Received: 22 October 2019 | Revised: 17 March 2020 | Accepted: 27 March 2020

DOI: 10.1002/env.2628

SPECIAL INVITED ARTICLE

WILEY

A high-resolution bilevel skew-*t* stochastic generator for assessing Saudi Arabia's wind energy resources

Felipe Tagle[1] | **Marc G. Genton**[2] | **Andrew Yip**[3] | **Suleiman Mostamandi**[3] | **Georgiy Stenchikov**[3] | **Stefano Castruccio**[1]

[1]Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, Notre Dame, Indiana

[2]Statistics Program, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

[3]Atmospheric and Climate Modeling Group, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

Correspondence
Stefano Castruccio, Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, Notre Dame, IN 46556.
Email: scastruc@nd.edu

Abstract

Saudi Arabia has recently established its renewable energy targets as part of its "Vision 2030" proposal, which represents a roadmap for reducing the country's dependence on oil over the next decade. This study provides a foundational assessment of the wind resource in Saudi Arabia that serves as a guide for the development of the outlined wind energy component. The assessment is based on a new high-resolution weather simulation of the region generated with the Weather Research and Forecasting (WRF) model. Furthermore, we propose a spatiotemporal stochastic generator of daily wind speeds that assists in characterizing the uncertainty of the energy estimates. The stochastic generator considers a vector autoregressive structure in time, with innovations from a novel biresolution model based on a skew-*t* distribution with a low-dimensional latent structure. Estimation of the spatial model parameters is performed using a Monte Carlo expectation-maximization (EM) algorithm, which achieves inference over approximately 184 million points and enables to capture the spatial patterns of the higher order moments that typically characterize high-resolution wind fields. Our results identify regions along the western mountain ranges and central escarpments that are suitable for the deployment of wind energy infrastructure. According to the assessment, between 30 and 70% of the national electricity demand could be met by wind energy.

KEYWORDS

expectation maximization, non-Gaussian, skew *t*, space time model

1 | INTRODUCTION

The anticipated adverse impacts of climate change have led policymakers worldwide to adopt mitigating strategies and actively promote the development of renewable energy sources. Wind energy is an important component of this effort, and the total world capacity has increased steadily for the past several decades. Recent surveys indicate that the current installed wind-energy capacity can fulfill 5% of the world's energy demand and, for an increasing number of countries, already represents a significant share of their energy mix. For example, wind energy meets 43% of the energy needs in Denmark (WWEA, 2018). In fact, wind surpassed coal in terms of total power-generating capacity in 2016, becoming the second largest power supply in the European Union (WindEurope, 2018). Originating from the uneven heating of the Earth's atmosphere by the sun, wind energy represents a clean energy resource of nearly infinite supply. However, despite its benefits, wind energy poses operational challenges for traditional electricity markets. Wind power is produced only as the wind blows, and no cost-effective means exist to store it; therefore, wind energy cannot be scheduled at will (Hering & Genton, 2010; Pinson, 2013; Zhu & Genton, 2012).

In the Middle East and North Africa (MENA) region, renewable energy sources provided only 3.3% of the total power generated (~1,200 TWh) in 2011*. However, renewable energy investments are expected to increase driven by energy security considerations and energy demand growth (Bryden, Riahi, & Zissler, 2013). Saudi Arabia, ubiquitously known for its vast oil reserves, is also embracing renewable energy sources. In 2016, Crown Prince Mohammad bin Salman issued "Vision 2030," a comprehensive plan to reduce the country's oil dependence. One of the main components of the plan is the installation of 9.5 GW of renewable energy by 2023 (Vision 2030, 2016). As of July 2018, the government is evaluating several proposals for the installation of the first 400 MW of wind power in the country in Dumat Al Jandal (Dumat Al Jandal, 2020).

Several studies have sought to quantify the abundance of the wind resource across the Arabian Peninsula (e.g., Rehman & Ahmad, 2004; Rehman et al., 2007; Shaahid, Al-Hadhrami, & Rahman, 2014); however, these have been plagued by sparse and temporally inconsistent observational data. The lack of a dense observational network has led other studies to rely on reanalysis data products, which use data assimilation techniques to combine numerical model output with observations to produce gridded datasets with usually global coverage. Yip, Gunturu, and Stenchikov (2016), using the modern-era retrospective analysis for research and applications (MERRA, Rienecker et al., 2011) reanalysis, reported on the variability and intermittency of the wind resource over the Arabian Peninsula, highlighting the abundance of the wind resource over the western mountains of Saudi Arabia relative to most of the Red Sea coastal areas, and its high variability in coastal areas along the Arabian Gulf. More recently, Chen, Castruccio, Genton, and Crippa (2018) analyzed wind power potential based on the Middle East North Africa coordinated regional climate downscaling experiment (MENA CORDEX) dataset and identified high wind energy potential areas over western Saudi Arabia, which will likely persist at least until the middle of the 21st century.

Wind energy is harnessed by turbines at hub heights ranging from 50 to 140 m, making it highly sensitive to the local surface topography. This sensitivity is compounded by the relationship between available energy and wind speed, which is proportional to the cube of the wind speed. Analyses based on reanalysis products or global models are limited in their ability to identify areas of high wind-energy potential, since the reanalysis data are typically only available at coarse resolution (e.g., 0.5°, see Figure 1a) and, therefore, can only represent large-scale processes and topographical features. Numerical model simulations with higher resolutions, on the order of 1–10 km, usually generated by numerical weather prediction (NWP) models are required for such purposes. In this work, we develop a new, unique dataset, the first with 5-km resolution over Saudi Arabia, to provide the first high-resolution, nation-wide assessment of the abundance and variability of wind resources in the region (see Figure 1b). This dataset was produced using a state-of-the-art mesoscale NWP model, the weather research and forecasting (WRF, Skamarock et al., 2008) model, which was developed at the National Center for Atmospheric Research (NCAR) Mesoscale and Microscale Meteorology Laboratory. WRF is currently in operational use for real-time forecasting at national meteorological centers as well as laboratories, universities, and private companies worldwide.

To reduce the computational demands of performing a numerical simulation at such a fine resolution, the temporal dimension is limited to 6 years, from 2009 to 2014. However, a proper quantification of wind-related statistics requires longer simulations and/or more trajectories, which are not possible to perform given the severe costs of running WRF at high resolution. A possible solution would be to develop a statistical approximation to WRF output and generate surrogate simulations of spatiotemporally resolved synthetic data to assess the variability of wind energy. In order to provide results that are relevant for policymaking, this *stochastic generator* (Jeong, Castruccio, Crippa, & Genton, 2018) should generate output at time-scales at which wind energy is usually integrated in electrical grids, that is, daily at a minimum. Modeling daily winds over such a large region presents two main challenges.

1. From a methodological standpoint, the assumption of Gaussianity has traditionally afforded an unparalleled degree of analytical tractability and flexibility in spatial and spatiotemporal statistics (Gelfand & Schliep, 2016). This assumption, while appropriate for highly aggregated data in time, such as annual mean wind (Jeong et al., 2018), fails to apply at lower levels of aggregation. For monthly data, a trans-Gaussian process such as the Tukey *g*-and-*h* (Xu & Genton, 2017) would be sufficient to capture the tail and skewness behavior (Jeong, Yan, Castruccio, & Genton, 2019). However, at

*Throughout this work, we refer to W as Watts and Wh for watt-hour, along with their usual convention of a previous letter M for Mega ($\times 10^6$) and G for Giga ($\times 10^9$), and T for Tera ($\times 10^{12}$).

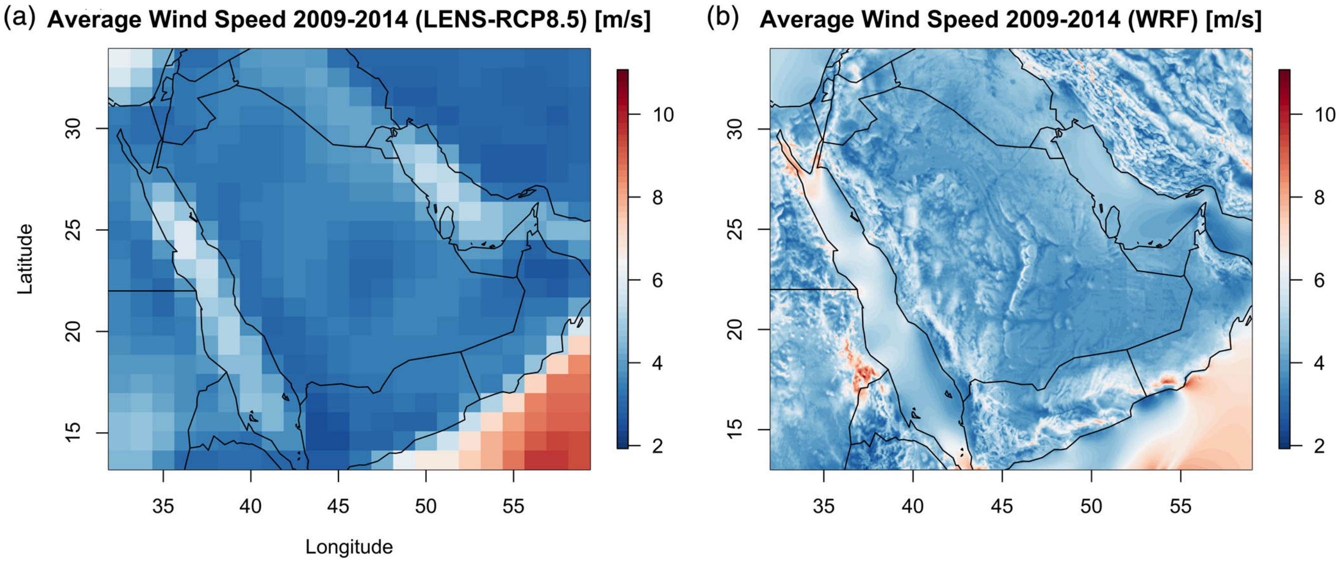

FIGURE 1 Average of daily wind speeds during the period 2009–2014 for (a) a CESM version 1 simulation, as represented member #1 of the large ensemble project (LENS, Kay et al., 2015) developed at NCAR and (b) the WRF simulation under consideration. Details of the simulation are available in Section 2

the daily time-scale, a model that can account for the severe skewness and kurtosis as well as the dependence across different scales is required.

For daily data, biresolution models based on the multivariate skew-t distribution have recently been proposed for modeling daily winds over Saudi Arabia from an Earth system model simulation (Tagle, Castruccio, Crippa, & Genton, 2019; Tagle, Castruccio, & Genton, 2019). The multivariate skew-t is a generalization of the traditional Student-t distribution that belongs to the skew-symmetric family of distributions (see Genton, 2004; Azzalini & Capitanio, 2014 for overviews). These models have been applied, however, at much lower resolutions; for example, 149 points at a 1° resolution, in contrast to the approximately 84,000 points here.

2. On the computational side, the increase of the spatial resolution of two orders of magnitude from previous work calls for the development of a model that is flexible enough both to achieve inference in an affordable time and to capture spatial and temporal dependence, as well as non-Gaussian behavior. Since previous biresolution skew-t models (Tagle, Castruccio, Crippa, & Genton, 2019) relied heavily on multivariate latent spatial processes, a more affordable but equally flexible model with a lower dimensional latent structure is necessary.

In this work, we propose a new biresolution skew-t stochastic generator that is suitable for inference at the spatial resolution of the WRF simulation. After removal of the temporal structure, which we assume follows a second-order autoregressive structure, we perform inference on the residuals with an expectation-maximization algorithm (EM, Dempster, Laird, & Rubin, 1977) and provide semi-closed-form expressions for both the E and the M steps, where the lower dimension of the latent factors allows considerably faster inference relative to other available models. We also demonstrate in a simulation study that this simplification implies only a small loss in efficiency compared with a ground truth from a fully structured (and considerably more expensive) model, and it vastly outperforms all other models when performing inference against the same model. As part of our results, we provide estimates of potential wind energy generation based on the stochastic generator output and a hypothetical build-out of wind farms along those areas deemed most suitable, which can serve as guidelines for the renewable energy goals of the Saudi Arabian government for 2030 and beyond.

The remainder of the article is organized as follows. Section 2 introduces the wind dataset used in the study. Section 3 describes the construction of the biresolution model used to capture the purely spatial variation and presents a simulation study that compares the performance of this model against alternatives in the literature. Section 4 provides details of the autoregressive model used for the temporal component of the proposed spatiotemporal model and describes the assessment metrics. Section 5 analyzes the model-generated wind field simulations over Saudi Arabia and presents the results pertaining to availability, variability, and intermittency of the wind resources and the associated produced energy with a standard wind turbine. Section 6 concludes with a discussion.

2 | THE WIND DATASET

We consider daily wind speeds over Saudi Arabia from January 1, 2009 to December 31, 2014, as simulated by WRF at a 5 km resolution. WRF relies on global atmospheric models that represent large-scale atmospheric phenomena or reanalysis products to provide boundary conditions from which the governing physical equations of finer-scale processes are solved. We performed our daily wind simulation using boundary conditions obtained from the European Center for Medium-range Weather Forecast (ECMWF) operational analysis at 15 km resolution. The dataset is comprised of 6 years of daily data over approximately 84,000 spatial locations, which corresponds to approximately 184 million data points (see Yip, 2018 for additional details).

Figure 1 compares the daily average wind speeds over the years 2009–2014, simulated by the CESM version 1, at approximately 1° resolution, and the present WRF simulation. The comparison illustrates the extent to which fine-scale features of a wind field become obscured at coarser resolutions. We note that this is not just an effect of averaging a spatial process from one grid onto a coarser grid, but rather a result of the limitations of Earth system models to represent relevant atmospheric processes that occur at both large and small spatial and temporal scales.

3 | A NEW SPATIAL MODEL FOR SKEWED DATA

In this section, we describe the construction of the skew-*t* model that will be used to model the innovations terms in the spatiotemporal model for wind speeds described in the next section. To motivate the construction, we briefly review general approaches to non-Gaussianity in space and time, as well as the skew-normal and skew-*t* distributions and a previous construction designed for low-resolution community Earth system model (CESM, Hurrell et al., 2013) runs as shown in Figure 1a, which inspired the proposed model. Finally, we present a simulation study benchmarking the proposed model against this other model and a Gaussian specification. Details of the EM-based inference are provided in the supplementary material.

3.1 | Approaches to model non-Gaussianity in space and time

Although work to capture non-Gaussianity in space and time is comparatively smaller than the vast literature available on Gaussian processes (see Heaton et al., 2019 for the most recent review and comparison), it can be generally divided in three categories. When the departure from Gaussianity is not too severe, a marginal transformation can be proposed to allow a parametric control over the higher moment (e.g., skewness and kurtosis) behavior of the process. There trans-Gaussian approaches have been extensively studied at least since Noël Cressie's first seminal book on spatial statistics (Cressie, 1993), and more recently new classes of transformations have been proposed to more flexibly control higher moments (Xu & Genton, 2017; Yan & Genton, 2019). An alternative approach to capture more severe degrees of non-Gaussianity is to model the data conditionally, and hence to control the process dynamics through a latent Gaussian model, thus achieving high modeling flexibility while avoiding computational issues for large data. This approach has seen a surge in popularity in the recent decade due to its natural tractability in the Bayesian framework. An alternative approach is to perturb the normal distribution directly. Recently, Tadayon and Torabi (2019) proposed to add an explicit term from a skewed distribution, and McDermott and Wikle (2019) expressed the data as a function of a latent process with nonlinear, machine learning inspired dynamics.

Our work is aligned with this last area of research: We propose a direct perturbation of the normal with a skew-*t* distribution, which by virtue of a representation theorem can be expressed in conditional stages, and hence is amenable to conditional inference with an EM algorithm, an inferential approach, which has been extensively developed in the Gaussian case (Bessac, Ailliot, & Monbet, 2015; Fassò & Finazzi, 2011; Zhang, 2007).

3.2 | Skew-symmetric distributions

The seminal work of Azzalini (1985) introduced the skew-normal family of distributions based on the concept of perturbing the normal distribution by a term to allow for varying degrees of skewness, prompting the development of entirely new classes of models (see the books of Genton, 2004 and Azzalini & Capitanio, 2014). The multivariate skew-normal distribution (SN, Azzalini & Dalla Valle, 1996) denoted by $\mathcal{SN}_d(\mathbf{0}, \overline{\boldsymbol{\Omega}}, \boldsymbol{\alpha})$ has a probability density function (pdf) given by

$$2\phi_d(\mathbf{x}; \overline{\boldsymbol{\Omega}})\Phi(\boldsymbol{\alpha}^\top \mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^d,$$

where ϕ_d is the pdf of a d-dimensional normal distribution with mean zero and positive definite correlation matrix $\overline{\boldsymbol{\Omega}}$, Φ is the cumulative distribution function of a standard normal distribution, and $\boldsymbol{\alpha}$ is a d-dimensional vector of parameters that controls the skewness (the null vector corresponds to the multivariate normal distribution). The above standard form can be generalized to a location-scale family, denoted by $\mathcal{SN}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha})$ by the transformation, $\mathbf{X} = \boldsymbol{\xi} + \boldsymbol{\omega}\tilde{\mathbf{X}}$, with $\tilde{\mathbf{X}}$ denoting a random vector with the standard skew-normal distribution, $\boldsymbol{\Omega} = \boldsymbol{\omega}^\top \overline{\boldsymbol{\Omega}} \boldsymbol{\omega}$ and $\boldsymbol{\omega}$ is a diagonal matrix collecting the individual standard deviation terms.

A natural extension of the SN is the multivariate skew-t distribution (ST, Azzalini & Capitanio, 2003), denoted by $\mathcal{ST}_d(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, \nu)$, which supplements the former by adding the flexibility to account for heavy-tailed behavior. Its construction is analogous to that of the multivariate Student-t distribution, as it is defined by the ratio $\mathbf{Y} = \mathbf{X}/\sqrt{Z}$, where $\mathbf{X}$ is SN, $Z \sim \chi^2_\nu/\nu$, χ^2_ν denoting a chi-squared distribution with ν degrees of freedom, and $\mathbf{X}$ is independent from Z. As with the Student-t distribution, ν controls the heavy-tailed behavior, and the ST tends to the SN as $\nu \to \infty$.

3.3 | Biresolution skew-*t* model

Since the work of Kim and Mallick (2004), the first to develop a spatial model with skew-normal finite-dimensional distributions (FDDs), several studies have followed proposing alternative spatial models that capture marginal skewness and possibly excess kurtosis; notably, Zhang and El-Shaarawi (2010) developed the so-called skew-Gaussian model, which addressed the issue of nonvanishing spatial dependence of Kim and Mallick (2004), at the expense of only having marginal distributions belong to the SN class. Mahmoudian (2017) considered a model whose FDDs were SN, based on the Sahu, Dey, and Branco (2003) parameterization, with log-normal scale mixing, as in Palacios and Steel (2006), to capture heavy-tailed behavior. Morris, Reich, Thibaud, and Cooley (2017) proposed a model with ST FDDs for the modeling of spatial extremes, using a randomized partitions of the domain to permit long-distance asymptotic independence. Tagle, Castruccio, Crippa, and Genton (2019) considered independent multivariate ST distributions in a regional gridded setting. More recently, Bevilacqua, Caamaño-Carrillo, Arellano-Valle, and Morales-Onñate (2020) extended the work of Zhang and El-Shaarawi (2010), by adding scale mixing in the form of a gamma process, thus achieving ST marginals.

Tagle, Castruccio, and Genton (2019) proposed an ST spatial model, where a large-scale effect interacted additively with locally stationary spatial processes whose finite dimensional distributions where multivariate SN, which after gamma scale mixing became multivariate ST distributed. More formally, given a partition of the spatial domain $\{\mathcal{D}_r\}_{r=1,\dots,R}$, the model considered a random vector $\mathbf{U}_0 = (U_{0,1}, \dots, U_{0,R})^\top \sim \mathcal{N}_R(\mathbf{0}, \boldsymbol{\Sigma}_0)$, interpreted as a large-scale effect, and a collection of regionally referenced random variables $U_{1,r}$, Z_r, and Gaussian processes η_r with isotropic correlation functions $C_{\psi_r}(\cdot)$, all independent, such that for a given location $\mathbf{s} \in \mathcal{D}_r$,

$$Y(\mathbf{s}) = \frac{\rho_r U_{0,r} + \lambda_r |U_{1,r}| + \eta_r(\mathbf{s})}{\sqrt{Z_r}}, \tag{1}$$

with $\rho_r \geq 0$ and $\lambda_r \in \mathbb{R}$. The sum $\lambda_r|U_{1,r}| + \eta_r(\mathbf{s})$, with $U_{1,r} \sim \mathcal{N}(0,1)$ independent across r, has an SN distribution; and by the properties of the SN distribution, the linear combination with $U_{0,r}$ is also SN distributed. The assumption that $Z_r \sim \text{Gamma}(\nu_r/2, \nu_r/2)$ completed the model, endowing $Y(\mathbf{s})$ with an ST distribution. While the model intuitively expresses the biresolution interaction, inference becomes challenging for datasets beyond several hundred locations. For the high-resolution dataset that motivates this work, we examine an alternative construction that retains the flexibility in capturing interregional spatial dependence, but with a reduction in the number of latent processes. Specifically, we remove the hierarchy represented by $\mathbf{U}_0$, introducing the vector directly within the SN construction, that is,

$$Y(\mathbf{s}) = \frac{\lambda_r |U_r| + \eta(\mathbf{s})}{\sqrt{Z_r}}. \tag{2}$$

Here for simplicity we have used the notation U_r instead of $U_{1,r}$, although while in (1) this latent effect was assumed independent across regions, now it has a dependence structure across regions similar to $U_{0,r}$, as will be explained in detail later. It is easy to show that $\sqrt{1+\lambda_r^2}Y(\mathbf{s})$ has a standard univariate skew-t distribution, denoted by $\mathcal{ST}(0,1,\alpha)$, where here $\alpha=\lambda_r$. The mean and variance of $Y(\mathbf{s})$, as well as its kth moment, then follow from the standard results of the skew-t distribution (see chap 4 of Azzalini & Capitanio, 2014). We include the former here for completeness,

$$\begin{aligned}\mathrm{E}[Y(\mathbf{s})] &= \frac{(\nu_r/2)^{1/2}\Gamma((\nu_r-1)/2)}{\Gamma(\nu_r/2)}\lambda_r\sqrt{\frac{2}{\pi}}\\ \mathrm{Var}[Y(\mathbf{s})] &= \frac{\nu_r}{\nu_r-2}\left(\lambda_r^2+1\right)-\mathrm{E}[Y(\mathbf{s})]^2.\end{aligned}$$

The kth moment for a standard skew-t random variable X is given by

$$\mathrm{E}[X^k]=\mathrm{E}[Z^{-k/2}]\mathrm{E}[V^k],$$

where we define V as having standard skew-normal distribution and Z is equal in distribution with the Z_r. Regarding the first term, it is well known that

$$\mathrm{E}[Z^{-k/2}]=\frac{(\nu/2)^{k/2}\Gamma((\nu-k)/2)}{\Gamma(\nu/2)},$$

while for the second,

$$\mathrm{E}[X^k]=\sqrt{\frac{2}{\pi}}\frac{\mathrm{sng}(\alpha)}{\alpha^{k+1}}B_k(\alpha^{-2}),$$

where for $h>0$,

$$B_k(h)=\frac{k-1}{h}B_{k-2}(h)+\frac{\beta_{k-1}}{h(1+h)^{k/2}},\quad \text{for } k=2,3,\dots,\text{and}$$

$$B_0(h)=\sqrt{\frac{\pi}{2h}},\quad B_1(h)=\frac{1}{h\sqrt{1+h}},$$

and β_k is the kth moment of the standard normal distribution, $\beta_k=(k-1)!!$, for $k=2,4,6,\dots$ and 0 otherwise. From this it follows that

$$\mathrm{E}[Y(\mathbf{s})^k]=(1+\lambda_r^2)^{-k/2}\frac{(\nu_r/2)^{k/2}\Gamma((\nu_r-k)/2)}{\Gamma(\nu_r/2)}\sqrt{\frac{2}{\pi}}\frac{\mathrm{sng}(\alpha_r)}{\alpha_r^{k+1}}B_k(\alpha_r^{-2}).$$

The covariance at locations $\mathbf{s}_1,\mathbf{s}_2\in\mathcal{D}_r$ is given by

$$\mathrm{Cov}(Y(\mathbf{s}_1),Y(\mathbf{s}_2))=\frac{\nu_r}{\nu_r-2}\left(\lambda_r^2+C_{\theta_r}(h)\right)-\mathrm{E}[Y(\mathbf{s}_1)]^2,$$

while if $\mathbf{s}_i\in\mathcal{D}_{r_i}, i=1,2$, that is, the two locations belong to different regions,

$$\mathrm{Cov}(Y(\mathbf{s}_1),Y(\mathbf{s}_2))=\rho_{r_1}\rho_{r_2}\frac{(\nu_{r_1}/2)^{1/2}\Gamma((\nu_{r_1}-1)/2)}{\Gamma(\nu_{r_1}/2)}\frac{(\nu_{r_2}/2)^{1/2}\Gamma((\nu_{r_2}-1)/2)}{\Gamma(\nu_{r_2}/2)}\mathrm{Cov}(|U_1|,|U_2|),$$

with

$$\mathrm{Cov}(|U_1|,|U_2|)=\frac{2}{\pi}\left(\sqrt{1-\rho_{12}^2}+\rho_{12}\arcsin(\rho_{12})-1\right),\quad \rho_{12}=(\boldsymbol{\Sigma}_0)_{1,2}.$$

Given a collection of points $\mathbf{s}_1, \dots, \mathbf{s}_{n_r} \in \mathcal{D}_r$, the finite dimensional distribution of $\mathbf{Y} = (Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_{n_r}))^\top$ is multivariate ST, and analogously to the univariate case, $\sqrt{1+\lambda_r^2}\,\mathbf{Y}$ is $\mathcal{ST}_{n_r}(\mathbf{0}, \boldsymbol{\Omega}_r, \boldsymbol{\alpha}_r, \nu_r)$, with

$$\boldsymbol{\alpha}_r^\top = \frac{\delta_r}{1-\delta_r^2} \frac{\mathbf{1}_{n_r}^\top \boldsymbol{\Omega}_r^{-1}}{\left(1 + \frac{\delta_r^2}{1-\delta_r^2} \mathbf{1}_{n_r}^\top \boldsymbol{\Omega}_r^{-1} \mathbf{1}_{n_r}\right)^{1/2}},$$

$$\boldsymbol{\Omega}_r = (1-\delta_r^2)\left(\boldsymbol{\Sigma}(\boldsymbol{\psi}_r) + \frac{\delta_r^2}{1-\delta_r^2}\mathbf{1}_{n_r}\mathbf{1}_{n_r}^\top\right),$$

$\boldsymbol{\Sigma}(\boldsymbol{\psi}_r)$ is the correlation matrix associated with η_r, and $\delta_r = \lambda_r/\sqrt{1+\lambda_r^2}$. The collection of vectors $\mathbf{Y} = (\mathbf{Y}_1^\top, \dots, \mathbf{Y}_R^\top)^\top$ has a convenient hierarchical representation,

$$\begin{aligned} &\mathbf{Y}_r | U_r = u_r, Z_r = z_r \overset{\text{iid}}{\sim} \mathcal{N}_{n_r}\left(\frac{\lambda_r |u_r|}{\sqrt{z_r}}, \frac{1}{z_r}\boldsymbol{\Sigma}(\boldsymbol{\psi}_r)\right), \\ &\mathbf{U} \sim \mathcal{N}_R(\mathbf{0}, \boldsymbol{\Sigma}_0), \\ &Z_r \overset{\text{iid}}{\sim} \chi_\nu^2/\nu_r. \end{aligned} \tag{3}$$

If we assume that $\mathbf{Z} = (Z_1, \dots, Z_R)^\top$ and $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_0(\boldsymbol{\psi}_0)$, then the model proposed in (3) involves a vector $(\mathbf{U}^\top, \mathbf{Z}^\top)^\top$ of latent random variables of length $2R$; and the parameters of the model are $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \dots, \boldsymbol{\theta}_R^\top, \boldsymbol{\psi}_0)^\top$, where $\boldsymbol{\theta}_r = (\lambda_r, \nu_r, \boldsymbol{\psi}_r)^\top$.

3.4 | Simulation study

We conduct a simulation study in order to measure the impact of using a traditional Gaussian random field model (henceforth GAU) in place of the proposed model (SKT) when the latter represents the true data generating process. Our interest in characterizing wind energy potential suggests the use of a model-selection metric more appropriate to this setting, for which we choose wind power density (WPD), which corresponds to the wind power that is available per unit area swept by the turbine, and is given by $\frac{1}{2}\rho w^3$, where ρ represents the air density and w is the wind speed.

We consider a spatial design consisting of $R = 50$ regions, each composed of $n_r = 15$ points, whose locations are chosen at random from a 1×1 square. Taken together, the 50 regions are arranged within an 8×7 grid. We assume the same region-specific parameter values across the 50 regions for the true SKT model. In particular, we fix $\lambda_r = 0.8$, $\nu_r = 8$, and assume an exponential correlation function for the η_r, with range parameter $\phi_r = 0.2$; clearly, this implies that the intraregional FDDs are the same. As noted above, these are multivariate ST whose marginal distributions can be readily obtained, and for the given parameter values are of the form depicted in Figure 2a along with a comparison with a Gaussian with the same mean and variance. Figure 2b depicts the intraregional correlation function.

The choice of parameter values, with their implied skewness of 0.42 and excess kurtosis of 1.79 aim to be representative of those found in the Saudi wind dataset. Finally, the range parameter governing the large-scale effect is fixed at $\phi_0 = 2$. From this SKT model, we generate 130 independent spatial replicates that will serve as the true data, and to which both the SKT and GAU model are fit. We generated several sets of simulations but found that the analysis below was largely insensitive to their sampling variability. From each of the fitted models, 2,190 spatial replicates were then generated, equivalent to 6 years of daily data, so as to match the temporal dimension of the Saudi wind dataset. The procedure whereby these innovations are transformed into wind speed, or rather, how wind speed is deconstructed to arrive at such a set of innovations, is explained in detail in the next section; here we provide a brief summary with some simplifying assumptions. First, the time series of wind speed at each location is standardized assuming an annual cycle in both the mean and standard deviation. The resulting vector of residuals is assumed to follow a vector autoregressive process of order 2 in the actual dataset, but here we overlook the effect of neighboring locations and consider a univariate AR(2) process at each location, whose innovations are either the SKT or GAU innovations generated above. Thus, working backward, from the set of innovations we reconstruct the AR(2) process at each location and apply the inverse of the noted standardization. The particular values of the AR(2) coefficients and the annual cycle in the mean

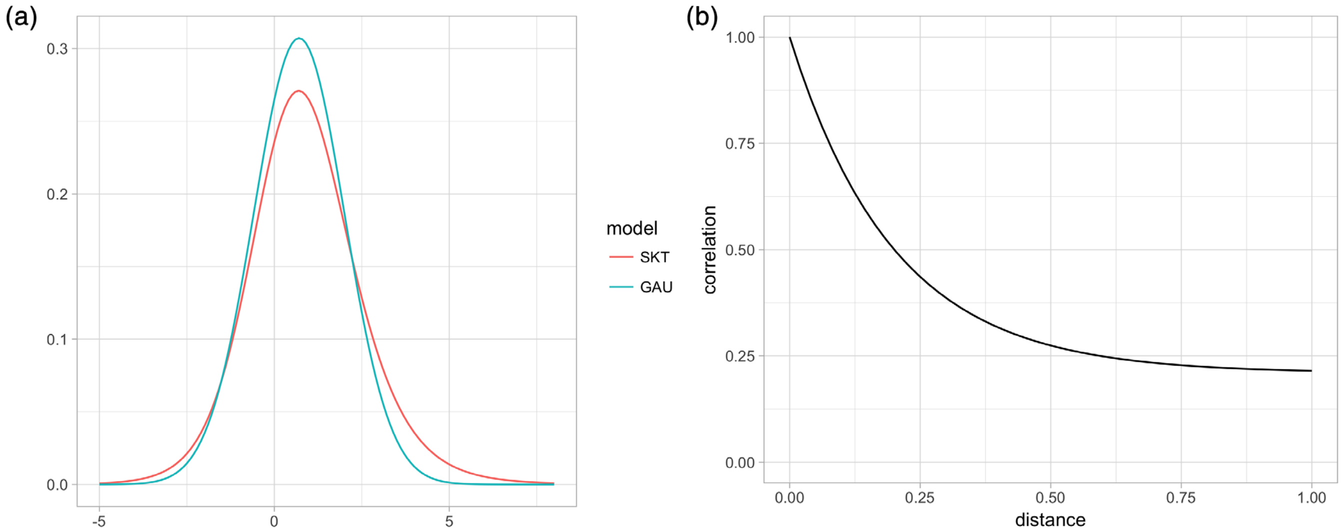

FIGURE 2 (a) Marginal distribution of the (red) SKT model with parameters $\lambda_r = 0.8$, $\nu_r = 8$ and, for comparison, a Gaussian distribution with matching mean and standard deviation and (b) the implied correlation function of the SKT model

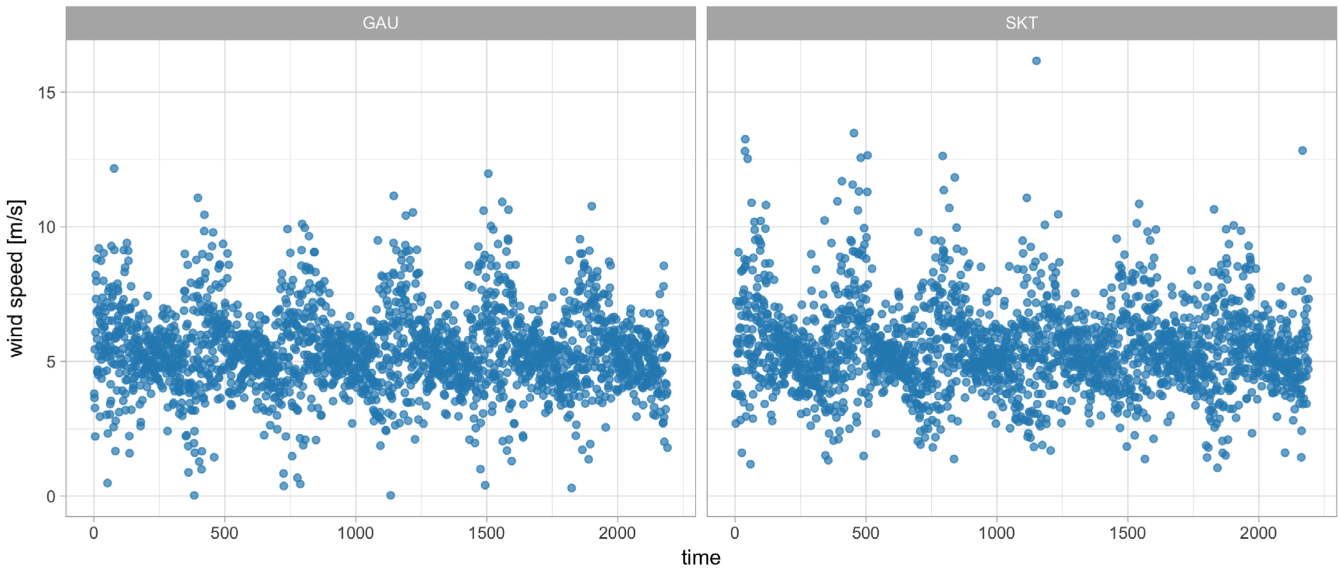

FIGURE 3 Sample time series of the synthetically generated 100-m wind speed for the GAU (left panel) and SKT models (right panel)

and standard deviation are chosen, as in the case of the true SKT model, to achieve a certain degree of comparability with the Saudi wind dataset. With this in mind, we selected five sets of annual cycles and AR(2) coefficients from across the country to capture different wind regimes (see Figure S.1 and Table S.1). Since these synthetic time series of wind speed are based on a mean structure estimated from near surface wind speed, these are extrapolated to 100 m—a standard hub height of present day turbines—using the power law method (Pryor & Barthelmie, 2011), whose details are provided in the next section. Figure 3 displays a sample of the wind speed time series from both models, where the influence of a heavier right tail in the case of the SKT manifests itself through the presence of the occasional outlier in excess of 12.5 m/s. As a final step, the 100-m wind speeds are transformed to WPD using said formula, with a fixed value of $\rho = 1.225\,\text{kg/m}^3$, which has been used in previous wind-related studies (e.g., Tagle, Castruccio, Crippa, & Genton, 2019).

Figure 4 presents the kernel density estimates for the mean WPD at each location for both models across the five sets of mean structure configurations. We see that in all cases, the WPD for the GAU model systematically underestimates WPD, with the mean WPD on average falling below that of the SKT by between 7 and 12%.

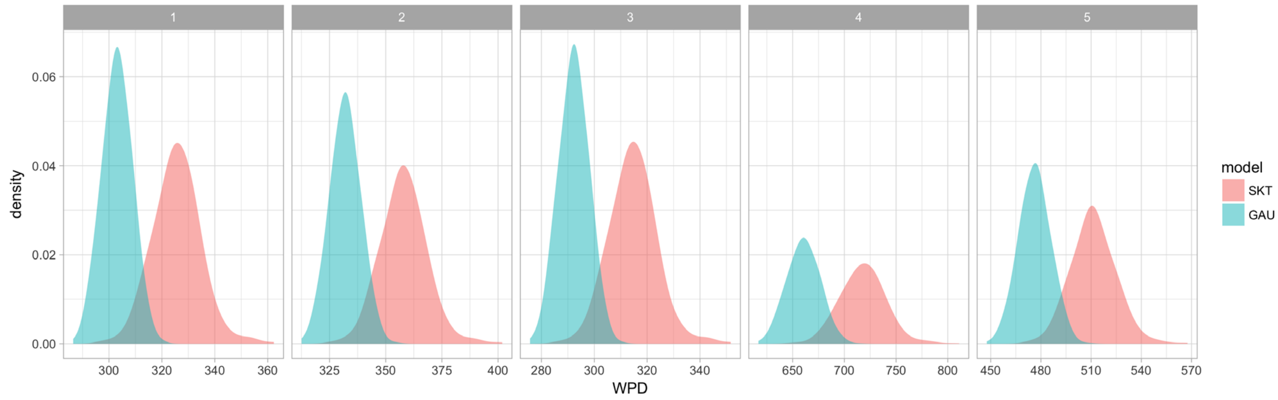

FIGURE 4 Kernel density estimates of the mean WPD at each location, for the GAU (green) and SKT (red) models

4 | ASSESSING WIND ENERGY OVER SAUDI ARABIA

We now proceed with the application with the data presented in Section 2. Our approach will focus on modeling the spatiotemporal dependence of this simulation, by first removing seasonalities for each location, then modeling the temporal dependence, and then applying the biresolution spatial skew-t model introduced in Section 3 to the innovations.

4.1 | Spatiotemporal model

Our focus is on the n-vector ($n = \sum_{r=1}^{R} n_r$) of daily wind speeds over the domain at day t, denoted by $\mathbf{W}_t = (W_{1,t}, \dots, W_{n,t})^\top$. As a preliminary step, we standardize the vector $\mathbf{W}_t$ by a location-scale transformation of the form $(W_{i,t} - \mu_{i,t})/\sigma_{i,m(t)}$, $i = 1, \dots, n$, where $\mu_{i,t}$ represents an annual cycle and $\sigma_{i,m(t)}$ is a monthly indexed scale factor. The annual cycle is approximated by regressing the time series of daily wind speed at each grid point on a small set of harmonics (1–5 cycles per year) that capture both annual and seasonal variation. The scale factor corresponds to the standard deviation of the daily residuals over each month; the notation $m(t)$ indicates a mapping from t to the respective calendar month. Henceforth, for notational purposes, we assume that such a transformation has been applied when referencing the vector $\mathbf{W}_t$. We propose the following second-order vector-autoregressive model (VAR(2)):

$$\mathbf{W}_t = \mathbf{A}_1\mathbf{W}_{t-1} + \mathbf{A}_2\mathbf{W}_{t-2} + \boldsymbol{\varepsilon}_t, \quad t = 3, \dots, T, \tag{4}$$

where $\mathbf{A}_k$, $k = 1, 2$, are $n \times n$ coefficient matrices and $\boldsymbol{\varepsilon}_t$ corresponds to the time t innovations. We find that a second-order autoregressive structure adequately captures the short-term serial dependence in wind speeds, in agreement with other wind speed modeling studies (e.g., Ailliot & Monbet, 2012; Brown, Katz, & Murphy, 1984), and we found no strong indication of temporal nonstationarity (see Figure S.2). Estimation of the entries of $\mathbf{A}_k$, $k = 1, 2$ has long been a problematic issue when the dimension is large. As in Tagle, Castruccio, Crippa, and Genton (2019), we find that a first-order stencil neighborhood scheme, that is, first-order neighbors for $\mathbf{A}_1$ and a diagonal form for $\mathbf{A}_2$, adequately represents the latter dependence structure. Figure S.3 in the Supplementary Material displays estimates for the diagonal elements of both autoregressive matrices. Figure 5 displays skewness and excess kurtosis estimates for the time series of residuals at each gridpoint. We perform D'Agostino (D'Agostino, 1970) and Anscombe-Glynn (Anscombe & Glynn, 1983) tests to assess the significance of the respective sample estimates (see Figure S.4). We find that the p-values for the sample skewness are generally all below 1%, and those for kurtosis, over 70% are also below 1%, thus motivating the use of a non-Gaussian modeling approach to the collection of VAR(2) residuals $\boldsymbol{\varepsilon}_t$. As apparent in the panels, both quantities closely follow the topography of the country. The winds with high skewness and excess kurtosis are located in the western mountain ranges (Hejaz and Asir), near the Red Sea, while winds in the flatlands, such as the Empty Quarter on the border with Oman, do not show strong indications of skewness or kurtosis.

An assumption of stationarity for a spatial domain of this size would naturally be inappropriate. To account for the spatial nonstationarity, we proceed to partition the domain into regions where such an assumption would be plausible, whose spatial behavior we model with the skew-t model described in Section 3. The EM algorithm assigns a multivariate normal

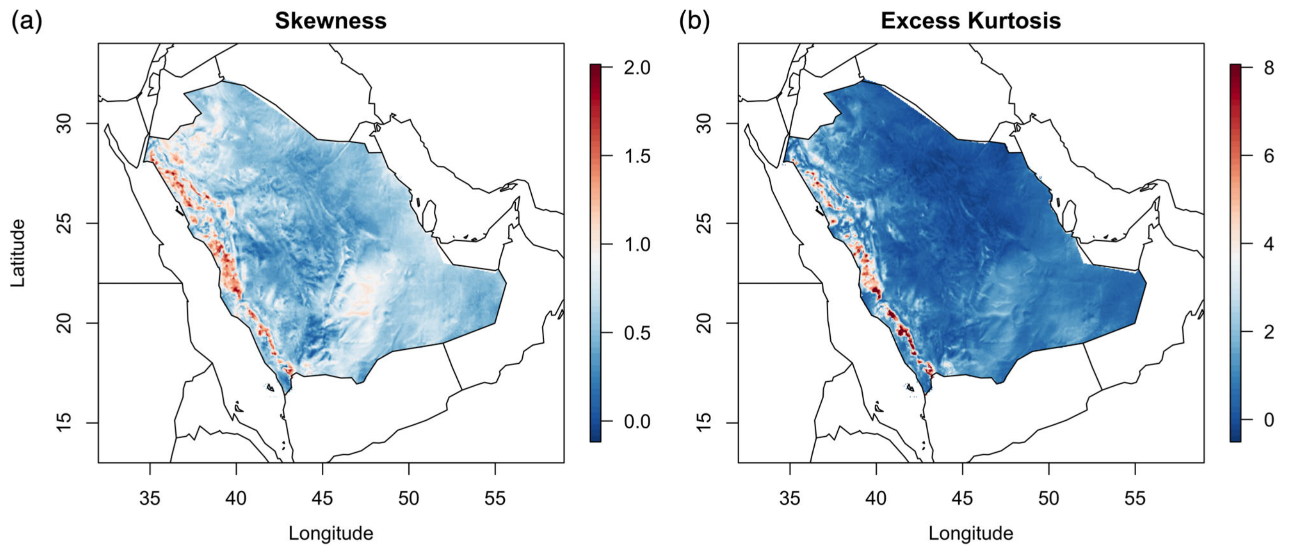

FIGURE 5 (a) Skewness and (b) excess kurtosis of VAR(2) residuals over Saudi Arabia

distribution to each subdomain; thus, the computational aspects of operating with the associated covariance matrices become relevant at this stage. At the same time, the subdomains should be large enough to ensure the identifiability of the stationary covariance model parameters. Exploratory analysis reveals a certain difficulty in meeting these two conditions over the south-eastern part of the country, where nonnegligible spatial dependence persists at distances implying computationally infeasible cluster sizes. While this implies a small amount of model misspecification, the considered area is entirely inhabited and comprises entirely of sand desert (*Rub' al Khali*, or *Empty Quarter* in English), so this does not affect our conclusions as no wind turbine can be physically installed there. We use a Ward's hierarchical clustering (Ward Jr., 1963) for the choice of clusters; 200 clusters achieves a reasonable compromise among the noted considerations (see Figure S.5). Thus, the vector $\boldsymbol{\varepsilon}_t$ is divided into $R = 200$ subvectors, $\boldsymbol{\varepsilon}_t = (\boldsymbol{\varepsilon}_{1,t}^\top, \ldots, \boldsymbol{\varepsilon}_{R,t}^\top)^\top$, where each $\boldsymbol{\varepsilon}_{r,t}$ is composed of n_r gridpoints, each with an $\mathcal{ST}_{n_r}(\boldsymbol{\xi}_r, \boldsymbol{\Omega}_r, \boldsymbol{\alpha}_r, \nu_r)$. Finally, before proceeding with the model fitting, we standardize the residuals by the $\boldsymbol{\xi}_r$ and $\boldsymbol{\omega}_r$ parameters, from $\boldsymbol{\varepsilon}_t = \boldsymbol{\xi}_r + \boldsymbol{\omega}_r\tilde{\boldsymbol{\varepsilon}}_t$, where $\tilde{\boldsymbol{\varepsilon}}_t \sim \mathcal{ST}_{n_r}(\mathbf{0}, \overline{\boldsymbol{\Omega}}_r, \boldsymbol{\alpha}_r, \nu_r)$, with $\boldsymbol{\Omega}_r = \boldsymbol{\omega}_r^\top\overline{\boldsymbol{\Omega}}_r\boldsymbol{\omega}_r$. However, these do not correspond in an explicit manner to the mean and standard deviation of the time series of residuals. In fact, for the ST parameterization, these parameters do not bare any direct relationship with the first two moments of the distribution.

To address this issue, we consider an alternative parameterization based on the first four cumulants of the ST distribution, which does provide estimates for $\boldsymbol{\xi}_r$ and $\boldsymbol{\omega}_r$, where $r = 1, \ldots, R$ (see Supplementary Material for details). This same procedure provides initial estimates of λ_r, $\boldsymbol{\psi}_r$, and ν_r. We assume a Matérn correlation function for both $\boldsymbol{\Sigma}_0$ and each η_r and fix the value of its smoothness parameter at 1.5. The distances used in the computation of $\boldsymbol{\Sigma}_0$ correspond to the regional centroid distances, while the initial values of the range parameter are obtained analogously considering the regionally averaged time series and associated empirical estimate of the correlation matrix.

From the fitted model, we generate 100 simulations of length 6 years, replicating the time period of the original dataset, from which the following assessment metrics are computed.

4.2 | Assessment metrics

As in the simulation study, we use WPDs as the primary measure to assess the wind energy potential over Saudi Arabia. Recall that to extrapolate near surface wind measurements to the different hub heights we use the power law method; this method assumes that the vertical wind profile at height z is given by $w(z) = w(z_r)(z/z_r)^a$, where z_r is a reference height and a is the power law exponent, which we fix at $1/7$ as it is appropriate over open land surfaces (Pryor & Barthelmie, 2011) and used in prior studies on wind in Saudi Arabia (Rehman et al., 2007). The value of the air density, ρ, is extracted from the modern-era Retrospective analysis for research and applications, version 2 (MERRA2, Gelaro et al., 2017) 2D surface turbulent-flux diagnostics at an hourly frequency, which we then average to obtain a daily mean value.

We compute the daily WPD for each of the 100 simulations at three hub heights, 50, 80, and 140 m, representing the hub heights of different generations of turbines. From these time series, we derive the relevant assessment metrics of the wind resource as suggested by Gunturu and Schlosser (2012) and analyzed in Yip et al. (2016) for the Arabian Peninsula, and consider their pointwise averages. In particular, we examine the abundance of the wind resource as characterized traditionally by the time-averaged daily WPD over the record period. In addition, we consider the variability of the wind energy resource, represented by the robust coefficient of variation (rCV), defined as

$$\mathrm{rCV} = \frac{\mathrm{median}(|\mathrm{WPD}_t - \mathrm{WPD}_{\mathrm{med}}|)}{\mathrm{WPD}_{\mathrm{med}}},$$

where $\mathrm{WPD}_{\mathrm{med}}$ represents the median of WPD time series. Finally, we consider the intermittency, which consists in the availability and persistence of the wind resource. Availability is defined as the fraction of the record period when the WPD exceeds a threshold deemed representative of a lower bound for commercial-scale power production. Here, we assign 200 $\mathrm{W/m^2}$ as the threshold, based on recent advances in low-wind turbine technology. Persistence supplements the notion of availability by measuring the median episode length (MEL), defined as the median of all episodes where the WPD was consecutively above the noted threshold.

5 | RESULTS

Yip et al. (2016) identified the Asir and Hejaz mountain ranges in western Saudi Arabia and their respective eastern plateaus as potentially rich wind-resource areas, with an average WPD between 140 and 200 $\mathrm{W/m^2}$ at 50-m above ground level, in contrast with a former consensus based on meteorological station data that posited coastal areas possessed more abundant wind resources (e.g., Rehman, 2005). The finer spatial resolution of the WRF dataset provides a higher degree of spatial variability that enables us to refine their findings by identifying the interior (eastern) escarpments of the latter mountain ranges as well the high ridges of the adjacent plateaus as the most wind-abundant areas in these regions (Figure 6a,b, see Figures S.6a,b and S.7a,b for results at 50- and 140-m hub heights). These areas cover a wide extent of the country, and the average WPD is found to be higher than originally estimated by the aforementioned work, with values exceeding 400 $\mathrm{W/m^2}$. The stochastic generator, as described in Section 3, allows us to assess the uncertainty of our estimates and their variability across space, as measured by the standard deviation across the noted simulations (see Figure 6c,d, and Figures S.6c,d and S.7c,d for 50 and 140 m). Over the areas of greatest abundance, the variability reaches 70 $\mathrm{W/m^2}$, though the variability notably decreases outside of the Asir mountain region. To further illustrate the value of the stochastic generator, we focus on the data for two spatial locations (indicated in the detail of Figure 6b,d) over the dates July 1–20, which falls during a period of seasonal high winds caused by the Asian summer monsoon. We compare the daily winds from the WRF output and the 100 realizations of the stochastic generator with a Q–Q plot. The results shown in Figure 6e,f highlight how the statistical model is able to adequately capture the true variability of the WRF-simulated wind speed.

Compared with Yip et al. (2016), the WPD rCV is of lower magnitude, which can be attributed to the use of daily instead of the hourly frequency considered therein. The rCV averages 0.54 across the domain, with the exception of the ridges along the border with Jordan where it exceeds 0.8 (see Figure S.8). The most salient spatial feature is the marked differences between the western escarpments of the Asir and Hejaz mountains and their eastern plateaus. The variability across simulations, as computed from the stochastic generator of this metric, were below 1% of the point value and, hence, are negligible given the spatial variability of the rCV.

The availability of the wind resource is an estimate of the percentage of time that a wind turbine would be operational. The spatial pattern closely resembles that of abundance (Figure 7a,b), with availability exceeding 50% along the areas of greatest abundance. As in the case of abundance, the effect of hub height plays an important role, with availability exceeding 80% in many areas for the highest hub height (see Figures S.9a,b and S.10a,b). Here, persistence of the wind-energy potential is measured in days as opposed to the more traditional hours because the daily frequency of the model output. In Figure 7c,d, we find that the median length for the more abundant areas is about 3–4 days at 50-m hub height (Figure S.9c,d), and up to 7 days at 140-m hug height (Figure S.10c,d). By comparing the annual and seasonal cycles, we find that the uncertainty is negligible, so the maps are not reported.

5.1 | Feasibility of the "Vision 2030" energy plan

In order to provide the most appropriate assessment of the uncertainty in the "Vision 2030" energy plan, parametric estimation uncertainty must be accounted for. The most theoretically appropriate approach to account for (asymptotic)

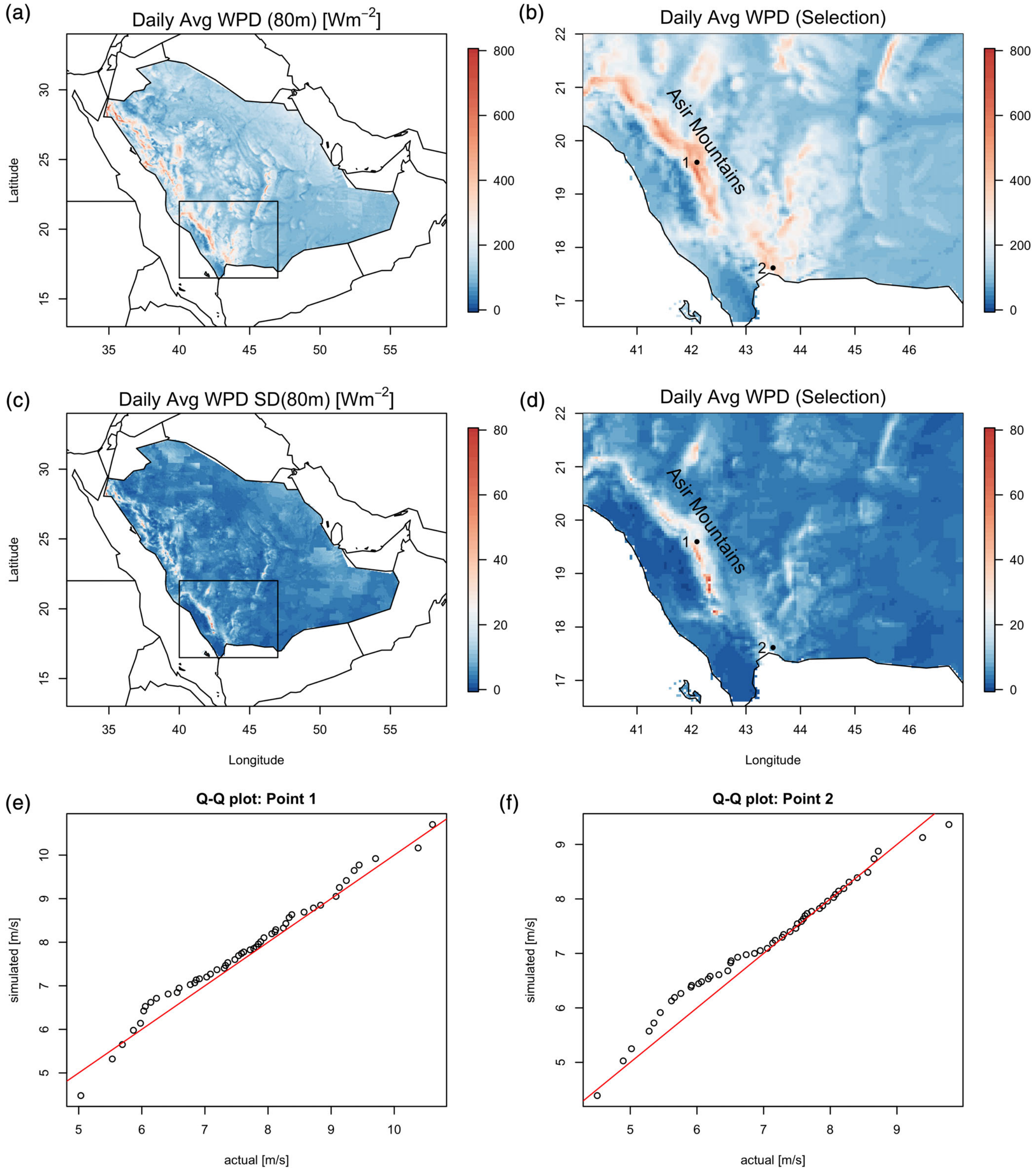

FIGURE 6 (a, b) Daily average and (c, d) standard deviation of wind-power density (WPD) at 80 m for 2009–2014, with detail. (e, f) Q–Q plot for the actual and simulated daily wind speeds (July 1–20, for all years), for the two locations indicated in (b) and (d), which are representative of regions with high WPD

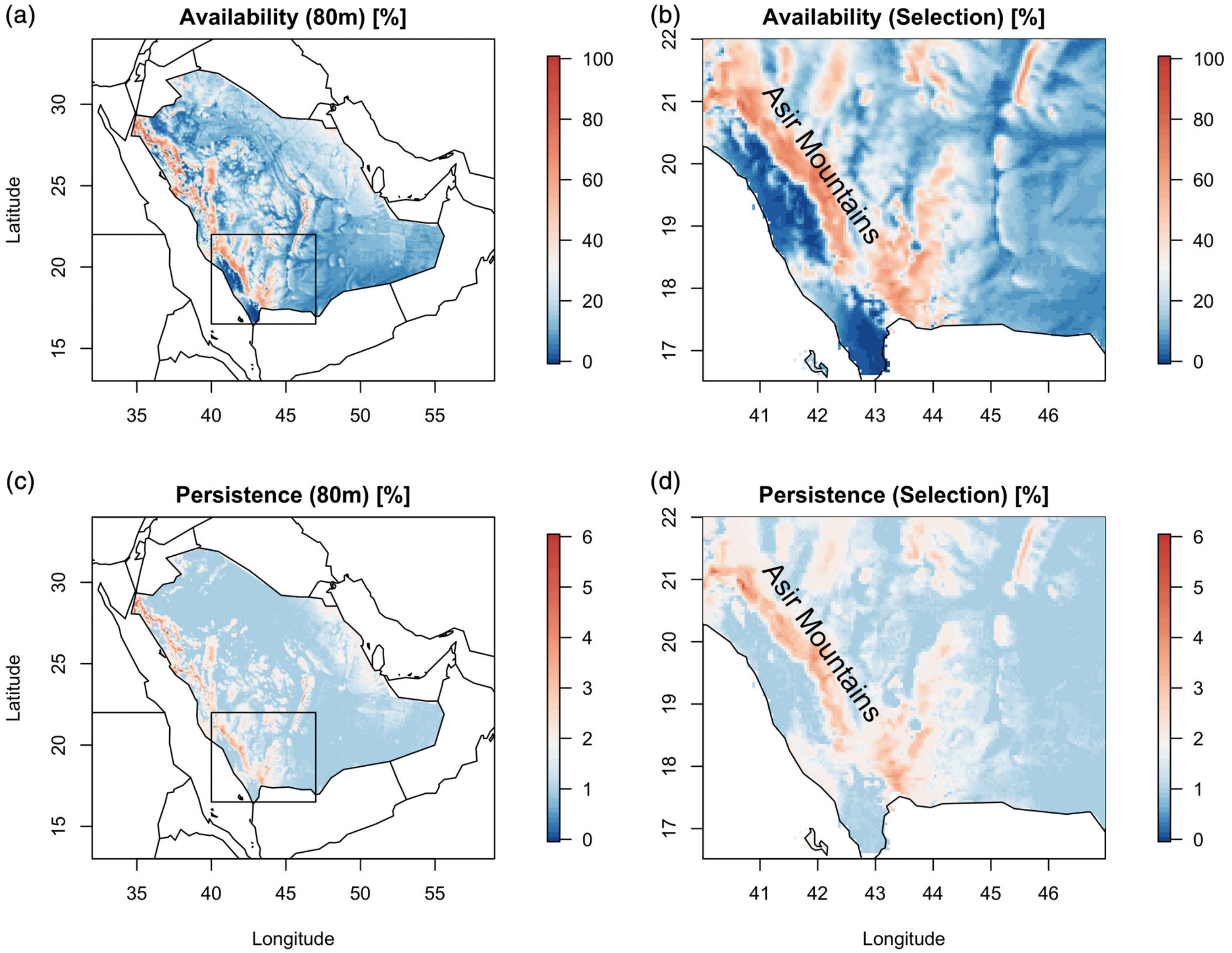

FIGURE 7 (a, b) Availability and (c, d) persistence of WPD, defined as the fraction of time during which the WPD exceeds a threshold of 200 W/ m^2, and the median length of time the WPD remains above the threshold, respectively, at an 80-m hub height, with detail

parameter uncertainty is to implement a supplemented expectation-maximization algorithm (Meng & Rubin, 1991), but this approach is too computationally expensive for a dataset of this size. Therefore, we propose a more practical approach, which exploits the model structure. Since $\boldsymbol{\theta}_r$ is expected to have similar value in regions neighboring r, once $\hat{\boldsymbol{\theta}}_r$ is computed, it is assumed to be the value at the centroid, and kriging with a Matérn function is applied throughout the regions. The parametric uncertainty is now reflected in the range of possible values of the parameters across the region. The simulation from the statistical model is now performed by sampling $\hat{\boldsymbol{\theta}}_r$ across this range of values, and then conditional on this value, a surrogate simulation is performed.

In 2016, the estimated electricity demand of Saudi Arabia was 287.7 TWh (SAMA, 2016). Here, we provide an estimate of the wind-power output that would be available if wind farms were placed in those areas identified above as most abundant, that is, where WPD exceeds 350 W/m^2. At each gridcell, we assume the installation of a wind farm with 10 Vestas V126-3.45MW turbines (rotor diameter = 126 m), the current version of a model widely deployed in nearby countries (e.g., Vestas, 2016, see Figure S.11 for the power curve) from the global leader in manufacturing wind turbines. Typical wind-farm configurations consider a spacing between wind turbines of seven times wider than the turbine-rotor diameter (7D); however, recent research suggests that a spacing of up to 15D may be more suitable for modern wind farms (Meyers & Meneveau, 2012). We adopt a slightly larger spacing of approximately 17D to account for the complex terrain of these areas, which results in the 10 turbine configuration per gridcell and a total installed capacity of 4.4 GW.

Figure 8a displays the daily average wind-power output in giga watt-hour at each gridpoint of the subdomain, assuming the above wind farm set up. Contours denote the areas with high potential wind-energy output, largely in excess of

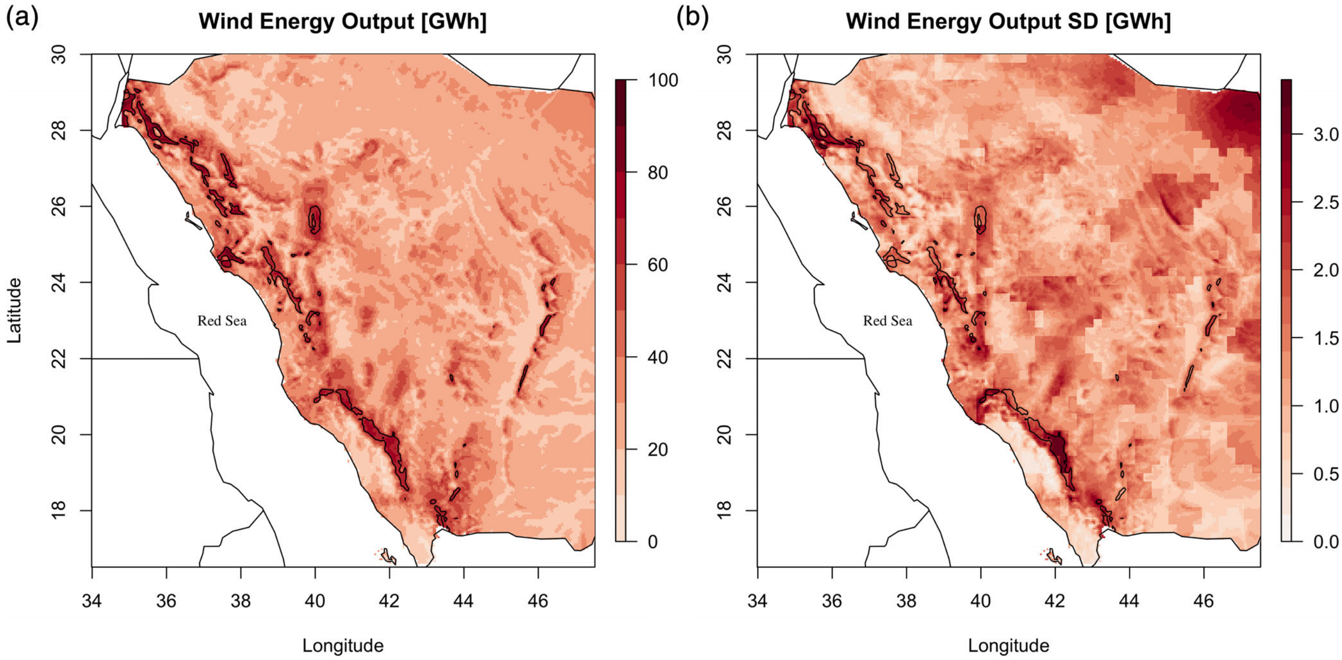

FIGURE 8 (a) Daily average wind-energy output and (b) standard deviation based on a wind farm configuration consisting of 10 Vestas V126-3.45MW turbines at each gridcell. The contours denote areas where the daily average WPD is above 350 W/m^2 (see Figure 6)

40 GWh. Considered together, these areas represent only about 1.5%, of the country's surface area, and could produce approximately 88 TWh, or over 30% of Saudi Arabia's annual electricity needs. If we increase the number of turbines per gridcell by reducing the spacing from 17D to 10D, this estimate would increase to over 70%. Since this estimate is sensitive to internal climate variability, as captured by the variability across the simulations, we find that the 2.5 and 97.5 percentiles, respectively, correspond to 29.0 and 32.1% of Saudi Arabia's annual electricity needs for 17D, and 66.8 and 73.9% for 10D. Under the more conservative setting of 17D, 12,800 turbines could be installed across the whole abundant area for a total capacity of 44.2 GW. Only 2,750 turbines would be required to reach 9.5 GW, the Vision 2030 goal for renewable energy from all sources, using wind energy alone.

The proposed model allows for spatially coherent surrogate simulations. For some metrics focused on the marginal distributions, such as the daily wind averages and its uncertainty in Figure 7, it is in principle possible to produce similar results with a spatially independent model. The final assessment of the aggregate potential contribution of wind energy to the country's portfolio from Figure 8, however, necessitates a proper modeling of the spatiotemporal dependence structure. Therefore, besides the visual appeal of spatially correlated simulations, as opposed to unphysical simulations with no spatial dependence, the proposed model allows an appropriate assessment of the uncertainty of aggregated metrics.

While these results strongly support the feasibility of the Vision 2030 target, one intrinsic limitation of this study is the temporal resolution. Indeed, for a more detailed assessment of turbine power hourly data instead of daily would be necessary, as intraday variability due to temperature change and local turbulence could potentially bias our estimates.

6 | CONCLUSIONS

In this work, we developed a biresolution stochastic generator of daily wind fields designed for an extremely large spatiotemporal dataset arising from a high-resolution weather simulation. Our stochastic generator enables the modeling of both small-scale features and large-scale interactions, while also allowing a flexible skew-t distribution for the marginals. We applied the model in a wind-energy assessment of Saudi Arabia using a high-resolution dataset derived from WRF, a state-of-the-art NWP model, and ECMWF reanalysis data for the boundary conditions.

Our results identified the eastern escarpments of the Asir and Hejaz mountain ranges and the ridges of the Jabal Tuwayq as the most abundant wind-resource areas. Together, these areas have the potential to provide enough wind energy to meet the goals as stated in the flagship "Vision 2030" blueprint. The uncertainties of all the assessment metrics were computed using the aforementioned stochastic generator; the resulting standard deviation maps and marginal histograms provide additional information for the engineers and policymakers reviewing the upcoming feasibility study regarding the installation of wind turbines.

Our findings indicate a wind-resource abundance that in many areas exceeded the estimates in previous studies using coarser resolution datasets (e.g., Yip et al., 2016). While large magnitudes are expected at a higher resolution, our estimates may have been partially contaminated by the use of $1/7$ as the exponent in the WPD calculation, which previously led to the overestimation of wind resources over areas with complex surface characteristics (Schwartz & Elliott, 2006). Future studies will incorporate some degree of uncertainty in the extrapolation from wind surface to hub height.

As these aspects are closely related to the operation of wind turbines, it would be beneficial but challenging to extend the model to an hourly frequency. This might provide more relevant assessments of the metrics, but would imply a 24-fold increase in size of an already massive dataset, as well as the need to model more complex spatiotemporal behavior arising at shorter time-scales. In situ short-term wind measurement campaigns are necessary to validate the results of our study, though these are not likely to be performed in the near future. Microscale numerical models are also necessary to better capture the details of the complex topography in the mountainous areas, achieve a more accurate representation of the intense surface heating in the region, and properly parameterize the stability of the boundary layer.

ACKNOWLEDGEMENT

The research reported in this publication was supported by funding from the KAUST Office of Sponsored Research (OSR) Competitive Research Grants (CRG) under award no. OSR-2015-CRG4-2640.

ORCID

Marc G. Genton https://orcid.org/0000-0001-6467-2998
Stefano Castruccio https://orcid.org/0000-0002-6728-965X

REFERENCES

Ailliot, P., & Monbet, V. (2012). Markov-switching autoregressive models for wind time series. *Environmental Modelling & Software*, *30*, 92–101.

Anscombe, F. J., & Glynn, W. J. (1983). Distribution of the kurtosis statistic b 2 for normal samples. *Biometrika*, *70*(1), 227–234.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, *12*, 171–178.

Azzalini, A., & Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*(2), 367–389.

Azzalini, A., & Capitanio, A. (2014). *The skew-normal and related families IMS monograph series*. Cambridge, MA: Cambridge University Press.

Azzalini, A., & Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, *83*(4), 715–726.

Bessac, J., Ailliot, P., & Monbet, V. (2015). Gaussian linear state-space model for wind fields in the North-East Atlantic. *Environmetrics*, *26*(1), 29–38.

Bevilacqua, M., Caamaño-Carrillo, C., Arellano-Valle, R. B., & Morales-Oññate, V. (2020). Non-gaussian geostatistical modeling using (skew) t processes. *Scandinavian Journal of Statistics*. https://doi.org/10.1111/sjos.12447Y.

Brown, B. G., Katz, R. W., & Murphy, A. H. (1984). Time series models to simulate and forecast wind speed and wind power. *Journal of Climate and Applied Meteorology*, *23*(8), 1184–1195.

Bryden, J., Riahi, L., & Zissler, R. (2013). MENA renewables status report. *Report for REN21*, *2013*, 1–40.

Chen, W., Castruccio, S., Genton, M. G., & Crippa, P. (2018). Current and future estimates of wind energy potential over Saudi Arabia. *Journal of Geophysical Research: Atmospheres*, *123*(12), 6443–6459.

Cressie, N. (1993). *Statistics for spatial data*. New York, NY: Wiley.

D'Agostino, R. B. (1970). Transformation to normality of the null distribution of g1. *Biometrika*, *57*(3), 679–681.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, *39*, 1–38.

Dumat Al Jandal (2020). Retrieved from https://dumataljandal.com/.

Fassò, A., & Finazzi, F. (2011). Maximum likelihood estimation of the dynamic coregionalization model with heterotopic data. *Environmetrics*, *22*(6), 735–748.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454.

Gelfand, A. E., & Schliep, E. M. (2016). Spatial statistics and gaussian processes: A beautiful marriage. *Spatial Statistics*, *18*, 86–104.

Genton, M. G. (2004). *Skew-elliptical distributions and their applications: A journey beyond normality*. Boca Raton, FL: Chapman & Hall/CRC Press.

Gunturu, U. B., & Schlosser, C. A. (2012). Characterization of wind power resource in the United States. *Atmospheric Chemistry and Physics*, *12*(20), 9687–9702.

Heaton, M. J., Datta, A., Finley, A., Furrer, R., Guinness, J., Guhaniyogi, R., et al. (2019). A case study competition among methods for analyzing large spatial data. *Journal of Agricultural Biological and Environmental Statistics*, *24*(3), 398–425.

Hering, A. S., & Genton, M. G. (2010). Powering up with space-time wind forecasting. *Journal of the American Statistical Association*, *105*(489), 92–104.

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P., et al. (2013). The community earth system model: A framework for collaborative research. *Bulletin of the American Meteorological Society*, *94*(9), 1339–1360.

Jeong, J., Castruccio, S., Crippa, P., & Genton, M. G. (2018). Reducing storage of global wind ensembles with stochastic generators. *Annals of Applied Statistics*, *12*(1), 490–509.

Jeong, J., Yan, Y., Castruccio, S., & Genton, M. G. (2019). A stochastic generator of global monthly wind energy with Tukey *g*-and-*h* autoregressive processes. *Statistica Sinica*, *29*, 1105–1126.

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., et al. (2015). The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, *96*(8), 1333–1349.

Kim, H.-M., & Mallick, B. K. (2004). A Bayesian prediction using the skew Gaussian distribution. *Journal of Statistical Planning and Inference*, *120*(1), 85–101.

Mahmoudian, B. (2017). A Skewed and heavy-tailed latent random field model for spatial extremes. *Journal of Computational and Graphical Statistics*, *26*(3), 658–670.

McDermott, P. L., & Wikle, C. K. (2019). Deep echo state networks with uncertainty quantification for spatio-temporal forecasting. *Environmetrics*, *30*(3), e2553.

Meng, X.-L., & Rubin, D. B. (1991). Using EM to obtain asymptotic variance-covariance matrices: The SEM algorithm. *Journal of the American Statistical Association*, *86*(416), 899–909.

Meyers, J., & Meneveau, C. (2012). Optimal turbine spacing in fully developed wind farm boundary layers. *Wind Energy*, *15*(2), 305–317.

Morris, S. A., Reich, B. J., Thibaud, E., & Cooley, D. (2017). A space-time skew-t model for threshold exceedances. *Biometrics*, *73*(3), 749–758.

Palacios, M. B., & Steel, M. F. J. (2006). Non-Gaussian Bayesian geostatistical modeling. *Journal of the American Statistical Association*, *101*(474), 604–618.

Pinson, P. (2013). Wind energy: Forecasting challenges for its operational management. *Statistical Science*, *28*(4), 564–585.

Pryor, S. C., & Barthelmie, R. J. (2011). Assessing climate change impacts on the near-term stability of the wind energy resource over the united states. *Proceedings of the National Academy of Sciences*, *108*(20), 8167–8171.

Rehman, S. (2005). Prospects of wind farm development in Saudi Arabia. *Renewable Energy*, *30*(3), 447–463.

Rehman, S., & Ahmad, A. (2004). Assessment of wind energy potential for coastal locations of the Kingdom of Saudi Arabia. *Energy*, *29*(8), 1105–1115.

Rehman, S., El-Amin, I., Ahmad, F., Shaahid, S., Al-Shehri, A., & Bakhashwain, J. (2007). Wind power resource assessment for Rafha, Saudi Arabia. *Renewable and Sustainable Energy Reviews*, *11*(5), 937–950.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., et al. (2011). MERRA: NASA's modern-era retrospective analysis for research and applications. *Journal of Climate*, *24*(14), 3624–3648.

Sahu, S. K., Dey, D. K., & Branco, M. D. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *Canadian Journal of Statistics*, *31*(2), 129–150.

SAMA (2016). Annual statistics. Economic Report, Saudi Arabian Monetary Authority. http://www.sama.gov.sa/en-us/economicreports/pages/yearlystatistics.aspx

Schwartz, M. and Elliott, D. (2006). Wind shear characteristics at central plains tall towers: Preprint, Technical report, National Renewable Energy Laboratory (NREL), Golden, CO.

Shaahid, S., Al-Hadhrami, L. M., & Rahman, M. (2014). Potential of establishment of wind farms in western Province of Saudi Arabia. *Energy Procedia*, *52*, 497–505.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Wang, W. and Powers, J. G. (2008). A description of the advanced research WRF version 3', University Corporation for Atmospheric Research. NCAR Technical note -475+STR.

Tadayon, V., & Torabi, M. (2019). Spatial models for non-gaussian data with covariate measurement error. *Environmetrics*, *30*(3), e2545.

Tagle, F., Castruccio, S., Crippa, P., & Genton, M. G. (2019). A non-Gaussian spatio-temporal model for daily wind speeds based on a multi-variate skew-*t* distribution. *Journal of Time Series Analysis*, *40*, 312–326.

Tagle, F., Castruccio, S., & Genton, M. G. (2019). A hierarchical bi-resolution spatial skew-*t* model. *Spatial Statistics*, *35*, 100398.

Vestas (2016). Vestas wins 89 MW EPC project in Jordan. *Press Release*. Retrieved from https://www.vestas.com/en/media/~/media/e2d7dabc808242a2a09104788861a458.ashx

Vision 2030 (2016). http://vision2030.gov.sa/en/node/87.

Ward, J. H., Jr. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*, 236–244.

WindEurope (2018). Wind power in 2017.

WWEA (2018). Wind power capacity reaches 539 gw, 52.6gw added in 2017'.

Xu, G., & Genton, M. G. (2017). Tukey *g*-and-*h* random fields. *Journal of the American Statistical Association*, *112*, 1236–1249.

Yan, Y., & Genton, M. G. (2019). Non-Gaussian autoregressive processes with Tukey *g*-and-*h* transformations. *Environmetrics*, *30*(2), e2503.

Yip, C. M. A. (2018). *Statistical characteristics and mapping of near-surface and elevated wind resources in the Middle East* (PhD thesis). King Abdullah University of Science and Technology.

Yip, C. M. A., Gunturu, U. B., & Stenchikov, G. L. (2016). Wind resource characterization in the Arabian Peninsula. *Applied Energy*, *164*, 826–836.

Zhang, H. (2007). Maximum-likelihood estimation for multivariate spatial linear coregionalization models. *Environmetrics*, *18*(2), 125–139.

Zhang, H., & El-Shaarawi, A. (2010). On spatial skew-Gaussian processes and applications. *Environmetrics*, *21*(1), 33–47.

Zhu, X., & Genton, M. G. (2012). Short-term wind speed forecasting for power system operations. *International Statistical Review*, *80*(1), 2–23.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Tagle F, Genton MG, Yip A, Mostamandi S, Stenchikov G, Castruccio S. A high-resolution bilevel skew-*t* stochastic generator for assessing Saudi Arabia's wind energy resources. *Environmetrics*. 2020;31:e2628. https://doi.org/10.1002/env.2628