Spatiotemporal probabilistic wind vector forecasting over Saudi Arabia

<table>
<thead>
<tr>
<th>Item Type</th>
<th>Article</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authors</td>
<td>Lenzi, Amanda; Genton, Marc G.</td>
</tr>
<tr>
<td>Eprint version</td>
<td>Publisher's Version/PDF</td>
</tr>
<tr>
<td>DOI</td>
<td>10.1214/20-AOAS1347</td>
</tr>
<tr>
<td>Publisher</td>
<td>Institute of Mathematical Statistics</td>
</tr>
<tr>
<td>Journal</td>
<td>Annals of Applied Statistics</td>
</tr>
<tr>
<td>Rights</td>
<td>Archived with thanks to Annals of Applied Statistics</td>
</tr>
<tr>
<td>Download date</td>
<td>29/09/2020 05:42:49</td>
</tr>
<tr>
<td>Link to Item</td>
<td><a href="http://hdl.handle.net/10754/665293">http://hdl.handle.net/10754/665293</a></td>
</tr>
</tbody>
</table>
Spatio-Temporal Probabilistic Wind Vector Forecasting over Saudi Arabia

Amanda Lenzi\(^1\) and Marc G. Genton\(^\dagger\)

September 28, 2020

Abstract

Saudi Arabia has recently begun promoting renewable energy as a potential alternative to fossil fuels for domestic power generation. In order to efficiently connect wind energy to the existing power grids, reliable wind forecasts and an accurate way of quantifying the uncertainties of these forecasts are required. Motivated by a data set of hourly wind speeds from 28 stations in Saudi Arabia, we build spatio-temporal models for short-term probabilistic forecasts of wind vectors. Traditionally, wind speed and wind direction have been considered independently, without taking dependencies into account. However, in many situations, e.g. energy management, it is essential to have information on the bivariate nature of the wind. We compare a coregionalization model for the wind vector with a univariate spatio-temporal model for the transformed wind speed in terms of sharpness and calibration. In both cases, the linear predictor is a function of covariates, a smooth function to capture the daily seasonality in the wind and a latent Gaussian field to model the spatial and temporal dependencies. Substantial improvements in reliability are observed when modelling the full bivariate structure instead of only considering speed. Furthermore, the bivariate model has the advantage of also producing forecasts for the wind direction. A Bayesian framework is used to obtain forecasts that are accurate and reliable, even at stations without observations, with relatively low computational cost. Simulated high-resolution data from a computer model are used to validate spatio-temporal forecasts. A detailed analysis on this case study shows how increasing the number of locations can improve the forecast performance.

Keywords: Integrated nested Laplace approximation, Linear model of coregionalization, Probabilistic forecast, Spatio-temporal modeling, Wind vector

Short title: Wind Vector Forecasting Over Saudi Arabia

\(^1\)Statistics Program, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia. E-mail: lenzi.amanda88@gmail.com, marc.genton@kaust.edu.sa

This publication is based on research supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No: OSR-2018-CRG7-3742.
1 Introduction

Saudi Arabia, like other countries, is undertaking massive efforts to create a more sustainable future. With the rapid population and industrial growth, the country is looking for alternatives to reduce its dependence on oil while still meeting its domestic demand in energy without having to shrink the volume of oil available for exports. As part of the “Vision 2030”, a package of economic and social polices designed to diversify the country’s economy, Saudi Arabia has set the ambitious goal of installing 9GW of wind energy as part of the 54GW renewable energy totally installed by 2032 (Vision2030, 2018). When seeking investors and sources of financing, estimates of the potential wind energy generation and advanced forecast methodologies are essential.

Due to the inherent variability of the wind, the efficient integration of wind energy into existing power grids requires high-quality forecasts, together with a quantification of the uncertainties (Ackermann and Thomas, 2005). However, accurately forecasting wind and quantifying the uncertainties of the forecasts is a challenging problem. The first challenge relates to the distribution of wind speed, which is bounded below by zero and significantly right-skewed. As a result, wind speed cannot be modeled directly by using the Gaussian distribution. A second difficulty is the lack of data bases from monitoring wind stations in Saudi Arabia. Data on the power generated are not yet available and the wind data are currently sparse in space, with only a few locations being consistently monitored throughout time. Until now, the literature on wind forecasting in Saudi Arabia is limited to systems that are optimized for individual locations separately; therefore, conclusions can only be reached where data are available (Mohandes et al., 1998, 2004). Other studies attempted to quantify the abundance of wind resources in the Arabian peninsula by using wind speed data from a maximum of 20 locations (Rehman and Ahmad, 2004, Rehman et al., 2007, Shaahid et al., 2014). Gridded analysis can be produced with reanalysis data, a combination of numerical model outputs and observational data. Because of the limited data from stations, reanalyzes data have been recently used for quantifying the potential wind speed and wind energy in Saudi Arabia (Yip et al., 2016, Tagle et al., 2017, Chen et al., 2018).
Treating the speed and the direction of wind as a bivariate process, i.e., by its west-east and north-south components, is needed yet has been mostly overlooked. The need for simultaneously modeling is not unique to the production of wind energy, but also applies widely to air traffic control, ship routing and energy management (Pinson and Madsen, 2009). Failure to model simultaneously with the exclusion of wind direction and the reliance solely on speed brings a systematic error into forecasting (Mandic et al., 2005; Alexiadis et al., 1999). Ensemble forecasts that combine multiple estimates of Numerical Weather Prediction (NWP) models initialized with different values were used to forecast wind vectors at medium horizons (i.e., 6 hours to 1 day ahead) by Schuhen et al. (2012), Pinson (2012), McLean Sloughter et al. (2013) and Eide et al. (2017). Hering and Genton (2010) modeled the wind vectors at each location separately with a bivariate skew-t distribution, which was only intended for a small number of stations, three, in their case. They considered each station separately and could therefore only forecast in time at the stations where data were available, whereas we build a spatio-temporal model and obtain forecasts even at stations without observations. Subsequently, we validate our method on a much more extensive portfolio of up to 242 stations.

Wind is affected by many physical factors with complex structures in space and time. Concerning spatial dependence, prevailing weather patterns usually propagate in a particular direction resulting in non-stationary properties, which can either be modelled through the mean or the covariance structure. Accounting for non-stationarity in the covariance function is difficult and can lead to model over-fitting (Fuglstad et al., 2015). From a practical point-of-view, however, forecasting systems designed for short ranges need to be able to generate efficient online updates whenever new data become available. In this paper, we use a set of covariates and a non-linear trend to capture the spatially varying structure through the mean function and choose a relatively simple covariance model to overcome the computational burden of fitting complex covariance structures. In terms of temporal dependence, it is desirable to have a model that adapts to external factors such as the diurnal cycle. For this purpose, models based on exponentially
forgetting past observations were applied on wind power data in Möller et al. (2008) and Pinson and Madsen (2012), and an adaptive filter for forecasting hourly wind speed was proposed in Dowell et al. (2014). We achieve adaptivity by allowing the parameters to vary depending on the data set that has been modeled using a sliding window technique similarly to Dowell and Pinson (2016). The model parameters are updated based only on the most recently available data and each time these data are made available. In this way, we can obtain computationally efficient spatio-temporal forecasts for the full wind vector, even for a large number of locations.

Motivated by a data set with hourly observations from 28 wind stations in Saudi Arabia, we present statistical models for probabilistic forecasting of wind vectors at several sites and future times (denoted here by lead time). The bivariate spatio-temporal process is modeled as a function of a common intercept, covariates, a smooth function to explain the daily seasonality in wind and a latent Gaussian field to capture the spatial and temporal dependencies. The cross-covariance function is based on a linear combination of independent univariate spatio-temporal processes, also known as linear model of coregionalization (LMC). This approach provides valid covariance models that are richer and more flexible than the separable or intrinsic structures because different parameters represent each component of the bivariate process (Matheron, 1982). The use of the LMC for modeling spatial data can be found in a variety of applications, e.g., Schmidt and Gelfand (2003), Gelfand et al. (2004), Banerjee et al. (2014), and in the review by Genton and Kleiber (2015). Fully Bayesian inference and prediction are performed by means of a stochastic partial differential equation (SPDE) approach and the Integrated nested Laplace approximation (INLA) (Rue et al., 2009). We compare the performance of the bivariate model with a univariate spatio-temporal model, where wind direction is introduced through covariates. Besides the originality of considering the wind components jointly, the main novelty of this work lies in the formulation of a spatio-temporal LMC as part of a more general model, which utilizes the state-of-the-art INLA-SPDE techniques to construct the Gaussian random fields (GRFs) that encode the dependence structures. The combination of these two approaches results in low-cost
computations and still provides calibrated probabilistic spatio-temporal forecasts.

In particular for the case of Saudi Arabia, where data from only a few stations are available, it is essential to provide spatio-temporal forecasts, i.e., at stations that do not contribute to the validation set. The early wind literature has only provided point forecasts of wind speed. However, decision-making problems in which the loss function is asymmetric may benefit from different quantiles, hence requiring probabilistic forecasts as the input (Bremnes 2004; Pinson and Kariniotakis 2010). We compute a full forecast distribution for each location and lead time and construct maps of wind speed for the different quantiles at a high spatial resolution for the entire country and several hours ahead with minimal computational time. These maps are crucial for operational problems, in which observations are limited to a number of locations, even though decision-making problems require an outline of the wind speed at every station inside a region.

In a second scheme, we validate spatio-temporal forecasts from our bivariate model on high-resolution simulated data from a computer model. We use a unique data set from a 5-kilometer resolution state-of-the-art NWP model, the Weather Research and Forecast (WRF) model built at the National Center for Atmospheric Research (NCAR), and run at the King Abdullah University of Science and Technology (KAUST). It is the first of its kind to be available at a spatial resolution fine enough to be able to represent small-scale topology, allowing a detailed analysis on how increasing the number of locations improves the spatio-temporal forecast performance. This study provides key insights that can be used to make recommendations for how many stations are necessary to have a good assessment of the wind process in an entire country.

The remainder of this paper is organized as follows. The wind data from Saudi Arabia used in our case study are described in Section 2. In Section 3, we present the univariate and bivariate hierarchical Bayesian models, as well as the scheme for obtaining probabilistic forecasts of wind speed and direction. We give details of the implementation and explain how we assess reliability and sharpness of the forecasts’ distributions in Section 4. Results from fitting our models to the...
wind data from Saudi Arabia are described in Section 5 and a forecast validation study is given in Section 6. Finally, conclusions and a discussion of future work are provided in Section 7.

2 Wind Data from Saudi Arabia

We analyze historical data records from the project Integrated Surface Database (ISD), which was initiated by the National Climatic Data Center (NCDC) of the National Oceanic and Atmospheric Administration (NOAA). The goal of this project was to merge a large number of hourly data sets into a common format, thus providing a single collection of global hourly data that is continuously updated and freely available. The data set is accessible via FTP (ftp://ftp.ncdc.noaa.gov/pub/data/noaa/).

Besides wind speed at 10m height and the geographic location (longitude and latitude coordinates), the data set contains meteorological information such as visibility, temperature, air pressure, elevation and wind direction. Specifically, the data set from Saudi Arabia contains hourly measurements from 88 stations starting from 1946 up-to-this day. The start and end dates of recordings differ, depending on the station. Here, we select locations with measurements from June 2010, which is one of the months with strongest wind potentials in Saudi Arabia (Chen et al., 2018), and at most 10% of missing values, resulting in 697 observations from 28 stations. The locations of the stations are presented in Figure 1 (left). Figure 1 (right) shows the wind speed diurnal cycle, which is calculated as the mean wind speed over the 28 stations, for all 30 days of June 2010, at every hour of the day. We find that speed measurements are higher during the day when the heat from the sun creates convection currents, and lower at night when the temperature drops (Weisser and Foxon, 2003).

Figure 2 (left) gives the mean sample autocorrelation function of the residuals obtained after removing the daily seasonality of the wind speed, at the 28 stations shown in Figure 1 (left), up to a lag of 24 hours. For each station, we model the daily seasonality by fitting a set of Fourier basis functions with a period equal to 24 hours, using 365 days of hourly wind speed data. The 95%
confidence interval of the autocorrelation function shows that there is no significant correlation after 8 hours (see dashed line in the same plot). We then examine the correlation in space, at a temporal lag equal to zero. Figure 2 (right) illustrates the empirical correlations for all pairs of stations, together with the fitted ones from a Matérn model. Parameters are estimated by minimizing the weighted least squares using a quasi-Newton method. The strongest correlations are from stations that are close to each other. From the fitted model shown in Figure 2 (right), we observe that the effective range at which correlation drops to 0.05 is approximately 1144 km. This can be partly explained by the landscape of Saudi Arabia, characterized mainly by sandy deserts and isolated mountain clumps that are less steep, with the exception of the mountains near the Red Sea in the Western part, where the elevation rises up to 2990 meters (see Figure 1 (left)).

Wind vectors represent wind speed and direction simultaneously through the length and the angle of the vector. For converting the data on wind speed and direction into Cartesian coordinates, we denote by $U$ the horizontal (west-east) wind component and by $V$ the vertical (north-south) wind component. These components are oriented, i.e., $U > 0$ represents the wind
blowing from west to east and $V > 0$ is the wind from south to north. Then

$$U(s, t) = -X(s, t) \times \sin\{\theta(s, t)(\pi/180)\},$$
$$V(s, t) = -X(s, t) \times \cos\{\theta(s, t)(\pi/180)\},$$

where $X(s, t)$ is the wind speed and $\theta(s, t)$ is the angle of the wind direction measured in degrees clockwise from north at station $s$ and time $t$.

Figure 3 shows scatter plots of wind vectors at the 4 stations indicated by a circle (○) in Figure 1 (left) during five days in June 2010. We see that at Station 4 the $U$ and $V$ components are highly directional, whereas for Station 1, they are spread over. For Stations 2 and 3, most of the $U$ values are positive, indicating that the wind flows from west to east. Specifically for Station 2, the two wind components show a correlation in the northwestern direction.

3 Spatio-Temporal Probabilistic Models for Wind

In this section, we discuss two different models for probabilistic wind forecasting. The first is a univariate Gaussian model for the square-root transformed wind speed data. This transformation has been shown to be well-suited for wind data because it resolves both the homogeneity of the
variances and the normality in wind measurements (Haslett and Raftery, 1989). We then consider a multivariate spatio-temporal model based on LMC for wind vectors. The section ends with the estimation procedure and the methodology used to obtain inference and probabilistic forecasts.

3.1 Univariate model

Let $Y(s, t)$ denote the square-root transformed wind speed at station $s$ and time $t$. We assume a normal distribution: $Y(s, t) \sim \mathcal{N}(\mu(s, t), \sigma^2)$, where $\sigma^2$ is the variance of a Gaussian white noise, and $\mu(s, t)$ is the mean function. We follow the classical regression approach for Gaussian data, where the variance only accounts for the measurement error (the nugget) and the mean includes the fixed and random effects:

$$
\mu(s, t) = \beta_0 + \beta_1 \text{Temp}(s, t) + \beta_2 \text{Elev}(s) + \beta_3 \sin\{\theta(s, t)\} + \beta_4 \cos\{\theta(s, t)\} + f\{H(t)\} + W(s, t),
$$

where $\beta_0$ is the intercept, which is constant in space and time, Temp is the covariate temperature (in degrees Celsius) that varies in space and time, Elev is the elevation (in meters) that is constant in time, and $\theta$ represents the corresponding hourly wind direction (in radians) for each location and time. We define $f\{H(t)\}$ as a smooth function that repeats each hour, also known as the “time of day” effect. Finally, the term $W$ is a spatio-temporal model that evolves in time with
first order autoregressive dependence

\[ W(s, t) = aW(s, t - 1) + Z(s, t), \quad (3) \]

with \( t = 2, \ldots, T \) and \( |a| < 1 \). The term \( Z(s, t) \) is a realization of a Gaussian random field (GRF) with mean zero. GRFs are the most common choices for modeling spatial and spatio-temporal dependencies (Cressie and Wikle, 2015). We assume that the covariance function is

\[
\text{cov}\{Z(s, t), Z(s', t')\} = \begin{cases} 
\sigma^2_z C(h; \tilde{\theta}), & \text{if } t = t', \\
0, & \text{if } t \neq t', 
\end{cases} \quad (4)
\]

for \( s \neq s' \). The correlation function \( C \) between two locations \( s \) and \( s' \) depends only on the spatial lag \( h = \|s - s'\| \). Therefore, the process is second-order stationary and isotropic (Cressie, 1992). The term \( \sigma^2_z \) denotes the marginal variance, and the Matérn correlation function, defined by the following

\[
C(h; \tilde{\theta}) = \frac{1}{\Gamma(\nu)2^{\nu-1}}(kh)^\nu \mathcal{K}_\nu(kh),
\]

where \( \tilde{\theta} = (\kappa, \nu)^\top \) and \( \mathcal{K}_\nu \) is the modified Bessel function of the second kind of order \( \nu \). The smoothness, as well as the mean-square differentiability of the underlying process, is controlled by the parameter \( \nu \), and \( \kappa > 0 \) is related to the range, defined as \( \phi = \sqrt{8\nu}/\kappa \). The range parameter \( \phi \) is the distance at which the spatial correlation is approximately 0.13. We follow the common practice of setting the parameter \( \nu \) as equal to one for identifiability purposes and to reduce the computational time. This choice of parameter was shown to be flexible enough to cover a large class of spatial variations (Rue et al., 2009). Examples where the smoothness parameter \( \nu \) is fixed can be found in Cameletti et al. (2013), Munoz et al. (2013) and Ingebrigtsen et al. (2014).

Interactions between space and time in the data can be achieved either through a flexible model for the mean structure or the covariance function. A high degree of flexibility in the covariance structure is not only computationally expensive, but when the information in the data is weak, it can result in overfitting (Fuglstad et al., 2015). Because our focus is on fast
computations and predictions, we take a more straightforward covariance function; and use spatially and time-varying covariates (temperature and direction) to capture the interaction between space and time in the data.

3.2 Bivariate model

In Section 3.1, the extra information of wind direction for forecasting wind speed is included as a covariate in the modeling of the mean of the distribution; here we focus on jointly modelling the bivariate structure of the wind vector that results from speed and direction.

The LMC components are linear combinations of independent processes used to produce valid covariance structures. We can write this linear combination as

\[ W_j(s, t) = \sum_{k=1}^{p} a_{kj} W_{kj}(s, t). \]

The \( W_j(s, t), j = 1, \ldots, p \), are independent and identically distributed univariate spatio-temporal processes, each with a marginal variance \( \sigma^2 \) and a correlation function \( \rho(h, r; \theta) \), where \( h \) and \( r \) are the spatial and temporal separations, respectively. The resulting cross-covariance function takes the form

\[ C_{ij}(h, r; \theta) = \rho(h, r; \theta) \sum_{k=1}^{p} a_{ik} a_{jk} \rho(h, r; \theta_k). \]

The intrinsic or separable specification assumes a common set of parameters for the spatio-temporal processes \( W_j(s, t) \). A more flexible class of LMC arises if the \( W_j(s, t) \) are independent but no longer identically distributed. Then, the cross-covariance takes the form

\[ C_{ij}(h, r; \theta_1, \ldots, \theta_p) = \sum_{k=1}^{p} a_{ik} a_{jk} \rho(h, r; \theta_k). \]

In matrix form, it becomes

\[ C(h, r; \theta) = \sum_{k=1}^{p} \rho(h, r; \theta_k) T_k, \]

where \( T_k = a_k a_k^T \), with \( a_k \) is the \( k \)th column of \( A \). Since \( A_k A_k^T = T_k \), there are \( p(p + 1)/2 \) parameters to be estimated in \( A \). A common reduction in the number of parameters is to make \( A \) lower triangular (Schmidt and Gelfand, 2003).

A wind vector is determined by wind speed and wind direction, or by its west-east (horizontal) and north-south (vertical) components, which we denote by \( U \) and \( V \), respectively (see (1)). Alternatively, instead of using the Cartesian system, the wind speed and direction variables can be considered in polar coordinates. However, this is more complex to do, as it requires to specify a circular model for the wind direction. Here, we develop LMCs as a component of a more general spatio-temporal model for wind vectors. The bivariate probability density function is assumed
to be Gaussian. We assume that $U(s,t)$ and $V(s,t)$ have the following marginal distributions

$$U(s,t) \sim \mathcal{N}(\mu_u(s,t), \sigma_u^2),$$

$$V(s,t) \sim \mathcal{N}(\mu_v(s,t), \sigma_v^2),$$

where $\sigma_u^2$ and $\sigma_v^2$ are the variances of the white noise and $\mu_u(s,t)$ and $\mu_v(s,t)$ are the mean of the random processes, which can be defined with different linear predictors to produce hierarchical models based on LMCs.

We develop a model for the wind vector in a hierarchical framework, where at the first stage $U(s,t)$ and $V(s,t)$ are considered to be conditionally independent, with their own set of parameters. We start by introducing two independent spatio-temporal processes at location $s$ and time $t$, which are analogue to the process in (3),

$$W_u(s,t) = a_u W_u(s, t-1) + Z_u(s,t),$$

$$W_v(s,t) = a_v W_v(s, t-1) + Z_v(s,t),$$

where $Z_u(s,t)$ and $Z_v(s,t)$ are realizations of zero-mean GRFs with covariance function given as in (4). We assume that $W_u$ and $W_v$ are independent processes, but allow the mean function of $\mu_v(s,t)$ to depend on both $W_u$ and $W_v$ through the scale parameter $\lambda_{21}$. That is,

$$\mu_u(s,t) = \beta_{0,u} + \beta_{1,u} \text{Temp}(s,t) + \beta_{2,u} \text{Elev}(s) + f\{H(t)\} + W_u(s,t),$$

$$\mu_v(s,t) = \beta_{0,v} + \beta_{1,v} \text{Temp}(s,t) + \beta_{2,v} \text{Elev}(s) + f\{H(t)\} + W_v(s,t) + \lambda_{21} W_u(s,t),$$

where the $\beta_0$’s, Temp, Elev and $f\{H(t)\}$ are the same fixed and random effects as in (2), and $W_u(s,t)$ and $W_v(s,t)$ are the spatio-temporal Gaussian processes defined as in (6).

The LMC makes it possible to obtain the posterior distributions for $U$ and $V$ separately, since the likelihood factors into two independent terms, and the prior distributions are also assumed to be independent. This specification makes calculations relatively simple, and therefore computationally manageable. A drawback is that the smoothness of any component of the multivariate process is reduced to that of the roughest underlying univariate random field. To allow flexibility depending on the data set being modeled, we considered four different versions of
the bivariate model: \( \mu_u \) and \( \mu_v \) treated independently, \( W_u(s, t) \) affecting \( \mu_v \), \( W_v(s, t) \) affecting \( \mu_u \), and finally, \( W_u(s, t) \) affecting \( \mu_v \) and \( W_v(s, t) \) affecting \( \mu_u \) simultaneously. We chose to present only the model in (7) over the other bivariate models because it resulted in a good trade-off between accuracy and computational time in terms of probabilistic forecast results for modeling the data described in Section 2. A comparison of forecasts performances among the different forms of the bivariate model can be found in the Supplementary Material.

### 3.3 Inference and prediction

Fitting advanced space-time models in a computationally tractable manner is not a simple task. When dealing with GRFs, inference is generally performed using the covariance matrix \( \Sigma \). However, for a large number of observations, such as for high temporal resolutions with thousands of points, it turns out that computations with dense covariance matrices are not feasible. The challenge is how to represent the covariance structure of the models in Section 3 in a computationally feasible way to obtain fast and accurate inference and predictions. For instance, the classical kriging-based approach is time consuming due to the factorization of dense covariance matrices, and does not take into account the uncertainty in the covariance structure, providing only point estimate forecasts. To overcome this challenge, Lindgren et al. (2011) used a framework based on the connection between latent GRF models and Gaussian Markov random field (GMRF) models. The key property is the use of GRF models to represent dense covariance structures, while computations are implemented on GMRFs with a sparse precision matrix where only close neighbors are non-zero. Based on previous work (Lau et al., 2010; Dowell and Pinson, 2015) and on the satisfactory forecasting results from our data, we find that the autoregressive process, which belongs to the class of GMRFs, combined with a random effect for the daily seasonality (see (2)), is well suited for modelling wind speed dependencies in time.

A spatio-temporal model can be constructed starting from the following basis representation:

\[
\eta(s, t) = \sum_{k=1}^{K} \psi_k(s, t) \gamma_k, \tag{8}
\]
where each basis function is the product of a spatial and a temporal basis function, \( \psi_k(s, t) = \psi_s^i(s) \psi_t^j(t) \) and the \( \gamma_k \)'s are Gaussian-distributed random weights. Using the representation in Jones and Zhang (1997) and Krainski (2018), a space-time SPDE can be constructed by considering two separate processes: the first is a one-dimensional SPDE to describe the temporal dynamics and the second SPDE models the driving noise of the first process as spatially correlated:

\[
\left[ \frac{\partial}{\partial t} + \varphi \right] \eta(s, t) = \xi(s, t), \quad \tau(\kappa^2 - \Delta)^{\alpha/2} \xi(s, t) = W(s, t), \quad (s, t) \in \Omega \times \mathbb{R},
\]

where \( \varphi \) corresponds to the autoregressive coefficient of order one, while \( \kappa > 0 \) and \( \alpha = \nu + d/2, \nu > 0, \) are the scale and smoothness parameters of the spatial process, respectively. The innovation term \( \xi \) is independent in time and spatially correlated, whereas \( W \) is a Gaussian white noise.

The spatial dimension is denoted by \( d \); \( \Delta \) represents the Laplacian. The space-time SPDE in (9) generates a precision matrix for the vector obtained from (8) as \( \mathbf{Q} = \mathbf{Q}_S \otimes \mathbf{Q}_T \), where \( \otimes \) denotes the Kronecker product; \( \mathbf{Q}_S \) is the precision matrix for the purely spatial model, and \( \mathbf{Q}_T \) is the precision matrix for the purely temporal model. Currently, there is an ongoing work on developing computationally efficient non-separable space-time models directly from the principles of diffusion processes in space-time (van Niekerk et al., 2019), but it is not ready yet for implementation and the computational time is unknown. While separability is usually an idealization of a real phenomenon, the broad use of separable models is mainly because it decreases the number of parameters in the covariance matrix, resulting in a substantial reduction of the computation time, when working with large space-time data sets. For a data set of Irish wind speeds, Genton (2007) found small losses of accuracy for the separable approximations, while allowing significant computational savings. For fine-scale spatio-temporal structures, Ezzat et al. (2018) argued that local wind fields are asymmetric and that the use of non-separable covariances improves forecasting performances. However, the improved accuracy of a more complex model has to be weighed against its computational cost, depending on the specific needs of the application. Because our focus is on short and medium-term forecasts (i.e., from 1 to 24 hours ahead), we need models
that can run and produce reliable forecasts in less than one hour.

The key issue is to have a GMRF modelled at the triangulation vertices with a dimension $K$ and a response variable observed at $N$ locations. The basis function representation of $\xi(s, t)$ is used to define a sparse matrix of weights $B$ (also called projector matrix) with dimension $N \times K$, that defines the value of the GMRF for each observation. Thus, the $B$ matrix is responsible for projecting the process from the triangulation vertices to the observed locations. One advantage of using the continuous basis function expansion, in comparison with classical geostatistics, is the possibility to evaluate the solution at any given point, without requiring interpolation techniques. An illustration of the triangulation of the spatial domain for the data in Saudi Arabia is shown in the Supplementary Material.

We perform inference and prediction from the models in Section 3 using the R-INLA framework detailed in Rue et al. (2009). The basic idea is to take advantage of the Markov structure in the latent variables and use a Laplace approximation to solve the integral that appears in the calculation of the marginal posterior distributions of the parameters. For further information on the R-INLA package, see http://www.r-inla.org.

We exploit the possibility in R-INLA of combining several likelihoods for different parts of the observations, which makes it possible to construct multivariate models through a shared component structure (Bakka et al., 2018; Krainski et al., 2019) and to jointly model the response. For the model covariates, each likelihood may have its own set of parameters or a common set that can be shared between multiple likelihoods. Specifically, for the random part of the model, the copy feature in the R-INLA package allows us to link the same GRF to more than one likelihood. This feature is particularly useful when fitting the LMC in Section 3.2.

### 3.4 Prior specification

Our choice for the priors is in general motivated by some characteristics in the data. We construct informative priors using Penalized Complexity (PC) (Simpson et al., 2017), which penalises the
complexity over a simpler base model. For instance, the temporal correlations $a$’s are specified through the PC prior assuming $P(a > 0) = 0.9$ because negative correlations are unlikely and less intuitive in wind data. The prior distributions of $\kappa$ and $\tau$, the parameters related to the SPDE in (9), are defined using the parameterization $\log(\kappa)$ and $\log(\tau)$. This parametrization has the advantage of giving well-behaved posteriors, but it is difficult to interpret. Joint PC priors for the range $\phi$ and the marginal standard deviation $\sigma_z$ was introduced in Fuglstad et al. (2019). Using this idea, we only need to specify the tail probabilities $P(\sigma_z > \sigma_0) = \alpha_1$ and $P(\phi < \phi_0) = \alpha_2$, which are chosen such that the range of the field is approximately 30% of the domain size with 1% chance ($\alpha_1 = 0.01$), whereas the prior marginal variance is set to be smaller than the variance of the data with a probability of 0.01 ($\alpha_2 = 0.01$). We use an almost flat prior for the intercepts and for the covariates’ coefficients, which is implemented as a zero mean Gaussian distribution with large variance. A slightly informative prior is used for the scale parameter $\lambda_{21}$, which is Gaussian with mean zero and variance equal to 0.1. For $\log(\sigma^2_\epsilon)$, $\log(\sigma^2_u)$ and $\log(\sigma^2_v)$, we assume a log-Gamma prior with parameters 1 and 0.00005.

### 4 Evaluation Scheme

#### 4.1 Implementation

To evaluate the spatio-temporal forecasting performance of the models in Section 3, the data are split into a training set of 22 stations and a testing set with the remaining 6 stations. The locations are shown on the map in Figure 1 (left). Five days of hourly data, equivalent to 120 points, are used for parameter estimation, and forecasts are obtained for lead times 1, …, 24, i.e., from 1 hour to 24 hours ahead. When comparing different sizes of training sets with respect to the mean absolute error (MAE), root mean squared error (RMSE) and continuous ranked probability score (CRPS), we find that scores differ little with the length of the training period. We choose to use a training window consisting of the most recent five days as it provides a satisfactory trade-off between accuracy and computational cost.
The training period is a sliding window used to estimate the parameters in the model and rolled ahead by one observation, so that the parameters can be reestimated for each new forecasting period. June 2010 data are used as an example to evaluate the performance of the predictions, but we find that other months lead to similar results. In total, the results provide an analysis of $6 \times 30 \times 24 = 4,320$ forecasts, for each lead time. The code is parallelized using the PARDISO sparse library (Kourounis et al., 2018) with 18 INLA threads on 18 cores. The code optimization is done by INLA so that different processors get different hyper-parameter sets to calculate for. In real life applications, the values obtained from the previous hour could be used as initial values for the calculations in the next time step which would make the running time considerably faster.

In terms of parameter specification, each component in (7) may have its own set of covariates with its own coefficient vector. However, here, the coefficients $\beta$'s, as well as the $f\{H(t)\}$, are the same for both mean functions $\mu_u(s, t)$ and $\mu_v(s, t)$, since we obtain no improvement of the CRPS values by treating them separately. Furthermore, it makes sense to consider that elevation, temperature and time of the day effect are expected to have a similar influence on both wind components.

To generate forecasts in space and time, it is necessary to define the values of the covariates temperature and elevation at the respective locations and lead times. The digital data set of elevation from the Shuttle Radar Topography Mission (SRTM), originally produced by NASA, are downloaded, using the R package raster (Hijmans, 2017). Hourly temperature data are obtained from the WRF Mesoscale numerical model (see Section 6.1).

Using the estimated parameters from our models as described in Section 3.3, we are able to obtain samples from the probabilistic forecast distributions at each lead time both for observed and unobserved locations. We generate 5,000 samples from the forecast distributions of the models described in Section 3.1 and 3.2. Here, we are interested not only in forecasting the full wind vector, but also specific derived quantities, such as speed. From the empirical distribution
of these simulated samples of vector forecasts, we can generate any derived quantity and use it as our forecast distribution. This is convenient, since forecasts for marginal wind speed and wind direction become readily available from the bivariate model in Section 3.2. We use the vector forecasts from the bivariate model to generate forecasts for the marginal wind speed and compare them with those generated from the univariate model in Section 3.1.

We compare the different models by combining the forecasts over the entire test period. This provides a more robust ranking of the different models than focusing only on a single data set. In addition, the scores are computed for each lead time and for each model.

4.2 Verification methods

The primary challenge in evaluating probabilistic forecasts is that the distribution of the actual data generating process is unknown. Therefore, the diagnostic approach should be able to compare a probability distribution from the forecast with the practical observation that is real-valued. For these situations, it is common to employ a framework based on calibration of probabilistic forecasts, also referred to as reliability. Calibration is critical in practical applications because it measures the consistency between the forecast system and the observations. For instance, if an event is predicted to occur with probability 10%, then it should, on average, happen in about 10% of the time.

A widely used diagnostic to assess the calibration of univariate forecast distributions is the probability integral transform (PIT) value, proposed by Dawid (1984). For a forecast distribution of wind speed with probability density $f$, the PIT value at time $t$ and location $s_{ij}$ is

$$\text{PIT}(s_{ij}, t) = \int_0^{x(s_{ij}, t)} f(x) \, dx,$$

for $i = 1, \ldots, D$ and $j = 1, \ldots, N$, where $D$ is the number of validation sets and $N$ is the number of stations. The actual outcome is denoted by $x(s_{ij}, t)$, which may be perceived as a random number generated from an unknown true underlying distribution $g$. If $f$ is continuous
and the forecasts are ideal, i.e., $f$ coincides with $g$. \cite{Diebold2008} showed that the series of PIT values are uniformly distributed in the interval $[0, 1]$. To have a numerical measure and to better compare different forecasting techniques, we compute the square-root sum of the squared distances (RSSD) from the PIT histogram to the horizontal line that corresponds to the density at $y = 1$. This score is negatively oriented, i.e., the lower, the better.

Another popular visual tool to assess calibration for continuous variables is the reliability diagram (see, e.g., \cite{Pinson2010}). The ideal scenario is when points are aligned with the diagonal in the reliability diagram, meaning that the nominal coverage rate, matches the observed frequencies precisely. For each quantile forecast $\hat{q}^{(\alpha)}(s_{ij}, t)$ at time $t$ and station $s_{ij}$, we define an indicator variable $I^{(\alpha)}(s_{ij}, t)$ that is equal to one if the observed value, $x(s_{ij}, t)$, is below the $\alpha$ quantile forecast (hit) and zero (miss) otherwise:

$$I^{(\alpha)}(s_{ij}, t) = \begin{cases} 1, & \text{if } x(s_{ij}, t) \leq \hat{q}^{(\alpha)}(s_{ij}, t), \\ 0, & \text{otherwise.} \end{cases}$$

Next, let $n^{(\alpha)}_{t,1}$ denote the sum of hits and $n^{(\alpha)}_{t,0}$ the sum of misses,

$$n^{(\alpha)}_{t,1} = \sum_{i=1}^{D} \sum_{j=1}^{N} I^{(\alpha)}(s_{ij}, t) \quad \text{and} \quad n^{(\alpha)}_{t,0} = DN - n^{(\alpha)}_{t,1}.$$  

The nominal coverage $c^{(\alpha)}_h$ can be estimated as the average of $I^{(\alpha)}(s_{ij}, h)$:

$$\hat{c}^{(\alpha)}_t = \frac{1}{DN} \sum_{i=1}^{D} \sum_{j=1}^{N} I^{(\alpha)}(s_{ij}, t) = \frac{n^{(\alpha)}_{t,1}}{n^{(\alpha)}_{t,1} + n^{(\alpha)}_{t,0}}.$$  

In practical applications, a probabilistic forecast should be sharp, meaning that the density forecast should be as tight as possible. To address this issue, \cite{Gneiting2007} suggest maximizing the sharpness conditional on calibration as a way to compare different forecasting models. Sharpness diagrams are useful for assessing the spread of the forecast distributions. For each nominal quantile, this type of diagram can be obtained from the size of the central predicted interval that is centered in probability around the median. The interval length is the mean length
of the forecast intervals from all the locations and validation sets and can be written as

\[ \bar{\delta}^\alpha(F, t) = \frac{1}{DN} \sum_{i=1}^{D} \sum_{j=1}^{N} \left\{ \hat{q}(s_{ij}, t)^{(1-\alpha/2)} - \hat{q}(s_{ij}, t)^{(\alpha/2)} \right\}, \]

where \( \hat{q}(s_{ij}, t)^{(\alpha)} \) is the quantile forecast with nominal coverage \( \alpha \) estimated at \( s_{ij} \) and time \( t \).

The continuous ranked probability score (CRPS) is particularly appealing when evaluating probabilistic forecasts because it is a proper scoring rule that rewards calibration and sharpness simultaneously. Similarly, the directional CRPS (dCRPS) is a generalization of the CRPS that can be used to evaluate circular variables (Grimit et al., 2006; McLean Sloughter et al., 2013):

\[ \text{dCRPS}(F, \theta) = E_F \{ \gamma(\Theta, \theta) \} - \frac{1}{2} E_F \{ \gamma(\Theta, \Theta') \}, \quad (10) \]

where \( F \) is the forecast distribution, \( \gamma(\theta_1, \theta_2) \) is the angular distance between two directions \( \theta_1 \) and \( \theta_2 \) on the circle \([-\pi, \pi]\), and \( \Theta \) and \( \Theta' \) are independent random variables with distribution \( F \). For each time \( t \), a computationally efficient approximation to (10) based on random samples of the directions \( \theta^{(1)}, \ldots, \theta^{(n)} \) from the corresponding forecast distribution can be evaluated using

\[ \hat{\text{dCRPS}}(F, t) = \frac{1}{DN} \sum_{i=1}^{D} \sum_{j=1}^{N} \left\{ \frac{1}{n} \sum_{k=1}^{n} \| \hat{\theta}^{(k)}(s_{ij}, t) - \theta(s_{ij}, t) \| - \frac{1}{2n^2} \sum_{k,l=1}^{n} \| \hat{\theta}^{(k)}(s_{ij}, t) - \hat{\theta}^{(l)}(s_{ij}, t) \| \right\}. \]

5 Results

The techniques proposed here are implemented on a case study based on the wind data from Saudi Arabia described in Section 2. We use the vector forecasts from the bivariate model in Section 3.2 to generate spatio-temporal forecasts for marginal wind speed. The performance of the bivariate model is compared to that of the univariate model for predicting wind speed described in Section 3.1. Details of the implementation, as well as the scores used to quantify the skill of probabilistic forecasts, can be found in Section 4.

We assess forecast performances of the wind speed from all wind stations and data sets generated from the univariate and bivariate models, detailed in Sections 3.1 and 3.2 respectively.
Figure 4: PIT histograms of spatio-temporal wind speed forecasts for the univariate model in Sections 3.1 (top row) and for the bivariate model in Section 3.2 (bottom row). Each column (from first to last) represents lead times 1, 24 and all lead times combined. The overall calibration of the bivariate model is the best, with smaller RSSD values on top of the histograms for all lead times. The dashed lines at $y = 1$ represent the uniform density.

Figure 4 displays the histograms of the PIT values for wind speed forecasts from the univariate (top row) and bivariate (bottom row) models at lead times 1 (left column), 24 (middle column) and all lead times combined (right column). These lead times have been selected to illustrate the reliability for both shorter and longer lead times; similar results have been found for other horizons. While the shape of the three PIT histograms for the univariate model at the top row indicates that the spread of the underlying forecast density is too small (also called underdispersed), we see that the PIT histograms for the bivariate model at the bottom row are approximately uniform, hence indicating calibration. The smaller RSSD values on top of each histogram confirm the visual inspection that the bivariate model outperforms the univariate
model in terms of calibration for all lead times.

![Reliability and Sharpness Diagrams](image)

**Figure 5:** Reliability (left column) and sharpness (right column) diagrams of spatio-temporal wind speed forecasts at lead times 1 (top row) and 24 (bottom row). Diagrams were calculated using the univariate model in Section 3.1 (◦) and the bivariate model in Section 3.2 (●).

Reliability diagrams for the forecasts from the univariate and bivariate models at lead times 1 and 24 are shown in the top and bottom rows, respectively, in the left column of Figure 5.

We expect a calibrated forecast with nominal coverage α to contain the observations α% of the time, resulting in points aligned on the diagonal. Although, for most quantiles, the points are close to the diagonal for both models, we can see that the bivariate model clearly improves the forecasting calibration at both lead times. The underdispersion indicated by the shape of the PIT histograms for the univariate model (see top row of Figure 4) when compared to the bivariate model, is also shown from the sharpness diagrams on the right column of Figure 5.
Indeed, the bivariate model produces wider forecast densities than the univariate model, which, in turn, results in calibrated forecasts for both, shorter and longer lead times.

The Spatial Prediction Comparison Test (SPCT) introduced in Hering and Genton (2011) is used to formally compare the spatio-temporal forecasts from the univariate and the bivariate models in terms of point forecast. Details on the SPCT test are given in Section 3 of the Supplementary Material. This test shows that wind speed forecasts from the univariate and bivariate models are not significantly different in terms of root mean squared error (RMSE) and mean absolute error (MAE); see Figure 4 of the Supplementary Material. The bivariate model still has the advantage of producing the wind direction forecast, and it is better calibrated than the univariate model in terms of PIT and reliability/sharpness (see Figures 4 and 5). The SPCT does not compare the whole forecast distribution; hence, it cannot be used beyond point forecasts.

Figure 6 shows an example of the wind speed data over Saudi Arabia, together with maps of the mean, 5% and 95% quantiles of the wind speed forecast distribution, using our bivariate model. The evolution in time can be seen from the rows in the same figure, which correspond to lead times 1 to 4. These maps show that the forecast distribution is able to capture the wind speed pattern throughout most of Saudi Arabia, with the 90% central part of the forecast distribution always containing the observed values.

6 Spatio-Temporal Forecast Validation

Section 5 shows the analyses from 28 wind stations located in Saudi Arabia. Here, we use simulated data from physical models to validate the performance of the spatio-temporal forecasts, and to evaluate the impact of adding stations. More specifically, we show how increasing the number of stations used to fit the model influences the accuracy of spatio-temporal predictions.
Figure 6: Examples of wind speed data together with mean, 5% and 95% quantiles (first, second, third and fourth columns respectively) of the forecast distribution of wind speed obtained using the bivariate model in Section 3.2. Each row (from first to last) corresponds to lead times 1, up to 4 hours.

6.1 The WRF simulated data set

We consider a new and unique data set from numerical model simulations of wind at a 5-kilometer resolution over the region of Saudi Arabia. Mesoscale numerical models such as WRF rely on large-scale atmospheric phenomena or meteorological reanalysis to provide boundary conditions and solve physical equations driving the real process on a fine scale. For the WRF data, the boundary conditions used to simulate data are obtained from the Modern-Era Retrospective 54
analysis for Research and Applications (MERRA, Rienecker et al. (2011)), a reanalysis product developed at NASA’s Global Modeling and Assimilation Office (GMAO), using the Goddard Earth Observing System Version 5 (GEOS-5) general circulation model (GCM), in conjunction with satellite and surface observations through a data assimilation system (see Yip (2018) for details).

The WRF data set covers the 2009-2014 period. Each measurement corresponds to hourly data of the $U$ and $V$ components, as well as the temperature on a regular grid of $769 \times 659$ points in space roughly bounded by $5-35^\circ$N and $30-65^\circ$E. In total, there are 506,771 spatial locations, and each one is measured at 52,560 different time points, resulting in more than 26 billion space-time points.

6.2 Forecast validation

The main idea is to fit our bivariate model to a subset of the simulated WRF data set, and to then compare the forecasts at the 6 locations used for validation in the analysis performed in Section 5 (see stations indicated with a cross ($\times$) in the map from Figure 1) with the actual WRF data at those locations. First, we fit our model to the WRF data locations that correspond to the 22 wind stations used as the training set (see stations marked as $\bullet$ and $\circ$ in the map from Figure 1). For stations in-between grid points of the WRF data, we use bilinear interpolation to determine the “WRF” values, i.e., interpolating linearly first in the $x$-direction, then in the $y$-direction. Next, we increase the number of locations used to fit the models by randomly picking new points from the WRF data set. We continue to grow the number of locations by 20 until we reach 242, and notice that only a negligible improvement is achieved after that.

Figure 7 illustrates the mean dCRPS values computed for different numbers of stations used to fit the model at lead times 1, 2 and 3. For each number of stations and lead time, the mean is taken over the locations used for validation and all validation sets from the moving-window approach described in Section 4. Similarly to the analysis of the wind data from stations in Saudi
Arabia presented in Section 2, an estimation of the parameters is performed on observations from a training period, which we assume here to be the 120 most recent observations, i.e., 5 days. As expected, when the number of stations increases, the mean dCRPS tends to decrease. The most significant improvement occurs when increasing the training set from 122 to 142 stations, for which the variation in the mean dCRPS is about 8% at lead time 1.

![Figure 7: Mean dCRPS among the 6 validation stations at all training sets with respect to the number of stations used to fit the model at lead times 1, 2 and 3.](image)

To illustrate the application of reproducing maps of wind speed in Saudi Arabia using different sizes of training set, we select hourly data from a particular day in June of 2010. Figure 8 shows an example of a map of wind speeds according to the simulated WRF data, mean, 5% and 95% quantiles (first, second, third and fourth columns, respectively) of the forecast distribution of marginal wind speed, obtained by fitting the data to our bivariate model. The four rows correspond to lead times 1, 2, 3 and 4, respectively. To illustrate the forecasts obtained with our
method, we consider a training set of 242 stations inside Saudi Arabia, which is the maximum number of stations considered to fit our model. This figure shows that the posterior mean is well-estimated, particularly for high values of wind speed. From the maps of the tails (third and fourth columns), we see that the forecast distribution is wide enough to cover the true value. A movie with the wind speed motion during the 24 hours of the day, starting from midnight, is provided in the Supplementary Material.

Figure 8: Example of maps of hourly wind speed over Saudi Arabia for a specific day in June 2010 from the WRF simulated data together with mean, 5% and 95% quantiles (first, second, third and fourth columns, respectively) of the forecast distribution of wind speed obtained using 242 stations to fit the bivariate model in Section 3.2. Each row (from first to last) corresponds to lead time 1, up to 4.
7 Discussion

In this study, our first goal was to present model-building strategies for wind speed forecasting that take into account the available information on the wind speed and its direction, across space and time. For this purpose, an LMC is a critical component of a more complex spatio-temporal model for wind vectors. Instead of characterizing the full space-time bivariate covariance, we specified the model components in a hierarchical way so that they could be fitted within reasonable computational time, with an SPDE approach that allows full Bayesian inference along with a full forecast distribution, not only for the wind vectors, but also for the univariate wind speed, through an Integrated nested Laplace approximation. Specifically, the mean computation time for fitting the univariate and the bivariate models to the NOAA data are 13 and 22 minutes, respectively, on a machine with Intel Xeon e5-2680 (2.8 GHz,) ten-core processors and 118GB memory.

Our motivation for proposing a bivariate model for the wind vector stems from the fact that the more classical univariate approaches may be uncalibrated and underdispersive for short-term forecasts. A comparison between the spatio-temporal forecast results of wind speed from the LMC with those from a univariate model fitted directly to the transformed wind speed showed that the LMC improved the calibration of probabilistic forecasts. This improvement comes from the correction of the consistently too narrow forecast intervals when using the univariate model, which resulted in a lack of sufficient reliability. Our results provide evidence that considering speed and direction as a bivariate process have the potential to improve the forecasts significantly. Furthermore, another clear advantage of the bivariate specification is the ability to forecast wind direction, in addition to speed.

Unlike the focus of most papers in the wind literature that issue forecasts for each location individually, here we modeled the space-time correlation between observed values in order to have the ability to forecast at several locations simultaneously, even when data are missing. Our models also included covariates such as the elevation and the temperature, as well as a
smooth function to model the daily seasonality present in wind. The proposed models were used to produce probabilistic forecasts of the wind speed over stations in Saudi Arabia. Another important feature is their ability to produce fast and accurate maps of wind speed forecasts, not only at stations where data are observed, but also for the entire country.

Our next objective was to validate the accuracy of the produced maps of wind forecasts from our spatio-temporal model, using a case study based on simulated data from a state-of-the-art physical model at a high spatial resolution. By increasing the number of stations used to fit the model, we can assess improvement in the forecasts and give recommendations for the number of stations needed in order to reach the desired accuracy. Our case study showed that the largest improvement, in terms of dCRPS for forecasting wind speed is obtained by increasing the number of stations in the training set, from 122 to 142. This is a reasonable number of locations where to have wind stations, or, in the future, installed wind farms.

The main rationale for this work is the clear need for producing reliable probabilistic forecasts of wind, at several locations and lead times simultaneously, in order to efficiently integrate wind energy into the existing power grids. Our focus, here, was Saudi Arabia, as it starts planning for a post-oil future and establishing ambitious targets for renewable energy. Building on the work presented here, future studies may involve modeling the wind vector using a bivariate Matérn model. Following Hering and Genton (2010), a LMC could be constructed using a bivariate skew-t distribution as the likelihood. In terms of nonstationarity for the bivariate models, at least two specifications of the LMC could be developed. The first is to let each $W_j(s, t)$ be a nonstationary process. Alternatively, one could allow the coefficients of the coregionalization matrix $A$ to be varying spatially.

**References**


