Energy Reports 10 (2023) 4781–4790

Contents lists available at ScienceDirect

Energy Reports

journal homepage: www.elsevier.com/locate/egyr

Research paper

Evaluating ERA5 reanalysis predictions of low wind speed events around the UK

Panit Potisomporn *, Thomas A.A. Adcock, Christopher R. Vogel

Department of Engineering Science, University of Oxford, Parks Road, Oxford, OX1 3PJ, United Kingdom

ARTICLE INFO

Keywords:
ERA5
Reanalysis
Low wind speed events
Wind energy

ABSTRACT

Low wind speed events represent one of the biggest challenges in fully de-carbonising the electricity system due to the growing proportion of wind energy in the UK energy mix. While reanalysis products are a useful tool to study the spatio-temporal characteristics of these occurrences, their performance and limitations should be understood prior to usage. In this study, hourly 10 m ERA5 reanalysis wind speed data were evaluated against in-situ wind speed measurements from 205 onshore and offshore observation stations around the UK. It was found that ERA5 has biases in mean wind speed of 0.166 m/s and −0.136 m/s for onshore and offshore domains respectively, and biases in hourly wind speed standard deviation of −0.354 m/s and −0.425 m/s for onshore and offshore domains respectively. Both errors are more pronounced in autumn and winter. These errors lead to an underestimation of the percentage frequency of short-duration low wind speed events. Furthermore, the findings suggest that the largest errors are from sites which are situated in coastal and mountainous regions where short-range topographical variability and local wind effects may not be resolved by ERA5. Despite these shortcomings, ERA5 nevertheless outperforms its global reanalysis counterparts in the UK domain and therefore, can provide valuable information in the context of low wind speed events prediction.

1. Introduction

In March 2021, the United Kingdom experienced the longest "cold calm spell" in over a decade; for 11 consecutive days, wind farms operated at 11% of their rated capacity, in a month where low temperatures drove up heating and electricity demand. In response, the electricity system escalated gas power output to compensate for the lack of wind power generation, resorting to the very energy source that the country has been endeavouring to curb. Such an occurrence, neither an isolated incident nor limited to the UK, potentially represents one of the biggest challenges in fully de-carbonising the electricity system (Staffell et al., 2021). Hence, there is a growing need to characterise the UK's wind energy resource, particularly these low wind speed events.

A trend that has emerged in the past decade is the growing prevalence of reanalysis data being used as a substitute for observed wind speed or measured power output in the characterisation of low wind speed events (Cannon et al., 2015; Bloomfield et al., 2016; Dawkins, 2019). Otherwise known as hindcasts, reanalysis datasets integrate an atmospheric circulation model with historical climate observations through the process of data assimilation. Reanalyses offer advantages over observed data in two main aspects: data consistency and coverage, as they are provided as regularly gridded datasets that cover a large area, often the whole world, and are also provided at regular time steps. Such consistency overcomes the shortcomings of measured data which are irregularly spatially distributed and tend to decrease in number further into the past. These advantages thus allow reanalyses to perform two tasks that observed data do not permit: the consideration of long-term trends in wind speed and the investigation of the wind resource over a wide geographical domain.

However, several studies have highlighted a number of issues associated with using reanalysis data. First, significant biases in mean wind speed have been observed in reanalysis datasets which stem from underlying systemic model errors (Bloomfield et al., 2016). Second, generation variability represented by reanalyses tends to be underestimated over short time scales, which may lead to misrepresentation of ramp events (Cannon et al., 2015). Third, the coarse spatial resolution of reanalyses is not able to resolve complex topography and the ensuing effects on airflow (Dawkins, 2019). Hence, a thorough evaluation of a reanalysis dataset is necessary to understand and mitigate the uncertainties which arise from these issues.

Hence, a substantial number of evaluations of the accuracy of reanalysis wind speed have been conducted in various locations in both onshore and offshore environments. For example, Miao et al. (2020) evaluated the accuracy of European Centre for Medium-Range Weather

* Corresponding author.
E-mail address: panit.potisomporn@eng.ox.ac.uk (P. Potisomporn).

https://doi.org/10.1016/j.egyr.2023.11.035
Received 3 January 2023; Received in revised form 30 October 2023; Accepted 13 November 2023
Available online 21 November 2023

Forecasts Reanalysis Interim (ERA-Interim), Japanese 55-year Reanalysis (JRA-55), Climate Forecast System Reanalysis (CFSR), Modern-Era Retrospective analysis for Research and Applications 2 (MERRA-2), and National Centers for Environmental Prediction (NCEP) reanalyses against observed surface wind speed data in the Northern Hemisphere, Carvalho et al. (2014) assessed the performances of ERA-Interim, CSFR, NCEP reanalysis 2, and MERRA reanalyses in Portugal and the Iberian Peninsula, Kumar and Anandan (2009) evaluated NCEP reanalysis 2 in India, while Stopa and Cheung (2014) similarly validated ERA-Interim and CSFR reanalysis in United States waters. Ideally 100 m wind speed would be used for validation as this is closer to the hub height of modern multi-MW wind turbines. However, the majority of these evaluations resorted to using observed wind speed from observations at 10 m for validation due to the much greater data availability at this elevation. Only a few validation studies of reanalysis wind speed at 100 m have been made (Kiss et al., 2009; Liléo and Petrik, 2000). Regardless, while most studies have concluded that most reanalyses show a decent correlation (mean Pearson $R > 0.75$) of wind speed with observed data, they also found significant mean biases of magnitude greater than 0.3 m/s and found RMSEs higher than 3 m/s.

Only a limited number of European Centre for Medium-Range Weather Forecasts Reanalysis 5 (ERA5) reanalysis wind speed evaluations exist in literature due to its relatively recent release. Of those that have been performed, Gualtieri (2021) validated ERA5 wind speed against 100 m tower observations in America, Europe, the Middle East, Africa, and Oceania, Fan et al. (2020) compared ERA5 against other reanalyses using more than 1000 observation stations across all six continents, Kardakaris et al. (2021) evaluated ERA5 wind speed against in-situ measurements in the Greek seas, while Soukissian and Sotiriou (2022) compared the long-term variability of wind speed and direction in the Mediterranean Basin with results obtained from ERA-Interim. As with evaluations of other reanalysis datasets, these studies reported mean wind speed biases that range from −0.84 m/s to 2.28 m/s while RMSEs were found to reach as high as 3.69 m/s. Furthermore, significant spatial variations in mean wind speed bias were also observed on a global scale. However, due to the superior spatial resolution, these studies found that ERA5 globally exhibits better performance, with respect to mean wind speed bias and RMSE, compared to other reanalysis datasets.

There exist a number of gaps within these evaluations of reanalysis wind speed data that are relevant to the scope of this study. First, ERA5 wind speed has never been investigated on a large scale in the UK (Gualtieri, 2022). Second, only a few studies (e.g., (Sharp et al., 2015)) have investigated the sources of errors which may stem from physical factors like topography or model uncertainties. Last and more importantly, the exclusive usage of error metrics like RMSE, mean wind speed bias, and Pearson correlation in most studies means the performance of ERA5 is only captured in very general terms. While this may be sufficient for applications like energy yield calculations, this study's focus on low wind speed events requires a further examination of the dataset in terms of low wind speed events which is generally lacking in the literature. Studies that mention low wind speed validation only did so with respect to wind speed percentiles that do not stem from physical reasons and although Cannon et al. (2015) briefly touched on the comparison between MERRA reanalysis and observed sustained low wind speed events, it was not the main purpose of the study.

Therefore, the aim of this study is to provide a comprehensive evaluation of ERA5 wind speed in the UK. In addition to evaluating ERA5's performance using standard metrics for general accuracy of wind speed prediction, this study also presents an assessment focused on low wind speed events, as well as the sources of underlying uncertainties.

2. Methodology and data

2.1. Reanalysis data

Released in 2019, ERA5 (Hersbach et al., 2020) is the latest generation reanalysis model produced by ECMWF using a four-dimensional variational (4D-VAR) assimilation scheme. This offers significant advantages over a three-dimensional variational (3D-VAR) assimilation scheme of other widely used reanalyses such as MERRA-2 and CSFR (Jamet and Loisel, 2013) in terms of its ability to incorporate observations at the exact time of measurement. At a spatial resolution of 0.25° latitude-longitude, or approximately 27.5 km, and a maximum temporal resolution of 1-h time steps, ERA5 offers the highest spatiotemporal resolution out of all reanalyses that provide global coverage. The fact that ERA5 extends back to 1950 is also advantageous because of the rarity of extreme low wind speed events which necessitates a long dataset and to capture inter-decadal variability in wind speeds. For example, the UK average wind speed during 1960–1980 was significantly lower than that in 1980-2000 (Watson et al., 2015). Furthermore, in addition to wind data, which are available in the form of Easterly and Northerly wind speed components at 10 m and 100 m elevations, other meteorological variables such as 2 m temperature, boundary layer height etc. are also provided by ERA5. As the focus of the study is on the region containing the UK's wind farm fleet, the scope of ERA5 dataset used herein extends from 48°N to 61°N and from 10°W to 4°E in the form of gridded data as illustrated in Fig. 1.

2.2. Observation data

In this study we use in-situ 10 m wind speed measurements taken from weather observation stations around the UK to validate the ERA5 reanalysis data. While it is preferable to validate ERA5 at 100 m, which is representative of the average wind turbine hub height, most wind speed measurements are typically taken at 10 m. The majority of the observed data were obtained from the British Atmospheric Data Centre's (BADC) Met Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations dataset (Met Office, 2012), which comprises hourly weather measurements from the Met Office station network from 1853 to the present, and represents the largest database of UK wind speed measurements available in public archives. Because most MIDAS stations are located onshore, additional station data were obtained from Crown Estate's Marine Data Exchange (MDE) database, which comprises weather data collected from met-masts near UK offshore wind farms.

Although the actual number of observation stations from these datasets exceeds 300, only those whose data length is longer than one year and with data availability greater than 90% were used, so that seasonal variations could be observed. It is important to note that for those stations that met these criteria, while the available data record can extend to more than 40 years, MIDAS station anemometers had a cut-in wind speed of 4–5 m/s before the Met Office's upgrade campaign reduced the cut-in wind speed to 0.5 m/s in 1997 (Sloan and Clark, 2012). Since the focus of the present study is placed on low wind speed events, only data after 1997 were therefore used. The filtering procedure resulted in 205 time series with an average data length of 9.5 years within 1997–2020, 172 of which are from onshore stations and 33 from offshore stations whose locations are illustrated in Fig. 1 and summarised in Table 1.

Due to the size of the dataset, the data cleaning process was automated based on the criteria employed by Staffell and Green (2014). First, data points that did not pass the Met Office's quality checks were removed. Second, non-zero identical repeated values that extended over 24 h were deemed measurement instrument errors, and therefore removed. Third, anomalously high readings, namely wind speeds above 125 knots, were removed. Finally, data whose units were in knots were converted to m/s. For all 205 observation locations, data availability remained greater than 98% after the removal of data points by the quality check, especially data from the Marine Data Exchange dataset that are of consistently higher quality than those from the MIDAS dataset, both in terms of measurement precision and data completeness. Therefore, no effort was made to recreate missing or removed observations to avoid introducing uncertainty into the data.

In the present study we treat this quality-controlled dataset of field measurements as a ground truth against which to compare the reanalysis data.

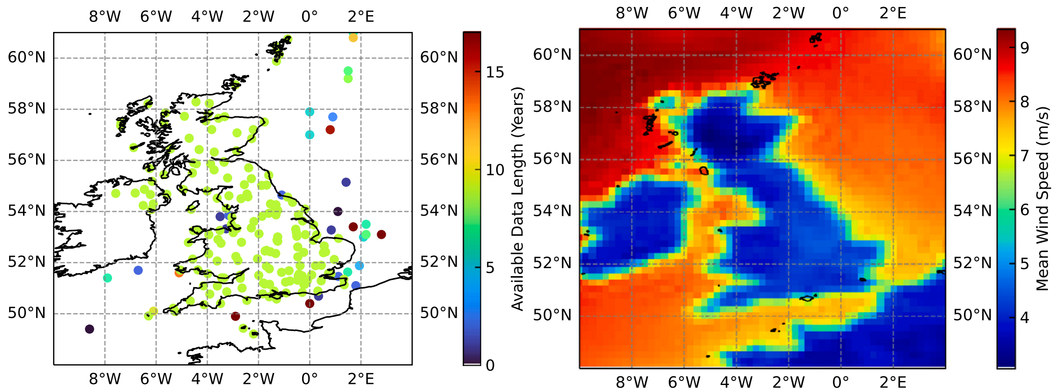

Fig. 1. Primary datasets used in the study. (Left) Locations of meteorological observation stations from the MIDAS dataset and the Marine Data Exchange dataset. (Right) Mean 10-m wind speed of the ERA5 dataset for 2000–2020.

Table 1
Observation station datasets used in the study. Note that 0.51 m/s corresponds to 1 knot, which is the precision of the MIDAS measurement devices.

Dataset	Domain	Precision	No. of Stations
MIDAS Land Observations	Onshore	0.51 m/s	172
MIDAS Marine Observations	Offshore	0.51 m/s	24
Marine Data Exchange	Offshore	0.01 m/s	9

2.3. Methodology

To evaluate the performance of ERA5 in capturing observed wind speed, the ERA5 wind speed time series was compared against the observed wind speed time series at the location of each of the 205 observation stations. This represents a total of 17,157,190 hourly data points drawn from the period 1997–2020. ERA5 wind speed time series at each specific location was obtained by separately interpolating Easterly and Northerly wind speed components u_x and u_y from the grid points to the station's coordinates. Multiple interpolation schemes were considered: bi-linear interpolation, inverse distance weighting interpolation, and bi-cubic interpolation. However, it was found the choice of interpolation scheme does not significantly affect the resultant interpolated time series, and hence, the less computationally expensive bi-linear interpolation was used. Using four grid points $(x_1, y_1), (x_1, y_2), (x_2, y_1)$, and (x_2, y_2) whose corresponding wind speed components are $u_{i,11}, u_{i,12}, u_{i,21}, u_{i,22}$, the bi-linearly interpolated wind speed component u_i at (x, y) is given by:

$$\begin{aligned} u_i = & \frac{(x_2 - x)(y_2 - y)}{(x_2 - x_1)(y_2 - y_1)} u_{i,11} + \frac{(x - x_1)(y_2 - y)}{(x_2 - x_1)(y_2 - y_1)} u_{i,21} \\ & + \frac{(x_2 - x)(y - y_1)}{(x_2 - x_1)(y_2 - y_1)} u_{i,12} + \frac{(x - x_1)(y - y_1)}{(x_2 - x_1)(y_2 - y_1)} u_{i,22}. \end{aligned} \tag{1}$$

The interpolated wind speed u and direction ϕ at the desired coordinate were then computed by:

$$u = \sqrt{u_x^2 + u_y^2}, \tag{2}$$

$$\phi = \frac{180}{\pi} \arctan\left(\frac{u_y}{u_x}\right). \tag{3}$$

At each observation station location, the observed wind speed time series u_{OBS} was compared against the interpolated ERA5 wind speed time series u_{ERA}, both of length N, using three metrics, namely: Root Mean Squared Error (RMSE) as a general error metric, Mean Error to capture bias, and Standard Deviation Error to capture the hourly variability as given by:

$$\text{RMSE (m/s)} = \sqrt{\frac{\Sigma_{i=1}^{N}(u_{ERA,i} - u_{OBS,i})^2}{N}}, \tag{4}$$

$$\text{Mean Error (m/s)} = \bar{u}_{ERA} - \bar{u}_{OBS}, \tag{5}$$

$$\text{StDev Error (m/s)} = \sigma_{ERA} - \sigma_{OBS}. \tag{6}$$

Error metrics related to the Weibull distribution, which is given by

$$f(u, \lambda, k) = \begin{cases} \frac{k}{\lambda}\left(\frac{u}{k}\right)^{k-1} e^{-(u/\lambda)^k} & ; \ u \geq 0 \\ 0 & ; \ u < 0 \end{cases}, \tag{7}$$

which are associated with the distribution's two governing parameters, namely the scale factor λ and the shape factor k were also used as performance metrics as given by:

$$\text{Scale Factor Error (\%)} = \frac{\lambda_{ERA} - \lambda_{OBS}}{\lambda_{OBS}}, \tag{8}$$

$$\text{Shape Factor Error (\%)} = \frac{k_{ERA} - k_{OBS}}{k_{OBS}}. \tag{9}$$

Finally, to capture the discrepancy between observed wind speed and ERA5 wind speed concerning low wind speed events, the error in the percentage frequency of wind speed less than the typical cut-in wind speed of a wind turbine of 4 m/s was used as a metric:

$$P_{Error}(u \leq 4) = P_{ERA}(u \leq 4) - P_{OBS}(u \leq 4), \tag{10}$$

where P_i is the respective probability density function.

A caveat that comes with such an approach is that there is no physical basis to support the use of the Weibull distribution to approximate wind speed. In the context of the present study, this is often the case for observed wind speed distributions for which there is a non-zero frequency of calm conditions (below the cut-in threshold of measurement devices), whereas a zero frequency of 0 m/s wind speeds is enforced by the two-parameter Weibull distribution. Although this fact has been acknowledged, these distribution-based error metrics are predicated on the commonly adopted practice of using the Weibull distribution as a standard engineering tool to approximate wind speed distributions (Burton et al., 2011; Carta and Ramírez, 2007; Seguro and Lambert, 2000). Hence, taking into account these limitations and given the scope of this study, the use of the Weibull distribution was deemed as adequate.

3. Results and discussions

3.1. Errors

3.1.1. Overview

Fig. 2 provides an overview of the errors between the 205 observed–ERA5 wind speed time series pairs in terms of the relationship between mean error and RMSE. With a mean RMSE of 1.849 m/s and a mean of mean error of 0.119 m/s, a small but not insignificant positive bias

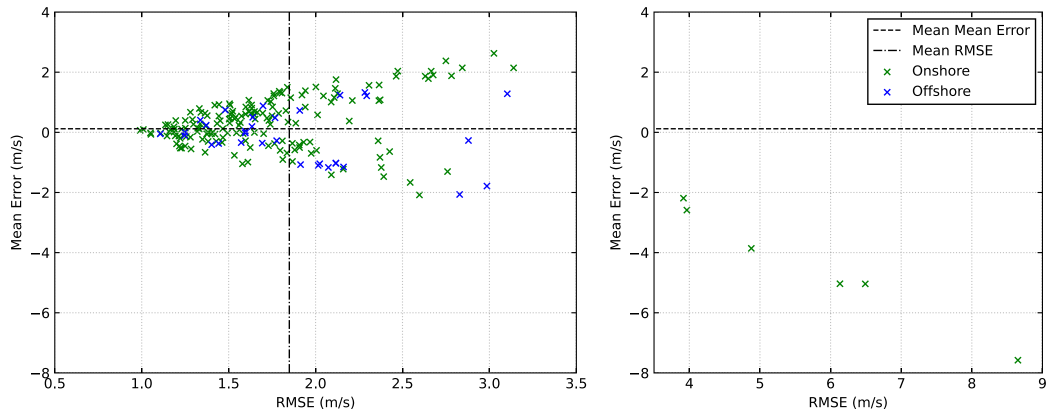

Fig. 2. Mean errors plotted against Root Mean Square Errors for each observed-ERA5 time series pair. Each cross represents errors from a single validation site, with green crosses indicating onshore locations and blue crosses indicating offshore locations. Note the scaling of the x-axis on the right to clearly display extreme errors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

is observed in ERA5 wind speed. While there is a clear correlation between RMSE and mean error, for locations that exhibit insignificant mean error, the non-zero RMSE when mean wind speed error is 0 m/s shows that a significant proportion of RMSE is also driven by wind speed noise. The relationship presented also shows that the largest RMSEs exhibited are caused by the underestimation of wind speed by ERA5 and that such an effect is only observed in the onshore domain; all locations with an RMSE greater than 3.5 m/s were onshore.

While the error profile presented here indicates that the reanalysis dataset is imperfect, ERA5 demonstrates superior performance in terms of these error metrics when compared with other datasets. For example, ERA-Interim, CSFR, and MERRA reanalyses exhibit mean RMSEs and means of mean errors upwards of 2 m/s and 0.34 m/s respectively when validated against observations in Northern Europe, the Iberian Peninsula, and Portugal (Carvalho et al., 2014; Gualtieri, 2021; Murcia et al., 2022). While these studies were conducted in different locations, a more direct comparison can be made between the present evaluation and the validation of the CSFR reanalysis dataset by Sharp et al. (2015). Like the present study, this was conducted in the UK against MIDAS stations data. The study reported mean RMSE and mean of mean error of 2.35 m/s and 0.33 m/s respectively, both of which are significantly higher than the ERA5 errors presented in this study.

To further evaluate the performance of ERA5 in the UK, ranges and distributions of each error metric are presented in Figs. 3 and 4. Moreover, due to the differences between onshore and offshore topography and weather regimes, it was also of interest to consider the performance of ERA5 separately in each environment. It was found that offshore sites demonstrate a smaller range of RMSEs due both to the extreme onshore errors previously mentioned and also that RMSE at the majority of sites lies between 1 and 2 m/s, with a mean of 1.889 m/s and a median of 1.628 m/s for onshore locations, and a mean of 1.841 m/s and a median of 1.771 m/s for offshore locations.

There are two major differences in mean errors between onshore and offshore sites. First, offshore sites exhibit fewer extreme errors than their onshore counterparts and, second, while there is a 0.166 m/s positive bias of mean wind speed in the onshore environment, a 0.136 m/s negative bias is observed in the offshore environment. Furthermore, both the onshore and offshore mean error distributions are loosely centred around their respective mean, while also displaying considerable variance. This observation is also reflected in the error distribution of scale factor, as scale factor is approximately proportional to mean wind speed for a given location. Similarly, the distribution of standard deviation error are also loosely centred around their respective mean, while also displaying considerable magnitudes of variance., albeit centred on a value that is below zero for both the onshore and offshore environments. This is reflected in the shape error distribution, which shows a very low frequency of negative error, suggesting that ERA5 consistently underestimates the hourly variability of wind speed. As a result, the ERA5 wind speed time series shows a clear tendency, especially in the onshore environment, to underestimate the percentage frequency of low wind speeds, as shown in the distribution of $P(u \leq 4)$ errors whose magnitude can approach 50%.

3.1.2. *Seasonal variation*

Seasonal variation of ERA5's predictions also has important implications on the reanalysis' ability to accurately simulate observed wind speed. Wind speed time series from all observation sites were sorted into each season as defined by a three-month period (spring starts on 1st of March and ends on 31st of May etc.), and the results are presented in Fig. 5. There is a clear difference between errors present in spring and summer and those in winter and autumn, where the former have an average mean error of −0.05 and 0.00 m/s whereas the latter pair's average mean errors are 0.28 m/s and 0.23 m/s respectively. At 1.78 m/s and 1.65 m/s, the RMSEs in spring and summer are also markedly lower than the RMSEs of 2.04 m/s and 1.90 m/s in winter and autumn. As well as the similar pattern observed for standard deviation errors, these trends lead to the fact that the percentage frequency of low wind speed, defined as that below 4 m/s, is only underestimated on average by 0.29% and 0.42% in spring and summer, compared to 5.46% and 5.81% in autumn and winter. These results are not unexpected as wind speeds tend to be higher in winter and autumn than in spring and summer due to the increase in temperature gradient and hence, the pressure gradient, between the approaching and the displaced air masses (Sinden, 2007; Baker et al., 1990). Therefore, on average, ERA5 would be expected to predict higher wind speeds in these months. Such an error trend is significant because winter and autumn are the two seasons where low temperatures lead to high electricity demand for heating, hence, the underestimation of low wind speed percentage frequency may lead to an overestimation of wind energy production's ability to match demand during these periods.

3.1.3. *Low wind speed events*

$P(u \leq 4)$ error is included as an error metric as the focus of this study is on low wind speed events. Although the percentage frequency of low wind speed $P(u \leq 4)$ provides a convenient measure to compare the performance of ERA5 in terms of extreme events across multiple locations, the duration of these events is of greater importance, as long periods of wind droughts can have more significant impacts on wind energy production than multiple scattered short-term events. There are multiple ways to quantify these longer-term events, as presented in each row of Fig. 6 for both the onshore and offshore environments.

First, low wind speed events, defined as periods where the wind speed is continuously below 4 m/s, from all 205 locations were sorted

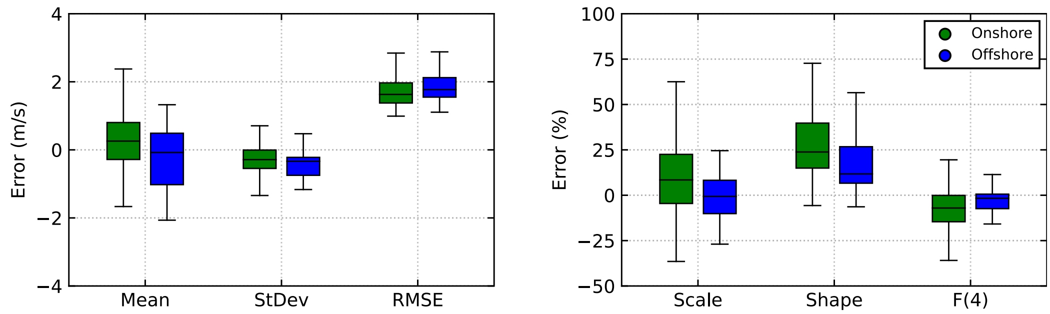

Fig. 3. Onshore (green) and offshore (blue) errors of ERA5 wind speed compared to observations at 205 validation sites. The left plot shows hourly wind speed mean error, standard deviation error, and RMSE while the right plot shows errors in scale parameter, shape parameter, and percentage frequency of low wind speed. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

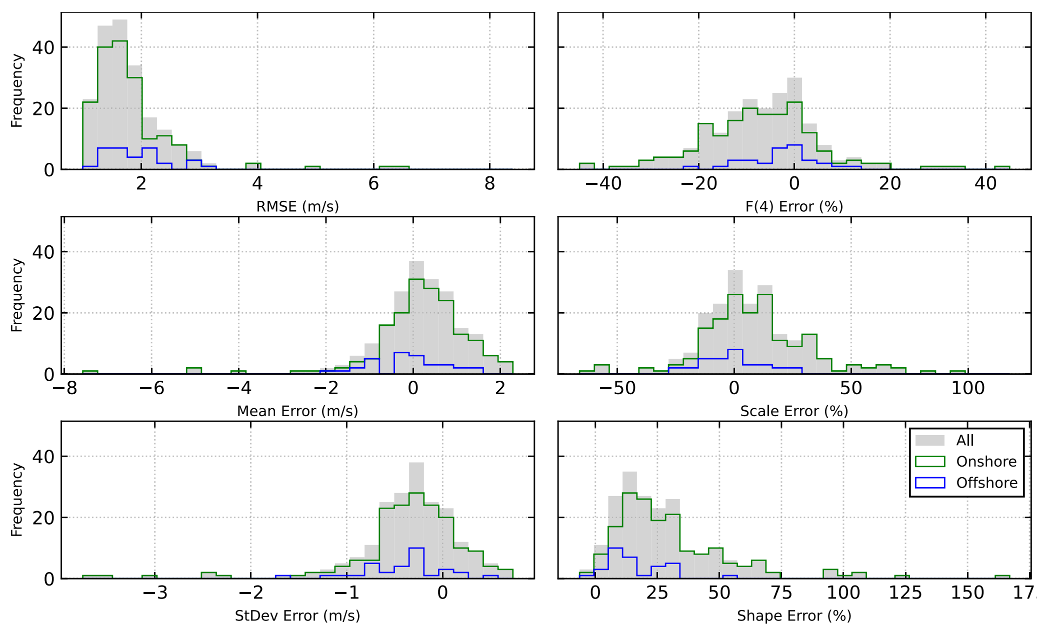

Fig. 4. Distributions of onshore (green), offshore (blue), and combined (grey) errors of ERA5 wind speed compared to observations at 205 validation sites. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

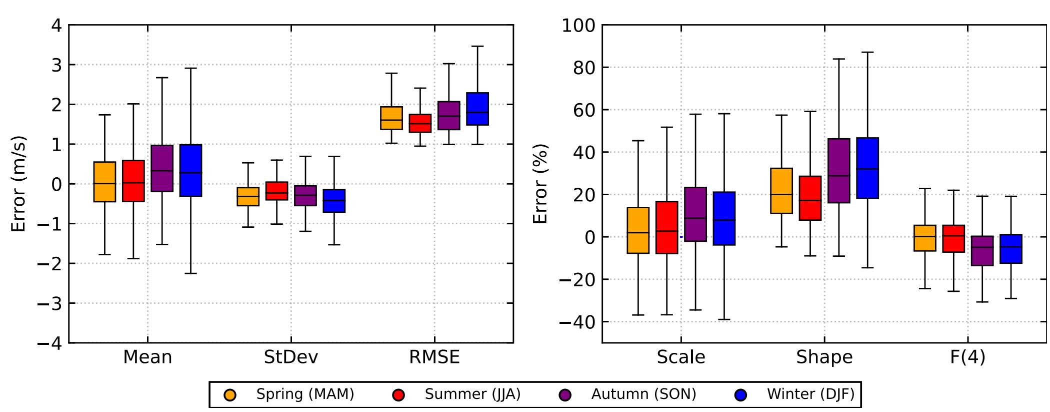

Fig. 5. ERA5 errors sorted into seasons. The left plot shows hourly wind speed mean error, standard deviation error, and RMSE while the right plot shows errors in scale parameter, shape parameter, and percentage frequency of low wind speed.

based on their duration and plotted against their number of observations. The first notable finding is the periodicity in the number of observations of onshore low wind speed events that is not seen in the offshore environment. This effect is explained by the clearer and more consistent diurnal variations of wind speed on land that arise from the presence of short time-scale changes in land temperature and their absence in sea surface temperature (Barthelmie, 1993). However, a more important trend that can be observed for both environments is that the number of observations of short-term (below 25 h for onshore and 15 h for offshore) low wind speed events is significantly underestimated by ERA5. For example, 1-h events are underestimated by over 260,000 observations. On the other hand, the number of observations

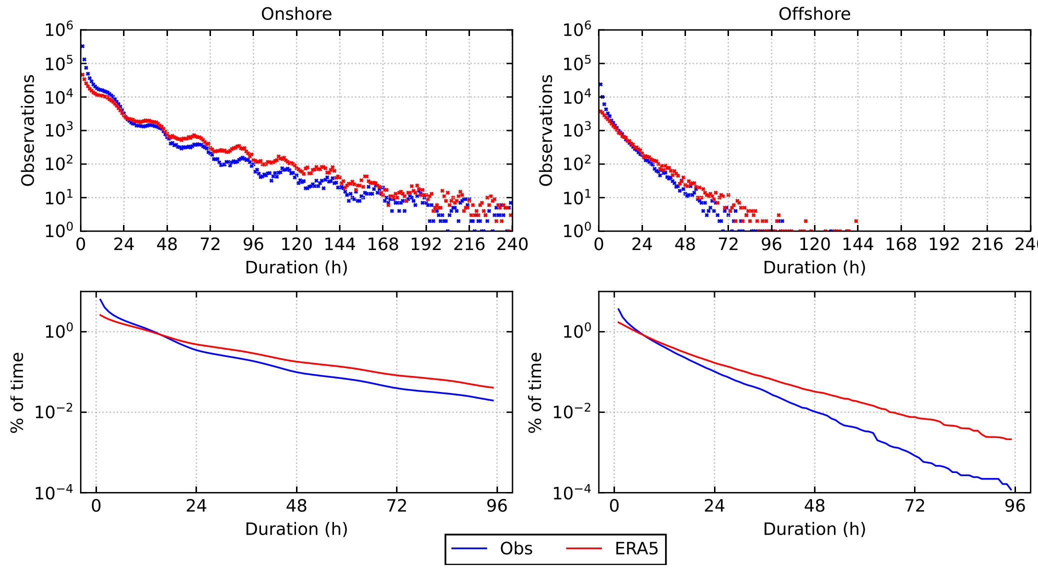

Fig. 6. Errors in terms of the persistence of low wind speed events from observation (blue) and ERA5 (red) data. The left column shows errors from onshore locations while the right column shows errors from offshore locations. The top row considers the raw frequency of these sustained low wind speed events while the bottom row considers the cumulative frequency. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of longer low wind speed events is overestimated by ERA5, though by a smaller margin. For example, ERA5 overestimates the observations of a five-day-long low wind speed event by 53 observations.

A more useful representation of low wind speed events, which has been employed in previous studies (Cannon et al., 2015; Potisomporn and Vogel, 2022), is to present these events as a cumulative sum and as a percentage of time rather than the raw number of observations, as illustrated in the second row of Fig. 6. For example, for the onshore environment, low wind speed events that persist for at least 15 h make up 1.1% of the data record. Here, it can be seen more clearly that considering the severity of events, the observed and ERA5 data show a decent level of agreement for wind speed events over 16 h for onshore and 6 h for offshore, while shorter wind speed events are significantly underestimated.

While the cumulative density at a low wind speed threshold ($P(u \leq 4)$) demonstrates that ERA5 tends to underestimate the number of observations of low wind speed, a further investigation into the duration of these events indicates that this discrepancy is heavily influenced by the underestimation of low wind speed events shorter than one day. This observed discrepancy can be explained by ERA5's underestimation of the hourly wind variability caused by small-scale fluctuations that are not captured by the macro-scale-driven ERA5 simulation, hence, the inability to capture small, frequent drops in wind speed that give rise to short events. Overall, it can be argued that ERA5's performance in capturing longer low wind speed events, which have a greater impact on wind energy's generation than short, intermittent ones, is more significant than its inability to capture the number of shorter events.

3.2. *Sources of uncertainty*

3.2.1. *Coastal and elevated sites*

Fig. 7 illustrates the spatial variation of the differences between ERA5 and observed wind speeds to gain insight into the errors and uncertainty associated with ERA5 reanalysis around the UK. While there is no apparent spatial pattern of error distribution across the UK, there exist some features of interest. First, a number of sites with high underestimation of mean and standard deviation, and hence high RMSE and overestimation of $P(u \leq 4)$, are observed onshore of North Western England and Western Scotland. This is a high-elevation region that typically experiences high wind speed but is represented by the lowest mean wind speed in ERA5 as shown in Fig. 1. Second, sites with large magnitude of mean wind speed errors and hence, high errors in low wind speed percentage frequency are often situated along the coast, particularly Great Britain's Southern coast.

Two observed trends are that coastal and elevated areas give rise to particularly significant errors, consistent with Sharp et al. (2015)'s evaluation of CFSR reanalysis wind speed. To further clarify this observation, the 205 sites were sorted into four categories, namely coastal locations (45 sites), elevated locations (22 sites), both coastal and elevated locations (3 sites), and locations that are neither coastal nor elevated (135 sites), where coastal locations are defined as those within 27.5 km of the coastline (the resolution of ERA5) and elevated locations are those situated over 200 m above sea level. The results presented in Figs. 8 and 9 corroborate this finding in a number of ways, as exemplified by the mean RMSE of non-coastal-non-elevated sites of 1.67 m/s compared to the RMSE of 3.17 m/s from sites in the other three categories.

First, elevated and elevated/coastal sites were found to be negatively biased whereas it was previously observed that ERA5 wind speed tends to be positively biased in the UK. High elevation implies mountainous regions that are predominantly situated in Scotland (e.g., the Grampion mountains, the Maell Gorm peak etc.) and some of the sites where ERA5 has the highest underestimation of mean wind speed are situated in such region, therefore implying the effects of topography on these discrepancies. Wind in mountainous regions is subjected to both the ridge-acceleration effect, where wind layers are accelerated over a peak, and the Venturi effect, where passes and valleys parallel to the wind direction increase the wind speed (Mathew, 2006). However, these topographic features are too small to be captured by the resolution of ERA5 grids, hence, physical features affecting the flow are under-resolved, leading to the underestimation of wind speed in these regions. This is further corroborated by Fig. 9 where underestimations of mean wind speed by ERA5 are more significant for elevated sites with higher observed wind speed than overestimations of mean wind speed at non-elevated sites with lower observed mean wind speed, hence implying localised, topography-induced acceleration. Therefore, rather than indicating a flaw in ERA5's performance, these outlying errors point to the placement of certain observation stations in locations where the winds are highly localised and greater resolution (e.g. through down-scaling) would be required. Regardless, it should be noted that these elevated sites, especially those with the largest errors in the North, are situated in remote mountainous regions that do not

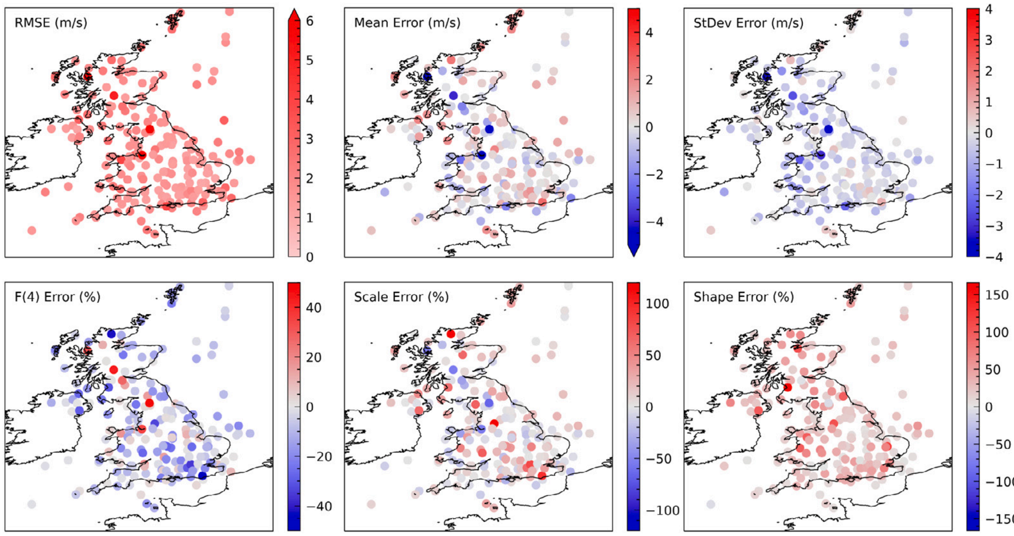

Fig. 7. Spatial distribution of ERA5 errors around the UK. Each circle represents the location of the observation station used to validate ERA5 data in this study.

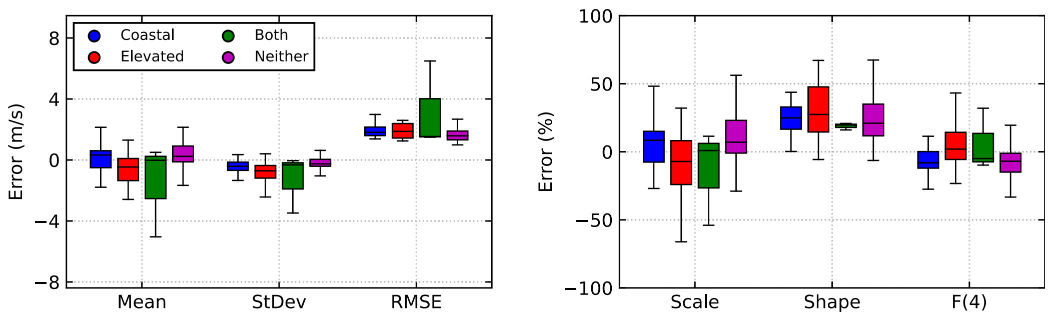

Fig. 8. Errors according to the classification of each site which is one of: coastal (blue), elevated (red), coastal and elevated (green), and neither (purple). Coastal sites are defined as those within 27.5 of the coastline (the resolution of ERA5) and elevated sites are defined as those over 200 m above sea level. The left plot shows hourly wind speed mean error, standard deviation error, and RMSE while the right plot shows errors in scale parameter, shape parameter, and percentage frequency of low wind speed. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

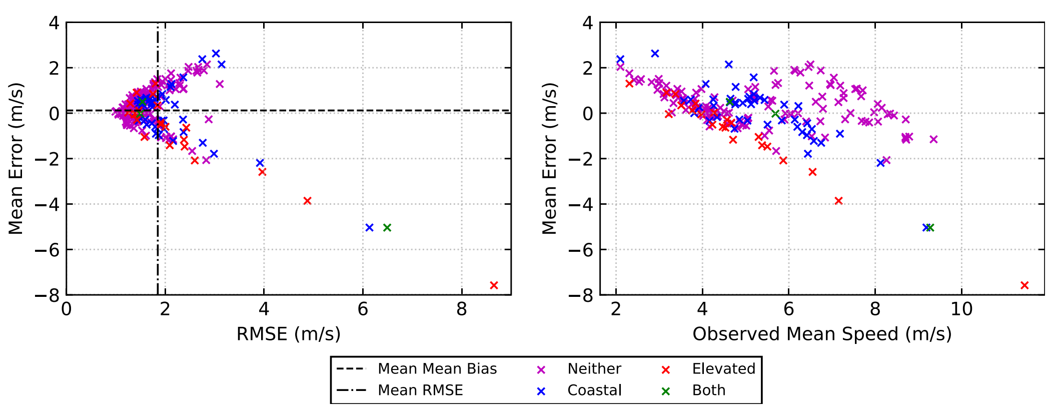

Fig. 9. (Left) Relationship between mean error and RMSE. (Right) Relationship between mean error and observed mean speed. Each cross represents errors from a single validation site sorted into the classification of each site (coastal (blue), elevated (red), coastal and elevated (green), and neither (purple)). Coastal sites are defined as those within 27.5 of the coastline (the resolution of ERA5) and elevated sites are defined as those over 200 m above sea level). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

currently accommodate nor are particularly suitable for wind farms and so will not have significant effects on the analysis of the UK's wind resource.

Second, although coastal sites exhibit a similar range of mean wind speed errors to sites that are neither coastal nor elevated, they contain more extreme errors. Characterised by the variability of topography, surface roughness, and thermal gradient, coastal areas have always been regarded as a modelling challenge (Beaucage et al., 2007; Sharp et al., 2015). This is, again, partly due to the resolution of reanalyses which cannot resolve the changes that occur over the short distance that is the interface between the offshore and the onshore environment. But unlike elevated sites whose terrains are not always suitable for wind farms, coastal regions, especially the West coast, are home to several wind farms both offshore and onshore and hence, are more

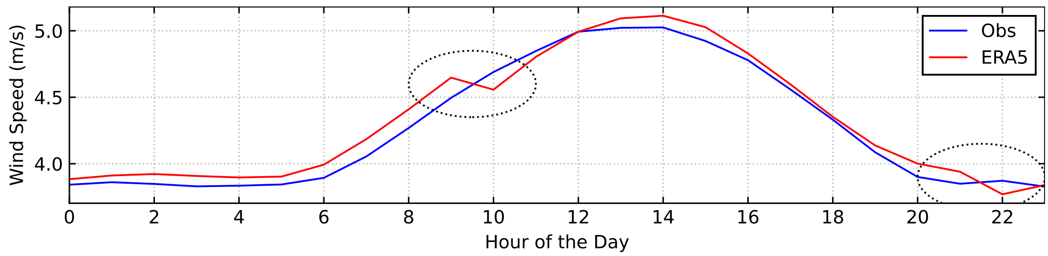

Fig. 10. Diurnal variation of observed and ERA5 wind speed taken from a validation site at 50.86, −3.24, with the discontinuities at 0900–1000 and 2100–2200 highlighted.

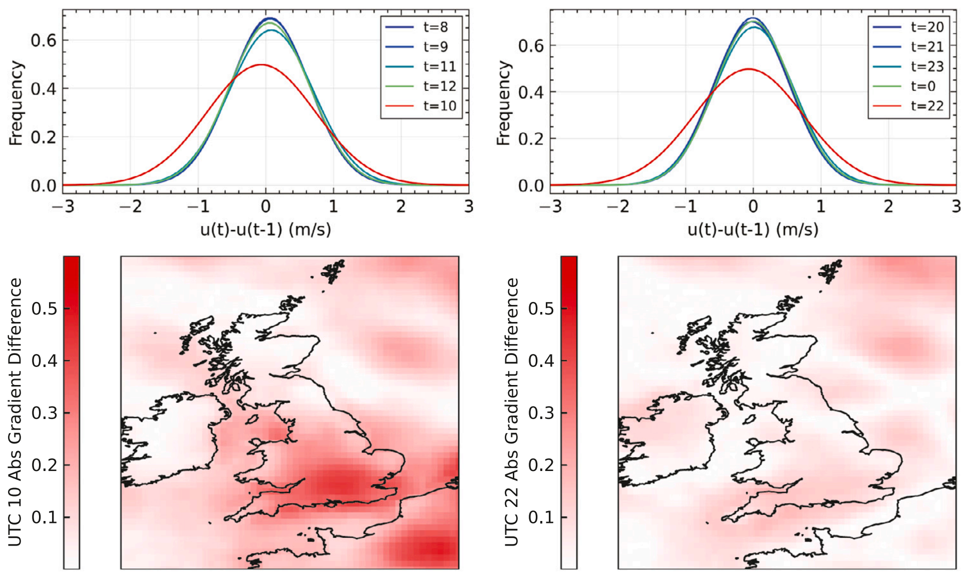

Fig. 11. Distributions of hourly wind speed gradient from (a) 0700–1200 UTC (b) and 1900–0000 UTC. (Bottom) Spatial difference of mean gradient difference (c) between 0900–1000 UTC and the mean of 0700–0900 UTC and 1000–1200 UTC, and (d) between 2100–2200 UTC and the mean of 1900–2100 UTC and 2200–0000 UTC.

consequential when evaluating reanalysis data. However, at an average RMSE of 2.05 m/s, ERA5's performance in coastal areas already surpasses other global reanalyses (Sharp et al., 2015).

3.2.2. Assimilation window error

Another major source of uncertainty stems from the underlying data assimilation model of ERA5 reanalysis. Reanalyses rely on the process of data assimilation to incorporate observed data to correct the forecasts made by the underlying weather prediction model. This process takes place in assimilation windows which extend for a short period of time, within which the cost function that captures the difference between the forecast and the observed data is minimised. In a 3D-VAR model where time is not involved, observations within the assimilation window are only used to correct a single point in the time domain Fujiwara et al. (2017). However, in ERA5's underlying 4D-VAR assimilation model, including time as the fourth dimension means that the initial state of the assimilation window is adjusted in the process to optimise the forecast for the entire window (Hersbach et al., 2020). More specifically, ERA5 utilises two 12-h assimilation windows per 24-h cycle where one begins at 1000 UTC and another at 2200 UTC. Since the initial state of these windows is also subjected to adjustments in a 4D-VAR model, discontinuities between 0900–1000 UTC and 2100–2200 UTC are present in several climate parameters, including u_x and u_y components of wind speed.

Although such an issue has been recognised by the ECMWF (European Centre for Medium-range Weather Forecasts, 2018), to the author's knowledge, no previous evaluation of ERA5 wind speed has investigated the impacts that these discontinuities have on data reliability especially in the context of low wind speed events. Hence, to observe such effect on wind speed, Fig. 10 illustrates the diurnal variation of wind speed from a typical observation site, averaged over the period 2010–2020, showing clearly the discontinuities in wind speed at the junction of these 12-h assimilation windows. A question that arises is whether these discontinuities are the result of consistent drops in wind speed. The top row of Fig. 11 shows the distribution of the hourly wind speed gradients at these two junctions. While the wind speed gradients around these two periods (two hours before and after) approximate a normal distribution (p-value > 0.9 by Kolmogorov–Smirnov test) with a positive mean, the gradient at these two time junctions approximate a normal distribution with a negative mean and a relatively larger variance, thus indicating that the discontinuities are the result of a random process rather than a consistent drop in wind speed. It can also be clearly seen that the effect is more pronounced at the 0900–1000 UTC discontinuity than that at 2100–2200 UTC, though the larger-than-usual variance of wind speed gradients at the later junction is still a cause for concern. To investigate the spatial variation of these discontinuities, the mean difference between the gradient at each junction and the two adjacent time steps are shown in the bottom row of Fig. 11. It is apparent that the effect is more pronounced at the 0900–1000 UTC than at the 2100–2200 UTC discontinuity and more importantly, that some locations are more susceptible to this effect, most visibly Southern Great Britain and the Southern North Sea.

There are two major implications of this modelling error. First, ERA5's representation of wind speed diurnal variation will not be representative of reality. Second, and more importantly, this error could lead to a misrepresentation of low wind speed events. Unlike a natural diurnal pattern where the likelihood of a low wind speed event starting at 1000 UTC is moderately low, ERA5's diurnal pattern exhibits a spike in likelihood at this hour of the day. Similarly, a significant drop in the likelihood of a low wind speed event ending at 1000 UTC was also observed. The likelihood of low wind speed event starting or ending at adjacent hours of the day remain comparable to the natural diurnal pattern. Hence, this implies that, for shorter events, this error will only increase the percentage frequency of one-hour low wind speed events and that for longer events, it will only extend the duration of low wind speed events by one hour. On balance, while such model error is undesirable, it is also unavoidable, and given that the impacts are limited, the utilisation of the fourth dimension in a 4D-VAR system still offers more advantages than disadvantages.

4. Conclusion

Prolonged low wind speed events, such as that in March 2021 in the United Kingdom, represent one of the big challenges in fully decarbonising the electricity system, hence the need to quantify these occurrences. While their superior temporal and spatial coverage, compared to observation data, make reanalysis data useful in assessing these rare events, their accuracy in predicting wind speed should be assessed prior to usage.

This paper evaluated ERA5 reanalysis wind speed data at 10 m against observation data from 205 weather stations, both onshore and offshore, around the UK from the year 1997 to 2020. ERA5 exhibits a positive mean wind speed bias in the onshore domain of 0.166 m/s and a negative mean wind speed bias of −0.136 m/s in the offshore domain. For both domains, ERA5 reanalysis under-predicts the hourly variability of wind speed by −0.365 m/s. Furthermore, it has been demonstrated that these errors are more pronounced in winter and autumn than in summer and spring. The spatial distribution of these errors indicates that the most extreme errors come from onshore validation sites that are situated in coastal or mountainous regions where ERA5's resolution is unable to resolve localised wind effects. While mountainous regions, which are not suitable for wind farms, may be less consequential, coastal regions do accommodate a number of wind farms and hence, point out to the need for higher-resolution models that can well capture this interface.

These errors, therefore, lead to an underestimation of the percentage frequency of low wind speed. However, a further investigation into the discrepancy between observed and ERA5 percentage frequency of low wind speed shows that this is mainly influenced by a large underestimation of short-term low wind speed events that persist for less than 12 h whereas longer events, which have higher significance on the electricity system, are shown to be markedly more accurately represented.

Reanalysis data are less than perfect and these findings show that, in agreement with evaluations of other reanalysis datasets, ERA5 is also subject to a significant level of bias, topography-induced errors, a misrepresentation of short-duration low wind speed events. However, the mean levels of mean wind speed bias and RMSE exhibited by the 205 validation sites demonstrate that despite these errors, ERA5 still outperforms its global reanalysis counterparts in the UK domain compared to results reported by previous studies. Hence, despite the shortcomings highlighted in this paper, ERA5 proves to be an important tool which can provide valuable information in the context of low wind speed events prediction.

CRediT authorship contribution statement

Panit Potisomporn: Methodology, Data curation, Software, Writing – original draft, Validation, Writing – review & editing, Formal analysis, Visualization, Investigation. **Thomas A.A. Adcock:** Conceptualization, Writing – review & editing, Methodology, Supervision, Resources. **Christopher R. Vogel:** Conceptualization, Writing – review & editing, Methodology, Supervision, Resources.

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.

Data availability

Data will be made available on request.

Acknowledgements

This research was funded in whole, or in part, by CRV's UKRI Future Leaders Fellowship MR/V02504X/1. For the purpose of Open Access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

References

Baker, R.W., Walker, S.N., Wade, J.E., 1990. Annual and seasonal variations in mean wind speed and wind turbine energy production. Sol. Energy 45, 285–289. http://dx.doi.org/10.1016/0038-092X(90)90013-3.

Barthelmie, R.J., 1993. Prospects for offshore wind energy: The state of the art and future opportunities: Offshore wind climate and wind energy applications prospects for offshore wind energy: The state of the art and future opportunities offshore wind climate and wind energy applications. Source: Wind Eng. 17, 86–99.

Beaucage, P., Glazer, A., Choisnard, J., Yu, W., Bernier, M., Benoit, R., Lafrance, G., 2007. Wind assessment in a coastal environment using synthetic aperture radar satellite imagery and a numerical weather prediction model. Can. J. Remote Sens. 33, 368–377. http://dx.doi.org/10.5589/m07-043.

Bloomfield, H.C., Brayshaw, D.J., Shaffrey, L.C., Coker, P.J., Thornton, H.E., 2016. Quantifying the increasing sensitivity of power systems to climate variability. Environ. Res. Lett. 11, 124025. http://dx.doi.org/10.1088/1748-9326/11/12/124025.

Burton, T., Jenkins, N., Sharpe, D., Bossanyi, E., 2011. Wind Energy Handbook. John Wiley & Sons.

Cannon, D.J., Brayshaw, D.J., Methven, J., Coker, P.J., Lenaghan, D., 2015. Using reanalysis data to quantify extreme wind power generation statistics: A 33 year case study in Great Britain. Renew. Energy 75, 767–778. http://dx.doi.org/10.1016/j.renene.2014.10.024.

Carta, J.A., Ramírez, P., 2007. Use of finite mixture distribution models in the analysis of wind energy in the Canarian Archipelago. Energy Convers. Manage. 48, 281–291. http://dx.doi.org/10.1016/j.enconman.2006.04.004.

Carvalho, D., Rocha, A., Gómez-Gesteira, M., Santos, C.S., 2014. WRF wind simulation and wind energy production estimates forced by different reanalyses: Comparison with observed data for Portugal. Appl. Energy 117, 116–126. http://dx.doi.org/10.1016/j.apenergy.2013.12.001.

Dawkins, L.C., 2019. Weather and climate related sensitivities and risks in a highly renewable UK energy system: A literature review. URL www.metoffice.gov.uk.

European Centre for Medium-range Weather Forecasts, 2018. ERA5 data documentation. URL https://confluence.ecmwf.int/display/CKB/ERA5%3A+data+documentation#ERA5:datadocumentation-Knownissues.

Fan, W., Liu, Y., Chappell, A., Dong, L., Xu, R., Ekström, M., Fu, T.M., Zeng, Z., 2020. Evaluation of global reanalysis land surface wind speed trends to support wind energy development using in situ observations. J. Appl. Meteorol. Climatol. 60, 33–50. http://dx.doi.org/10.1175/JAMC-D-20-0037.1.

Fujiwara, M., Wright, J.S., Manney, G.L., Gray, L.J., Anstey, J., Birner, T., Davis, S., Gerber, E.P., Harvey, V.L., Hegglin, M.I., Homeyer, C.R., Knox, J.A., Krüger, K., Lambert, A., Long, C.S., Martineau, P., Molod, A., Monge-Sanz, B.M., Santee, M.L., Tegtmeier, S., Chabrillat, S., Tan, D.G., Jackson, D.R., Polavarapu, S., Compo, G.P., Dragani, R., Ebisuzaki, W., Harada, Y., Kobayashi, C., McCarty, W., Onogi, K., Pawson, S., Simmons, A., Wargan, K., Whitaker, J.S., Zou, C.Z., 2017. Introduction to the SPARC reanalysis intercomparison project (S-Rip) and Overview of the reanalysis systems. Atmos. Chem. Phys. 17, 1417–1452. http://dx.doi.org/10.5194/acp-17-1417-2017.

Gualtieri, G., 2021. Reliability of ERA5 reanalysis data for wind resource assessment: A comparison against tall towers. Energies 14, 4169. http://dx.doi.org/10.3390/en14144169.

Gualtieri, G., 2022. Analysing the uncertainties of reanalysis data used for wind resource assessment: A critical review. Renew. Sustain. Energy Rev. 167, 112741.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G.D., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J.N., 2020. The ERA5 global reanalysis. Q. J. R. Meteorol. Soc. 146, 1999–2049. http://dx.doi.org/10.1002/qj.3803.

Jamet, C., Loisel, H., 2013. Data Assimilation Methods. Wiley Blackwell, pp. 303–317. http://dx.doi.org/10.1029/2008GM000763.

Kardakaris, K., Boufidi, I., Soukissian, T., 2021. Offshore wind and wave energy complementarity in the Greek seas based on ERA5 data. Atmosphere 12 (10), 1360.

Kiss, P., Varga, L., Jánosi, I.M., 2009. Comparison of wind power estimates from the ECMWF reanalyses with direct turbine measurements. J. Renew. Sustain. Energy 1, 033105. http://dx.doi.org/10.1063/1.3153903.

Kumar, M.S., Anandan, V.K., 2009. Comparision of the NCEP/NCAR reanalysis II winds with those observed over a complex terrain in lower atmospheric boundary layer. Geophys. Res. Lett. 36, http://dx.doi.org/10.1029/2008GL036246.

Liléo, S., Petrik, O., 2000. Investigation on the use of NCEP/NCAR, MERRA and NCEP/CFSR reanalysis data in wind resource analysis. Sigma 1.

Mathew, S., 2006. Wind Energy: Fundamentals, Resource Analysis and Economics, Vol. 1. Springer.

Met Office, 2012. Met office integrated data archive system (MIDAS) land and marine surface stations data (1853-Current). URL http://catalogue.ceda.ac.uk/uuid/220a65615218d5c9cc9e4785a3234bd0.

Miao, H., Dong, D., Huang, G., Hu, K., Tian, Q., Gong, Y., 2020. Evaluation of northern hemisphere surface wind speed and wind power density in multiple reanalysis datasets. Energy 200, 117382.

Murcia, J.P., Koivisto, M.J., Luzia, G., Olsen, B.T., Hahmann, A.N., Sørensen, P.E., Als, M., 2022. Validation of European-scale simulated wind speed and wind generation time series. Appl. Energy 305, 117794. http://dx.doi.org/10.1016/j.apenergy.2021.117794.

Potisomporn, P., Vogel, C.R., 2022. Spatial and temporal variability characteristics of offshore wind energy in the United Kingdom. Wind Energy 25 (3), 537–552.

Seguro, J.V., Lambert, T.W., 2000. Modern estimation of the parameters of the Weibull wind speed distribution for wind energy analysis. J. Wind Eng. Ind. Aerodyn. 85, 75–84.

Sharp, E., Dodds, P., Barrett, M., Spataru, C., 2015. Evaluating the accuracy of CFSR reanalysis hourly wind speed forecasts for the UK, using in situ measurements and geographical information. Renew. Energy 77, 527–538. http://dx.doi.org/10.1016/j.renene.2014.12.025.

Sinden, G., 2007. Characteristics of the UK wind resource: Long-term patterns and relationship to electricity demand. Energy policy 35, 112–127. http://dx.doi.org/10.1016/j.enpol.2005.10.003.

Sloan, C., Clark, M.R., 2012. A comparison of three met office wind observing systems. Atmos. Sci. Lett. 13, 283–288. http://dx.doi.org/10.1002/asl.396.

Soukissian, T., Sotiriou, M.-A., 2022. Long-term variability of wind speed and direction in the Mediterranean Basin. Wind 2 (3), 513–534.

Staffell, I., Green, R., 2014. How does wind farm performance decline with age? Renew. Energy 66, 775–786. http://dx.doi.org/10.1016/j.renene.2013.10.041.

Staffell, I., Green, R., Green, T., Jansen, M., 2021. Electric insights quarterly.

Stopa, J.E., Cheung, K.F., 2014. Intercomparison of wind and wave data from the ECMWF reanalysis interim and the NCEP climate forecast system reanalysis. Ocean Model. 75, 65–83. http://dx.doi.org/10.1016/j.ocemod.2013.12.006.

Watson, S., Kritharas, P., Hodgson, G., 2015. Wind speed variability across the UK between 1957 and 2011. Wind Energy 18 (1), 21–42.