A diffusion-based spatio-temporal extension of Gaussian Matérn fields

Finn Lindgren[1], Haakon Bakka[2], David Bolin[3], Elias Krainski[3]
and Håvard Rue[3,4]

Abstract

Gaussian random fields with Matérn covariance functions are popular models in spatial statistics and machine learning. In this work, we develop a spatio-temporal extension of the Gaussian Matérn fields formulated as solutions to a stochastic partial differential equation. The spatially stationary subset of the models have marginal spatial Matérn covariances, and the model also extends to Whittle-Matérn fields on curved manifolds, and to more general non-stationary fields. In addition to the parameters of the spatial dependence (variance, smoothness, and practical correlation range) it additionally has parameters controlling the practical correlation range in time, the smoothness in time, and the type of non-separability of the spatio-temporal covariance. Through the separability parameter, the model also allows for separable covariance functions. We provide a sparse representation based on a finite element approximation, that is well suited for statistical inference and which is implemented in the R-INLA software. The flexibility of the model is illustrated in an application to spatio-temporal modeling of global temperature data.

[1] School of Mathematics, The University of Edinburgh, Scotland. As part of the EUSTACE project, Finn Lindgren received funding from the European Union's Horizon 2020 Programme for Research and Innovation, under Grant Agreement no. 640171.

[2] Kontali, Oslo, Norway.

[3] CEMSE Division, King Abdullah University of Science and Technology, Kingdom of Saudi Arabia.

[4] Address for correspondance: Professor Håvard Rue, King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia. Email: haavard.rue@kaust.edu.sa
Received: June 2023

1. Introduction

1.1. Modelling spatio-temporal data

Statistical models for spatio-temporal data have applications in areas ranging from the analysis of environmental data (Cameletti et al., 2013) and climate data (Wood et al., 2004; Fuglstad and Castruccio, 2020), to resource and risk modeling (e.g., of wildfires, Serra et al. (2014)), disease modeling (Bhatt et al., 2015; Moraga, 2019), and ecology (Yuan et al., 2017; Zuur, Ieno and Saveliev, 2017). These models typically use spatio-temporal random effects, defined as Gaussian spatio-temporal stochastic processes and rely on a large body of theoretical and methodological literature (Stein, 2012; Gelfand et al., 2010; Cressie and Wikle, 2011, and references therein).

At best, this theory is carefully studied when the spatio-temporal model is constructed, so that the model with the most appropriate assumptions can be used. In practice, however, users of statistical software often choose a model based on convenience. If there are available code examples, the choices made in these will often be carried forward into future analyses. For example, users of R-INLA (Rue, Martino and Chopin, 2009, 2017; van Niekerk et al., 2021; van Niekerk and Rue, 2024; Gaedke-Merzhäuser et al., 2022; van Niekerk et al., 2023) construct space-time models through Kronecker products of a spatial Matérn model, and first- or second-order autoregressive models in time, following the code examples in Krainski et al. (2019). This paper is aimed at improving the general practice of space-time data analysis, by providing a new family of spatio-temporal stochastic processes for use as random effects in statistical software.

We will mainly discuss stochastic processes $u(\boldsymbol{s},t)$ that are stationary and spatially isotropic, i.e., the covariance function can be written as $\operatorname{cov}(u(\boldsymbol{s}_1,t_1),u(\boldsymbol{s}_2,t_2)) = R(h_s,h_t)$, where $h_s = ||\boldsymbol{s}_1 - \boldsymbol{s}_2||$ and $h_t = |t_1 - t_2|$, but will also extend these process models to spatial non-stationarity and processes on general manifolds. We consider these stochastic processes in the context of hierarchical models, as a latent model component, observed through some measurement process, with no direct measurements of the stochastic process itself. Consider, for example, a model with a linear predictor

$$\eta(\boldsymbol{s},t) = \sum_{i=1}^{m} X_i(\boldsymbol{s},t)\beta_i + f_1\{z_1(\boldsymbol{s},t)\} + \ldots + f_k\{z_k(\boldsymbol{s},t)\} + u(\boldsymbol{s},t), \tag{1}$$

that is connected to the response *y* through some likelihood or loss function (Bissiri, Holmes and Walker, 2016) such that $\mathrm{E}\{y(\boldsymbol{s},t)\} = g\{\eta(\boldsymbol{s},t)\}$ for some fixed and known function *g*. Here X_i and z_j are covariates that vary over both space and time, β_i the regression coefficient for the fixed effects, and $f_j(z_j)$ are random effects. Typical examples are splines and latent Gaussian processes used to approximate the effect of altitude or distance to coastline. This common situation with a stochastic process as a model component impacts the methodological considerations we make. The predictor is also a spatio-temporal stochastic process, with a covariance function that can be deduced from the assumptions on the model components. However, properties of the predictor that we

may discover by investigating the covariance function of the predictor may not be shared by the spatio-temporal model component *u* because of the other factors. Hence, we may have little prior information about the covariance structure of the spatio-temporal model component, except that it should be physically realistic, and should mimic the dependency structure in models of physical processes.

Users of software for spatio-temporal modelling most often use separable models (see, e.g., Bakka et al. (2018); Krainski et al. (2019)), i.e., models where *u* has a covariance function of the form $R(h_s, h_t) = R_s(h_s)R_t(h_t)$, for some spatial and temporal marginal covariance functions $R_s(\cdot)$ and $R_t(\cdot)$. This is typically not because this is a desired property, but since such models are readily available in statistical software, and there are many good arguments for why models should not be assumed separable, see Stein (2005), Cressie and Huang (1999), Fonseca and Steel (2011), Rodrigues and Diggle (2010), Gneiting (2002), Sigrist, Künsch and Stahel (2015), Wikle (2015).

1.2. The Matérn family of covariance functions

The most well known family of covariance functions for stationary random fields on $\mathbb{R}^d$ is the Matérn covariance,

$$R_M(h) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\kappa h)^{\nu}K_{\nu}(\kappa h),\qquad(2)$$

where $\nu, \kappa > 0$ are smoothness and scale parameters, σ^2 is the variance of the corresponding random field, K_{ν} is the Bessel function of the second kind of order ν, and Γ is the Gamma function. An important property of this covariance family is that it allows for explicit control of the differentiability of the corresponding stochastic process through the parameter ν. It further allows for control of the practical correlation range $r = \sqrt{8\nu}/\kappa$ (Lindgren et al., 2011). The covariance function is usually attributed to Matérn (1960), and it was advocated early by Handcock and Stein (1993) and Stein (2012). See Guttorp and Gneiting (2006) for a historical account of the covariance function and its connections to various areas in physics.

The goal of this paper is to extend the Matérn covariance function to a family of spatio-temporal covariance functions. One way of doing this would be to extend the covariance function to a spatio-temporal covariance. However, we argue that it is better to base the extension on some of the other equivalent mathematical representations, or *views*, of Gaussian Matérn fields. One such alternative representation is the stochastic partial differential equation (SPDE) representation by Whittle (1963). Specifically, a Gaussian Matérn field on $\mathbb{R}^d$ solves the SPDE

$$(\kappa^2 - \Delta)^{\alpha/2}u = \mathcal{W},\qquad(3)$$

where $\kappa > 0$, Δ is the Laplacian, $\mathcal{W}$ is Gaussian white noise, and $\alpha = \nu + d/2$. Via the SPDE representation, we note that a Gaussian Matérn field has precision operator $Q = (\kappa^2 - \Delta)^{\alpha}$. The precision operator (as well as the pseudo-differential operator $(\kappa^2 -$

$\Delta)^{\alpha/2}$) are defined in terms of Fourier transforms (Lindgren et al., 2011), and informally, we get the Fourier transform of the precision operator by replacing derivatives with d-dimensional wave-numbers $\boldsymbol{w}$. For any precision operator which is a polynomial in the Laplacian, $Q = \text{Poly}(-\Delta)$, such as the Matérn operator with $\alpha \in \mathbb{N}$, this results in a polynomial $\mathcal{F}(Q) = \text{Poly}(\|\boldsymbol{w}\|^2)$. This function is the reciprocal of the spectrum of the Gaussian process, illustrating why many common spectrums are the reciprocal of an even polynomial. In fact, Rozanov (1977) showed that a stationary stochastic process on $\mathbb{R}^d$ is Markov if and only if the spectral density is the reciprocal of a polynomial, and more generally, a stochastic process is Markov if the precision operator is a local operator, which is the case for integer powers of the Laplacian. For further details on the theory of the SPDE representation, see Kelbert, Leonenko and Ruiz-Medina (2005); Prévôt and Röckner (2007); Lindgren et al. (2011); Bolin and Kirchner (2020); Bolin et al. (2023).

We could also represent a Gaussian Matérn field as a stochastic integral with respect to white noise. For Gaussian Matérn fields, the kernel in the integral representation is the Green's function of the differential operator (see, e.g., Bolin, 2014). This representation can be used to define other valid covariance functions by replacing the Green's function with some other kernel (see, e.g., Fuentes, 2002; Higdon, 2002; Rodrigues and Diggle, 2010).

The modeling approaches stemming from these different views of the Gaussian Matérn fields can be thought of as implicit and explicit. In implicit approaches such as the covariance-based representation, one does not have a direct formulation of the process itself, and properties of interest need to be derived from the covariance function. In explicit, or constructive, approaches one directly defines the process through, e.g., an SPDE or a stochastic integral with the desired properties encoded. In this paper we follow the explicit approach to construct a stochastic process based on diffusion processes. Other properties, such as covariance non-separability, are then merely consequences of the explicit construction.

1.3. SPDE-based spatio-temporal generalisations of the Matérn covariance family

There is a large literature on spatio-temporal covariance models see, e.g., Porcu, Furrer and Nychka (2021) and the references within). Broadly, models for spatio-temporal Gaussian random fields can be divided into two categories; the implicit second-order covariance based models and explicit dynamical models (Cressie and Wikle, 2011; Roques, Allard and Soubeyrand, 2022). It should be noted that Porcu et al. (2021), contrary to this terminology, classifies the SPDE-based methods as implicit since they do not explicitly specify the covariance function. However, the covariance functions is merely a property of the process, and only indirectly defines the process family, whereas dynamical models directly determine the spatial and temporal evolution of the process. As shown by Lindgren et al. (2011), the covariance does not have an inherent advantage over spectral and precision operator/matrix methods, for practical applications and computations.

The second-order model specifications specify the Gaussian process properties by specifying its first two moments, and are thus based on formulating valid spatio-temporal covariance functions. In dynamical model specifications, the evolution of the Gaussian process is explicitly described either by specifying the conditional distributions of the current state of the process given its past through conditional distributions (e.g., Storvik, Frigessi and Hirst, 2002), or by specifying the process as the solution to an SPDE (Cressie and Wikle, 2011). One of the advantages with the dynamical approach is that it avoids the difficulties with formulating flexible and yet valid spatio-temporal covariance functions that can possess features such as non-separability or non-stationarity. In this work we focus on dynamical models specified through SPDEs, which makes extension to non-stationary fields and manifold models straightforward.

Several papers have been using the SPDE view to suggest models for spatio-temporal stochastic processes. A common extension of Matérn covariance fields to space-time is to use it as the spatial component in a separable model. Jones and Zhang (1997) discuss how separable covariance functions can be understood through differential operators, written as $L = L_s L_t$, where L_s is a purely spatial operator and L_t is a purely temporal operator. In agreement with Jones and Zhang (1997), we note that these operators are almost never encountered when modeling physical reality, hence, separable models are typically not physically motivated models for the spatio-temporal process.

Whittle (1963) considered a spatio-temporal stochastic process formulated as a solution to

$$\frac{\partial u}{\partial t} + (\kappa^2 - \Delta)u(\boldsymbol{s},t) = \varepsilon(\boldsymbol{s},t), \tag{4}$$

where $\varepsilon(\boldsymbol{s},t)$ is a stationary spatio-temporal noise process. Whittle (1986) denoted the model as a "diffusion-injection model" since it is a diffusion process with stochastic variability "injected" through the noise process on the right-hand side. Despite being a natural spatio-temporal extension of the Matérn model (3) with $\alpha = 2$, the model does not have any flexibility in terms of differentiability in space or time. Jones and Zhang (1997) proposed a generalization, with greater flexibility for the marginal spatial covariances, by considering the fractional SPDE

$$\left(\frac{\partial}{\partial t} + \left(\kappa^2 - \Delta\right)^{\alpha/2}\right)u(\boldsymbol{s},t) = \mathrm{d}\mathcal{E}(\boldsymbol{s},t), \tag{5}$$

where $\mathrm{d}\mathcal{E}$ is space-time Gaussian white noise. When requiring spatial operator order $\alpha > d$, this SPDE has regular continous solutions. In order to allow smaller operator orders α, such as a dampened ordinary diffusion operator with $\alpha = 2$ on $\mathbb{R}^2$, as in Whittle (1963), the driving noise process would need to have spatial dependence. We will make this precise in later sections. An advantage with (5) is that the spatial smoothness can be controlled, since the solutions on the spatial domain $\mathbb{R}^d$ have smoothness $v_s = \alpha - d/2$. The disadvantage is that the temporal smoothness also is determined by α. As we will see later, the marginal temporal differentiability of the solution is $v_t = (1 - d/\alpha)/2$.

A model with general differentiability in both space and time was formulated by Stein (2005), who considered Gaussian spatio-temporal models specified through the spectrum

$$S(\boldsymbol{w}_s, w_t) = \{c_1(a_1^2 + \|\boldsymbol{w}_s\|^2)^{\alpha_1} + c_2(a_2^2 + |w_t|^2)^{\alpha_2}\}^{-\nu}, \tag{6}$$

where $c_1 > 0, c_2 > 0, a_1, a_2$ are scale parameters, $a_1^2 + a_2^2 > 0$, α_1, α_2 and ν are smoothness parameters with further restrictions in order to obtain a model with finite variance. For example, on a two-dimensional spatial and one-dimensional temporal domain, $2/\alpha_1 + 1/\alpha_2 < 2\nu$ is required. Stein's model can also be stated as an SPDE driven by space-time white noise,

$$\left(c_1(a_1^2 - \Delta)^{\alpha_1} + c_2\left(a_2^2 - \frac{\partial^2}{\partial t^2}\right)^{\alpha_2}\right)^{\nu/2} u(\boldsymbol{s}, t) = \mathrm{d}\mathcal{E}(\boldsymbol{s}, t), \tag{7}$$

see Krainski (2018) and Vergara, Allard and Desassis (2022). A related model based on spectral densities, which also has separable models as a special case, was considered by Fuentes, Chen and Davis (2008).

The case $\alpha = 2$ of (5) for general dimension was considered in (Lindgren et al., 2011, Section 3.5), suggesting the generalisation

$$\left(\frac{\partial}{\partial t} + \kappa^2 + \boldsymbol{m} \cdot \nabla - \nabla \cdot \boldsymbol{H} \nabla\right) u(\boldsymbol{s}, t) = \mathrm{d}\mathcal{E}_Q(\boldsymbol{s}, t), \tag{8}$$

where $\boldsymbol{H}$ is a constant diffusion matrix, $\boldsymbol{m}$ is an advection (transport) vector field, and the innovation process $\mathrm{d}\mathcal{E}_Q(\boldsymbol{s}, t)$ white noise in time but is sufficiently smooth in space to generate regular solutions $u(\boldsymbol{s}, t)$; see Lindgren et al. (2011) and Sigrist et al. (2015, Sec 2.2). Physically, this model might be interpreted as a dampened advection-diffusion process, with the driving mechanism of the space-time field, such as introducing new mass (or, particles) into the system, having positive spatial correlation. See also Liu, Yeo and Lu (2022); Clarotto et al. (2022).

In this work, we introduce another generalisation of the models by Jones and Zhang (1997) and Lindgren et al. (2011) that intersects, but is otherwise distinct from, the Stein model family.

1.4. Outline

In Section 2 we introduce a new family of SPDE-based spatio-temporal stochastic processes. Model properties such as spatial and temporal differentiability, and parameter interpretations, are presented in Section 3.We present a sparse basis function representation in Section 4, and an implementation in R-INLA (Rue et al., 2009) in Supplementary Materials, which allows us to construct models with different likelihoods and several random effects in a generalised additive model context. In Section 5, we present a forecasting example that illustrates clearly the difference between separable models and non-separable diffusion-based models, and an application to a global temperature dataset. The article concludes with a discussion in Section 6.

2. A diffusion-based family of spatio-temporal stochastic processes

In this section we define a diffusion-based extension of the Gaussian Matérn fields to a family of spatio-temporal stochastic processes (abbreviated DEMF). The main property we aim for is that the process should be a Gaussian Matérn field when considered for a fixed time point in $\mathbb{R}^d$. That is, when the process is considered on the spatial domain $\mathcal{D} = \mathbb{R}^d$, the spatial marginalisations of the process have Matérn covariances. When the models are considered on a general (compact) manifold $\mathcal{D}$, the spatial marginalisations are solutions to a generalised spatial Whittle-Matérn model on $\mathcal{D}$ (Lindgren, Bolin and Rue, 2022).

Consider the operator $L_s = \gamma_s^2 - \Delta$, $\gamma_s > 0$, on a spatial domain $\mathcal{D}$, including any boundary conditions needed for compact domains. Let the precision operator for the generalised Whittle-Matérn covariances be $Q(\gamma_s, \gamma_e, \alpha) = \gamma_e^2 L_s^\alpha$, corresponding to solutions $v(\boldsymbol{s})$ to the spatial stochastic SPDE

$$\gamma_e L_s^{\alpha/2} v(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \qquad \boldsymbol{s} \in \mathcal{D} \tag{9}$$

where $\mathcal{W}$ is a spatial white noise process, as discussed by Whittle (1963) and Lindgren et al. (2011). When $\mathcal{D} = \mathbb{R}^d$, and a stationary condition is imposed, these processes are regular Matérn processes. We then define a noise process $\mathrm{d}\mathcal{E}_Q(\boldsymbol{s}, t)$ as Gaussian noise that is white in time but correlated in space, with precision operator $Q = Q(\gamma_s, \gamma_e, \alpha_e)$ for some non-negative α_e. For $a > 0$, the cumulative time-integral process

$$\mathcal{E}_Q(\boldsymbol{s}, (0, a]) = \int_{t=0}^a \mathrm{d}\mathcal{E}_Q(\boldsymbol{s}, t) \tag{10}$$

is a Q-Wiener process (Da Prato and Zabczyk, 2014), with spatial precision operator Q/a.

The case of a separable covariance model with a Matérn covariance in space and an exponential covariance in time is obtained from the stationary solutions to

$$\left(\frac{\partial}{\partial t} + \kappa \right) u(\boldsymbol{s}, t) = \mathrm{d}\mathcal{E}_Q(\boldsymbol{s}, t), \qquad (\boldsymbol{s}, t) \in \mathcal{D} \times \mathbb{R}. \tag{11}$$

This is a spatial generalisation of the Ornstein-Uhlenbeck processes. We aim to produce a space-time model with diffusive behaviour. For this, we replace the dampening coefficient κ in (11) with a power of the dampended diffusion operator L_s, defining a model family of the time-stationary solutions to iterated diffusion-like processes

$$\left(\gamma_t \frac{d}{dt} + L_s^{\alpha_s/2} \right)^{\alpha_t} u(\boldsymbol{s}, t) = \mathrm{d}\mathcal{E}_Q(\boldsymbol{s}, t), \qquad (\boldsymbol{s}, t) \in \mathcal{D} \times \mathbb{R}. \tag{12}$$

When $\mathcal{D} = \mathbb{R}^d$, the space-stationary solutions are used. For compact manifolds with boundary, the operators L_s and Q are equipped with suitable boundary conditions on $\partial \mathcal{D}$.

In total, the model has three non-negative smoothness parameters $(\alpha_t, \alpha_s, \alpha_e)$ and three positive scale parameters $(\gamma_t, \gamma_s, \gamma_e)$. It is not immediately obvious how the definition (12) would be interpreted for non-integer powers α_t. However, by taking advantage of the spectral properties of the operators, we define the following model, which has an operator that more clearly allows fractional powers α_t, as

$$\left(-\gamma_t^2 \frac{d^2}{dt^2} + L_s^{\alpha_s} \right)^{\alpha_t/2} u(\boldsymbol{s},t) = \mathrm{d}\mathcal{E}_Q(\boldsymbol{s},t), \qquad (\boldsymbol{s},t) \in \mathcal{D} \times \mathbb{R}. \qquad (13)$$

Theorem 2.1. *For $\mathcal{D} = \mathbb{R}^d$, as well as for other domains where L_s has well defined positive powers, the definitions* (12) *and* (13) *of the Gaussian process $u(\boldsymbol{s},t)$ coincide for $\alpha_t \in \mathbb{N}$.*

Proof. This can be seen by applying the techniques developed in Vergara et al. (2022). Alternatively, the transfer function $G(\omega_t)$ (see Lindgren, 2012, Chapter 4) for the temporal linear filter defined by the operator in (12) is $G(\omega_t) = (i\gamma_t \omega_t + L_s^{\alpha_s/2})^{\alpha_t}$, well-defined for positive integers α_t, and has $|G(\omega_t)|^2 = (\gamma_t^2 \omega_t^2 + L_s^{\alpha_s})^{\alpha_t}$. The transfer function $H(\omega_t)$ for the temporal linear filter defined by the operator in (13) is $H(\omega_t) = (\gamma_t^2 \omega_t^2 + L_s^{\alpha_s})^{\alpha_t/2}$, well-defined for positive α_t. We see that $|G(\omega_t)|^2 = |H(\omega_t)|^2$, so the spectral properties of the two process definitions coincide for positive integer α_t values.   ∎

It should be noted that it would be possible to give a more direct definition of the model (12) with fractional α_t, but this would require more sophisticated mathematical tools, which is outside the scope of this work.

The two representations make it clear that the model with $\alpha_e = 0$ is a special case of the Stein (2005) model family, with $a_1 = 0$ and $\alpha_1 = 1$ in (7), and that the model of Jones and Zhang (1997) is obtained by setting $\alpha_e = 0$ and $\alpha_t = 1$ in (12).

The use of the same spatial operator L_s in the left hand side of (13) as in the precision operator on the right hand side is what causes the spatial marginalisation of the process to be Matérn fields in the simplest case, as will be shown in Section 3. The parameters α_t, α_s, and α_e determine the differential operator orders involved in the SPDE operator and therefore also the smoothness properties of the process, as shown in Section 3.

The model can be further generalised by allowing the γ parameters to vary across space. This is most straightforward for γ_s, since that only alters the L_s operator. For complex domains, as well as when L_s is generalised to vary across space, the resulting solutions are not space-stationary, but still have marginal spatial properties defined by powers of L_s. The practical precision construction in Section 4 can be generalised to separable non-stationarity, where γ_t is allowed to depend on time and γ_s and γ_e depends on space, since that retains commutativity between the temporal and spatial operators.

2.1. Compact domains and manifolds

For compact domains, the model definitions include some form of boundary conditions. These boundary conditions induce boundary effects near the domain boundary, and as shown in Lindgren et al. (2011), if such effects are undesirable, one can extend the domain by at least the spatial range. By taking advantage of a non-stationary spatial operator, the barrier method introduced by Bakka et al. (2019) can also be used to nearly eliminate boundary effects, as well as to obtain models that appropriately take complex geography into account. Similarly, all the common extensions onto curved manifolds, such as the globe, can be implemented using the same approaches as for $\mathbb{R}^d$. This includes the finite element methods used in Section 4, but also Fourier-like spectral basis function expansions given by the eigenfunctions of the Laplacian, either given in closed form, e.g., spherical harmonics on the globe, or obtained numerically from finite element eigenfunction computations. See Lindgren et al. (2022) for an overview of the literature on these alternative methods.

3. Parameter interpretations and model properties

In this section we discuss marginal spatial and temporal properties of the diffusion-based model (12). In order to simplify the exposition, we focus on the ordinary Matérn covariance case when the spatial domain is $\mathcal{D} = \mathbb{R}^d$. In this case, the space-time spectral density of the stationary solutions $u(\boldsymbol{s},t)$ to (13) is

$$S_u(\boldsymbol{\omega}_s, \omega_t) = \frac{1}{(2\pi)^{d+1}\gamma_e^2[\gamma_t^2\omega_t^2 + (\gamma_s^2 + \|\boldsymbol{\omega}_s\|^2)^{\alpha_s}]^{\alpha_t}(\gamma_s^2 + \|\boldsymbol{\omega}_s\|^2)^{\alpha_e}}, \qquad (14)$$

for $(\boldsymbol{\omega}_s, \omega_t) \in \mathbb{R}^d \times \mathbb{R}$. The space-time covariance function is given by the Fourier integral

$$R_u(\boldsymbol{s},t) = \int_{\mathbb{R}} \int_{\mathbb{R}^d} \exp[i(\boldsymbol{\omega}_s \cdot \boldsymbol{s} + \omega_t t)] S_u(\boldsymbol{\omega}_s, \omega_t) \, \mathrm{d}\boldsymbol{s} \, \mathrm{d}t \qquad (15)$$

for spatial lags $\boldsymbol{s}$ and temporal lags t.

3.1. Sample path continuity and differentiability theory

For fields with Matérn covariance functions, the degree of differentiability is encoded in the smoothness index ν. For models with space-time spectral density given by (14), the marginal covariance in time is not generally of the Matérn class, so we need to use more general conditions for determining the smoothness.

The differentiability of a stationary process $x(t)$, $t \in \mathbb{R}$, is determined by the decay rate of its spectral density. If $S(\omega) \sim \omega^{-\gamma}$ for some $\gamma > 0$ for large ω, then the process is a times mean square differentiable for all $a < \frac{\gamma-1}{2}$ (Stein, 2005).

For stationary Gaussian processes, stronger statements of almost sure sample path continuity of derivatives and Hölder continuity can be made. The technical details can be found in Section 9.3 of Cramér and Leadbetter (1967) and Scheuerer (2010), and are

summarised in Appendix A, including a more formal characterisation of the smoothness index. The results show that Gaussian processes with spectral densities satisfying $S(\boldsymbol{\omega}) \sim \|\boldsymbol{\omega}\|^{-2\nu-d}$ for some $\nu > 0$ and large $\|\boldsymbol{\omega}\|$ have smoothness index ν. This means that the sample paths have almost surely continuous derivatives of order up to and including $k = \lceil \nu \rceil - 1$, and that the derivatives of order k are Hölder of index a for any $0 < a < \nu - k$. Further, the sample paths are almost surely in the Sobolev spaces $W^{b,2}$ for any $b < \nu$, on finite subsets of $\mathbb{R}^d$. Although these results are derived specifically for $\mathbb{R}^d$, it is clear that sample path properties of Whittle-Matérn fields on more general but smooth domains will have similar, and usually identical, local differentiability properties, based on the decay rate of the eigenspectrum of the Laplacian. In particular, the spectral Fourier representations on the 2D sphere $\mathbb{S}^2$ lead to series that converge under the same conditions as the continuous spectra on $\mathbb{R}^2$.

The smoothness index ν can be interpreted as the smallest value for which some form of weak continuity does *not* hold. For a process on a multidimensional domain with potentially different smoothness in different directions, Theorem A.1 and smoothness definition in Appendix A will be applied to the one-dimensional marginals of the process.

3.2. Properties of the spatio-temporal model

We can now show that the spatial marginals of $u(\boldsymbol{s},t)$, i.e. for fixed t, are Matérn covariance fields, given that the smoothness parameters are chosen appropriately. To keep some notational brevity, we first define the unit variance and range Matérn covariance function $R_\nu^M(t)$,

$$R_\nu^M(t) = \frac{1}{\Gamma(\nu)2^{\nu-1}} t^\nu K_\nu(t), \quad t \geq 0, \tag{16}$$

and the scaling constants

$$C_{\mathbb{R}^d,\alpha} = \frac{\Gamma(\alpha - d/2)}{\Gamma(\alpha)(4\pi)^{d/2}},$$

for $d = 1, 2, 3, \ldots$ and $\alpha > d/2$. These appear as variance scaling constants for the regular Whittle-Matérn SPDE models.

Proposition 3.1. *Define the effective spatial marginal operator order* $\alpha = \alpha_e + \alpha_s(\alpha_t - 1/2)$ *and assume that* $\alpha > d/2$. *Then the solution* $u(\boldsymbol{s},t)$ *to* (13) *has marginal spatial covariance function*

$$\mathrm{cov}(u(\boldsymbol{s}_1,t), u(\boldsymbol{s}_2,t)) = \sigma^2 R_{\nu_s}^M(\gamma_s \|\boldsymbol{s}_2 - \boldsymbol{s}_1\|)$$

where $\nu_s = \alpha - d/2$ *is the spatial smoothness index and*

$$\sigma^2 = \frac{C_{\mathbb{R},\alpha_t} C_{\mathbb{R}^d,\alpha}}{\gamma_e^2 \gamma_t \gamma_s^{2\alpha-d}}. \tag{17}$$

Proof. See Appendix D.1, that also includes a derivation of the marginal spatial cross-spectra for different time lags. ∎

Proposition 3.2. *Assume* $\alpha_t, \alpha_s, \alpha_e$ *satisfy* $\alpha > d/2$. *Then the temporal smoothness index of the solutions* $u(\boldsymbol{s},t)$ *to* (13) *is* $\nu_t = \min\left[\alpha_t - \frac{1}{2}, \frac{\nu_s}{\alpha_s}\right]$, *and for* $d = 2$, *the marginal temporal spectrum is*

$$S_t(\omega_t) \propto {}_2F_1\left(\alpha_t, \frac{\alpha_e - 1}{\alpha_s} + \alpha_t, \frac{\alpha_e - 1}{\alpha_s} + \alpha_t + 1; -\omega_t^2 \gamma_t^2 / \gamma_s^{2\alpha_s}\right),$$

where ${}_2F_1$ *denotes the hypergeometric function.*

Proof. See Appendix D.2. ∎

For integer values of the operator orders, the hypergeometric function can be expressed using elementary functions. When $\alpha_t = \alpha_s = 2$ and $\alpha_e = 0$ for $d = 2$, we obtain

$$S_t(\omega_t) \propto \int_0^\infty \frac{1}{(\widetilde{\omega}_t^2 + (1+v)^2)^2} \, \mathrm{d}v = \frac{\arctan(\widetilde{\omega}_t)}{2\widetilde{\omega}_t^3} - \frac{1}{2\widetilde{\omega}_t^2(\widetilde{\omega}_t^2 + 1)}, \tag{18}$$

where $\widetilde{\omega}_t = \omega_t \gamma_t / \gamma_s^{\alpha_s}$, showing that the marginal temporal covariance is not a Matérn covariance. The exception is the separable case, where the temporal covariance function is a Matérn covariance function with smoothness index $\alpha_t - 1/2$.

Corollary 3.2.1. *Assume that* $\alpha_s = 0$, $\alpha_t > 1/2$, *and* $\alpha_e > d/2$. *Then the stationary solutions* $u(\boldsymbol{s},t)$ *to* (13) *have a separable space-time covariance function where the spatial covariance is given by Proposition 3.1 and the marginal temporal covariance function is*

$$C(u(\boldsymbol{s},t_1), u(\boldsymbol{s},t_2)) = \sigma^2 R_{\nu_t}^M(\gamma_t^{-1}|t_2 - t_1|),$$

where $\nu_t = \alpha_t - 1/2$ *and* σ^2 *is given by* (17) *with* $\alpha = \alpha_e$.

Proof. Follows directly from the product form of the space-time spectrum (14). ∎

In Table 1, we summarise the general smoothness results, as well as some important special cases. The special cases denoted *diffusion* are generalised analogues of the diffusion-injection model (4), and the special *critical diffusion* model is later used in Sections 4 and 5. The general conditions on the α parameters that give well defined solutions are encoded in the spatial and temporal smoothness conditions $\nu_s > 0$ and $\nu_t > 0$, and can also be written as the conditions $\alpha = \alpha_e + \alpha_s(\alpha_t - 1/2) > d/2$ and $\alpha_t > 1/2$.

Table 1. *Summary of the smoothness properties of the solutions $u(\boldsymbol{s},t)$ for different values of the parameters $\alpha_t, \alpha_s, \alpha_e$, together with some examples. Here ν_t and ν_s respectively denote the temporal and spatial smoothnesses of the process.*

α_t	α_s	α_e	Type	ν_t	ν_s
α_t	α_s	α_e	General	$\min\left[\alpha_t - \frac{1}{2}, \frac{\nu_s}{\alpha_s}\right]$	$\alpha_e + \alpha_s(\alpha_t - \frac{1}{2}) - \frac{d}{2}$
α_t	0	α_e	Separable	$\alpha_t - \frac{1}{2}$	$\alpha_e - \frac{d}{2}$
α_t	α_s	$\frac{d}{2}$	Critical	$\alpha_t - \frac{1}{2}$	$\alpha_s(\alpha_t - \frac{1}{2})$
α_t	α_s	0	Fully non-separable	$\alpha_t - \frac{1}{2} - \frac{d}{2\alpha_s}$	$\alpha_s(\alpha_t - \frac{1}{2}) - \frac{d}{2}$
1	2	$\alpha_e > \frac{d}{2}$	Sub-critical diffusion	$1/2$	$\alpha_e + 1 - \frac{d}{2}$
1	2	$\frac{d}{2}$	Critical diffusion	$1/2$	1
1	2	$\frac{d}{2} - 1 < \alpha_e < \frac{d}{2}$	Super-critical diffusion	$\nu_s/2$	$\alpha_e + 1 - \frac{d}{2}$
1	0	2	Separable	$1/2$	$2 - \frac{d}{2}$
$3/2$	2	0	Fractional diffusion	$1 - \frac{d}{4}$	$2 - \frac{d}{2}$
2	2	0	Iterated diffusion	$\frac{3}{2} - \frac{d}{4}$	$3 - \frac{d}{2}$

3.2.1. Quantifying non-separability

From Table 1 we can see that the α_e parameter controls the type of non-separability. An important case is $\alpha_e = 0$, which we refer to as fully non-separable models. The spectral density for such models is a subfamily of the Stein (2005) spectral model family. The degree of non-separability can be quantified by the relation between α_e and the effective marginal spatial operator order α. We introduce the non-separability parameter $\beta_s = 1 - \alpha_e/\alpha = 1 - \alpha_e/(\nu_s + d/2) \in [0,1]$, where $\beta_s = 0$ gives a separable model, and $\beta_s = 1$ gives a "maximally non-separable" model. Assuming given values for the temporal smoothness $\nu_t > 0$, spatial smoothness $\nu_s > 0$, and non-separability $\beta_s \in [0,1]$, we can find the corresponding values of $(\alpha_t, \alpha_s, \alpha_e)$. Let $\beta_*(\nu_s, d) = \frac{\nu_s}{\nu_s + d/2}$. Then

$$\alpha_t = \nu_t \max\left(1, \frac{\beta_s}{\beta_*(\nu_s, d)}\right) + \frac{1}{2},$$

$$\alpha_s = \frac{\nu_s}{\nu_t} \min\left(\frac{\beta_s}{\beta_*(\nu_s, d)}, 1\right) = \frac{1}{\nu_t}\min\left[(\nu_s + d/2)\beta_s, \nu_s\right],$$

$$\alpha_e = \frac{1 - \beta_s}{\beta_*(\nu_s, d)}\nu_s = (\nu_s + d/2)(1 - \beta_s).$$

The critical branching point $\beta_s = \beta_*(\nu_s, d)$ motivates the term *critical* for such models. Models with $\beta_s < \beta_*(\nu_s, d)$ are *sub-critical* and models with $\beta_s > \beta_*(\nu_s, d)$ are *super-critical*. The critical models have $\alpha_t = \nu_t + 1/2$, $\alpha_s = \nu_s/\nu_t$, and $\alpha_e = d/2$. The diffusion models in Table 1 with $\alpha_t = 1$ and $\alpha_s = 2$ are of particular interest, as they arise from a basic heat equation.

Notably, the fully non-separable diffusion model ($\alpha_t = 1, \alpha_s = 2, \alpha_e = 0$) requires $d = 1$ to ensure $\nu_s > 0$, whereas the fully non-separable twice iterated diffusion model ($\alpha_t = 2, \alpha_s = 2, \alpha_e = 0$) is valid for $d \in \{1, 2, 3, 4, 5\}$.

3.2.2. Scale parameter interpretation

To improve the interpretability of the scale parameters, we define σ, r_s, and r_t via

$$\sigma^2 = \frac{C_{\mathbb{R}, \alpha_t} C_{\mathbb{R}^d, \alpha}}{\gamma_t \gamma_e^2 \gamma_s^{2\alpha - d}} \tag{19}$$

$$r_s = \gamma_s^{-1} \sqrt{8\nu_s} \tag{20}$$

$$r_t = \gamma_t \gamma_s^{-\alpha_s} \sqrt{8(\alpha_t - 1/2)}, \tag{21}$$

where r_s is the correlation range as in Lindgren et al. (2011), giving approximately correlation of 0.13 at r_s distance in space (keeping time fixed). Similarly, r_t controls the temporal correlation range for the separable model. In the non-separable cases, it is the temporal correlation range for the evolution of the spatial eigenfunction corresponding to the smallest eigenvalue of the Laplacian, i.e. a constant function over space, evolving in time. Eigenfunctions for larger spatial eigenvalues have shorter temporal correlation range, so the combined effective range will typically be smaller than the nominal r_t value would indicate.

3.3. Examples

Table 2. *Four specific DEMF models on $\mathbb{R}^2$ or $\mathbb{S}^2$.*

Model	α_t	α_s	α_e	Type	ν_t	ν_s
A: DEMF(1,0,2)	1	0	2	Separable order 1	1/2	1
B: DEMF(1,2,1)	1	2	1	Critical diffusion	1/2	1
C: DEMF(2,0,2)	2	0	2	Separable order 2	3/2	1
D: DEMF(2,2,0)	2	2	0	Iterated diffusion	1	2

To simplify notation, we denote by $\text{DEMF}(\alpha_t, \alpha_s, \alpha_e)$ the model with given values for $(\alpha_t, \alpha_s, \alpha_e)$. The four models we will consider on $\mathbb{R}^2$ are defined in Table 2. For these, we choose γ_s, γ_e and γ_t, so that $\sigma = 1$, $r_s = 1$ and $r_t = 1$. This enables us to compare the non-Matérn behaviour of the temporal correlation to the spatial Matérn behaviour.

In general, the covariances are not available in closed form, but since the temporal covariance for each spatial frequency is of Matérn type, the spatial cross-spectra (derived in Appendix D.1) can be inverted numerically to obtain the cross-covariance. Specifically, the cross-covariance can be computed numerically with a 2D fast Fourier transform (FFT) computation for each fixed temporal lag (see Appendix B). This technique is related to the half-spectral space-time covariance models from Horrell and Stein (2017).

There, they focus on models where the temporal spectrum is known for each spatial location, $\mathcal{F}_t R(\boldsymbol{s},t) = f(\omega_t) g(\boldsymbol{s}, \omega_t)$, but the theory also covers the case of known spatial spectrum for each time point, $\mathcal{F}_s R(\boldsymbol{s},t) = f(\boldsymbol{\omega}_s) g(\boldsymbol{\omega}_s, t)$, that we use here.

In Figure 1 we show the spatio-temporal covariance function for these four models, and the marginal spatial covariances are shown in Figure 2. There is a clear difference between the spatio-temporal covariances, even though the marginal spatial covariances are identical for the first three models.
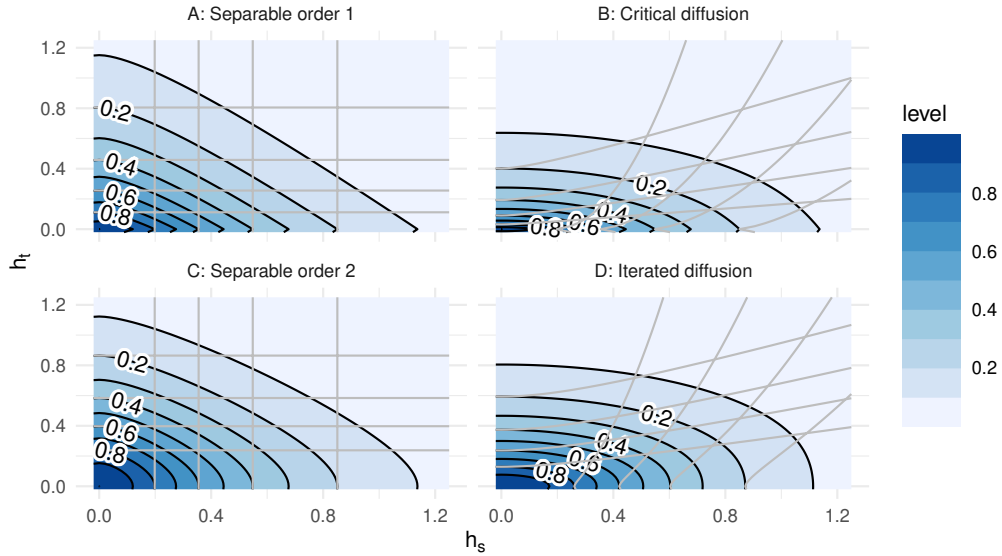
Figure 1. *The space-time covariance functions for spatial dimension $d = 2$, for the four models in Table 2, Section 3.3. The grey overlayed curves are level curves of the relative decay of the spatial and temporal covariances in relation to the marginal covariances. The non-separable models have non-orthogonal decay.*

3.4. Spheres and other manifolds

As noted earlier, the marginal spatial covariance properties of the DEMF models on general manifolds are rooted in the properties of the Whittle-Matérn operator, and depend on the specific geometry. However, the temporal structure is linked to each spatial frequency in the same way for every manifold, so we can focus on the effects on the spatial properties. Smoothness properties intuitively follow from the local properties of the differential operator on smooth manifolds, which locally behave like $\mathbb{R}^d$, so that is not the main obstacle to determining the process properties. Instead, it is the effect of the manifold's intrinsic curvature that prevents general closed form expressions for the covariance functions to be derived. On a compact manifold $\mathcal{D}$, the covariance function for models based on $L_s^{\alpha/2} = (\gamma_s^2 - \Delta)^{\alpha/2}$ (where $\alpha = \alpha_e + \alpha_s(\alpha_t - 1/2)$ in the DEMF
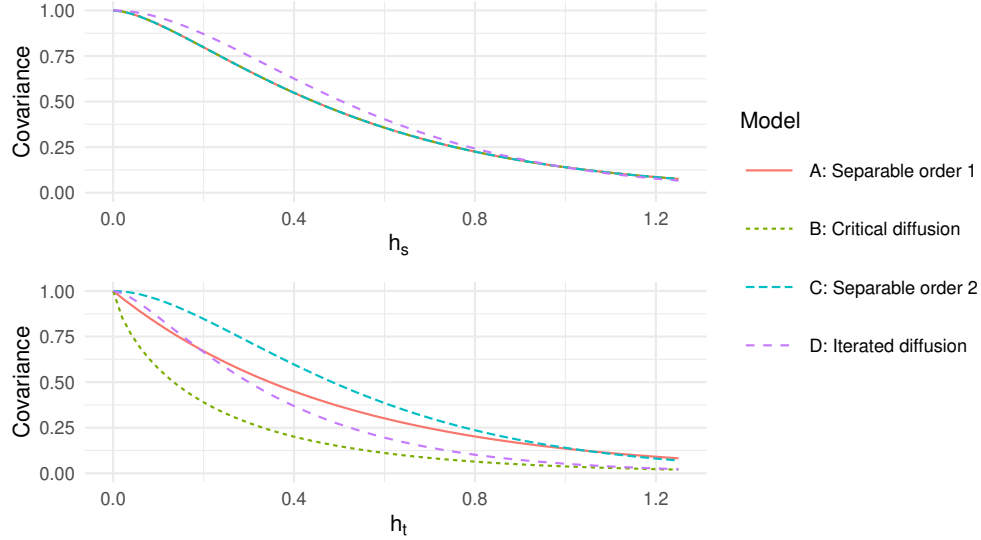
Figure 2. *The marginal spatial and temporal covariances for spatial dimension $d = 2$, for the four models in Table 2, Section 3.3. The spatial correlation is approximately 0.13 when the distance equals the range r_s. For the temporal correlations, that relationship to r_t only holds for the contribution from the evolution of a spatial constant, and the effective range has a more complex structure, depending on the combined model parameter.*

models) takes the form

$$R(\boldsymbol{s},\boldsymbol{s}') = \sum_{k=0}^{\infty} C_k \frac{1}{(\gamma_s^2 + \lambda_k^2)^{\alpha}} E_k(\boldsymbol{s}) E_k(\boldsymbol{s}'), \quad \boldsymbol{s} \in \mathcal{D},$$

where (λ_k, E_k) are the eigenvalue/function pairs of the $-\nabla \cdot \nabla$ (negated Laplace-Beltrami) operator on $\mathcal{D}$, and C_k are scaling constants that depend on potential scaling of the eigenfunctions and multiplicity of eigenvalues. This was used in Lindgren et al. (2011) to show that the finite element constructions for Whittle-Matérn fields work on general manifolds. On the sphere, the eigenfunctions are the spherical harmonics, with eigenvalues $\lambda_k = k(k+1)$ with multiplicity $2k+1$. With the spherical harmonic definitions in Appendix C, the resulting covariance can be simplified to

$$R_{\mathbb{S}^2,\alpha}(\boldsymbol{s},\boldsymbol{s}';\gamma_s) = \sum_{k=0}^{\infty} \frac{2k+1}{4\pi[\gamma_s^2 + k(k+1)]^{\alpha}} P_{k,0}(\boldsymbol{s} \cdot \boldsymbol{s}'), \tag{22}$$

where $P_{k,0}(\cdot)$ are Legendre polynomials of order k, and the factor $\frac{2k+1}{4\pi}$ comes from the eigenvalue multiplicity and Fourier-Bessel transform theory on the sphere (see Appendix C). It follows from the construction that the infinite series for the covariances of the process derivatives that the differentiability properties on the sphere are the same as on $\mathbb{R}^2$, as the terms $\lambda_k^a \frac{2k+1}{[\gamma_s^2 + \lambda_k]^{\alpha}}$ decay at the same rate as required for the smoothness criteria on $\mathbb{R}^2$ from Appendix A.

Due to the wraparound effects on the sphere, the spatial variance contribution to the overall field variance is not the same as on $\mathbb{R}^2$, and the factor $C_{\mathbb{R}^d,\alpha}/\gamma_s^{2\alpha-d}$ in (17) needs to be replaced by a function of γ_s defined by

$$C_{\mathbb{S}^2,\alpha}(\gamma_s) = \sum_{k=0}^{\infty} \frac{2k+1}{4\pi[\gamma_s^2 + k(k+1)]^\alpha}, \tag{23}$$

obtained from the spectral representation of a spherical Whittle-Matérn field. The overall variance can then be written as $\text{var}[u(\boldsymbol{s},t)] = \frac{C_{\mathbb{R},\alpha_t}}{\gamma_e^2 \gamma_t} C_{\mathbb{S}^2,\alpha}(\gamma_s)$, and the asymptotic behaviour of $C_{\mathbb{S}^2,\alpha}(\gamma_s)$ as γ_s approaches 0 or ∞ is given by

$$C_{\mathbb{S}^2,\alpha}(\gamma_s) = \sum_{k=0}^{\infty} \frac{2k+1}{4\pi[\gamma_s^2 + k(k+1)]^\alpha} \sim \begin{cases} \frac{1}{4\pi\gamma_s^{2\alpha}}, & \gamma_s \to 0, \\ \frac{1}{4\pi(\alpha-1)\gamma_s^{2\alpha-2}}, & \gamma_s \to \infty. \end{cases}$$

This shows that for large γ_s, i.e. short spatial ranges, the variance of the field $u(\boldsymbol{s},t)$ on the sphere is the same as on $\mathbb{R}^2$, but for small γ_s, i.e. long spatial ranges, the spherical geometry leads to larger variance than on $\mathbb{R}^2$. For intermediate γ_s values, the upper tail of the infinite series can be bounded by tractable integrals, which also allows bounding the relative error in numerical covariance and variance evaluation, by replacing the upper series tail from $k = K$ by the integral $\int_{K+1/2}^{\infty} \frac{2k+1}{4\pi[\gamma_s^2 + k(k+1)]^\alpha}\, dk$. More details are given in Appendix C.2.

4. Hilbert space representation

The discussion up to this point has focused on the general continuous domain properties of the proposed model class. We will now discuss aspects of numerical implementations, suitable for inclusion in generalised additive latent Gaussian models, as available in the `INLA` and `inlabru` packages for `R`. The general construction is applicable to a wide range of basis function representations. In practice, we will use the finite element approach from Lindgren et al. (2011) due to its computational convenience, in particular in the unstructured spatial observation location and manifold domain contexts.

4.1. Hilbert space approximation

We consider general Kronecker product basis expansions

$$u(\boldsymbol{s},t) = \sum_{i=1}^{n_s} \sum_{j=1}^{n_t} \psi_i(\boldsymbol{s})\phi_j(t)u_{ij}, \tag{24}$$

where $\{\psi_i(\boldsymbol{s}); i = 1,\ldots,n_s\}$ and $\{\phi_j(t); j = 1,\ldots,n_t\}$ are finite basis sets for Hilbert spaces on a spatial domain $\mathcal{D}$ and a time interval $[T_0, T_1] \subset \mathbb{R}$, respectively. We will show that projection onto the resulting Kronecker function space only involve integrals of the form $\langle \phi_j, \phi_{j'} \rangle$, $\langle (-\Delta)^{k/4}\phi_j, (-\Delta)^{k/4}\phi_{j'} \rangle$, and $\langle L_s^{k/2}\psi_i, L_s^{k/2}\psi_{i'} \rangle$. This is possible due

to the lack of interaction in the individual model operators; the operator as a whole is non-separable, but each operator term is space-time separable. This also extends to the case of a non-stationary L_s operator, as mentioned in Section 2.

Different choices of spatial and temporal basis functions have benefits and drawbacks depending on the specific modelling and data context. A natural choice for the spatial domain is local piecewise linear basis functions. Such functions were used in Lindgren et al. (2011) to construct model representations with sparse precision matrix structure for the basis expansion coefficients, via Gaussian Markov random fields (GMRF). This allows a large number of basis functions to be used, and pointwise georeferenced observations will not alter the sparseness of the posterior precision matrix, making this a versatile approach, that can also be used in combination with sparse matrix solvers developed for ordinary deterministic PDE computations. For very smooth processes, the piecewise linear basis functions can in principle be replaced by higher order local polynomials (Liu, Guillas and Lai, 2016), but this can be difficult to implement. For non-stationary $L_s = \gamma_s(\boldsymbol{s}) - \Delta$, the spatially varying $\gamma_s(\boldsymbol{s})$ values only have a local influence on the finite element construction, so the additional computational complexity lies mainly on the increased number of parameters needed to represent the spatial variation of $\gamma_s(\cdot)$.

An alternative to piecewise linear basis functions are harmonic basis functions based on the eigenfunctions of the Laplacian. These can be very efficient on domains that admit fast Fourier inversion algorithms, such as $\mathbb{R}^d$ and partially on $\mathbb{S}^2$. However, the diagonal precision matrix structure implied by the basic models is broken by scattered georeferenced observations, as the resulting posterior precision matrix becomes dense, so the utility is greatest for very smooth processes that can cut off the harmonics at a long spatial range. So-called *conditioning by kriging* can also be applied in such cases, but this is computationally expensive for large numbers of observations unless the number of basis functions is kept small. A further complication on general domains and manifolds is the lack of closed form expressions for the harmonics. Computing them with finite element methods, for example, is as expensive as applying the piecewise linear basis GMRF representations directly. They are also impractical for non-stationary operators, since the precision matrices will typically become dense instead of diagonal.

A third alternative is Karhunen-Loève expansions, which yield better approximations for fewer basis functions than harmonic basis. They can handle non-stationary operators, but need recomputing the basis for each set of parameter values, making inference expensive. For irregular data, the same problem exists of turning a sparse prior precision matrix into a dense posterior precision matrix. However, for given parameters, it can in principle be applied to the posterior distribution instead. Unfortunately, the numerical computations for each eigenfunction is at least as expensive as computing the posterior expectation using the same numerical method (e.g., finite elements) as in the GMRF computations, making the full computation much more expensive, and best suited to special cases such as computing a compact representation of a given, fixed, distribution.

Despite their practical numerical cost and other related problems, the harmonic basis and Karhunen-Loève e expansions are excellent tools for theoretical analysis, and their discrete domain formulations are essential in the theoretical proofs of the general discretisation construction below. See Lindgren et al. (2022) for further discussion on the relative merits of different basis choices.

The above considerations largely apply to the temporal basis function choice as well, with a few useful differences. First, in addition to piecewise linear basis functions, B-spline basis functions of higher order can readily be applied, and in particular second order B-splines (piecewise quadratic basis functions) provide immediate benefits with only minimal extra effort. Where piecewise linear basis functions require some form of mass lumping for operator order 2, second order B-splines can be applied with least squares finite element projection, and the resulting discretised Laplacian operator matrix has the same non-sparsity as for piecewise linear basis functions. In addition, when applied to order 1 operators, temporal interpolation in the finite-dimensional representation exhibits less quasi-deterministic fluctuations than for piecewise linear basis functions. Second, harmonic basis functions are useful for smooth cyclic processes, e.g. seasonal effects, but otherwise suffer from the same issues as in space.

4.2. Precision matrix construction

In this section we represent the stochastic processes $\text{DEMF}(\alpha_t, \alpha_s, \alpha_e)$ using general Kronecker basis Hilbert space representations. Define $u(s,t)$ on $\mathcal{D} \times \mathbb{R}$, for some polygonal domain $\Omega \subset \mathbb{R}^d$, as the solution to (13) with some boundary conditions on $\partial \mathcal{D}$. The particular choice of boundary conditions does not matter much in what follows as long as they lead to a well defined precision operator for the solutions of the equation posed on the bounded domain. However, in most practical situations one would use homogeneous Neumann boundary conditions on the spatial domain.

For implementations, we restrict the temporal domain to an interval, and we then also need to impose temporal boundary conditions. However, temporal boundary effects can be handled by direct calculations for the resulting AR(2) dependence structure for the temporal coefficients in the approximation; see Appendix E.

The projection of the solutions onto the finite Hilbert space result in a discretised model where the coefficients u_{ij} in (24) have a precision matrix that is expressed as a sum of Kronecker products. As in Lindgren et al. (2011), the approximation properties of the discretisation is directly linked to the expressiveness of the finite-dimensional Hilbert space spanned by the Kronecker basis $\{\psi_i(\boldsymbol{s})\phi_j(t), i = 1, \ldots, n_s, j = 1, \ldots, n_t\}$.

We provide the following theorem that links the continuous domain DEMF models to finite-dimensional Hilbert space representations. The theorem focuses on the link between the continuous domain precision operator and the precision matrix, necessarily assuming unique solutions with a unique covariance function. This makes it applicable, in principle, to more esoteric models involving various forms of intrinsic stationarity, i.e. non-stationary models with stationary properties with respect to some contrast filters. However, the details of such models are beyond the scope of the presentation.

Theorem 4.1. *Let $\alpha_t \in \mathbb{N}$ and consider the equation*

$$\left(-\gamma_t^2 \frac{\partial^2}{\partial t^2} + L_s^{\alpha_s}\right)^{\alpha_t/2} u(\boldsymbol{s},t) = \mathrm{d}\mathcal{E}_{\gamma_e^2 L_s^{\alpha_e}}(\boldsymbol{s},t) \quad \text{on } \mathcal{D} \times [T_0, T_1], \tag{25}$$

where $[T_0, T_1] \subset \mathbb{R}$ is a bounded interval, L_s is some spatial differential operator, and some boundary conditions on $\partial \mathcal{D}$ and at T_0 and T_1 are assumed such that the precision operator for the solutions of (25) is well defined. Let $\{\psi_i(\boldsymbol{s}), i = 1, \ldots, n_s\}$ and $\{\phi_j(t), j = 1, \ldots, n_t\}$ be bases for finite-dimensional Hilbert spaces on $\mathcal{D}$ and $[T_0, T_1]$, respectively, chosen such that the product basis set $\{\psi_i(\boldsymbol{s})\phi_j(t), i = 1, \ldots, n_s, j = 1, \ldots, n_t\}$ form a basis for a finite-dimensional Hilbert space $V_h \subset V$, and let $u(\boldsymbol{s},t) = \sum_{i,j} \psi_i(\boldsymbol{s})\phi_j(t)$ $u_{i,j} \in V_h$ be a finite-dimensional representation of a solution to (25). Assume the following two conditions:

(i) Let $v(t) = \sum_{j=1}^{n_t} \phi_j(t)v_j$ be a finite-dimensional approximation of a solution to

$$b^{1/2}\left(-\frac{\partial^2}{\partial t^2} + \kappa^2\right)^{\alpha_t/2} v(t) = \mathcal{W}(t), \quad \text{on } [T_0, T_1],$$

for some $b > 0$, $\kappa > 0$, and $\alpha_t = 1, 2, \ldots$, and the boundary conditions at T_0 and T_1. Assume that the precision matrix for the weights vector $\boldsymbol{v} = (v_1, \ldots, v_{n_t})$ takes the form

$$b \sum_{k=0}^{2\alpha_t} \kappa^{2\alpha_t - k} \boldsymbol{J}_{\alpha_t, k/2}$$

for some symmetric matrices $\boldsymbol{J}_{\alpha_t, 0}$, $\boldsymbol{J}_{\alpha_t, 1/2}$, to $\boldsymbol{J}_{\alpha_t, \alpha_t}$.

(ii) Let $w(\boldsymbol{s}) = \sum_{i=1}^{n_s} \psi_i(\boldsymbol{s})w_i$ be a finite-dimensional approximation of a solution to

$$L_s^{a/2} w(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}) \quad \text{on } \mathcal{D},$$

where L_s is equipped with the boundary conditions on $\partial \mathcal{D}$, for some $a \geq 0$. Assume that the precision matrix for $\boldsymbol{w} = (w_1, \ldots, w_{n_s})$ is $\boldsymbol{K}_a = \boldsymbol{C}^{1/2}\left(\boldsymbol{C}^{-1/2}\boldsymbol{K}_1 \boldsymbol{C}^{-1/2}\right)^a \boldsymbol{C}^{1/2}$ for some symmetric positive definite matrix $\boldsymbol{K}_1$.

Assume additionally that the temporal precision construction in condition (i) is valid for all $\kappa \geq \lambda_0^{\alpha_s/2}/\gamma_t$, where λ_0 is the smallest eigenvalue in the generalised eigenvalue problem $\boldsymbol{K}_1\boldsymbol{e} = \boldsymbol{C}\boldsymbol{e}\lambda$. Then, the precision matrix for the collected coefficient vector $\boldsymbol{u} = (u_{1,1}, u_{2,1}, \ldots)$ is given by

$$\boldsymbol{Q_u} = \gamma_e^2 \sum_{k=0}^{2\alpha_t} \gamma_t^k \boldsymbol{J}_{\alpha_t, k/2} \otimes \boldsymbol{K}_{\alpha_s(\alpha_t - k/2) + \alpha_e}.$$

Proof. The result follows from discretising the spatial dimension, diagonalising the resulting operator matrices, and applying the temporal precision structure condition to the resulting independent temporal equations. A detailed proof is given in Appendix D.3.

∎

The existence of finite-dimensional representations fulfilling conditions (i) and (ii) for certain choices of basis functions follows directly from the general constructions in Lindgren et al. (2011).

For the regular Whittle-Matérn operator $L_s = \gamma_s^2 - \Delta$ on $\mathcal{D}$ we have $\boldsymbol{K}_1 = \gamma_s^2 \boldsymbol{C} + \boldsymbol{G}$. For triangulated domains with local piecewise linear basis functions with $\sum_{i=1}^{n_s} \psi_i(\boldsymbol{s}) \equiv 1$ on $\mathcal{D}$, we can take $\boldsymbol{C}$ to be the diagonal mass lumped mass matrix with $C_{i,i} = \langle \psi_i, 1 \rangle$ and symmetric sparse structure matrix $\boldsymbol{G}$ with $G_{i,j} = \langle \nabla \psi_i, \nabla \psi_j \rangle$. For domains where the orthogonal harmonic eigenfunctions of Δ are available, such as rectangular subdomains of $\mathbb{R}^d$ and spherical harmonics on $\mathbb{S}^2$, the full mass and structure matrices $\boldsymbol{C}$ and $\boldsymbol{G}$ are both diagonal, with $C_{i,i} = \langle \psi_i, \psi_i \rangle$.

In the temporal case, the same technique applies, but higher order B-spline basis functions are more easily applied, allowing, e.g., second order B-splines to be used without mass lumping. For temporal Neumann boundary conditions, $\boldsymbol{J}_{\alpha_t, k/2} = \boldsymbol{0}$ for odd $k = 1, 3, \ldots, 2\alpha_t - 1$ and $[\boldsymbol{J}_{\alpha_t, k/2}]_{i,j} = \langle (-\Delta)^{k/4} \phi_i, (-\Delta)^{k/4} \phi_j \rangle$ (or non-conformal approximations for non-smooth basis functions) for even $k = 0, 2, \ldots, 2\alpha_t$. Lemma E.1 in Appendix E can be used for first and second order B-spline basis functions for $\alpha_t = 1$ and 2 to provide approximate stationary boundary conditions by modifying the $\boldsymbol{J}_{\alpha_t, k/2}$ matrices for $k = 0, 1, \ldots, 2\alpha_t$. When such temporal boundary corrections are used, fractional orders appear in $\boldsymbol{K}_{\alpha_s(\alpha_t - k/2) + \alpha_e}$ for odd k unless α_s is an even integer. For the spatial piecewise linear finite element constructions, this would break sparsity, but for orthogonal harmonic function representations, $\boldsymbol{K}_a$ is diagonal for all $a \geq 0$, allowing the fractional powers to be used without loss of the diagonal property.

In the proof of Theorem 4.1, we see that it is sufficient that the initial temporal precision structure is valid for $\kappa \geq \kappa_0 = \gamma_s^{\alpha_s/2} / \gamma_t$. By taking a Taylor expansion for the boundary precision elements with respect to κ and κ_0, the approximation would be improved, compared with taking the Taylor expansion at $\kappa = 0$, as the expansion would be closer to the exact expression for a wider range of relevant temporal frequencies. This improvement would however come at the expense of making the matrix constructions dependent on the γ_s and γ_t parameters directly.

5. Applications

5.1. Separable vs non-separable forecasting

The difference between using separable and the non-separable models is most clearly seen when doing forecasting. To illustrate this, we simulated spatial data for time $t = 0$, and compute the posterior conditional expectation for $t = 0$, 1, and 2. For the simulation,

we used the four Matérn models defined in Table 2 with one percent (standard deviation) nugget effect added. The parameters were set to $r_s = 4.0$, $r_t = 2.5$ for the separable models and $r_t = 4.5$ for the non-separable models, and $\sigma = 1$. The scaling difference for r_t in the non-separable models compensates for the difference in parameter interpretation illustrated in Section 3.3. In the estimation, the nugget precision and the temporal range parameters r_t were kept fixed, so that only the marginal standard deviation σ and the spatial range parameter r_s were estimated for each model.
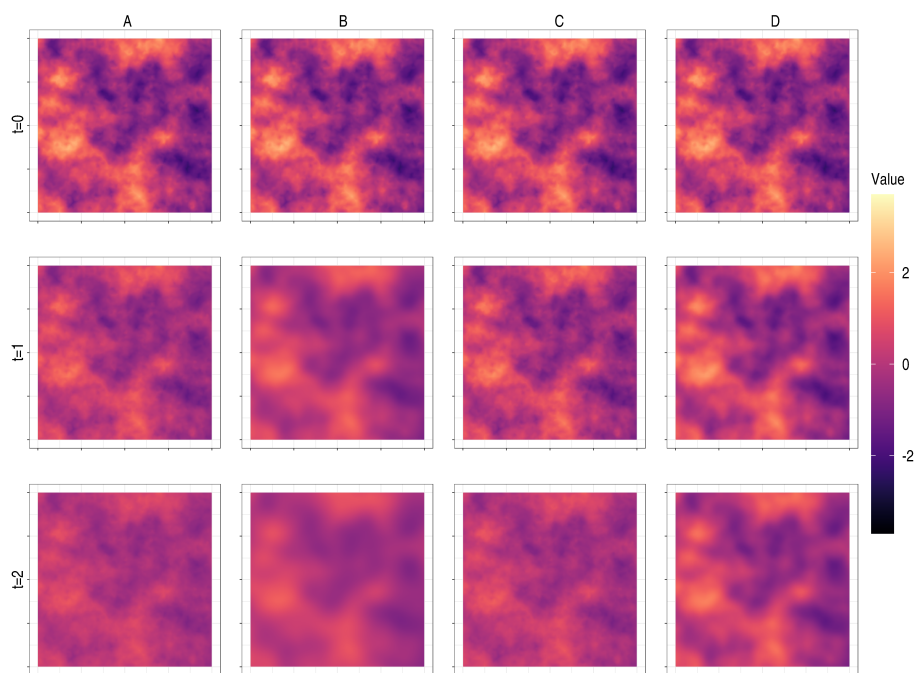
Figure 3. *Predictions from each model (A, B, C and D) when conditioned on a spatially dense dataset at t = 0, and no observations for t = 1 and t = 2.*

Figure 3 displays the predictions from the four models in Table 2. For $t = 0$, the results are similar for the four models, due to the highly informative data. For the predictions for $t = 1$ and $t = 2$, we see how the separable model A and C only reduce the fields point-wise towards zero, and that non-separable models B and D, exhibit spatial diffusion, as expected. This behaviour was part of the theoretical motivation of Whittle (1954, 1963), and also a major motivation for developing the DEMF family. It is also noteworthy that since the forecasts are conditional expectations based on a finite set of observations, they are smoother than the process realisations. For the separable models, this effect is not visible, since there this effect only appears on smaller spatial scales than shown, but it is clearly visible for the non-separable model. In all four cases, the posterior process realisations however have their ordinary, lower, smoothness. This is important to take into account when considering probabilistic forecasts, in particular for prediction of non-linear functionals of the process.
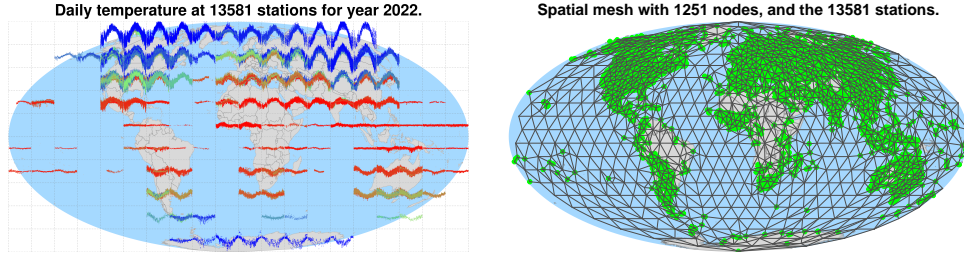
Figure 4. *Daily average temperature time series shown grouped near the corresponding locations (left), with colours based on the year average, for each station, from blue (cold), to red (warm). The locations (green) and the mesh used for the spatio-temporal model components v and u (right). The computational mesh is defined by spherical triangles directly on the globe surface, and triangles crossing the $\pm 180°$ longitude curve are not shown in this Mollweide projection.*

5.2. Global temperature dataset

This section presents some results analysing daily temperature data, where all the code for the data cleaning, model fitting and plots are included in the supplementary material.

5.2.1. Data and model structure

We used daily data for year 2022, using minimum (TMIN) and maximum (TMAX) daily temperatures, as described in Menne et al. (2012). We cleaned the data for inconsistencies before the analysis. In particular, values beyond 7 standard deviations from the mean were treated as missing. We computed the mean of these two variables for each day at each one of 13 567 stations world-wide, a total of 4 951 955 data entries. Figure 4 (top) shows this data as time series grouped by location.

The model includes an overall level μ, the elevation in kilometres $E(\boldsymbol{s})$, a smoothed deviation from the overall mean jointly over latitude and time $b(\boldsymbol{s},t)$, a spatio-temporal random field $v(\boldsymbol{s},t)$ varying slowly in time, and a spatio-temporal random field, $u(\boldsymbol{s},t)$, capturing the daily variability. The $b(\boldsymbol{s},t)$ function is allowed to vary by latitude and time, but is fixed to zero at the equator. The linear predictor expression is

$$\eta(\boldsymbol{s},t) = \mu + \alpha E(\boldsymbol{s}) + b(\boldsymbol{s},t) + v(\boldsymbol{s},t) + u(\boldsymbol{s},t). \tag{26}$$

Each observation y_i is modelled with additive Gaussian noise with a common variance parameter, σ_e^2, so that $y_i = \eta(\boldsymbol{s}_i,t_i) + e_i$, where $(\boldsymbol{s}_i,t_i)$ is observation i, $i = 1,\dots n$, and $e_i \sim \mathrm{N}(0,\sigma_e^2)$.

5.2.2. Model discretisation and estimation

For the $b(\boldsymbol{s},t)$ and $v(\boldsymbol{s},t)$ functions in the predictor expression (26), we defined temporal basis functions 1, $\cos[(t-1)\cdot 2\pi/365]$, and $\sin[(t-1)\cdot 2\pi/365]$. For $b(\boldsymbol{s},t)$, these were

multiplied with two quadratic basis function in $\sin(\text{latitude} \cdot \pi/180)$, which guarantees smooth behaviour with respect to the location, s, at the two poles, giving a total of six basis functions. For $v(s,t)$, each of the three temporal basis functions were instead multiplied by stationary spatial Whittle-Matérn fields over the sphere forming a model term that captures the seasonal local deviation from the basic seasonal pattern described by $b(s,t)$.

The reported results were estimated using a spatial mesh with $1\,251$ nodes (median node distance ~ 587km), shown in Figure 4, both for the spatial coefficients in v and for u. For u, we discretised the time domain with first order basis functions with one knot per day. This setting gives a spatio-temporal model for u with size $456\,615$. In vector form we have

$$ y = 1\mu + E\alpha + Bb + A_v v + A_u u + e, $$

where B is a six-column matrix of the evaluated basis functions for $b(s,t)$ at the observation locations and times, and A_v and A_u contains the evaluated basis functions, respectively, for $v(s,t)$ and $u(s,t)$. The vectors b, v, and u contain the corresponding basis weights.

We used independent priors for all the model parameters. We used a flat prior for μ and a Gaussian with mean zero and variance 100 for α and each element in b. The three spatial fields in v are assumed as independent realizations each one modelled using Eq. 3 with a common spatial range r_v, and common marginal variance σ_v^2. The $u(s,t)$ term is a spatio-temporal field using one of the four models in Table 2. In total, we have six variance/range parameters to estimate. We used penalized complexity priors for all these parameters (Simpson et al., 2017; Fuglstad et al., 2018), applied to the marginal properties of the models.

To define the PC-prior for σ_e we used $\Pr(\sigma_e \geq 5) = 0.01$ and the same for σ_v and σ. We used $\Pr(r_v \leq 1000\text{km}) = 0.01$ for r_v, $\Pr(r_s \leq 600\text{km}) = 0.01$ for r_s. For r_t we used $\Pr(r_t \leq 1 \text{ days}) = 0.01$ in models A and C and $\Pr(r_t \leq 2 \text{ days}) = 0.01$ in models B and D.

5.2.3. Model fitting results

Attributing the relative contributions to each model component is non-trivial due to the posterior correlation between the components. However, a basic linear model variance decomposition, $\text{SQT} = \sum_i (y_i - \bar{y})^2$ and $\text{SQR} = \sum_i (y_i - \text{E}(\eta_i \mid y))^2$, can be obtained to define $R^2 = 1 - \text{SQR}/\text{SQT}$. We have that the predictor model $\eta(s,t)$ captures 97.18% of the variability with model B. Table 3 reports DIC, WAIC, and goodness-of-fit statistics for within-sample and leave-one-out assessment (leave-one-out log predictive density score, LCPO, see Held, Schrodle and Rue (2010), for each of the five fitted models. For within-sample assessment, R^2, mean squared error (MSE), and mean absolute error (MAE) assess the posterior mean and median only, whereas the log predictive density score (LPO), CRPS, and SCRPS assess the full predictive distribution (Gneiting et al., 2005; Bolin and Wallin, 2023). The model M_0 includes the fixed effects and $v(s,t)$,

whereas models A, B, C and D all include $\boldsymbol{u}$, using the four models in Table 2. When considering R^2, LPO, MSE, MAE, CRPS and SCRPS model B performed a slightly better. When considering DIC, WAIC and LCPO model C was slightly better.

Table 3. *Summary statistics for each estimated model. The LCPO score is the average negated log-predictive density for leave-one-out predictions, and LOP is its within-sample version. The MAE, MSE, CRSP, SCRPS were all computed as within-sample scores based on a Gaussian approximation of the posterior predictive distribution for each data point.*

Model	M_0	A	B	C	D
R^2	0.8649	0.9718	**0.9718**	0.9718	0.9718
DIC	5.8103	4.3217	4.3222	**4.3212**	4.3222
WAIC	5.8101	4.3123	4.3128	**4.3118**	4.3128
LPO	2.9046	2.1332	**2.1330**	2.1335	2.1331
LCPO	2.9051	2.1572	2.1574	**2.1569**	2.1574
MSE	19.5151	4.1163	**4.1136**	4.1193	4.1155
MAE	3.3236	1.4667	**1.4657**	1.4678	1.4661
CRPS	2.4288	1.0947	**1.0943**	1.0952	1.0945
SCRPS	1.7905	1.3947	**1.3945**	1.3949	1.3946

Table 4. *The posterior mean and standard deviation (in brackets) for each of the model parameters.*

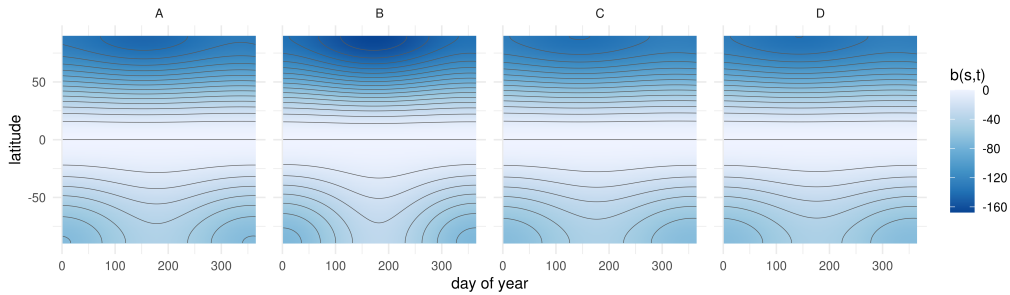|	A	B	C	D
σ_e	2.06 (0.001)	2.06 (0.001)	2.06 (0.001)	2.06 (0.001)
r_v	5402 (217)	15304 (10683)	5242 (266)	5346 (486)
σ_v	8.5 (0.4)	20.9 (14.0)	7.9 (0.4)	8.3 (0.8)
r_s	1322 (6)	2079 (19)	1329 (7)	1365 (6)
r_t	5.60 (0.04)	42.07 (0.91)	3.92 (0.02)	6.94 (0.05)
σ	2.73 (0.01)	3.67 (0.03)	2.63 (0.01)	2.80 (0.01)

Figure 5. *The posterior mean of the smoothed seasonal latitude effect $b(\boldsymbol{s}, t)$ for each model.*
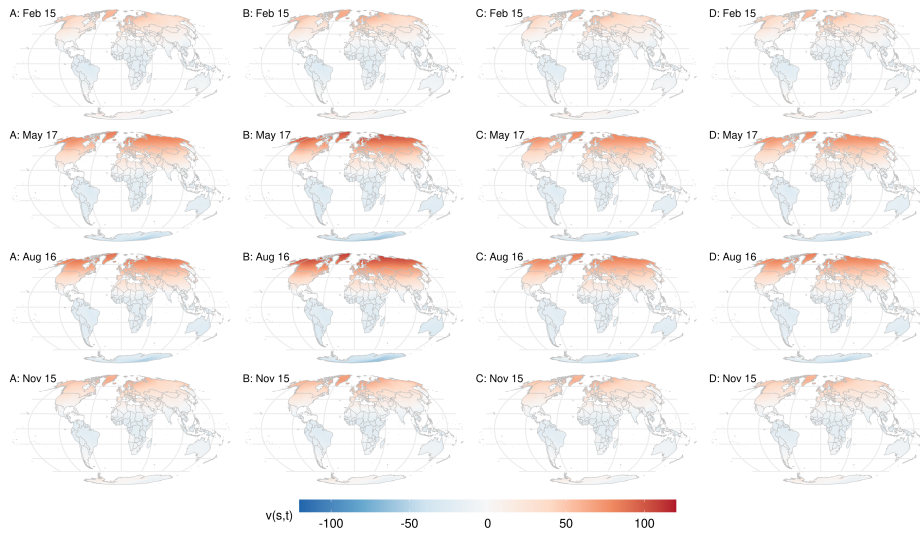
Figure 6. *The posterior mean for $v(\boldsymbol{s},t)$, at some time points, for each model. From left to right, models A, B, C, and D. From top to bottom, the time points are 46, 137, 228, and 319, corresponding to the day of year in 2022 as labelled at the top left of each plot.*
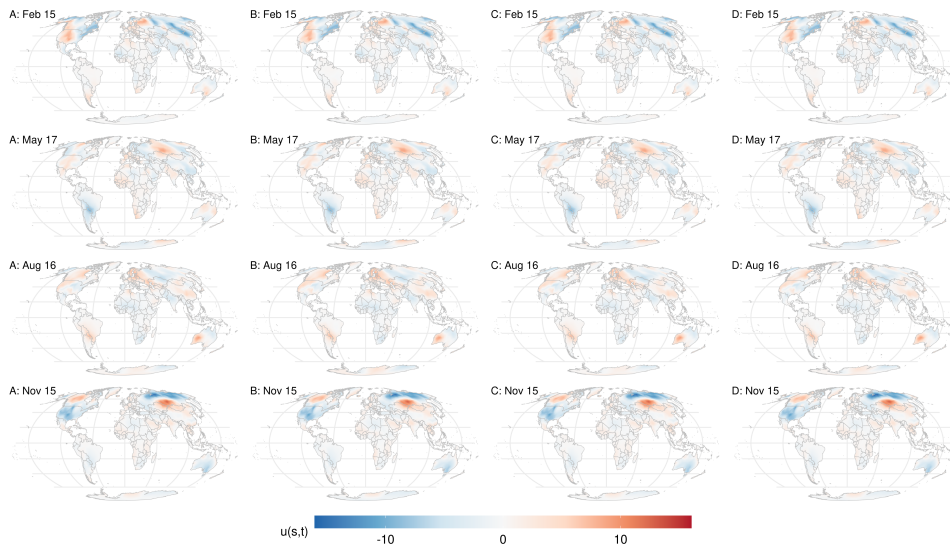

Figure 7. *The posterior mean for $u(\boldsymbol{s},t)$, at some time points, for each model. From left to right, models A, B, C, and D. From top to bottom, the time points are 46, 137, 228, and 319, corresponding to the day of year in 2022 as labelled at the top left of each plot.*

For B (using model B for $u(\boldsymbol{s},t)$), the posterior mean for μ is 34.74 and for α is -4.70. The posterior mean of $b(\boldsymbol{s},t)$ for all four models are shown in Figure 5, which displays the temperature over time and latitude. The seasonal pattern is clear, with summer and winter temperatures in the two hemispheres standing out, with in particular lower temperatures (blue) in each hemisphere's respective winter. The model range/variance parameter estimates are summarised in Table 4. For the posterior mean, we have $\mathrm{E}(r_s \mid \boldsymbol{y}) = 2078.66$ km and $\mathrm{E}(r_t \mid \boldsymbol{y}) = 42.07$ days for model B, and smaller for the other models. In particular the temporal range estimates are much smaller for the other models, ranging from 4 to 7 days, which is more realistic. These values can be interpreted through Figure 2. The posterior mean for the spatio-temporal field $v(\boldsymbol{s},t)$ for some days in 2022 is shown in Figure 6. This term is intended to capture slowly temporally varying spatial variation from the overall mean, elevation effect and the basic seasonal latitude parts of the model. It is clear that the desired interpretation is confounded with the interpretation of $b(\boldsymbol{s},t)$, shown in Figure 5. The posterior mean for the spatio-temporal field $u(\boldsymbol{s},t)$ for some days in 2022 is shown in Figure 7. This term captures the remaining spatio-temporal variation of the temperature field around the other parts of the model.

5.2.4. Forecast evaluation

As was already apparent from the diagnostic scores in Table 3, despite the temporal range parameters being different for the four models, particularly for model B, they are nearly indistinguishable with respect to direct and leave-one-out prediction distributions. Since the space-time non-separability effect is unclear in the leave-one-out setting, we extend the assessment by computing multi-horizon temporal predictions. We used the first 14 days of the data from each month to predict the following 7 days. These forecasts were done while keeping the covariance parameters and the long term spatio-temporal components $b(\boldsymbol{s},t)$ and $v(\boldsymbol{s},t)$ fixed to their posterior modes from the full joint model estimates, so that only the short-term spatio-temporal field $u(\boldsymbol{s},t)$ was reestimated for each scenario. This generated forecasts for each model for 12 different weather and seasonal conditions over the year.

Figure 8(top) shows the mean absolute error (MAE), mean squared error (MSE), mean Dawid-Sebastiani (DS, equivalent to log-score for Gaussian predictions, see Gneiting et al., 2005), and mean SCRPS summarized for each prediction horizon (1–7 days) for each of the 12 scenarios. Figure 8(bottom) shows the difference between the scores for each model to those of model D, to more clearly highlight the differences between the models. The prediction errors all exhibit increasing variability for longer forecast horizons, as well as a generally increasing trend, that mostly levels off around 6 days, which is compatible with the estimated temporal correlation length parameter r_t for models A, C, and D. For 1-day ahead forecasts, model D achieved the lowest scores, and it appears more stable than the other models for longer forecast horizons. Model B has large score variability, and is doing worse than the other three models for long forecast horizons, in particular for the scores that take forecast uncertainty into account. For more details see
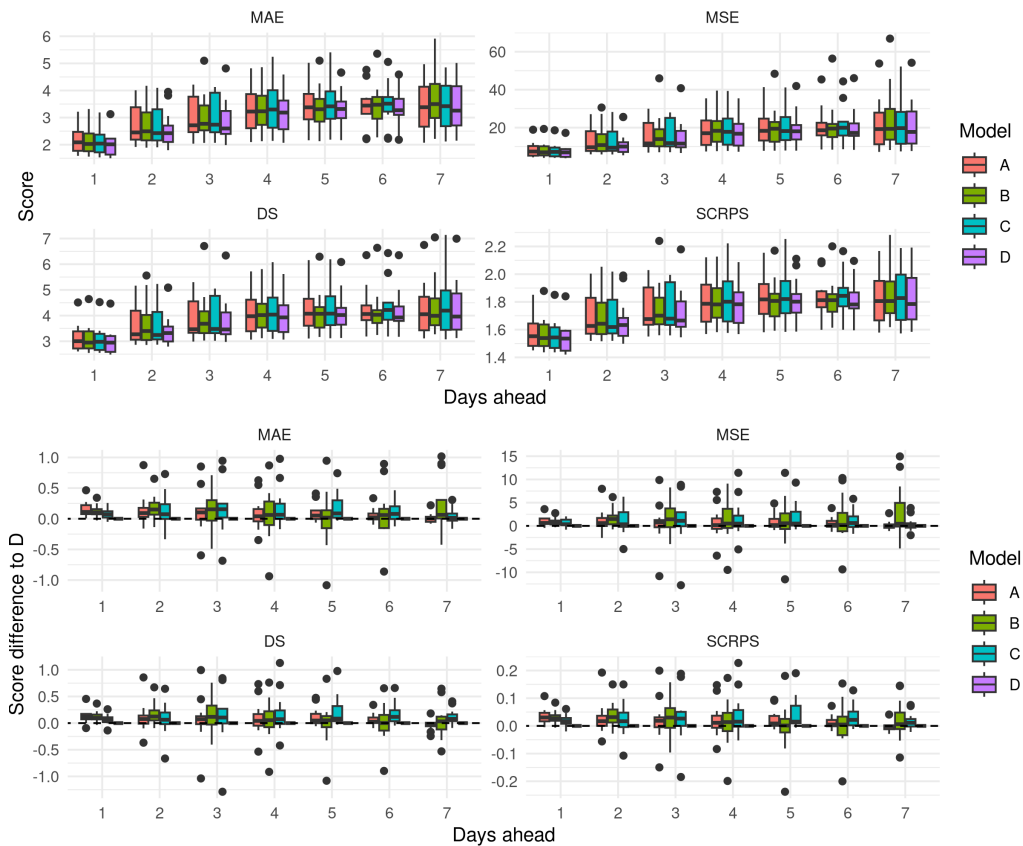
Figure 8. *Top: Multi-horizon (1–7 days) and multi scenario (one for each month of year 2022) forecast scores for predicting one week ahead. Lower scores indicate a better forecast. Bottom: The differences in scores compared with model D.*

Appendix F, where one can see that the scores are generally worse in the start and end of the year, indicating an unmodelled aspect of seasonality, e.g. in weather variability.

6. Discussion

We have developed a spatio-temporal extension of the Gaussian Matérn fields based on a fractional and stochastic version of the physical diffusion equation considered by Whittle (1954, 1963). We named the new family the diffusion-based extension of the Matérn field (DEMF), and showed that it has several useful properties: The spatial marginals are Gaussian Matérn fields; the family contains Markovian diffusion processes with clear physical interpretations; and we can control the smoothness in space and in time, the degree of non-separability, and interpret all the parameters. The family can also be extended to non-stationary models and be defined on curved manifolds.

The DEMF family contains several important subfamilies; 1) separable models, 2) Markov models, 3) partially separable models, 4) a fully non-separable subfamily of the Stein (2005) family, and 5) spatially non-stationary model dynamics. This provides a rich outset for studying the practical and methodological impacts these assumptions have.

An important special case in the DEMF family is the DEMF(1,2,1) model, which in two-dimensional space, is the closest stochastic process analogue to the diffusion equation (see (4)), and hence a natural default choice for spatio-temporal model components. The non-separable DEMF(1,2,1) model has the same smoothness in space and in time as the separable DEMF(1,0,2) model, which has a covariance function that is a Kronecker product of a Matérn covariance in space and an exponential covariance in time. Of particular interest is also the non-separable DEMF(2,2,0) model which can be viewed as an iterated diffusion model.

Although the proposed model family includes non-separable models, which in itself might be desirable from considerations about covariance properties, another view-point is that the non-separability here arises as a direct and natural consequence of the physics-inspired dynamical diffusion construction. Most importantly, the results shed light on which types of non-separability would occur naturally under certain assumptions on the spatio-temporal dynamics and properties of the driving noise process. Although there are strong arguments in the literature against using a separable model, the space of non-separable models is vastly larger than the space of separable models. Hence we need to consider which types of non-separable models are more, and which are less, appropriate than the separable alternatives. As illustrated by the practical example in Section 5.2, it is important to assess models in a context relevant to the intended use case. In particular, non-separability is unlikely to make a difference for space-time interpolation, as assessed by e.g. leave-one-out cross-validation, but can make a difference in full space-time forecasting settings.

It is natural to view the model class as an example of building models via building blocks with precision operator space-time separability. The most basic form of separability is *functional separability*, where a spatial and temporal processes are added or multiplied, which can be viewed as having $S + T$ degrees of freedom, where S and T are the spatial and temporal effective dimensions of the functions. The next form is *covariance separability*, where the model is formed from a sum of covariances (giving the same as functional separability) or a product of covariances, where the latter gives $S \cdot T$ degrees of freedom. These covariance product models are covariance separable but functionally non-separable. For precision models, plain products are equivalent to covariance separable models, but sums of precision products give covariance non-separability. In both the covariance and precision cases, non-stationarity in the spatial and temporal operators can be introduced, as long as the operator separability is kept. This distinguishes this type of non-separability from fully non-separable non-stationary models that cannot be written as precision sums and products. The key is to

retain commutativity between the spatial and temporal operators within each product: $(\boldsymbol{Q}_t \otimes \boldsymbol{I}_s)(\boldsymbol{I}_t \otimes \boldsymbol{Q}_s) = (\boldsymbol{I}_t \otimes \boldsymbol{Q}_s)(\boldsymbol{Q}_t \otimes \boldsymbol{I}_s) = \boldsymbol{Q}_t \otimes \boldsymbol{Q}_s$.

With the GMRF representation presented herein, the computational costs of the separable and non-separable models are similar, as the sparsity structure of posterior precisions, given irregularly spaced observations in generalised latent Gaussian models, is only marginally affected by the non-separability, and can even be more sparse in the non-separable cases; the separable precision neighbourhood structures are space-time prisms, whereas the non-separable neighbourhood structures are double-cones. Together with interpretable parameters, this makes the non-separable models as practically accessible as the separable models. In the supplementary materials we provide an implementation with examples in `R-INLA`.

In this paper we mainly focused on stationary fields, but also showed how very little in the theory and computational construction changes for models with curved manifolds or spatially non-stationary operators, as already discussed by Lindgren, Rue and Lindström (2011). Although the initial practical implementation only covers a subset of the general model class, we believe that the general results can and will be applied in more general contexts in the future.

Supplementary Material

R code for the examples: The examples were computed with the `INLAspacetime` package, using the `cgeneric` method from the `R-INLA` software for computationally efficiency, via the `inlabru` interface Bachl et al. (2019). See also Niekerk et al. (2021) for a similar example. Code for the figures and examples is available at https://github.com/finnlindgren/spacetime-paper-code, and the `INLAspacetime` R package (https://github.com/eliaskrainski/INLAspacetime) implements a subset of the models.

References

Bachl, F. E., Lindgren, F., Borchers, D. L. and Illian, J. B. (2019). inlabru: an R package for Bayesian spatial modelling from ecological survey data. *Methods in Ecology and Evolution*, 10, 760-766.

Bakka, H., Rue, H., Fuglstad, G.-A., Riebler, A., Bolin, D., Illian, J., Krainski, E., Simpson, D. and Lindgren, F. (2018). Spatial modeling with R-INLA: A review. *Wiley Interdisciplinary Reviews: Computational Statistics*, 10, e1443.

Bakka, H., Vanhatalo, J., Illian, J. B., Simpson, D. and Rue, H. (2019). Non-stationary Gaussian models with physical barriers. *Spatial Statistics*, 29, 268-288.

Bhatt, S., Weiss, D. J., Cameron, E., Bisanzio, D., Mappin, B., Dalrymple, U., Battle, K. E., Moyes, C. L., Henry, A., Eckhoff, P. A., Wenger, E. A., Briët, O., Penny, M. A., Smith, T. A., Bennett, A., Yukich, J., Eisele, T. P., Griffin, J. T., Fergus, C. A., Lynch, M., Lindgren, F., Cohen, J. M., Murray, C. L. J., Smith, D. L., Hay, S. I., Cibul-

skis, R. E. and Gething, P. W. (2015). The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nature*, 526, 207-211.

Bissiri, P. G., Holmes, C. C. and Walker, S. G. (2016). A general framework for updating belief distributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78, 1103-1130.

Bolin, D. (2014). Spatial Matérn fields driven by non-Gaussian noise. *Scandinavian Journal of Statistics*, 41, 557-579.

Bolin, D. and Kirchner, K. (2020). The Rational SPDE Approach for Gaussian Random Fields With General Smoothness. *Journal of Computational and Graphical Statistics*, 29, 274-285. Publisher: Taylor & Francis.

Bolin, D., Simas, A. B. and Xiong, Z. (2024). Covariance-based rational approximations of fractional SPDEs for computationally efficient Bayesian inference. *Journal of Computational and Graphical Statistics*, 33, 64-74.

Bolin, D. and Wallin, J. (2023). Local scale invariance and robustness of proper scoring rules. *Statistical Science*, 38, 140-159.

Cameletti, M., F. Lindgren, D. Simpson, and H. Rue (2013). Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *AStA Advances in Statistical Analysis*, 97, 109-131.

Clarotto, L., Allard, D., Romary, T. and Desassis, N. (2022). The SPDE approach for spatio-temporal datasets with advection and diffusion. *arXiv 2208.14015*.

Cramér, H. and Leadbetter, M. R. (1967). *Stationary and related stochastic processes: Sample function properties and their applications*. Wiley. Also available as a Dover reprint (2004).

Cressie, N. and Huang, H.-C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, 94, 1330-1339.

Cressie, N. and Wikle, C. K. (2011). *Statistics for spatio-temporal data*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ.

Da Prato, G. and J. Zabczyk (2014). *Stochastic equations in infinite dimensions*. Cambridge university press.

Erdélyi, A. (1953). Higher transcendental functions. *Higher Transcendental Functions, Edited by A. Erdelyi. Vol. I*, p. 59. McGraw-Hill.

Fonseca, T. C. and Steel, M. F. (2011). A general class of nonseparable space–time covariance models. *Environmetrics*, 22, 224-242.

Fuentes, M. (2002). Spectral methods for nonstationary spatial processes. *Biometrika*, 89, 197-210.

Fuentes, M., Chen, L. and Davis, J. M. (2008). A class of nonseparable and nonstationary spatial temporal covariance functions. *Environmetrics: The official journal of the International Environmetrics Society*, 19, 487-507.

Fuglstad, G.-A. and Castruccio, S. (2020). Compression of climate simulations with a nonstationary global spatiotemporal SPDE model. *Annals of Applied Statistics*, 14, 542-559.

Fuglstad, G. A., Simpson, D., Lindgren, F. and Rue, H. (2018). Constructing priors that penalize the complexity of Gaussian random fields. *Journal of the American Statistical Association*, 0, 1-8.

Gaedke-Merzhäuser, L., van Niekerk, J., Schenk, O. and Rue, H. (2022). Parallelized integrated nested laplace approximations for fast bayesian inference. *Statistics and Computing*, 33.

Gelfand, A., Diggle, P., Guttorp, P. and Fuentes, M. (2010). *Handbook of Spatial Statistics*. CRC Press.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space–time data. *Journal of the American Statistical Association*, 97, 590-600.

Gneiting, T., Raftery, A., Westveld III, A. and Goldman, T. (2005). Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Monthly Weather Review*, 133, 1098-1118.

Guttorp, P. and Gneiting, T. (2006). Studies in the history of probability and statistics XLIX On the Matérn correlation family. *Biometrika*, 93, 989-995.

Handcock, M. S. and Stein, M. L. (1993). A Bayesian analysis of kriging. *Technometrics*, 35, 403-410.

Held, L., Schrodle, B. and Rue, H. v. (2010). Posterior and Cross-validatory Predictive Checks: A Comparison of MCMC and INLA. In *Statistical Modelling and Regression Structures*, pp. 111–131. Springer.

Higdon, D. (2002). Space and space-time modeling using process convolutions. In C. W. Anderson, V. Barnett, P. C. Chatwin, and A. H. El-Shaarawi (Eds.), *Quantitative Methods for Current Environmental Issues*, London, pp. 37–56. Springer London.

Horrell, M. T. and Stein, M. L. (2017). Half-spectral space-time covariance models. *Spatial Statistics*, 19, 90-100.

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

Kelbert, M. Y., Leonenko, N. N. and Ruiz-Medina, M. D. (2005). Fractional random fields associated with stochastic fractional heat equations. *Advances in Applied Probability*, 37, 108-133.

Krainski, E. T. (2018). *Statistical Analysis of Space-time Data: New Models and Applications*. Ph. D. thesis, Norwegian University of Science and Technology.

Krainski, E. T., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilio, D., Simpson, D., Lindgren, F. and Rue, H. (2019). *Advanced Spatial Modeling with Stochastic Partial Differential Equations using R and INLA*. New York: Chapman and Hall/CRC. Github version www.r-inla.org/spde-book.

Lindgren, F., Bolin, D. and Rue, H. (2022). The SPDE approach for Gaussian and non-Gaussian fields: 10 years and still running. *Spatial Statistics*, 50, 100599.

Lindgren, F., Rue, H. and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation

approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73, 423-498.

Lindgren, G. (2012). *Stationary Stochastic Processes: Theory and Applications*. Chapman and Hall.

Liu, X., Guillas, S. and Lai, M.-J. (2016). Efficient Spatial Modeling Using the SPDE Approach With Bivariate Splines. *Journal of Computational and Graphical Statistics*, 25, 1176-1194.

Liu, X., Yeo, K. and Lu, S. (2022). Statistical Modeling for Spatio-Temporal Data From Stochastic Convection-Diffusion Processes. *Journal of the American Statistical Association*, 117, 1482-1499. Publisher: Taylor & Francis.

Matérn, B. (1960). Spatial variation - stochastic models and their application to some problems in forest surveys and other sampling investigations. *Meddelanden från statens skogsforskningsinstitut*, 49.

Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E. and Houston, T. G. (2012). An overview of the global historical climatology network-daily database. *Journal of Atmospheric and Oceanic Technology*, 29, 897-910.

Moraga, P. (2019). *Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny*. CRC Press.

Niekerk, J. V., Bakka, H., Rue, H. and Schenk, O. (2021). New Frontiers in Bayesian Modeling Using the INLA Package in R. *Journal of Statistical Software*, 100, 1-28.

Porcu, E., Furrer, R. and Nychka, D. (2021). 30 years of space-time covariance functions. *Wiley Interdisciplinary Reviews. Computational Statistics (WIREs)*, 13, e1512, 24.

Prévôt, C. and Röckner, M. (2007). *A concise course on stochastic partial differential equations*, Volume 1905. Springer.

Rodrigues, A. and Diggle, P. J. (2010). A class of convolution-based models for spatio-temporal processes with non-separable covariance structure. *Scandinavian Journal of Statistics*, 37, 553-567.

Roques, L., Allard, D. and Soubeyrand, S. (2022). Spatial statistics and stochastic partial differential equations: A mechanistic viewpoint. *Spatial Statistics*, 50, 100591.

Rozanov, J. A. (1977). Markov random fields and stochastic partial differential equations. *Mathematics of the USSR-Sbornik*, 32, 515-534.

Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71, 319-392.

Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P. and Lindgren, F. (2017). Bayesian computing with INLA: a review. *Annual Review of Statistics and Its Application*, 4, 395-421.

Scheuerer, M. (2010). Regularity of the sample paths of a general second order random field. *Stochastic Processes and their Applications*, 120, 1879-1897.

Serra, L., Saez, M., Juan, P., Varga, D. and Mateu, J. (2014). A spatio-temporal Poisson hurdle point process to model wildfires. *Stochastic Environmental Research and Risk Assessment*, 28, 1671-1684.

Sigrist, F., Künsch, H. R. and Stahel, W. A. (2015). Stochastic partial differential equation based modelling of large space–time data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77, 3-33.

Simpson, D. P., Rue, H., Riebler, A., Martins, T. G. and Sørbye, S. H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors (with discussion). *Statistical Science*, 32, 1-28.

Stein, M. L. (2005). Space–time covariance functions. *Journal of the American Statistical Association*, 100, 310-321.

Stein, M. L. (2012). *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media.

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

van Niekerk, J., Bakka, H., Rue, H. and Schenk, O. (2021). New frontiers in Bayesian modeling using the INLA package in R. *Journal of Statistical Software*, 100, 1-28.

van Niekerk, J., Krainksi, E., Rustand, D. and Rue, H. (2023). A new avenue for Bayesian inference with INLA. *Computational Statistics & Data Analysis*, 181, 107692.

van Niekerk, J. and Rue, H. (2024). Low-rank variational bayes correction to the Laplace method. *Journal of Machine Learning Research*.

Vergara, R. C., Allard, D. and Desassis, N. (2022). A general framework for SPDE-based stationary random fields. *Bernoulli*, 28, 1-32. Publisher: Bernoulli Society for Mathematical Statistics and Probability.

Wahba, G. (1981). Spline Interpolation and Smoothing on the Sphere. *SIAM Journal on Scientific and Statistical Computing*, 2, 5-16. Publisher: Society for Industrial and Applied Mathematics.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, 41, 434-449.

Whittle, P. (1963). Stochastic-processes in several dimensions. *Bulletin of the International Statistical Institute*, 40, 974-994.

Whittle, P. (1986). *Systems in stochastic equilibrium*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Ltd., Chichester.

Wikle, C. K. (2015). Modern perspectives on statistics for spatio-temporal data. *Wiley Interdisciplinary Reviews: Computational Statistics*, 7, 86-98.

Wood, A. W., Leung, L. R., Sridhar, V. and Lettenmaier, D. P. (2004). Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic Change*, 62, 189-216.

Yuan, Y., Bachl, F. E., Lindgren, F., Borchers, D. L., Illian, J. B., Buckland, S. T., Rue, H. and Gerrodette, T. (2017). Point process models for spatio-temporal distance sampling

data from a large-scale survey of blue whales. *The Annals of Applied Statistics*, 11, 2270-2297.

Zuur, A. F., Ieno, E. N. and Saveliev, A. A. (2017). Spatial, temporal and spatial-temporal ecological data analysis with R-INLA. *Highland Statistics Ltd*, 1.

Appendices

A. Almost sure sample path continuity

We start by rephrasing the main theorem of Section 9.3 of Cramér and Leadbetter (1967), and giving a formal definition of the smoothness index.

Definition A.1 (Cramér and Leadbetter, Section 2.5, generalised). *A stochastic process* $x(t)$ *on some domain* $\mathcal{D}$, *is* equivalent *to another process* $y(t)$ *on* $\mathcal{D}$, *if for each fixed* $t \in \mathcal{D}$, $x(t) = y(t)$, *with probability one. This means that x differs from y on at most a set with measure zero, and that they have the same finite dimensional distributions.*

This technical definition allows us to view equivalent processes as an equivalence class that encapsulates some of the finer details of probabilistic measure theory for sample path continuity of stochastic processes.

Theorem A.1 (Cramér and Leadbetter, Section 9.3). *Let* $S^*(\omega)$ *be the spectral measure of a stationary Gaussian process* $x(t)$ *on* $t \in \mathbb{R}$, *and let*

$$I_{a,b} = \int_0^\infty \omega^{2a} \left[\log(1+\omega)\right]^b \, \mathrm{d}S^*(\omega)$$

for $a, b \geq 0$. *For spectral measures that admit a spectral density* $S(\omega)$, *replace* $\mathrm{d}S^*(\omega)$ *in* $I_{a,b}$ *with* $S(\omega)\,\mathrm{d}\omega$.

1. *If* $I_{a,b} < \infty$ *for some* $b > 3$ *and some a in the range* $[k, k+1)$ *for some* $k \in \mathbb{N}$, *then* $x(t)$ *is equivalent to a process* $y(t)$ *that has a continuous sample derivative of order k, with probability one.*

2. *If* $I_{a,1} < \infty$ *for some a in the range* $(k, k+1]$ *for some* $k \in \mathbb{N}$, *then* $x(t)$ *is equivalent to a process* $y(t)$ *whose sample derivative of order k is Hölder continuous with exponent* $a - k \in (0, 1]$, *with probability one.*

For the case $k = 0$, *the sample derivative of order zero refers to the sample path of the process itself.*

Proof. The results follow directly from the main theorem of Section 9.3 of Cramér and Leadbetter (1967). ∎

Results from Scheuerer (2010) show that under a similar condition for d-dimensional domains,

$$\int_{\mathbb{R}^d} \|\boldsymbol{\omega}\|^{2a} \left[\log(1 + \|\boldsymbol{\omega}\|)\right]^b \, \mathrm{d}S^*(\boldsymbol{\omega}) < \infty,$$

for all $a < \nu$ and some $b > 1$, the sample paths on $\mathbb{R}^d$ belong to any Sobolev space $W^{a,2}$ of order $a < \nu$, on any bounded subdomain, with probability one. For isotropic spectra, this translates to $I_{a,b} < \infty$ for all $a < \nu$ and some $b > 1$, when applied to the one-dimensional marginal spectra.

The integral criteria above motivate the following characterisation of the *smoothness index ν*, in particular when applied to models with power law spectral density tails.

Definition A.2. *The smoothness index ν of a stationary Gaussian process $x(t)$, $t \in \mathbb{R}$, is $\nu = \sup_a \{a; I_{a,1} < \infty\}$, where $I_{a,1}$ is defined as in Theorem A.1.*

B. Numerical evaluation of covariances

When spatio-temporal spectral density is available in closed format on $\mathbb{R}^d \times \mathbb{R}$, the covariance function can be obtained to close numerical accuracy using fast Fourier transformation (FFT). In order to reduce the memory requirements for isotropic models on high-dimensional spatial domains, the marginal space-time spectrum along a single spatial dimension can be evaluated first. For general models, evaluating spatial FFT transformations for each time lag further reduces the memory footprint if only some of the covariances are stored.

The idea is construct the folded spectrum resulting from spatial/temporal discretisation, and then discretise it onto a finite regular lattice. The resulting integral approximations can be evaluated with standard FFT implementations, and the numerical approximation error in the covariance evaluation is determined by the frequency resolution and smoothness of the spectral density. The brief theory behind the construction presented below is based on Lindgren (2012).

B.1. Spectral folding

The exact spectral representation of the covariance evaluated on a discrete infinite lattice can be derived from the continuous domain representation. For simplicity, assume the same lattice spacing h in each direction. A stationary covariance function $R(\boldsymbol{s})$ evaluated

at lattice points $\boldsymbol{j}h$, $\boldsymbol{j} \in \mathbb{Z}^d$ is given by

$$R(\boldsymbol{s}) = \int_{\mathbb{R}^d} \exp(i\boldsymbol{\omega} \cdot \boldsymbol{s}) S(\boldsymbol{\omega}) \, d\boldsymbol{\omega},$$

$$R(\boldsymbol{j}h) = \int_{\mathbb{R}^d} \exp(i\boldsymbol{\omega} \cdot \boldsymbol{j}h) S(\boldsymbol{\omega}) \, d\boldsymbol{\omega}$$

$$= \int_{[-\pi/h,\pi/h]^d} \sum_{\boldsymbol{k} \in \mathbb{Z}^d} \exp(i(\boldsymbol{\omega} + 2\pi\boldsymbol{k}/h) \cdot \boldsymbol{j}h) S(\boldsymbol{\omega} + 2\pi\boldsymbol{k}/h) \, d\boldsymbol{\omega}$$

$$= \int_{[-\pi/h,\pi/h]^d} \exp(i\boldsymbol{\omega} \cdot \boldsymbol{j}h) \widetilde{S}(\boldsymbol{\omega}) \, d\boldsymbol{\omega}, \tag{27}$$

where

$$\widetilde{S}(\boldsymbol{\omega}) = \sum_{\boldsymbol{k} \in \mathbb{Z}^d} S(\boldsymbol{\omega} + 2\pi\boldsymbol{k}/h), \quad \boldsymbol{\omega} \in [-\pi/h, \pi/h]^d.$$

If instead the spatial discretisation should be interpreted as the *cell averages* (which is the more usual case for PDE discretisations and e.g. satellite data, rather than pointwise values), the spectrum is altered by a multiplicative frequency filter with a squared sinc function:

$$\widetilde{S}(\boldsymbol{\omega}) = \sum_{\boldsymbol{k} \in \mathbb{Z}^d} S(\boldsymbol{\omega} + 2\pi\boldsymbol{k}/h) \prod_{l=1}^{d} \left\{ \frac{\sin[(\omega_l + 2\pi k_l/h)/2]}{(\omega_l + 2\pi k_l/h)/2} \right\}^2, \quad \boldsymbol{\omega} \in [-\pi/h, \pi/h]^d.$$

B.2. Discrete Fourier transformation

To approximate the integral in (27) with FFT, choose a positive integer M. This gives a numerical integration approximation

$$\widehat{R}(\boldsymbol{j}h) = \left(\frac{\pi}{hM}\right)^d \sum_{\boldsymbol{k} \in [-M,M)^d} \exp\left(i\boldsymbol{k} \cdot \boldsymbol{j} \frac{2\pi}{2M}\right) \widetilde{S}\left(\boldsymbol{k}\frac{\pi}{hM}\right), \quad \boldsymbol{j} \in [-M,M)^d, \tag{28}$$

which is of the form that can be evaluated using FFT.

B.2.1. Sampling

With the above theory, sampling from the model can be expressed as an integral with respect to continuous domain complex valued white noise process, $dZ(\boldsymbol{\omega})$, with conjugate symmetry:

$$x(\boldsymbol{j}h) = \int_{[-\pi/h,\pi/h]^d} \exp(i\boldsymbol{\omega} \cdot \boldsymbol{j}h) \widetilde{S}(\boldsymbol{\omega})^{1/2} \, dZ(\boldsymbol{\omega}), \quad \boldsymbol{j} \in \mathbb{Z}^d,$$

where $\overline{dZ(-\boldsymbol{\omega})} = dZ(\boldsymbol{\omega})$, $\mathsf{Cov}(dZ(\boldsymbol{\omega}), dZ(\boldsymbol{\omega}')) = \delta(\boldsymbol{\omega} - \boldsymbol{\omega}') \, d\boldsymbol{\omega}$. This can be discretised with a lattice of frequencies in much the same way as for computing the covariance function, with noise variances equal to the cell area/volume $\left(\frac{\pi}{hM}\right)^d$ of each frequency lattice point. When the outer pairwise opposing cells are discretised, the combined complex noise contributions are real, and should be assigned to the $-M$ indices, which ensures that the resulting field has no non-zero imaginary components.

C. Spherical harmonics

C.1. Definition and standard properties

In $\mathbb{R}^2$, the harmonic functions, sine and cosine, play an important role as basis functions in spectral representations of functions and random fields. On the sphere, this role is instead taken by the *spherical harmonics*. This section presents the basic results needed for spectral representation theory for stationary processes on the sphere.

Definition C.1. *The* spherical harmonic $Y_{k,m}(\boldsymbol{u})$, $\boldsymbol{u} = \left[u_1, u_2, u_3\right]^\top \in \mathbb{S}^2 \subset \mathbb{R}^3$, *of order* $k = 0, 1, 2, \dots$ *and mode* $m = -k, \dots, k$ *is defined by*

$$Y_{k,m}(\boldsymbol{u}) = \sqrt{(2k+1) \cdot \frac{(k-|m|)!}{(k+|m|)!}} \cdot \begin{cases} \sqrt{2}\sin(m\phi)P_{k,-m}(\cos\theta) & -k \leq m < 0, \\ P_{k,0}(\cos\theta) & m = 0, \\ \sqrt{2}\cos(m\phi)P_{k,m}(\cos\theta) & 0 < m \leq k, \end{cases}$$

where ϕ is the longitude and $\theta = \arccos(u_3)$ is the colatitude, and $P_{k,|m|}(u_3)$ are associated Legendre functions ($P_{k,0}(u_3)$ are Legendre polynomials). Note that $\sin\phi = u_2/\sqrt{u_1^2 + u_2^2}$, $\cos\phi = u_1/\sqrt{u_1^2 + u_2^2}$, and $\cos\theta = u_3$.

Standard property results for spherical harmonics, following Wahba (1981), building the basis of spherical Fourier theory:

1. The spherical harmonics form an orthogonal basis for functions on the unit sphere, $\mathbb{S}^2$:

$$\langle Y_{k,m}, Y_{k',m'} \rangle_{\mathbb{S}^2} = \begin{cases} 4\pi, & k' = k, m' = m, \\ 0, & \text{otherwise.} \end{cases}$$

2. The addition formula for spherical harmonics is

$$\sum_{m=-k}^{k} Y_{k,m}(\boldsymbol{u})Y_{k,m}(\boldsymbol{v}) = (2k+1)P_{k,0}(\boldsymbol{u}^\top \boldsymbol{v}).$$

3. The spherical harmonics are eigenfunctions to the Laplacian on $\mathbb{S}^2$,

$$\Delta Y_{k,m}(\boldsymbol{u}) = -k(k+1)Y_{k,m}(\boldsymbol{u}).$$

4. Let $\phi(\boldsymbol{u})$ be a square-integrable function on $\mathbb{S}^2$. Then $\phi(\boldsymbol{u})$ has series expansion

$$\phi(\boldsymbol{u}) = (\mathcal{F}^{-1}\widehat{\phi})(\boldsymbol{u}) = \sum_{k=0}^{\infty} \sum_{m=-k}^{k} \widehat{\phi}(k,m)Y_{k,m}(\boldsymbol{u}),$$

with *Fourier Bessel* coefficients $\widehat{\phi}(k,m) = (\mathcal{F}\phi)(k,m) = \frac{1}{4\pi}\langle \phi(\boldsymbol{u}), Y_{k,m}(\boldsymbol{u})\rangle_{\mathbb{S}^2(\mathrm{d}\boldsymbol{u})}$. Also, $\langle \phi, 1 \rangle_{\mathbb{S}^2} = 4\pi\widehat{\phi}(0,0)$ and $\langle \phi, \phi \rangle_{\mathbb{S}^2} = 4\pi \sum_{k,m} \widehat{\phi}(k,m)^2$.

C.2. Spherical variance approximation

Define

$$F_{a,b} = \sum_{k=a}^{b} \frac{2k+1}{4\pi[\gamma_s^2 + k(k+1)]^\alpha},$$

so that $F_{0,\infty}$ gives the variance in (23). With

$$I_{a,b} = \int_a^b \frac{2x+1}{4\pi[\gamma_s^2 + x(x+1)]^\alpha} \, \mathrm{d}x$$

$$= \frac{1}{4\pi(\alpha-1)} \left(\frac{1}{[\gamma_s^2 + a(a+1)]^{\alpha-1}} - \frac{1}{[\gamma_s^2 + b(b+1)]^{\alpha-1}} \right),$$

choose K so that the terms in the sum (23) are decreasing for $k \geq K$. This holds for any $K \geq K_0$, where $K_0 = 0$ if $\gamma_s \leq 1/2$, and $K_0 = \left\lceil \sqrt{\frac{\gamma_s^2 - 1/4}{2\alpha - 1}} - \frac{1}{2} \right\rceil$ for $\gamma_s > 1/2$. Then the full sum $F_{0,\infty}$ can be bounded by a partial sum $F_{0,K}$ and tail integrals:

$$F_{0,K} + I_{K+1,\infty} \leq F_{0,\infty} \leq F_{0,K} + I_{K,\infty}.$$

Tighter bounds can in principle be obtained for the approximation $F_{0,\infty} \approx F_{0,K} + I_{K+1/2,\infty}$. Let f_x denote the integrand for $I_{a,b}$. Then a second order Taylor expansion around each $x = k$ gives the error bound

$$F_{0,K} + I_{K+1/2,\infty} - F_{0,\infty} = I_{K+1/2,\infty} - F_{K+1,\infty} \leq \frac{1}{24} \sum_{k=K+1}^{\infty} \sup_{x \in (k-1/2, k+1/2)} f_x'' .$$

It may be possible to construct a bound for this series using another integral bound, but the practical utility of doing so is unclear.

D. Collected proofs

D.1. Proof of Proposition 3.1

The covariance function for spatial lag $s = s_2 - s_1$ and temporal lag t can be written as a nested integral,

$$\mathrm{cov}[u(\mathbf{0},0), u(s,t)] = \int_{\mathbb{R}^d} \int_{\mathbb{R}} \exp[i(s \cdot \boldsymbol{\omega}_s + t\omega_t)] S_u(\boldsymbol{\omega}_s, \omega_t) \, \mathrm{d}\omega_t \, \mathrm{d}\boldsymbol{\omega}_s$$

$$= \int_{\mathbb{R}^d} \exp(is \cdot \boldsymbol{\omega}_s) \left\{ \int_{\mathbb{R}} \exp(it\omega_t) S_u(\boldsymbol{\omega}_s, \omega_t) \, \mathrm{d}\omega_t \right\} \mathrm{d}\boldsymbol{\omega}_s.$$

$$= \int_{\mathbb{R}^d} \exp(is \cdot \boldsymbol{\omega}_s) S_u(\boldsymbol{\omega}_s; t) \, \mathrm{d}\boldsymbol{\omega}_s,$$

where the inner integral $S_u(\boldsymbol{\omega}_s; t)$ is the marginal spatial cross-spectrum for time lag t.

Let $\lambda = \gamma_s^2 + \|\boldsymbol{\omega}_s\|^2$ and $\kappa^2 = \lambda^{\alpha_s}/\gamma_t^2$. Then, integrating over ω_t, we get

$$
\begin{aligned}
S_u(\boldsymbol{\omega}_s;t) &= \frac{1}{(2\pi)^d \gamma_e^2 \lambda^{\alpha_e} \gamma_t^{2\alpha_t}} \int_{\mathbb{R}} \frac{\mathrm{e}^{it\omega_t}}{2\pi(\omega_t^2 + \lambda^{\alpha_s}/\gamma_t^2)^{\alpha_t}} \, \mathrm{d}\omega_t \\
&= \frac{1}{(2\pi)^d \gamma_e^2 \lambda^{\alpha_e} \gamma_t^{2\alpha_t}} \frac{C_{\mathbb{R},\alpha_t}}{\kappa^{2(\alpha_t - 1/2)}} R^M_{\alpha_t - 1/2}(\kappa t) \\
&= \frac{C_{\mathbb{R},\alpha_t}}{\gamma_e^2 \gamma_t} \frac{1}{(2\pi)^d (\gamma_s^2 + \|\boldsymbol{\omega}_s\|^2)^{\alpha}} R^M_{\alpha_t - 1/2}\left\{ t\sqrt{\gamma_s^2 + \|\boldsymbol{\omega}_s\|^2}/\gamma_t \right\},
\end{aligned}
$$

where $R^M_\nu(t)$ is the standard Matérn correlation with smoothness ν, defined in (16), and the additional scaling was given in Lindgren et al. (2011). For $t = 0$, the temporal constriction factor is 1, and we recognize the resulting expression as the spectral density corresponding to a spatial Matérn covariance function with range parameter γ_s and smoothness parameter $\nu_s = \alpha - d/2$, and marginal variance equal to the sought value σ^2 in the proposition. We then also know that the marginal spectrum for $t = 0$ in any single spatial dimension is proportional to $(\gamma_s^2 + \omega^2)^{-\nu_s + 1/2}$, which shows that the conditions on a in Theorem A.1 are fulfilled if and only if $a < \nu_s$, so ν_s is the smoothness index.

D.2. Proof of Proposition 3.2

Let ν_t be the smoothness index for the marginal temporal process $u(\boldsymbol{s},t)$. We need to identify for which values of a the integral $I_{a,1} = \int_0^\infty \omega_t^{2a} \log(1 + \omega_t) S_u(\omega_t) \, \mathrm{d}\omega_t$ in Theorem A.1 is finite. We start by integrating out the spatial spectral dimensions and reparameterising the resulting integral:

$$
\begin{aligned}
S_u(\omega_t) &\propto \int_{\mathbb{R}^d} [\gamma_t^2 \omega_t^2 + (\gamma_s^2 + \|\boldsymbol{\omega}_s\|^2)^{\alpha_s}]^{-\alpha_t} (\gamma_s^2 + \|\boldsymbol{\omega}_s\|^2)^{-\alpha_e} \, \mathrm{d}\boldsymbol{\omega}_s \\
&\propto \int_0^\infty r^{d-1} [\gamma_t^2 \omega_t^2 + (\gamma_s^2 + r^2)^{\alpha_s}]^{-\alpha_t} (\gamma_s^2 + r^2)^{-\alpha_e} \, \mathrm{d}r \\
&\propto \int_0^\infty v^{(d-2)/2} (1 + v)^{-\alpha_e} (\widetilde{\omega}_t^2 + (1 + v)^{\alpha_s})^{-\alpha_t} \, \mathrm{d}v
\end{aligned}
\tag{29}
$$

where we in the second step changed to polar coordinates and in the third set $v = r^2/\gamma_s^2$ and $\widetilde{\omega}_t = \omega_t \gamma_t / \gamma_s^{\alpha_s}$. The integral (29) is finite for all $\widetilde{\omega}_t$ when $\alpha_e + \alpha_s \alpha_t > d/2$. Assuming $a < \nu_t$, we can then write the integral in the smoothness criterion as

$$
\begin{aligned}
I_{a,1} &= \int_0^\infty \omega_t^{2a} \log(1 + \omega_t) S_u(\omega_t) \, \mathrm{d}\omega_t \\
&= C_0 \int_0^\infty \widetilde{\omega}_t^{2a} \log\left(1 + \frac{\widetilde{\omega}_t \gamma_s^{\alpha_s}}{\gamma_t}\right) \int_0^\infty v^{(d-2)/2} (1 + v)^{-\alpha_e} (\widetilde{\omega}_t^2 + (1 + v)^{\alpha_s})^{-\alpha_t} \, \mathrm{d}v \, \mathrm{d}\widetilde{\omega}_t
\end{aligned}
$$

for some constant C_0. Let $\varepsilon > 0$ such that $a + \varepsilon < \nu_t$. Then $\log\left(1 + \frac{\widetilde{\omega}_t \gamma_s^{\alpha_s}}{\gamma_t}\right) \le C_\varepsilon \widetilde{\omega}_t^{2\varepsilon}$ for all $\widetilde{\omega}_t > 0$ for some $C_\varepsilon > 0$. We can then bound $I_{a,1}$ and change the order of integration

since the integrands are positive:

$$I_{a,1} \leq C_0 C_\varepsilon \int_0^\infty \widetilde{\omega}_t^{2(a+\varepsilon)} \int_0^\infty v^{(d-2)/2}(1+v)^{-\alpha_e}(\widetilde{\omega}_t^2 + (1+v)^{\alpha_s})^{-\alpha_t} \, dv \, d\widetilde{\omega}_t$$

$$= C_0 C_\varepsilon \int_0^\infty v^{(d-2)/2}(1+v)^{-\alpha_e} \int_0^\infty \frac{\widetilde{\omega}_t^{2(a+\varepsilon)}}{(\widetilde{\omega}_t^2 + (1+v)^{\alpha_s})^{\alpha_t}} \, d\widetilde{\omega}_t \, dv.$$

The change of variables $w = \frac{\widetilde{\omega}_t}{(1+v)^{\alpha_s/2}}$ in the inner integral gives

$$I_{a,1} \leq C_0 C_\varepsilon \int_0^\infty v^{(d-2)/2}(1+v)^{-\alpha_e} \int_0^\infty \frac{w^{2(a+\varepsilon)}(1+v)^{(a+\varepsilon-\alpha_t)\alpha_s}}{(w^2+1)^{\alpha_t}}(1+v)^{\alpha_s/2} \, dw \, dv$$

$$= C_0 C_\varepsilon \int_0^\infty v^{(d-2)/2}(1+v)^{-\alpha_e - \alpha_s(\alpha_t - a - \varepsilon - 1/2)} \int_0^\infty \frac{w^{2(a+\varepsilon)}}{(w^2+1)^{\alpha_t}} \, dw \, dv.$$

In this expression, the inner integral is a finite constant, C_w, when $2\alpha_t - 2a - 2\varepsilon > 1$, i.e., when $a + \varepsilon < \alpha_t - 1/2$. Since ε can be chosen arbitrarily small, we can make C_w finite for all $a < \alpha_t - 1/2$. The remaining integral has an integrable singularity at $v = 0$ for $d = 1$, and the integral is finite when $\alpha_e + \alpha_s(\alpha_t - a - \varepsilon - 1/2) - (d-2)/2 > 1$. Solving for a and again recognising that ε can be chosen arbitrarily small, we have now shown that $I_{a,1} < \infty$ when both $a < \alpha_t - 1/2$ and $a < \frac{\alpha_e + (\alpha_t - 1/2) - d/2}{\alpha_s} = \frac{\nu_s}{\alpha_s}$ hold. Therefore the temporal smoothness is given by $\nu_t = \min(\alpha_t - 1/2, \frac{\nu_s}{\alpha_s})$.

We now turn to the special case $d = 2$, where we can derive an explicit expression for the spectral density. Let $B(x,y)$ be the beta function,

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1} \, dt.$$

Making the change of variables $1 + x = (1+v)^{\alpha_s}$ in (29) the marginal temporal spectrum becomes

$$S_u(\omega_t) \propto \int_0^\infty (1+x)^{-\frac{\alpha_e-1}{\alpha_s}-1}(\widetilde{\omega}_t^2 + 1 + x)^{-\alpha_t} \, dx \qquad [\text{formula 3.197.9 in G\&R (p317)}]$$

$$\propto B\left(\frac{\alpha_e - 1}{\alpha_s} + \alpha_t, 1\right) {}_2F_1\left(\alpha_t, \frac{\alpha_e - 1}{\alpha_s} + \alpha_t, \frac{\alpha_e - 1}{\alpha_s} + \alpha_t + 1; -\widetilde{\omega}_t^2\right),$$

because $\frac{\alpha_e - 1}{\alpha_s} + \alpha_t = \frac{\nu_s}{\alpha_s} + \frac{1}{2} > 0$. Finally we verify that this spectrum yields the smoothness parameter implied by the general dimension result. Assuming that $a - b$ is not an integer, the hypergeometric function ${}_2F_1(a, b; c, z)$ for large values of z behaves like

$$_2F_1(a,b,c,z) \sim c_1 z^{-a} + c_2 z^{-b} + \mathcal{O}(z^{-a-1}) + \mathcal{O}(z^{-b-1})$$

as $z \to \infty$. If $a - b$ is an integer we have to multiply z^{-a} or z^{-b} with $\log(z)$ (Erdélyi (1953) volume 1, section 2.3.2, page 76). This extra logarithmic factor will not make a difference for the final smoothness. Thus, we may write

$$S_t(\omega_t) = \mathcal{O}(\omega_t^{-2\alpha_t}) + \mathcal{O}\left(\omega_t^{-2\left(\frac{\alpha_e-1}{\alpha_s}+\alpha_t\right)}\right) = \mathcal{O}\left(\omega_t^{-2\left(\alpha_t + \frac{1}{\alpha_s}\min(0,\alpha_e-1)\right)}\right)$$

for large ω_t. This decay rate is such that the conditions in Theorem A.1 are if and only if $a < v_t$ with

$$v_t = \frac{2(\alpha_t + \frac{1}{\alpha_s}\min(0, \alpha_e - 1)) - 1}{2} = \alpha_t + \frac{1}{\alpha_s}\min(0, \alpha_e - 1) - \frac{1}{2} = \min\left[\alpha_t - \frac{1}{2}, \frac{v_s}{\alpha_s}\right],$$

which completes the proof.

D.3. Proof of Theorem 4.1

Define the eigenvector matrix $\boldsymbol{V}$ and the eigenvalue (diagonal) matrix $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_{n_s})$ solving the generalised eigenvalue problem $\boldsymbol{K}_1\boldsymbol{V} = \boldsymbol{C}\boldsymbol{V}\boldsymbol{\Lambda}$. Since $\boldsymbol{K}_1$ and $\boldsymbol{C}$ are symmetric and $\boldsymbol{K}_1$ is positive definite, the eigenvectors can be chosen so that $\boldsymbol{V}^\top\boldsymbol{C}\boldsymbol{V} = \boldsymbol{I}$. For general $a = 0, 1, 2, \ldots$, $\boldsymbol{K}_{a+1} = \boldsymbol{K}_a\boldsymbol{C}^{-1}\boldsymbol{K}_1$, so that $\boldsymbol{K}_{a+1}\boldsymbol{V} = \boldsymbol{K}_a\boldsymbol{V}\boldsymbol{\Lambda}$. Recursion shows that $\boldsymbol{K}_a\boldsymbol{V} = \boldsymbol{C}\boldsymbol{V}\boldsymbol{\Lambda}^a$, which also holds for general $a \geq 0$, and $\boldsymbol{V}^\top\boldsymbol{K}_a\boldsymbol{V} = \boldsymbol{\Lambda}^a$.

For $\alpha_t = 1$, the temporal evolution of the spatial Hilbert space discretisation of (25) is determined by

$$\left(\gamma_t\boldsymbol{C}\frac{\partial}{\partial t} + \boldsymbol{K}_{\alpha_s/2}\right)\boldsymbol{u}(t) = \boldsymbol{C}\,\mathrm{d}\mathcal{E}_{\gamma_e^2\boldsymbol{K}_{\alpha_e}}(t), \quad t \in \mathbb{R}.$$

A multivariate change of variables $\boldsymbol{u}(t) = \boldsymbol{V}\boldsymbol{z}(t)$ and multiplication by $\boldsymbol{V}^\top$ on both sides gives

$$\left(\gamma_t\boldsymbol{I}\frac{\partial}{\partial t} + \boldsymbol{\Lambda}^{\alpha_s/2}\right)\boldsymbol{z}(t) = \boldsymbol{V}^\top\boldsymbol{C}\,\mathrm{d}\mathcal{E}_{\gamma_e^2\boldsymbol{K}_{\alpha_e}}(t) = \mathrm{d}\mathcal{E}_{\gamma_e^2\boldsymbol{\Lambda}^{\alpha_e}}(t),$$

where the precision of the driving noise process follows from

$$\gamma_e^2\left(\boldsymbol{V}^\top\boldsymbol{C}\boldsymbol{K}_{\alpha_e}^{-1}\boldsymbol{C}\boldsymbol{V}\right)^{-1} = \gamma_e^2\boldsymbol{V}^{-1}\boldsymbol{C}^{-1}\boldsymbol{K}_{\alpha_e}\boldsymbol{C}^{-1}\boldsymbol{V}^{-\top}$$
$$= \gamma_e^2\boldsymbol{V}^\top\boldsymbol{K}_{\alpha_e}\boldsymbol{V} = \gamma_e^2\boldsymbol{\Lambda}^{\alpha_e}.$$

For $\alpha_t = 2$, the same technique yields

$$\left(-\gamma_t^2\boldsymbol{C}\frac{\partial^2}{\partial t^2} + \boldsymbol{K}_{\alpha_s}\right)\boldsymbol{u}(t) = \boldsymbol{C}\,\mathrm{d}\mathcal{E}_{\gamma_e^2\boldsymbol{K}_{\alpha_e}}(t)$$

and

$$\left(-\gamma_t^2\boldsymbol{I}\frac{\partial^2}{\partial t^2} + \boldsymbol{\Lambda}^{\alpha_s}\right)\boldsymbol{z}(t) = \mathrm{d}\mathcal{E}_{\gamma_e^2\boldsymbol{\Lambda}^{\alpha_e}}(t).$$

Using the solutions for $\alpha_t = 1$ and 2 as the driving noise processes on the right hand side, the recursive construction technique from Lindgren et al. (2011) gives the general

spatial discretisations

$$\left(-\gamma_t^2 \boldsymbol{C}\frac{\partial^2}{\partial t^2} + \boldsymbol{K}_{\alpha_s}\right)^{\alpha_t/2} \boldsymbol{u}(t) = \boldsymbol{C}\, \mathrm{d}\mathcal{E}_{\gamma_e^2 \boldsymbol{K}_{\alpha_e}}(t),$$

$$\left(-\gamma_t^2 \boldsymbol{I}\frac{\partial^2}{\partial t^2} + \boldsymbol{\Lambda}^{\alpha_s}\right)^{\alpha_t/2} \boldsymbol{z}(t) = \mathrm{d}\mathcal{E}_{\gamma_e^2 \boldsymbol{\Lambda}^{\alpha_e}}(t),$$

for any $\alpha_t = 1, 2, \ldots$. Since the evolution of $\boldsymbol{z}(t)$ is independent between the vector components, we get

$$\left(-\gamma_t^2 \frac{\partial^2}{\partial t^2} + \lambda_i^{\alpha_s}\right)^{\alpha_t/2} z_i(t) = \frac{1}{\gamma_e \lambda_i^{\alpha_e/2}} \mathcal{W}_i(t), \quad \text{for } i = 1, \ldots, n_s,$$

where λ_i is the i:th generalised eigenvalue of $\boldsymbol{K}_1$, and $\mathcal{W}_i(\cdot)$ are white noise processes, independent across all i. Rearranging factors, we get

$$\gamma_e \lambda_i^{\alpha_e/2} \gamma_t^{\alpha_t} \left(-\frac{\partial^2}{\partial t^2} + \gamma_t^{-2}\lambda_i^{\alpha_s}\right)^{\alpha_t/2} z_i(t) = \mathcal{W}_i(t), \quad \text{for } i = 1, \ldots, n_s.$$

Applying the temporal condition of the theorem with $b_i = \gamma_e^2 \lambda_i^{\alpha_e} \gamma_t^{2\alpha_t}$ and $\kappa_i = \lambda_i^{\alpha_s/2}/\gamma_t$ then gives a the temporal discretisation precision for each $z_i(t)$ as

$$\boldsymbol{Q}_{z_i} = \sum_{k=0}^{2\alpha_t} b_i \kappa_i^{2\alpha_t - k} \boldsymbol{J}_{\alpha_t, k/2}.$$

Collecting the processes gives the joint precision as

$$\boldsymbol{Q}_z = \sum_{k=0}^{2\alpha_t} \boldsymbol{J}_{\alpha_t, k/2} \otimes \mathrm{diag}(b_i \kappa_i^{2\alpha_t - k}) = \gamma_e^2 \sum_{k=0}^{2\alpha_t} \gamma_t^k \boldsymbol{J}_{\alpha_t, k/2} \otimes \boldsymbol{\Lambda}^{\alpha_e + (2\alpha_t - k)\alpha_s/2}.$$

The joint discretisation vector in the original parameterisation is given by $\boldsymbol{u} = (\boldsymbol{I} \otimes \boldsymbol{V})\boldsymbol{z}$, with covariance $\boldsymbol{Q}_{\boldsymbol{u}}^{-1} = (\boldsymbol{I} \otimes \boldsymbol{V})\boldsymbol{Q}_{\boldsymbol{z}}^{-1}(\boldsymbol{I} \otimes \boldsymbol{V}^\top)$. We note that $\boldsymbol{V}^{-\top}\boldsymbol{\Lambda}^a \boldsymbol{V}^{-1} = \boldsymbol{K}_a$, so that the joint precision matrix becomes

$$\boldsymbol{Q}_{\boldsymbol{u}} = (\boldsymbol{I} \otimes \boldsymbol{V}^{-\top})\boldsymbol{Q}_z(\boldsymbol{I} \otimes \boldsymbol{V}^{-1}) = \gamma_e^2 \sum_{k=0}^{2\alpha_t} \gamma_t^k \boldsymbol{J}_{\alpha_t, k/2} \otimes \boldsymbol{K}_{\alpha_e + (\alpha_t - k/2)\alpha_s},$$

which completes the proof.

E. Temporal GMRF representation with stationary boundary conditions

We present precision matrices for stationary AR(2) (autogregressive order 2) processes, and then show how this can be used to construct stationary boundary conditions for GMRF representations of 1st and second order Whittle-Matérn type stochastic differential equations.

Lemma E.1. *Let u_k be a stationary AR(2) process with evolution*

$$a_0 u_k + a_1 u_{t-k} + a_2 u_{k-2} = e_k,$$

with $a_0 > 0$ and e_k independent, $e_k \sim N(0,1)$. Then, the precision matrix $\boldsymbol{Q}$ for $(u_1, \ldots, u_n)$ is quint-diagonal, and, except for the upper left and lower right 2×2 corners, $\boldsymbol{Q}$ has diagonal elements $q_0 = a_0^2 + a_1^2 + a_2^2$ and off-diagonal elements $q_1 = a_1(a_0 + a_2)$ and $q_2 = a_0 a_2$. Further, the corner elements are given by

$$Q_{0,0} = Q_{n,n} = a_0^2, \qquad\qquad Q_{1,1} = Q_{n-1,n-1} = a_0^2 + a_1^2,$$
$$Q_{0,1} = Q_{n,n-1} = a_1 a_0, \qquad\qquad Q_{1,0} = Q_{n-1,n} = a_1 a_0.$$

Conversely, if the inner elements q_0, q_1, and q_2 are known, the a_0, a_1, and a_2 values can be recovered, and hence the corner elements be constructed: Define the constants

$$b_+ = \sqrt{q_0 + 2q_1 + 2q_2}, \quad b_- = \sqrt{q_0 - 2q_1 + 2q_2}, \quad b_s = \frac{b_+ + b_-}{2}.$$

Then,

$$a_0 = \frac{1}{2}\left(b_s + \sqrt{b_s^2 - 4q_2}\right), \quad a_1 = \frac{b_+ - b_-}{2}, \quad a_2 = \frac{1}{2}\left(b_s - \sqrt{b_s^2 - 4q_2}\right).$$

Proof. Follows by direct computation. ∎

Let $\Phi_t = \{\phi_1(t), \ldots, \phi_{N_t}(t)\}$ be a set of piecewise linear basis functions in time, on a regular grid, and consider precision matrices on the coefficients for a linear combination of these basis functions. We want to obtain a GMRF representation of a stationary process Ornstein-Uhlenbeck process $z(t)$, such that

$$\kappa z(t) + \frac{d}{dt} z(t) = b^{-1/2} \varepsilon(t), \quad t \in \mathbb{R} \tag{30}$$

where ε is white noise. However, we can instead use the equivalent stochastic process model

$$\left(\kappa^2 - \frac{d^2}{dt^2}\right)^{1/2} z(t) = b^{-1/2} \varepsilon(t), \quad t \in \mathbb{R}. \tag{31}$$

Under stationarity, these two models are equivalent in the sense that they have the same covariance function. Let $\boldsymbol{M}_0 = (\langle \phi_i, \phi_j \rangle)_{i,j}$, $\boldsymbol{M}_2 = (\langle \nabla \phi_i, \nabla \phi_j \rangle)_{i,j}$. Assuming Neumann boundary conditions on a finite interval, and (31), the precision matrix is $\boldsymbol{Q} = b(\kappa^2 \boldsymbol{M}_0 + \boldsymbol{M}_2)$, see Lindgren et al. (2011, Sec 2.3). This matrix does not represent a stationary process on the finite interval. However, it is quint-diagonal, and can be corrected to give a stationary GMRF by adding $b\kappa\sqrt{1 + h^2\kappa^2/4} \approx b\kappa$, to the first and the last entries of

the matrix $\boldsymbol{Q}$, per the previous lemma. Here, h is the step-size in the mesh, and we assume that $h\kappa$ is small. Let $\boldsymbol{M}_1$ be a matrix of zeroes, except the first and last elements which are $1/2$. We then have a stationary GMRF representation of the process (31) with precision matrix

$$b(\kappa^2\boldsymbol{M}_0 + 2\kappa\boldsymbol{M}_1 + \boldsymbol{M}_2). \tag{32}$$

For second order B-spline basis functions, a similar adjustment can be made to the initial and final 2-by-2 blocks of the matrix. In both cases, Taylor expansion of the boundary correction at a specific $\kappa_0 > 0$ is likely preferable when the temporal construction is applied to the space-time construction in Theorem 4.1.

F. Application details

We performed the computations for the temperature case study on a single node machine with 52 cores (26 dual-socket Intel Xeon Gold 6230R CPU) and 755 GB of main memory. After preliminary model fitting with lower resolution spatial mesh we fitted the model with 1 251 mesh nodes. The parallel computations were performed with `inlabru` (version 2.10.0), `INLAspacetime` (version 0.1.7) via `R-INLA` (version 23.11.26) with the `PARDISO` library, using 4 parallel evaluations of the posterior log-density, each one using 8 threads. The average time per function evaluation were 52.94 seconds, 67.96 seconds, 92.06 seconds and 186.05 seconds, respectively for models A, D, C and D. The respective number of evaluations of the posterior density were 479, 1114, 333 and 607, and the total computing time 7.12 hours, 21.09 hours, 8.58 hours and 31.45 hours. The real memory peak were 114.75 GB, 178.78 GB, 219.25 GB and 211.43 GB, respectively. These timings can be significantly reduced by using a different parallelization strategy along with better starting values. Starting values can be set from fits of each one of these models but using lower resolution meshes which allows faster computations.

The computed results were used for the within-sample and leave-one-out prediction scores in Table 3, as well as for the multi-horizon forecast assessment in Section 5.2.4. Details of the multi-horizon forecast scores are shown in Figures 9 and 10, including the mean error (ME, estimated forecast bias), mean absolute error (MAE), mean squared error (MSE), mean Dawid-Sebastiani scores (DS), mean continuous ranked probability score (CRPS), and scale-invariant CRPS (SCRPS).
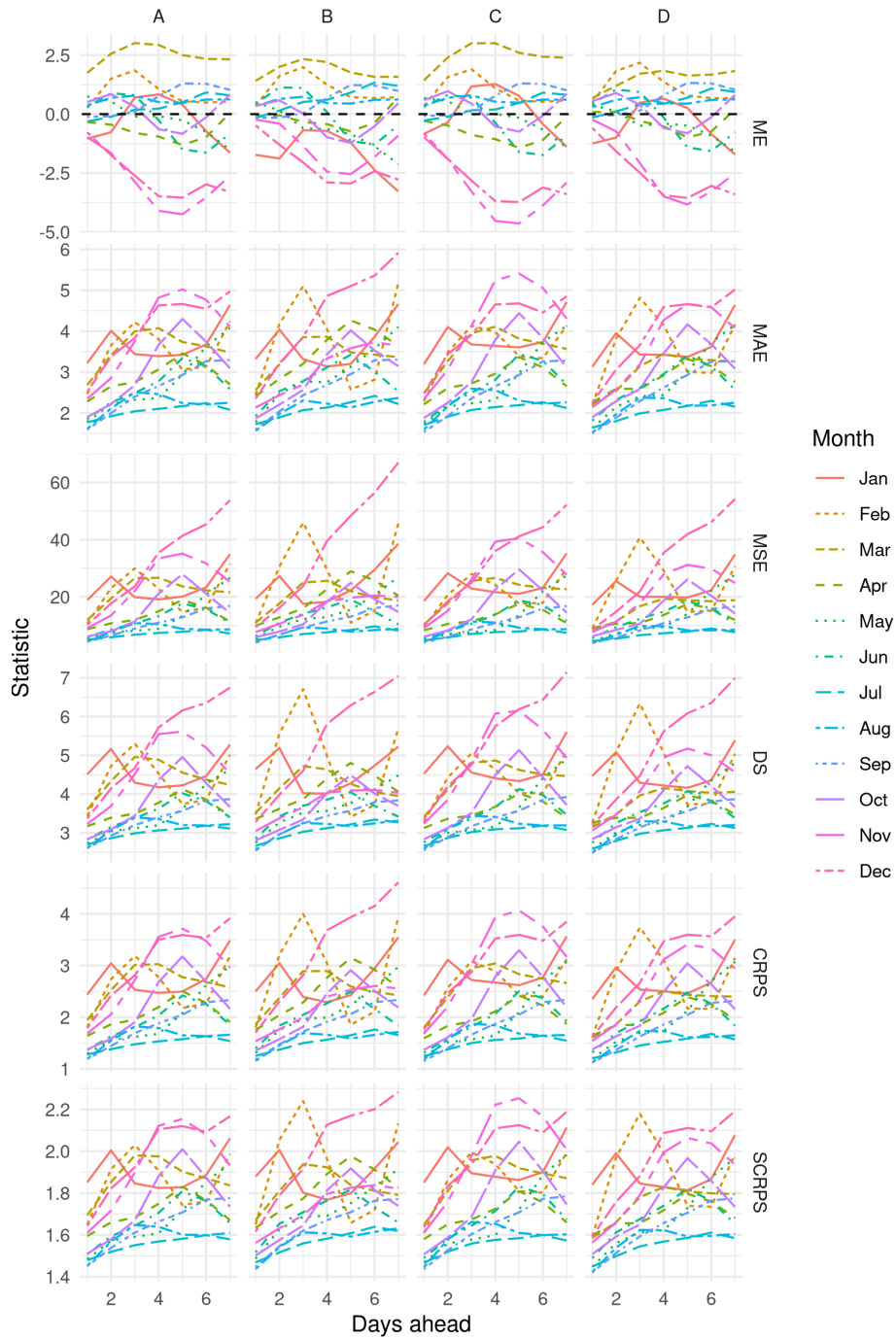
Figure 9. *Mean error and prediction score averages for each model, for each forecast horizon (1–7) and each month of the year, for the multi-horizon multi-scenario setting.*
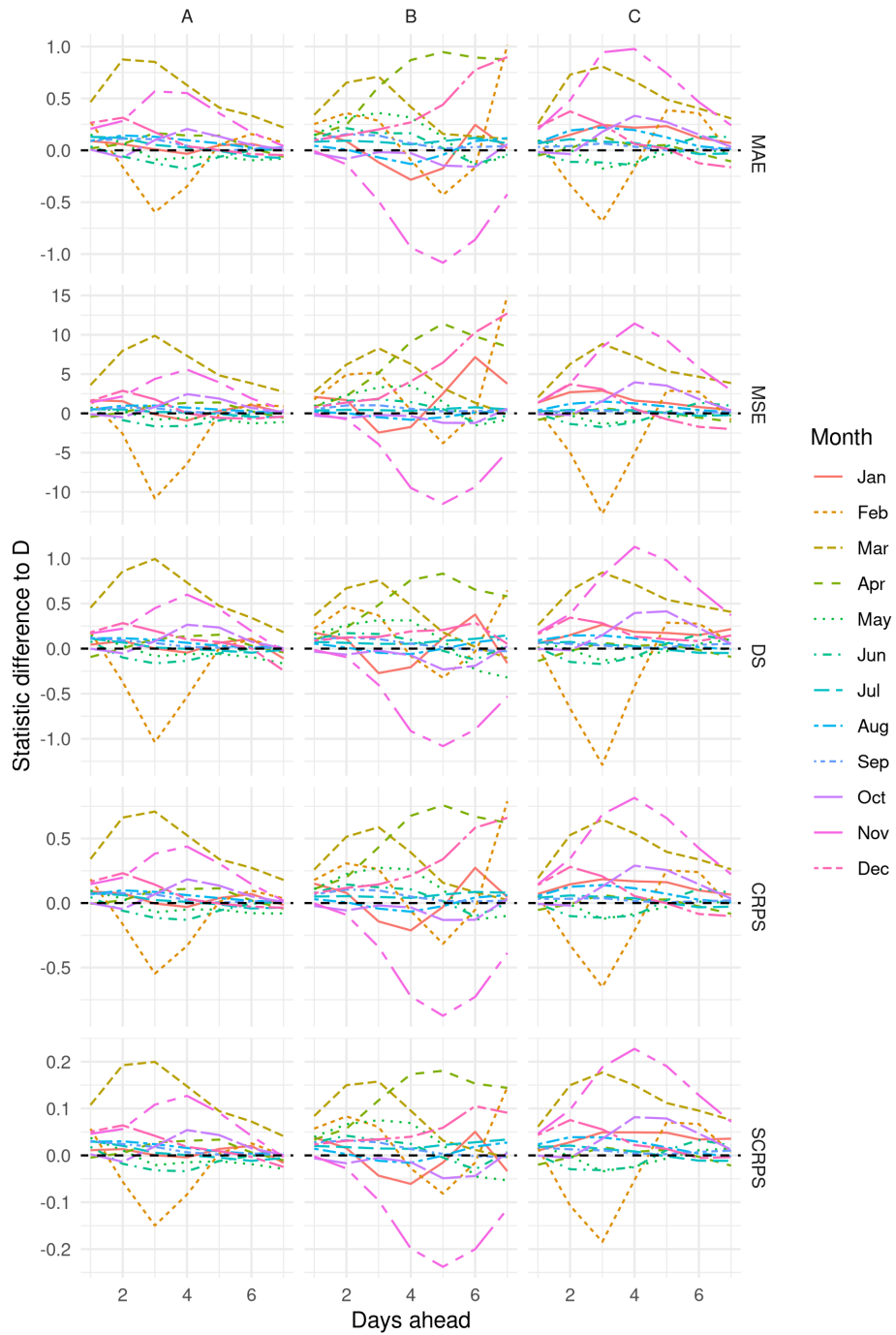
Figure 10. *Prediction score averages for each model with the scores for model D subtracted, for each forecast horizon (1–7) and each month of the year, for the multi-horizon multi-scenario setting.*

Discussion of "A diffusion-based spatio-temporal extension of Gaussian Matérn fields"
by Finn Lindgren, Haakon Bakka, David Bolin, Elias Krainski and Håvard Rue

Christopher K. Wikle
Department of Statistics
University of Missouri, USA

Nathan B. Wikle
Department of Statistics and Actuarial Science
University of Iowa, USA

We congratulate the authors on their spatio-temporal extension of the Matérn spatial process. As outlined in this comprehensive paper, the connection between the Matérn covariance function and its SPDE representation remains of foundational importance to the understanding and construction of geospatial stochastic processes. This paper joins the authors' previous efforts in increasing the awareness (and ease of use) of so-called "constructive" approaches to spatial modeling, most notably their SPDE approach for Gaussian fields (Lindgren, Rue and Lindström, 2011) and its implementation via INLA (Rue et al., 2017). We share the belief that such constructive approaches typically function best when relevant physical processes are incorporated in the model, often to the benefit of the theorist and practitioner alike. Furthermore, the extension from a purely spatial to spatio-temporal model is both elegant and (at least conceptually) simple: specify a dynamical system in the SPDE construction. It is surprising, then, that until Jones and Zhang's (1997) specification of a stationary space-time process via a fractional diffusion equation, this approach received so little attention after its introduction by Heine (1955) and Whittle (1963). Indeed, Jones and Zhang (1997) was ahead of its time; it is nice to see that work getting the attention it deserves and that it has helped to motivate the extension to spatio-temporal Matérn processes presented here. We have little to add in this regard beyond what is said in the paper. Instead, we will focus our brief discussion on the broader SPDE modeling approach.

One interesting question, which might be overlooked upon a first reading of this paper, is whether there is a distinct advantage to specifying the dynamical system in continuous, rather than discrete, time. For example, Cressie and Wikle (2015, Chap. 6) compare the space-time covariance from the diffusion-injection SPDE model of Heine (1955) and the marginal space-time covariance implied by the associated conditional discrete-time dynamic spatio-temporal model (DSTM). Not surprisingly, they are nearly indistinguishable except near the space and time origins. Thus, one obvious advantage of the marginal SPDE approach is the adaptability of the continuous-time model to irregularly observed and/or high-frequency data, although the exact scope of the advantage would depend on the number of observations and the complexity of the specified SPDE. Yet, when the data are reasonably considered to be discrete in time, the conditional DSTM may provide much greater flexibility in deep hierarchical frameworks as it allows parameters

that control the advection, diffusion, and injection processes to themselves be specified as dependent processes (or conditioned on spatio-temporal covariates). This deep conditional modeling, motivated by mechanistic and stochastic parameter processes, is the foundation for the physical-statistical modeling approach pioneered by Mark Berliner and colleagues (e.g., Berliner, 2003; Wikle and Hooten, 2010).

Perhaps a less obvious advantage of continuous SPDE representations in general, and the addition of the temporal smoothness parameter in the model presented here specifically, is its potential adaptability to changes in temporal support. For example, spatial data may represent observations from a time-averaged or aggregated dynamic process (e.g., monthly precipitation totals, annual average pollution concentrations, etc.). In such cases, the SPDE approach can suggest novel marginal spatial dependence structures for the aggregated observations: define the relevant continuous-time process and integrate the solution over the observational time span. This idea has been considered for advection-diffusion processes with certain simplifying assumptions (see Wikle et al. (2022) or Hanks (2017) for a similar approach). However, to our knowledge this has not been attempted with the more general fractional diffusion-like processes considered in this paper. Such an approach may yield relevant marginal spatial dependence structures beyond the fixed-time Gaussian Matérn fields considered here, and continues the broader trend of relating conditional spatial models to relevant spatio-temporal processes found elsewhere in the geospatial literature. We recall D.R. Cox's comment to the classic Besag (1974) work on Markov random fields (MRFs):

> *Nevertheless, understanding of the conditional models may be helped by relating them to temporal-spatial models, and in particular to their stationary distributions. It would be interesting to know what general connections can be established between Mr Besag's auto-models and stationary distributions of simple temporal-spatial processes.* (Besag, 1974, p.225)

Another exciting avenue suggested by the present work concerns extensions to nonlinear spatio-temporal processes. Most biological/environmental/geophysical processes are, at least at some temporal scale, nonlinear. That is, there are explicit interactions between spatio-temporal scales of variability that lead to nonlinear phenomena (density-dependent growth, fronts, predator-prey dynamics, disease spread, etc.). Given that such processes imply higher-order marginal dependence, the constructive approach may motivate new statistical models for such processes. Wikle and Hooten (2010) present a discrete time DSTM approach for quadratic nonlinear processes motivated by PDEs, but the resulting models can be limited by the curse of (parameter) dimensionality and computational tractability. These are the same issues one faces with nonlinear deep neural approaches for modeling nonlinear spatio-temporal processes (see the recent review in Wikle and Zammit-Mangion, 2023). Whether the SPDE approach can overcome these issues is unclear, but it provides, at least, a possible connection (or alternative) to the increasingly popular physically-informed neural network approaches (e.g., Cuomo et al., 2022; Ren et al., 2023).

Another area of future research motivated by the present work is to expand the general manifold ideas herein to suggest new general classes of dependence models for graphical models such as those motivated by time-varying networks (e.g., Ghosh et al., 2022). Indeed, the well-known connection between MRFs and graphical models, and the similar connection between SPDE models and MRF-based implementation methods championed by the authors, suggests a potential to provide mechanistically-motivated process dependence models for time-varying networks, both in their classical representation and in graph neural methods (e.g., Zhou et al., 2020).

In conclusion, we thank the authors for a well-written and important extension of their SPDE/INLA approach to flexible spatio-temporal processes. Not only does this provide a useful suite of models for many real-world data sets, but it has suggested to us several interesting connections and possible extensions to our own work.

References

Berliner, L. M. (2003). Physical-statistical modeling in geophysics, *Journal of Geophysical Research: Atmospheres*, 108.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems, *Journal of the Royal Statistical Society: Series B (Methodological)*, 36, 192-225.

Cressie, N. and Wikle, C. K. (2015). *Statistics for Spatio-Temporal Data*, John Wiley & Sons.

Cuomo, S., Di Cola, V. S., Giampaolo, F., Rozza, G., Raissi, M. and Piccialli, F. (2022). Scientific machine learning through physics–informed neural networks: Where we are and what's next, *Journal of Scientific Computing*, 92, 88.

Ghosh, D., Frasca, M., Rizzo, A., Majhi, S., Rakshit, S., Alfaro-Bittner, K. and Boccaletti, S. (2022). The synchronized dynamics of time-varying networks, *Physics Reports*, 949, 1-63.

Hanks, E. M. (2017). Modeling spatial covariance using the limiting distribution of spatio-temporal random walks, *Journal of the American Statistical Association*, 112, 497-507.

Heine, V. (1955). Models for two-dimensional stationary stochastic processes, *Biometrika*, 42, 170-178.

Jones, R. H. and Zhang, Y. (1997). Models for continuous stationary space-time processes, *in* P. D. E. R.-C. W. W. R. W. T.G. Gregoire, D.R. Brillinger, ed., *Modelling Longitudinal and Spatially Correlated Data*, Springer, New York, NY, 289-298.

Lindgren, F., Rue, H. and Lindström, J. (2011). An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach, *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73, 423-498.

Ren, P., Rao, C., Liu, Y., Ma, Z., Wang, Q., Wang, J.-X. and Sun, H. (2023). Physr: Physics-informed deep super-resolution for spatiotemporal data, *Journal of Computational Physics*, 492, 112438.

Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P. and Lindgren, F. K. (2017). Bayesian computing with inla: A review, *Annual Review of Statistics and Its Application*, 4, 395-421.

Whittle, P. (1963). Stochastic-processes in several dimensions, *Bulletin of the International Statistical Institute*, 40, 974-994.

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

Wikle, C. K. and Zammit-Mangion, A. (2023). Statistical deep learning for spatial and spatiotemporal data, *Annual Review of Statistics and Its Application*, 10, 247-270.

Wikle, N. B., Hanks, E. M., Henneman, L. R. F. and Zigler, C. M. (2022). A mechanistic model of annual sulfate concentrations in the united states, *Journal of the American Statistical Association*, 117, 1082-1093.

Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C. and Sun, M. (2020). Graph neural networks: A review of methods and applications, *AI open*, 1, 57-81.

Jorge Mateu
Department of Mathematics,
University Jaume I of Castellón, Spain
email: mateu@uji.es

Francesco Serafini
School of Mathematics,
Statistics and Physics, Newcastle
email: francesco.serafini@ncl.ac.uk

The authors are to be congratulated on a valuable and thought-provoking contribution in the rapidly developing field of space-time modelling with tangible grounds in statistics, mathematics and computer sciences. Since the first INLA proposal, thousands of scientific contributions have appeared in the literature, using, adapting and enhancing this Bayesian-based inference framework, starting with spatial contexts, and lately extending this paradigm into space-time in combination with more complex developments. In this context, we underline the undoubtful fact that Gaussian random fields with Matérn covariance functions are not only popular (as written by the authors) but necessary building block models in spatial statistics and machine learning. These two scientific communities are coming together to solve joint problems that time ago were solved separately.

This paper extends the Matérn covariance function to a family of spatio-temporal covariance functions, and the way the authors have chosen their way is through the alternative mathematical representation of stochastic partial differential equations (SPDE), rather than following analytical representations of covariance functions. By doing this, this paper enhances, and builds upon, existing SPDE developments and are able to delineate and define the process with desired properties already encoded in the process.

We acknowledge this diffusion-based extension of the Gaussian Matérn fields to a family of spatio-temporal stochastic processes (DEMF) for a number of reasons, very much in particular for the possibilities of improving existing methodologies in the field of spatio-temporal point processes. Our discussion centres around two main points as follows.

1. Applications to point processes

We have found the sentence *'In practice, however, users of statistical software often choose a model based on convenience. If there are available code examples, the choices made in these will often be carried forward into future analyses.'* particularly true in applied contexts, and it highlights the importance of having accessible code for complex

models. An interesting class of models that is not mentioned in the paper but that is already implemented in `inlabru` (Bachl et al., 2019) is the Log-Gaussian Cox Process (LGCP) model. LGCP models are implemented using the methodology described in Simpson et al. (2016) and we believe there is no problem in combining the new class of random effects illustrated in the paper with this point process methodology. Having the ability to use non-separable spatio-temporal GMRFs as random effects in LGCP models is appealing in a number of different applied contexts. In this section, we will provide examples of interesting applications and use of the proposed class of random effects. Most of these examples concern earthquake occurrence but can be used to illustrate different problems.

A first application that comes to mind regarding earthquakes modelling and forecasting is concerned with the magnitude distribution. Two widely popular choices consist in setting a magnitude of completeness M_0, and assuming that the observed magnitudes $m \leq M_0$ follow an exponential distribution; this corresponds to the standard Gutenberg-Richter law (Gutenberg and Richter, 1956), or a truncated Pareto distribution on the seismic moment (Kagan, 1991). In this context, there is great interest in determining whether the parameters of the magnitude distribution are varying over time and space (Herrmann, Piegari and Marzocchi, 2022; El-Isa and Eaton, 2014; Kamer and Hiemer, 2015). These variations are usually estimated partitioning the space-time domain and producing different estimates for each element of the partition. The ability of expressing the parameters as a Gaussian Markov random field (GMRF) with non-separable covariance functions and with a physical interpretation could be very relevant in this context. This will not only be useful in testing more complex hypothesis on the magnitude distribution but also to build better performing models in terms of their operational capabilities (Hiemer and Kamer, 2016). Still in the context of the magnitude distribution, it would be very interesting to express the magnitude threshold M_0 using a spatio-temporal GMRF. This would be important because the magnitude of completeness changes over time and space depending on the quality of the seismographic network and, therefore, it would be appropriate having a non-separable, possibly non-stationary, random effect.

The combination with LGCP models can prove fruitful for modelling long (years, decades) and short (days, weeks, months) term seismicity. Regarding long-term seismicity, Bayliss et al. (2020, 2022) used the SPDE approach and `inlabru` to build models for earthquakes occurrence incorporating covariates and using the random effect to explain the spatial correlation due to unobserved phenomena. Having easy access to non-separable spatio-temporal random effects could greatly improve this type of models. Indeed, these effects would reflect the spatial and temporal change in long-term seismicity, which are mainly due to the variables changing over time and space, such as distance from faults, the level of stress accumulated in the earth crust, and heat flow.

Regarding short-term seismicity models, Serafini, Lindgren and Naylor (2023) adapted the technique used for LGCP to approximate Hawkes process models (Hawkes,

1971), which have intensity of the form

$$\lambda(\mathbf{x}|\mathcal{H}_{\mathbf{x}}) = \mu(\mathbf{x}) + \sum_{\mathbf{x}_i \in \mathcal{H}_{\mathbf{x}}} g(\mathbf{x} - \mathbf{x}_i), \qquad (1)$$

where $\mathbf{x}$ is a point in space-time and possibly equipped with one or more marks (e.g., magnitude, depth), and $\mathcal{H}_{\mathbf{x}}$ is the history of the process up to $\mathbf{x}$. The Hawkes process can be seen as the superposition of a background Poisson process $\mu(\mathbf{x})$ representing events occurring spontaneously, and different offspring processes $g(\mathbf{x} - \mathbf{x}_i)$ representing the points induced by observation $\mathbf{x}_i$. In this context, having non-separable, possibly non-stationary, spatio-temporal random effects may be useful in various ways. First of all, similarly to long-term seismicity, it can be used for the background rate $\mu(\mathbf{x})$. This is usually considered constant in time but varying in space and it is factorised as $\mu(\mathbf{x}) = \mu\nu(\mathbf{x})$, where $\nu(\mathbf{x})$ is just the spatial variation integrating to one over the spatial domain. This is usually estimated independently from μ and the parameters of the triggering function ($g(\mathbf{x} - \mathbf{x}_i)$) (Ogata, 2011). Therefore, a natural first extension would be to use an LGCP model for $\nu(\mathbf{x})$ with non-separable spatio-temporal effects to have a spatio-temporally varying background rate. Another important application would be to use such random effects to define the parameters of the triggering function. As before, this is usually done by partitioning the observations. In contrast, having this type of diffusion models would first provide a more mathematically appealing formulation of models with triggering functions that have spatio-temporally varying parameters, and secondly, will likely provide more capable models in terms of forecasting future occurrences. In general, providing accessible code to implement these random effects in combination with LGCP models will enable many applied researchers to formulate and test more complex hypotheses on the earthquake generation process with the potential of shedding light on new aspects of the process, and improve forecasting models used for operational purposes.

Another potential field of application is cosmology and research devoted in studying the evolution of the spatio-temporal correlation between galaxies locations (Kerscher, 2001; Simon, 2007; Friedrich et al., 2021). Indeed, they also use LGCPs to model the locations of galaxies and, therefore, having a non-separable spatio-temporal process would be useful in this context as opposed to the separable models that are in use now. The problems that cosmologists face are similar to the distance sampling problem described on the `inlabru` website[1]. One caveat would be that they usually are interested in covariance functions defined as power laws of the distance between galaxies. For this field, it would be useful to know if there are special settings of the DEMF parameters able to approximate a power law, and if there is the possibility to generalising the approach proposed in the paper to obtain such covariance functions. Even if this would not be possible, providing accessible code for LGCP to be used in combination with DEMF random effects would be beneficial for this field for two reasons, the first one is having the ability to use non-separable models, and secondly, it can leverage the advantages in

---

[1]https://inlabru-org.github.io/inlabru/articles/2d_lgcp_distancesampling.html

computational time provided by `inlabru`. Indeed, their data are usually composed of petabytes of recordings which usually makes the use of MCMC or similar techniques unfeasible.

2. Non-Euclidean spaces, and anisotropy

Although mathematically very convenient, Euclidean spaces (and Euclidean distances) are arguably not always the best option, and many times they are not even correct in certain contexts. Spatial data living on networks is gaining importance with the development of technology able to provide such type of data. This applies to all types of spatial data, in particular to geostatistical and point pattern data. Linking with the previous section, space-time point pattern models such as LGCPs and Hawkes processes living on networks require accurate while fast inferential frameworks to be able to provide real and helpful solutions to events living on networks.

A number of papers have dealt with the analysis of crime data using self-exciting point process theory, after the analogy drawn by Mohler et al. (2011) between aftershock ETAS models and crime. Zhuang and Mateu (2019) proposed a spatio-temporal Hawkes-type point process model, which includes a background component with daily and weekly periodisation, and a clustering component that is triggered by previous events. However, as crime events are naturally constrained to occur on the streets structure of a city, we advocate the use of point processes on linear networks. A network, or a graph, is a collection of vertices joined by edges. A linear network is a union of finitely many line segments in the plane where different edges only possibly intersect with each other at one of their vertices.

Statistical analysis of network data presents severe challenges (Baddeley et al., 2021). A network is not spatially homogeneous, which creates geometrical and computational complexities and leads to new methodological problems, with a high risk of methodological error. Real network data, as crime data, can also exhibit an extremely wide range of spatial scales. These problems pose a significant challenge to the classical methodology of spatial statistics based on stationary processes, which is largely inapplicable to data on a network. Note also that the choice of distance metric on the network is pivotal in the theoretical development and in the analysis of real data.

As commented above, a Hawkes process can be interpreted as a generalised Poisson cluster process associating to centres, of rate μ, a branching process of descendants. The spatio-temporal Hawkes process has a conditional intensity of the form

$$\lambda(x,y,t) = \mu(x,y,t) + \sum_{t_i; t_i < t} g(x - x_i, y - y_i, t - t_i), \tag{2}$$

where $\mu(x,y,t)$ is the background rate, and $g(x,y,t)$ is the rate of occurrence triggered by an event at time 0 and location at the origin. The triggering density governs the spatio-temporal distance of triggered events from their antecedent events and is usually modelled to decay with distance from the origin over time and space.

Although `inlabru` makes good progress on fitting Hawkes point processes, there is yet a gap in providing a strong framework when network data comes into play. Adapting the diffusion approach proposed in this paper to enhance `inlabru` with fitting capabilities on networks would be welcome by the research community.

Also in this network context, there are some further points to be considered. One is the possibility to generalise equation (24) in the main paper with different basis functions that support processes on non-continuous, non-Euclidean spaces (e.g., processes on networks). Piece-wise linear basis functions, Harmonic basis functions or Karhunen-Loève expansions are still to be extended to these supports, opening theoretical avenues of research. Another aspect is the idea of latent embeddings of point process excitations. When specific events seem to spur others in their wake, marked Hawkes processes enable us to reckon with their statistics. The under-determined empirical nature of these event-triggering mechanisms hinders estimation in the multivariate setting. Spatio-temporal applications alleviate this obstacle by allowing relationships to depend only on relative distances in particular (non-)Euclidean spaces; in this case, we can embed arbitrary event types in a new latent space following the idea of diffusion maps (DM). We might posit a diffusion process across event types. Random walk methods yield approximate manifold embeddings, proven helpful in deep representations. Constructed as graph affinities, the triggering influences guide a Markovian random walk of which diffusion maps may be approximated via spectral decomposition. Indeed, asymmetrical DM embeddings serve as an adequate initial condition, but are not always conducive to stable learning in conjunction with dynamic kernel bases; this approximation builds the DM approach.

A final aspect needed in the context of LGCPs or Hawkes processes is how to exploit inherent properties of existing main directions in the events leading to a clear anisotropic spatial or space-time structure. The paper on discussion focuses on separability and provides a unified framework to deal with separable covariance functions as well as non-separable ones, depending on the value of the parameters. In doing so, the covariance functions they report (e.g. Preposition 3.1, Corollary 3.2.1, Figure 1) are always isotropic. Formulating anisotropic covariance functions is a known exercise and these could drive anisotropic LGCPs. However, using the diffusion-based representation might help in getting deeper the process structure itself while providing comprehensive and flexible anisotropic structures. If, in addition, this can be implemented for network-support point processes, we will be able to handle and model a larger body of more complex problems.

References

Bachl, F. E., Lindgren, F., Borchers, D. L. and Illian, J. B. (2019). inlabru: an R package for Bayesian spatial modelling from ecological survey data. *Methods in Ecology and Evolution*, 10, 760-766.

Baddeley, A., Nair, G., Rakshit, S., McSwiggan, G. and Davies, T. (2021). Analysing point patterns on networks – review. *Spatial Statistics*, 42, 100435.

Bayliss, K., Naylor, M., Illian, J. and Main, I. G. (2020). Data-driven optimization of seismicity models using diverse data sets: Generation, evaluation, and ranking using inlabru. *Journal of Geophysical Research: Solid Earth*, 125, e2020JB020226.

Bayliss, K., Naylor, M., Kamranzad, F. and Main, I. (2022). Pseudo-prospective testing of 5-year earthquake forecasts for California using inlabru. *Natural Hazards and Earth System Sciences*, 22, 3231-3246.

El-Isa, Z. H. and Eaton, D. W. (2014). Spatiotemporal variations in the *b*-value of earthquake magnitude-frequency distributions: Classification and causes. *Tectonophysics*, 615, 1-11.

Friedrich, O., Andrade-Oliveira, F., Camacho, H., Alves, O., Rosenfeld, R., Sanchez, J., Fang, X., Eifler, T. F., Krause, E., Chang, C., et al. (2021). Dark energy survey year 3 results: covariance modelling and its impact on parameter estimation and quality of fit. *Monthly Notices of the Royal Astronomical Society*, 508, 3125-3165.

Gutenberg, B. and Richter, C. F. (1956). Earthquake magnitude, intensity, energy, and acceleration: (second paper). *Bulletin of the seismological society of America*, 46, 105-145.

Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58, 83-90.

Herrmann, M., Piegari, E., and Marzocchi, W. (2022). Revealing the spatiotemporal complexity of the magnitude distribution and b-value during an earthquake sequence. *Nature Communications*, 13, 5087.

Hiemer, S. and Kamer, Y. (2016). Improved seismicity forecast with spatially varying magnitude distribution. *Seismological Research Letters*, 87, 327-336.

Kagan, Y. (1991). Seismic moment distribution. *Geophysical Journal International*, 106, 123-134.

Kamer, Y. and Hiemer, S. (2015). Data-driven spatial b value estimation with applications to california seismicity: To b or not to b. *Journal of Geophysical Research: Solid Earth*, 120, 5191-5214.

Kerscher, M. (2001). Constructing, characterizing, and simulating gaussian and higher-order point distributions. *Physical Review E*, 64, 056109.

Mohler, G. O., Short, M., Brantingham, P., Schoenberg, F. and Tita, G. (2011). Self-exciting point process modeling of crime. *Journal of the American Statistical Association*, 106, 100-108.

Ogata, Y. (2011). Significant improvements of the space-time etas model for forecasting of accurate baseline seismicity. *Earth, planets and space*, 63, 217-229.

Serafini, F., Lindgren, F. and Naylor, M. (2023). Approximation of bayesian hawkes process with inlabru. *Environmetrics*, 34, e2798.

Simon, P. (2007). How accurate is limber's equation? *Astronomy & Astrophysics*, 473, 711-714.

Simpson, D., Illian, J. B., Lindgren, F., Sørbye, S. H. and Rue, H. (2016). Going off grid: computationally efficient inference for log-gaussian cox processes. *Biometrika*, 103, 49-70.

Zhuang, J. and Mateu, J. (2019). A semiparametric spatiotemporal hawkes-type point process model with periodic background for crime data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182, 919-942.

Rejoinder

We thank the discussants for their insightful comments and suggestions. We will address some of the points raised, by topic.

As noted by Wikle & Wikle in their comment, an important aspect of the DEMF models presented in the paper is their physical interpretability. In addition to providing a straightforward method for defining valid space-time stochastic processes without requiring direct access to the resulting covariance function, the continuous definition allows them to form building blocks of complex models. Hierarchical models with multiple observation sources and types of data are easily defined, and the same hierarchical model can involve point-referenced measurements as well as spatially and temporally aggregated information. Moreover, just as in the spatial case handled in Simpson et al. (2016) and extended by Bachl et al. (2019), the processes can be used as building blocks for log-Gaussian Cox process observation models, by constructing numerical integration schemes based on the basis functions used to numerically represent or approximate the theoretical process realisations.

In their comment, Wikle & Wikle also raise the question of whether the increased flexibility of time-discrete models may be advantageous. While such flexibility can be useful in specific situations where the processes exhibit variation and structure on a shorter time scale than the numerical representation can handle in a continuous-time approach, in general we believe that the greater interpretability of time-continuous models outweighs this; see also Simpson, Lindgren and Rue (2011) for a related discussion. One reason is that some of the flexibility of valid time-discrete models constitutes non-physical behaviour that is sensitive to changes in the temporal resolution. In addition, the possibilities for time-discrete models that do have physical meaning, such as advection, diffusion, and injection processes, are equally applicable in the time-continuous setting. The only real obstacle to such extensions, including allowing the model parameter themselves to be spatial, temporal, or spatio-temporal processes, is the practical numerical implementation and inference, both in terms of computational cost and the need for sufficient amounts of information in the available data. Indeed, the comment mentions that the curse of dimensionality also affects the time-discrete approaches. In practice, the modeller needs to decide what temporal spatial resolution, and structural flexibility is needed for any given problem.

Both discussion comments bring up the idea of extending the models beyond $\mathbb{R}^d$ and $\mathbb{S}^2$. One advantage of the SPDE approach to defining stochastic processes and using finite elements or related methods for practical computation is that extensions to curved spaces is relatively straightforward. For sufficiently smooth manifolds, much of the theory and practical implementation details remain the same. The only real difference is the need to define the differential operators in terms of the metric of the manifold. For purely

spatial fields, this was covered for Whittle-Matérn fields in Lindgren, Rue and Lindström (2011), allowing models to be defined on curved compact manifolds (Coveney et al., 2020). As can be seen by the proof of the precision structure construction in the paper, both non-stationary and manifold extensions of the DEMF models are straightforward, as each term in the precision matrix sum only involves Kronecker products of separable time and space operators; See Kirchner and Willems (2023) for recent theoretical results. We agree that it would be interesting to consider extensions of the proposed models to networks or graphs, and this is a topic that we are currently investigating, and we discuss this in further detail below.

In their comments, both Mateu & Serafini and Wikle & Wikle touch upon different aspects of either extending the models to non-linear behaviour or to use them as building blocks in other models, such as Hawkes point process. Continuing the theme of taking inspiration from physics, we note that even in the deterministic case, non-linear PDE models can be extremely challenging, both from a theoretical and practical point of view. For example, the Navier-Stokes equations still lack a general existence and smoothness theory. However, a potentially fruitful avenue may be to extend linearisation techniques to the stochastic setting, by considering linear SPDEs whose solutions approximate the statistical properties of the original non-linear models.

When it comes to using the models as building blocks, such as modelling spatio-temporally varying parameters of a Hawkes process excitation kernel, the need for non-separable models is less clear. In point process settings, the observations generally are only weakly informative about the model parameters, and the non-separability would likely only be a higher-order effect that cannot be reliably estimated or used, and that covariance product separable models are likely to be sufficient. The main utility of the DEMF models is likely to be as primary models in a hierarchical model where the parameters are modelled in a more parsimonious way, ideally in combination with observed covariates with at least qualitatively known impact on the processes.

We agree with Mateu & Serafini that there are exciting opportunities to create new models by finding SPDE operators that generate fields with other properties than those of the Whittle-Matérn class, such as power law covariance functions. Spatially oscillating fields have already been constructed without direct reference to the wave equation Lindgren et al. (2011), but directly adapting the wave equation could have useful applications and increased interpretability, including intermediate models between the heat and wave equations. Another possibility is to exploit the same technique as used by Bolin, Simas and Xiong (2023) and Sørbye, Myrvoll-Nilsen and Rue (2019) to approximate Whittle-Matérn models with fractional exponents. In this approach, the target model is approximated by a sum of just a few high-order Markov processes, but the resulting processes are non-Markovian.

We now turn our attention to the issue of how to define space-time processes on networks or graphs. A suitable space to introduce this extension on is a metric graph Γ, which is defined in terms of a set of vertices $\mathcal{V}$ and a set of edges $\mathcal{E}$ connecting the vertices. The difference to a regular graph is that the edges are defined as rectifiable

curves, and a position $s \in \Gamma$ can be represented as (e,t), where $e \in \mathcal{E}$ denotes an edge, and t is a position on that edge. Thus, these spaces contain linear networks as a special case.

Recently, we introduced Whittle–Matérn fields on metric graphs as the solution to

$$(\kappa^2 - \Delta_\Gamma)^{\alpha/2}(\tau u) = \mathcal{W}, \quad \text{on } \Gamma \tag{1}$$

where $\alpha > 1/2$, $\mathcal{W}$ is Gaussian white noise on Γ and Δ_Γ is the so-called Kirchhoff Laplacian on Γ, which is an operator that acts as the second derivative on the edges (Bolin, Simas and Wallin, 2023a).

This model is well-posed for any compact metric graph, and α controls the sample path regularity of the solutions in the same way as it does when the model is posed on Euclidean domains. Thus, the model can be used to define differentiable Gaussian processes on general metric graphs, and this is as far as we know, the only construction that can do so. Further, if $\alpha \in \mathbb{N}$, these models are Markov random fields (Bolin, Simas and Wallin, 2023b) and this can be used to perform exact and computationally efficient likelihood-based inference (Bolin, Simas and Wallin, 2023d). For fractional α, and for generalised Whittle–Matérn fields where κ and τ are spatially varying, FEM approximations can be used to obtain computationally efficient approximations (Bolin et al., 2023).

These spatial models are implemented in the R package MetricGraph (Bolin, Simas and Wallin, 2023c), which also contains an implementation of LGCPs on metric graphs. An example of a simulated point pattern based on a Whittle–Matérn field with $\alpha = 2$, and an estimate of the log-intensity using R-INLA can be seen in Figure 1.
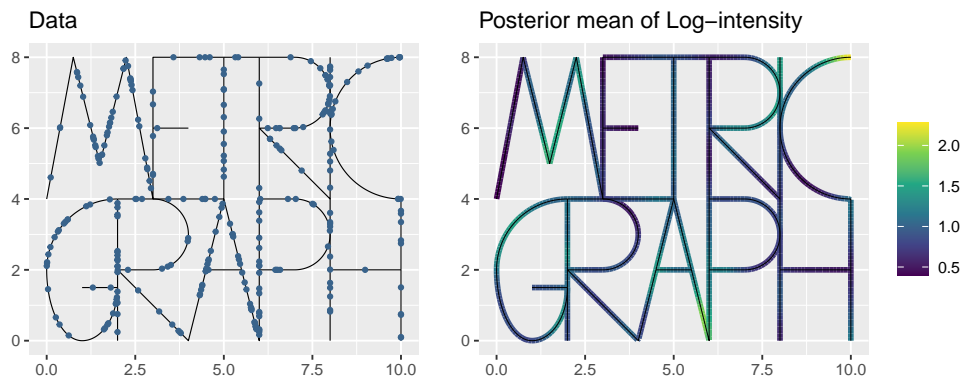

Figure 1. *Simulated LGCP on a metric graph and estimate of log-intensity based on R-INLA and the MetricGraph package.*

We are working on the extension of the spatio-temporal models introduced in this work to the setting of metric graphs. In this case, it is important to consider an extension

of the model which includes an advection term,

$$(\gamma_t \frac{d}{dt} + (\kappa^2 + \rho \, \mathrm{d}_s - \Delta_\Gamma)^{\alpha_s/2})^{\alpha_t} u = \mathrm{d}\mathcal{E}_Q, \quad \text{on } \Gamma \times [0, T], \tag{2}$$

because many datasets on metric graphs, such as river systems, have a clear transport direction. It should be noted that even though the spatio-temporal models considered in this work result in isotropic covariances when posed on $\mathbb{R}^d$, this is not the case for the metric graph setting: Even if $\rho = 0$ in (2), the model is anisotropic on general metric graphs. Thus, we agree that isotropy typically is not a realistic property on metric graphs.

Finally, there are indeed close connections between graphical models and Whittle–Matérn fields on metric graphs, because certain models based on the graph Laplacian can be viewed as finite difference approximations to (1) (Bolin et al., 2023d). Finding similar connections between the spatio-temporal model (2) and spatio-temporal graphical models is an interesting topic for future work.

References

Bachl, F. E., Lindgren, F., Borchers, D. L. and Illian, J. B. (2019). inlabru: an R package for Bayesian spatial modelling from ecological survey data. *Methods in Ecology and Evolution*, 10, 760-766.

Bolin, D., Kovács, M., Kumar, V. and Simas, A. B. (2023). Regularity and numerical approximation of fractional elliptic differential equations on compact metric graphs. *Math. Comp.* In press.

Bolin, D., Simas, A. B. and Wallin, J. (2023a). Gaussian Whittle-Matérn fields on metric graphs. *Bernoulli*. In press.

Bolin, D., Simas, A. B. and Wallin, J. (2023b). Markov properties of Gaussian random fields on compact metric graphs. Preprint, arxiv:2304.03190.

Bolin, D., Simas, A. B. and Wallin, J. (2023c). *MetricGraph: Random fields on metric graphs*. R package version 1.1.2, https://CRAN.R-project.org/package=MetricGraph.

Bolin, D., Simas, A. B. and Wallin, J. (2023d). Statistical inference for Gaussian Whittle–Matérn fields on metric graphs. Preprint, arXiv:2304.10372.

Bolin, D., Simas, A. B. and Xiong, Z. (2023). Covariance-based rational approximations of fractional SPDEs for computationally efficient Bayesian inference. *J. Comput. Graph. Statist.* In press.

Coveney, S., Corrado, C., Roney, C. H., Wilkinson, R. D., Oakley, J. E., Lindgren, F., Williams, S. E., O'Neill, M. D., Niederer, S. A. and Clayton, R. H. (2020). Probabilistic interpolation of uncertain local activation times on human atrial manifolds. *IEEE transactions on bio-medical engineering*, 67, 99-109.

Kirchner, K. and Willems, J. (2023). Regularity theory for a new class of fractional parabolic stochastic evolution equations. *Stochastics and Partial Differential Equations: Analysis and Computations*.

Lindgren, F., Rue, H. and Lindström, J. (2011). An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation ap-

proach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73, 423-498.

Simpson, D., Illian, J. B., Lindgren, F., Sørbye, S. H. and Rue, H. (2016). Going off grid: computationally efficient inference for log-Gaussian Cox processes. *Biometrika*, 103, 49-70.

Simpson, D. P., Lindgren, F. K. and Rue, H. (2011). Think continous: Markovian Gaussian models in spatial statistics. *Spatial Statistics*, 1, 16-29.

Sørbye, S. H., Myrvoll-Nilsen, E. and Rue, H. (2019). An approximate fractional Gaussian noise model with $O(n)$ computational cost. *Statistics and Computing*, 29, 821-833.