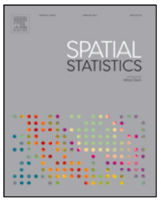
Spatial Statistics 53 (2023) 100718

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta



Surface time series models for large spatio-temporal datasets☆

Israel Martínez-Hernández [a,b,*], Marc G. Genton [b]

[a] Department of Mathematics and Statistics, Lancaster University, Lancaster, United Kingdom
[b] Statistics Program, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

ARTICLE INFO

Article history:
Received 22 August 2022
Accepted 23 November 2022
Available online 5 December 2022

Keywords:
Finite element method
Functional dynamic factor model
Gaussian Markov random field
Large-scale computations
Spatio-temporal modeling
Wind speed

ABSTRACT

The data observed in many phenomena have a spatial and a temporal component. Due to the rapid development of complex, performant technologies, spatio-temporal data can now be collected on a large scale. However, the statistical modeling of large sets of spatio-temporal data involves several challenging problems. For example, it is computationally challenging to deal with large datasets and spatio-temporal nonstationarity. Therefore, the development of novel statistical models is necessary. Here, we present a new methodology to model complex and large spatio-temporal datasets. In our approach, we estimate a continuous surface at each time point, and this captures the spatial dependence, possibly nonstationary. In this way, the spatio-temporal data result in a sequence of surfaces. Then, we model this sequence of surfaces using functional time series techniques. The functional time series approach allows us to obtain a computationally feasible methodology, and also provides extensive flexibility in terms of time-forecasting. We illustrate these advantages through a Monte Carlo simulation study. We also test the performance of our method using a high-resolution wind speed simulated dataset of over 4 million values. Overall, our method uses a new paradigm of data analysis in which the random fields are considered as a single entity, a very valuable approach in the context of big data.

☆ This research was supported by the King Abdullah University of Science and Technology (KAUST), Saudi Arabia.
* Corresponding author at: Department of Mathematics and Statistics, Lancaster University, Lancaster, United Kingdom.
E-mail address: i.martinezhernandez@lancaster.ac.uk (I. Martínez-Hernández).

https://doi.org/10.1016/j.spasta.2022.100718

1. Introduction

We are in an era where data are important assets in many contexts, including decision making. Due to the rapid development of complex, performant technologies, data can now be collected on a large scale, resulting in high-dimensional and high-frequency data. Statistical methods are expected to work on these complex and large datasets. However, the statistical analysis of complex and large datasets involves computational challenges, sometimes necessitating high-performance computing, which could be a limiting factor for the practitioner. These challenges have led us to a new paradigm of data analysis. One approach to overcome these challenges is to assume that observations have characteristics that vary along a continuum, e.g., curves or surfaces. That is, a curve or a surface is considered as a single point observation. Here, we use this continuous approach to propose a methodology to model and forecast complex and large spatio-temporal data that are dense in space and dense in time.

There has been an increasing amount of interest in space–time modeling. Traditionally, statistical methods for spatio-temporal data focus on modeling the space–time covariance function $C(\mathbf{s}_1, t_1, \mathbf{s}_2, t_2)$, with $(\mathbf{s}_i, t_i)$, $i = 1, 2$, being two space–time locations. Therefore, researchers have put considerable work and effort into studying valid models for the covariance function (see, e.g., the recent review by Chen et al., 2021, and references therein). A simple and widely used space–time covariance function is the product of purely spatial and purely temporal covariance functions, $C(\mathbf{s}_1, t_1, \mathbf{s}_2, t_2) = C_{\mathrm{S}}(\mathbf{s}_1, \mathbf{s}_2) \cdot C_{\mathrm{T}}(t_1, t_2)$, where C_{S} and C_{T} are the purely spatial and purely temporal covariance functions, respectively. Such a space–time covariance function is known as *separable*. Although separability does not allow interaction between space and time, it is often appealing because it leads to computationally feasible methods in some scenarios (Genton, 2007). However, the approach of modeling the space–time covariance function can be unfeasible for large datasets. For example, the computational cost to factorize dense $N \times N$ (covariance) matrices is $O(N^3)$, which is challenging for large N. Some approaches to overcome the computational cost are covariance tapering (Furrer et al., 2006; Kaufman et al., 2008) and fixed rank kriging (Cressie and Johannesson, 2008; Nguyen et al., 2014); see the reviews by Sun et al. (2012), Heaton et al. (2019), and Huang et al. (2021). Covariance tapering uses a compactly supported approximation for the covariance function, and fixed rank kriging aims to reduce the parameter dimensionality. For nonstationary spatio-temporal data, see Nychka et al. (2018) and Kuusela and Stein (2018).

Another approach to space–time modeling is combining techniques from time series and spatial statistics. This approach provides statistical models that are spatially descriptive and temporally dynamic (Wikle and Cressie, 1999; Gelfand et al., 2005; Wikle and Hooten, 2010; Sigrist et al., 2012). The temporal modeling mostly accounts for the spatio-temporal variation. In other words, the spatio-temporal data are considered as a multivariate time series whose components are correlated in space. With this approach, many classes of time series models can be used, but it is not suitable for spatially rich data. Instead, if the locations where data are observed are dense in space, then, at each time point, the data can be considered as a continuous surface (a continuous realization of a random field). Then, the surfaces can be modeled using (functional) time series methodologies. This continuous approach gains computational advantages by approximating the continuous random field in a finite-dimensional subspace (dimensional reduction). Here, we use this second approach; namely, we combine techniques from functional data analysis (FDA) and spatial statistics.

FDA assumes that observations have characteristics that vary along a continuum, e.g., curves or surfaces (Ramsay and Silverman, 2005; Kokoszka and Reimherr, 2017). Thus, FDA deals with data that are defined in an intrinsically infinite-dimensional space. Usually, FDA focuses on data that are curves. Due to its capability to model large and complex data, the methodology of FDA has been extended to spatio-temporal data. One possible extension assumes that a curve is observed at each spatial location (Giraldo et al., 2011; Menafoglio et al., 2016). Another extension assumes that the spatial data are a realization of a deterministic continuous random field (Sangalli et al., 2013; Bernardi et al., 2017). Here, we use a similar idea to the latter approach. Furthermore, we combine techniques from functional time series and continuous random field models. Expressly, at each time, we assume that a random field's realization is a single point observation (called surface data), so that the spatio-temporal data result in a sequence of surfaces; we call them *surface time series*.

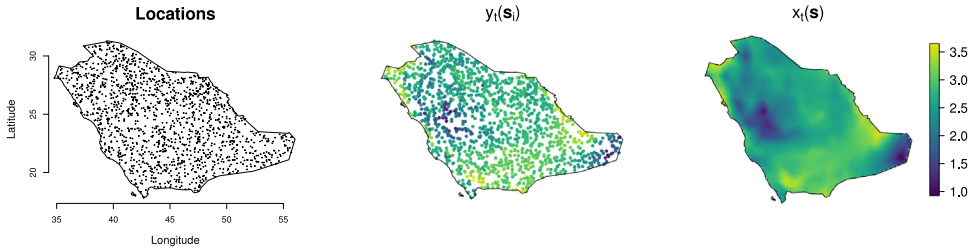


Fig. 1. Left: Locations in Saudi Arabia where wind speed is observed. In total, we have 1500 locations. Center: Snapshot of the square-root of wind speed measurements at time t. Right: The estimated continuous random field $x_t(\mathbf{s})$ in terms of basis functions.

With this approach, spatial dependence is captured through the continuity of the surface, while temporal dependence is modeled with time series methods (see, e.g., the recent review by Martínez-Hernández and Genton, 2020). In practice, the continuous surface data need to be estimated. This continuous surface estimate can be related to kriging in classical spatial data analysis. Here, we use similar ideas to those utilized for one-dimensional functional data to estimate the continuous surface. Specifically, we represent the continuous random field with finite basis functions.

The estimation of continuous surfaces via finite basis function representation is a common approach. An example of a basis function can be the tensor product of univariate B-splines (Wood, 2006; Xiao et al., 2013). Then, the estimation can be formulated as the minimization of the sum of squared errors with a penalization term. The latter method is known as basis-penalty smoother (Wood, 2017). Another choice of basis function is the finite element method, see, e.g., Sangalli et al. (2013). An alternative estimation procedure considers a stochastic partial differential equation (SPDE). In Lindgren et al. (2011), the link between Gaussian random fields and a specific SPDE was studied. In that paper, a Gaussian Markov random field (GMRF) was used to approximate the solution to the SPDE. We detail this approximation in Section 2.1. Its advantage is the sparse structure of the precision matrix of the GMRF, which has computational benefits, especially in large datasets. Then, we use the SPDE approach and the Finite Element Method (FEM) to estimate the continuous representation of the surface. Note that in principle, we can use any of the above methods to estimate the continuous surface since our methodology is not limited to the specific case of the SPDE approach.

Our aim in this paper is to propose a new methodology for analyzing and forecasting complex and large spatio-temporal data that is computationally feasible and implementable with existing R packages (R Core Team, 2022). We assume that the data are realizations of a time series taking values in a continuous space. Specifically, for each discrete time, we estimate the continuous random field. Thus, the (nonstationary) spatial dependence is considered in the continuous estimation of the random field. Then, we model the temporal dependence assuming that the data are surface time series data. Finally, we use time series techniques to describe temporal dependence and then forecast the spatio-temporal data. An advantage of our proposed methodology is that it can handle spatio-temporal data with different numbers of spatial locations and different locations for each time t since we estimate the continuous random field at each discrete time.

Our proposal is motivated by high-resolution simulated wind data from the Weather Research and Forecasting (WRF) model. Fig. 1 illustrates our proposal. In the left panel, 1500 locations $\mathbf{s}_i$ over Saudi Arabia are shown at which the wind speed data are assumed to be observed. Let $y_t(\mathbf{s}_i)$ be the observed data displayed in Fig. 1 (center) at locations $\mathbf{s}_i$ and a fixed time t, $t = 1, \ldots, T$. For each time t and using the observed data $y_t(\mathbf{s}_i)$, we estimate the continuous random field $x_t(\mathbf{s}) = \sum_k b_{k,t}\psi_k(\mathbf{s})$ in terms of finite basis functions $\{\psi_k\}$. Fig. 1 (right) shows the resulting continuous random field. Then, we model the temporal-dependence through the coefficients $\{b_{k,t}\}$ of the basis functions using multivariate functional time series methodology. The latter uses past information, including neighbors, for each location; this is detailed in Section 2.2. Thus, we allow interaction between space and time through the coefficients. Because of the latter, our approach is

computationally more feasible than the covariance modeling approach. Furthermore, our approach can model data with spatial locations changing from time to time since x_t is considered continuous for each time point t. Hence, this paper provides practitioners with an alternative method of modeling large spatio-temporal data without necessitating high-performance computing.

The remainder of our paper is organized as follows: In Section 2, we present our proposed methodology. This section describes the estimation of the continuous surfaces and the modeling of the surface time series. In Section 3, we describe how to estimate the parameters involved in our proposed methodology. Also, we describe how to obtain forecasts. In Section 4, we conduct a simulation study to evaluate the performance of the proposed methodology under different simulation settings. In Section 5, we analyze a dataset of hourly wind speed from the Weather Research and Forecasting (WRF) model. In Section 6, we present some discussion.

2. Surface time series models

We consider the analysis of complex and large spatio-temporal datasets, where the spatial component is assumed to be continuous and the temporal component is assumed to be discrete. Specifically, we assume that the observed data have the form $\{y_1(\mathbf{s}), y_2(\mathbf{s}), \ldots, y_T(\mathbf{s})\}$, where for each $t \in \{1, \ldots, T\}$, $y_t : D \subset \mathbb{R}^2 \to \mathbb{R}$ is a continuous function, $\mathbf{s} \in D$ represents the location, and D represents the study area (e.g., a country). In practice, data are observed on a finite set of points in D, i.e., for each $t = 1, \ldots, T$, we observe m_t points $\mathbf{y}_t = \{y_t(\mathbf{s}_1), \ldots, y_t(\mathbf{s}_{m_t})\}^\top$. Additionally, data may be observed with measurement errors. Therefore, we assume that the observed data are such that

$$\begin{aligned} y_t(\mathbf{s}_i) &= x_t(\mathbf{s}_i) + \varepsilon_t(\mathbf{s}_i), \\ x_t(\mathbf{s}_i) &= \mu_t(\mathbf{s}_i) + z_t(\mathbf{s}_i), \end{aligned} \tag{1}$$

where $\mu_t(\mathbf{s}_i)$ is the mean representing large scale variation, and $\boldsymbol{\varepsilon}_t = \{\varepsilon_t(\mathbf{s}_1), \ldots, \varepsilon_t(\mathbf{s}_{m_t})\}^\top \sim \mathcal{N}_{m_t}(\mathbf{0}, \tau_{\varepsilon,t}^{-1}\mathbf{I}_{m_t})$ is the i.i.d. random noise that accounts for the measurement error (the nugget). Here, we assume $z_t(\mathbf{s}_i)$ is a realization of an unknown continuous Gaussian random field (GRF), the most common choice for modeling spatial (and spatial–temporal) dependencies (Cressie and Wikle, 2011).

Our goal is to model and forecast the continuous surface time series $\{x_t;\ t = 1, \ldots, T\}$. For that, we use the two-step methodology that is widely used in functional time series; see, e.g., Hyndman and Ullah (2007) and Aue et al. (2015). Explicitly, we propose extending this methodology to surface time series: 1) estimate the continuous function x_t; and 2) model the surface time series $\{x_t;\ t = 1, \ldots, T\}$. In the following, we detail these two steps.

2.1. Estimating the continuous random field

The first step in our methodology is to estimate the continuous random field. To simplify notation, we assume a constant mean at each time, i.e., $\mu_t(\mathbf{s}_i) = \mu_t$; thus, we focus on z_t. For each $t = 1, \ldots, T$, the continuous random field $z_t(\mathbf{s})$ accounts for the spatial dependence. Therefore, we estimate $z_t(\mathbf{s})$ using the Matérn covariance function (Matérn, 1986; Stein, 1999, p. 31), which is defined for two locations $\mathbf{s}_1$ and $\mathbf{s}_2$ as

$$\operatorname{cov}\{z_t(\mathbf{s}_1), z_t(\mathbf{s}_2)\} = \frac{\sigma_{z,t}^2}{\Gamma(\nu_t)2^{\nu_t-1}}(\kappa_t\|\mathbf{s}_1 - \mathbf{s}_2\|)^{\nu_t}K_{\nu_t}(\kappa_t\|\mathbf{s}_1 - \mathbf{s}_2\|),$$

where $\|\cdot\|$ is the Euclidean distance, K_{ν_t} is the modified Bessel function of the second kind of order ν_t, and $\sigma_{z,t}^2 = 1/\tau_{z,t}$ is the marginal variance. The parameter ν_t controls the mean-square differentiability of the underlying process. The parameter $\kappa_t > 0$ is a scaling parameter related to the range ρ_t. The empirically derived definition of the range is $\rho_t = \sqrt{8\nu_t}/\kappa_t$ with ρ_t corresponding to the distance where the spatial correlation for each ν_t is close to 0.1. In practice, the smoothness parameter ν_t is often fixed based on our *a priori* belief of the smoothness of the underlying process. Here, we fix $\nu_t = 1$ for all t, similarly to Lindgren et al. (2011) and Cameletti et al. (2013). The choice of ν_t is flexible enough to cover a large class of spatial variations (see Rue et al., 2009).

From model (1), we have $\mathbf{y}_t|\mathbf{x}_t, \tau_{\varepsilon,t} \sim \mathcal{N}_{m_t}(\mathbf{x}_t, \tau_{\varepsilon,t}^{-1}\mathbf{I}_{m_t})$, where $\mathbf{x}_t = \{x_t(\mathbf{s}_1), \ldots, x_t(\mathbf{s}_{m_t})\}^\top$. Usually, the inference of GRF is performed using the covariance matrix $\mathbf{\Sigma}_t$ of $\mathbf{z}_t$. However, for large datasets, this is not computationally feasible because the covariance matrix is completely dense and, therefore, difficult to work with. To reduce the computational cost, we notice that a Gaussian spatial process with Matérn covariance is the solution to the SPDE model

$$\tau_t(\kappa_t^2 - \Delta)^{\alpha_t/2} z_t(\mathbf{s}) = W_t(\mathbf{s}), \tag{2}$$

where $\Delta = \sum_{j=1}^{2} \partial^2/\partial s_j^2$ is the Laplacian operator, $\alpha_t = \nu_t + 1$ controls the smoothness of the spatial process, and $\tau_t > 0$ controls the variance. The process W_t is a spatial Gaussian white noise. Following this SPDE approach, Lindgren et al. (2011) used an FEM to build the Gaussian Markov random field (GMRF) that best approximates the solution of (2). Given a triangulation of the domain D, the solution to Eq. (2) is expressed as

$$z_t(\mathbf{s}) = \sum_{k=1}^{K} b_{k,t}\,\psi_k(\mathbf{s}), \tag{3}$$

where $\{\psi_k\}$ are piecewise linear basis functions defined on the triangular mesh, and $\{b_{k,t}\}$ are Gaussian-distributed random weights. The solution (3) is found based on equality in distribution criterion, resulting in a GMRF. The selection of K depends on each data application, and should be specified by the user; for more details, see Lindgren et al. (2011). Thus, we approximate the continuous random field z_t in (1) with a finite element representation (3).

As mentioned above, our methodology is not limited to the specific case of the SPDE approach in the sense that we can use other methods to estimate the coefficients $\{b_{k,t}\}$ in (3). In the following, we assume that data are a sequence of continuous surfaces with representations as in (3) regardless of how these surfaces are estimated.

2.2. Modeling temporal dependence

The second step in our methodology is to model temporal dependence taking into account the spatial dependence. In this section, without loss of generality, we assume that observations are continuous random fields observed over time with the form (3) (i.e., the mean component μ_t is zero). Thus, $\mathbf{b}_t = (b_{1,t}, \ldots, b_{K,t})^\top$ is a K-dimensional vector coefficient that serves as a time-dependent weight on the basis functions that are spatially correlated. The time series $\{\mathbf{b}_t\}_t$ can result in a high-dimensional and nonstationary time series. The dimension K will depend on how fine the mesh of the region D is. In our data analysis (Section 5), we use $K = 364$. The corresponding mesh of D is illustrated in Fig. 3. In this paper, we assume that the time series $\{\mathbf{b}_t\}_t$ is high-dimensional and includes regions D that may be large or have complex shapes.

We model $\mathbf{b}_t$ with a multivariate functional time series model. To that end, we assume that, for each $k = 1, \ldots, K$, $\{b_{k,t}\}_t$ is a realization of a functional time series $\{f_n^{(k)}(u);\ u \in \mathcal{T},\ n = 1, \ldots, N\}$, where

$$f_n^{(k)}(u) = b_{k,u}\mathbb{1}\{u \in [(n-1)\delta, n\delta], n \in \mathbb{N}\}, \tag{4}$$

with $\mathbb{1}(\cdot)$ being the indicator function, and δ representing the period of time that encompasses the continuous function $f_n^{(k)}(\cdot)$. For example, δ can represent days, months, or years, and is defined by the user depending on the data and on the problem being addressed. In Section 5, we define δ to represent days, so, for each k, $\{f_n^{(k)}(u)\}_n$ is a daily functional time series in which n represents the indices of the days and u represents the time within a day. Then, with representation (4), we have $\{\mathbf{f}_n\}_n$ as a K-dimensional functional time series, where $\mathbf{f}_n = \{f_n^{(1)}(u), \ldots, f_n^{(K)}(u)\}^\top$. This approach allows us to forecast different time horizons, e.g., hours, days, weeks, months, or years.

Because of the high dimensionality, we propose modeling $\{\mathbf{f}_n\}_n$ with a functional dynamic factor model (Hays et al., 2012; Gao et al., 2019; Martínez-Hernández et al., 2022). This model uses eigenfunctions of the long-run covariance of $\{\mathbf{f}_n\}_n$. Specifically, we approximate the functional time

series with a dynamic Functional Principal Component (FPC) model (Hörmann et al., 2015), and then we model the multivariate time series of scores using a factor model:

$$\begin{aligned}
&\mathbf{f}_n(u) = \boldsymbol{\Lambda}(u)\boldsymbol{\beta}_n + \boldsymbol{\eta}_n(u),\\
&\boldsymbol{\beta}_{p,n} = \mathbf{A}_p\boldsymbol{\theta}_{p,n} + \mathbf{e}_{p,n},\\
&\boldsymbol{\theta}_{p,n} - \sum_{i=1}^{q_1}\mathbf{G}_i\boldsymbol{\theta}_{p,n-i} = \boldsymbol{\epsilon}_n + \sum_{j=1}^{q_2}\mathbf{M}_j\boldsymbol{\epsilon}_{p,n-j},
\end{aligned} \tag{5}$$

where $\boldsymbol{\Lambda}(u)$ is the $K\times p_0K$ block matrix with $1\times p_0$ diagonal blocks $\boldsymbol{\lambda}^{(k)}(u) = \{\lambda_1^{(k)}(u), \ldots, \lambda_{p_0}^{(k)}(u)\}^\top$ of eigenfunctions of the long-run covariance of $\{f_n^{(k)}\}$, where $k = 1, \ldots, K$, p_0 is the number of principal components, and zero elsewhere; $\boldsymbol{\beta}_n = (\beta_{1,n}^{(1)}, \ldots, \beta_{p_0,n}^{(1)}, \beta_{1,n}^{(2)}, \ldots, \beta_{p_0,n}^{(K)})^\top$ is the p_0K-dimensional vector containing the FPC scores of all K functional time series; and $\boldsymbol{\eta}_n(u)$ represents the remaining principal components terms from p_0+1 to infinity. Unlike the classical FPC analysis, the dynamic FPC analysis takes into account the temporal dependence of the functional time series $\mathbf{f}_n$.

For each principal component $p = 1, \ldots, p_0$, $\boldsymbol{\beta}_{p,n} = (\beta_{p,n}^{(1)}, \beta_{p,n}^{(2)}, \ldots, \beta_{p,n}^{(K)})^\top$ is the vector of the pth FPC score of all K functional time series; $\mathbf{A}_p$ is the $K\times L$ constant matrix of factor loadings; $\boldsymbol{\theta}_{p,n}$ is the L-dimensional vector of the factor time series; the error term $\mathbf{e}_{p,n}$ represents what is not explained by the factor time series; $\mathbf{G}_i$ and $\mathbf{M}_j$ are the $L\times L$ coefficient matrices representing the temporal evolution of the factor time series; and $\boldsymbol{\epsilon}_{p,n}$ is a white noise vector. The dynamic FPC analysis reduces the functional time series to vector time series consisting of FPC scores. It also takes into account the temporal dependence, unlike the classical FPC analysis.

Models (3) and (5) provide extensive flexibility in the predictions, which is attractive for several applications. For example, the time scales of predictions from wind speed data range from minutes to days (short, medium, and long-term predictions). Most research focuses on short, medium, or long-term forecasts, but none predicts each of these different time scales using only one model. A model that can forecast different time scales is computationally convenient. It provides flexibility in the interpretation with no need to estimate multiple models. With our models (5), short and medium-term predictions are obtained with a one-step-ahead prediction $\widehat{f}_{N+1}$. The long-term forecast is obtained with an h-step ahead prediction $\widehat{f}_{N+h}$, where $h > 1$ (see details in Section 5).

Remark 1. If the mean values μ_t are not zero, then $\{\mu_t\}$ can be modeled together with the coefficient $\mathbf{b}_t$. That is, instead of using $b_{k,u}$ in (4), we can use $\mu_u + b_{k,u}$. Alternatively, one can model and predict $\{\mu_t\}$ separately from $\mathbf{b}_t$, e.g., using a univariate functional time series if $\mu_t(\mathbf{s}) = \mu_t$, where the univariate functional time series is constructed similarly as in (4).

3. Inference and forecasting

3.1. Inference

To estimate the continuous random field x_t, we need to estimate the coefficients $b_{k,t}$ in (3) and the mean μ_t (possibly dependent on $\mathbf{s}$). For this estimation, we use the R-INLA framework (Rue et al., 2009), which requires that we set priors on the hyperparameters $\tau_{\varepsilon,t}$, $\sigma_{x,t}$, and ρ_t. The choice of priors depends on the characteristics of the data. We assign a joint prior on $\sigma_{x,t}$ and ρ_t in the same way as described in Fuglstad et al. (2019). That is, the joint prior is specified indirectly using the concept of penalized complexity (PC) (see Simpson et al., 2017). We only need to specify the tail probabilities $P(\rho_t < \rho_0) = p_{\rho_t}$ and $P(\sigma_{x,t} > \sigma_0) = p_{\sigma_{x,t}}$. In other words, we only need to set the lower tail quantile ρ_0 and probability p_{ρ_t} for the range, and the upper tail quantile σ_0 and probability $p_{\sigma_{x,t}}$ for the standard deviation. For $\tau_{\varepsilon,t}$, we assume a vague Gamma prior with parameters 1 and 0.00005.

For the model (5), we need to estimate $\boldsymbol{\Lambda}(u)$, $\boldsymbol{A}_p$, $\mathbf{G}_i$, and $\mathbf{M}_j$. To estimate $\boldsymbol{\Lambda}(u)$, we use the dynamic FPC analysis. Once $\boldsymbol{\Lambda}(u)$ is estimated, we obtain $\boldsymbol{\beta}_n$, and then we estimate the components of the factor model. A common assumption in factor models is that the factors are independent. Here, we also make this assumption. Therefore, we model each factor time series $\{\theta_{p,n}^{(l)}\}_n$, $l = 1, \ldots, L$,

separately. Consequently, the matrices $\mathbf{G}_i$ and $\mathbf{M}_j$ are diagonal matrices in model (5). For this estimation, we use the *ftsa* package (Hyndman and Shang, 2020). In the following, we detail the estimation procedures for model (5).

Let $c^{(k)}(u,v) = \sum_{h=-\infty}^{\infty} c_h^{(k)}(u,v)$ be the long-run covariance of the functional time series $\{f_n^{(k)}\}$, with $c_h^{(k)}(u,v) = \text{cov}\{f_n^{(k)}(u), f_{n+h}^{(k)}(v)\}$ being the auto-covariance function at lag h. We define the long-run covariance operator as $C^{(k)}(f)(u) = \int_{\mathcal{T}} c^{(k)}(u,v) f(v) \mathrm{d}v$. An estimator $\widehat{\lambda}_p^{(k)}$ for $\lambda_p^{(k)}$ is defined as the eigenfunction of the empirical long-run covariance operator. That is, $\widehat{\lambda}_p^{(k)}$ is such that $\widehat{C}^{(k)}(\widehat{\lambda}_p^{(k)})(u) = \widehat{\zeta}_p^{(k)}\widehat{\lambda}_p^{(k)}(u)$, where $\widehat{\zeta}_p^{(k)}$ is the corresponding eigenvalue in descending order. The estimated operator $\widehat{C}^{(k)}$ is obtained using a bandwidth selection method described in Rice and Shang (2017). Explicitly, $\widehat{c}^{(k)}(u,v) := \sum_{|h|\leq h} W(h/q)\widehat{c}_h^{(k)}(u,v)$, where $\widehat{c}_h^{(k)}(u,v)$ is the empirical auto-covariance function at lag h, and W is a symmetric weight function. Then, the components of $\boldsymbol{\beta}_n$ in (5) are obtained as $\tilde{\beta}_{p,n}^{(k)} = \int_{\mathcal{T}} f_n^{(k)}(u)\widehat{\lambda}_p^{(k)}(u)\mathrm{d}u$.

We use the estimated $\tilde{\boldsymbol{\beta}}_{p,n}$ vector to estimate $\boldsymbol{A}_p$, $\mathbf{G}_i$, and $\mathbf{M}_j$. A common approach is to estimate $\boldsymbol{A}_p$ and $\boldsymbol{\theta}_{p,n}$ by the eigendecomposition of a matrix involving the covariance and cross-covariance of $\boldsymbol{\beta}_{p,n}$, $\boldsymbol{\theta}_{p,n}$, and $\mathbf{e}_{p,n}$, at different lags h (see, e.g., Lam et al., 2011; Gao et al., 2019). Let $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{(p)}(h) = \text{cov}(\boldsymbol{\beta}_{p,n+h}, \boldsymbol{\beta}_{p,n})$, $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(p)}(h) = \text{cov}(\boldsymbol{\theta}_{p,n+h}, \boldsymbol{\theta}_{p,n})$, and $\boldsymbol{\Sigma}_{\mathbf{e}}^{(p)}(h) = \text{cov}(\mathbf{e}_{p,n+h}, \mathbf{e}_{p,n})$ be the covariance matrices. Similarly, let $\boldsymbol{\Sigma}_{\boldsymbol{\theta},\mathbf{e}}^{(p)}(h)$ be the cross-covariance matrix between $\boldsymbol{\theta}_{p,n}$ and $\mathbf{e}_{p,n}$ at lag h. From model (5), we have $\mathbf{L}^{(p)} = \mathbf{L}^{(p)*} + \mathbf{E}^{(p)}$, where $\mathbf{L}^{(p)} = \sum_{h=1}^{h_0} \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{(p)}(h)\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{(p)}(h)^\top$, $\mathbf{L}^{(p)*} = \mathbf{A}_p\{\sum_{h=1}^{h_0} \mathbf{B}(h)\mathbf{B}(h)^\top\}\mathbf{A}_p^\top$, and $\mathbf{E}^{(p)} = \mathbf{A}_p\mathbf{B}(h)\boldsymbol{\Sigma}_{\mathbf{e}}^{(p)}(h)^\top + \boldsymbol{\Sigma}_{\mathbf{e}}^{(p)}(h)\mathbf{B}(h)\mathbf{A}_p^\top + \boldsymbol{\Sigma}_{\mathbf{e}}^{(p)}(h)\boldsymbol{\Sigma}_{\mathbf{e}}^{(p)}(h)^\top$, with $\mathbf{B}(h) = \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(p)}(h)\mathbf{A}_p^\top + \boldsymbol{\Sigma}_{\boldsymbol{\theta},\mathbf{e}}^{(p)}(h) + \boldsymbol{\Sigma}_{\mathbf{e},\boldsymbol{\theta}}^{(p)}(h)$. Then, a natural estimator for $\mathbf{A}_p$ is defined as $\widehat{\mathbf{A}}_p = (\widehat{\mathbf{a}}_{p,1}, \ldots, \widehat{\mathbf{a}}_{p,L})$, where $\widehat{\mathbf{a}}_{p,l}$ is the lth eigenvector of $\widehat{\mathbf{L}}^{(p)}$, and $\widehat{\mathbf{L}}^{(p)} = \sum_{h=1}^{h_0} \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}^{(p)}(h)\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}^{(p)}(h)^\top$, with $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}^{(p)}(h)$ being the empirical covariance matrix of $\tilde{\boldsymbol{\beta}}_{p,n}$ at lag h. Finally, the estimated factor time series $\widehat{\boldsymbol{\theta}}_{p,n}$ is obtained as $\widehat{\boldsymbol{\theta}}_{p,n} = \widehat{\mathbf{A}}_p^\top \tilde{\boldsymbol{\beta}}_{p,n}$. Then, we fit Autoregressive Moving Average (ARMA) models to each component of $\widehat{\boldsymbol{\theta}}_{p,n}$ using a maximum likelihood approach. We use the Akaike information criterion (AIC) to select the orders of the ARMA model.

3.2. Forecasting

Here, we describe how to forecast the continuous surface time series $x_t(\mathbf{s})$ at time $T+h_1$, $h_1 > 0$. Once again, without loss of generality, let us assume $\mu_t(\mathbf{s}) = 0$. Thus, we need to forecast $\widehat{f}_{N+h_2}^{(k)}$, where $h_2 = \lceil h_1/\delta \rceil$. From model (5), we have that the temporal dependence is contained in the factor process $\{\boldsymbol{\theta}_{p,n}\}_n$. Since the factors are mutually uncorrelated, we model each component of $\{\boldsymbol{\theta}_{p,n}\}_n$ separately for $p = 1, \ldots, p_0$. This implies estimating Lp_0 ARMA models. With the ARMA models, we obtain the forecasted factors $\widehat{\boldsymbol{\theta}}_{p,N+h_2}$; thus, we obtain $\widehat{\tilde{\boldsymbol{\beta}}}_{p,N+h_2}$. Then, we obtain the h_2-step-ahead forecast at time N of the functional time series: $\widehat{f}_{N+h_2}^{(k)}(u) = \widehat{\boldsymbol{\lambda}}^{(k)}(u)^\top \widehat{\tilde{\boldsymbol{\beta}}}_{N+h_2}^{(k)}$. The forecasted K functional time series $\{\widehat{f}_{N+h_2}^{(k)}(u);\ k = 1, \ldots, K\}$ contain the coefficient values $b_{k,T+h_1}$, which are needed to obtain $\widehat{x}_{T+h_1}(\mathbf{s})$. Expressly, the forecasted coefficients are defined as

$$\widehat{b}_{k,T+h_1} = \widehat{f}_{N+\lceil h_1/\delta \rceil}^{(k)}(h_1 - \delta\lfloor h_1/\delta \rfloor), \quad k = 1, \ldots, K. \tag{6}$$

Finally, the mean prediction of the continuous random field is

$$\widehat{x}_{T+h_1}(\mathbf{s}) = \sum_{k=1}^{K} \widehat{b}_{k,T+h_1}\psi_k(\mathbf{s}). \tag{7}$$

It is possible to construct a prediction interval of the surface $\widehat{x}_{T+h_1}$. This can be obtained from the prediction intervals of the curve $\widehat{f}_{N+h_2}^{(k)}$. Let $\widehat{f}_{N+h_2,\alpha}^{(k)}$ and $\widehat{f}_{N+h_2,1-\alpha}^{(k)}$ be the prediction bounds for $\widehat{f}_{N+h_2}^{(k)}$, where α is the level of significance. Then, $\widehat{x}_{T+h_1}^{\alpha}(\mathbf{s}) = \sum_{k=1}^{K} \widehat{f}_{N+\lceil h_1/\delta \rceil,\alpha}^{(k)}(h_1 - \delta\lfloor h_1/\delta \rfloor)\psi_k(\mathbf{s})$ and $\widehat{x}_{T+h_1}^{1-\alpha}(\mathbf{s}) = \sum_{k=1}^{K} \widehat{f}_{N+\lceil h_1/\delta \rceil,1-\alpha}^{(k)}(h_1 - \delta\lfloor h_1/\delta \rfloor)\psi_k(\mathbf{s})$ are the lower and upper prediction bounds,

respectively, with level of significance α. To obtain $\widehat{f}^{(k)}_{N+h_2,\alpha}$ and $\widehat{f}^{(k)}_{N+h_2,1-\alpha}$, we use the bootstrapping approach described in Gao et al. (2019).

Remark 2. If $\mu_t(\mathbf{s}) \neq 0$ and varies across time t, then the forecasting of $\mu_t(\mathbf{s})$ is similar to the forecasting of $z_t(\mathbf{s})$. If, in addition, $\mu_t(\mathbf{s})$ is constant in D, then $\widehat{\mu}_{T+h_1}$ can be obtained using a univariate functional time series forecasting method (Martínez-Hernández and Genton, 2021).

The procedure described above is implementable with the R-INLA and *ftsa* packages.

4. Simulation study

Here, we illustrate the advantage of our method. We present two simulation studies and compare our method with a separable space–time model that is often used in practice (e.g., Lenzi and Genton, 2020). The separable space–time model is defined as an SPDE model for the spatial domain and an AR(1) model for the time dimension. Explicitly, this model is defined as follows:

$$\begin{aligned} y_t(\mathbf{s}) &= \mu + \xi_t(\mathbf{s}) + \varepsilon_t(\mathbf{s}) \\ \xi_t(\mathbf{s}) &= a\xi_{t-1}(\mathbf{s}) + W_t(\mathbf{s}), \end{aligned} \tag{8}$$

where $\varepsilon_t(\mathbf{s})$ is the measurement error, $\xi_t(\mathbf{s})$ is a latent process following a first-order autoregressive dynamic model, and $W_t(\mathbf{s})$ is a temporally independent realization of a Gaussian random field with mean zero. The spatially correlated innovations $W_t(\mathbf{s})$ are characterized by the spatial covariance function, and this covariance function is defined by the Matérn function. We use the R-INLA implementation to fit model (8). In practice, model (8) is widely used for its computational advantages; see Cameletti et al. (2013) for more details.

In each simulation study, we compare our proposed methodology with model (8) in terms of their forecasting accuracy. For each study, we simulated 1000 Monte Carlo datasets.

4.1. Simulation setting

We simulate spatio-temporal data $\{y_t(\mathbf{s}); \mathbf{s} \in D, t = 1, \ldots, T\}$ from model (1). For that, we specify the mean μ_t and simulate $z_t(\mathbf{s})$ from (3), for which we must specify the domain D and the triangulation of D. We define D as the country of Saudi Arabia (see Fig. 1). Then, we obtain $\psi_1, \ldots, \psi_{K=78}$, which are the basis functions defined on the triangular mesh of D. We simulate the measurement error such that $\varepsilon_t(\mathbf{s}_i) \sim$ i.i.d. $\mathcal{N}(0, 0.25)$. We consider $T = 960$ time points. To fit the model, we evaluate $y_t(\mathbf{s})$ at 100 locations, $\mathbf{s}_1, \ldots, \mathbf{s}_{100}$, for $t = 1, \ldots, T$. These 100 locations are uniformly distributed over D. We have chosen these values to be able to compare our model and model (8). Although our model can deal with a larger dataset, we could not estimate model (8) for larger values of T.

To evaluate the ability of the model to correctly forecast, we split the data into a training set of 912 time points and a testing set with the remaining 48 time points. That is, we forecast the remaining 48 time points. Then, we quantify the mean squared error (MSE) forecasting at the 100 locations: $\sum_{i=1}^{100}\{x_t(\mathbf{s}_i) - \widehat{x}_t(\mathbf{s}_i)\}^2/100$, $t = 913, \ldots, 960$.

4.2. Simulation 1

First, we simulate the coefficients of $z_t(\mathbf{s})$ using model (5) and set $\delta = 24$, which results in $N = T/\delta = 40$ curves. That is, we simulate multivariate functional time series $\{\mathbf{f}_n(u); u \in [0, 1], n = 1, \ldots, N = 40\}$, with components $f^{(k)}_n(u) = \sum_{p=1}^{3} \beta^{(k)}_{p,n}\lambda^{(k)}_p(u)$. The eigenfunctions are defined as $\lambda^{(k)}_1(u) = \sin(2\pi u + \pi k/2)$, $\lambda^{(k)}_2(u) = \cos(2\pi u + \pi k/2)$, and $\lambda^{(k)}_3(u) = \sin(4\pi u + \pi k/2)$. The coefficients $\boldsymbol{\beta}_{p,n} = (\beta^{(1)}_{p,n}, \ldots, \beta^{(K)}_{p,n})^\top$ are simulated as $\boldsymbol{\beta}_{p,n} = \mathbf{A}_p\boldsymbol{\theta}_{p,n}$, where $\mathbf{A}_p = (a^{(p)}_{ij})_{1\leq i,j\leq K}$ with $a^{(p)}_{ij} = K^{-1/4}b^{(p)}_{ij}$, $b^{(p)}_{ij} \sim$ i.i.d. $\mathcal{N}(2, 4)$ for $p = 1, 2$, and $b^{(3)}_{ij} \sim$ i.i.d. $\mathcal{N}(0, 0.04)$. The factor time series $\boldsymbol{\theta}_{p,n}$ are simulated from an autoregressive model of order 1 and order 2, i.e, AR(1) and AR(2). Each component $\theta^{(l)}_{p,n}$ of the factor time series $\boldsymbol{\theta}_{p,n}$ is simulated as follows: $\theta^{(1)}_{p,n} = 0.5\theta^{(1)}_{p,n-1} + 0.2\theta^{(1)}_{p,n-2} + \epsilon_n$.

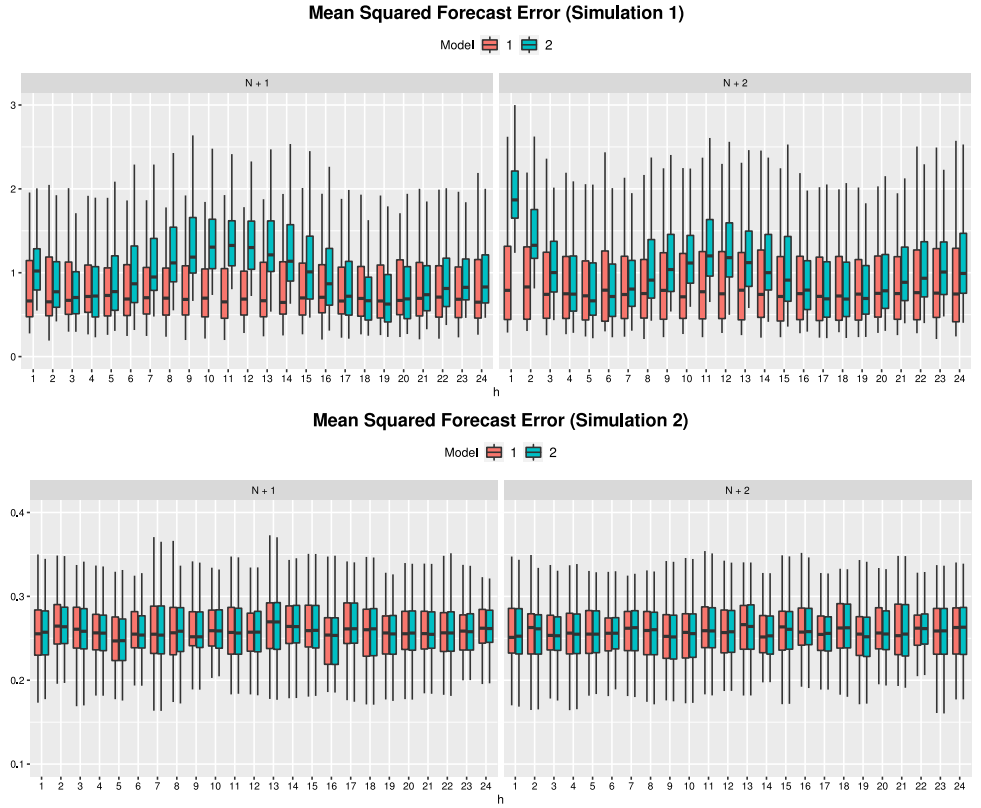


Fig. 2. Boxplots of MSE forecasting values from simulations 1 and 2. The MSE values correspond to 48 time points forecasted with Model 1 (our proposal) and Model 2. The index $N+1$ represents the first 24 time points and $N+2$ represents the rest. In simulation 1, Model 1 presents the lowest MSE values; in simulation 2, both models perform similarly.

For $l = 2, \ldots, K$, $\theta_{p,n}^{(l)} = K^{-1} z_n^{(l)}$, with $z_n^{(l)} = 0.2 z_{n-1}^{(l)} + \epsilon_n$ and ϵ_n representing Gaussian white noise with unit variance in both cases. Second, we define the mean component μ_t as evaluations of the mean function $\mu(u) = 3u\cos(u\pi) + 3.6u + 12$, $u \in [0, 1]$. Finally, the random field is obtained as $x_t(\mathbf{s}) = \mu_t + \sum_{k=1}^{K} b_{k,t}\psi_k(\mathbf{s})$, where $b_{k,t}$ are evaluations of $f_n^{(k)}(u)$.

Notice that we simulate factors as many as the number of bases, $L = K$, where the kth-factor contribution is down-weighted by $1/K$, except for $l = 1$. This simulation provides a more complex structure mimicking real datasets. However, when estimating model (5), we estimate only three factors, $L = 3$.

4.3. Simulation 2

For this scenario, we simulate $x_t(\mathbf{s})$ from model (8). That is, we define $x_t(\mathbf{s}) = \mu + \xi_t(\mathbf{s})$, where $\mu = 10$. The process $\xi_t(\mathbf{s})$ is simulated as $\xi_t(\mathbf{s}) = 0.8\xi_{t-1}(\mathbf{s}) + \sqrt{1 - 0.8^2}\, W_t(\mathbf{s})$, where $\xi_1(\mathbf{s}) = W_1(\mathbf{s})$. We generate the T independent realizations of the spatial model $W_t(\mathbf{s})$ using a Whittle covariance function with variance 1 and scaling parameter 0.1. This corresponds to a range of 28.2, approximately. We add the term $\sqrt{1 - 0.8^2}$ to make the process stationary in time. Although our method does not require stationarity in time, model (8) does.

4.4. Simulation results

Here, we present the simulation results from the two models in the two scenarios described above. Fig. 2 shows the MSE forecasting values for the 48 time points forecasted. In each panel (scenario), we present the performance of both models side-by-side. Overall, our methodology (labeled Model 1) has a good performance compared with model (8) (labeled Model 2).

In simulation 1 (first row of Fig. 2), we observe that Model 1 (our proposal) presents lower MSE forecasting values in all periods ($N+1$ and $N+2$). These periods correspond to the 48 time points forecasted. The average of the MSE values for Model 1 in periods $N+1$ and $N+2$ are 0.915 and 1.012, respectively. For Model 2, the average of the MSE values in periods $N+1$ and $N+2$ are 1.24 and 1.228, respectively. Therefore, we conclude that Model 1 outperforms Model 2. This conclusion comes as no surprise since the data-generating process is Model 1.

In the scenario of simulation 2 (second row of Fig. 2), both models perform similarly in all periods ($N+1$ and $N+2$), i.e., across all 48 time points. The average of the MSE values for Model 1 in periods $N+1$ and $N+2$ are 0.26 and 0.259, respectively. For Model 2, we obtain the same average values as Model 1. In this case, the data-generating process is Model 2. Thus, the fact that our methodology performs as well as Model 2 can be seen as an advantage of our model.

In conclusion, based on the two simulations, our methodology outperforms the spatio-temporal model most commonly used in practice. Furthermore, our methodology provides extensive flexibility to model nonstationary, complex, and large datasets.

5. Wind data analysis

5.1. Data description

We use wind data simulated by Yip (2018) from the Weather Research and Forecasting (WRF) model. Each measurement corresponds to hourly wind speed (m/s) from the year 2010 over Saudi Arabia. Originally, the wind speeds were simulated on a regular grid of points in space, namely at 5-km resolution. For our aim, we select 1500 locations randomly over Saudi Arabia and analyze four months of data: May, June, July, and August. Thus, our dataset contains 1500 locations and 2952 time points, so a total of $4,428,000$ values. The 1500 locations are shown in the left panel of Fig. 1. Let $\text{ws}_t(\mathbf{s}_i)$ denote the wind speed at location $\mathbf{s}_i$ and time t. Then, for each $\mathbf{s}_i$, $i = 1, \ldots, 1500$, $\{\text{ws}_t(\mathbf{s}_i)\}_t$ is an hourly time series of length $T = 2952$. To use models (2) and (3), we consider the square-root transformation of the wind speed. The transformed wind speed resembles the Gaussian distribution (Haslett and Raftery, 1989). We denote by $y_t(\mathbf{s}_i) = \sqrt{\text{ws}_t(\mathbf{s}_i)}$ the square-root transformed wind speed.

The understanding of wind speed behavior is important for renewable energy. Wind energy is one of the most efficient renewable energy sources available, but the availability of this resource depends on location. To ensure reliability and quality of a power system, it is important to develop highly accurate wind speed prediction methods. Wind speed forecasting is classified depending on the time-scale horizons to be forecasted. These can be grouped into short-term (30 min to 6 h ahead), medium-term (6 h to 1 day ahead), and long-term (1 day to 1 week ahead) forecasting. Several approaches to wind speed prediction models have been proposed in the literature, such as models with time series techniques, spatial statistics techniques, or both that focus on predicting a specific time-scale horizon. A few attempts to propose a model for forecasting short and medium-term time scales can be found in the literature (see, e.g., Lenzi and Genton, 2020, and references therein). In contrast, our model can forecast short, medium, and long-term time scales without any additional modification. Our goal is to accurately forecast wind speed data for short, medium, and long-term time scale horizons.

The wind speed dataset that we consider here is large and complex in space and time. Thus, these data fit into our proposed methodology described in Section 2. The advantage of our method (or a statistical model) over the WRF model is mainly the computational cost. We define the time period δ in (4) such that n represents days, i.e., $\delta = 24$ (hours). This choice of δ adapts to the diurnal cycle, an external factor, which is desirable in the model (Weisser and Foxon, 2003). To evaluate

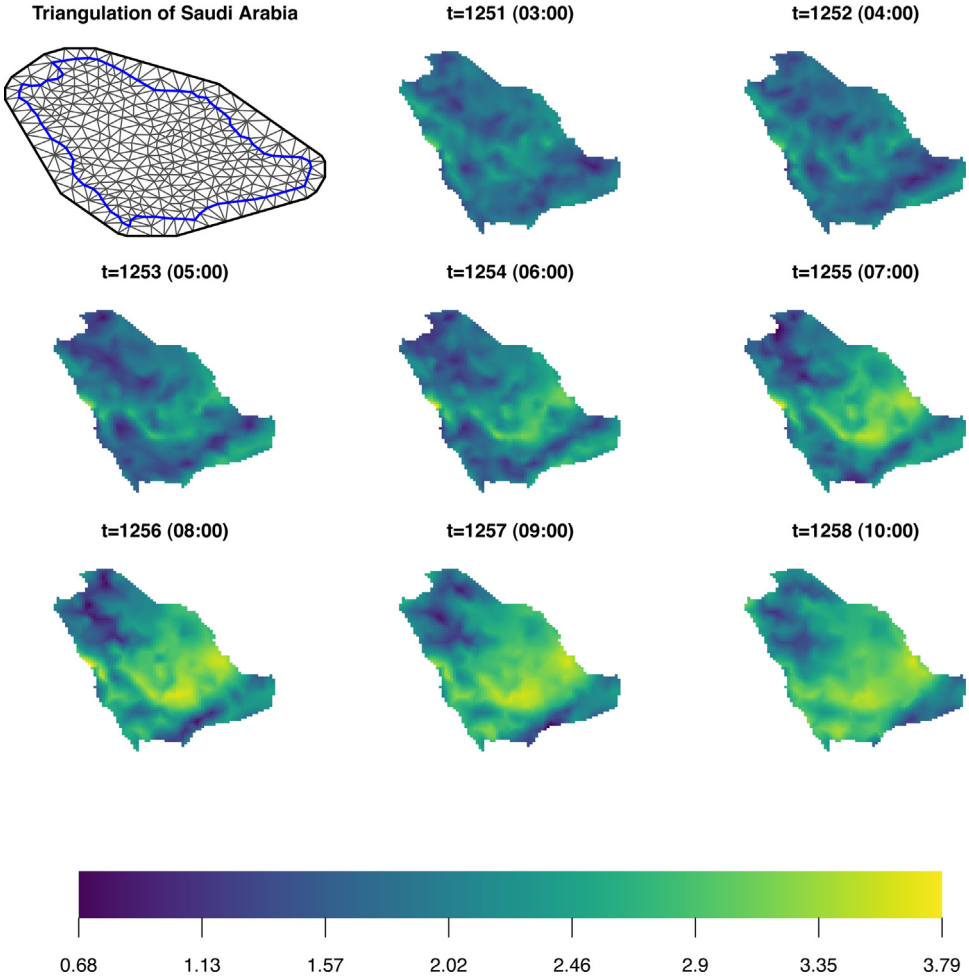


Fig. 3. A mesh for Saudi Arabia with 364 nodes: $\psi_1, \ldots, \psi_{364}$ (top left), and eight consecutive continuous random fields, corresponding to June 22, 2010, estimated with the basis representation (3). Each continuous random field has the form $x_t(\mathbf{s}) = \mu_t + \sum_{k=1}^{364} b_{k,t}\psi_k(\mathbf{s})$, with $t = 1251, \ldots, 1258$ representing hours.

the performance of the forecast, we use the four months of data to fit the model, and then we forecast the first three days of September, i.e., we forecast 72 h ahead. The first step is to define a triangulation of the domain, which is the country of Saudi Arabia. The triangulation used in the data analysis is displayed in the first panel of Fig. 3. Once we have defined a triangulation of our domain, we proceed to estimate the continuous random field. We detail this estimation in the following.

5.2. Model specifications

To estimate the continuous wind speed random field, we use 364 basis functions, $\psi_1, \ldots, \psi_{364}$, determined by the triangulation. We set the range prior such that $P(\rho_t < 500) = 0.5$, and the standard deviation such that $P(\sigma_{x,t} > 10) = 0.01$. The range prior is based on the characteristics of Saudi Arabia's landscape. Fig. 3 shows the result of eight consecutive continuous random fields: $x_t(\mathbf{s}) = \mu_t + \sum_{k=1}^{364} b_{k,t}\psi_k(\mathbf{s})$, with t representing hours and corresponding to June 22, 2010, at 03:00,

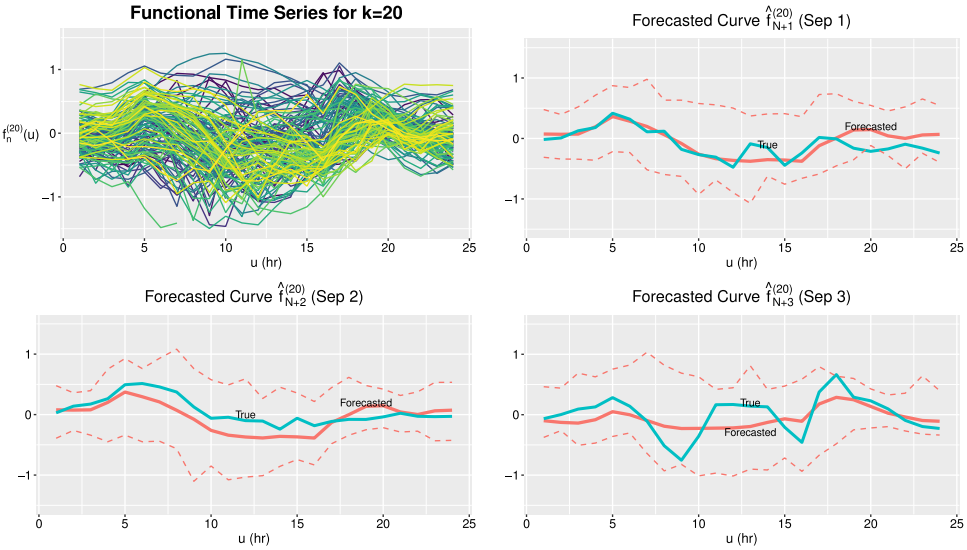


Fig. 4. A functional time series and its forecasts. Top left: Functional time series $\{f_n^{(20)}\}_n$ that corresponds to the basis function ψ_{20}, and $n = 1, \ldots, N = 123$ (days). Top right and second row: The forecasted curves corresponding to September 1, 2, and 3, 2010, for the basis component $k = 20$. These forecasted curves are obtained from model (5). The dashed curves represent the lower and upper prediction bounds.

04:00, …, 10:00 am. After estimating the continuous random fields, we transform the coefficients $\{b_{k,t}\}$ into a functional time series f_n^k. As mentioned above, we define the time period as $\delta = 24$ (hours) to account for the diurnal cycle. Thus, the sample size of the functional time series is $N = 123$. Fig. 4 shows one resulting functional time series corresponding to the basis component $k = 20$, i.e., the functional time series $\{f_n^{(20)}(u);\ u \in \mathcal{T}, n = 1, \ldots, 123\}$. In the top left panel, we observe that this functional time series shows a repeated pattern over many days that fluctuates throughout the day. This time (day)-dependent behavior is observed across all functional time series $\{f_n^{(k)}\}_n$, $k = 1, \ldots, K = 364$. Then, we proceed to describe the temporal dependence with a functional time series model, and then we forecast the next three days: September 1, 2, and 3, 2010.

We use the functional time series $\mathbf{f}_n = \{f_n^{(1)}(u), \ldots, f_n^{(364)}(u)\}^\top$ to model the temporal dependence. In the dynamic FPC representation, model (5), we set $p_0 = 3$ and the number of factors $L = 3$. The selection of p_0 is based on the proportion of the variability of the functional time series explained. In general, it is common to have values of p_0 smaller than five in real data applications. As mentioned above, the ARMA model is estimated using likelihood and the orders are selected via AIC. Using the forecasted values $\widehat{\widetilde{\boldsymbol{\beta}}}_{p,N+1}$, $\widehat{\widetilde{\boldsymbol{\beta}}}_{p,N+2}$, and $\widehat{\widetilde{\boldsymbol{\beta}}}_{p,N+3}$, we obtain the forecasted curves $\hat{f}_{N+1}$, $\hat{f}_{N+2}$, and $\hat{f}_{N+3}$ corresponding to September 1, 2, and 3, 2010.

5.3. Results

Fig. 4 shows the forecasted curves of the functional time series $f_n^{(20)}$. For each forecasted curve, we also plot the "true" coefficient curves. These true curves are the curves obtained from the coefficients of the estimated surfaces $x_{T+h}(\mathbf{s})$ for September 1, 2, and 3, 2010. The forecasted curves are expected to be close to the true curves. Thus, we can evaluate the performance of our model. From Fig. 4, we observe that the forecasted curves mimic the main behavior of the true curves. This means that the forecasted surfaces are expected to be accurate in representing the spatial dependence and temporal dependence.

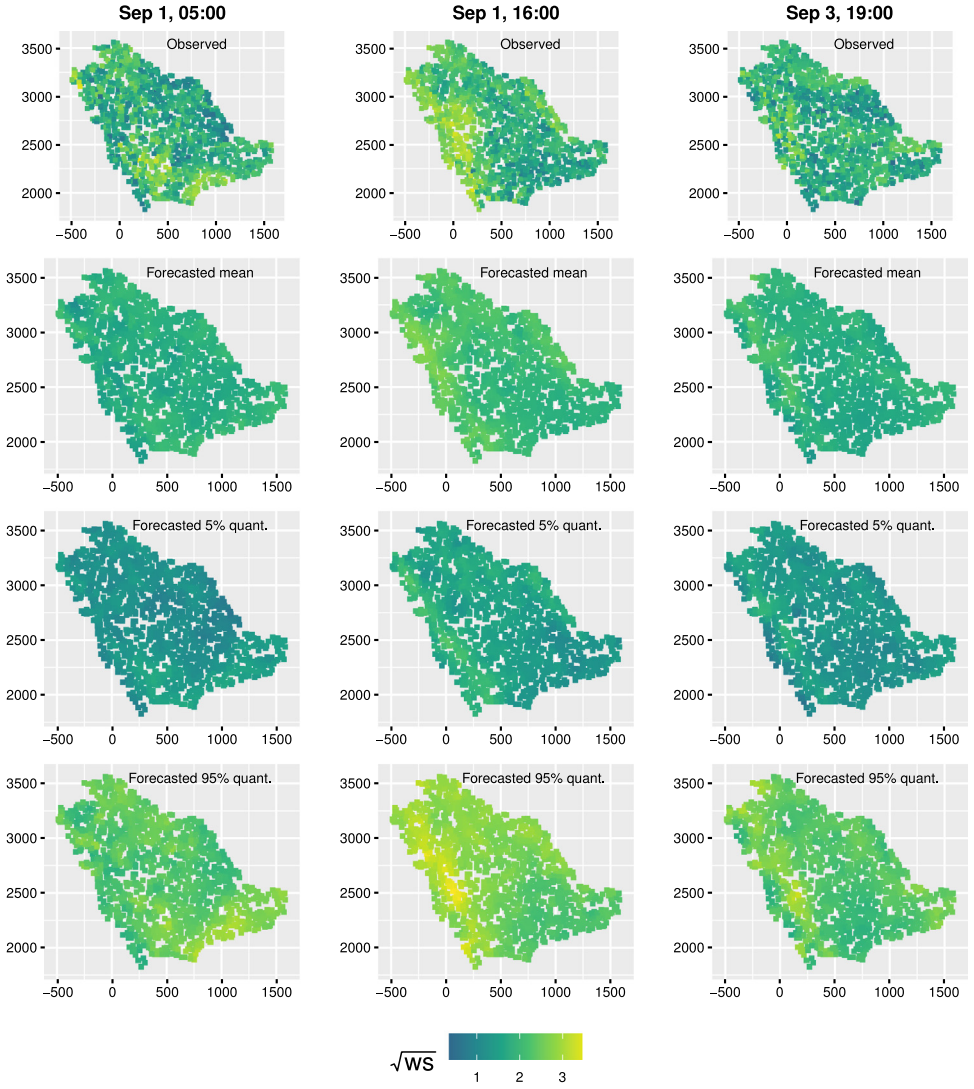


Fig. 5. Results of the forecasted values at three different times. The first row shows the square-root of the observed wind speed values at 1500 locations. The second row shows the forecasted values which preserve the same spatial pattern as the observed values. The third and fourth rows show the 5% and 95% quantiles of the forecast values, respectively.

We can now evaluate the forecasted curves at any value of the domain $\mathcal{T}$ to obtain the forecasted continuous wind speed field. We present three forecasted continuous wind speed fields at different times: September 1 at 05:00, September 1 at 16:00, and September 3 at 19:00. This means that we evaluate $\widehat{\mathbf{f}}_{N+1}$ at $u = 5$ and $u = 16$, and $\widehat{\mathbf{f}}_{N+3}$ at $u = 19$. These three time horizons correspond to short, medium, and long-term predictions, respectively. Fig. 5 shows an evaluation of the three surfaces forecasted at the 1500 locations where data are observed. Also, Fig. 5 shows the corresponding real wind speed data (which was not used to fit the model). We observe that the forecasted values present the same variation as the original data. For example, the observed wind speed values for September 1 at 16:00 (second column) are high in the northwest of the country (area of Medina). These high values extend along the coast of the Red Sea; whereas in the center of

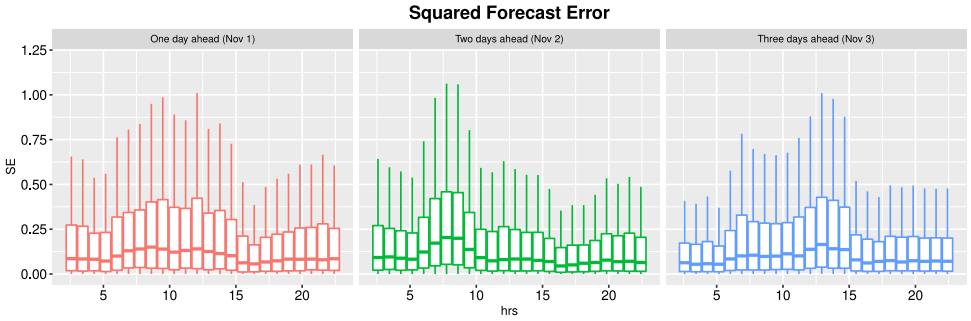


Fig. 6. Error values from the forecasted values. At each hour on September 1, 2, and 3, 2010, we present a boxplot of the values $\{y_t(\mathbf{s}_i) - \widehat{x}_t(\mathbf{s}_i)\}_i^2$, where $\widehat{x}_t$ is the forecasted surface.

the country, the wind speed values are low. We obtain this same variation on the forecasted mean values. We see that the highest values forecasted are in the northwest region and along the Red Sea coast. This conclusion holds for all 72 forecasted hours.

To gain a better perspective on our method's performance, we compute the squared forecast error at each hour on September 1, 2, and 3, 2010. That is, we compute $\{y_t(\mathbf{s}_i) - \widehat{x}_t(\mathbf{s}_i)\}_i^2$ for each t, where i is the location's index where data are observed, and $\widehat{x}_t$ is the forecasted surface. Fig. 6 shows the boxplots of the squared forecast error values at each hour for one, two, and three days ahead, corresponding to September 1, 2, and 3, 2010, respectively. We observe that, for September 1 and 3, the highest error values are obtained from these boxplots around noon, whereas for September 2, the highest error values are obtained between 6:00 am and 9:00 am. This could be due to high wind speed when the sun's heat creates convection currents causing the observed and forecasted values to differ more. Despite these error values, the general behavior of the data is recovered, as illustrated in Fig. 5. Thus, we conclude that our methodology has a good performance.

6. Discussion

An important limitation of many spatio-temporal models is their inability to work with large datasets. Because of this inability, among other things, a new paradigm of data analysis is required. Along this line, we have proposed a new approach to model large and complex spatio-temporal data. This approach assumes that at each time point, the observation is a continuous random field. The continuous random fields are estimated using a finite basis representation, where the coefficients of the basis functions are estimated assuming that the random fields have a Matérn covariance structure. Then, we model the sequence of surfaces using multivariate functional time series methodology.

Our methodology can be applied to a large class of data and is not limited to the specific case of wind speed data. Furthermore, it can model data that are nonstationary in space and time. This is because the continuous surface estimation does not necessarily require stationarity in space, and the coefficients of the basis functions are modeled with a functional time series methodology (the stationarity condition is across the index n and not across t). Additionally, our methodology can handle changes in locations and number of locations over time.

A limitation of our methodology is the uncertainty quantification. For example, with our methodology, we cannot obtain a closed-form expression of the variance of the error prediction $x_{T+h_1}(\mathbf{s}) - \widehat{x}_{T+h_1}(\mathbf{s})$; this is in contrast to kriging, where we have a closed-form expression. One possible and logical solution is to use a bootstrap method to obtain the uncertainty quantification. Another limitation of our methodology is the estimation of the number of factors L to be used. Although it is common in practice to set $L = 2$ or 3, which usually works well, it is necessary to have a statistical method to select L.

We conclude that our methodology presents a valuable alternative to model large spatio-temporal data.

References

Aue, A., Norinho, D.D., Hörmann, S., 2015. On the prediction of stationary functional time series. J. Amer. Statist. Assoc. 110 (509), 378–392.

Bernardi, M.S., Sangalli, L.M., Mazza, G., Ramsay, J.O., 2017. A penalized regression model for spatial functional data with application to the analysis of the production of waste in Venice province. Stoch. Environ. Res. Risk Assess. 31 (1), 23–38.

Cameletti, M., Lindgren, F., Simpson, D., Rue, H., 2013. Spatio-temporal modeling of particulate matter concentration through the SPDE approach. AStA Adv. Stat. Anal. 97 (2), 109–131.

Chen, W., Genton, M.G., Sun, Y., 2021. Space–time covariance structures and models. Annu. Rev. Stat. Appl. 8 (1), 191–215.

Cressie, N., Johannesson, G., 2008. Fixed rank kriging for very large spatial data sets. J. R. Stat. Soc. Ser. B Stat. Methodol. 70 (1), 209–226.

Cressie, N., Wikle, C.K., 2011. Statistics for Spatio-Temporal Data. In: Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., Hoboken, NJ.

Fuglstad, G.-A., Simpson, D., Lindgren, F., Rue, H., 2019. Constructing priors that penalize the complexity of Gaussian random fields. J. Amer. Statist. Assoc. 114 (525), 445–452.

Furrer, R., Genton, M.G., Nychka, D., 2006. Covariance tapering for interpolation of large spatial datasets. J. Comput. Graph. Statist. 15 (3), 502–523.

Gao, Y., Shang, H.L., Yang, Y., 2019. High-dimensional functional time series forecasting: An application to age-specific mortality rates. J. Multivariate Anal. 170, 232–243.

Gelfand, A.E., Banerjee, S., Gamerman, D., 2005. Spatial process modelling for univariate and multivariate dynamic spatial data. Environmetrics 16 (5), 465–479.

Genton, M.G., 2007. Separable approximations of space–time covariance matrices. Environmetrics 18 (7), 681–695.

Giraldo, R., Delicado, P., Mateu, J., 2011. Ordinary kriging for function-valued spatial data. Environ. Ecol. Stat. 18 (3), 411–426.

Haslett, J., Raftery, A.E., 1989. Space–time modelling with long-memory dependence: Assessing Ireland's wind power resource. J. R. Stat. Soc. Ser. C. Appl. Stat. 38 (1), 1–50.

Hays, S., Shen, H., Huang, J.Z., 2012. Functional dynamic factor models with application to yield curve forecasting. Ann. Appl. Stat. 6 (3), 870–894.

Heaton, M.J., Datta, A., Finley, A.O., Furrer, R., Guinness, J., Guhaniyogi, R., Gerber, F., Gramacy, R.B., Hammerling, D., Katzfuss, M., Lindgren, F., Nychka, D.W., Sun, F., Zammit-Mangion, A., 2019. A case study competition among methods for analyzing large spatial data. J. Agric. Biol. Environ. Stat. 24 (3), 398–425.

Hörmann, S., Kidziński, L.u., Hallin, M., 2015. Dynamic functional principal components. J. R. Stat. Soc. Ser. B Stat. Methodol. 77 (2), 319–348.

Huang, H., Abdulah, S., Sun, Y., Ltaief, H., Keyes, D.E., Genton, M.G., 2021. Competition on spatial statistics for large datasets. J. Agric. Biol. Environ. Stat. 26 (4), 580–595.

Hyndman, R.J., Shang, H.L., 2020. ftsa: Functional time series analysis. R package version 5.8.

Hyndman, R.J., Ullah, M.S., 2007. Robust forecasting of mortality and fertility rates: a functional data approach. Comput. Statist. Data Anal. 51 (10), 4942–4956.

Kaufman, C.G., Schervish, M.J., Nychka, D.W., 2008. Covariance tapering for likelihood-based estimation in large spatial data sets. J. Amer. Statist. Assoc. 103 (484), 1545–1555.

Kokoszka, P., Reimherr, M., 2017. Introduction to Functional Data Analysis. In: Texts in Statistical Science Series, CRC Press, Boca Raton, FL.

Kuusela, M., Stein, M.L., 2018. Locally stationary spatio-temporal interpolation of argo profiling float data. Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 474 (2220), 20180400.

Lam, C., Yao, Q., Bathia, N., 2011. Estimation of latent factors for high-dimensional time series. Biometrika 98 (4), 901–918.

Lenzi, A., Genton, M.G., 2020. Spatiotemporal probabilistic wind vector forecasting over Saudi Arabia. Ann. Appl. Stat. 14 (3), 1359–1378.

Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J. R. Stat. Soc. Ser. B Stat. Methodol. 73 (4), 423–498.

Martínez-Hernández, I., Genton, M.G., 2020. Recent developments in complex and spatially correlated functional data. Braz. J. Probab. Stat. 34 (2), 204–229.

Martínez-Hernández, I., Genton, M.G., 2021. Nonparametric trend estimation in functional time series with application to annual mortality rates. Biometrics 77 (3), 866–878.

Martínez-Hernández, I., Gonzalo, J., González-Farías, G., 2022. Nonparametric estimation of functional dynamic factor model. J. Nonparametr. Stat. 34 (4), 895–916.

Matérn, B., 1986. Spatial Variation, second ed. In: Lecture Notes in Statistics., vol. 36, Springer-Verlag, Berlin, With a Swedish summary.

Menafoglio, A., Grujic, O., Caers, J., 2016. Universal kriging of functional data: Trace-variography vs cross-variography? Application to gas forecasting in unconventional shales. Spatial Stat. 15, 39–55.

Nguyen, H., Katzfuss, M., Cressie, N., Braverman, A., 2014. Spatio-temporal data fusion for very large remote sensing datasets. Technometrics 56 (2), 174–185.

Nychka, D., Hammerling, D., Krock, M., Wiens, A., 2018. Modeling and emulation of nonstationary Gaussian fields. Spatial Stat. 28, 21–38.

R Core Team, 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Ramsay, J.O., Silverman, B.W., 2005. Functional Data Analysis, second ed. In: Springer Series in Statistics., Springer.
Rice, G., Shang, H.L., 2017. A plug-in bandwidth selection procedure for long-run covariance estimation with stationary functional time series. J. Time Series Anal. 38 (4), 591–609.
Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. R. Stat. Soc. Ser. B Stat. Methodol. 71 (2), 319–392.
Sangalli, L.M., Ramsay, J.O., Ramsay, T.O., 2013. Spatial spline regression models. J. R. Stat. Soc. Ser. B Stat. Methodol. 75 (4), 681–703.
Sigrist, F., Künsch, H.R., Stahel, W.A., 2012. A dynamic nonstationary spatio-temporal model for short term prediction of precipitation. Ann. Appl. Stat. 6 (4), 1452–1477.
Simpson, D., Rue, H., Riebler, A., Martins, T.G., Sørbye, S.H., 2017. Penalising model component complexity: A principled, practical approach to constructing priors. Statist. Sci. 32 (1), 1–28.
Stein, M.L., 1999. Interpolation of Spatial Data: Some Theory for Kriging. In: Springer Series in Statistics, Springer-Verlag, New York.
Sun, Y., Li, B., Genton, M.G., 2012. Geostatistics for large datasets. In: Porcu, E., Montero, J.-M., Schlather, M. (Eds.), Advances and Challenges in Space–Time Modelling of Natural Events. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 55–77.
Weisser, D., Foxon, T., 2003. Implications of seasonal and diurnal variations of wind velocity for power output estimation of a turbine: a case study of grenada. Int. J. Energy Res. 27 (13), 1165–1179.
Wikle, C.K., Cressie, N., 1999. A dimension-reduced approach to space–time Kalman filtering. Biometrika 86 (4), 815–829.
Wikle, C.K., Hooten, M.B., 2010. A general science-based framework for dynamical spatio-temporal models. TEST 19 (3), 417–451.
Wood, S.N., 2006. Low-rank scale-invariant tensor product smooths for generalized additive mixed models. Biometrics 62 (4), 1025–1036.
Wood, S., 2017. Generalized Additive Models: An Introduction with R, second ed. Chapman and Hall/CRC.
Xiao, L., Li, Y., Ruppert, D., 2013. Fast bivariate *P*-splines: the sandwich smoother. J. R. Stat. Soc. Ser. B Stat. Methodol. 75 (3), 577–599.
Yip, C.M.A., 2018. Statistical Characteristics and Mapping of Near-Surface and Elevated Wind Resources in the Middle East (Ph.D. thesis). King Abdullah University of Science and Technology.