

# A temporal model for vertical extrapolation of wind speed and wind energy assessment

Paola Crippa<sup>a,\*</sup>, Mariana Alifa<sup>a</sup>, Diogo Bolster<sup>a</sup>, Marc G. Genton<sup>b</sup>, Stefano Castruccio<sup>c</sup>

<sup>a</sup> Department of Civil and Environmental Engineering and Earth Sciences, University of Notre Dame, Notre Dame, IN, USA

<sup>b</sup> Statistics Program, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

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

## HIGHLIGHTS

- A temporal model to extrapolate hourly winds to turbine hub height is proposed.
- The proposed model outperforms traditional vertical extrapolation methods.
- A time-varying variance model improves extrapolation for stable/unstable regimes.
- The temporal model reduces the absolute bias in capacity factors by 58–64%.

## ARTICLE INFO

### Keywords:

Atmospheric stability  
Power law  
Temporal model  
Wind energy assessment  
Wind extrapolation  
Wind shear coefficient

## ABSTRACT

Accurate wind speed estimates at turbine hub height are critical for wind farm operational purposes, such as forecasting and grid operation, but also for wind energy assessments at regional scales. Power law models have widely been used for vertical wind speed profiles due to their simplicity and suitability for many applications over diverse geographic regions. The power law requires estimation of a wind shear coefficient,  $\alpha$ , linking the surface wind speed to winds at higher altitudes. Prior studies have mostly adopted simplified models for  $\alpha$ , ranging from a single constant, to a site-specific constant in time value. In this work we (i) develop a new model for  $\alpha$  which is able to capture hourly variability across a range of geographic/topographic features; (ii) quantify its improved skill compared to prior studies; and (iii) demonstrate implications for wind energy estimates over a large geographical area. To achieve this we use long-term high-resolution simulations by the Weather Research and Forecasting model, as well as met-mast and radiosonde observations of vertical profiles of wind speed and other atmospheric properties. The study focuses on Saudi Arabia, an emerging country with ambitious renewable energy plans, and is part of a bigger effort supported by the Saudi Arabian government to characterize wind energy resources over the country. Results from this study indicate that the proposed model outperforms prior formulations of  $\alpha$ , with a domain average reduction of the wind speed RMSE of 23–33%. Further, we show how these improved estimates impact assessments of wind energy potential and associated wind farm siting.

## 1. Introduction and motivation

The adverse impacts of fossil fuels on the environment and human health have led policymakers worldwide to adopt mitigating strategies and actively promote the replacement of these traditional energy sources with wind, solar, wave and tidal energy. Of these renewable energy sources, wind energy has seen one of the largest deployments [1]. Wind energy already represents a significant share of some countries' energy

portfolio, with Denmark at the forefront at 43% of the country's energy needs [2]. Many others have policies in place to increase its share over coming decades. For instance, several countries in Europe aim to generate 50% or more of their energy consumption from renewable energy sources, which include significant shares of wind power. Similarly in the United States, the electricity produced from renewable energy sources is projected to surpass coal by 2030 with wind energy being a major contributor [3].

\* Corresponding author at: Department of Civil and Environmental Engineering and Earth Sciences, 156 Fitzpatrick Hall, University of Notre Dame, Notre Dame, IN 46556, USA.

E-mail address: [pcrippa@nd.edu](mailto:pcrippa@nd.edu) (P. Crippa).

<https://doi.org/10.1016/j.apenergy.2021.117378>

Received 13 January 2021; Received in revised form 16 June 2021; Accepted 4 July 2021

Available online 28 July 2021

0306-2619/© 2021 Elsevier Ltd. All rights reserved.

Accurate wind resource assessments and identification of optimal turbine heights require a detailed characterization of vertical wind speed profiles [4,5], as wind power density varies with the cube of wind speed [6]. Challenges in characterizing wind speed at hub height relate to (i) difficulties in direct observations and (ii) limited high resolution model simulations. High resolution measurements of vertical wind speed profiles are limited in both space and time, and the sparsity of such observations remains a key challenge when seeking to identify optimal locations for wind resource harvesting on regional/national scales. Global model simulations and reanalysis products have also been used in the context of wind energy applications [6–9]; however, significant limitations remain due to spatial resolution, both horizontal (i.e., grid spacing ranging from 10 to 100 km) and vertical (i.e., generally coarser than 100 m). High resolution models, applied over a region of interest (i.e., regional models), represent an important tool to overcome the observation sparsity issue, although high computational cost often precludes applications over large areas for sufficiently long times [10]. There is thus a need to develop empirical statistical models capable of extrapolating wind speed observations from the surface to hub height. Fast and accurate models for vertical extrapolation allow for the assessment of suitability for wind farming over large areas in a more cost-effective fashion than measurement campaigns with wind towers, especially in regions with an emerging wind energy portfolio where such campaigns are not presently available.

Prior studies have sought to identify the relationship between surface and higher altitude winds through the development of different theoretical or empirical models, as well as artificial neural networks in different regions of the world, including over complex terrain, to improve estimates of wind speed at hub height [11,12]. A recent comprehensive review of more than 300 applications suggests that the power law is the most reliable and common model for extrapolation to typical turbine hub heights for wind energy applications [13]. A key parameter of the power law is the wind shear coefficient,  $\alpha$ , which is known to vary with atmospheric stability [14]. Most wind energy assessment studies have adopted  $\alpha = 1/7$  for simplicity [15], although the underlying assumption of near-neutral atmospheric stability on a flat terrain is rarely satisfied and thus may lead to erroneous extrapolated wind profiles [13]. Constant wind shear exponents have mostly been adopted in prior studies; these prove to be generally suitable for wind power estimates over long time scale assessments (e.g., yearly), while a temporally varying  $\alpha$  is needed to produce accurate wind power estimates at sub-daily level (e.g., hourly) [16]. A previous study [17] showed that the wind shear exponent exhibits significant diurnal and seasonal variability, based on 10-minute LiDAR observations at a site along Lake Erie, and found that a site-specific calibrated  $\alpha$ , able to account for such variations, significantly improved wind extrapolation. Similar results were found when the hourly averaged  $\alpha$  was used to extrapolate wind speeds at 60 m from measured 10 m wind speed at three sites in Serbia [18].

Logarithmic law extrapolation models have also been widely used in the past. In these models, based on Monin-Obukhov similarity theory (MOST), the wind speed at hub height is a function of the terrain's friction velocity and roughness length, with additional possible parameterization of the Monin-Obukhov length dependent on stability conditions [19]. While the logarithmic law provides good physical interpretability as it is based on MOST, the need for additional surface measurements presents limitations to large-scale applications where this information may not be available. Moreover, a comprehensive review [13] found that such logarithmic models, despite their theoretical foundation, have a more limited extrapolation range capability than power law models. Non-parametric extrapolation methods have also gained popularity thanks to recent advances in machine learning and artificial intelligence applications. Recent literature studies have tested the performance of artificial neural networks both on their own [12] and in hybrid models with genetic algorithms [11]. Random forest [20] and symbolic regression methods [21] have also been tested. While these

novel methods outperform the traditional power law, their non-parametric nature does not allow for a physical interpretation of the model results, which is of interest for the present project.

Numerical weather prediction models are powerful tools to investigate physical mechanisms dictating vertical wind speed profiles. For example, simulations of the Weather Research and Forecasting (WRF) model have revealed a sensitivity of vertical wind profiles to planetary boundary layer (PBL) parameterizations [22] and a dependence on stability [23]. WRF skills in capturing wind speeds are found to depend on stability conditions, with higher performance during unstable conditions [24], as well as on horizontal resolution and terrain complexity, which impact wind speed extrapolation at hub height [22].

Capturing temporal variability in  $\alpha$  is thus critical to provide accurate estimates of extrapolated wind speeds. Prior studies have mostly focused on specific sites/limited areas and on short time periods, thus preventing assessments of the full range of variability of  $\alpha$  during different atmospheric conditions and seasons, at national scales. This study aims to fill a current literature gap by proposing a general statistical, yet physically-interpretable, model for  $\alpha$ , capable of capturing its hourly variability, as opposed to prior studies that have adopted fixed  $\alpha$  in time and/or space. The proposed model focuses not just on the ability to extrapolate wind speeds, but also to capture different hourly variability by accounting for a temporally varying variance. Further, it provides robust conclusions as it is tested over a range of elevations, land use, and synoptic conditions over an entire country. With this work we address the following research objectives:

- 1) Develop a novel statistical temporal model which generalizes the power law, and is able to both extrapolate the wind speed at hub height and explain its variability with a sub-daily time varying (heteroskedastic) error.
- 2) Quantify the role of atmospheric stability and geographic features in dictating vertical wind speed profiles.
- 3) Quantify the impact of the proposed relationship for applications in wind energy estimates, and how they vary in space and time from hourly to seasonal scales.

To achieve these, we analyze vertical wind speed profiles from both direct observations and yearlong high resolution WRF simulations performed over Saudi Arabia, a large country characterized by a range of geographic features, including complex terrain and coastal regions, which to date has been only sparsely investigated [9,25], despite the recent growing commitment to install extensive wind farms [26,27]. The paper is structured as follows. Section 2 describes the observed and simulated data used for this analysis, Section 3 introduces the proposed model, and Section 4 presents key research results and implications for regional scale wind energy assessments. We include concluding remarks in Section 5. The R code for the model implementation for a sample dataset is available in a public GitHub repository at the following link: [github.com/Env-an-Stat-group/21.Crippa.APEN](https://github.com/Env-an-Stat-group/21.Crippa.APEN).

## 2. Observed and simulated meteorological data

A yearlong simulation with the WRF model [28] is performed and analyzed in this work. Specifically, the non-hydrostatic dynamics core of WRF is applied at 6 km horizontal resolution with 40 vertical layers up to 50 hPa over the entire Arabian Peninsula for all of 2016. The model defines pressure levels spaced by around 20 m, starting at around 10 m from the surface, to capture wind speed variability with height and near hub-height. Lateral boundary conditions for meteorology are imposed using the operational high-resolution European Centre for Medium-Range Weather Forecast model (HRES-ECMWF, [28,29]) every 6 h. Key model parameterizations adopted are summarized in Table S1 and are chosen to follow a setup that enables high skill in reproducing wind speed profile observations [4].

In this work we analyze hourly vertical wind speed profiles up to



To select  $K$ , Fig. S3a and S3b shows the map and domain average of change in a model selection metric, the Bayesian Information Criterion (BIC), for incremental increases in harmonics. Here it appears evident that the BIC of  $K = 5$  is lower than the one of all the previous  $K$  and very similar to the BIC of higher  $K$ , therefore  $K = 5$  harmonics results in a sufficiently flexible yet parsimonious model.

- 2) It allows for incorporation of other site-specific, time-varying covariates  $X_j(t)$ . In this work, we consider the case with no covariates and the case of  $P = 1$  covariate  $X_1(t) = \frac{d\theta}{dz}(t)$ , the vertical potential temperature gradient. While this information is in practice unavailable and not simple to estimate, we consider this model as an idealized case to investigate whether a proxy for atmospheric stability, expressed here as potential temperature gradient, provides additional predictive power. Geostrophic winds have also been found to improve wind forecasting [37], and could thus be included as a covariate as well when this information is available.
- 3) The model variance is assumed to also vary temporally with a periodic daily pattern controlled by  $\tilde{K}$  harmonics. The case for a heteroskedastic model can be gleaned from the considerable difference in variance of the wind shear coefficient between stable and unstable regimes as seen in Fig. 2, and is further shown in Fig. S4, where the variance of the wind detrended from model (2) is shown. In the same figure, we show the fitted daily variance pattern as a function of the number of harmonics, and throughout this work we fix  $\tilde{K} = 5$  as it captures the major patterns in daily variations. The addition of a heteroskedastic term for the error allows for considerably more flexible and realistic uncertainty quantification (and hence confidence intervals) for the extrapolated wind, as will be shown in Section 4.

The parameters that the model fits are therefore,

1. For  $\alpha(t)$ : a mean component  $\alpha_0$ , coefficients  $\beta_i$  and  $\beta_i'$  for the harmonics, and coefficients  $\gamma_j$  for the influence of any covariates (in our case, atmospheric stability).
2. For the error variance of  $\alpha(t)$ ,  $\sigma_t^2$ : a mean component  $r_0$  and coefficients  $r_i$  and  $r_i'$  for the harmonics.

All parameters are real unbounded values and unknown. We expect that the uncertainty in the estimation of  $\alpha(t)$  is going to be non-negligible compared to the magnitude of  $\sigma_t^2$ . A proper and complete study would require assumptions on the asymptotic distribution of the parameter estimates and a prohibitively expensive computational study, beyond the scope of this work. While the new model as introduced in (2) does not have an explicit spatial component, the spatial innovation in this study lies in the training of the model with high-resolution, gridded

wind speed data from WRF over the entire country of Saudi Arabia, spanning a range of terrain and climate conditions, allowing for practically continuous results of sub-daily  $\alpha(t)$  over the whole territory, which is unprecedented.

### 3.2. Inference and methodology for validation

Inference for model (2) is performed for every grid cell independently and in two steps. First, wind shear harmonics  $\alpha(t)$  are estimated with weighted least squares, with weights inversely proportional to the hourly variance of the wind speed aloft. Once the trend is estimated, the residuals are computed, and the harmonics for  $\sigma_t^2$  are estimated. The process could in principle be performed in one step, either with full or profile maximum likelihood, but this would have resulted in a more computationally burdensome inference, with little to no additional gain in predictive power. The normality assumption underpinning the model (2) can be directly assessed through a quantile-quantile (QQ) plot of the model's standardized residuals for two representative locations in Fig. S5.

We fit the model to the WRF wind speeds at level 1 and 6 and their respective heights  $z_1$  and  $z_2$  (approximately 10 and 110 m, respectively) using the hourly data for each day. Specifically, 80% of the days of the simulated year are randomly selected as the training set, while the remaining 20% of the days are used as test set, i.e., to evaluate the skills of proposed model (2). The comparison against the standard power law (1) is performed similarly, focusing on the same sampled days for training and testing. Model performance is quantified in terms of Root Mean Squared Error (RMSE), Mean Fractional Bias (MFB) and Mean Absolute Error (MAE). In order to produce robust statistics of model performance, we perform an ensemble of 100 iterations of the random splitting of training and testing sets for our model evaluation, and compute the ensemble mean of the model performance metrics of interest. In order to assess the model's ability to quantify the uncertainty, we compute the empirical coverage, i.e., the number of elements in the testing set whose true level 6 wind speed falls inside the confidence intervals (computed from the training set) for the same height. The model is able to correctly quantify the uncertainty to the extent that the empirical coverage is approximately equal to the nominal confidence level of the interval, in this case 95%.

Our model evaluation also includes comparison against a stability-specific wind shear coefficient which varies as a function of atmospheric stability based on the Pasquill definition [38]. We first use WRF simulated wind speed at 10 m and varying patterns for the daytime insolation, to classify each hour of 2016 and each grid cell into one of the Pasquill stability categories, and also compute the  $\alpha$  value between WRF levels 1 and 6 using the power law fit (1) at each grid cell and hour of 2016. We then compute the domain average  $\alpha$  for each different Pasquill

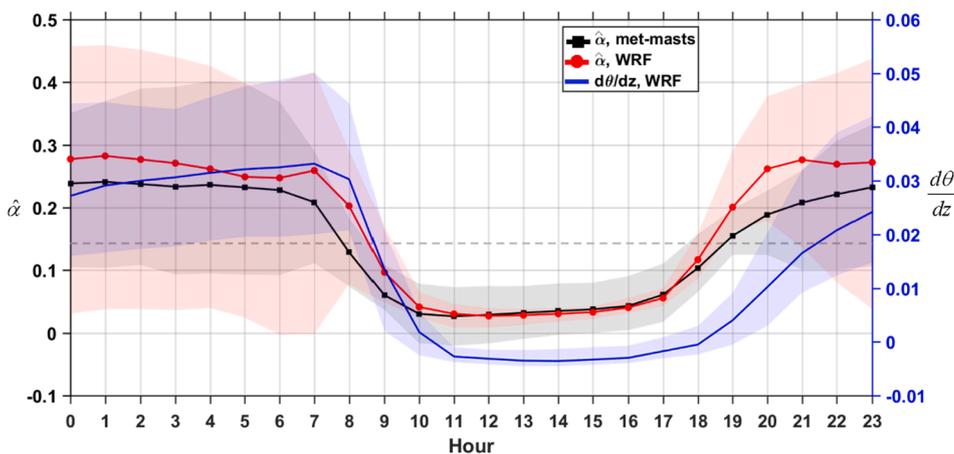


Fig. 2. Diurnal variability of the estimated wind shear exponent ( $\hat{\alpha}$ , left axis) and atmospheric stability (expressed as  $\frac{d\theta}{dz}$  [ $^{\circ}\text{C m}^{-1}$ ], right axis) from KA.CARE observations and WRF simulations at the same sites.  $\hat{\alpha}$  is obtained from the power law in equation (1) to hourly observed and modeled wind speeds at 40, 60, 80, and 100 m averaged over the year 2016.  $\frac{d\theta}{dz}$  is shown only from the WRF output as the sites do not measure vertical profiles of temperature. Solid lines represent averages and the shadings indicate the full variability (min-max) across the ten measurement sites. The dashed grey line indicates the reference value of  $\alpha = 1/7$ .

stability category (Table S4) based on the  $\alpha$  values derived from the power law fit at each grid cell and hour. These Pasquill-dependent average  $\alpha$  will be used to extrapolate the wind speed from 10 m to 110 m. Insolation patterns are divided into two seasons, with the summer season defined from April to September and the winter season for the remaining 6 months of the year, based on the sunrise and sunset times in Saudi Arabia during different months. Specifically, we define the two hours after sunrise as “Slight”, 2–3 h after that as “Moderate”, 4–5 h in the middle of daytime as “Strong” insolation, and then back to “Moderate” and “Slight” until sunset. Due to the typical low cloudiness of the region [39], the daytime insolation pattern remains constant within each season and nighttime hours are assumed to be mostly clear (cloud cover  $\leq 3/8$ ).

#### 4. Results and discussion

In this section, we assess the spatio-temporal variability of the wind shear coefficient for the WRF dataset and compare it with met-mast observations, as well as its empirical relationship with atmospheric stability validated against radiosonde data. Our proposed model for wind speed extrapolation with height is then evaluated using WRF simulated data and its skills are compared to approaches widely used in the literature and industrial practice: a model with constant  $\alpha = 1/7$ ; a model with site-specific but constant-in-time  $\alpha$ ; and a model with time-variable  $\alpha$  based on Pasquill stability. In order to distinguish between the parameters for the power laws in (1) and (2) and their estimates, we denote by  $\hat{\alpha}$  the wind shear coefficient inferred directly from the WRF data using all levels up to  $\sim 110$  m (i.e., levels 1 through 6).

##### 4.1. Diurnal variability of the wind shear coefficient and atmospheric stability

Here we focus on the standard power law for every hour, and characterize the diurnal variability exhibited by the wind shear exponent in both observations and simulated wind speed profiles. Both the WRF data and the met-masts indicate a strong diurnal pattern at KA.CARE sites, which suggests an important role played by daily atmospheric dynamics in dictating vertical wind speed profiles. Highest wind shear exponents (i.e.,  $\hat{\alpha} > 0.2$ ) occur during more stable atmospheric conditions, mostly during 7 pm to 7 am (Fig. 2). Some locations, including Al-Jouf in the North and Sharurah in the South (see Fig. 2 and Table S2 for their location) exhibit a more pronounced diurnal cycle with  $\hat{\alpha}$  oscillating between 0.35–0.40 and 0.03–0.04 during the most stable and unstable conditions, respectively.

The WRF simulation reproduces both the diurnal cycle and the range of the wind shear coefficient from the observations well, although it slightly overestimates the spread among sites. Sites closer to the coast present lower  $\hat{\alpha}$  (Table S2), possibly as a result of the role of low level jets [40,41] associated with land and sea breezes which weaken the power law relationship between surface and elevated winds [42,43]. We observe more variability in the temperature gradient during nighttime hours, corresponding to stable, well-stratified atmospheric layers, while higher turbulence in the daytime results in a very well mixed layer, where the potential temperature is almost constant with height, with a gradient close to 0. Early mornings (late afternoons) see the transition between stable and unstable (unstable and stable for the transition in the evening) regimes as indicated by lower/higher temperature gradients compared to adjacent hours (Fig. 2 and Fig. S6). The exact occurrence of these transition hours shifts throughout the day by a couple of hours depending on the season and times of sunrise and sunset. This analysis thus suggests that the assumption of near-neutral stability and associated  $\alpha = 1/7$  is rarely observed over the Arabian Peninsula, hence highlighting the need for more flexible approaches.

As the KA.CARE sites do not record vertical temperature profiles, we quantify WRF skills in capturing potential temperature gradients from

the surface up to 100 m by comparison with data from eight radiosonde sites. Our statistical analysis indicates high skill of WRF with low values of RMSE = 0.03, MFB = 0.09 and MAE = 0.02, [44] and a correlation coefficient 0.4 when averaged among sites during nighttime (00 UTC or 3 am local time, Table S3). Similar results are found during daytime (12 UTC, 3 pm local time), although lower correlations are observed (Table S3). During such hours of strong convection and mixing, the potential temperature gradients are very close to zero and the variations between the observed and simulated values, responsible for the low correlation coefficients, are negligible compared to the diurnal variability in  $\frac{d\theta}{dz}$ , that is indeed well captured by the model.

To further explore the temporal patterns of  $\hat{\alpha}$  and its relationship with atmospheric stability, we compare its map with that of  $\frac{d\theta}{dz}$  calculated using the potential temperatures at layer 1 ( $\sim 10$  m) and layer 6 ( $\sim 110$  m). The wind shear exponent (Fig. S7) follows a similar diurnal and spatial pattern to the temperature gradient. As expected, yearly averaged values show that the daytime hours are mostly unstable, and the nighttime hours are stable. The unstable hours, with deep atmospheric mixing during the daytime, show  $\hat{\alpha}$  close to 0, meaning that wind speed is essentially constant with height. The stable hours of the night show higher  $\hat{\alpha}$  in the places where higher  $\frac{d\theta}{dz}$  are also observed, further suggesting the existence of a strong association between the two.

##### 4.2. Spatial variability of the wind shear coefficient

The range of variability in  $\hat{\alpha}$  during a stable hour (i.e., 3 am) is shown in Fig. 3. While large portions of the domain present low/negative values of  $\hat{\alpha}$  (but with a low coefficient of determination,  $R^2$ , see Fig. S8) when looking at the 2.5th percentile (domain average  $\hat{\alpha} = -0.22$  and  $R^2 = 0.11$ ), a good fit is instead found for the 50th percentile (domain average  $\hat{\alpha} = 0.34$  and  $R^2 = 0.95$ ) and the 97.5th percentile (domain average  $\hat{\alpha} = 0.78$  and  $R^2 \sim 1$ ). This indicates that for specific days/events, more complex vertical wind speed structures/dynamics occur; thus the power law relationship is not always an appropriate model to describe such profiles. Small portions of the domain present extremely low or extremely high values of the wind exponent (above  $|1|$ ) for the 2.5th and 97.5th percentiles respectively, which would result in unrealistic variations of the wind speed over the lowest 100 m of the atmosphere. Fig. S9 shows that such cases, possibly resulting from terrain complexity or a failure of the power-law model to capture the true profile, are generally outliers across the domain while the great majority of the territory does exhibit physical behavior.

Conversely, very unstable hours do not show large variability in the distribution of  $\hat{\alpha}$  as the domain averaged 2.5th and 97.5th percentiles are  $-0.0097$  ( $R^2 = 0.40$ ) and  $0.0708$  ( $R^2 \sim 1$ ), respectively (Fig. S10). This may be at least partially attributed to fully mixed wind speed profiles and constant temperature gradient. An example of the hourly variability through the analyzed year in the goodness of fit of the power law in model (1) at two KA.CARE sites (Jeddah and Hafar-Al-Batin) is presented in Fig. S11. As mentioned earlier, the power law fit (expressed as  $R^2$ ) is generally very good (median  $R^2 \sim 1$ ) for very stable and very unstable hours, while variable performance is found for transition hours. Further, Fig. S11 also highlights the large intra-daily variability in the  $R^2$ , which implies that the power law is not always the most suitable model to capture the vertical wind speed profile.

Over the analyzed year and throughout the simulated domain, the atmosphere is unstable/neutral for  $\sim 40\%$  of the time and stable for the remaining time. The coastal regions show a lower percentage of unstable hours, particularly along the Persian Gulf coast. Grid cells with high Terrain Ruggedness Index (TRI, Fig. 1) also present the weakest fits of  $\hat{\alpha}$  (Fig. 3, Fig. S10). This indicates terrain influence on the wind speed profile, attributable to turbulence by surface friction. The negative values of  $\hat{\alpha}$  are observed most frequently along the places with highest TRI (Fig. 3b) and do not negatively impact wind energy applications in the region, as prior work has deemed these sites unsuitable for wind

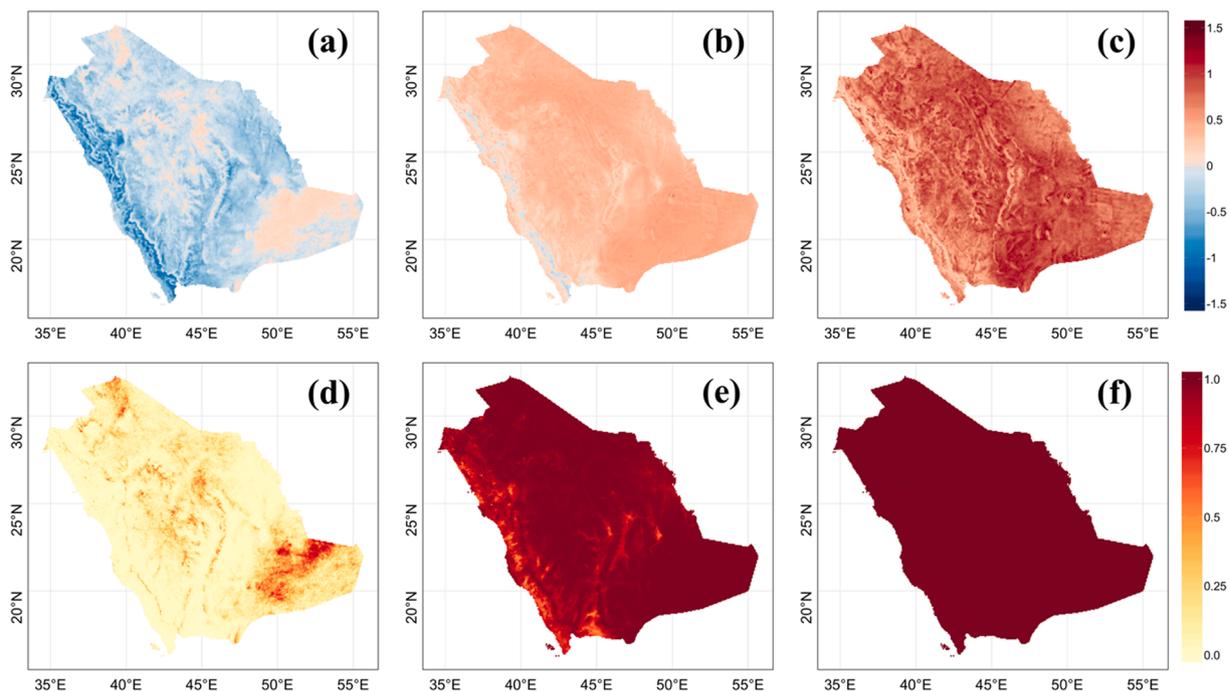


Fig. 3. Range of variability in the wind shear exponent and coefficient of determination ( $R^2$ ) at 3 am local solar time (stable conditions). The top row indicates the (a) 2.5th, (b) 50th, and (c) 97.5th percentile of  $\hat{\alpha}$  for each grid cell, while the bottom row indicates the (d) 2.5th, (e) 50th, and (f) 97.5th percentile for the corresponding  $R^2$  over all days in 2016.

farm operations [4].

### 4.3. Model fit, evaluation and inter-comparison

Here we fit our proposed new model as described in (2) for capturing a heteroskedastic wind with time-varying  $\alpha$ , with the final aim of showing its superior extrapolation performance against the standard model in (1).

We quantify model skills in predicting wind speed at  $\sim 110$  m based on the RMSE, MFB and MAE computed for all points belonging to specific ranges of TRI and distance from the coast (Table S5). The RMSE ranges between 1.5 and 1.9, while both the MFB and MAE are relatively small independently of geographic location. The very low MFB statistics presented are the result of compensating errors occurring at different hours of the day (see example in Fig. S12), with hours in the middle of the day, corresponding to unstable boundary layer, presenting the

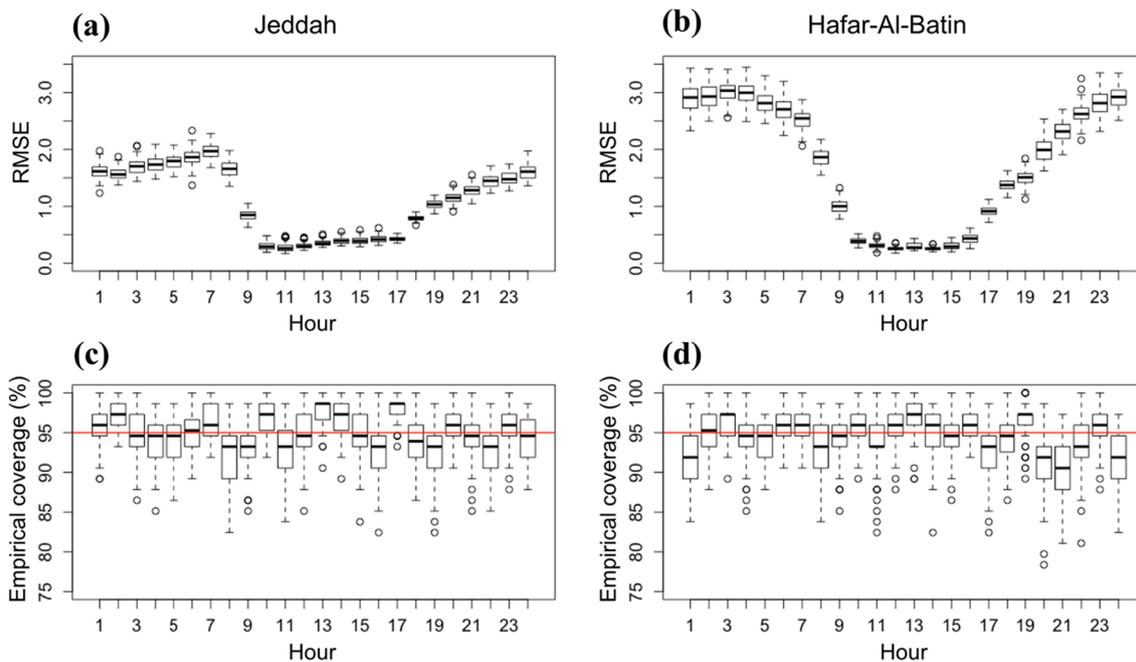


Fig. 4. Diurnal variability of the RMSE between our model and the actual WRF wind speed at level 6 (a-b) and the associated empirical coverage percentage (c-d) for two representative sites: Jeddah (a,c) and Hafar-Al-Batin (b,d). The spread in the boxplots derives from the 100 random samples of 20% of the data used for evaluation purposes, as described in Section 3.2.

lowest and least variable values. A similar pattern is observed for the RMSE in the first row of Fig. 4, albeit with larger values. The coefficient of determination  $R^2$  for the power law (1) is reported in Table S5, and it indicates that coastal locations and areas of complex terrain are characterized by lower values. This can be explained by the occurrence of low-level jets from the ocean breeze and more complex flow over elevated peaks that are thus not consistent with the expectation of monotonic increase of wind speed with height. Regardless of the irregularity of the profile, our model still captures the winds aloft well, as indicated by the low MFB and MAE. As land/sea breezes are characterized by relatively consistent temporal patterns (i.e., they generally occur at similar hours through the day at each location), the time component of our model appears to be able to account for those irregular wind vertical profiles (Table S5). Coastal locations may also see less inter-seasonal variability due to ocean proximity. This is further confirmed by a sample coastal location, Jeddah, which shows less variability in the hourly variance of the wind speed at WRF level 6 than the inland site of Hafar-Al-Batin (Fig. S4).

The model is also able to correctly assess the uncertainty of the extrapolation. The second row of Fig. 4 shows the empirical coverage of a nominal 95% confidence interval throughout all 24 h for both Jeddah and Hafar-Al-Batin, with every boxplot representing the variability throughout different days selected by the 100 random sampling iterations as described in Section 3.2. Overall, the empirical coverage is very close to the nominal confidence level, and the median of each boxplot is aligned with 95%.

We also compared the performance of our proposed model in (2) (henceforth  $\alpha_{ST}$ ) against the performance of the standard power law in (1) using (i) a constant value of  $\alpha = 1/7$  (henceforth  $\alpha_C$ ), valid for neutral stability and flat terrain; (ii) a site-specific wind shear coefficient constant in time (henceforth  $\alpha_S$ ); and (iii) a stability-varying wind shear coefficient based on the site's hourly Pasquill stability classification (henceforth  $\alpha_P$ ), as detailed in Section 3.2. We computed RMSE, MFB and MAE for model performance with respect to the WRF simulated level 6 wind speeds, as reported in Table 1. As one might expect  $\alpha_C$  performs the worst and shows the largest RMSE and MFB, whereas site-specific coefficients,  $\alpha_S$  and  $\alpha_P$ , provide slightly improved performance. Our proposed model instead has median RMSE and MAE which are 40% smaller than  $\alpha_C$ , and median MFB which is almost an order of magnitude smaller. These results highlight that accounting for the temporal variability in  $\alpha$  is crucial to properly extrapolate wind speed at different altitudes and assuming a power law constant in time leads to significant biases that may impact wind energy resources assessments (see more discussion in Section 4.4).

We also test an idealized model (2) with  $P = 1$  covariate  $X_1(t) = \frac{d\theta}{dz}(t)$ , thereby incorporating temperature gradient as additional information about atmospheric stability (henceforth  $\alpha_{ST-S}$ ). As discussed in Section 4.1, the temporal evolution of atmospheric stability and  $\alpha$  are highly correlated, and the temperature gradient is the driver of atmospheric dynamics and turbulence features which inevitably impact the wind vertical profiles. Model  $\alpha_{ST-S}$  outperforms  $\alpha_{ST}$ , the same model without temperature gradient, although this difference in performance

**Table 1**

Performance metrics of different models in extrapolating wind speed at 110 m (level 6) from 10 m (level 1) compared to the WRF simulated wind speed at level 6. The median value in space is reported for each metric (the 2.5th and 97.5th percentiles are in parentheses). The performance metrics are computed based on 100 samples of 20% of randomly extracted days.

Model	RMSE	MFB $\times 10^2$	MAE
$\alpha_C$	2.86 (1.81; 3.59)	-16.09 (-27.73; 22.07)	2.30 (1.51; 2.90)
$\alpha_S$	2.82 (1.68; 3.62)	0.35 (-0.8; 1.20)	2.42 (1.26; 3.16)
$\alpha_P$	2.63 (1.68; 3.26)	-15.95 (-26.60; 21.05)	1.96 (1.31; 2.44)
$\alpha_{ST-S}$	1.70 (1.24; 2.16)	-1.18 (-6.41; 9.10)	1.13 (0.85; 1.42)
$\alpha_{ST}$	1.84 (1.30; 2.28)	-2.52 (-8.59; 9.11)	1.24 (0.89; 1.50)

is small compared to the improvement from previous models  $\alpha_S$ ,  $\alpha_P$  and  $\alpha_C$ . While this provides compelling evidence for the use of the temperature gradient, in a real-life scenario such information needs to be either measured at hub height, hence effectively being as challenging to measure as the model output itself, or estimated from the surface with additional uncertainty. For these reasons, our energy assessment application will focus on the more practical case of  $\alpha_{ST}$ .

#### 4.4. Implications for wind energy resource assessments

Fig. 5 (panels a-d) shows the spatial patterns of the RMSE of the extrapolated wind speeds at  $\sim 110$  m compared to the original WRF output for the four models:  $\alpha_C$ ,  $\alpha_S$ , and  $\alpha_P$ , as well as our time-varying heteroskedastic model  $\alpha_{ST}$ . The traditional approximation  $\alpha_C$  for neutral conditions shows a systematic, large, positive bias (i.e., the neutral approximation greatly underpredicts the wind speeds around 110 m), over most of the domain (Fig. 5a), a condition only marginally improved by  $\alpha_S$  and  $\alpha_P$  (Fig. 5b and c), while our  $\alpha_{ST}$  model performs better for inland and lower-elevation areas; the degree of improvement is smaller for elevated regions and coastal areas (Fig. 5d). Models  $\alpha_C$ ,  $\alpha_S$ , and  $\alpha_P$  present similar spatial patterns, with largest errors in the south-eastern and in portions of the central-east regions, both of which are extensively characterized by low elevation and relatively flat terrain. Despite the relatively flat terrain, for which we would expect simple models for  $\alpha$  to work relatively well, this result instead suggests that these areas experience major temporal variability in the wind patterns. We hypothesize that the large values may be attributable to the strong seasonal variability in wind patterns (both speed and direction). For example, during summer (and occasionally in winter) the Shamal trade winds are responsible for a northwesterly flow which mostly impacts the north-central regions of the peninsula, while a southeasterly flow, associated with the onset of the Indian monsoon, occurs in spring and primarily impacts wind fields in the south-east of the Peninsula [9]. Our model outperforms the other approaches systematically over the entire domain, with a domain average reduction of the RMSE of 33%, 27% and 23% compared to  $\alpha_C$ ,  $\alpha_P$ , and  $\alpha_S$ , respectively.

Besides its superior predictive ability, the temporal heteroskedasticity of  $\alpha_{ST}$  also allows for the correct quantification of uncertainty in the daily variation of wind, as previously demonstrated in Fig. 4, where  $\alpha_C$ ,  $\alpha_S$ , and  $\alpha_P$  would have produced overly small or large confidence intervals, depending on the hour of the day. Fig. 5 (panels e-h) show the map of the empirical coverage for all four methods. While  $\alpha_C$  and  $\alpha_P$  result in an overall underestimation of the uncertainty (Fig. 5e and Fig. 5g), noticeably in the south-west areas,  $\alpha_S$  results in severe overestimation, with the vast majority of the domain showing overly large confidence intervals with 100% empirical coverage (Fig. 5f). Our  $\alpha_{ST}$  model is instead able to capture the correct extent of the uncertainty for the majority of the country, with exception of the complex and elevated terrain in the western part of the country (Fig. 5h). This lack of correct uncertainty quantification is attributable to the same causes of the relatively poor fit of the power law for sites with high TRI, as highlighted in Table S5. Finally, we used the four models presented here to extrapolate winds at WRF level 9 ( $\sim 200$  m), to explore model performance at higher altitudes than the ones available for training. Analogously to the results in Fig. 5,  $\alpha_{ST}$  outperforms the models  $\alpha_C$ ,  $\alpha_P$  and  $\alpha_S$  when extrapolating to 200 m (Fig. S13).

The improved predictive abilities of  $\alpha_{ST}$  on wind speed are expected to be reflected, to a larger extent, on wind energy potential quantification, as biases in the wind speed scale up with a cubic relationship when considering the energy produced. Fig. 6 presents the difference in annual capacity factors between those derived using different extrapolation models and the ones derived using the WRF output (Fig. S14). Specifically, for each grid cell of our domain, we compute the annual capacity factors, defined as the annual average of the hourly wind power generated divided by the turbine's rated peak power. We do not account for other losses (e.g. due to wake effect or decay in efficiency during

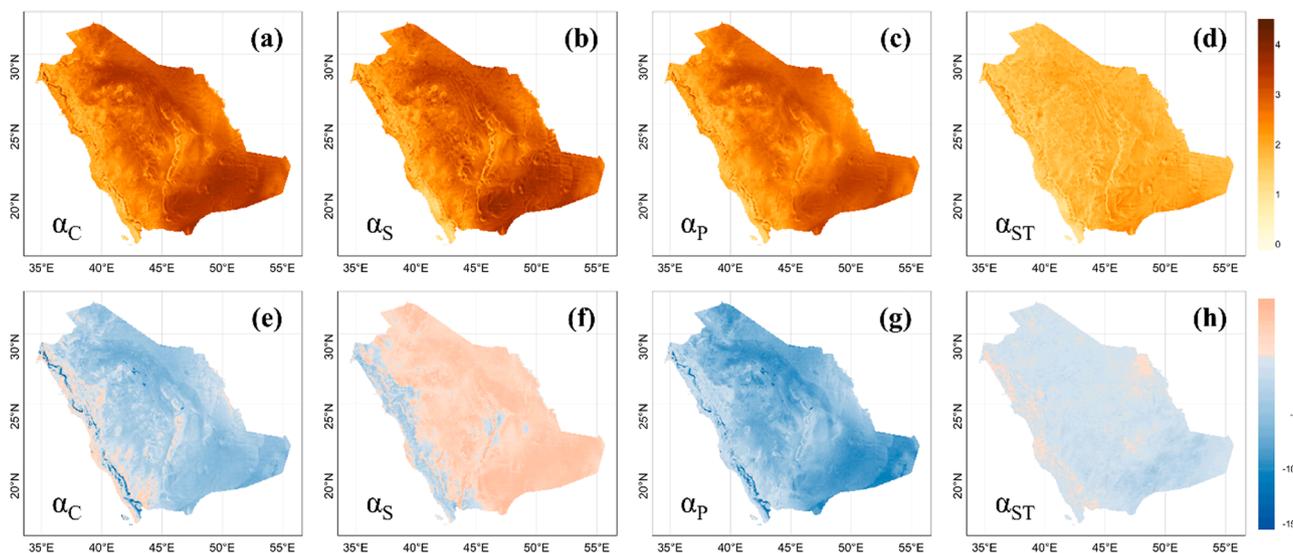


Fig. 5. First row: RMSE computed on 20% of data and 100 iterations for winds extrapolated to 110 m using (a)  $\alpha_C$ , (b)  $\alpha_S$ , (c)  $\alpha_P$ , and (d)  $\alpha_{ST}$  relative to the original WRF output. The second row shows the associated coverage difference.

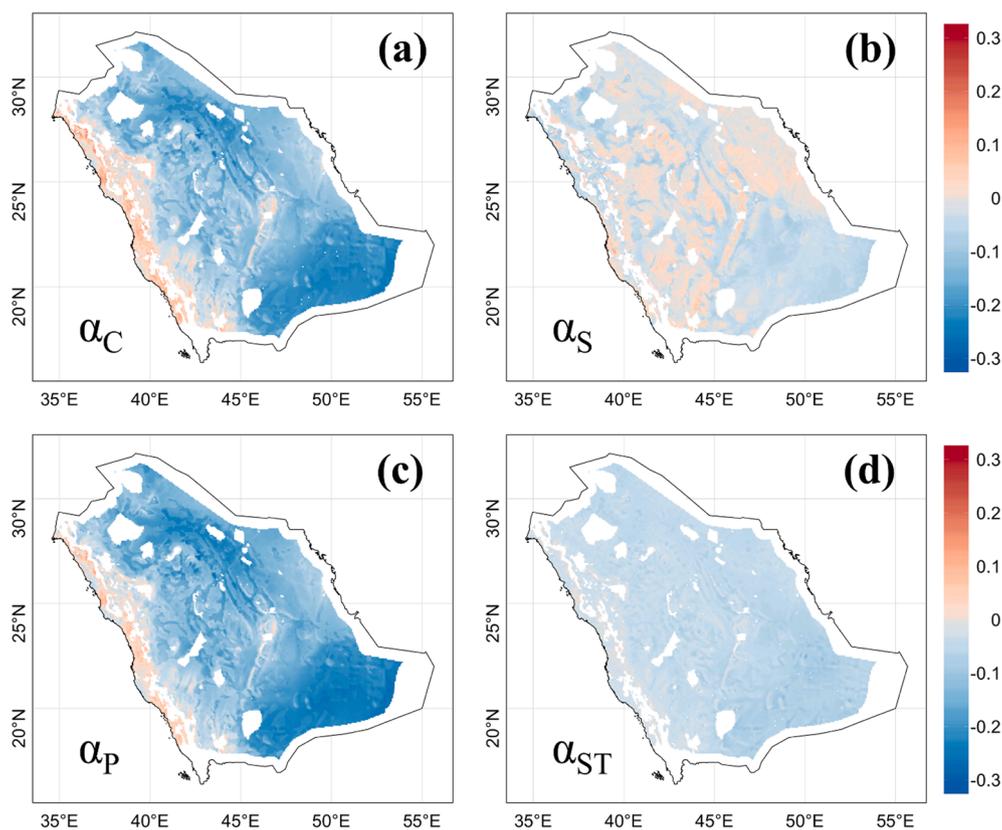


Fig. 6. Difference between capacity factors computed based on (a)  $\alpha_C$  model, (b)  $\alpha_S$  model, (c)  $\alpha_P$  model (d)  $\alpha_{ST}$  model and the reference capacity factors computed based on the WRF output. The white areas are excluded from the analysis as not suitable for wind farm location (see [4] for further details).

turbines’ lifetime) which may occur during real operations, thus the presented capacity factors may be larger than the operational ones. In this analysis, we consider a range of commercially-available wind turbines and we assume optimal turbine type and height (and thus the associated power curve) for each location, as identified in prior work based on the same WRF simulations as the ones presented here [4]. The extrapolation to hub height is performed with  $\alpha_C$ ,  $\alpha_S$ ,  $\alpha_P$ , and  $\alpha_{ST}$  (Fig. 6 panels a, b, c, and d, respectively), while the WRF output is derived at the exact turbine hub height via linear interpolation between the two

layers that contain the hub height. It is readily apparent how our model systematically underestimates the actual annual capacity factors and displays a median spatial absolute bias of 0.05. Conversely, the traditional extrapolation methods present more complex spatial patterns of the bias, both over- and under-estimating the capacity factor depending on the region (Fig. 6) and showing more spread among sites, with interquartile range for  $\alpha_C$  (0.1),  $\alpha_P$  (0.093), and  $\alpha_S$  (0.047) being much larger than that of  $\alpha_{ST}$  (0.022, Table S6 and Fig. S15). Quantifying such biases is key as they may lead to significant economic losses from the

wind energy providers [9] and grid operation management.

## 5. Conclusions

This work introduces a new method to extrapolate wind speeds from the surface to higher altitudes, a topic particularly relevant for wind energy applications, as an accurate characterization of the wind speed at hub height is crucial for both accurate wind energy assessments and improved wind power forecasts needed by wind farm operators. The proposed model is flexible enough to capture the observed strong hourly variability of the wind shear coefficient  $\alpha$ , which can be explained with boundary layer dynamics and atmospheric stability and generalizes standard literature approaches based on a power law method. We validate our model using WRF hourly output, applied at 6 km resolution over Saudi Arabia for an entire calendar year. The large spatial domain and relatively large temporal scale enable testing under different seasonal conditions and a variety of geographical/terrain features including coastlines, desert areas, complex terrain, and plateaus. Previous studies using both modelled and reanalysis wind data have found a low inter-annual variability of wind speeds in the Arabian Peninsula, meaning that the applicability and conclusions of our study may be extended beyond the year of data used for model development. Further, we perform a comparative analysis between our model and prior literature studies. Our model is found to outperform the power law under the traditional assumption of  $\alpha = 1/7$ , which is valid for near-neutral atmospheric conditions on open flat terrain, conditions that are rarely satisfied throughout the day and over different geographical regions. Our model also presents higher skills than a power law with site specific  $\alpha$  (either constant through time or varying based on atmospheric stability) with significant implications for the predictive skills of wind capacity factors. Indeed, we estimate that our model reduces the absolute bias in the estimated capacity factor on average by 58% and 64% over the entire country, compared to the assumptions of  $\alpha = 1/7$  and a Pasquill-stability varying  $\alpha$ . While the domain averaged mean absolute bias of our model is similar to that of a spatially varying yet time invariant  $\alpha$ , our new approach shows a systematic negative bias and a smaller spread of error, facilitating country-scale bias correction. The proposed work is general and applicable to any country where sub-daily information on the vertical wind speed profiles is available for a certain time period. Our new model's improved accuracy in hub-height extrapolation also offers an opportunity to revisit the implications for future wind energy potential in Saudi Arabia discussed in prior studies that adopted the neutral assumption for the power law.

The method presented in this work is arguably more flexible than traditional methodologies, yet it is still reliant on the power law. Evidence suggests that the use of a predefined functional form may be limiting, and a fully non-parametric specification with neural networks may result in higher predictive skills. Despite these improvements in point prediction, wind extrapolation from the surface to hub height is bound to have high uncertainty by the nature of the problem, so further improvements should always stress the importance of correctly assessing model uncertainty.

Future work will be focused on quantifying the model skills in other geographical regions and climate to provide a broader and more general assessment of the range of improvements that this approach brings against the standard power law. For countries where limited observations and no simulations of wind speed profiles are readily available, it is still possible to provide a coarse assessment of our model with global data products, such as reanalysis data from the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) which is characterized by 6-hourly data and 50 km resolution global coverage and possibly integrated with high resolution vertical profile information (such as from radiosondes).

## CRedit authorship contribution statement

**Paola Crippa:** Conceptualization, Supervision, Investigation, Writing – original draft. **Mariana Alifa:** Formal analysis, Data curation, Writing – review & editing. **Diogo Bolster:** Conceptualization, Supervision, Writing – original draft. **Marc G. Genton:** Supervision, Writing – original draft. **Stefano Castruccio:** Conceptualization, Supervision, Methodology, Writing – original draft.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This publication is based upon work supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under Award No: OSR-CRG 7 2018-3742.2. The authors thank the King Abdullah City for Atomic and Renewable Energy (K.A. CARE) for providing the wind speed observational data. HRES-ECMWF operational analysis data were downloaded from the ECMWF data portal (<https://www.ecmwf.int/en/forecasts/datasets/set-i>) through the KAUST ECMWF licence.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.apenergy.2021.117378>.

## References

- [1] REN21, ed. Renewables 2020: Global Status Report. Renewable Energy Policy Network for the 21st Century; Paris; 2020. p. 302.
- [2] WWEA. Wind power capacity reaches 539 GW, 52.6 GW added in 2017. World Wind Energy Association; 2018.
- [3] EIA. Annual Energy Outlook 2017 with projections to 2050, U.S.E.I. Administration, Editor. 2017, U.S. Energy Information Administration; 2017. p. 64.
- [4] Giani P, et al. Closing the gap between wind energy targets and implementation for emerging countries. *Appl Energy* 2020;269:115085.
- [5] Lantz EJ et al. Increasing Wind Turbine Tower Heights: Opportunities and Challenges. United States; 2019.
- [6] Lu X, McElroy MB, Kiviluoma J. Global potential for wind-generated electricity. *Proc Natl Acad Sci* 2009;106(27):10933–8.
- [7] Jeong J, Castruccio S, Crippa P, Genton MG. Reducing storage of global wind ensembles with stochastic generators. *Ann Appl Statist* 2018;12(1). <https://doi.org/10.1214/17-AOAS110510.1214/17-AOAS1105SUPP>.
- [8] Miller LM, Kleidon A. Wind speed reductions by large-scale wind turbine deployments lower turbine efficiencies and set low generation limits. *PNAS* 2016; 113(48):13570–5.
- [9] Tagle F, Castruccio S, Crippa P, Genton MG. A Non-Gaussian Spatio-Temporal Model for Daily Wind Speeds Based on a Multi-Variate Skew-t Distribution. *J Time Ser Anal* 2019;40(3):312–26.
- [10] Gutowski WJ, Ullrich PA, Hall A, Leung LR, O'Brien TA, Patricola CM, et al. The Ongoing Need for High-Resolution Regional Climate Models: Process Understanding and Stakeholder Information. *Bull Am Meteorol Soc* 2020;101(5): E664–83.
- [11] Islam MS, Mohandes M, Rehman S. Vertical extrapolation of wind speed using artificial neural network hybrid system. *Neural Comput Appl* 2017;28(8):2351–61.
- [12] Vassallo D, Krishnamurthy R, Fernando HJS. Decreasing Wind Speed Extrapolation Error via Domain-Specific Feature Extraction and Selection. *Wind Energ Sci Discuss* 2019;2019:1–17.
- [13] Gualtieri G. A comprehensive review on wind resource extrapolation models applied in wind energy. *Renew Sustain Energy Rev* 2019;102:215–33.
- [14] Gualtieri G. Atmospheric stability varying wind shear coefficients to improve wind resource extrapolation: A temporal analysis. *Renewable Energy* 2016;87:376–90.
- [15] Peterson EW, Hennessey JP. On the Use of Power Laws for Estimates of Wind Power Potential. *J Appl Meteorol* 1978;17(3):390–4.
- [16] Kubik ML, Coker PJ, Barlow JF, Hunt C. A study into the accuracy of using meteorological wind data to estimate turbine generation output. *Renewable Energy* 2013;51:153–8.
- [17] Li J, Wang X, Yu X. Use of spatio-temporal calibrated wind shear model to improve accuracy of wind resource assessment. *Appl Energy* 2018;213:469–85.

- [18] Durišić Ž, Mikulović J. A model for vertical wind speed data extrapolation for improving wind resource assessment using WAsP. *Renewable Energy* 2012;41:407–11.
- [19] Panofsky HA. Determination of stress from wind and temperature measurements. *Q J R Meteorol Soc* 1963;89(379):85–94.
- [20] Bodini N, Optis M. The importance of round-robin validation when assessing machine-learning-based vertical extrapolation of wind speeds. *Wind Energy Sci* 2020;5(2):489–501.
- [21] Valsaraj P, et al. Symbolic regression-based improved method for wind speed extrapolation from lower to higher altitudes for wind energy applications. *Appl Energy* 2020;260:114270.
- [22] Stuta D, West G, Stull R. WRF Hub-Height Wind Forecast Sensitivity to PBL Scheme, Grid Length, and Initial Condition Choice in Complex Terrain. *Weather Forecasting* 2017;32(2):493–509.
- [23] Draxl C, Hahmann AN, Peña A, Giebel G. Evaluating winds and vertical wind shear from Weather Research and Forecasting model forecasts using seven planetary boundary layer schemes. *Wind Energy* 2014;17(1):39–55.
- [24] Fernández-González S, Martín ML, García-Ortega E, Merino A, Lorenzana J, Sánchez JL, et al. Sensitivity Analysis of the WRF Model: Wind-Resource Assessment for Complex Terrain. *J Appl Meteorol Climatol* 2018;57(3):733–53.
- [25] Chen W, Castruccio S, Genton MG, Crippa P. Current and Future Estimates of Wind Energy Potential Over Saudi Arabia. *J Geophys Res: Atmos* 2018;123(12):6443–59.
- [26] NREP. Saudi Arabia Renewable Energy Targets and Long Term Visibility. National Renewable Energy Program; 2018.
- [27] Nurunnabi M. Transformation from an Oil-based Economy to a Knowledge-based Economy in Saudi Arabia: the Direction of Saudi Vision 2030. *J Knowledge Econ* 2017;8(2):536–64.
- [28] Skamarock WC et al. A Description of the Advanced Research WRF Version 3 2008; 27:3–27.
- [29] ECMWF. ECMWF IFS CY41r2 High-Resolution Operational Forecasts. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory: Boulder, CO; 2016.
- [30] Stull RB. *An Introduction to Boundary Layer Meteorology*, vol. 13. Dordrecht: Springer; 1988.
- [31] MEASNET. Evaluation of Site-Specific Wind Conditions - Measuring Network of Wind Energy Institutes (MEASNET). Madrid, Spain; 2009.
- [32] Counihan J. Adiabatic atmospheric boundary layers: A review and analysis of data from the period 1880–1972. *Atmos Environ* 1975;9(10):871–905.
- [33] DeMarrais GA. Wind-speed profiles at Brookhaven National Laboratory. *J Meteorol* 1959;16(2):181–90.
- [34] Davenport AG. *Rationale for Determining Design Wind Velocities*; 1960.
- [35] Sedghi M, Boroushaki M, Hannani SK. Modeling changes in wind speed with height in Iran's cities and its impact on the energy production. *J Renewable Sustainable Energy* 2015;7(2):023132. <https://doi.org/10.1063/1.4919083>.
- [36] Panofsky HA. Atmospheric turbulence: models and methods for engineering applications / Hans A. Panofsky, John A. Dutton, ed. J.A. Dutton. New York: Wiley; 1984.
- [37] Zhu X, Bowman KP, Genton MG. Incorporating geostrophic wind information for improved space–time short-term wind speed forecasting. *Ann Appl Statist* 2014;8(3):1782–99.
- [38] Pasquill F. The Estimation of Dispersion of Windborne Material. *Meteorol Magazine* 1961;90(33).
- [39] Dasari HP, et al. High-resolution assessment of solar energy resources over the Arabian Peninsula. *Appl Energy* 2019;248:354–71.
- [40] Giannakopoulou EM, Toumi R. The Persian Gulf summertime low-level jet over sloping terrain. *Q J R Meteorol Soc* 2012;138(662):145–57.
- [41] Samman AE, Gallus WA. A climatology of the winter low-level jet over the Red Sea. *Int J Climatol* 2018;38(15):5505–21.
- [42] Wang H, et al. Profiles of Wind and Turbulence in the Coastal Atmospheric Boundary Layer of Lake Erie. *J Phys Conf Ser* 2014;524:012117.
- [43] Emeis S. Current issues in wind energy meteorology. *Meteorol Appl* 2014;21(4):803–19. <https://doi.org/10.1002/met.1472>.
- [44] Emery C, Tai E, Yarwood G. Enhanced meteorological modeling and performance evaluation for two Texas ozone episodes. Prepared for the Texas natural resource conservation commission, by ENVIRON International Corporation; 2001.