ASSESSING THE RELIABILITY OF WIND POWER OPERATIONS UNDER A CHANGING CLIMATE WITH A NON-GAUSSIAN BIAS CORRECTION

BY JIACHEN ZHANG[1,*], PAOLA CRIPPA[2], MARC G. GENTON[3] AND
STEFANO CASTRUCCIO[1,†]

[1]*Department of Applied and Computational Mathematics and Statistics, University of Notre Dame,* *jzhang19@nd.edu;*
†*scastruc@nd.edu*

[2]*Department of Civil and Environmental Engineering and Geosciences, University of Notre Dame, pcrippa@nd.edu*

[3]*Statistics Program, King Abdullah University of Science and Technology, marc.genton@kaust.edu.sa*

Facing increasing societal and economic pressure, many countries have established strategies to develop renewable energy portfolios whose penetration in the market can alleviate the dependence on fossil fuels. In the case of wind, there is a fundamental question related to the resilience and hence profitability of future wind farms to a changing climate, given that current wind turbines have lifespans of up to 30 years. In this work we develop a new non-Gaussian method to adjust assimilated observational data to simulations and to estimate future wind, predicated on a trans-Gaussian transformation and a clusterwise minimization of the Kullback–Leibler divergence. Future winds abundance will be determined for Saudi Arabia, a country with a recently established plan to develop a portfolio of up to 16 GW of wind energy. Further, we estimate the change in profits over future decades using additional high-resolution simulations, an improved method for vertical wind extrapolation and power curves from a collection of popular wind turbines. We find an overall increase in daily profit of $272,000 for the wind energy market for the optimal locations for wind farming in the country.

1. Introduction. The vast evidence of the negative effects of fossil fuel emissions (IPCC (2014)) calls for a major systemic change in current strategies to produce and distribute energy throughout the world. Societies worldwide are adapting by developing renewable alternatives to reduce dependence on fossil fuels and to align with the standards imposed by the Paris Agreement (Kinley (2017)). Wind has been the natural resource with the largest share of power generation worldwide, with the United States and China being the two major leaders (REN21 Secretariat (2018)). In the United States, renewable energies, with wind energy being a major contributor, are predicted to surpass coal in terms of share of the energy market by the end of this decade. Similarly, while European countries have an overall smaller absolute installed capacity, the penetration of the energy in the national grid is in percentage considerable, with a peak of 42% in Denmark.

The wind energy sector is at its early stages in developing countries such as Saudi Arabia. Despite being one of the countries with the highest per capita energy consumption (World Bank (2020)), the sixth largest consumer of oil worldwide (British Petroleum (2020)) and the largest in the Gulf Cooperating Council (GCC) (International Renewable Energy Agency (2018)), the transition to renewable energy is very recent and almost exclusively focused on solar energy (International Renewable Energy Agency (2019)). Currently Saudi Arabia's contribution to the GCC's renewable energy portfolio amounts to 16% of the total capacity, with only 0.2% available for sharing and only 2% generated from wind (International Renewable Energy Agency (2018)). As part of the recently outlined "Vision 2030" plan (Nurunnabi

(2017)), Saudi Arabia aims to generate 16 GW of wind energy (NREP (2018)), positioning the country as one of the major global wind energy suppliers and considerably contributing to the planned emissions reduction targets stipulated in the United Nations Framework Convention on Climate Change (UNFCCC (2020)). Investments in wind energy would also result in the increased reliability of renewable energy; for example, solar energy is only available during daytime, whereas winds often peak at nighttime.

Under the aforementioned scenario, a comprehensive analysis must be conducted to identify optimal sites for developing wind farms based on a cost-benefit analysis. Several recent studies with global climate models at annual (Jeong et al. (2018)), at monthly (Jeong et al. (2019)) scale and later at daily level (Tagle et al. (2019, 2020)), including validation of observational and simulated data sets (Chen et al. (2018)) and extreme wind conditions (Chen, Castruccio and Genton (2021)), provided initial evidence about the availability of sufficient wind for installing wind turbines. Very recently, Giani et al. (2020) conducted the first full feasibility study with a new high-resolution ensemble and identified the optimal locations and types of turbine based on maintenance and operational costs. While their study provided the first detailed assessment of the country's resources, it relied on only four years of data, from 2013 to 2016, due to computational and storage constraints. This timescale is inconsistent with the current lifespan of wind turbines, which can be as long as 30 years, and the multidecadal effects of climate change could have an impact on the wind resource availability.

In order to provide a pathway for a feasible and robust implementation of a wind energy portfolio, the question of resilience under a changing climate must be addressed. Given the computational impossibility to simulate high resolution numerical simulations for decades, an alternative strategy focused on publicly available data must be devised. In this work we focus on developing a methodology for assessing future winds, predicated upon the estimation of the relationship between simulated and observed data for a historical period at the same spatial and temporal scales. Under the assumption of an enduring relation between these two data sets, future wind behaviors are estimated by applying the same relation to future climate simulations.

Methods for adjusting observations and simulations have a long history in geoscience (see, e.g., Kim, Kwon and Han (2015), Yuan et al. (2019) and Hawkins et al. (2013), Ho et al. (2012) for a general review) and bear some resemblance but also substantial differences with methodologies for calibrating observations with numerical simulations. Adjustment methods can be generally divided into two categories: observation-driven and simulation-driven.

Observation-driven approaches focus on adjusting observations based on estimated changes in simulations. This *Delta method* or *change factor* (Hawkins et al. (2013)) has been subject to some recent generalizations beyond a simple mean correction. Indeed, Leeds, Moyer and Stein (2015) proposed a spectral-based approach to the adjustment of time series, thus resulting in an adjustment of the implied covariance structure. At the core of this method is the notion of (penalized) estimation of the ratio between the spectral densities of observations and simulations. Poppick et al. (2016) proposed a generalization of the aforementioned approach to a transient climate and multiple simulations under different future scenarios.

Simulation-driven methods are instead focused on estimating an empirical relationship between historical observation and simulation and use it to correct future simulations. This *bias correction* approach, in its simplest form, focuses on an individual time series, estimates the differences between observational and simulated historical climate and applies this difference to future simulations to obtain the future observations. Slightly more sophisticated approaches also involve the use of variance (see Section 3 for a comprehensive review) or transforming the quantiles of the distribution (*quantile mapping*, Cannon, Sobie and Murdock (2015)). In more recent years, more studies have acknowledged the need for methods of bias adjustment for spatially distributed (and possibly multivariate) data. Nguyen, Mehrotra

and Sharma (2019) proposed a frequency based adjustment of spatial data. Among others, Cannon (2018), Mehrotra and Sharma (2016) and Cannon, Piani and Sippel (2020) extended quantile mapping to multivariate time series, and François et al. (2020) provided a recent overview and comparison of multivariate bias correction methods.

In this work we propose a new bias correction approach based on a non-Gaussian clusterwise spatial transformation which is estimated by minimizing the distance between the joint distribution of the observational and of the simulated data field. Our proposed approach is based on three fundamental ideas: 1) non-Gaussianity can be accounted for through a marginal transformation which is slowly varying in space; 2) the transformation to Gaussianity can be inferred by minimizing the distributional distance between the two fields as long as this value is comparable and, possibly, within the confidence interval of the maximum likelihood estimate and 3) transformation to Gaussianity allows a simple adjustment of first and second moments and a back-transformation to the original scale.

Section 2 introduces the data sets used for this study, validates the historical wind speeds across Saudi Arabia and proposes a model for the mean and temporal dependence. Section 3 reviews current approaches to observation-simulation corrections. Section 4 introduces the proposed methodology for non-Gaussian adjustment. Section 5 validates the model with a simulation study with non-Gaussian random fields and historical data. Section 6 applies the proposed methodology, along with vertical wind extrapolation and power curve evaluation, to estimate the change in daily profits from future winds in Saudi Arabia. Section 7 concludes with a discussion. The code for this work is available in the Supplementary Material (Zhang et al. (2021)) and online at the following GitHub repository: github.com/Env-an-Stat-group/21.Zhang.AoAS.

2. Data and preprocessing. In this study we focus on daily wind speed data at 10 meters above the ground level in Saudi Arabia which is bounded approximately by longitudes of 34°E–56°E and latitudes of 16°N–33°N (see Figure S1(a–b)). Wind speed is derived as the Euclidean norm from the zonal and meridional velocity (i.e., over the x and y axes). We present the data set of the power curves along with the information related to the cost of energy in the Supplementary Material.

2.1. *Observational data.* We use the Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2, Gelaro et al. (2017)), available from 1980 to the present day. Reanalysis data consists of observational data assimilated to a numerical weather model and, in the geoscience community, this data product is considered the best representation of the state of the Earth's system. MERRA-2 is the reference observational data set used in our study and is available on a regular grid with a resolution of $0.625° \times 0.5°$ in longitude and latitude, respectively. We only use daily wind speed data from 1980 to 2005, for a total of 26 years, to match the simulation data sets presented in Section 2.2. There are $n = 614$ locations in Saudi Arabia at MERRA-2 resolution for a total of $T = 9497$ days. Throughout this manuscript we will denote the MERRA-2 wind fields as $W_O(\mathbf{s}_i, t_j)$, where the subscript O indicates the observation.

2.2. *Regional simulations.* We use the simulations from the Coordinated Regional Climate Downscaling Experiment (CORDEX) which is a set of coordinated regional experiments. Specifically, we focus on the Middle East North Africa (MENA) CORDEX Program, and among the five available simulations, we select CORDEX-4, which exhibits the best agreement with MERRA-2 in Saudi Arabia, as demonstrated by Chen et al. (2018). The simulation resolution is $0.22° \times 0.22°$, finer than MERRA-2, and boundary conditions for the historical and future periods are provided by the Geophysical Fluid Dynamics Laboratory

global coupled climate-carbon Earth System Models (GFDL-ESM2M, Dunne et al. (2013)). Future GFDL-ESM2M runs are simulated under the Representative Concentration Pathways 8.5 (RCP 8.5, van Vuuren et al. (2011)) scenario in which the global radiative forcing is assumed to increase by 8.5 W/m^2 by 2100 (Taylor, Stouffer and Meehl (2012)). We regrid and upscale the MENA CORDEX data to the MERRA-2 resolution by considering the average of locations in MENA CORDEX that are within the range of two consecutive MERRA-2 grids. Even though the historical run spans from 1950 to 2005, we only consider data from 1980 to 2005 to align the observation window with MERRA-2; for future runs we only consider simulations in the near future for the same number of years, that is, from 2025 to 2050. We will denote the MENA-CORDEX wind fields as $W_S(\mathbf{s}_i, t_j)$, where the subscript S indicates the simulation.

2.3. *Model for the mean and temporal dependence.* We assume a periodic climatology described by K harmonics in order to account for the interannual wind variability. If we denote the wind speed for location $\mathbf{s}$ and time t by $W(\mathbf{s}, t)$, its interannual variability across the year can be written as

$$W(\mathbf{s}_i, t_j) = \mu(\mathbf{s}_i, t_j) + \sum_{i'=1}^{P} \phi_{i',i} \varepsilon(\mathbf{s}_i, t_{j-i'}) + \varepsilon(\mathbf{s}_i, t_j),$$

(1)

$$\mu(\mathbf{s}_i, t_j) = \omega_i * \mathrm{yr}(t_j) + \sum_{k=1}^{K} \left\{ \beta_{k,i} \sin\left(\frac{2\pi k t_j}{\delta}\right) + \beta'_{k,i} \cos\left(\frac{2\pi k t_j}{\delta}\right) \right\},$$

where $\delta \in \{365, 366\}$, depending on whether the year is nonleap or leap, $\mathbf{s}_i = (x_i, y_i)$, $i = 1, \ldots, n$, $j = 1, \ldots, T$, and $\mathrm{yr}(t)$ represents the year of day t. Thus, the model assumes a location-specific linear annual trend with K harmonics to explain interannual variability and P autoregressive coefficients. We further assume that $\varepsilon(\mathbf{s}_i, t_j)$ is independent across t_j with a spatial dependence that will be specified later. The mean and the autoregressive coefficients are estimated through site-specific inference, initially by assuming Gaussian errors and independence in time for estimating $\{\omega_i, \beta_{k,i}, \beta'_{k,i}\}$ and subsequently estimating ϕ_i by maximum likelihood. In the Supplementary Material we present diagnostics to show that: 1) the linear trend ω_i is significant for the majority of points in both data sets (Figure S3); 2) three harmonics ($K = 3$) are sufficient to explain the climatology for both data sets (Figure S4); 3) the parameters $\beta_{k,i}$, $\beta'_{k,i}$, ϕ_i are constant in time (Figures S5, S6 and S7) and 4) the residuals of (1) are uncorrelated in time (Figure S8).

3. Review of adjustment approaches. MERRA-2 represents the state of the system assimilated from observations; it can be observed up to the present, and its future needs to be estimated. The simulations obtained from MENA CORDEX can be used as a proxy to assess the future wind speed. From the preliminary analysis in the Supplementary Material, a systematic mismatch between the two data sets is apparent. The objective of this study is to provide an adjustment of this mismatch in the historical period and, employing the available future MENA CORDEX simulations, use it to predict future MERRA-2 observations. While the methods are general, for consistency with the previous section we will denote the spatiotemporal wind fields as $\mathbf{W}_O^{(H)}$ and $\mathbf{W}_S^{(H)}$ for MERRA-2 (O for observations) and MENA CORDEX (S for simulations), respectively, for the historical period (1980–2005). The future data will be denoted as $\mathbf{W}_O^{(F)}$ and $\mathbf{W}_S^{(F)}$. Throughout this section we will refer to MERRA-2 as "observations" and MENA CORDEX as "simulations" to emphasize the generality of our methodology.

3.1. *Correction for marginal distributions.* The simplest approach to achieve adjusted future simulations is bias correction, which assumes a simple additive bias between observations and simulations, estimates it in the historical period and then adjusts future simulations. Bias correction has been widely used in a number of studies in geoscience (see, e.g., Chen, Pryor and Li (2012), Hemer, McInnes and Ranasinghe (2012) for wind applications). This approach assumes that there is no difference in marginal variance, spatial covariance or other high-order moments. Formally, we assume that the observation-corrected data for a location $\mathbf{s}_i$ can be expressed as

$$(2) \qquad W_O^{(F)}(\mathbf{s}_i, t_j) = W_S^{(F)}(\mathbf{s}_i, t_j) + \{\mu_O^{(H)}(\mathbf{s}_i, t_j) - \mu_S^{(H)}(\mathbf{s}_i, t_j)\},$$

where t_j refers to the jth time for the historical or future time period and $\mu_\ell^{(H)}(\mathbf{s}_i, t_j)$ for $\ell = \{O, S\}$ is the mean in the historical period (from 1980 to 2005), which is estimated via the location-specific inference of the mean parameters in (1), as described in Section 2.

Two data sets typically exhibit differences also in their temporal variability (see Figure S2), which cannot be accounted for by a simple bias correction. Therefore, a relatively more articulated approach focuses on adjusting both the mean and variance (Li et al. (2019), Teutschbein and Seibert (2012)). Formally,

$$(3) \qquad W_O^{(F)}(\mathbf{s}_i, t_j) = \mu_O^{(H)}(\mathbf{s}_i, t_j) + \frac{\sigma_O^{(H)}(\mathbf{s}_i)}{\sigma_S^{(H)}(\mathbf{s}_i)}\{W_S^{(F)}(\mathbf{s}_i, t_j) - \mu_S^{(H)}(\mathbf{s}_i, t_j)\},$$

where $\sigma_\ell^{(H)}(\mathbf{s}_i)$ for $\ell = \{O, S\}$ is the standard deviation in the historical period and is estimated from the parameters of the autoregressive process in (1).

3.2. *Covariance adjustment.* In the previous section the adjustment was performed independently for every grid point, hence focusing only on the marginal distribution, without considering the potential dependence across multiple locations. However, there is strong evidence of spatial dependence in the observed and simulated field. Indeed, Figure S10 indicates a strong empirical spatial correlation between two neighboring points for both data sets. It is, therefore, of interest to develop methods to correct not just for the mean and variance but also for the spatial correlation.

A simple generalization of (3) allows for adjustment of both the mean and covariance matrix (therefore, including variance), and, under the assumption of Gaussianity for both observations and simulations, this would be sufficient to fully characterize a transformation for the joint distribution. Indeed, future observations can be expressed as

$$(4) \qquad \mathbf{W}_O^{(F)}(t_j) = \boldsymbol{\mu}_O^{(H)}(t_j) + (\boldsymbol{\Sigma}_O^{(H)})^{1/2}\{(\boldsymbol{\Sigma}_S^{(H)})^{-1/2}\}^\top\{\mathbf{W}_S^{(F)}(t_j) - \boldsymbol{\mu}_S^{(H)}(t_j)\},$$

where $\boldsymbol{\mu}_\ell^{(H)}(t_j) = (\mu_\ell^{(H)}(\mathbf{s}_1, t_j), \ldots, \mu_\ell^{(H)}(\mathbf{s}_n, t_j))^\top$ (with the same convention being used for the other quantities) and $\boldsymbol{\Sigma}_\ell^{(H)}$ for $\ell = \{O, S\}$ denotes the covariance matrices for the observational and simulated data sets in the historical period; the superscript $1/2$ denotes the Cholesky decomposition. While $\boldsymbol{\Sigma}_\ell^{(H)}$ could be estimated using a simple nonparametric sample covariance matrix, the relatively large number of locations ($n = 614$) would result in an unstable estimate. Instead, we assume spatial dependence via two parametric models in increasing order of complexity. First, we consider a stationary and isotropic covariance function modeled by the Matérn correlation (Stein (1999)), that is, if two measurement are separated by distance h, then their correlation is

$$(5) \qquad C(h) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\sqrt{2\nu}\frac{h}{\rho}\right)^\nu K_\nu\left(\sqrt{2\nu}\frac{h}{\rho}\right),$$

where $\rho > 0$ is the range parameter, $\nu > 0$ is the smoothness of the process and K_ν is the Bessel function of the second kind of order ν. Second, we consider a more articulated nonstationary covariance resulting from the kernel convolution representation of the random field (Higdon (2002)), which can be written in closed form under a general class of kernels as (Paciorek and Schervish (2006))

$$(6) \qquad C(\mathbf{s}, \mathbf{s}') = \sigma(\mathbf{s})\sigma(\mathbf{s}') \frac{|\boldsymbol{\Sigma}(\mathbf{s})|^{1/4}|\boldsymbol{\Sigma}(\mathbf{s}')|^{1/4}}{|\frac{\boldsymbol{\Sigma}(\mathbf{s})+\boldsymbol{\Sigma}(\mathbf{s}')}{2}|^{1/2}} g(\sqrt{Q(\mathbf{s}, \mathbf{s}')}),$$

where

$$Q(\mathbf{s}, \mathbf{s}') = (\mathbf{s} - \mathbf{s}')^\top \left(\frac{\boldsymbol{\Sigma}(\mathbf{s}) + \boldsymbol{\Sigma}(\mathbf{s}')}{2} \right)^{-1} (\mathbf{s} - \mathbf{s}').$$

In this study the function g is specified to be the exponential function (although several other alternatives are possible) with a spatially varying range. All the spatially varying parameters are defined through the mixture of fixed knots for A fixed locations $\mathbf{b}_a$, $a = 1, \dots, A$. Herein, we select $A = 4$, as a larger number of knots would imply a considerable increase in computational cost. The spatially varying variance is defined as

$$\sigma(\mathbf{s}) = \sum_{a=1}^{A} w_a(\mathbf{s})\sigma_a, \qquad w_a(\mathbf{s}) \propto \exp\left\{ -\frac{\|\mathbf{s} - \mathbf{b}_a\|^2}{2\lambda_\sigma} \right\},$$

where the weights are normalized and λ_σ could be estimated, but for simplicity it is here assumed to be fixed at one half of the minimum distance between the knots. A similar approach is used to define $\boldsymbol{\Sigma}(\mathbf{s})$ and the exponential function range. Inference is performed through local likelihood; see Risser and Calder (2017) for details.

4. Adjusting for non-Gaussian spatial data. The methods presented in Section 3 allow to correct the joint distribution under the assumption that the mean and covariance are the only quantities for which the adjustment is necessary. This is appropriate only for physical variables whose temporal aggregation is sufficiently large to ensure, at least, approximate Gaussianity. Wind data aggregated at daily or subdaily level are generally expected to exhibit non-Gaussian behaviors and, in particular, to be skewed to the right, owing to occasionally high values because of local meteorological events, such as storm fronts or persistent wind gusts. This is indeed the case in our application. Figure 1(a) shows the histogram of the MENA CORDEX data from a location at the northwest end of Saudi Arabia represented by a cross in Figure S1(a–b). The superimposed best fit for a Gaussian distribution is vastly inadequate, as it fails to capture the aforementioned right skew. Furthermore, the red boxplots in Figure 1(c–d) show the skewness and excess kurtosis for all locations in MENA CORDEX with a characteristic right skew and some degree of excess kurtosis.

We propose a correction method for non-Gaussian data based on marginal, spatially varying transformations. We denote this transformation to Gaussianity by g_λ, a function depending on λ, and we assume the same class of transformation by g_λ with $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)^\top$, indicating the elementwise transformation at each location. Future observations are then obtained as

$$(7) \quad \mathbf{W}_O^{(F)}(t_j) = g_{\lambda_O}^{-1}[\boldsymbol{\mu}_O^{(H)}(t_j) + (\boldsymbol{\Sigma}_O^{(H)})^{1/2}\{(\boldsymbol{\Sigma}_S^{(H)})^{-1/2}\}^\top \{g_{\lambda_S}(\mathbf{W}_S^{(F)})(t_j) - \boldsymbol{\mu}_S^{(H)}(t_j)\}],$$

where now $\boldsymbol{\mu}_\ell^{(H)}(t_j)$ and $\boldsymbol{\Sigma}_\ell^{(H)}$ for $\ell = \{O, S\}$ represent the mean and covariance matrix of the *transformed* process $g_{\lambda_\ell}(\mathbf{W}_\ell^{(H)})(t_j)$, respectively. In the next sections we will discuss the particular choice of transformation g_λ and the assumptions on $\boldsymbol{\lambda}$ to allow an approximately optimal choice across the entire spatial domain.
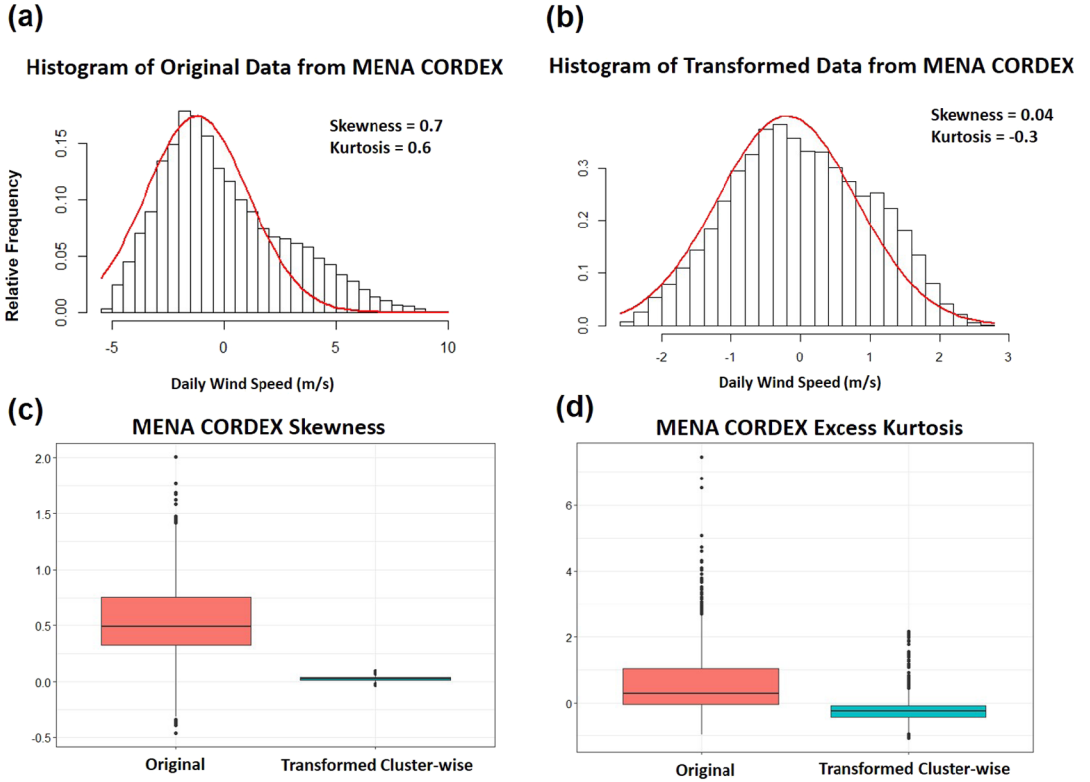
(a)

Histogram of Original Data from MENA CORDEX

Skewness = 0.7
Kurtosis = 0.6

(b)

Histogram of Transformed Data from MENA CORDEX

Skewness = 0.04
Kurtosis = -0.3

(c)

MENA CORDEX Skewness

(d)

MENA CORDEX Excess Kurtosis

FIG. 1. *Comparison of the MENA CORDEX data from* 1980 *to* 2005 *before and after transformation. Histogram of the* (*a*) *original and* (*b*) *transformed data at one selected location* (*see the cross in Figure S*1(*a–b*))*. The red line represents the Normal distribution that best fits the data. Boxplot of* (*c*) *skewness and* (*d*) *excess kurtosis at all locations.*

4.1. *Relative merits of the proposed approach and the Kennedy and O'Hagan framework.* The methodology we propose in this work bears some similarities but also differences with a widely used framework for coupling data and simulations initially proposed by Kennedy and O'Hagan (2000) and Kennedy and O'Hagan (2001). Indeed, their model can be written as

$$(8) \qquad W_O(\mathbf{s}) = \rho \cdot W_S(\mathbf{s}) + \delta(\mathbf{s}) + \epsilon(\mathbf{s}),$$

where the subscripts O and S indicate observations and simulations, respectively. In its simplest form this is a deterministic correction with ρ and $\delta(\mathbf{s})$, which is identifiable as long as ρ is constant (Tuo and Wu (2015)). Equation (4) resembles this expression, but it is more general, as it proposes a linear transformation of the entire spatial field. Indeed, the closest equivalent in the Kennedy and O'Hagan (2000) framework would be

$$(9) \qquad \mathbf{W}_O = \rho \cdot \mathbf{W}_S + \delta + \epsilon,$$

where $\mathbf{W}_O$, $\mathbf{W}_S$, δ and ϵ are n-dimensional vectors and $\rho = \rho \mathbf{I}_n$ is an $n \times n$ matrix. Our approach as detailed in equation (7) is more general, as it does not just assume a linear transformation but also a nonlinear function for both observations and simulations to achieve Gaussianity and hence, theoretically, justifies our transformation. It is, therefore, more similar in spirit to some of the most recent work in calibration using nonlinear functions, such as deep neural networks (Bhatnagar et al. (2020)).

Another important difference is that, while (8) articulates a stochastic relationship, our model is instead deterministic, as the adjustment parameters are chosen by simple linear

algebra operations to transform a Normal vector into another one and a (nonlinear) transformation to normality. Our choice of a deterministic relationship was mostly justified by the need to align our work with the current practice in bias correction.

On the computational side the choice of our deterministic transformation has the advantage of requiring an estimation of parameters from observations and simulations separately. In the context of our problem, where the observational data set has $n \times T = 614 \times 9497 \approx 5.8 \times 10^6$ observations, this is a very desirable feature especially for a nonstationary spatiotemporal model already considerably time consuming such as the one we propose.

Finally, our work focuses on bias correcting observations with a computer simulation, such as MENA CORDEX, for which we do not have control over the input parameters. This approach is extremely common in geoscience, and a wide range of literature can be found, mostly focused on Earth system models. The approach by Kennedy and O'Hagan (2000, 2001) is generally applied for calibration of computer models, that is, determining the best choice of numerical model input so that the simulations would resemble the observations. Calibration is typically performed assuming tens, or even hundreds of simulations are available for a given problem with input specified from some design criterion, such as a Latin hypercube design. This number is simply not achievable here; we only have five MENA CORDEX simulations with an extremely large input space, so in Chen et al. (2018) we chose the best one according to some in situ observations.

4.2. *The Yeo–Johnson transformation.* The standard Box–Cox transformation (Erdin, Frei and Künsch (2012)) cannot be used for g_λ, as our objective is to transform residuals, which are not necessarily nonnegative. Instead, we rely on a similar function, the Yeo-Johnson transformation (Yeo and Johnson (2000)),

$$g_\lambda(x) = \begin{cases} [(x+1)^\lambda - 1]/\lambda, & x \geq 0, \lambda \neq 0, \\ \log(x+1), & x \geq 0, \lambda = 0, \\ -[(-x+1)^{2-\lambda} - 1]/(2-\lambda), & x < 0, \lambda \neq 2, \\ -\log(-x+1), & x < 0, \lambda = 2. \end{cases}$$

While the standard approach for estimating λ is to maximize the likelihood of transforming the data to Gaussianity, in this work we propose a different inference approach for both observation and simulation so that their (joint) distributional distance, which in this work is measured as the Kullback–Leibler divergence (KL divergence) (Kullback and Leibler (1951); see Section 4.3)) is minimized. The key idea of our proposed approach is that, even if the parameter is estimated by minimization of KL divergence and not maximum likelihood estimator (MLE), their difference is not substantial. Indeed, a sizable number of sites (50% and 37% for observations and simulations, respectively) have clusterwise transformation parameters within the (asymptotic, likelihood based) 95% MLE confidence interval including the parameter obtained via distance minimization, despite its narrow range owing to the sample size from $T = 9497$ days (see Figure S11). Once the parameter has been inferred for both data sets and the transformation has been applied, the bias and covariance from the correction method in Section 3.2 are estimated under the assumption of Gaussianity, and the results are applied to the future simulated data to obtain an estimate for future observational data, as explained in equation (7). As an example, Figure 1(b) shows the result for the transformed wind field according to the method proposed in this study for one location, resulting in an approximately Gaussian distribution. Figure 1(c–d) show the skewness and excess kurtosis at all the locations before and after the transformation (with MLE, see Figure S12 for the results with KL divergence), and it is readily apparent how the transformation is sufficiently flexible to transform the data to Gaussianity at all locations.

4.3. *k-Nearest neighbors approximation of the Kullback–Leibler divergence.* The KL divergence is a measure of how one probability distribution differs from a second reference probability distribution. In this study we estimate the KL divergence between the simulations and observations, denoted as f_S and f_O, respectively,

$$(10) \qquad D_{\mathrm{KL}}(f_O \parallel f_S) = \int f_O(x) \log\left\{\frac{f_O(x)}{f_S(x)}\right\} dx.$$

In our case, a direct evaluation is computationally impossible, as comparing two multivariate distributions of dimension $n = 614$ would require an integration over the same number of dimension in (10). While a simplifying expression is available for the Gaussian distribution, it will not be used here as a comparison with different methods in the original non-Gaussian scale must be made. Therefore, we rely upon a numerical approximation of the integral using k-nearest-neighbor (k-NN) (Li et al. (2019), Wang, Kulkarni and Verdú (2009)). Let us assume that we have two n-dimensional random samples from two populations with probability density functions f_O and f_S. We define the distance between the ith sample from f_O and: 1) its k-NN in the same population as $\rho_{k'}(i)$ and 2) its k-NN in the other population as $\nu_{k'}(i)$. Then, we use the following approximation (Wang, Kulkarni and Verdú (2009)):

$$(11) \qquad \hat{D}_{\mathrm{KL}}(f_O \parallel f_S) = \frac{1}{n}\sum_{i=1}^{n} \log \frac{\hat{f}_O(X_i)}{\hat{f}_S(X_i)} = \frac{n}{m}\sum_{i=1}^{m} \log \frac{\nu_{k'}(i)}{\rho_{k'}(i)} + \log\frac{m'}{m-1},$$

where $\hat{f}_O^{(k')}(X_i) = \frac{k'}{m-1} \cdot \frac{1}{v_1(n)\rho_{k'}^d(i)}$ and $\hat{f}_S^{(k')}(X_i) = \frac{k'}{m'} \cdot \frac{1}{v_1(d)\nu_{k'}^d(i)}$ are the k-NN estimators of f_O and f_S, respectively, $v_1(n) = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)}$ and $\Gamma(\cdot)$ is the Gamma function.

In our case the dimension is equal to the number of locations, that is, $n = 614$, and the number of samples $m = m' = 4749$ (13 years of data in the testing set, as detailed in Section 2) since we evaluated MENA CORDEX and MERRA-2 for the same time period. One critical aspect of this approximation is the selection of k', that is, the number of nearest neighbors. In this study the standard convention of using the square root of the number of observations is used (Boltz, Debreuve and Barlaud (2007)) so that $k' = \sqrt{4749} \approx 70$.

4.4. *K-means clustering.* To determine the transformation in (7), λ_O and λ_S must be estimated. While it is possible in principle to estimate all $2n = 1222$ location-specific parameters jointly, this would lead to an unnecessary and severe overparametrization, as geographically close locations are expected to have similar estimates. In practice, a joint estimation is also computationally infeasible, since it would require a simultaneous optimization over all $2n$ parameters.

In order to reduce the parameter space and, consequently, the computational time, we assume that λ_O and λ_S are constant across some regions with a k-means clustering approach, that is, for some p-variate observations $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and a fixed k'', we aim to minimize the distance between elements of each cluster and its mean.

While our main interest is to obtain regions with similar λ_O, it would be desirable to have spatially coherent clusters for interpretability. Therefore, we assume that, for each site, we have $p = 3$ covariates: the MLE of the element of λ_O corresponding to that site, the latitude and the longitude. Preliminary exploratory analysis (not shown) indicates that a weighted version of k-means with even a small weight on latitude and longitude allows for spatially coherent clusters. Therefore, we will assume a weight of 0.98 on the MLE and a weight of 0.01 each for the latitude and longitude, and $k'' = 20$ clusters, each with thirty to fifty grids, as shown in Figure 2(b–c). The usage of clusters, instead of location-specific parameters, implies some loss of information. For MENA CORDEX, Figure 2(a) shows the map of the pointwise MLEs (see Figure S11(a) for MERRA-2), whereas Figure 2(b) shows the clusterwise MLEs.
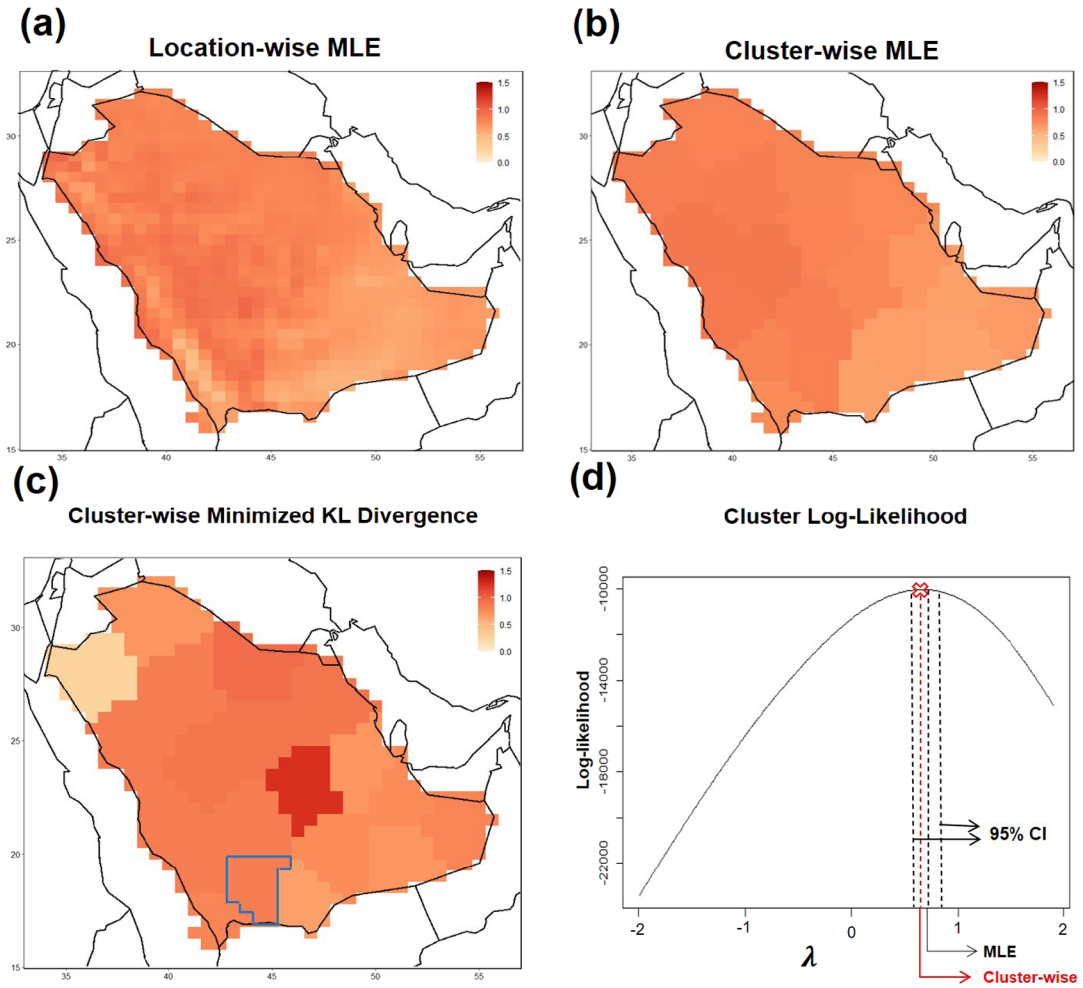
FIG. 2. *MENA CORDEX maps of:* (a) *the Yeo–Johnson MLE,* (b) *the clusterwise MLE and* (c) *the cluster-wise parameter estimated by minimizing the KL divergence.* (d) *The log-likelihood profile of the cluster indicated in* (c), *the red cross represents the cluster-wise parameter and the outer-most dashed lines indicate the* 95% *confidence interval of the MLE.*

Both the values and spatial patterns are visually similar. Figure S11(b–c) show the map of the points whose (asymptotic, likelihood-based) confidence interval for pointwise MLEs include the MLE for the corresponding cluster of both data sets. Despite the very small intervals resulting from $T = 9497$ observations, a sizable portion of the sites (50% MERRA-2 and 37% for MENA CORDEX) includes the cluster MLE.

Even though our objective is to estimate $\boldsymbol{\lambda}_\ell = (\lambda_{\ell;1}, \ldots, \lambda_{\ell;k''})^\top$, $\ell = \{O, S\}$ (i.e., the parameters for each cluster) such that the KL divergence is minimized, the resulting value must not be too different from the MLE to still achieve Gaussianity. Figures 2(b) and 2(c) compare the clusterwise MLE and KL minimizer for MENA CORDEX. Their patterns are very similar, albeit with a few noticeable differences. 12 clusters from among $k'' = 20$ have the KL minimizer inside the MLE confidence interval for both MENA CORDEX and MERRA-2, and Figure 2(d) shows an example of the cluster highlighted in the southern part of the country. In the case of pointwise MLEs, despite the narrow interval from the large number of observed days, the KL minimizers in panel (c) fall within the interval from panel (b).

5. Validation with simulations and real data. In this section we validate our proposed methodology with respect to traditional approaches in terms of the KL divergence

for: 1) two simulated non-Gaussian random fields and 2) the MENA CORDEX simulations and MERRA-2 data. Throughout this section we use M to denote the marginal bias correction with mean (2), MV to denote mean and variance (3), MC and MN to denote the mean and covariance in model (4) via the Matérn parameters in (5) and the nonstationary model (6), respectively. Further, we compare these methods with our proposed approach in (7) for two cases in which the nonstationary covariance (6) is used. In the first case, denoted as T1, a single transformation parameter is considered for each of the observations and simulations so that all locations are transformed by the same parameter, minimizing the KL divergence. In the second case, denoted as TC, we use a clusterwise parameter determined by k-means, as specified in Section 4.4, for each of the two data sets. Thus, the locations in the same cluster are transformed by the same parameter, and the vector of all cluster transformations minimizes the KL divergence across all the sites. Computational details are presented in the Supplementary Material.

5.1. *Simulated data.* We perform a simulation study to assess the model performance in capturing varying degrees of non-Gaussianity. Both data sets are simulated from two random fields that cannot be transformed to Gaussianity with a simple marginal transformation. Furthermore, to simplify the setting we do not assume any interannual or annual trend or temporal dependence (i.e., $\phi_{i',i} = \beta_{k,i} = \beta'_{k,i} = \omega_i = 0$ for all $i = 1, \ldots, n$, $i' = 1, \ldots, P$ and $k = 1, \ldots, K$ in (1)). The spatial domain of our simulation comprises of 200 locations divided into $R = 8$ spatial regions with 25 points each on a 5×5 grid in a square of length 1. For each simulation, 100 replicates were generated, among which the first 50 were considered the historical period and the last 50 were considered the future period. A total of 1000 simulations was performed.

As "observational" data, we simulated samples from a biresolution model with a non-Gaussian marginal distribution (Tagle, Castruccio and Genton (2020), Tagle et al. (2020)). For each region $r = 1, \ldots, R$, we have

$$W_O^{(\ell)}(\mathbf{s}) = \frac{\lambda^{\text{SKT}}|U_r^{\text{SKT}}| + \eta_r^{\text{SKT}}(\mathbf{s})}{\sqrt{Z_r^{\text{SKT}}}}, \tag{12}$$

where $\ell = \{H, F\}$ for historical and future time period, the temporal index has been omitted for simplicity, $Z_r^{\text{SKT}} \sim \text{Gamma}(\nu^{\text{SKT}}/2, \nu^{\text{SKT}}/2)$ independent and identically distributed and $\mathbf{U}^{\text{SKT}} = (U_1^{\text{SKT}}, \ldots, U_R^{\text{SKT}})^\top \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_0^{\text{SKT}})$ and $\eta_r^{\text{SKT}}(\mathbf{s})$ indicate a zero mean Gaussian random field independent across r with covariance matrix $\boldsymbol{\Sigma}_r^{\text{SKT}}$. The model assumes that, for each region r, the two effects Z_r^{SKT} and U_r^{SKT} are constant, and the small scale variation is accounted by the field $\eta_r^{\text{SKT}}(\mathbf{s})$. Across the regions the vector $\mathbf{U}^{\text{SKT}}$ characterizes the large-scale dependence. This model (12) exhibits a skew-t marginal distribution, that is, a perturbation of the t distribution accounting for skewed behavior (Azzalini and Capitanio (2003)), and it can be represented hierarchically, thereby allowing relatively fast frequentist or Bayesian inference. For our simulation we fix the parameters $\lambda^{\text{SKT}} = 0.8$ and $\nu^{\text{SKT}} = 8$ and assume that $\boldsymbol{\Sigma}_r^{\text{SKT}}$ and $\boldsymbol{\Sigma}_0^{\text{SKT}}$ are generated from exponential covariance functions with ranges of 0.2 and 0.5, respectively, corresponding to maximum correlations of 0.77 and 0.24, respectively. The large correlation within each region is in accordance with the previous results with biresolution models, where the majority of the dependence is explained by the small scale (Castruccio, Ombao and Genton (2018), Tagle et al. (2020)). Additional simulations with parameter choices, leading to weaker and stronger correlation, are shown in Figure S13.

The "simulation" data are obtained from a Gaussian-Log-Gaussian model (GLG, Palacios and Steel (2006)), which assumes

$$W_S^{(\ell)}(\mathbf{s}) = \frac{\eta^{\text{GLG}}(\mathbf{s})}{\sqrt{\xi^{\text{GLG}}(\mathbf{s})}} + \varepsilon^{\text{GLG}}(\mathbf{s}), \tag{13}$$
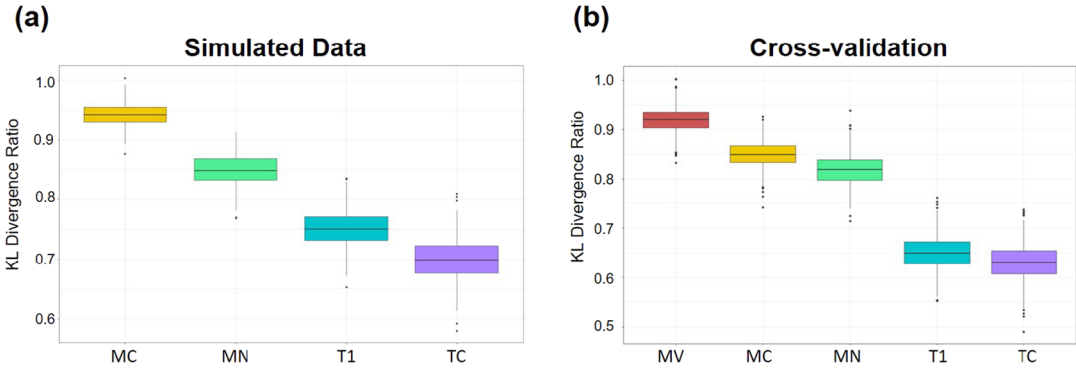
(a)

(b)

FIG. 3.   *Boxplot of KL divergences between* (a) *MV and the MC, MN, T1 and TC method for the simulated data and* (b) *M and the MV, MC, MN, T1 and TC method for the estimated and actual MERRA-2 from* 1993 *to* 2005. *Each of the* 1000 *elements of the boxplot represents an independent simulation in* (a) *and a random subsample of* 300 *locations in* (b).

where $\ell = \{H, F\}$ for historical and future time period and $\eta^{\mathrm{GLG}}(\mathbf{s})$ is an isotropic zero mean Gaussian field with covariance function C^{GLG}, which is considered to be an exponential with a range parameter of 0.2. The process $\xi^{\mathrm{GLG}}(\mathbf{s})$ is such that $\log\{\xi^{\mathrm{GLG}}(\mathbf{s})\}$ is a Gaussian field with mean $-\nu^{\mathrm{GLG}}/2$ and covariance function $\nu^{\mathrm{GLG}}C^{\mathrm{GLG}}$, where C^{GLG} is exponential with a range of 0.7, so that marginally $\xi^{\mathrm{GLG}}(\mathbf{s})$ is lognormal with mean 1 and variance $\exp(\nu^{\mathrm{GLG}}) - 1$ and $\nu^{\mathrm{GLG}} = 8$ in our simulation. The error term $\varepsilon^{\mathrm{GLG}}(\mathbf{s})$ represents an independent Gaussian noise with zero mean and variance $\tau^{2;\mathrm{GLG}} = 0.1$, representing the micro-scale variability or the nugget effect. As for the biresolution model (12), model (13) cannot be transformed to Gaussianity with a simple marginal transformation; thus, our proposed approach is not trained to obtain an exact transformation to Gaussianity for either the "observed" or "simulated" data sets.

Figure 3(a) shows the results in terms of the KL divergence ratio between the MV method and MC, MN, T1 and TC (M is not applied because the "simulated" data have zero mean by construction), where each element of the boxplots represents one of the 1000 simulations. Despite the considerably different and nontrivial non-Gaussian structures of the two data sets, the proposed approach can more closely estimate the true future observations against a Gaussian transformation with an improvement of 29% in the median using a single transformation in T1. The clusterwise transformation TC yields a further improvement to 31%, although at an increased computational cost, because the transformation parameters must be separately estimated for each region.

5.2. *Wind speed data.*   Since future observations cannot be used to validate our model, we only consider the historical data period, and we separated it into a training and testing set. We use the first 13 years of data (1980 to 1992) as the training set and the last 13 years (1993 to 2005) as the test set. Thus, the models are estimated according to the training set, and the predictions in the test period are made by assuming that only MENA CORDEX is available. Then, MERRA-2 is estimated and compared with the original data.

We compute the KL divergence (estimated and true MERRA-2 data in the testing set) ratio between the simple mean bias correction (M) and other approaches. The MV, MC and MN approaches result in ratios of 0.93, 0.88 and 0.81, respectively, indicating an increasingly faithful representation of MERRA-2 if the variance and covariance are estimated. Our proposed approach can further decrease the KL ratio to 0.58 and 0.51 for T1 and TC, respectively.

To assess the uncertainty in the KL ratio, for 1000 times we sample 300 points (from the total $n = 614$) from the spatial domain, perform the correction methods and estimate the

KL ratio for each sample. Each boxplot in Figure 3(b) represents the KL ratio between M and the other approaches for each subsample. For MC, MN and T1 methods, we estimate the covariance function parameters for each sample. In the case of the transformation, we used the same parameters as for the full data set for computational convenience. For TC we used k-means and a stratified sampling method to select 50% of the locations in each cluster. Overall, the improvement in the proposed model is apparent throughout the subsamples, with an increasingly small KL divergence and with our proposed T1 and TC methods yielding the smallest divergence.

6. Application. We use the proposed approach to estimate how the daily revenue from the current optimal turbines build-out in Saudi Arabia will be impacted by a changing climate over the next few decades. In Section 6.1 we describe our approach for extrapolating the predicted surface wind to hub height, whereas in Section 6.2 we assess the final change in profits implied by the corrected data extrapolated at hub height.

6.1. *Extrapolation to hub height.* In Figure 4(a) the daily surface wind speed according to MERRA-2 is presented. This map indicates spatial patterns of potential interest for wind harvesting, especially in the northwest corner, which is a site of particular interest, because of the ongoing project to build a selfsustainable city (NEOM, Farag (2019)). However, the surface wind does not necessarily represent the wind at higher altitudes. Indeed, the wind speed data from both MENA CORDEX and MERRA-2 are computed or observed at a reference height of 10 meters, whereas wind turbines normally operate at a height of 80–120 meters. Therefore, in order to assess the wind energy output the surface wind speed must be extrapolated to the height at which wind turbines operate. There is a vast literature on extrapolating wind speed from surface to a height within the boundary layer; see Emeis (2018) for a comprehensive review. In the vast majority of studies, the power law is used,

$$
W_O^{(F)}(\mathbf{s}_i, t_j, h_k) = W_O^{(F)}(\mathbf{s}_i, t_j, h_r) \left(\frac{h_k}{h_r} \right)^{\alpha_{ij}} e^{\eta(\mathbf{s}_i, t_j)},
$$

(14)

$$
\eta(\mathbf{s}_i, t_j) \sim \mathcal{N}(0, \sigma_i^2),
$$

where h_k is the height to which we want to extrapolate and h_r is the height at which data are available (in our case it is 10 meters). The α is the *shear coefficient* of the power law, and its value is assumed to change, depending on local spatial properties, such as surface roughness and thermal stability (i.e., the temperature gradient for the first layers in the boundary layer) (Gualtieri (2019)). In the absence of any meteorological information, the standard approach is to assume that $\alpha_{ij} = 1/7$ which corresponds to a value observed on flat terrain under neutral atmospheric conditions (Rehman et al. (2007), Tagle et al. (2019)). Direct estimation of α_{ij}, using MERRA-2 or MENA CORDEX, is impossible, as direct estimation of the power law from (14) would require sufficient vertical wind levels below 100 meters, and neither data sets is designed for this level of accuracy near the surface. In this study we rely instead on a high-resolution WRF ensemble and select a run by adopting a planetary boundary layer parameterization, resolution and boundary conditions, resulting in simulations closer to a few in situ data available (Giani et al. (2020)). The WRF simulation is specifically designed to capture the wind at a high resolution near the surface and resolves the wind speed vertical profile at six levels which are approximately equally spaced from the ground level to an altitude of 100 meters. Therefore, these data will be used to estimate α_{ij} and σ_i^2 in (14). Since we focus on daily data (neither MERRA-2 nor MENA CORDEX has hourly data) and a preliminary analysis (not shown) did not highlight temporal changes in the shear coefficient, $\alpha_{ij} = \alpha_i$ is estimated from the power law with a simple log regression.
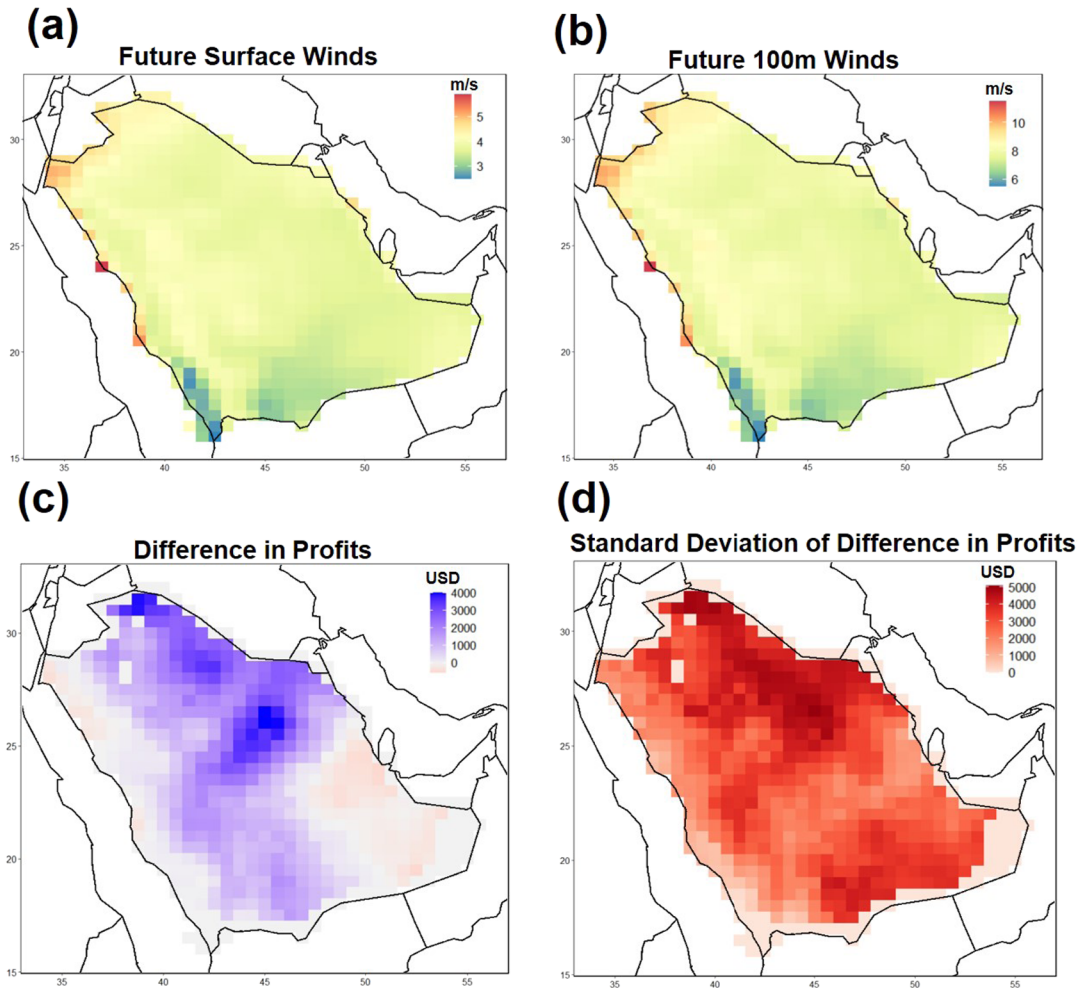
FIG. 4. *Maps of averaged daily wind speed at*: (*a*) *the surface and* (*b*) 100 *meters from* 2025 *to* 2050 *and* (*c*) *average differences between the future and historical wind energy profits and their standard deviations* (*d*).

The map of the estimated wind shear coefficients, along with their standard deviations and the coefficients of determination R^2 can be observed from Figure S14. The map of the estimates indicates a considerable spatial variability with very small and even negative values of the shear coefficient closely associated with the mountainous areas of the Hejaz region in the west. While slower wind at high altitudes is an unexpected (yet physically admissible) behavior within the boundary layer, it often occurs in areas exhibiting the smallest R^2 and the largest standard deviation, indicating that the power law (14) is likely not an appropriate model, as also indicated in many studies (Crippa et al. (2021), Gualtieri (2019)). Since the installation of wind turbines is not cost effective in the rough terrains of Saudi Arabia (Giani et al. (2020)), the model misspecification over these regions is not a major concern. Once the α_i are estimated, MERRA-2 data are then downscaled from the 50×50 km to the 6×6 km WRF resolution with ordinary kriging using another Matérn (5) covariance, and the wind speed at the desired hub height is computed. Finally, given the uncertainty associated with the determination of the shear coefficients, we performed 1000 extrapolations for each site by simulating both the shear coefficient with its variability, as well as the random noise from the parameters estimated in (14).

Figure 4(b) shows the extrapolated average daily wind at 100 meters for MERRA-2 during 2025–2050. Besides being uniformly faster than wind at the surface, the spatial patterns are

not drastically different from those in panel (a), as expected. Figure S15 shows the relative change of 100 meter wind against surface wind and highlights how the increase in wind speed ranges from 60% to about 85%.

6.2. *Assessing changes in future wind energy revenues.* Giani et al. (2020) showed that the most cost effective choice of turbines depends upon the site, because the construction and maintenance cost as well as the potential for harnessing the wind vary. Here, we use the same map of the most efficient turbines as indicated in Figure 5(a) of the aforementioned manuscript and extrapolate the surface wind to the hub height of each turbine type. The wind speed is translated into energy through a turbine-specific power curve which is a function that represents the ability of the turbine to translate the blade movement into energy. Power curves are zero until a given wind speed, increase up to a maximum rated power and are constant for any stronger wind. We make use of of proprietary database purchased from *The Wind Power* (www.thewindpower.net/about_en.php), providing the nominal power curves for the most popular turbines worldwide. The actual power curves are generally not available and require ad hoc methodology for their estimation (Ding (2019)). The total energy output per cell can be calculated by multiplying the power for a single turbine with the total number of turbines, is dependent on the diameter of the rotor blade and has to allow for sufficient spacing among turbines to avoid local turbulence (wake effect) which would reduce the overall efficiency.

Finally, the wind power is translated into actual revenue, using the current tariffs of the Saudi Electricity Company (www.se.com.sa/en-us/customers/Pages/TariffRates.aspx), which is approximately five U.S. cents per kWh. The current number is likely to be a sensible overestimation because the associated direct and indirect costs were not accounted for. The difference in revenue, as implied by the changes between future and present wind, is presented in Figure 4(c). Overall, the map indicates increased revenues of up to approximately $4000 per day for the vast majority of the sites. Most importantly, the sites with increased future revenues are those located near the coasts, hence with more potential for wind harvesting, especially the NEOM region in the north west. The regions toward the center of Saudi Arabia, including the neighborhoods of the capital Riyadh, would incur a loss in revenue from changes in wind if turbines were to be installed there, although it would not be as substantial as in the other areas. The uncertainty map is presented in Figure 4(d) and indicates large uncertainties in some areas, especially near the Persian Gulf. While these uncertainties are arguably large in some areas, they are a direct consequence of the considerable daily variability of the power law (14). The uncertainty in the use of the power law has been acknowledged (Crippa et al. (2021), Gualtieri (2019)), even though it has not been frequently used with the goal of uncertainty quantification but in relation to additional covariates, such as temperature gradients and stability metrics, which cannot be determined in this study given the coarse vertical structures of MERRA-2 and MENA CORDEX.

While the aforementioned maps provide an overall estimation of the difference in profits across the country, the areas of highest interest are the ones where the installation of wind turbines would be the most cost-effective choice. Therefore, we estimate the difference in profits for the 75 wind farms locations identified in Giani et al. (2020) to be the most promising sites. We add the difference in profits across the sites, and we obtain a total increase in profit of approximately $272,000 and a standard deviation of $14,000, hence lending additional support for the long-term profitability of the selected sites. In Figure S16 we have recalculated the daily profits using a simple MV approach, that is, by adjusting only mean and variance in the optimal sites, and we obtained a considerably decreased daily profit of $111,000 (standard deviation $26,000).

7. Conclusion. In this study we have addressed the issue of resilience of Saudi Arabia's current plan to diversify its energy portfolio with wind energy under changing climate conditions. The key element associated with this assessment is the estimation of future winds using reanalysis data. In order to provide that, we proposed a novel trans-Gaussian clusterwise adjustment model based on minimizing the KL divergence between reanalysis and simulated data. Once the model was properly validated with a simulation study and with historical data, we estimated future winds from 2025 to 2050. These estimates were extrapolated to the turbine hub height using the output from high resolution numerical model simulations specifically designed to capture the vertical structure in the boundary layer. Finally, these estimates were translated into wind power and changes in revenue from future to present winds. Results show a sizable increase of approximately $272,000 in daily revenues at the ideal construction sites, albeit with a standard deviation of $14,000 due to the uncertainty propagated by the extrapolation of wind from the surface to hub height.

From a methodological perspective, our model generalized some of the most common approaches used for correcting spatial fields to the non-Gaussian case and proposed an inferential approach that could be scaled to considerably higher spatial resolutions for future data sets with: 1) clustering of the trans-Gaussian transformation and 2) numerical approximation of the KL divergence with k nearest neighbors. The limitations of the current approach will likely become apparent if hourly or subhourly resolution were made available: in that case, the space-time interaction could not be reduced to a simple vector moving average model and would likely require nonseparable models, the application of which would considerably increase the computational cost. Furthermore, high-resolution wind data would: 1) result in sparse wind such that a simple trans-Gaussian model could not capture, and a more involved latent Gaussian model would be required and 2) invalidate extrapolation with the power law, which available literature suggests to use only up to hourly resolution. This would prompt the development of more sophisticated nonparametric approaches, such as neural networks (Vassallo, Krishnamurthy and Fernando (2020)).

From the applied perspective, even though extrapolation is widely acknowledged to be the most important source of uncertainty associated with the determination of wind energy from surface wind, this study did not account for other sources. First, MERRA-2 is only one of the available reanalysis data products, so further comparison with other products, such as the new ERA5 from the European Centre for Medium-Range Weather Forecasts (Hersbach et al. (2020)), could be performed. Second, power curves have been obtained from proprietary data, but the raw data used to determine them are confidential and would likely indicate some degree of uncertainty with respect to the determination of the said curve. Third, the translation of power into revenue is highly dependent upon policies and negotiations between the turbine operators and local authorities; therefore, this may vary across different regions of Saudi Arabia. Future work will focus on better assessing these sources of uncertainty by engaging both industrial partners and policymakers and discussing potential changes in the siting work conducted by Giani et al. (2020) in light of these new results.

Finally, we emphasize how the scope of this study could be generalized to any country with an emerging wind energy portfolio. Even though vertical extrapolation required a high-resolution ensemble focused on Saudi Arabia and energy costs are expected to change across different countries, the core of our methodology involves only publicly available data and hence can be used to inform strategies for other Gulf countries and beyond. Furthermore, in lieu of new data from a high-resolution ensemble, the power law could be simplified to the case of neutral atmospheric stability and flat terrain, a simplifying assumption widely accepted in wind power literature.

SUPPLEMENTARY MATERIAL

Supplementary document (DOI: 10.1214/21-AOAS1460SUPP; .zip). A document providing additional technical details and supplementary details.

REFERENCES

AZZALINI, A. and CAPITANIO, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 367–389. MR1983753 https://doi.org/10.1111/1467-9868.00391

BHATNAGAR, S., CHANG, W., KIM, S. and WANG, J. (2020). Computer model calibration with time series data using deep learning and quantile regression. Available at arXiv:2008.13066.

BOLTZ, S., DEBREUVE, E. and BARLAUD, M. (2007). kNN-based high-dimensional Kullback–Leibler distance for tracking. In *Eighth International Workshop on Image Analysis for Multimedia Interactive Services* (*WIAMIS '07*) 16–16.

BRITISH PETROLEUM (2020). BP statistical review of world energy. Available at www.bp.com/content/dam/bp/en/corporate/pdf/energy-economics/statistical-review/bp-stats-review-2018-full-report.pdf.

CANNON, A. J. (2018). Multivariate quantile mapping bias correction: An N-dimensional probability density function transform for climate model simulations of multiple variables. *Clim. Dyn.* **50** 31–49.

CANNON, A. J., PIANI, C. and SIPPEL, S. (2020). Chapter 5—bias correction of climate model output for impact models. In *Climate Extremes and Their Implications for Impact and Risk Assessment* (J. Sillmann, S. Sippel and S. Russo, eds.) 77–104. https://doi.org/10.1016/B978-0-12-814895-2.00005-7

CANNON, A. J., SOBIE, S. R. and MURDOCK, T. Q. (2015). Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? *J. Climate* **28** 6938–6959. https://doi.org/10.1175/JCLI-D-14-00754.1

CASTRUCCIO, S., OMBAO, H. and GENTON, M. G. (2018). A scalable multi-resolution spatio-temporal model for brain activation and connectivity in fMRI data. *Biometrics* **74** 823–833. MR3860703 https://doi.org/10.1111/biom.12844

CHEN, W., CASTRUCCIO, S. and GENTON, M. G. (2021). Assessing the risk of disruption of wind turbine operations in Saudi Arabia using Bayesian spatial extremes. *Extremes* **24** 267–292. MR4246278 https://doi.org/10.1007/s10687-020-00384-1

CHEN, L., PRYOR, S. C. and LI, D. (2012). Assessing the performance of intergovernmental panel on climate change AR5 climate models in simulating and projecting wind speeds over China. *J. Geophys. Res., Atmos.* **117** D24102. https://doi.org/10.1029/2012JD017533

CHEN, W., CASTRUCCIO, S., GENTON, M. G. and CRIPPA, P. (2018). Current and future estimates of wind energy potential over Saudi Arabia. *J. Geophys. Res., Atmos.* **123** 6443–6459. https://doi.org/10.1029/2017JD028212

CRIPPA, P., ALIFA, M., BOLSTER, D., GENTON, M. G. and CASTRUCCIO, S. (2021). A heteroskedastic time-varying model for improved hourly wind power forecasting. Under review.

DING, Y. (2019). *Data Science for Wind Energy*. CRC press, Boca Raton, FL.

DUNNE, J. P., JOHN, J. G., SHEVLIAKOVA, E., STOUFFER, R. J., KRASTING, J. P., MALYSHEV, S. L., MILLY, P. C. D., SENTMAN, L. T., ADCROFT, A. J. et al. (2013). GFDL's ESM2 global coupled climate–carbon Earth system models. Part II: Carbon system formulation and baseline simulation characteristics. *J. Climate* **26** 2247–2267. https://doi.org/10.1175/JCLI-D-12-00150.1

EMEIS, S. (2018). *Wind Energy Meteorology*, 2nd ed. Springer, New York. https://doi.org/10.1007/978-3-319-72859-9

ERDIN, R., FREI, C. and KÜNSCH, H. R. (2012). Data transformation and uncertainty in geostatistical combination of radar and rain gauges. *J. Hydrometeorol.* **13** 1332–1346. https://doi.org/10.1175/JHM-D-11-096.1

FARAG, A. A. (2019). The story of NEOM city: Opportunities and challenges. In *New Cities and Community Extensions in Egypt and the Middle East* (A. Ibrahim, S. Attia and Z. Shafik, eds.) 35–49. Springer, Cham, Switzerland.

FRANÇOIS, B., VRAC, M., CANNON, A. J., ROBIN, Y. and ALLARD, D. (2020). Multivariate bias corrections of climate simulations: Which benefits for which losses? *Earth Syst. Dyn.* **11** 537–562. https://doi.org/10.5194/esd-11-537-2020

GELARO, R., MCCARTY, W., SUÁREZ, M. J., TODLING, R., MOLOD, A., TAKACS, L., RANDLES, C. A., DARMENOV, A., BOSILOVICH, M. G. et al. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *J. Climate* **30** 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

GIANI, P., FELIPE, T., GENTON, M. G., CASTRUCCIO, S. and CRIPPA, P. (2020). Closing the gap between wind energy targets and implementation for emerging countries. *Appl. Energy* **269** 115085. https://doi.org/10.1016/j.apenergy.2020.115085

GUALTIERI, G. (2019). A comprehensive review on wind resource extrapolation models applied in wind energy. *Renew. Sustain. Energy Rev.* **102** 215–233. https://doi.org/10.1016/j.rser.2018.12.015

HAWKINS, E., OSBORNE, T. M., HO, C. K. and CHALLINOR, A. J. (2013). Calibration and bias correction of climate projections for crop modelling: An idealised case study over Europe. *Agric. For. Meteorol.* **170** 19–31. Agricultural prediction using climate model ensembles. https://doi.org/10.1016/j.agrformet.2012.04.007

HEMER, M., MCINNES, K. and RANASINGHE, R. (2012). Climate and variability bias adjustment of climate model-derived winds for a southeast Australian dynamical wave model. *Ocean Dyn.* **62** 87–104.

HERSBACH, H., BELL, B., BERRISFORD, P., HIRAHARA, S., HORÁNYI, A., MUÑOZ-SABATER, J., NICO-LAS, J., PEUBEY, C., RADU, R. et al. (2020). The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146** 1999–2049. https://doi.org/10.1002/qj.3803

HIGDON, D. (2002). Space and space–time modeling using process convolutions. In *Quantitative Methods for Current Environmental Issues* 37–56. Springer, London. MR2059819

HO, C. K., STEPHENSON, D. B., COLLINS, M., FERRO, C. A. T. and BROWN, S. J. (2012). Calibration strategies: A source of additional uncertainty in climate change projections. *Bull. Am. Meteorol. Soc.* **93** 21–26. https://doi.org/10.1175/2011BAMS3110.1

INTERNATIONAL RENEWABLE ENERGY AGENCY (2018). Renewable energy statistics. Available at https://irena.org/publications/2018/Jul/Renewable-Energy-Statistics-2018.

INTERNATIONAL RENEWABLE ENERGY AGENCY (2019). Renewable energy market analysis: GCC. Available at https://www.irena.org/publications/2019.

IPCC (2014). Part A: Global and sectoral aspects. In *AR5 Climate Change* 2014: *Impacts, Adaptation, and Vulnerability* Cambridge Univ. Press.

JEONG, J., CASTRUCCIO, S., CRIPPA, P. and GENTON, M. G. (2018). Reducing storage of global wind ensembles with stochastic generators. *Ann. Appl. Stat.* **12** 490–509. MR3773402 https://doi.org/10.1214/17-AOAS1105

JEONG, J., YAN, Y., CASTRUCCIO, S. and GENTON, M. G. (2019). A stochastic generator of global monthly wind energy with Tukey *g*-and-*h* autoregressive processes. *Statist. Sinica* **29** 1105–1126. MR3932511

KENNEDY, M. C. and O'HAGAN, A. (2000). Predicting the output from a complex computer code when fast approximations are available. *Biometrika* **87** 1–13. MR1766824 https://doi.org/10.1093/biomet/87.1.1

KENNEDY, M. C. and O'HAGAN, A. (2001). Bayesian calibration of computer models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** 425–464. MR1858398 https://doi.org/10.1111/1467-9868.00294

KIM, K. B., KWON, H.-H. and HAN, D. (2015). Bias correction methods for regional climate model simulations considering the distributional parametric uncertainty underlying the observations. *J. Hydrol.* **530** 568–579. https://doi.org/10.1016/j.jhydrol.2015.10.015

KINLEY, R. (2017). Climate change after Paris: From turning point to transformation. *Climate Policy* **17** 9–15. https://doi.org/10.1080/14693062.2016.1191009

KULLBACK, S. and LEIBLER, R. A. (1951). On information and sufficiency. *Ann. Math. Stat.* **22** 79–86. MR0039968 https://doi.org/10.1214/aoms/1177729694

LEEDS, W. B., MOYER, E. J. and STEIN, M. L. (2015). Simulation of future climate under changing temporal covariance structures. *Adv. Stat. Climatol. Meteorol. Oceanogr.* **1** 1–14. https://doi.org/10.5194/ascmo-1-1-2015

LI, D., FENG, J., XU, Z., YIN, B., SHI, H. and QI, J. (2019). Statistical bias correction for simulated wind speeds over CORDEX-East Asia. *Earth Space Sci.* **6** 200–211. https://doi.org/10.1029/2018EA000493

MEHROTRA, R. and SHARMA, A. (2016). A multivariate quantile-matching bias correction approach with auto- and cross-dependence across multiple time scales: Implications for downscaling. *J. Climate* **29** 3519–3539.

NGUYEN, H., MEHROTRA, R. and SHARMA, A. (2019). Correcting systematic biases across multiple atmospheric variables in the frequency domain. *Clim. Dyn.* **52** 1283–1298.

NREP (2018). Saudi Arabia renewable energy targets and long term visibility, national renewable energy program.

NURUNNABI, M. (2017). Transformation from an oil-based economy to a knowledge-based economy in Saudi Arabia: The direction of saudi vision 2030. *Journal of the Knowledge Economy* **8** 536–64.

PACIOREK, C. J. and SCHERVISH, M. J. (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics* **17** 483–506. MR2240939 https://doi.org/10.1002/env.785

PALACIOS, M. B. and STEEL, M. F. J. (2006). Non-Gaussian Bayesian geostatistical modeling. *J. Amer. Statist. Assoc.* **101** 604–618. MR2281244 https://doi.org/10.1198/016214505000001195

POPPICK, A., MCINERNEY, D. J., MOYER, E. J. and STEIN, M. L. (2016). Temperatures in transient climates: Improved methods for simulations with evolving temporal covariances. *Ann. Appl. Stat.* **10** 477–505. MR3480504 https://doi.org/10.1214/16-AOAS903

REHMAN, S., EL-AMIN, I. M., AHMAD, F., SHAAHID, S. M., AL-SHEHRI, A. M. and BAKHASHWAIN, J. M. (2007). Wind power resource assessment for Rafha, Saudi Arabia. *Renew. Sustain. Energy Rev.* **11** 937–950. https://doi.org/10.1016/j.rser.2005.07.003

REN21 SECRETARIAT (2018). Renewables 2018—global status report. Paris, France.

RISSER, M. and CALDER, C. (2017). Local likelihood estimation for covariance functions with spatially-varying parameters: The convoSPAT package for R. *J. Stat. Softw.* **81** 1–32. https://doi.org/10.18637/jss.v081.i14

STEIN, M. L. (1999). *Interpolation of Spatial Data*: *Some Theory for Kriging. Springer Series in Statistics*. Springer, New York. MR1697409 https://doi.org/10.1007/978-1-4612-1494-6

TAGLE, F., CASTRUCCIO, S. and GENTON, M. G. (2020). A hierarchical bi-resolution spatial skew-*t* model. *Spat. Stat.* **35** 100398, 12. MR4052623 https://doi.org/10.1016/j.spasta.2019.100398

TAGLE, F., CASTRUCCIO, S., CRIPPA, P. and GENTON, M. G. (2019). A non-Gaussian spatio-temporal model for daily wind speeds based on a multi-variate skew-*t* distribution. *J. Time Series Anal.* **40** 312–326. MR3946155 https://doi.org/10.1111/jtsa.12437

TAGLE, F., GENTON, M. G., YIP, A., MOSTAMANDI, S., STENCHIKOV, G. and CASTRUCCIO, S. (2020). A high-resolution bilevel skew-t stochastic generator for assessing Saudi Arabia's wind energy resources. *Environmetrics* **31** e2628, 16. MR4166851 https://doi.org/10.1002/env.2628

TAYLOR, K. E., STOUFFER, R. J. and MEEHL, G. A. (2012). An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93** 485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

TEUTSCHBEIN, C. and SEIBERT, J. (2012). Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *J. Hydrol.* **456–457** 12–29. https://doi.org/10.1016/j.jhydrol.2012.05.052

TUO, R. and WU, C. F. J. (2015). Efficient calibration for imperfect computer models. *Ann. Statist.* **43** 2331–2352. MR3405596 https://doi.org/10.1214/15-AOS1314

UNFCCC (2020). Intended nationally determined contributions: Kingdom of Saudi Arabia. Available at www4.unfccc.int/sites/submissions/INDC/Published%20Documents/Saudi%20Arabia/1/KSA-INDCs%20English.pdf.

VAN VUUREN, D. P., EDMONDS, J., KAINUMA, M., RIAHI, K., THOMSON, A., HIBBARD, K., HURTT, G. C., KRAM, T., KREY, V. et al. (2011). The representative concentration pathways: An overview. *Clim. Change* **109** 5–31.

VASSALLO, D., KRISHNAMURTHY, R. and FERNANDO, H. J. S. (2020). Decreasing wind speed extrapolation error via domain-specific feature extraction and selection. *Wind Energy Science* **5** 959–975. https://doi.org/10.5194/wes-5-959-2020

WANG, Q., KULKARNI, S. R. and VERDÚ, S. (2009). Divergence estimation for multidimensional densities via *k*-nearest-neighbor distances. *IEEE Trans. Inf. Theory* **55** 2392–2405. MR2729888 https://doi.org/10.1109/TIT.2009.2016060

WORLD BANK (2020). Energy use (kg of oil equivalent per capita). Available at https://data.worldbank.org/indicator/EG.USE.PCAP.KG.OE.

YEO, I.-K. and JOHNSON, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika* **87** 954–959. MR1813988 https://doi.org/10.1093/biomet/87.4.954

YUAN, Q., THORARINSDOTTIR, T. L., BELDRING, S., WONG, W. K., HUANG, S. and XU, C.-Y. (2019). New approach for bias correction and stochastic downscaling of future projections for daily mean temperatures to a high-resolution grid. *J. Appl. Meteorol. Climatol.* **58** 2617–2632. https://doi.org/10.1175/JAMC-D-19-0086.1

ZHANG, J., CRIPPA, P., GENTON, M. G. and CASTRUCCIO, S. (2021). Supplement to "Assessing the reliability of wind power operations under a changing climate with a non-Gaussian bias correction." https://doi.org/10.1214/21-AOAS1460SUPP