



# Assessing the risk of disruption of wind turbine operations in Saudi Arabia using Bayesian spatial extremes

Wanfang Chen<sup>1</sup> · Stefano Castruccio<sup>2</sup> · Marc G. Genton<sup>1</sup>

Received: 15 June 2019 / Revised: 23 June 2020 / Accepted: 24 June 2020 /

Published online: 22 July 2020

© Springer Science+Business Media, LLC, part of Springer Nature 2020

## Abstract

Saudi Arabia has been seeking to reduce its dependence on oil by diversifying its energy portfolio, including the largely underused energy potential from wind. However, extreme winds can possibly disrupt the wind turbine operations, thus preventing the stable and continuous production of wind energy. In this study, we assess the risk of disruptions of wind turbine operations, based on return levels with a hierarchical spatial extreme modeling approach for wind speeds in Saudi Arabia. Using a unique Weather Research and Forecasting dataset, we provide the first high-resolution risk assessment of wind extremes under spatial non-stationarity over the country. We account for the spatial dependence with a multivariate intrinsic autoregressive prior at the latent Gaussian process level. The computational efficiency is greatly improved by parallel computing on subregions from spatial clustering, and the maps are smoothed by fitting the model to cluster neighbors. Under the Bayesian hierarchical framework, we measure the uncertainty of return levels from the posterior Markov chain Monte Carlo samples, and produce probability maps of return levels exceeding the cut-out wind speed of wind turbines within their lifetime. The probability maps show that locations in the South of Saudi Arabia and near the Red Sea and the Persian Gulf are at very high risk of disruption of wind turbine operations.

**Keywords** Return levels · Spatial extremes · Spatial hierarchical modeling · Wind energy · Wind extremes · Wind turbines

**AMS 2000 Subject Classifications** 62M30 · 62M40

---

✉ Wanfang Chen  
wanfang.chen@kaust.edu.sa

Extended author information available on the last page of the article.

## 1 Introduction

Facing the rising demand in energy resulting from the industrial development, urbanization and growth of its population, Saudi Arabia is seeking to change the heavy reliance on fossil fuels for its energy needs by investing in its largely underused potential of renewable energy. According to KA-CARE (2012), Saudi Arabia is planning to build a capacity of 54 GW of renewable energy portfolio by 2032, of which 9 GW are expected to come from wind power. The NEOM project (NEOM Project 2017), initiated from the Saudi Vision 2030 (Vision 2030 2016), also aims at pursuing a very large, self-sustainable city with a substantial reliance on wind energy.

Wind has an intermittent nature, and hence implies a high level of uncertainty in assessing its intensity. Wind extremes are of particular interest as they can be destructive to the integrity of wind turbines, or force their shutdowns, thus bringing a threat to the stability of the production of wind power. Shutdowns on a large number of wind turbines, even on a small number of wind farms, can result in a deep and fast drop in wind power output, which can have a large impact on the local electrical grid. Therefore, an accurate assessment of wind extremes is crucial for risk management purposes. Recently, several studies have been carried out to evaluate the wind power potential over Saudi Arabia (Rehman and Ahmad 2004; Rehman et al. 2007; Shaahid et al. 2014; Yip et al. 2016; Chen et al. 2018; Giani et al. 2020), yet little work has focused on the risk assessment of extreme wind speeds. In this work, we aim to provide the first high-resolution risk assessment of disruption of wind turbine operations in Saudi Arabia by modeling the spatial extremes of wind speeds. To do so, we use a newly developed, unique dataset generated by the Weather Research and Forecasting (WRF) model with  $5\text{ km} \times 5\text{ km}$  spatial resolution over Saudi Arabia, produced by Yip (2018).

Historically, the generalized extreme-value (GEV) distribution and generalized Pareto distribution (GPD) in Extreme-Value Theory (EVT) have been used in the statistical modeling of extremes at individual locations. Since the univariate extreme value analysis typically only uses the very limited observations in the tail, it is useful to pool data from all locations to reduce the uncertainty of parameter estimates. Furthermore, since extreme values are usually spatially correlated, it is important to borrow information from neighboring locations. Recently, researchers have been increasingly focusing on the modeling of spatial extremes, i.e., modeling the joint tail behaviors of extremes in space by characterizing the spatial dependence of extremes between locations. The classical methodologies for modeling spatial extremes can be divided into three main classes: copula, max-stable and latent process models. The copula and max-stable models directly apply a spatial model to the extreme values to capture the spatial dependence, whereas the latent process models assume a latent spatial process embedded into the parameters of the marginal distribution at each location; see Davison et al. (2012) for a review and comparison of these three approaches. The so-called near-independence models (e.g., Ledford and Tawn 1996; Ledford and Tawn 1997; Heffernan and Tawn 2004; Ramos and Ledford 2009) for

spatial extremal data have also been developed to capture the decreasing dependence for increasing rare events that classical models are not able to capture. For example, the conditional extremes model of Heffernan and Tawn (2004) has been used in several studies on spatial extremes (e.g., Winter et al. 2016; Towe et al. 2019). An alternative method for modeling spatial extremes by borrowing strength across locations is regional frequency analysis (RFA) by Hosking and Wallis (1997) as applied in Weiss et al. (2014).

Many studies evaluating the wind extremes over a spatial domain have applied the extreme-value theory to data from climate model outputs, but assumed that the extreme values across locations are mutually independent (e.g., Kharin and Zwiers 2000; Nikulin et al. 2011; de Winter et al. 2013; Kumar et al. 2015). Some studies have modeled the spatial dependence in the extremes based on max-stable processes (e.g., Huser and Davison 2014) or copulas (e.g., Sang and Gelfand 2010), but the implementation of these models in large dimensions, such as for the spatial domain of a WRF model, can be very difficult. The latent process approach is most naturally performed in a Bayesian setting due to the hierarchical modeling procedure involved, where Markov chain Monte Carlo (MCMC) algorithms can be used to quantify the uncertainty (e.g., Banerjee et al. 2014). Davison et al. (2012) concluded that Bayesian hierarchical models are well suited for computing marginal properties such as return levels, which are often used to analyze the extreme behaviors of climate variables. Therefore, the latent process approach is retained in the present paper for assessing the wind speed return levels based on the large WRF dataset. We do not adopt the RFA for our application since it does not provide a satisfactory method of uncertainty quantification as stated in Cooley et al. (2007), who presented the first use of Bayesian hierarchical models to produce a map characterizing precipitation extremes across a geographic region with data from 56 stations. Later on, Cooley and Sain (2010) developed a spatial hierarchical model for extreme precipitation data from a Regional Climate Model (RCM) at 2,464 locations. Other similar applications, either using station data or gridded data from climate models, include Gaetan and Grigoletto (2007), Sang and Gelfand (2009), Turkman et al. (2010) and Jalbert et al. (2017).

To our knowledge, Cooley and Sain (2010) performed the first study on spatial hierarchical modeling of extremes for gridded data from an RCM. The spatial non-stationarity was captured by modeling the spatially varying parameters of the marginal extreme value distribution with latent Gaussian processes. Geophysically-based covariates, such as the longitude, latitude and elevation, were incorporated into the extremes modeling. The hierarchical procedure was implemented in a Bayesian framework, and inference was based on the MCMC posterior samples.

One limitation of the model presented by Cooley and Sain (2010) is that it can only be applied to small datasets, and is therefore not computationally feasible to our much larger dataset over the climatologically more diverse spatial domain. In this study, we apply a novel hierarchical spatial extremes model to wind speed data in the new high-resolution WRF dataset from Yip (2018). Our main contributions can be summarized as follows:

- 1) we provide smooth estimates of the spatial parameters by performing inference with neighboring spatial clusters;

- 2) we provide computationally affordable inference by means of appropriately-divided subregions and parallelization, instead of focusing on a full spatial analysis;
- 3) we provide the first high-resolution risk assessment of wind extremes over Saudi Arabia, with uncertainty quantified in the Bayesian framework;
- 4) we provide evidence for the inadequacy of using a proper multivariate conditional autoregressive (CAR) model to capture the spatial random effects in high spatial-resolution datasets, and we demonstrate that the improper multivariate intrinsic autoregressive (IAR) specification in the Bayesian framework is more appropriate for capturing the irregular and strong spatial dependence.

The remainder of this paper is organized as follows: In Section 2, we describe the new WRF dataset, as well as the preprocessing required to apply the spatial extremes model. In Section 3, we introduce the methodology used in this work, including the univariate extreme value analysis, the spatial clustering method and the spatial extremes model. Section 4 provides the computational details. In Section 5, we summarize the main results for parameter and return level estimations, and analyze the risk of disruption of wind turbine operations in Saudi Arabia. We then draw some conclusions and discuss future research directions in Section 6.

## 2 Data

Wind turbines operate at hub heights ranging from 50 to 140 meters and are highly sensitive to the local surface topography. Hence, assessing the near-surface wind resources and managing the risk of wind extremes require high-resolution data with detailed local surface information. For regions such as Saudi Arabia, with sparse and difficult access to data from monitoring stations, spatio-temporal gridded data simulated from General Circulation Model (GCM) experiments and Regional Climate Model (RCM) experiments are usually used to infer global or regional wind climatology. GCMs can only capture large-scale features of winds, due to the coarse spatial resolution (typically hundreds of kilometers) applied. RCMs are able to reproduce the localized wind variabilities, with spatial resolutions of tens of kilometers.

In this work, we use a newly obtained dataset with high spatial resolution over Saudi Arabia, produced by Yip (2018) in the Atmospheric and Climate Modeling group at KAUST. This new dataset was produced using the Weather Research and Forecasting (WRF) model (specifically, WRF-ARW version 3.9.1), a state-of-the-art mesoscale Numerical Weather Prediction (NWP) model developed at the National Center for Atmospheric Research (NCAR) Mesoscale and Microscale Meteorology Laboratory. The initial and boundary conditions are from the Integrated Forecasting System (IFS) analysis dataset at the European Center for Medium-range Weather Forecasting (ECMWF). The dataset is validated for various aspects of wind characteristics (the mean and variability of wind speed and wind power generation, mean episode length, etc.), using the Modern Era Retrospective-Analysis for Research and Applications (MERRA) reanalysis dataset as the reference. The original WRF data are provided on a spatial resolution of approximately  $5 \text{ km} \times 5 \text{ km}$  with  $769 \times 659 = 506,771$  locations covering the Arabian Peninsula, over a period ranging

from 2009 to 2014, with hourly temporal resolution. The temporal dimension is limited to six years, due to the high computational demands of performing a numerical simulation at such a fine spatial resolution. In this study, we focus on Saudi Arabia, with 83,981 gridded spatial locations, bounded approximately by longitudes  $34 - 56^{\circ}\text{E}$  and latitudes  $15.5 - 33^{\circ}\text{N}$ . In order to avoid modeling the complex non-stationarity over time due to seasonality, we restrict our attention to the three Summer months, June, July and August (JJA), characterized by the strongest wind regimes of the year (Yip et al. 2016; Tagle et al. 2019). Finally, we extrapolate the wind speed to the hub height of 140 meters, a reference height for the latest generation of turbines (Yip et al. 2016). For this, we adopt the commonly used power law method (Emeis 2005), which assumes that wind speed at a certain height  $z$  is approximated by:

$$w(z) = w(z_r) \left( \frac{z}{z_r} \right)^{\alpha},$$

where  $z_r$  is the reference height,  $w(z_r)$  is the wind speed at  $z_r$  and  $\alpha$  is the power law exponent. We assume that  $\alpha = 1/7$ , a condition implying a stable boundary layer, an appropriate condition over open land surfaces (Pryor and Barthelmie 2011) that has been used in previous studies on wind in Saudi Arabia (Rehman et al. 2007; Tagle et al. 2019).

### 3 Methodology

In this section, we illustrate the statistical methodology used in this work. First, we introduce the approaches for univariate extreme value analysis. One of these (the point process approach) is used as the first level of our hierarchical spatial extremes model. We then illustrate how the domain is divided into smaller regions with clustering. Lastly, we demonstrate how we can perform inference with our hierarchical model.

#### 3.1 Univariate extreme value analysis

The EVT (Fisher and Tippett 1928; Resnick 1987; Coles 2001; de Haan and Ferreira 2006) has played an important role in univariate extreme value analysis. For statistical modeling of extremes based on EVT, one approach relies on the limiting distribution families for sample block-maxima that satisfy the property of max-stability. Suppose  $Z_1, Z_2, \dots$  is a sequence of i.i.d. continuous scalar random variables and let  $M_n = \max\{Z_1, \dots, Z_n\}$ . The Extremal Types Theorem (Fisher and Tippett 1928; Gnedenko 1943) states that if there exist sequences of constants  $a_n > 0$  and  $b_n$  such that, as  $n \rightarrow \infty$ ,  $\mathbb{P}\{(M_n - b_n)/a_n \leq z\} \rightarrow G(z)$  for some non-degenerate distribution  $G$ , then  $G$  can only have the following GEV distributional form:

$$G(z) = \begin{cases} \exp \left\{ - \left( 1 + \xi \frac{z-\mu}{\sigma} \right)_+^{-1/\xi} \right\}, & \xi \neq 0, \\ \exp \left\{ - \exp \left( - \frac{z-\mu}{\sigma} \right) \right\}, & \xi = 0, \end{cases}$$

where  $a_+ = \max(0, a)$ ,  $\mu$  is the location parameter,  $\sigma > 0$  is the scale parameter, and  $\xi$  is the shape parameter determining the rate of tail decay, with  $\xi > 0$  giving the heavy-tailed (Fréchet) case,  $\xi = 0$  giving the light-tailed (Gumbel) case, and  $\xi < 0$  giving the short-tailed (reversed Weibull) case.

Another approach for modeling extremes is to investigate the exceedances over a given threshold. Suppose that the block maxima  $M_n = \max\{Z_1, \dots, Z_n\}$  are approximately  $\text{GEV}(\mu, \sigma, \xi)$ . Denote an arbitrary term in  $Z_1, Z_2, \dots$  as  $Z$ . We can then approximate the distribution of  $Z$  exceeding a well-chosen high threshold  $u$  by the Generalized Pareto Distribution  $\text{GPD}(\tau, \xi)$  (Pickands 1975; Davison and Smith 1990):

$$\mathbb{P}(Z - u < z | Z > u) = \begin{cases} 1 - (1 + \xi \frac{z}{\tau})_+^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp(-\frac{z}{\tau}), & \xi = 0, \end{cases}$$

where  $\xi$  and  $\tau > 0$  are shape and scale parameters, respectively. Hence, the distribution of  $Z$  is

$$\mathbb{P}(Z \leq z) = \begin{cases} 1 - \zeta_u (1 + \xi \frac{z-u}{\tau})_+^{-1/\xi}, & \xi \neq 0, \\ 1 - \zeta_u \exp(-\frac{z-u}{\tau}), & \xi = 0, \end{cases} \quad z > u,$$

where  $\zeta_u = \mathbb{P}(Z > u)$  is the probability of exceedance.

An alternative to the GPD approach for modeling threshold exceedances is the Point Process (PP) approach (Smith 1989). Suppose that we observe  $n_y$  (where  $y$  is shorthand for years) block maxima, which are approximately  $\text{GEV}(\mu, \sigma, \xi)$ . Then the point process  $\{(i, Z_i); i = 1, \dots, n\}$  is well approximated above some high threshold  $u$  by a Poisson point process with mean measure

$$\Lambda([t_1, t_2] \times (z, \infty)) = n_y(t_2 - t_1) \{1 + \xi(z - \mu)/\sigma\}_+^{-1/\xi}, \quad z > u, \quad 0 \leq t_1 < t_2 \leq 1.$$

Let  $u$  be a high threshold ensuring that the points in  $[0, 1] \times (u, \infty)$  are well approximated by a Poisson point process. The Poisson process likelihood is

$$L(\mu, \sigma, \xi | z_1, \dots, z_n) \propto \exp \left\{ -n_y (1 + \xi \frac{u-\mu}{\sigma})_+^{-1/\xi} \right\} \prod_{i=1}^{N_u} \left\{ \frac{1}{\sigma} \left( 1 + \xi \frac{z_{(i)} - \mu}{\sigma} \right)_+^{-1/\xi-1} \right\}, \quad (1)$$

where  $z_{(1)}, \dots, z_{(N_u)}$  are the  $N_u$  observations exceeding the threshold  $u$ .

### 3.2 Spatial clustering

The main difficulty of applying a spatial extremes model to our large spatial domain is the computational infeasibility. We choose to divide the study region into small clusters and fit the model to the subregions instead of the whole Saudi Arabia. This allows us to capture the locally stationary features of our model while also reducing the computational burden with parallel computing.

In this study, we use the simple and fast  $k$ -means clustering method that partitions  $N$  observations of dimension  $p$  into  $k$  clusters in which each observation

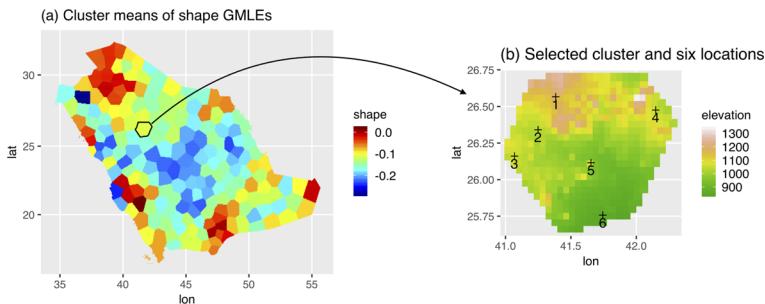
belongs to the cluster with the nearest mean, serving as a prototype of the cluster. Specifically, given a set of observations  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ , where each observation is a  $p$ -dimensional real vector,  $k$ -means clustering aims to partition the  $N$  observations into  $k$  ( $k \leq N$ ) sets,  $S = \{S_1, S_2, \dots, S_k\}$ , in order to minimize the within-cluster sum of squares; the objective is to find

$$\arg \min_S \sum_{i=1}^k \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2,$$

where  $\boldsymbol{\mu}_i$  is the mean of the points in  $S_i$ .

Due to the large and climatologically diverse spatial domain of our interest, a model with a constant shape parameter  $\xi$  would not be adequate, neither would a simple formula for characterizing the variability of  $\xi$ , such as a linear regression on the latitude and longitude, be adequate. Therefore, we spatially model  $\xi$ , in addition to the location parameter  $\mu$  and the scale parameter  $\sigma$ . In order to characterize the spatial variability of  $\xi$ , we use this property as our main factor for clustering. Specifically, we use the marginal estimates for  $\xi$ , such as the maximum likelihood estimate (MLE), obtained independently in each grid cell. However, empirical studies have shown that absurd values for  $\xi$  can be generated by MLEs, especially in small samples. In order to restrict  $\xi$  to a statistically and physically reasonable range, Martins and Stedinger (2000) suggested to use the generalized MLE (GMLE), obtained by maximizing the generalized log likelihood with a prior for  $\xi$  added to the log likelihood. We follow Martins and Stedinger (2000) by using a beta density prior, but instead of using the beta parameters (6, 9), as they did, we use the parameters (6, 4) which are better adapted to the wind-speed application.

In the  $k$ -means clustering, we impose contiguous spatial clusters for better interpretability, hence longitude and latitude are also included for clustering. We then choose the value for  $k$  and assign weights for the three clustering factors in order to yield appropriate contiguous clusters that are 1) big enough for the identifiability of model parameters, 2) small enough for computational feasibility for fitting the spatial extremes model in each of the  $k$  cluster neighbors (whose definition will be introduced later), and 3) similar in size for easier implementation of parallel computing. Extensive sensitivity analysis on the spatial clustering reveals that the spatial patterns of clusters are quite similar for different calibrated values of  $k$  and different assigned weights that put more weight on  $\xi$ ; see Fig. 8 in Appendix A. Table 2 shows the range of cluster sizes for each of the selected settings. We find that a configuration of  $k = 200$  with assigned weights (0.2, 0.2, 0.6) (for longitude, latitude and GMLEs of  $\xi$ , respectively) achieves a reasonable compromise among the above noted considerations. The choice of  $k = 250$  with the same weights is a potential alternative, which will be used for further sensitivity analysis on the end results from our Bayesian hierarchical model in Section 5. Figure 1a depicts the mean of marginal GMLEs for  $\xi$  within each cluster. We can see a clear non-stationarity and a continuously varying pattern of the shape parameter over the spatial clusters. Figure 1b provides a detailed representation of one randomly selected spatial cluster with six locations in that cluster, which will be used in Section 4.



**Fig. 1** **a** Cluster means for marginal GMLEs of the shape parameter  $\xi$  over Saudi Arabia. **b** Zoom-in elevation (in meters) map for the randomly selected cluster and six locations labeled as “+” and numbered as 1, 2, . . . , 6

Fitting the model to each of the  $k = 200$  spatial clusters independently and combining the results would produce discontinuities on the boundaries between clusters. Instead, we fit the model independently to cluster neighbors (i.e., the locations in each cluster as well as its neighboring clusters), and retain only the results for that single cluster. We illustrate how this algorithm works in one randomly selected cluster neighbor shown in Fig. 9 in Appendix A. In order to obtain the estimates for locations in the central cluster in Fig. 9, instead of fitting our model in that single cluster only, we fit our model in its cluster neighbor (i.e., all the seven clusters shown), but only the estimates in the central cluster are retained. By doing so, we can see in Section 5 that the boundary effect can be eliminated as there is no visible discontinuity on the boundaries of clusters, and the maps can be greatly smoothed with information borrowed from the neighboring clusters.

### 3.3 The hierarchical spatial extremes model

In this section, we introduce the Bayesian hierarchical extremes model to be applied to each of the subregions. This model consists of three levels: data, process and prior.

#### – Data level

The threshold exceedances at each location are assumed to follow the PP model with the modified Martins and Stedinger (2000) prior, as described in Section 3.2, and the observations are conditionally independent, given the marginal PP parameters. Hence, the generalized likelihood function linked to Eq. 1 is given by

$$\prod_{i=1}^N \left[ \exp \left\{ -n_y \left( 1 + \xi_i \frac{u_i - \mu_i}{\sigma_i} \right)_+^{-1/\xi_i} \right\} \prod_{k=1}^{N_{u_i}} \left\{ \frac{1}{\sigma_i} \left( 1 + \xi_i \frac{z_{i,k} - \mu_i}{\sigma_i} \right)_+^{-1-1/\xi_i} \right\} \times \frac{\Gamma(10)}{\Gamma(6)\Gamma(4)} (0.5 + \xi_i)^5 (0.5 - \xi_i)^3 \right],$$

where  $N$  is the number of grid cells in the spatial domain of interest (i.e., each of the aforementioned cluster neighbors),  $n_y$  is the number of years of observations at each location, and  $z_{i,k}, k = 1, \dots, N_{u_i}$ , are the observations exceeding the threshold  $u_i$  in grid cell  $i, i = 1, \dots, N$ . In this study, the range for  $N$  (the number of grid cells in each of the 200 cluster neighbors) is [963, 4063], and  $n_y = 6$  (the number of years of wind speed data at each location) for all the cluster neighbors. To set the thresholds  $u_i, i = 1, \dots, N$ , one fast way is by using a rule of thumb to choose the  $r$  largest observations, where the 90% quantile is commonly used (DuMouchel 1983; Scarrott and MacDonald 2012). In this work, since the amount of data is large (hourly wind speeds in Summer with 13,248 time points at each location), we choose a higher quantile, i.e., the 95% quantile as the threshold. The QQ-plots for marginal GPD fitting at 9 randomly selected locations (see Fig. 10 in Appendix A) show that the 95% quantile is sufficiently high for the GPD to adequately model the exceedances. Graphical approaches for choosing a threshold, such as the mean residual life plot and parameter stability plot, are far more time-consuming than using a rule of thumb, especially for our large data set. We further de-cluster the threshold exceedances and only the cluster maxima are used with a common length of the cluster interval of 24 hours (Smith 1985). The resulting time series (i.e., the cluster maxima), at each location, exhibit lengths ranging from 39 to 162 (i.e.,  $N_{u_i} \in [39, 162], i = 1, \dots, N$ ), and are found to be stationary, based on simple diagnostics such as the Auto-Correlation Function (ACF) plots, the Augmented Dickey-Fuller test and Ljung-Box test (results not reported here). To provide quantitative justification for using 95% quantile rather than 90% quantile as the threshold, we compare the values of threshold and goodness of fit for the GPD model used for marginal extremes based on the KL divergence at each location; see Fig. 11 in Appendix A. The thresholds are much lower and the KL divergences are much higher for most of the locations using 90% quantile than those using 95% quantile, which indicates that the common choice of 90% quantile may not be adequate here. A higher quantile than 95% could be chosen, but it would result in fewer exceedances and thus lead to higher uncertainty in the parameter estimates. In addition to the marginal fitting, the choice of threshold would also influence the end results of applying our spatial extremes hierarchical model, which will be discussed in Section 5.

#### – Process level

The three PP parameters are assumed to be spatially varying and modeled with latent Gaussian processes:

$$\begin{aligned}\mu_i &\sim \mathcal{N}(X_i^T \boldsymbol{\beta}_\mu + \phi_{i,\mu}, 1/\tau_\mu^2), i = 1, \dots, N, \\ \log(\sigma_i) &\sim \mathcal{N}(X_i^T \boldsymbol{\beta}_\sigma + \phi_{i,\sigma}, 1/\tau_\sigma^2), i = 1, \dots, N, \\ \xi_i &\sim \mathcal{N}(X_i^T \boldsymbol{\beta}_\xi + \phi_{i,\xi}, 1/\tau_\xi^2), i = 1, \dots, N,\end{aligned}$$

where  $X_i$  is a vector of covariate information for location  $i$ , and  $\boldsymbol{\beta}_\theta$  ( $\theta = \mu, \sigma$  or  $\xi$ ) is the vector of regression coefficients. We incorporate some geophysically-based covariates, i.e., longitude, latitude and elevation, all rescaled

to a range of  $[0, 1]$ . Here  $\tau_\theta^2$  ( $\theta = \mu, \sigma$  or  $\xi$ ) is a fixed precision, and  $\boldsymbol{\phi} = (\boldsymbol{\phi}_1^T, \boldsymbol{\phi}_2^T, \dots, \boldsymbol{\phi}_N^T)^T$  ( $\boldsymbol{\phi}_i = (\phi_{i,\mu}, \phi_{i,\sigma}, \phi_{i,\xi})^T, i = 1, 2, \dots, N$ ) is the spatial random effect.

The spatial random effects  $\boldsymbol{\phi}$  can be modeled with a proper multivariate conditional autoregressive (CAR) prior or an improper multivariate intrinsic autoregressive (IAR) prior (Besag et al. 1995; Besag and Kooperberg 1995; Gelfand and Vounatsou 2003; Kavanagh et al. 2016). In this study, we show that the proper CAR model is inadequate for capturing the irregular and strong spatial dependence given the high spatial resolution in our data; see discussion and Figs. 12 and 13 in Appendix A. Hence, we use the multivariate IAR model for  $\boldsymbol{\phi}$ , and impose a sum-to-zero constraint on  $\boldsymbol{\phi}$  as a remedy for the impropriety. The correlation structure is specified via the full conditionals, which are computationally convenient when using Gibbs sampling for drawing posteriors:

$$(\boldsymbol{\phi}_i | \boldsymbol{\phi}_{-i}, \mathbf{W}, \boldsymbol{\Sigma}) \sim \mathcal{N}_3 \left( \frac{\sum_{k=1}^N w_{ki} \boldsymbol{\phi}_i}{\sum_{k=1}^N w_{ki}}, \frac{\boldsymbol{\Sigma}}{\sum_{k=1}^N w_{ki}} \right),$$

where  $\boldsymbol{\phi}_{-i}$  denotes the set of spatial random effects except those at the  $i$ th location,  $\mathbf{W} = (w_{ij})$  is the adjacency matrix, and  $\boldsymbol{\Sigma}$  is the cross-variables covariance matrix. We use the common binary specification for  $\mathbf{W}$ , where its entry  $w_{ij} = 1$  if the grid cells  $i$  and  $j$  are adjacent, and is zero otherwise.

#### – Prior level

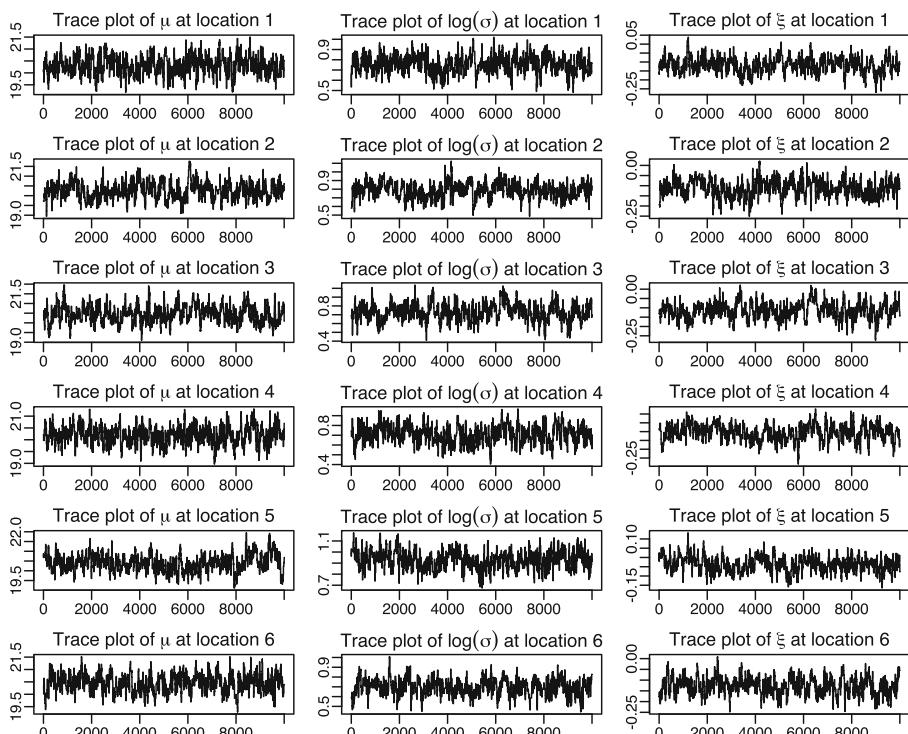
Each of the intercepts  $\beta_{0,\theta}$  ( $\theta = \mu, \sigma$ , or  $\xi$ ) is assigned an independent and conjugate Gaussian prior distribution, with mean being the GMLEs for each parameter  $\theta$  and with variance 100. For each regression coefficient  $\beta_{k,\theta}$  ( $\theta = \mu, \sigma$ , or  $\xi$ ), we assign a conjugate zero-mean Gaussian prior with variance 10. A conjugate inverse Wishart prior with 4 degrees of freedom is assigned to the matrix  $\boldsymbol{\Sigma}$ , as suggested by Lee (2013), with mean being the diagonal matrix with main diagonal  $(\text{var}(\hat{\mu}), \text{var}(\log(\hat{\sigma})), \text{var}(\hat{\xi}))$ , which corresponds to the respective empirical variabilities in the marginal GMLEs of the three PP parameters; here we use the data from the neighboring cluster for specifying the hyperparameters in  $\boldsymbol{\Sigma}$  to avoid using the data twice in a given cluster. Lastly, the fixed precisions in the regressions are set to be  $(\tau_\mu^2, \tau_\sigma^2, \tau_\xi^2) = (4, 200, 500)$ , which reflect the difference in scale between the three PP parameters, but are also chosen so that most of the variability would have to be explained by the spatial random effects  $\boldsymbol{\phi}$ .

## 4 Computational details

In this section, we provide computational details for the inference, using our data and the hierarchical spatial extremes model. For drawing posterior samples of the parameters, we adopt a Gibbs sampling scheme with Metropolis Hastings (MH) steps for the conditional distributions that do not have a closed form. Specifically, the parameters  $\mu_i, \sigma_i, \xi_i, i = 1, \dots, N$ , are updated with the MH algorithm, where the candidates for the three PP parameters are drawn as a block for each location from a uniform

distribution centered at current values. The parameters in  $\beta$  and  $\phi$  are updated as a block by sampling from the posterior multivariate normal distributions, using the methods depicted in Rue and Held (2005) (see their page 35 for the algorithm of sampling from a Gaussian Markov random field defined from its canonical representation). The parameters in  $\Sigma$  are updated as a block by sampling from the posterior inverse Wishart distribution. In addition, the starting values for the three PP parameters are set to be the marginal GMLEs because they may serve as a good guess and lead to faster convergence.

The procedure of fitting the model to each of the  $k = 200$  cluster neighbors independently can be performed in parallel, using multi-cores; this greatly improves the computational efficiency. Trace plot diagnostics (not shown) show a good convergence of all parameters after a burn-in period of 5,000 iterations. We then use the subsequent 5,000 iterations as posterior samples for inference purposes. For illustration, we randomly select one spatial cluster and six locations in that cluster, as shown in Fig. 1b. In Fig. 2, only those for the three PP parameters at the six locations within that cluster are represented. Trace plots for the other parameters are not shown, due to the length limit of this paper.

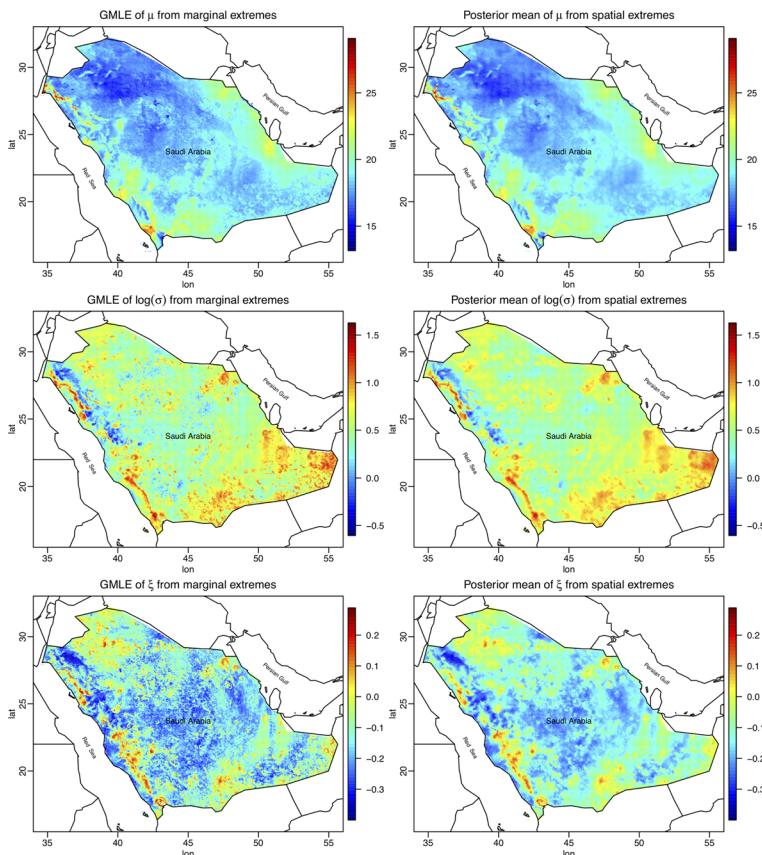


**Fig. 2** Trace plots of 10,000 MCMC iterations for the three PP parameters at six randomly selected locations

## 5 Results

In this section, we show the results of the model in Section 3.3 with the WRF wind dataset in Saudi Arabia. In Fig. 3, the posterior means of the three PP parameters from our spatial extremes modeling are shown on the right, and the marginal GMLEs are shown on the left for comparison. The estimations from spatial fitting are close to those from marginal fitting, while the former are smoother than the latter, especially for the shape parameter.

With the posterior distributions from MCMC samples for the three PP parameters, the inference on the wind extremes can then be made to assess the risk of disruption of wind turbine operations over Saudi Arabia. Extremal behaviors are usually summarized by return levels. The  $M$ -year return level is often described informally as an extreme event that would be expected to occur once in  $M$  years. More precisely, the  $M$ -year return level is the level which is exceeded in any one year with probability



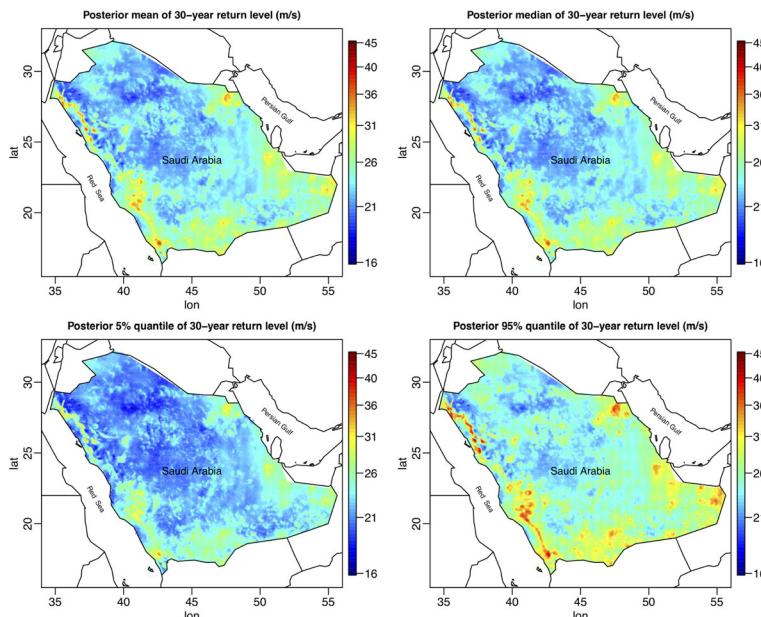
**Fig. 3** Left: GMLEs of the three PP parameters over Saudi Arabia from marginal fitting; Right: posterior means of the three PP parameters from MCMC samples with Bayesian hierarchical spatial extremes modeling

$1/M$ , which is simply the  $1 - 1/M$  quantile of the fitted PP. Specifically, the  $M$ -year return level at each location is given by

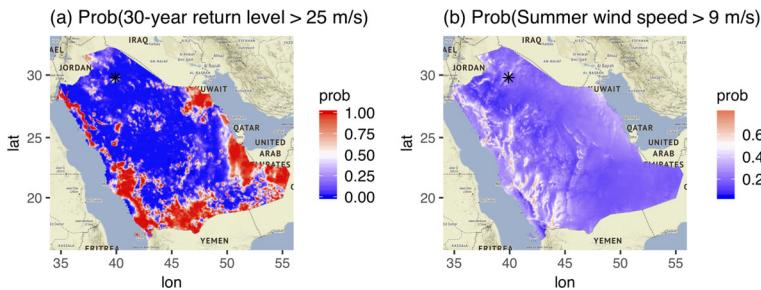
$$\hat{R}_i = \hat{\mu}_i + \frac{\hat{\sigma}_i}{\hat{\xi}_i} \left[ \{-\log(1 - 1/M)\}^{-\hat{\xi}_i} - 1 \right], i = 1, \dots, B,$$

where  $\hat{\mu}_i, \hat{\sigma}_i, \hat{\xi}_i, i = 1, \dots, B$ , are the posterior samples for the PP parameters from the Bayesian hierarchical spatial extremes model. In this work, we are interested in the risk of disruption of wind turbine operations within its lifetime, which is typically 30 years. Hence, we compute the 30-year return levels, and perform the inference based on the posterior distributions of return levels. Figure 4 represents the posterior mean, median, 5% quantile and 95% quantile of 30-year log return levels from MCMC samples.

Figure 5a shows the map of probabilities of the 30-year return levels that exceed a general cut-out wind speed from MCMC samples, at each location, reflecting the risk of disruption of wind turbine operations within their lifetime. The cut-out wind speed denotes the speed at which the turbine shuts down by some sort of stalling or braking mechanism to avoid damage from further operation in high winds. The general cut-out wind speed is 25 m/s for a typical modern wind turbine, as well as for some of the most promising turbines for the study of sustainability in Saudi Arabia, such as the GE Energy 3.4-137 turbine, the Nordex N117/3000 turbine and the Nordex N131/3600 turbine, whose power curves are shown in Fig. 6. The wind usually has to return to a much lower speed, called the cut-back-in wind speed, which is typically

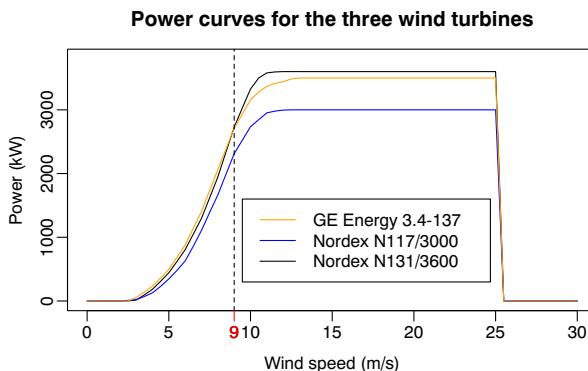


**Fig. 4** Posterior mean, median, 5% quantile and 95% quantile of 30-year return levels (m/s) from MCMC samples with Bayesian hierarchical spatial extremes modeling over Saudi Arabia. Maps are plotted in log scale



**Fig. 5** **a** Probability map over Saudi Arabia showing the risk of disruption of wind turbine operations from MCMC samples with Bayesian hierarchical spatial extremes modeling, using 25 m/s as the cut-out wind speed. **b** Probability map of the Summer wind speeds that exceed 9 m/s. “prob” stands for “probability”. The location of the first wind farm at Dumat Al-Jandal in Saudi Arabia is labeled as “\*”

20 m/s at which the blades are reverted to move in the wind, for a certain amount of time before the turbine will restart to generate power. Figure 5a indicates a large area over Saudi Arabia that would suffer from high risks of disruptions of wind turbine operations, mostly located in the South, and near the Red Sea and the Persian Gulf. To site a wind farm and improve the stability of wind power generation, one must take into account the risk of shutdowns of wind turbines over a possibly long period of time. Figure 5b shows the map of probabilities of the Summer wind speeds that exceed 9 m/s, at which the aforementioned three representative wind turbines would operate with at least half of the maximum power rate (called the rated output power). It shows that the west coastal regions exhibit high winds in Summer. These locations were also identified by Yip et al. (2016) and Chen et al. (2018) as having a high wind power potential. Saudi Arabia is planning to build its first wind farm at Dumat Al Jandal (labeled as “\*” in Fig. 5) with an installed capacity of 400-megawatts (MW) (Dumat Al Jandal wind farm 2017). The probability of disruption of wind turbine operations at this location is very low (0.72%) albeit with an average mean wind speed in Summer that is not too high (about 6 m/s). Hence, the country’s first



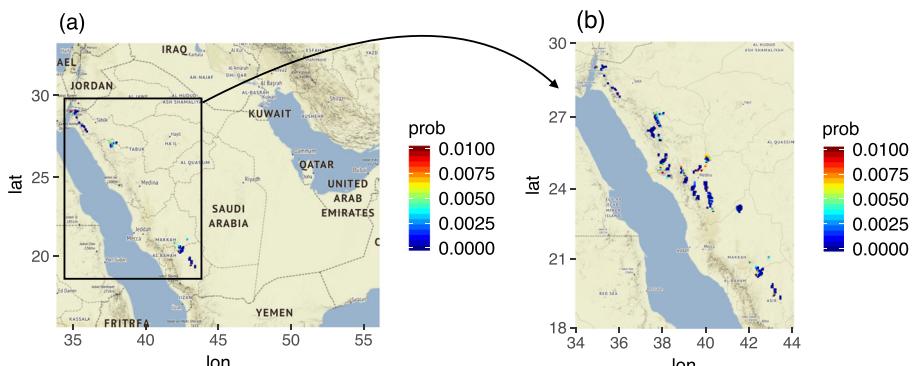
**Fig. 6** Power curves for the three representative wind turbines

wind farm would ensure stability of operations, and choosing appropriate turbines is important in order to increase the power production (e.g., choosing turbines that have smaller cut-in speed at which the turbines first start to rotate and generate power).

Figure 7 highlights the 375 grid cells over Saudi Arabia where the wind speeds exceed 9 m/s for at least half of the time in Summer, and the risk of disruption of wind turbine operations is lower than 1%. These locations may be potentially best sites to launch wind farms that could generate high wind power persistently with low risk of being disrupted from operations.

As we have mentioned in Sections 3.2 and 3.3, here we present a sensitivity analysis on the choices of  $k$  and the threshold. We compare the performance of the Bayesian hierarchical model in four scenarios with different choices of  $k$  (i.e.,  $k = 200$  and  $k = 250$ ) and threshold (i.e., 90% and 95% quantiles) based on the deviance information criterion (DIC) defined as  $DIC = 2\bar{D}(\theta) - D(\bar{\theta})$ , where  $\theta$  is the vector of the three GPD parameters and  $D(\theta) = -2 \log p(z|\theta) + C$  is the deviance function, where  $C$  is a constant that cancels out when comparing different models. Based on the posterior MCMC samples,  $\bar{D}(\theta)$  is calculated as the average of  $D(\theta)$  over the samples of  $\theta$ , and  $D(\bar{\theta})$  is the value of  $D(\theta)$  evaluated at the average of the samples of  $\theta$ . The scaled DIC (i.e., total DIC divided by 83,981, the number of grid cells) and the total computation time (with 25 cores on a cluster of 2.4 GHz processors) for each of the four scenarios is shown in Table 1. Interestingly, the total computation time is about the same for the four scenarios. Based on DIC, the choice of  $k = 200$  and 95% quantile as the threshold yields the lowest DIC, though the set-up of  $k = 250$  and threshold at 95% quantile performs comparatively well. For the end results of the “best” locations with high Summer winds and low risk of disruption of wind turbine operations, these two settings have identified 324 common locations, which consist of approximately 87% of the total number of the “best” locations (see Fig. 14 in Appendix A).

#### Locations with high summer winds and low risk



**Fig. 7** Potential best locations for siting wind farms over Saudi Arabia where the wind speeds exceed 9 m/s for at least half of the time in Summer, and the risk of disruption of wind turbine operations is lower than 1%. “prob” stands for “probability of the 30-year return levels that exceed 25 m/s”

**Table 1** DIC and computation time for the Bayesian hierarchical model using different values of  $k$  and thresholds. The lowest DIC is in bold

Scenarios	DIC	Computation time
$k = 200$ , Threshold = 90% quantile	143.5	6.7 h
$k = 200$ , Threshold = 95% quantile	<b>-144.2</b>	6.5 h
$k = 250$ , Threshold = 90% quantile	129.9	8.2 h
$k = 250$ , Threshold = 95% quantile	-141.0	6.1 h

## 6 Conclusion and discussion

In this work, we assessed the risk of disruption of wind turbine operations over Saudi Arabia, using a hierarchical spatial extremes model in a Bayesian setting. We used hourly wind speed data from a unique WRF dataset with high spatial resolution. For computational feasibility, we divided our study region into subregions, using the spatial clustering method, and then fitted the model to each subregion. Instead of fitting the model to one single spatial cluster at a time, we borrowed information from the neighboring spatial clusters for the purpose of spatial smoothing. The computational efficiency was largely improved by parallel computing on the subregions. Lastly, we showed the inadequacy of using a multivariate CAR prior for capturing the strong dependence in spatial random effects, and that the multivariate IAR appears to be more appropriate for our high-resolution data. The risk maps show that a large area over Saudi Arabia would suffer from high risks of disruptions of wind turbine operations, mostly in the South and near the Red Sea and Persian Gulf. We also identified locations with high winds in Summer and a low risk of turbine disruptions, which can be potentially good locations for harvesting a persistent wind energy.

One limitation for our methodology is that the uncertainty due to the clustering was not taken into account. Reich and Shaby (2019) proposed a method for areal extremes that accounts for spatial dependence using latent clustering of neighboring regions. However, there the spatial dependence parameter  $\phi$  in the Potts model for the cluster labels is difficult to estimate even for moderate sample size. Reich and Shaby (2019) proposed an informal way to estimate  $\phi$ , which is largely uncertain, and then fixed it for their remaining analysis. Methods to estimate  $\phi$ , such as MCMC or cross-validation, are cumbersome to implement for large datasets. Another restriction of the method in Reich and Shaby (2019) is that the uncertainty for the number of possible clusters,  $k$ , is not quantified; instead,  $k$  is fixed (as the number of grid cells) for simplicity, which may be inappropriate especially for data (such as our high-resolution WRF data) with strong spatial dependence where a far smaller number of possible clusters is sufficient. Furthermore, including the clustering method of Reich and Shaby (2019) into our model would largely increase the computation complexity and is almost infeasible for our large dataset. For one subregion (i.e., the NEOM region) in the northwest of Saudi Arabia with 2,254 locations in our dataset where we have tried to fit our model with the clustering included and  $k$  is fixed at a small value (i.e.,  $k = 50$ ), it takes around 3 hours per 1,000 iterations. Therefore, we decided to choose the value of  $k$  for spatial clustering prior to the Bayesian hierarchical

modeling. However, it may still be necessary to account for the uncertainty of the clustering in a more sophisticated way than the sensitivity analyses done in this study.

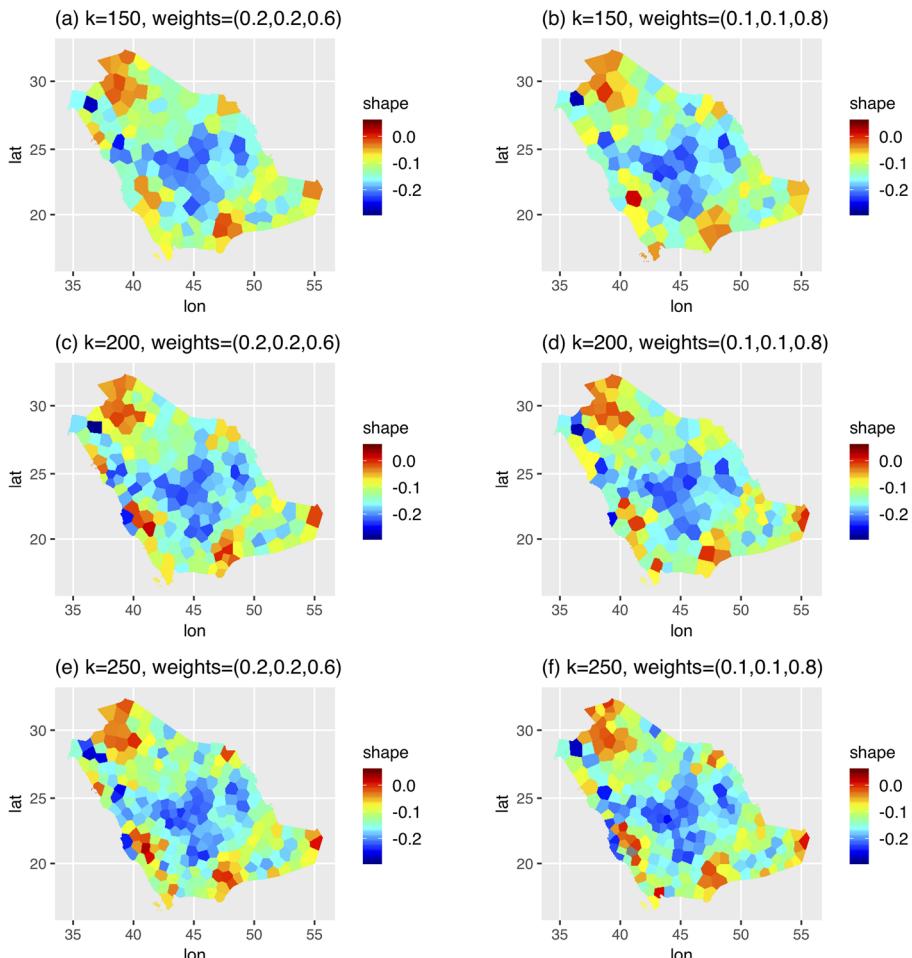
Another limitation is that the spatial clustering is performed using the marginal shape estimates  $\xi$ , which in practice, are largely uncertain, and this may influence the clustering result. Since the accuracy for a marginal estimate  $\xi_i$  at location  $i$  is negatively correlated with the length,  $N_{u_i}$ , of the temporal cluster maxima ( $N_{u_i} \in [39, 162]$ ) used to fit the GPD model marginally, we redid the clustering where locations with  $N_{u_i} < 50$  (i.e., with highly uncertain shape estimates) are removed. Figure 15 in Appendix A indicates that the clustering results are quite similar with or without those locations with highly uncertain marginal  $\xi$  estimates. In order to take the uncertainty of marginal  $\xi$  estimates into account in a more comprehensive way, clustering methods for uncertain data, such as those in Kriegel and Pfeifle (2005a, b), could be used, but they are computationally cumbersome for our large dataset. Since we are not directly interested in the clustering, and as we can see from our sensitivity analyses that the clustering result would not largely affect the end results for applying our model, we think that this is a minor issue in our study.

Finally, given only 6 years of data, there is a concern that many large-scale climate drivers (e.g., NAO, AMO) that would have some impact on the region of Saudi Arabia work on larger time scales (e.g., AMO works on a 60-year time scale), and therefore it is quite likely that the simulated WRF data are only from a particular regime. A more accurate quantification of wind extremes over Saudi Arabia may be achieved by combining the high spatial-resolution data with long-term climate simulations (e.g., the simulations from GCMs running for centuries at relatively coarse spatial resolutions) which are able to capture the large-scale dynamics of the atmosphere. However, Mann et al. (2020) found no evidence of internal multidecadal and interdecadal oscillations, such as an AMO, in climate model simulations, and they claimed that the only dynamic signal distinguishable from background noise was ENSO, which typically occurs every 2 to 7 years. These findings may provide justifications for our study based on only 6 years of data.

Future work directions could provide information about the duration of wind extremes events that would be helpful to understand how long a wind turbine will be out of action. Methods such as that in Telesca et al. (2018) could be used to analyze the temporal properties of wind extremes. A subsequent problem is an estimation of the economic loss due to the disruption of wind turbines in the electrical grid, using the power curves associated with individual turbines. Additionally, based on Zhu et al. (2014) and Xie et al. (2014) who incorporated advanced space-time wind forecasts into electric power system scheduling, we may include wind extremes in the power system dispatch problem. In addition to wind speed, many other environmental factors, such as temperature, air pressure, turbulence intensity, wind shear and humidity, can potentially disrupt wind turbine operations. Therefore, it would be valuable, in future research studies, to provide a risk assessment of wind turbine operations, using a multivariate spatial extremes model that incorporates multiple environmental factors.

**Acknowledgements** This publication is based on research supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No: OSR-2018-CRG7-3742 and in part by the Center of Excellence for NEOM Research at KAUST. We are grateful to Professor Georgiy Stenchikov's group, the Atmospheric and Climate Modeling group at KAUST, for producing and providing the high-resolution WRF dataset. Many thanks also to Professor Daniel Cooley for providing the codes used in the article by Cooley and Sain (2010).

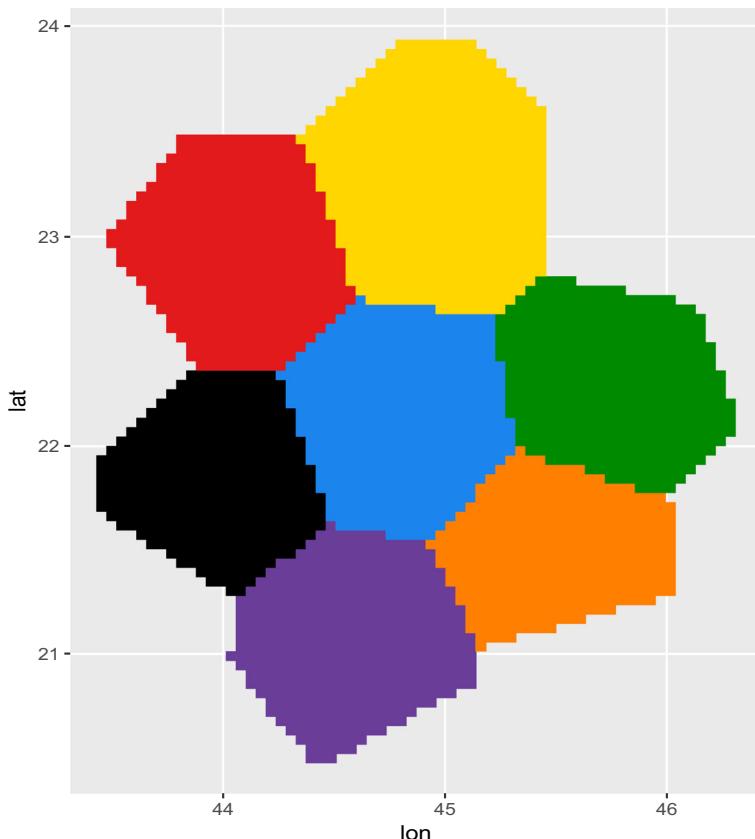
## Appendix A: Supplementary figures, tables and discussions



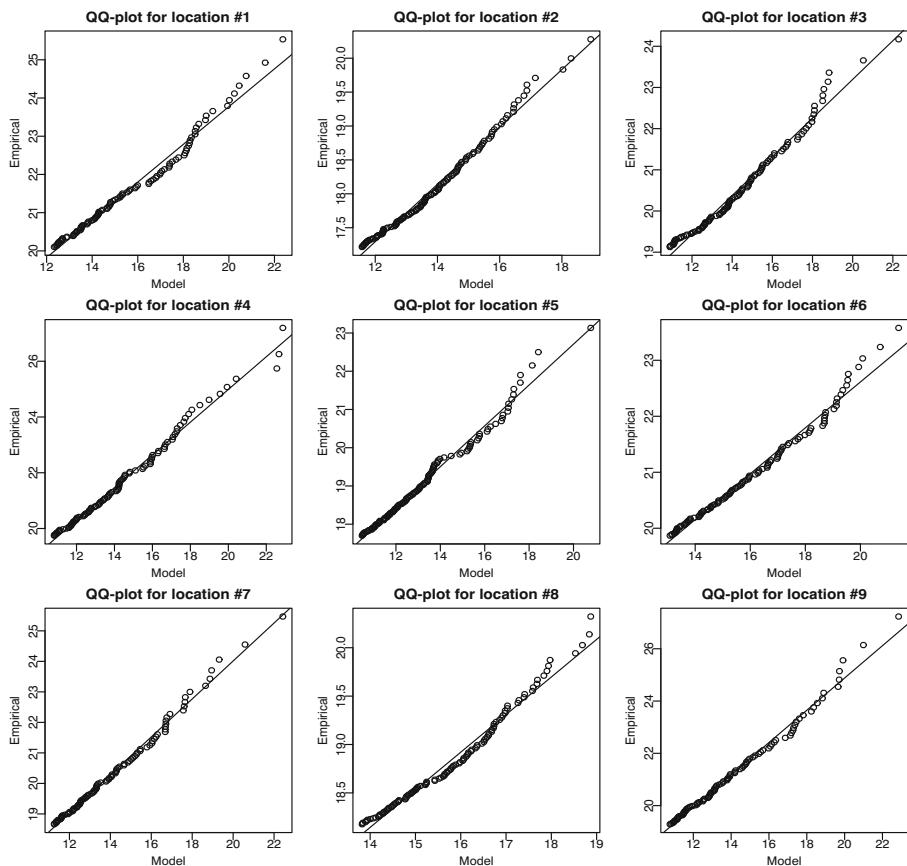
**Fig. 8** Cluster means for GMLEs of the shape parameter  $\xi$  with different values of  $k$  and different assigned weights for longitude, latitude and GMLEs for  $\xi$ , respectively

**Table 2** Range of cluster sizes with different values of  $k$  and different weights. The two “best” configurations are in bold

$k$	Weights			Range of cluster sizes
	longitude	latitude	$\xi$	
150	0.2	0.2	0.6	[283,1035]
<b>200</b>	<b>0.2</b>	<b>0.2</b>	<b>0.6</b>	<b>[242,695]</b>
<b>250</b>	<b>0.2</b>	<b>0.2</b>	<b>0.6</b>	<b>[190,560]</b>
150	0.1	0.1	0.8	[286,1006]
200	0.1	0.1	0.8	[153,862]
250	0.1	0.1	0.8	[154,673]



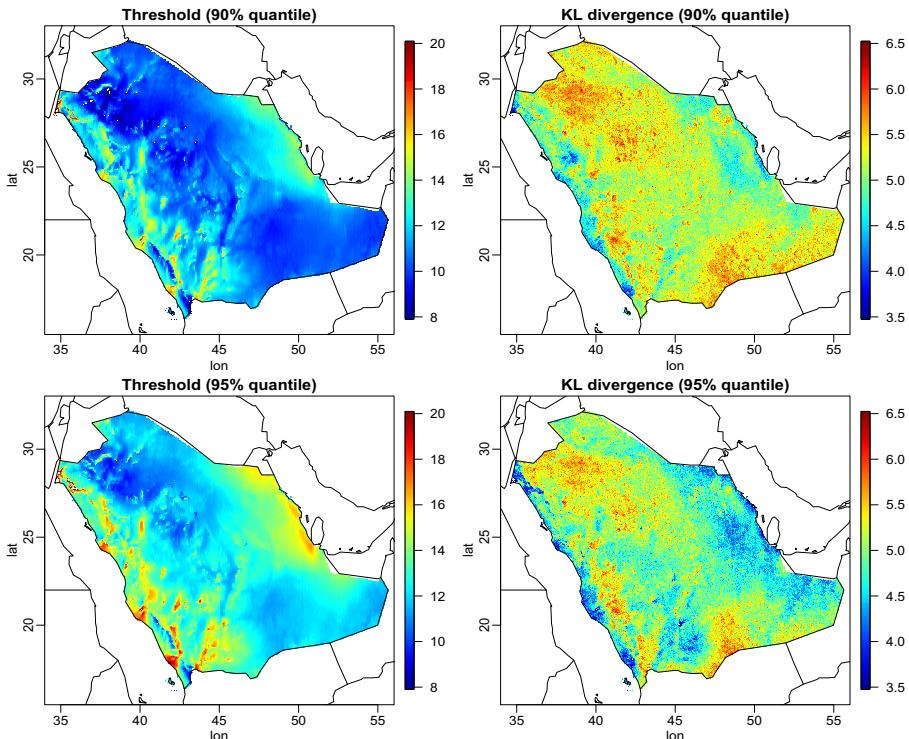
**Fig. 9** Example for a cluster neighbor. Each color represents one single cluster, and all the seven clusters represent the cluster neighbor of the central cluster in blue



**Fig. 10** QQ-plots of marginal GPD fitting for 9 randomly selected locations, using 95% quantile as the threshold

### Discussion on CAR and IAR specifications for spatial random effects

There have been discussions on whether the proper or improper CAR (referred to as IAR) specification should be used in practice (e.g., Besag et al. 1995, Besag and Kooperberg 1995, Gelfand and Vounatsou 2003 and Banerjee et al. 2014). As Besag et al. (1995) showed, the marginal maximal bivariate correlation that can be captured with a proper Gaussian field is around 0.6. Besag and Kooperberg (1995) pointed out a common disadvantage of a proper CAR that appreciable correlations between the spatial random effects at neighboring sites require parameter values extremely close to a particular boundary of the parameter space. Gelfand and Vounatsou (2003) demonstrated that IAR is analogous to the nonstationary or random walk case in familiar autoregressive time series models and can be advantageous in accommodating more irregular spatial behaviors. Banerjee et al. (2014) also claimed that the breadth of spatial patterns may be too limited if the proper CAR is used, and the improper IAR choice may actually enable a wider scope for posterior spatial patterns.

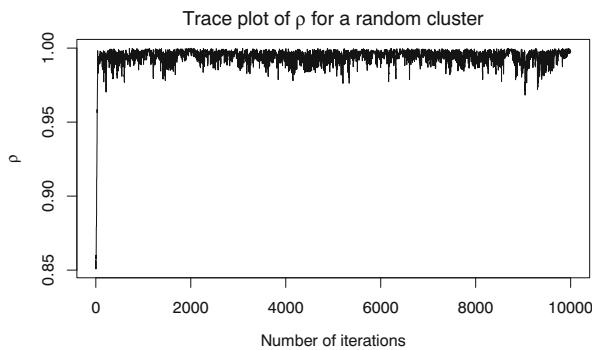


**Fig. 11** Values of threshold and KL divergence for marginal extremes using 90% (top) and 95% (bottom) quantile, respectively, as a rule of thumb in the GPD model

Here we first implement our hierarchical model with the proper CAR prior used for the spatial random effects  $\phi$ . The multivariate CAR model (Kavanagh et al. 2016) we used is given by:

$$\phi \sim \mathcal{N}_{3N} \left( \mathbf{0}, [\mathbf{Q}(\mathbf{W}, \rho) \otimes \boldsymbol{\Sigma}^{-1}]^{-1} \right),$$

where  $\mathbf{Q}(\mathbf{W}, \rho) = \rho[\text{diag}(\mathbf{W}\mathbf{1}) - \mathbf{W}] + (1 - \rho)\mathbf{I}$  ( $\mathbf{1}$  is the  $N \times 1$  vector of ones,  $\mathbf{I}$  is the  $N \times N$  identity matrix) is the  $N \times N$  precision matrix for the joint distribution corresponding to the CAR prior proposed by Leroux et al. (2000), while  $\boldsymbol{\Sigma}$  is a  $3 \times 3$  cross-variables covariance matrix. The matrix  $\mathbf{Q}(\mathbf{W}, \rho)$  controls the spatial autocorrelation structure of the random effects, and is based on a non-negative symmetric  $N \times N$  neighborhood (or adjacency) matrix  $\mathbf{W}$ , and a spatial dependence parameter  $\rho$ . We use the common binary specification for  $\mathbf{W}$ , where its entry  $w_{ij} = 1$  if the grid cells  $i$  and  $j$  are adjacent, and is zero otherwise. The parameter  $\rho$  is a spatial autoregressive parameter, with  $\rho$  close to one corresponding to strong spatial dependence and  $\rho = 0$  corresponding to independence in space. When  $\rho = 1$ , we



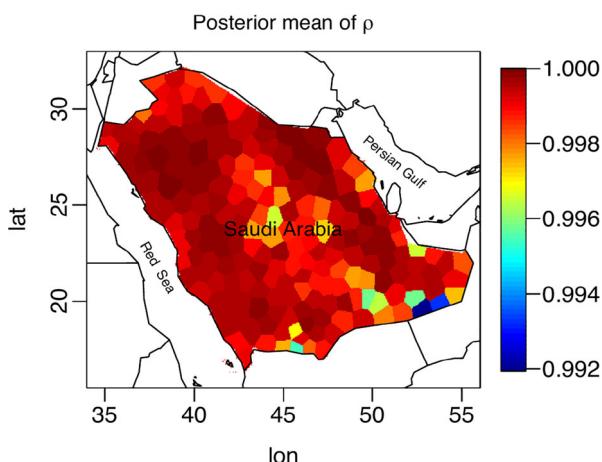
**Fig. 12** Trace plot of 10,000 MCMC iterations for the parameter  $\rho$  in the MCAR model in a random spatial cluster as selected in Fig. 1 in the main text

obtain the multivariate IAR model. The correlation structure is specified via the full conditionals:

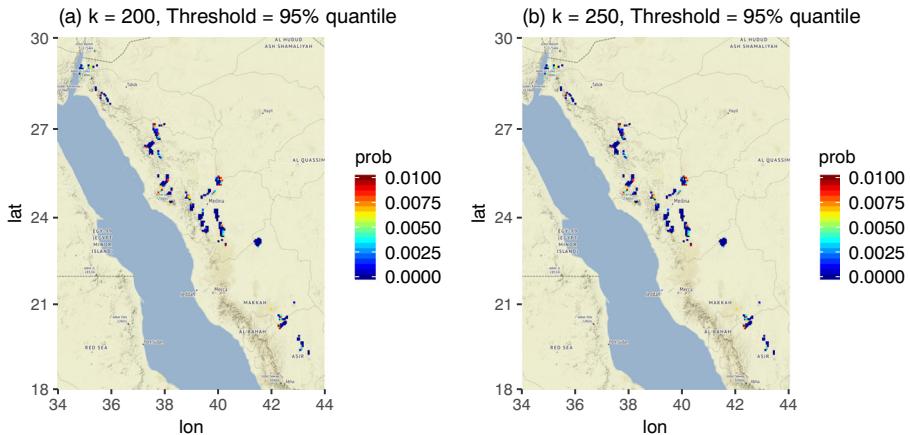
$$(\phi_i | \phi_{-i}, W, \Sigma, \rho) \sim \mathcal{N}_3 \left( \frac{\rho \sum_{k=1}^N w_{ki} \phi_i}{\rho \sum_{k=1}^N w_{ki} + 1 - \rho}, \frac{\Sigma}{\rho \sum_{k=1}^N w_{ki} + 1 - \rho} \right),$$

where  $\phi_{-i}$  denotes the set of spatial random effects except those at the  $i$ th location. With the choice of the matrix  $W$ , the conditional expectation of spatial random effect at one location is a weighted average of the random effects in its adjacent locations, and the covariance is weighted by the number of adjacent locations.

In our Bayesian hierarchical model, a Uniform[0, 1] prior is assigned to the spatial autoregressive parameter  $\rho$ , as the negative spatial autocorrelation is rarely seen in practice in spatial areal unit data (Tobler 1970), and  $\rho \in [0, 1]$  is a sufficient condition for the covariance matrix of the joint distribution to be nonsingular (Banerjee

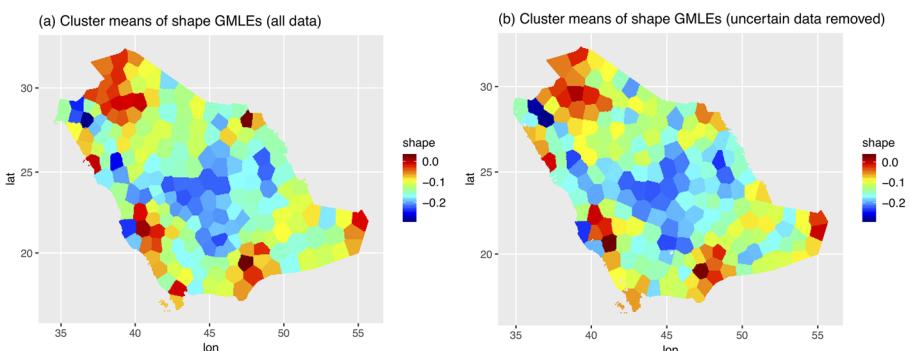


**Fig. 13** Posterior mean of the spatial autoregressive parameter  $\rho$  in the MCAR model from MCMC samples with Bayesian hierarchical spatial extremes modeling



**Fig. 14** Potential best locations for siting wind farms over Saudi Arabia where the wind speeds exceed 9 m/s for at least half of the time in Summer, and the risk of disruption of wind turbine operations is lower than 1% for the setting of **a**  $k = 200$ , Threshold = 95% quantile and **b**  $k = 250$ , Threshold = 95% quantile. “prob” stands for “probability of the 30-year return levels that exceed 25 m/s”

et al. 2014). The parameter  $\rho$  is updated with the MH algorithm, where the candidate for  $\rho$  is drawn from a truncated normal distribution in the unit interval so as to bound  $\rho$  in  $[0, 1]$ . Other settings for priors and computational details are the same as in the main text. The posterior density from MCMC samples for  $\rho$  is peaked near  $\rho = 1$  for all subregions (see Figs. 12 and 13), suggesting that there is more spatial dependence in the data than the model can capture. Therefore, we replace the multivariate CAR with the multivariate IAR in order to capture the irregular and strong spatial dependence in our high-resolution data. Although the IAR is improper, we are only using it as a prior; the posterior will typically still emerge as proper, so Bayesian inference can still proceed. On the other hand, we can impose a sum-to-zero constraint on  $\phi$  as a remedy of impropriety, which is numerically convenient in the MCMC sampling procedure.



**Fig. 15** Cluster means of shape GMLEs with **a** all data and **b** locations with highly uncertain marginal shape estimates (i.e., where the length of temporal cluster maxima,  $N_{u_t}$ , is less than 50) removed

## Appendix B: Supplementary R codes

The datasets generated and/or analyzed during the current study were used under license, and so are not publicly available. Data are available however from the corresponding author upon reasonable request. The R codes related to this article can be found online at the github repository: <https://github.com/wanruofenfang123/Bayesian-Hierarchical-Spatial-Extremes>.

## References

- Banerjee, S., Carlin, B.P., Gelfand, A.E.: Hierarchical Modeling and Analysis for Spatial Data, 2nd ed. Monographs on Statistics and Applied Probability. Chapman and Hall/CRC, Boca Raton (2014)
- Besag, J., Kooperberg, C.: On conditional and intrinsic autoregressions. *Biometrika* **82**(4), 733–746 (1995)
- Besag, J., Green, P., Higdon, D., Mengerson, K.: Bayesian computation and stochastic systems. *Stat. Sci.* **10**(1), 3–41 (1995)
- Chen, W., Castruccio, S., Genton, M.G., Crippa, P.: Current and future estimates of wind energy potential over Saudi Arabia. *J. Geophys. Res.: Atmos.* **123**(12), 6443–6459 (2018)
- Coles, S.: An Introduction to Statistical Modeling of Extreme Values, vol. 208. Springer, Berlin (2001)
- Cooley, D., Sain, S.R.: Spatial hierarchical modeling of precipitation extremes from a regional climate model. *J. Agric. Biol. Environ. Stat.* **15**(3), 381–402 (2010)
- Cooley, D., Nychka, D., Naveau, P.: Bayesian spatial modeling of extreme precipitation return levels. *J. Am. Stat. Assoc.* **102**(479), 824–840 (2007)
- Davison, A.C., Smith, R.L.: Models for exceedances over high thresholds. *J. R. Stat. Soc. Ser. B (Methodol.)* **52**(3), 393–442 (1990)
- Davison, A.C., Padoan, S.A., Ribatet, M.: Statistical modeling of spatial extremes. *Stat. Sci.* **27**(2), 161–186 (2012)
- de Haan, L., Ferreira, A.: Extreme Value Theory: an Introduction. Springer, New York (2006)
- de Winter, R., Sterl, A., Ruessink, B.: Wind extremes in the North Sea Basin under climate change: An ensemble study of 12 CMIP5 GCMs. *J. Geophys. Res.: Atmos.* **118**(4), 1601–1612 (2013)
- Dumat Al Jandal wind farm: <https://dumatajandal.com>, Accessed: 20 May 2019 (2017)
- DuMouchel, W.H.: Estimating the stable index  $\alpha$  in order to measure tail thickness: a critique. *Ann. Stat.* **11**(4), 1019–1031 (1983)
- Emeis, S.: How well does a power law fit to a diabatic boundary-layer wind profile? *DEWI Mag.* **26**, 59–62 (2005)
- Fisher, R.A., Tippett, L.H.C.: Limiting forms of the frequency distribution of the largest or smallest member of a sample. In: Mathematical Proceedings of the Cambridge Philosophical Society, vol. 24, pp. 180–190. Cambridge University Press (1928)
- Gaetan, C., Grigoletto, M.: A hierarchical model for the analysis of spatial rainfall extremes. *J. Agric. Biol. Environ. Stat.* **12**(4), 434–449 (2007)
- Gelfand, A.E., Vounatsou, P.: Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics* **4**(1), 11–15 (2003)
- Giani, P., Tagle, F., Genton, M.G., Castruccio, S., Crippa, P.: Closing the gap between wind energy targets and implementation for emerging countries. *Appl. Energy* **269**, 115085 (2020)
- Gnedenko, B.: Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.* **44**, 423–453 (1943)
- Heffernan, J.E., Tawn, J.A.: A conditional approach for multivariate extreme values (with discussion). *J. R. Stat. Soc.: Ser. B (Stat. Methodol.)* **66**(3), 497–546 (2004)
- Hosking, J., Wallis, J.R.: Regional Frequency Analysis: an Approach Based on L-Moments. Cambridge University Press, Cambridge (1997)
- Huser, R., Davison, A.: Space–time modelling of extreme events. *J. R. Stat. Soc.: Ser. B (Stat. Methodol.)* **76**(2), 439–461 (2014)

- Jalbert, J., Favre, A.C., Bélisle, C., Angers, J.F.: A spatiotemporal model for extreme precipitation simulated by a climate model, with an application to assessing changes in return levels over North America. *J. R. Stat. Soc.: Ser. C (Appl. Stat.)* **66**(5), 941–962 (2017)
- KA-CARE: King Abdullah City for Atomic and Renewable Energy: Building the Renewable Energy Sector in Saudi Arabia. Report (2012)
- Kavanagh, L., Lee, D., Pryce, G.: Is poverty decentralizing? Quantifying uncertainty in the decentralization of urban poverty. *Ann. Am. Assoc. Geogr.* **106**(6), 1286–1298 (2016)
- Kharin, V.V., Zwiers, F.W.: Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere–ocean GCM. *J. Clim.* **13**(21), 3760–3788 (2000)
- Kriegel, H.P., Pfeifle, M.: Density-based clustering of uncertain data. In: Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, pp. 672–677. ACM (2005a)
- Kriegel, H.P., Pfeifle, M.: Hierarchical density-based clustering of uncertain data. IEEE (2005b)
- Kumar, D., Mishra, V., Ganguly, A.R.: Evaluating wind extremes in CMIP5 climate models. *Clim. Dyn.* **45**(1–2), 441–453 (2015)
- Ledford, A.W., Tawn, J.A.: Statistics for near independence in multivariate extreme values. *Biometrika* **83**(1), 169–187 (1996)
- Ledford, A.W., Tawn, J.A.: Modelling dependence within joint tail regions. *J. R. Stat. Soc.: Ser. B (Stat. Methodol.)* **59**(2), 475–499 (1997)
- Lee, D.: CARBayes: an R package for Bayesian spatial modeling with conditional autoregressive priors, vol. 55 (2013)
- Leroux, B.G., Lei, X., Breslow, N.: Estimation of disease rates in small areas: a new mixed model for spatial dependence. In: Statistical Models in Epidemiology, the Environment, and Clinical Trials, pp. 179–191. Springer, Berlin (2000)
- Mann, E.M., Steinman, B.A., Miller, S.K.: Absence of internal multidecadal and interdecadal oscillations in climate model simulations. *Nat. Commun.* **11**(49), 1–9 (2020)
- Martins, E.S., Stedinger, J.R.: Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resour. Res.* **36**(3), 737–744 (2000)
- NEOM Project: <https://www.neom.com/>, Accessed: 28 Aug 2018 (2017)
- Nikulin, G., Kjellström, M.E., Hansson, U., Strandberg, G., Ullerstig, A.: Evaluation and future projections of temperature, precipitation and wind extremes over Europe in an ensemble of regional climate simulations. *Tellus A: Dyn. Meteorol. Oceanogr.* **63**(1), 41–55 (2011)
- Pickands, J.: Statistical inference using extreme order statistics. *Ann. Stat.* **3**(1), 119–131 (1975)
- Pryor, S., Barthelmie, R.: Assessing climate change impacts on the near-term stability of the wind energy resource over the United States. *Proc. Natl. Acad. Sci.* **108**(20), 8167–8171 (2011)
- Ramos, A., Ledford, A.: A new class of models for bivariate joint tails. *J. R. Stat. Soc.: Ser. B (Stat. Methodol.)* **71**(1), 219–241 (2009)
- Rehman, S., Ahmad, A.: Assessment of wind energy potential for coastal locations of the Kingdom of Saudi Arabia. *Energy* **29**(8), 1105–1115 (2004)
- Rehman, S., El-Amin, I., Ahmad, F., Shaahid, S., Al-Shehri, A., Bakhshwain, J.: Wind power resource assessment for Rafha, Saudi Arabia. *Renew. Sustain. Energy Rev.* **11**(5), 937–950 (2007)
- Reich, B.J., Shaby, B.A.: A spatial Markov model for climate extremes. *J. Comput. Graphic. Stat.* **28**(1), 117–126 (2019)
- Resnick, S.I.: Extreme Values, Regular Variation and Point Processes. Springer, New York (1987)
- Rue, H., Held, L.: Gaussian Markov Random Fields: Theory and Applications. Chapman and Hall/CRC, Boca Raton (2005)
- Sang, H., Gelfand, A.E.: Hierarchical modeling for extreme values observed over space and time. *Environ. Ecol. Stat.* **16**(3), 407–426 (2009)
- Sang, H., Gelfand, A.E.: Continuous spatial process models for spatial extreme values. *J. Agric. Biol. Environ. Stat.* **15**(1), 49–65 (2010)
- Scarrott, C., MacDonald, A.: A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT-Stat. J.* **10**(1), 33–60 (2012)
- Shaahid, S., Al-Hadhrami, L.M., Rahman, M.: Potential of establishment of wind farms in western province of Saudi Arabia. *Energy Procedia* **52**, 497–505 (2014)
- Smith, R.L.: Maximum likelihood estimation in a class of nonregular cases. *Biometrika* **72**(1), 67–90 (1985)

- Smith, R.L.: Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Stat. Sci.* **4**(4), 367–377 (1989)
- Tagle, F., Castruccio, S., Crippa, P., Genton, M.G.: A non-Gaussian spatio-temporal model for daily wind speeds based on a multivariate skew-t distribution. *J. Time Ser. Anal.* **40**, 312–326 (2019)
- Telesca, L., Guignard, F., Laib, M., Kanevski, M.: Analysis of temporal properties of wind extremes. *arXiv:180808847* (2018)
- Tobler, W.R.: A computer movie simulating urban growth in the Detroit region. *Econ. Geogr.* **46**(sup1), 234–240 (1970)
- Towe, R.P., Tawn, J.A., Lamb, R., Sherlock, C.G.: Model-based inference of conditional extreme value distributions with hydrological applications. *Environmetrics* **30**, e2575 (2019)
- Turkman, K.F., Turkman, M.A., Pereira, J.: Asymptotic models and inference for extremes of spatio-temporal data. *Extremes* **13**(4), 375–397 (2010)
- Vision 2030: <http://vision2030.gov.sa/en/node/87>, Accessed: 28 Aug 2018 (2016)
- Weiss, J., Bernardara, P., Benoit, M.: Modeling intersite dependence for regional frequency analysis of extreme marine events. *Water Resour. Res.* **50**(7), 5926–5940 (2014)
- Winter, H.C., Tawn, J.A., Brown, S.J., et al.: Modelling the effect of the El Nino-Southern Oscillation on extreme spatial temperature events over Australia. *Ann. Appl. Stat.* **10**(4), 2075–2101 (2016)
- Xie, L., Gu, Y., Zhu, X., Genton, M.G.: Short-term spatio-temporal wind power forecast in robust look-ahead power system dispatch. *IEEE Trans. Smart Grid* **5**(1), 511–520 (2014)
- Yip, C.M.A.: Statistical Characteristics and Mapping of Near-Surface and Elevated Wind Resources in the Middle East. PhD thesis, King Abdullah University of Science and Technology (2018)
- Yip, C.M.A., Gunturu, U.B., Stenchikov, G.L.: Wind resource characterization in the Arabian Peninsula. *Appl. Energy* **164**, 826–836 (2016)
- Zhu, X., Genton, M.G., Gu, Y., Xie, L.: Space-time wind speed forecasting for improved power system dispatch. *TEST* **23**(1), 1–25 (2014)

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Affiliations

Wanfang Chen<sup>1</sup>  · Stefano Castruccio<sup>2</sup> · Marc G. Genton<sup>1</sup>

Stefano Castruccio  
scastruc@nd.edu

Marc G. Genton  
marc.genton@kaust.edu.sa

<sup>1</sup> Statistics Program, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

<sup>2</sup> Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, Notre Dame, IN 46556, USA