian kernel IDE model is fit with all five parameters varying in space: two means, two variances, and the covariance. The bivariate stable model achieves a wider array of kernel shapes with four spatially varying parameters, thus offering a more general inference framework without sacrificing computational feasibility relative to the state-of-the-art Gaussian kernel model.

4 Posterior simulation

Here, we discuss posterior simulation for the model given in Equations (2) and (3), under the prior model of Section 3 for the stable kernel distribution. Although we focus on the general stable kernel IDE model, the approach is similar for the simpler ECS kernel model. Combining all proposed model components we have:

$$
\begin{aligned}
Y_t \mid a_t, \sigma^2 &\sim \mathrm{N}(\Psi_t a_t, \sigma^2 I_{n_t}), \qquad t = 1, ..., T \qquad (7)\\
a_t \mid a_{t-1}, \tau^2, \varphi &\sim \mathrm{N}(G_t B_\varphi a_{t-1}, \tau^2 G_t V_t G_t'), \qquad t = 1, ..., T\\
\gamma_s(z) &= c(s)\,(2\pi)^{-1} \sum_{m=1}^{M} w_m(s)\,\mathrm{be}(z/2\pi \mid m, M - m + 1)\\
w_m(s) &= \sum_{j=1}^{J} q(s)(1 - q(s))^{j-1} Z_{jm}, \quad (Z_{j1}, ..., Z_{jM}) \sim \mathrm{Multinomial}(1, (1/M, ..., 1/M))\\
\boldsymbol{\mu}_j &= \mu_{\mu_j} 1 + K_\zeta \zeta_{\mu_j}, \quad \zeta_{\mu_j}(u_i) \mid \sigma_{\mu_j} \sim \mathrm{N}(0, \sigma_{\mu_j}), \quad j = 1, 2\\
\log(\boldsymbol{c}) &= \mu_c 1 + K_\zeta \zeta_c, \quad \zeta_c(u_i) \mid \sigma_c \sim \mathrm{N}(0, \sigma_c)\\
\Phi^{-1}(\boldsymbol{q}) &= \mu_q 1 + K_\zeta \zeta_q, \quad \zeta_q(u_i) \mid \sigma_q \sim \mathrm{N}(0, \sigma_q)
\end{aligned}
$$

where $(\boldsymbol{c}, \boldsymbol{q}, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$, collect the spatially dependent kernel parameter values at all spatial locations.

The matrix V is an unscaled spatial covariance matrix. The matrix K_ζ maps latent variables $\zeta_r = (\zeta_r(u_1), ..., \zeta_r(u_Q))'$ vectors for $r \in \{\mu_1, \mu_2\, c, q\}$ to the processes governing the spatially dependent IDE kernel parameters. The values of K_ζ correspond to convolution kernel where the (i, j)-th element is $k_\zeta(s_i - u_j)$.

The parameters are estimated using Markov chain Monte Carlo. The state vectors $\{a_0, ..., a_T\}$ are sampled via forward filtering backward sampling (Frühwirth-Schnatter, 1994). Conditional on the state vectors, the kernel parameters are updated using Metropolis-Hastings within Gibbs sampling. Regarding the hyperparameters of the kernel process convolution for $\mu_1(s)$, $\mu_2(s)$, $\log(c(s))$, and $\Phi^{-1}(q(s))$, conditionally conjugate updates are available for scale parameters σ_{μ_1}, σ_{μ_2}, σ_c, σ_q under inverse gamma priors, as well as for location parameters μ_{μ_1}, μ_{μ_2}, μ_c, μ_q under normal priors.

To facilitate estimation of the tail parameter α, we recommend a discrete uniform prior on a grid between 1.05 and 2. This results in ready updates for α from its discretized posterior conditional distribution. Another advantage of discretizing α is that the integrals in the basis coefficients from Table 1 can be calculated prior to MCMC for each Beta density in (6) and for each possible value for α, resulting in a significant speed-up. For each set $(Z_{j1}, ..., Z_{jM})$, $j = 1, ..., J$, the single variable which is equal to 1 can be sampled from a discrete distribution. Letting l_j be an indicator variable, such that $l_j = m$ when $Z_{jm} = 1$, the conditional posterior probability for $l_j = m$ is proportional to the corresponding prior times the joint distribution from (3) evaluated at $l_j = m$. From a computational perspective, the discreteness of these variables lends itself to parallelization, as does sampling the discretized variable α. This is done by sending the calculations of the components of the discrete probabilities to different nodes and then collecting the proportional posteriors to calculate the probabilities. For a large data set, this can significantly increase the speed of MCMC computations. For this work, the parallelization was done in C++ using openMP.

Decomposing the process and the IDE kernel using a basis expansion requires truncation to a finite number of basis functions (Xu et al., 2005). In one dimension, the number of basis functions required for accurately representing the kernel is relatively small, ranging between 20 and 100. To achieve the same level of accuracy in two dimensions, hundreds of basis functions may be required. In the IDE literature, Wikle & Holan (2011) address this issue with Stochastic Search Variable Selection to reduce the number of basis functions. We utilize a method that also seeks to find basis functions that

can be removed from the analysis while maintaining smoothness of the Fourier approximation.

The main determining factor of the required number of basis functions is the width of the kernel compared to the range of the data. For example, when using a Gaussian kernel, a higher variance requires fewer basis functions. This implies that the required number of basis functions changes throughout the MCMC. Recall that $B_{\varphi,t}$ is a matrix where the (i,j)-th element is the coefficient of the j-th basis function at location s_i. By investigating the basis decomposition, we note that a small percentage of the largest basis coefficients are significantly larger than 0. Also, the size of the coefficient is not fully determined by the order of the frequencies. As basis function index increases, the frequency of the function increases. The general trend is that the coefficients get smaller, but several coefficients of smaller frequency basis functions appear to be less significant. By exploiting these facts the dimension of the state space can be intelligently decreased.

The MCMC can be adjusted at each iteration to include only the most important basis functions and decrease computational time of the algorithm. After calculating $B_{\varphi,t}$, the column maximum of $B_{\varphi,t}$ can be used to decide the number of basis functions. We threshold the columns using the generalized double Pareto distribution (Armagan et al., 2013). This approach maintains the continuity of the target function without over-shrinking. The method sets b_j equal to 0, when $|b_j| < \epsilon\sqrt{(\eta+1)}$. When $|b_j| > \epsilon\sqrt{(\eta+1)}$, the adjusted values b_j^* are

$$b_j^* = \begin{cases} 0.5\left\{b_j - \epsilon\sqrt{(\eta+1)} + [b_j^2 + 2b_j\epsilon\sqrt{(\eta+1)} - 3\epsilon^2(\eta+1)]^{1/2}\right\} & \text{if } b_j > 0 \\ 0.5\left\{b_j + \epsilon\sqrt{(\eta+1)} - [b_j^2 + 2b_j\epsilon\sqrt{(\eta+1)} - 3\epsilon^2(\eta+1)]^{1/2}\right\} & \text{if } b_j < 0 \end{cases} \tag{8}$$

where the value for ϵ defines the level of truncation and η controls the shrinkage; η was set to 1 for the data analyses of Section 5. To be able to predict the length of the learning algorithm, it may help to keep the number of basis functions constant. The only way to accomplish this would be to change the thresholding level, ϵ. This approach also ensures that a reasonable number of basis functions will always be present. We select ϵ that yields a fixed number of basis functions for each MCMC iteration. If the column maximum of $B_{\varphi,t}$ is less than $\epsilon\sqrt{(\eta+1)}$, then the column is not used and the dimensionality is reduced, otherwise, the values are adjusted as in Equation (8).

5 Sea surface temperature data analysis

Sea surface temperature (SST) in the tropical Pacific Ocean has been useful in predicting a number of climate phenomena (Philander, 1985). The most prominent of these phenomena is El Niño, which is a warming of ocean temperature occurring between $-5°$ and $5°$ Latitude and $180°$ and $240°$ E Longitude (see http://www.elnino.noaa.gov). This warming results in a shift of nutrients in the water which can affect agriculture and fisheries in several countries. It can also have profound climate effects that include increased precipitation in some areas of the Americas, as well as very dry conditions in some others (Ropelewski & Halpert, 1987). Figure 3 shows SST anomalies, i.e., observed differences from long-term averages, in the Pacific ocean in July 2015. The large red portion in the middle suggests an El Niño event that was known to have occurred during that year.

Data for sea surface temperature has been collected for a number of decades. Prior to the 1980s it was collected monthly using instruments attached to buoys or ships (Anding & Kauth, 1970). Since then it has mainly been measured using satellite observations and is collected more frequently (Dozier (1981), see also https://podaac.jpl.nasa.gov/SeaSurfaceTemperature). Using monthly SST data collected across the entire spatial field shown in Figure 3, a spatio-temporal model can be built to describe and predict its behavior.

There is a rich history of work dedicated to prediction of El Niño, as well as of its counterpart La Niña, which is the cooling that usually follows. El Niño follows a 2 to 7 year cycle and typically begins in Autumn, staying as long as a year. While many deterministic physical models have been developed to explain and predict the occurrences (Cane et al., 1986; Jan van Oldenborgh et al., 2005), much success has come from simply using SST data over a large region in the Pacific Ocean in a stochastic model (Barnston et al., 1999; Berliner et al., 2000). These methods include linear systems (Penland & Magorian, 1993), as well as non-linear methods (Wikle & Hooten, 2010; Cressie & Wikle, 2011).

We apply the IDE model with the spatially dependent bivariate stable kernel to SST data. The specific data set includes 9692 locations on a $1°$ by $1°$ resolution grid. It is collected monthly from December 1980 to July 2015. Seasonality is accounted for by modelling the anomalies, which are differences between the data and the long-term monthly average for each location. The purpose of this analysis is to illustrate the ability to apply the bivariate stable kernel IDE model to a substantive real data example, and to compare predictive accuracy between the two types of stable kernels (general

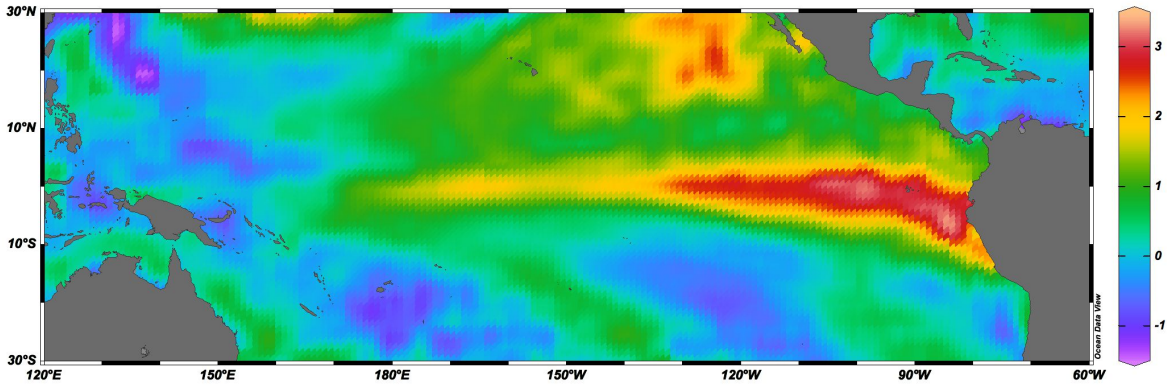


Figure 3: Sea surface temperature anomalies as measured in July 2015.

stable and ECS kernel) and the Gaussian kernel. We do not claim a novel modelling strategy for SST data analysis or for predicting El Niño, and we therefore forgo attempts at scientific conclusions.

5.1 Analysis under the general stable kernel IDE model

We apply the stable kernel IDE model to the SST monthly anomalies. The parameter space comprises the observational variance σ^2, the process variance τ^2, and the stable distribution kernel parameters, which include the latent variables involved in the kernel convolution for processes $\mu_1(s)$, $\mu_2(s)$, $\log(c(s))$, and $\Phi^{-1}(q(s))$, along with the associated hyperparameters (μ_r, σ_r) for $r \in \{\mu_1, \mu_2\, c, q\}$. The full model is given in Equation (7) in Section 4. Because the data lies on a regular grid, many components of do not need to be indexed by time, in particular, $G_t = G$ and $\Psi_t = \Psi$. The model is completed with inverse gamma priors for σ^2, τ^2, and σ_{μ_1}, σ_{μ_2}, σ_c, σ_q, normal priors for μ_{μ_1}, μ_{μ_2}, μ_c, μ_q, and a discrete uniform prior for α on a grid between 1.05 and 2 with step size 0.05.

The model was applied with $M = 40$; inference results were unaffected for M larger than 30. We estimated $\mathrm{E}(q)$ across the spatial domain based on the kernel process convolution for $\Phi^{-1}(q(s))$, and used the approach of Section 3 to set $J = 40$ for the truncation level of the geometric weights prior.

The matrix V is the unscaled spatial covariance matrix of a Matern covariance function with $\kappa = 1.5$ and an effective range of 2, based on an exploratory analysis of the covariance structure. The priors for σ^2 and τ^2 are $\mathrm{IG}(5, 3)$, which are informed by the expected range of the data. The matrix K_ζ is built using a Matern function for the convolution kernel with $\kappa = 2.5$ and an effective range of 4 for all the parameters, forcing a smooth evolution of the parameter processes across the domain. The

knots are chosen on a 20 by 20 grid, resulting in a dimension reduction of the variables representing the spatially varying parameter processes from 9692 to 400. The priors for the location parameters are $\mu_{\mu_j} \sim \mathrm{N}(0,1)$, $j = 1,2$, $\mu_c \sim \mathrm{N}(0,4)$, and $\mu_q \sim \mathrm{N}(-1, 0.5)$. The scale parameters are given priors of $\mathrm{IG}(4,3)$ for σ_c and σ_{μ_j}, $j = 1,2$, and $\mathrm{IG}(10,6)$ for σ_q. These priors are dispersed with means driven by the scale of the data and reasonable kernel shapes. Inference for process $q(s)$ is sensitive to the prior choice, especially for very diffuse priors. There are several combinations of $q(s)$ and the latent Z_{jm} variables which result in the same values for w_m, which are identifiable. The other parameters are not overly sensitive to the prior.

Using the MCMC algorithm discussed in Section 4, 15,000 samples are drawn from the posterior distribution, with the first 5,000 discarded as burn-in, leaving 10,000 samples. The kernel parameters are sampled 200 times between iterations of sampling the state space to speed up convergence and improve the MCMC mixing. Convergence of the MCMC algorithm was checked using trace plots of kernel density values at various locations throughout the data domain, suggesting convergence before 5,000 iterations. The good mixing and relatively quick convergence observed for the SST data may be due to the large amounts of data. Additional considerations may be necessary when dealing with smaller data sets, such as more informative priors or longer MCMC chains.

Posterior inference for the IDE kernel density and the associated measure of the bivariate stable distribution can reveal information about the nature of dependence between locations. The posterior mean estimate for the IDE kernel density at three spatial locations is shown in Figure 4 along with estimates for the associated scaled density for measure Γ_s. Differences in the measure across the three locations are much more easily detected than differences in the stable kernel densities.

To detect trends, we devise a method of measuring symmetry across the spatial field. Based on Lemma 2, we use $sym_{ell}(s) = \int_0^{\pi} \{\gamma_s(z) - \gamma_s(z+\pi)\}^2 \, \mathrm{d}z$ as a metric of elliptical symmetry. Similarly, a metric of spherical symmetry can be defined as $sym_{sph}(s) = \int_0^{2\pi} \{\gamma_s(z)/c(s) - (2\pi)^{-1}\}^2 \, \mathrm{d}z$. Estimates for these symmetry metrics are plotted as a function of spatial location in Figure 5. The two symmetry maps are generally similar. The kernels at locations south of the equator are symmetric whereas the kernels for locations north of the equator are non-symmetric. This illustrates the utility of a skewed, spatially varying IDE kernel to enable flexible asymmetric propagation of the spatio-temporal process. The difference in the symmetry of the stable kernel on different sides of the equator suggests that the relationship between the spatial and temporal components of the process behaves differently on the

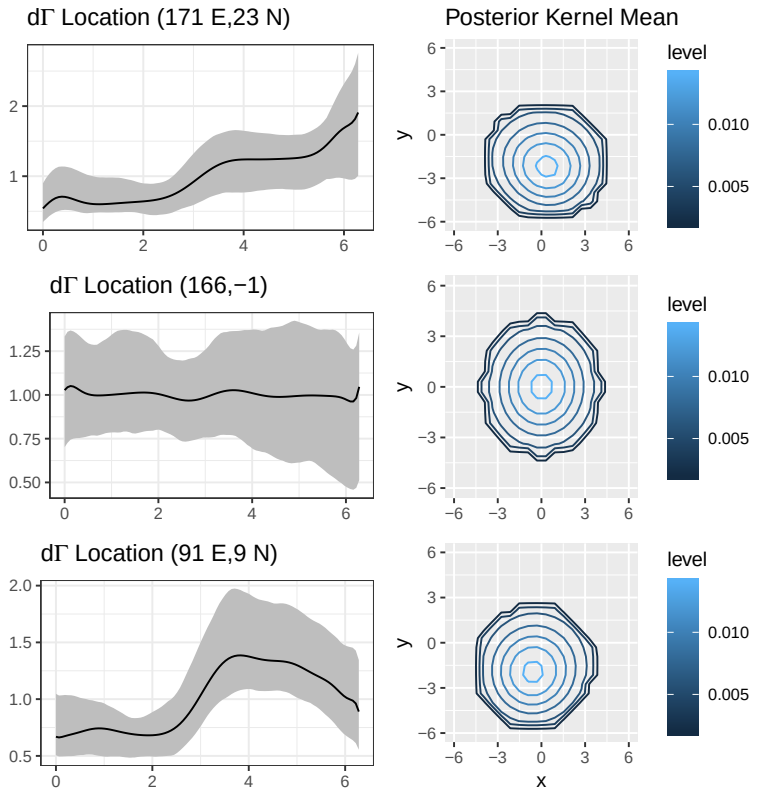


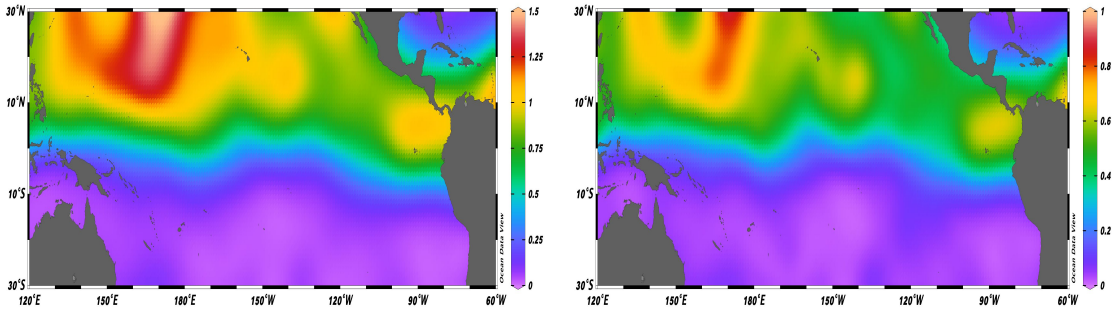
Figure 4: Posterior mean estimates and 95% credible bands for the scaled density γ_s corresponding to the stable kernel distribution measure (left panels) and posterior mean estimates for the IDE kernel density (right panels) for three spatial locations.

two sides of the equator. The physical interpretation of this is not clear.

Recall that the Gaussian distribution is a special case of the stable distribution when $\alpha = 2$. Because the parameter α was discretized on an even grid between 1.05 and 2 with step size of 0.05, we can find the posterior probability of Gaussianity. The prior probability $\Pr(\alpha = 2) = 0.05$ is revised to $\Pr(\alpha = 2 \mid \text{data}) = 0.006$, suggesting that it is unlikely that a Gaussian IDE kernel is appropriate. The posterior mode of α was 1.70 with a 95% posterior credible interval of $[1.55, 1.85]$, suggesting thicker tails for the IDE kernel than those of the Gaussian distribution.

5.2 Predictive inference and model comparison

In addition to the general stable kernel IDE model, we consider three more modelling methods: IDE models with spatially varying ECS and Gaussian kernels, and a dynamic spatio-temporal model



(a) Elliptical Symmetry

(b) Spherical Symmetry

Figure 5: Estimates for the elliptical and spherical symmetry metrics across the spatial domain (see Section 5.1 for details). For both metrics, smaller values are associated with more symmetric kernels.

(DSTM) based on an unstructured (and unknown) evolution matrix (Wikle & Hooten, 2010).

The Fourier coefficients for the ECS model are given in Table 1. The ECS kernel shapes are controlled by five spatially varying processes: $\mu_1(s)$ and $\mu_2(s)$ for the components of the location parameter vector $\mu(s)$, and $\sigma_1^2(s)$, $\sigma_2^2(s)$ and $\sigma_{12}(s)$ for the terms of the scale matrix $\Sigma(s)$. To model these spatial processes we use process convolutions, all on the same grid. The prior structure for the location parameters is the same with the general stable model. The priors for the components of the scale matrix are built from Gaussian process convolutions after a logarithmic transformation. We impose the additional prior restriction of positive definiteness for $\Sigma(s)$, which does not affect the efficiency of the MCMC computing for this particular analysis. The prior for the ECS tail parameter α is discrete uniform supported on an even grid from 1.05 to 2 with grid step 0.05. (As expected from the results of the general stable IDE model, the prior probability of $\alpha = 2$ was again decreased given the data.) The special case of the ECS model fixing $\alpha = 2$ yields a Gaussian kernel IDE model with spatially varying parameters (Wikle, 2002; Xu et al., 2005). Details of the DSTM, including model fitting, can be found in Chapter 8 of Cressie & Wikle (2011). For the SST data, we used 45 basis functions, as they hold over 90% of the spatial information. Although we do not detail the prior specification for the DSTM, we note that inference was insensitive to the specific prior choices.

We assess model performance by scoring the predictions for each time point using energy scores of the form $m^{-1}\sum_{i=1}^{m}||y^{(i)}-y|| - 0.5m^{-2}\sum_{i=1}^{m}\sum_{j=1}^{m}||y^{(i)}-y^{(j)}||$, where $y^{(1)},...,y^{(m)}$ are samples from the posterior predictive distribution, and y denotes the data vector (Gneiting et al., 2008).

We also compare the models using K-step ahead forecasts for all spatial locations and time

Model	$K = 0$	$K = 1$	$K = 2$	$K = 3$
DSTM	0.53	0.96	1.29	1.72
Gaussian	0.56	0.79	1.00	1.68
ECS	0.54	0.71	0.96	1.68
Stable	0.54	0.70	0.96	1.65

Table 2: K-step ahead predicted energy scores, where $K = 0, 1, 2, 3$. Results are reported for the four models considered for the SST data: a dynamic spatio-temporal model (DSTM), and three IDE models with Gaussian, elliptically contoured stable (ECS), and general stable kernels, all three kernel choices involving spatially varying parameters.

points. This is done by propagating the state variables through the process level of the model, $a_t^* \sim \mathrm{N}(GB_\varphi a_{t-1}, \tau^2 GVG')$. Such propagation can be applied for several steps past the final time point, T, followed by drawing the K-step ahead prediction $Y_{T+K}^* \sim \mathrm{N}(\Psi a_{T+K}^*, \sigma^2 I)$. This method is a standard dynamic model forecasting tool (West & Harrison, 2006) and is used frequently in spatio-temporal dynamic models (e.g., Sigrist et al., 2012). SST data is known to be non-linear, and so prediction may be adjusted by fitting a lagged model to the data, as in Berliner et al. (2000). The scoring comparison for the K-step ahead predictions, including in-sample predictions for $K = 0$, are shown in Table 2 for the four models considered for the SST data. Focusing on the three IDE models, the stable distribution consistently outperforms the Gaussian in each of the forecasts, while the ECS is typically in between in performance. As the ECS kernel IDE model includes the Gaussian as a direct special case, and the general stable kernel IDE model holds the ECS as a special case, these results are perhaps to be expected for in-sample predictions. The fact that such results extend to K-step ahead forecasts suggests that there is practical utility from the added complexity of the stable IDE models. Finally, we note that in-sample prediction is (slightly) better using the DSTM relative to the IDE models. However, out of sample prediction is noticeably worse for the DSTM, with the best scores in all three cases arising under the general stable IDE model.

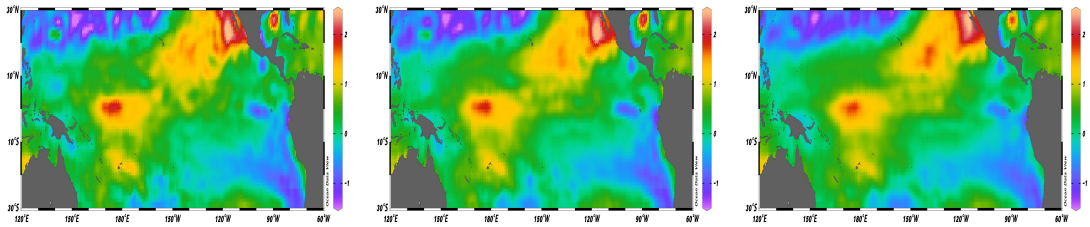
To perform true out of sample predictions, the last four months were left out of the analysis, resulting in 399 time points ending in February 2015 being included in the model fit. The posterior distributions for $Y_{400}^*, ..., Y_{403}^*$ are drawn as part of the MCMC algorithm. The means of these posterior distributions for the general stable and Gaussian kernel IDE models are shown in Figure 6 compared

against the truth for February through June 2015. The data depicts the El Niño which is known to have occurred in 2015. Predictions under both models capture the El Niño warming to some degree, but the intensity of the predicted warming using the stable kernel IDE model is much closer to the truth than under the model with the Gaussian kernel, especially closer to the coast of South America.

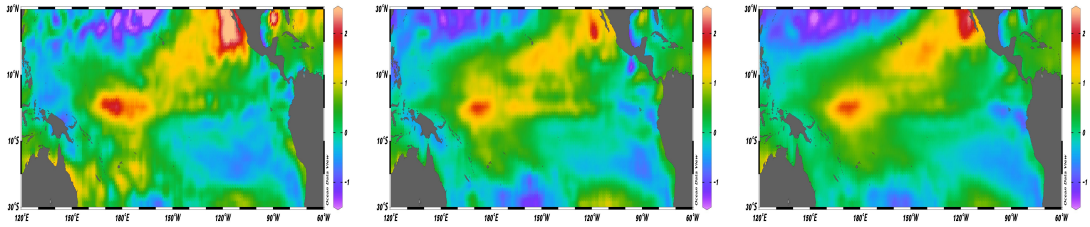
There are several ways to numerically declare an El Niño event. Most of these involve high anomalies in certain regions in the Pacific. For example, the official National Oceanic and Atmospheric Administration (NOAA) criterion involves the block average anomalies in El Niño region 3.4 to be above 0.05°C for 3 consecutive months, although this is modified to 5 consecutive months for the NOAA's Climate Prediction Center (Larkin & Harrison, 2005). The El Niño region 3.4 includes 120° to 170° W Longitude and $-5°$ to 5° Latitude. Based on the stable kernel IDE model, Figure 7 shows the posterior predictive probabilities that an El Niño will occur within the next 9 months for all months dating back to the beginning of the data set. These estimates are sampled from the K-step ahead posterior predictive distributions given all information through February 2015. The red X marks indicate when an El Niño did occur. The green bars extend from 9 months prior to an El Niño occurring to 3 months after. Therefore, blue dots that occur within the green bars are indications of correct El Niño predictions. The model seems to be able to predict an El Niño months in advance in most cases. Figure 7 also provides similar results for La Niña events. In this case, the predictions are more conservative, though using probabilities greater than 0.4 as a cutoff, they are still viable.

6 Summary

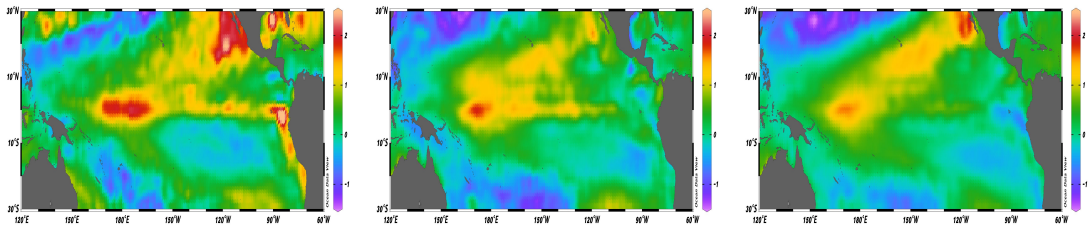
We have developed a modelling framework that explicitly considers space-time dynamics using a location-dependent linear evolution equation that is specified by a flexible family of infinitely-divisible kernels. In particular, we have explored integro-difference equation (IDE) space-time models with bivariate stable kernel distributions. The utility of the more flexible shape of such IDE kernels relative to the state-of-the-art Gaussian choice has been studied through model properties and demonstrated with the analysis of sea surface temperature data. A key feature of the model is that it includes the Gaussian kernel IDE model as a special case, thus enabling checking its validity in specific applications. The model can accommodate spatio-temporal non-stationarity, it incorporates observational error, and it can handle irregularly located observations and missing values.



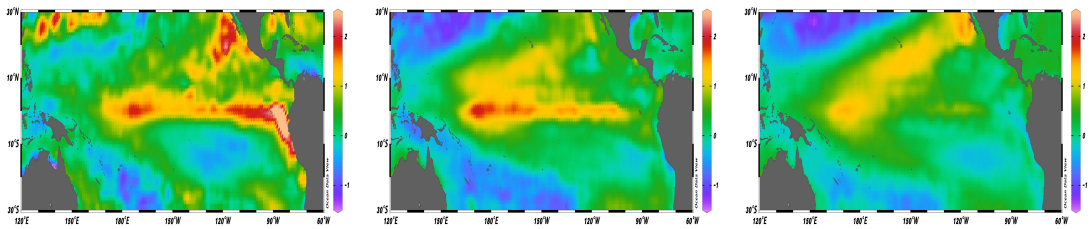
(a) February 2015 Truth (b) February 2015 Stable fit (c) February 2015 Gaussian fit



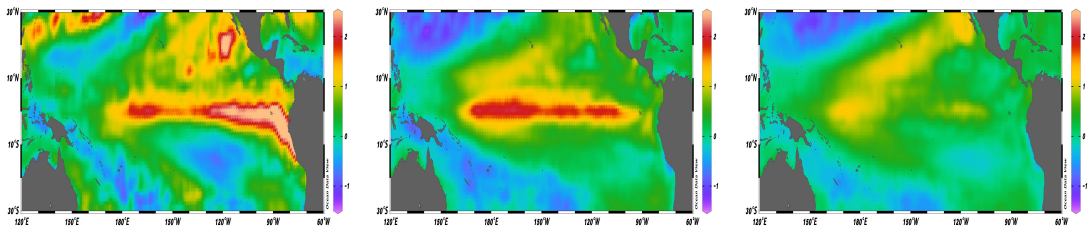
(d) March 2015 Truth (e) March 2015 Stable prediction (f) March 2015 Gaussian prediction



(g) April 2015 Truth (h) April 2015 Stable prediction (i) April 2015 Gaussian prediction

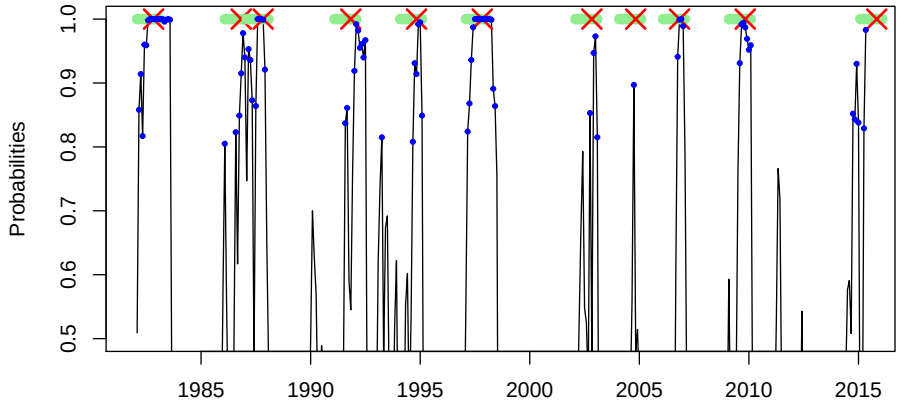


(j) May 2015 Truth (k) May 2015 Stable prediction (l) May 2015 Gaussian prediction

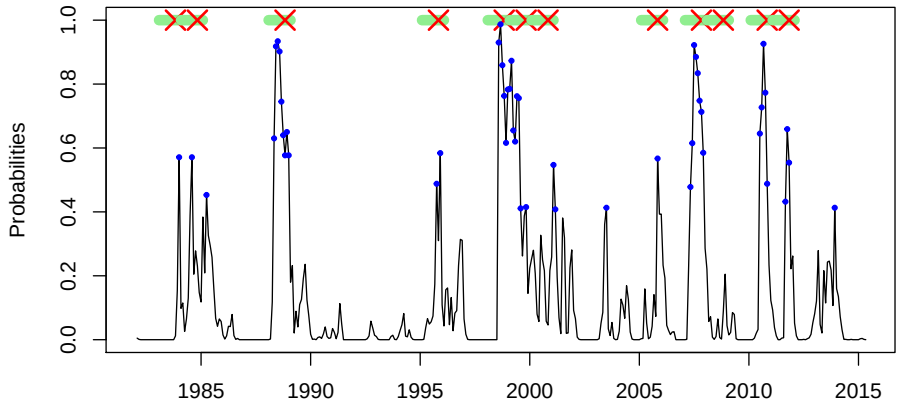


(m) June 2015 Truth (n) June 2015 Stable prediction (o) June 2015 Gaussian prediction

Figure 6: SST data and posterior K-step ahead predictions using the stable and Gaussian kernel IDE models. Data through February 2015 are included in the model fits, and the forecasts are shown for months March 2015 through June 2015.



(a) El Niño Predictions



(b) La Niña Predictions

Figure 7: Posterior probabilities that an El Niño or La Niña event will occur within the coming 9 months, under the general stable kernel IDE model. The red X marks indicate El Niño or La Niña occurrences and the green bars extend from 9 months before the events to 3 months after. Additionally, points with probabilities above 0.8 (0.4) in the El Niño (La Niña) plot are dotted blue.

To infer the characteristics of the IDE kernel and the states of the spatio-temporal process, we use Fourier basis representations for both the process and the kernel. A nonparametric prior is used for the measure that controls the shape of the bivariate stable IDE kernel. This, coupled with thresholding of the basis coefficients, produces a substantial dimension reduction. As a result, the model, even being more flexible, is more parsimonious than a spatially varying Gaussian kernel IDE model. The model's predictive performance is superior to that of the more standard model, based on the results from the non-trivial data example considered in the paper. Also, the use of a Bayesian hierarchical model

formulation allows for full probabilistic assessment of the prediction uncertainty. Finally, the structure of the proposed model allows for efficient parallelization of the posterior simulation algorithm.

Acknowledgements

The authors wish to thank Editor Simon Wood, an Associate Editor, and two reviewers for several comments that improved the presentation of the material in the paper.

References

Anding, D. & Kauth, R. (1970). Estimation of sea surface temperature from space. *Remote Sensing of Environment* **1**, 217–220.

Armagan, A., Dunson, D. B. & Lee, J. (2013). Generalized double Pareto shrinkage. *Statistica Sinica* **23**, 119–143.

Barnston, A. G., Glantz, M. H. & He, Y. (1999). Predictive skill of statistical and dynamical climate models in SST forecasts during the 1997–98 El Niño episode and the 1998 La Niña onset. *Bulletin of the American Meteorological Society* **80**, 217–244.

Berliner, L. M., Wikle, C. K. & Cressie, N. (2000). Long-lead prediction of Pacific SSTs via Bayesian dynamic modeling. *Journal of Climate* **13**, 3953–3968.

Brown, P. E., Roberts, G. O., Kåresen, K. F. & Tonellato, S. (2000). Blur-generated non-separable space–time models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62**, 847–860.

Cane, M. A., Zebiak, S. E. & Dolan, S. C. (1986). Experimental forecasts of El Niño. *Nature* **321**, 827.

Cressie, N. & Wikle, C. K. (2011). *Statistics for spatio-temporal data*. New York: John Wiley & Sons.

Dozier, J. (1981). A method for satellite identification of surface temperature fields of subpixel resolution. *Remote Sensing of Environment* **11**, 221–229.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis* **15**, 183–202.

Gneiting, T., Stanberry, L. I., Grimit, E. P., Held, L. & Johnson, N. A. (2008). Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds. *Test* **17**, 211–235.

Higdon, D. (1998). A process-convolution approach to modelling temperatures in the North Atlantic Ocean. *Environmental and Ecological Statistics* **5**, 173–190.

Jan van Oldenborgh, G., Balmaseda, M. A., Ferranti, L., Stockdale, T. N. & Anderson, D. L. (2005). Did the ECMWF seasonal forecast model outperform statistical ENSO forecast models over the last 15 years? *Journal of Climate* **18**, 3240–3249.

Larkin, N. K. & Harrison, D. (2005). On the definition of El Niño and associated seasonal average us weather anomalies. *Geophysical Research Letters* **32**, doi: 10.1029/2005GL022738.

Lemos, R. T. & Sansó, B. (2009). A spatio-temporal model for mean, anomaly, and trend fields of North Atlantic sea surface temperature. *Journal of the American Statistical Association* **104**, 5–18.

Mena, R. H., Ruggiero, M. & Walker, S. G. (2011). Geometric stick-breaking processes for continuous-time Bayesian nonparametric modeling. *Journal of Statistical Planning and Inference* **141**, 3217–3230.

Nolan, J. P. (2013). Multivariate elliptically contoured stable distributions: theory and estimation. *Computational Statistics* **28**, 2067–2089.

Nolan, J. P. (2014). Financial modeling with heavy-tailed stable distributions. *Wiley Interdisciplinary Reviews: Computational Statistics* **6**, 45–55.

Panorska, A. K. (1996). Generalized stable models for financial asset returns. *Journal of Computational and Applied Mathematics* **70**, 111–114.

Penland, C. & Magorian, T. (1993). Prediction of Niño 3 sea surface temperatures using linear inverse modeling. *Journal of Climate* **6**, 1067–1076.

Petrone, S. (1999). Bayesian density estimation using Bernstein polynomials. *Canadian Journal of Statistics* **27**, 105–126.

Philander, S. (1985). El niño and la niña. *Journal of the Atmospheric Sciences* **42**, 2652–2662.

Press, J. S. (1972). Multivariate stable distributions. *Journal of Multivariate Analysis* **2**, 444–462.

Richardson, R., Kottas, A. & Sansó, B. (2017). Flexible integro-difference equation modeling for spatio-temporal data. *Computational Statistics & Data Analysis* **109**, 182–198.

Richardson, R., Kottas, A. & Sansó, B. (2018). Bayesian non-parametric modeling for integro-difference equations. *Statistics and Computing* **28**, 87–101.

Richardson, R. A. (2017). Sparsity in nonlinear dynamic spatiotemporal models using implied advection. *Environmetrics* **28**, e2456.

Ropelewski, C. F. & Halpert, M. S. (1987). Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation. *Monthly Weather Review* **115**, 1606–1626.

Samorodnitsky, G. & Taqqu, M. S. (1997). Stable non-Gaussian random processes. *Econometric Theory* **13**, 133–142.

Scerri, K., Dewar, M. & Kadirkamanathan, V. (2009). Estimation and model selection for an IDE-based spatio-temporal model. *IEEE Transactions on Signal Processing* **57**, 482–492.

Sigrist, F., Künsch, H. R. & Stahel, W. A. (2012). A dynamic nonstationary spatio-temporal model for short term prediction of precipitation. *The Annals of Applied Statistics* **6**, 1452–1477.

Storvik, G., Frigessi, A. & Hirst, D. (2002). Stationary space-time Gaussian fields and their time autoregressive representation. *Statistical Modelling* **2**, 139–161.

West, M. & Harrison, J. (2006). *Bayesian forecasting and dynamic models*. Springer Science & Business Media.

Wikle, C. K. (2002). A kernel-based spectral model for non-Gaussian spatio-temporal processes. *Statistical Modelling* **2**, 299–314.

Wikle, C. K. & Cressie, N. (1999). A dimension-reduced approach to space-time Kalman filtering. *Biometrika* **86**, 815–829.

Wikle, C. K. & Holan, S. H. (2011). Polynomial nonlinear spatio-temporal integro-difference equation models. *Journal of Time Series Analysis* **32**, 339–350.

Wikle, C. K. & Hooten, M. B. (2010). A general science-based framework for dynamical spatio-temporal models. *Test* **19**, 417–451.

Xu, K., Wikle, C. K. & Fox, N. I. (2005). A kernel-based spatio-temporal dynamical model for nowcasting weather radar reflectivities. *Journal of the American Statistical Association* **100**, 1133–1144.