Energy 229 (2021) 120603

Contents lists available at ScienceDirect

Energy

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

Accurate long-term power generation model for offshore wind farms in Europe using ERA5 reanalysis

Liam Hayes*, Matthew Stocks, Andrew Blakers

College of Engineering & Computer Science, Australian National University, Australia

A R T I C L E I N F O

Article history:
Received 23 November 2020
Received in revised form
16 March 2021
Accepted 6 April 2021
Available online 18 April 2021

Keywords:
ERA5
Offshore wind farm
Wind power modelling
Reanalysis
Capacity factor
Hourly timeseries

A B S T R A C T

Accurate long-term wind speed data is important for understanding the role of offshore wind farms in future energy systems. Meteorological reanalyses, such as ERA5, are relied upon by the wind energy industry and researchers. Being unaffected by onshore topography and surface roughness, the historic generation of offshore wind farms can be accurately predicted using such weather reanalysis.

In this work we present a new method for using ERA5 weather data to model long term (>40 year) hourly wind generation for individual offshore wind farms. The model is validated against 57 offshore wind farms in Europe, and reduces the root mean squared error in hourly and daily capacity factor predictions by 10% and 18% respectively when compared to the Renewables Ninja. Further, 40 years (from 1980 to 2019) of ERA5 hourly wind speeds within 200 km of the coast is made easily available for energy system research on our accompanying website (windtlas.xyz).

1. Intro & background

Wind energy is the second leading source of new electricity generation capacity deployment globally after solar photovoltaics. Offshore wind is a key driver of wind energy growth, with the Global Wind Council predicting annual installations will reach 20 Gigawatts (GW) by 2025. This growth has corresponded with decreases of the levelised cost of off-shore wind by two thirds from US$225/MWh to US$83/MWh with forecasts of US$58/MWh in 2025 [1], increasing its economic attractiveness. Due to limited land availability and a generally superior wind resource, more densely populated regions such as Japan and Europe are likely to depend on offshore wind for the bulk of their electricity needs despite its higher cost. For example, the British Government has recently announced plans for 40GW of offshore wind by 2030, sufficient to power all UK homes [2].

Wind is a variable renewable electricity source, in that the generation is non-dispatchable and dependent upon the variation of the wind resource. As the fraction of variable generation increases in an electricity system, supply and demand needs to be balanced with energy storage and other dispatchable supply [3]. Good understanding of the variability of the generation spatially across the electricity system over extended time periods is important to determine the scale of dispatchable support needed. Accurate time series of potential offshore wind generation are required for both energy cost and balancing cost estimation. Long term energy traces with hourly resolution are a key input to optimisation models used to determine potential renewable energy mixes [4–7].

A popular method for estimating wind resource time series is by using the output of global weather simulations, referred to as reanalysis. An atmospheric reanalysis system comprises a global forecast model, input observations and an assimilation scheme which are used in combination to produce best estimates of past atmospheric states [8]. A number of these models have been developed and provide information at different temporal and spatial resolutions [9–12].

Commercial wind farm modelling products such as WaSP and Vortex FDC use these large scale, relatively spatially coarse reanalysis wind outputs as the boundary input for downscaling to achieve higher spatial resolution using course terrain data. This mesoscale modelling is then used as an input for microscale modelling with high resolution topography. While this approach provides high accuracy, it is highly complex and hence typically only the domain for wind farm developers. However, the wider community of energy researchers, market operators and policy makers stand to benefit from easily accessible data on wind energy

* Corresponding author.
E-mail addresses: liam.hayes@anu.edu.au, liam.hayes400@gmail.com (L. Hayes), matthew.stocks@anu.edu.au (M. Stocks), andrew.blakers@anu.edu.au (A. Blakers).

https://doi.org/10.1016/j.energy.2021.120603
0360-5442/

resources. This gap has been partially filled by two projects. Firstly, Global Wind Atlas [13] provides high spatial resolution data on wind direction, speed and power at various times of day and seasons with the aim to help “policymakers, planners, and investors identify high-wind areas for wind power generation”. However, the Global Wind Atlas provides no timeseries data. Secondly, Renewables Ninja [14] provides 40 years of hourly capacity factor predictions based on NASA's MERRA2 reanalysis, which has been widely used since its release in 2016.

The ERA5 reanalysis, produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), provides the most accurate wind speed data for modelling wind farm generation [15]. When compared to MERRA2, ERA5 has a higher spatial resolution of 31 km compared to 50 km, and contains wind speeds at heights closer to that of modern wind turbines with 7 different heights below 200 m compared to just three heights of 2 m, 10 m and 50 m. ERA5 is the reanalysis of choice for the Global Wind Atlas 3.0. Further, it has been shown to outperform MERRA and MERRA2 for modelling wind farm generation [15], reducing mean absolute error in hourly generation by 24%. Yet to the best of our knowledge, no additional wind energy modelling research is based on ERA5 data.

The main objective of this paper is to build on the Renewables Ninja project by providing 40 years of hourly offshore wind generation based on ERA5, the current best reanalysis for wind modelling through the accompanying website. Further, we validate our data against offshore European wind farm generation and compare the model with Renewables Ninja as the baseline.

The accompanying website makes high quality wind speed data and offshore wind farm energy traces more available to energy researchers and policy makers. The hourly generation predictions allow the interactions between historical offshore wind potential and historical demand, supply and storage patterns to be analysed. It can also help to estimate the energy cost and cost of balancing of potential offshore wind projects and to show correlations and anti-correlations between wind generation at diverse locations. Although the ERA5 reanalysis data is publicly available, it is not highly accessible to those interested in wind farms, since downloads are optimised to provide data covering a large area (the whole globe) for a short time (such as month). A subset of ERA5 data can be accessed through the Copernicus Climate Change Service. However to access ERA5 wind speeds at the most appropriate altitudes, the much slower ECMWF's MARS tape archive must be used. Hence to obtain 40 years of hourly wind speeds for wind farm modelling, users must submit hundreds of data requests to the tape archive over several weeks. The website provided by this project opens up ERA5 wind speed data to many potential users and uses.

This paper deviates from previous similar studies by focusing exclusively on offshore wind farms, and validating predictions against the generation of individual wind farms, rather than a country's aggregated wind generation. This approach allowed the changing capacity of individual wind farms to be identified and hence improved the accuracy of the validation.

The remainder of the paper is structured as follows. Firstly, the reanalysis and wind farm data are described in section 2. In section 3, the methods used are presented. Our results and a discussion of our findings are found in section 4.

2. Data

Several different data sources were used in this study, including the ERA5 reanalysis dataset (section 2.1), offshore wind farm generation traces (section 2.2), wind farm metadata (section 2.3) and capacity factor predictions from Renewables Ninja (section 2.4) which were used as the baseline predictions.

2.1. ERA5 reanalysis

Reanalysis combines historic weather observations with an atmospheric weather model to estimate a previous state of the global weather system. There are several reanalysis products available that provide global hourly wind speeds. MERRA and MERRA2 by NASA were released in 2009 and 2015 respectively and have been used in numerous wind research studies since [16–32]. In 2019 the European Centre for Medium-Range Weather Forecasts (ECMWF) released ERA5 data for the period from 1979 to within 5 days of real time, which contains hourly estimates of a comprehensive list of atmospheric parameters including wind speed and direction. ECMWF will soon make available ERA5 reanalysis data dating back to 1950. ERA5 outperforms MERRA2 in wind energy modelling [15] and MERRA2 wind energy data has already been made widely available [23]. For these reasons we chose to use the ERA5 reanalysis in this study.

This study focuses on offshore wind farm generation and hence benefits from the higher accuracy of offshore wind speeds in reanalysis models. Wind speeds over the ocean are not affected by local topography. In addition, wind shear due to local ocean surface roughness is accounted for by the coupling of the ERA5 ocean and atmospheric components [10]. On the other hand, onshore wind speeds are highly dependent on local topography and surface roughness and many wind farms are built on hills or ridges to take advantage of localised speed ups. Since these topographic features are much smaller than the reanalysis resolution, their effects are not captured in the reanalysis wind data. Fig. 1 compares the effect of local topography on the generation of an offshore and an onshore wind farm.

The analysis in this study makes use of ERA5 hourly wind speeds at three atmospheric levels corresponding to heights of 54 m, 107 m and 170 m, for a period of six years (2014–2019). This time period was chosen to cover all generation data sourced from ENTSO-E [33] (see section 2.2). Further, 40 years (from 1980 to 2019) of ERA5 hourly wind speeds within 200 km of the coast and generation traces of offshore wind farms are made easily available for wind energy research on our accompanying website (windatlas.xyz).

2.2. Individual offshore wind farm generation

Power generation for individual offshore wind farms were sourced from the ENTSO-E Transparency Platform. ENTSO-E (European Network of Transmission System Operators for Electricity) hosts any transparency data made available by its 42 constituent transmission system operators, and is the largest source of offshore wind generation data. The generation traces used hourly or sub-hourly time intervals. For this study, sub-hourly traces were averaged to produce hourly traces.

Due to the more recent uptake of offshore wind generation and differences in data availability, large sources of hourly generation traces for individual offshore wind farms were not identified outside Europe. For example, China has roughly 50 operational offshore wind farms, but no publicly available generation data could be found.

The 57 hourly generation traces obtained from ENTSO-E comprised 44 in the UK, 7 in Belgium and 6 in Denmark. The 57 generation traces corresponded to 40 unique wind farms (see Fig. 2) since several of the UK traces corresponded to different sections of the same wind farm. For example, London Array has four separate feed-in points and four distinct generations traces. The separate wind farm sections were modelled independently and are considered separate wind farms for the purpose of this paper. All generation traces cover the period December 2014 to December 2019, or from whenever the wind farm was first operational.

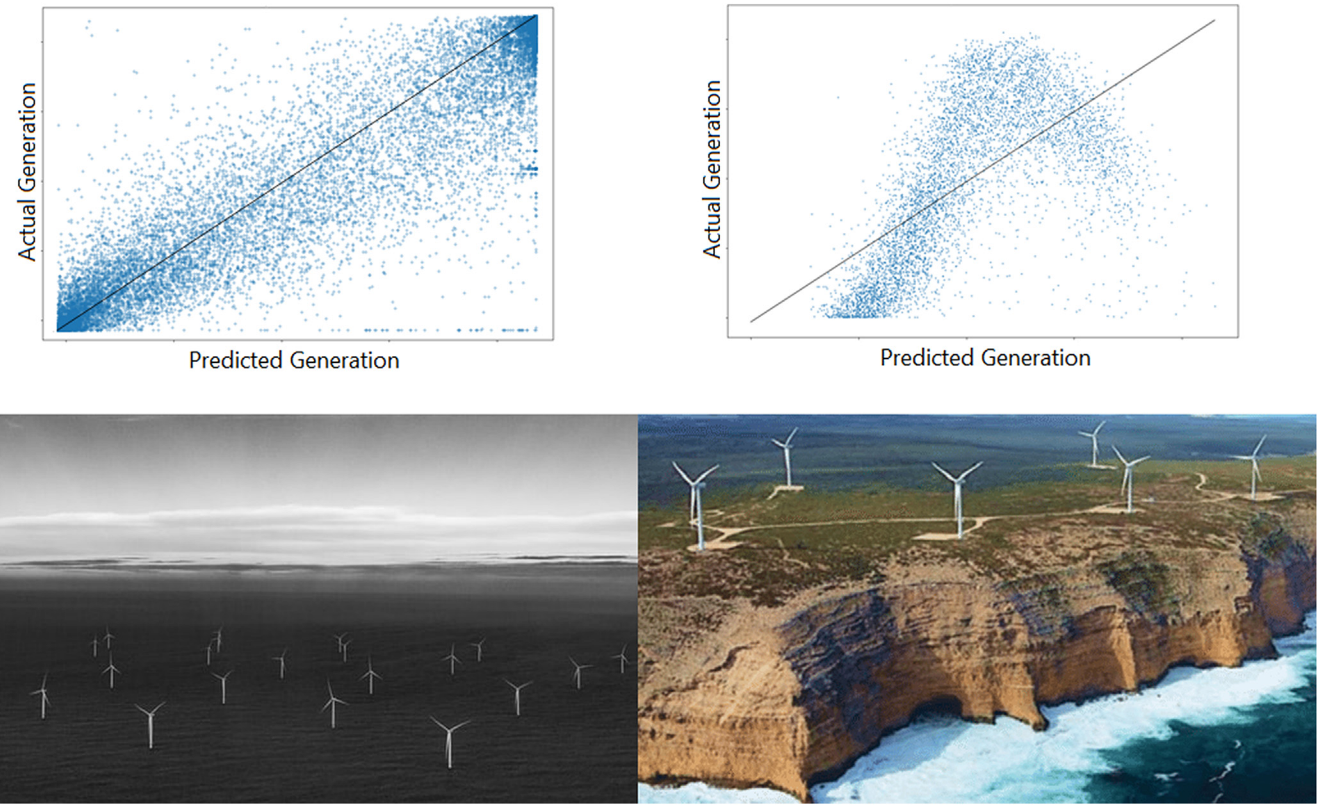

Fig. 1. The generation of Race Bank offshore wind farm (left) is highly predictable (photo by Nicholas Doherty on Unsplash). Cathedral Rocks wind farm in Australia (right) is built on top of a sea cliff which has a drastic effect on local wind speeds (photo from Coast Protection Board of South Australia). The plots show actual capacity factors against those predicted from ERA5 wind speeds. The Cathedral Rocks plot only shows generation for westerly winds to highlight the effect of the sea cliffs.

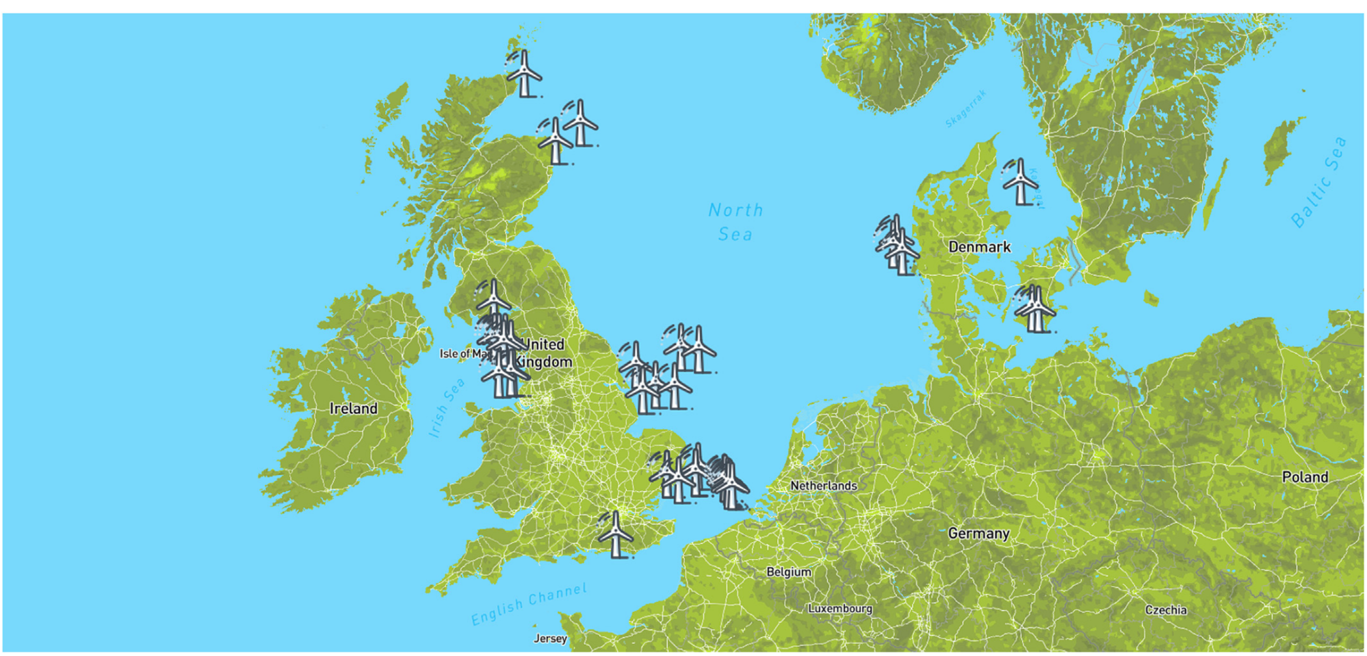

Fig. 2. Map of the 40 wind farms in this study.

2.3. Metadata

In order to compare ERA5 wind speeds with wind farm generation traces, the coordinates, hub height, and turbine power curve are needed for each wind farm. This information was obtained from thewindpower.net, with any gaps being filled from resources such as wikipedia.org and the wind farm developer websites. Where the power curve was not available, a power curve for a turbine with a similar power per swept area was used, preferably from the same manufacturer.

2.4. The Renewables Ninja

The Renewables Ninja website (renewables.ninja) provides 40 years of hourly wind speeds and capacity factors based on MERRA2 data. The website interpolates wind speeds to a specified location and extrapolates to a specified hub height. For 23 European countries it also applies wind speed correction factors (bias correction) to improve wind modelling accuracy. All the wind farms in this study are located in regions where the Renewables Ninja website applied bias correction.

For this study we downloaded the Renewables Ninja's bias corrected wind speeds, but not the capacity factors, instead we chose to calculate the capacity factors ourselves. The Renewables Ninja website dynamically calculates capacity factors based on the bias corrected wind speeds using the method described in the associated study [23] and its supplementary material. The paper states that capacity factors are calculated using power curves smoothed with a Gaussian filter of width $0.6 + 0.2w$ where w is the wind speed however the website's data is better recreated with a filter of twice that width ($1.2 + 0.4w$). We were able to accurately reproduce the capacity factors from the downloaded wind speeds when using this larger filter width, which confirmed we correctly followed the described methodology. The capacity factors calculated using the smaller filter width (as detailed in the paper) performed better than those calculated using the larger filter width (as downloaded from the website), hence we use the narrower filter with lower errors for this study. For example, for Thorntonbank 3 wind farm the root-mean-square error (RMSE) of predicted hourly capacity factors is 0.136 when using the smaller filter width, and 0.144 when using the larger filter width. A secondary advantage of calculating the Renewables Ninja capacity factors is that any turbine power curve can be used, rather than those on the predefined list available on the website. Applying our own power curves eliminated any discrepancy between power curves used on Renewables Ninja data and ERA5 data.

3. Method

This section describes how wind farm capacity factors were calculated from published generation (section 3.1) and how ERA5 wind speeds were used to predict the same capacity factors (section 3.2). Reasons for not adjusting to Global Wind Atlas data or correcting for spatial or temporal bias are also discussed.

3.1. Normalising offshore wind farm generation traces

The capacity factor, CF, of a wind farm at time t is defined as $CF(t) = \frac{G(t)}{C(t)}$, where $G(t)$ is the power generation and $C(t)$ is the capacity, or maximum possible output power, at time t.

One approach is to assume the wind farm's current capacity, C(t), is fixed at nameplate capacity from the date of commissioning [15,23]. Full downtime events can be detected and removed, as in Ref. [15]. However, this does not account for commissioning and partial downtime, where some of the wind farm is brought online or taken offline, an event found to be quite common. Fig. 3 shows that the Belwind II wind farm was brought online in stages and the London Array 1 wind farm experienced many events that limited its capacity without completely turning off. This highlights a significant source of inaccuracy when modelling wind generation at the country wide level and is the core reason we chose to validate our predictions against individual wind farms.

For this study, the wind farm capacity for time t was estimated from the generation trace to be the maximum power output in a 20-day rolling window centered at t, using Equation (1). The 20-day window (480 h) was chosen to balance underestimating the capacity during long periods of low wind, with overestimating the capacity for the first and last 20 days of a partial downtime event. This window size was chosen manually, by visually inspecting capacity estimates for several wind farms during extended periods of low wind speeds.

$$C(t) = max(G(t-x)),\ x \in \mathbb{Z}\ and\ |x| < = 240\ hours \tag{1}$$

3.2. Converting ERA5 wind speeds to capacity factors

The Virtual Wind Farm (VWF) method [23] was adapted to convert ERA5 wind speeds into hourly capacity factors for each farm. A summary of the method used and differences to the original VWF method is provided below.

For each location and hour, the wind speed at the wind farm's hub height is determined. Three wind speeds (w_1, w_2 and w_3) at three corresponding heights ($h_1 = 54$ m, $h_2 = 107$ m and $h_3 = 170$ m) are used to fit the following logarithmic relationship.

$$w = A + B \cdot ln(h) \tag{2}$$

The resulting constants, A and B, are used to calculate the wind speed (w_{hub}) experienced at the wind farm's hub height (h_{hub}).

The four surrounding spatial grid points in the ERA5 dataset are interpolated to the farm's location using inverse distance weighting, shown in Equation (3). A previous study [29] found there "may be some benefit in using … inverse distance weighting" for individual wind farms.

$$w_{farm} = \frac{\sum_{i=1}^{4} w_{hub,i} \Big/ d_i}{\sum_{i=1}^{4} 1 \Big/ d_i} \tag{3}$$

In Equation (2), $i=1,2,3,4$ represent the four surrounding grid points and d_i is the distance between the wind farm location and grid point i. This Equation provides w_{farm}, the wind speed for a particular hour interpolated to the appropriate height and location.

The order of the vertical and horizontal interpolation steps is reversed compared to the VWF method to reduce the data storage requirements of the published data. We found this ordering of interpolations had no significant impact on results.

We did not adjust our logarithmic fit to account for surface roughness, as done in the VWF method. As mentioned previously, the wind shear due to local ocean surface roughness is accounted for by the coupling of the ERA5 ocean and atmospheric components.

The interpolated wind speeds are converted to capacity factor using an adjusted turbine power curve. The power curve of the most appropriate turbine model for each wind farm (see section 2.3) is selected and then adjusted in a two step process, as visualised in Fig. 4. Firstly, the power curve for the wind farm's turbine model, $PC_{original}(w)$, is smoothed with a Gaussian filter of width $\sigma = 1.17$ m/s (Equation (4)) to simulate the distribution of wind speeds within a time period of an hour and between individual wind turbines in a farm. Secondly, the wind speeds are reduced by shifting the resulting power curve, $PC_{smoothed}(w)$, by a fixed value $w_0 = 0.71$ m/s (Equation (5)) to account for the wake loss experienced by turbines in a large wind farm. The adjusted power curve $PC_{adjusted}(w)$ is then employed to convert w_{farm}, the hourly wind speeds at the location and height of the wind farm, to a unitless

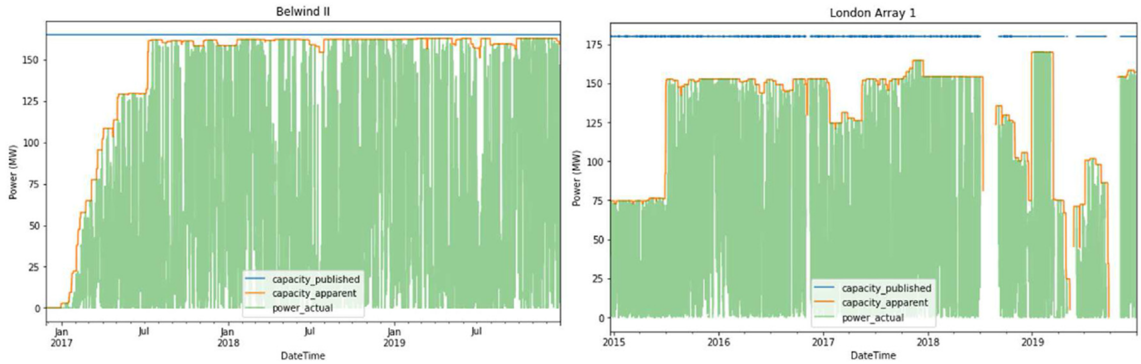

Fig. 3. The hourly generation (green) for two wind farms (Belwind II and London Array 1). The published capacity (blue) and apparent operating capacity (orange) are also shown. Note that the graphs present all the available generation and are not cropped to show a particular time period. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

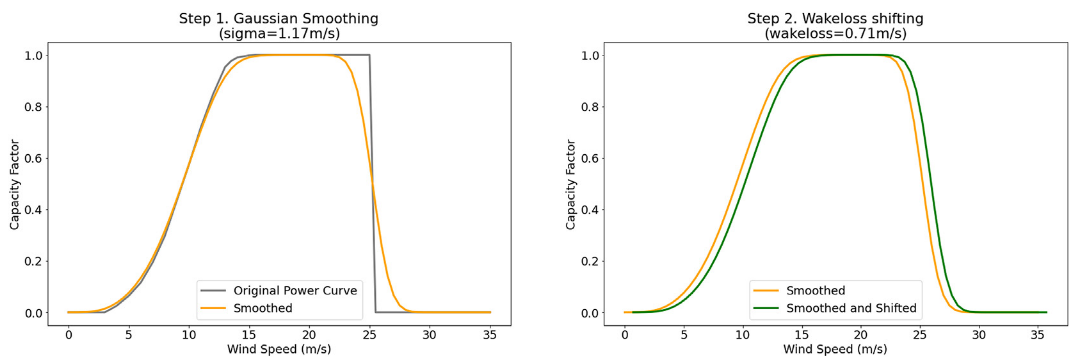

Fig. 4. Graphical representation of power curve adjustments applied to the power curve of the Vestas V112-3000 turbine.

capacity factor using linear interpolation (Equation (6)). The final value is the capacity factor predicted for our model for a particular location, hub height, historical hour, and turbine model.

$$PC_{smoothed}(w) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \sum_{x=w-4\sigma}^{w+4\sigma} PC_{original}(x) \cdot exp\left(-\frac{x^2}{2\sigma^2}\right) \text{ where } \sigma = 1.17\ m\Big/s \tag{4}$$

$$PC_{adjusted}(w) = PC_{smoothed}(w - w_0) \text{ where } w_0 = 0.71\ m\big/s \tag{5}$$

$$CF = PC_{adjusted}\left(w_{farm}\right) \tag{6}$$

This 'sigma-wakeloss' model was chosen for its ability to produce good capacity factor predictions, while remaining simple and physically explainable. The addition of further parameters, such as allowing sigma and wakeloss to vary linearly with wind speed, did little to improve the accuracy of the model but increased the difficulty of generalising appropriate parameter values to offshore locations outside this study.

The values of σ and w_0 were calculated to minimise the average duration curve error of the 57 wind farms in this study. The duration curve is a sorted sequence of the hourly capacity factors across the whole operational life of the wind farm, an example is presented in Fig. 5. Equation (7) shows how the RMSE of the duration curve is calculated for each wind farm. The model parameters, σ and w_0 were calculated to minimise the duration curve RMSE averaged across the 57 wind farms. Using the RMSE of the hourly capacity factor timeseries, instead of the duration curve, leads to systematic overestimation during periods of low capacity factors while high capacity factors are underestimated.

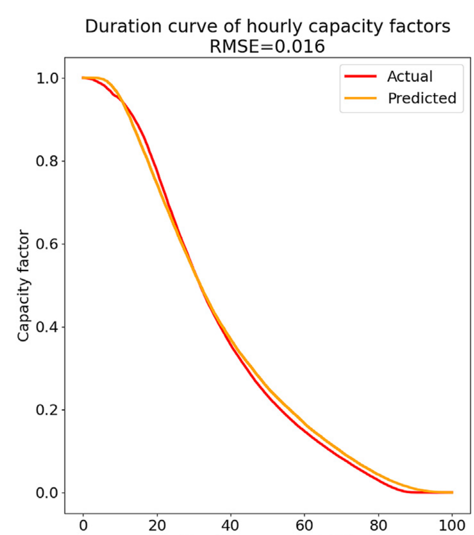

Fig. 5. An example of a duration curve, with the calculated RMSE shown in the title.

$$RMSE_{dcurve} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} \left[dcurve_{predicted}(x_i) - dcurve_{actual}(x_i) \right]^2} \tag{7}$$

We acknowledge that the wakeloss parameter does not necessarily suggest that offshore wind farms have slower wind speeds due to the wake effect by 0.71 m/s on average. The wakeloss parameter could be adjusting for a systematic error between ERA5 wind speeds and wind speed experienced by the actual wind farm. It could also be correcting for alternative energy losses such as turbine friction. Realistically it's likely to be a combination of the above, however the true source isn't critical. The key point is that the wakeloss parameter improves the accuracy of the model.

3.3. Bias correcting using Global Wind Atlas

We chose not to bias correct our wind speeds using data obtained from the Global Wind Atlas, as done in several previous studies [24,28,29,32,34]. The Global Wind Atlas provides the average wind speed at 250 m resolution, calculated using microscale fluid modelling which captures the effects of local topography at a fine scale. Adjusting ERA5 wind speeds to match the average wind speed from the Global Wind Atlas could, in principle, improve the accuracy of hourly wind speeds. The trade off is that the fluid modelling is done in separate 300 km tiles, resulting in sharp changes to the calculated average wind speeds at the boundaries of up to 15% (see Fig. 6). For most onshore locations this trade off is likely worthwhile. However, offshore wind speeds are mostly unaffected by local topography, hence the benefits of using Global Wind Atlas corrections are outweighed by the boundary uncertainties.

3.4. Seasonal and diurnal bias correction

Another bias correction method involves correcting for seasonal and diurnal bias in reanalysis wind speeds by applying monthly and hourly wind speed correction factors as in Ref. [21]. However, correction factors calculated directly from the generation data used in this study could lead to overfitting our model to the wind patterns of northern Europe. The Global Wind Atlas also provides seasonal and diurnal wind speed patterns, but this data can't be downloaded for the whole globe, only manually retrieved for location of interest. Further, Gruber et al. [29] concluded that adjusting for seasonal and diurnal trends did not improve results.

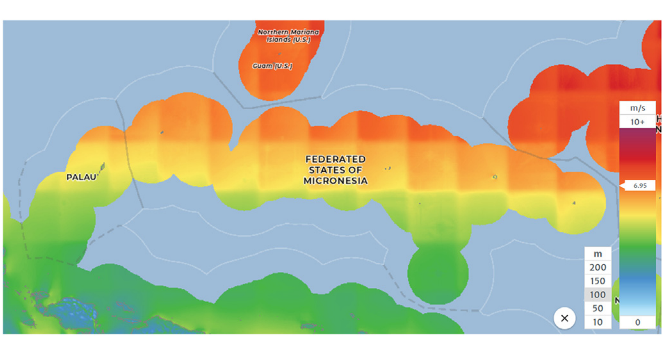

Fig. 6. A screenshot of the Global Wind Atlas showing a tiling effect in the average wind speed data over Micronesia.

3.5. Spatial bias

We did not attempt to correct for spatial bias in our model by allowing the wakeloss and sigma parameters to be regionally dependent. Spatial bias, in terms of reanalysis, is where a meteorological parameter is overestimated in one area, and underestimated in another, on a scale larger than the resolution of the reanalysis. The Renewables Ninja model accounts for spatial bias in the MERRA2 wind speeds in Europe using country dependent model parameters. Our model achieves a higher level of accuracy than the Renewables Ninja model for wind farms across Europe without the model parameters, wakeloss and sigma, being location dependent. This demonstrates that any spatial bias present in ERA5 wind speeds in Europe is not of great concern for offshore wind modelling.

4. Results

4.1. Validation results

This section compares the capacity factors predicted by our model and Renewables Ninja against the 57 ENTSO-E generation traces for the period from December 2014 to December 2019. The ENTSO-E generation trace for each farm was converted into an hourly capacity factor timeseries (section 3.2) which was used as the ground truth. The ERA5 wind speeds were converted to a second hourly capacity factor timeseries for each farm using that farm's location, hub height and turbine power curve (section 3.2), this is our model's prediction. The final hourly capacity factor timeseries was sourced from the Renewables Ninja website (section 2.4), this is the baseline prediction that we compare our model's performance against.

The results are presented as RMSE (root mean square error), which is based on the difference between the predicted and actual capacity factors (0.1 indicates an average root mean squared difference of 0.1, not a difference of 10%). The RMSE was calculated for the hourly capacity factors, daily capacity factors and the hourly capacity factor duration curve.

Table 1 summaries the mean and standard deviation of the RMSE results. Fig. 7 displays the RMSE results for each of the 57 wind farms. Our model, using ERA5 wind speeds, consistently outperforms the Renewables Ninja, based on MERRA2 wind speeds. Our model produces better hourly and daily capacity factor predictions and duration curves for the vast majority of wind farms.

4.2. Annual capacity factors predictions

In this section we present the accuracy of our long run capacity factor predictions for the 57 wind farms. The validation presented in section 4.1 focuses on hourly generation since accurate time series of potential offshore wind generation are important in energy system modelling. However, hourly timeseries are not always required by those using wind energy data; it is also important to know the accuracy of longer term averages.

Fig. 8 shows the difference between the predicted annual capacity factors and the actual annual capacity factors for each wind farm. This also represents the expected under- or overprediction of any given hourly capacity factor prediction for a given wind farm (the difference of annual averages equals the annual average of hourly differences). Our annual generation predictions have an error of 7.0% on average, whereas the average error in the Renewables Ninja predictions is 11.2%. Fig. 8 also shows that Renewables Ninja predictions tend to overestimate annual capacity factors (for 51 out of 57 of the wind farms).

Table 1
Summary of capacity factor prediction results.

RMSE	Our model (ERA5)	Renewables Ninja (MERRA2)	Average percentage Improvement in RMSE	Improvement (number of traces)
Hourly capacity factor	0.17 ± 0.03	0.19 ± 0.03	10.3%	54 out of 57
Daily capacity factor	0.11 ± 0.03	0.13 ± 0.03	18.2%	53 out of 57
Duration curve	0.05 ± 0.02	0.11 ± 0.02	57.0%	55 out of 57

Note that if the prediction error is purely random, one would expect the daily RMSE values to be smaller than the hourly RMSE values by the factor $\sqrt{24}$, since the predictions are aggregated over 24 time periods. Since this is not the case, the prediction error is predominantly due to model bias, indicating that a model tailored to an individual wind farm would perform much better than our more generalised model.

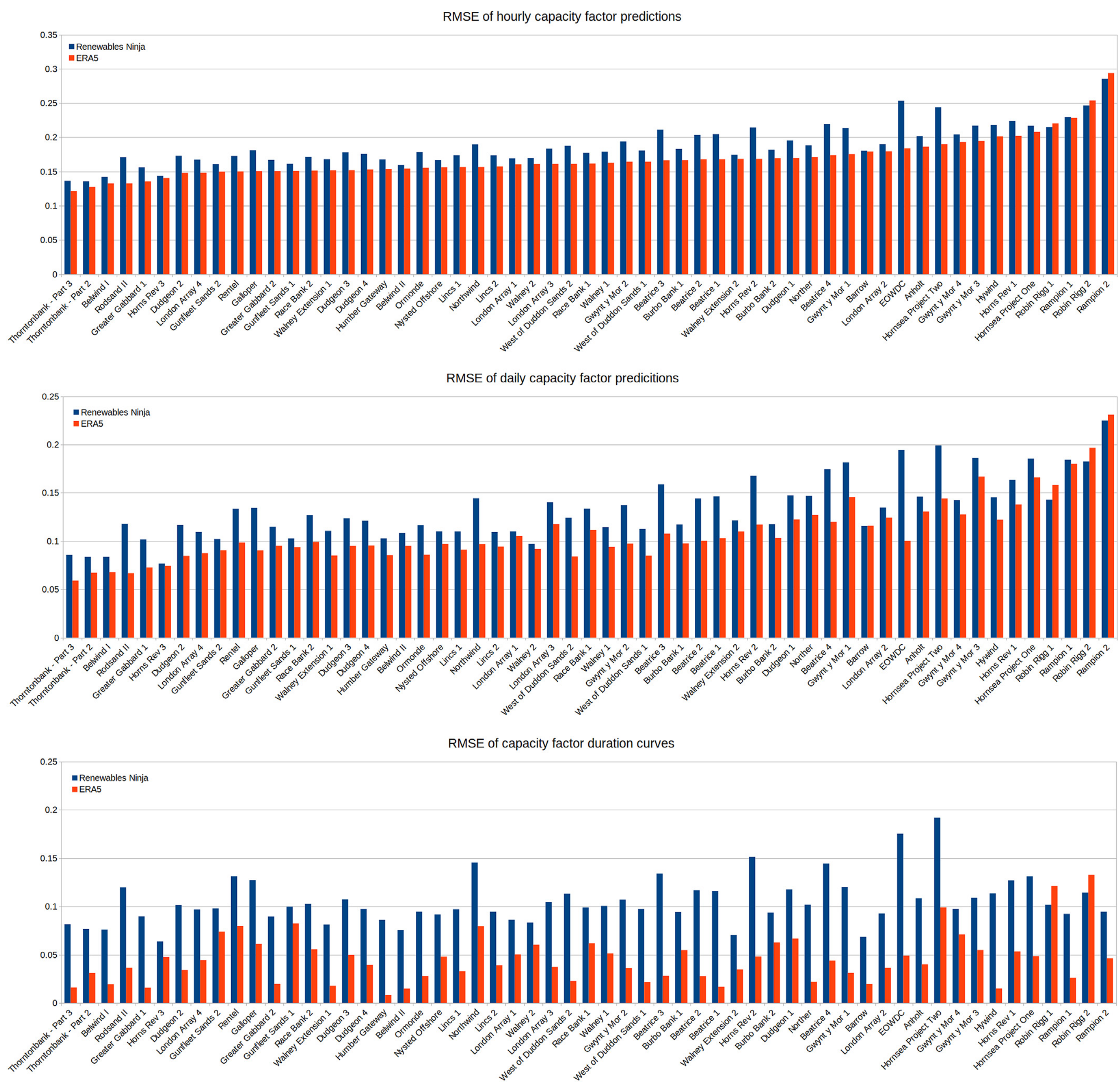

Fig. 7. Comparison of results. The RMSE of our predictions is shown in red, the RMSE of the baseline (Renewables Ninja) is shown in blue for the 57 offshore wind farms used in this study. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

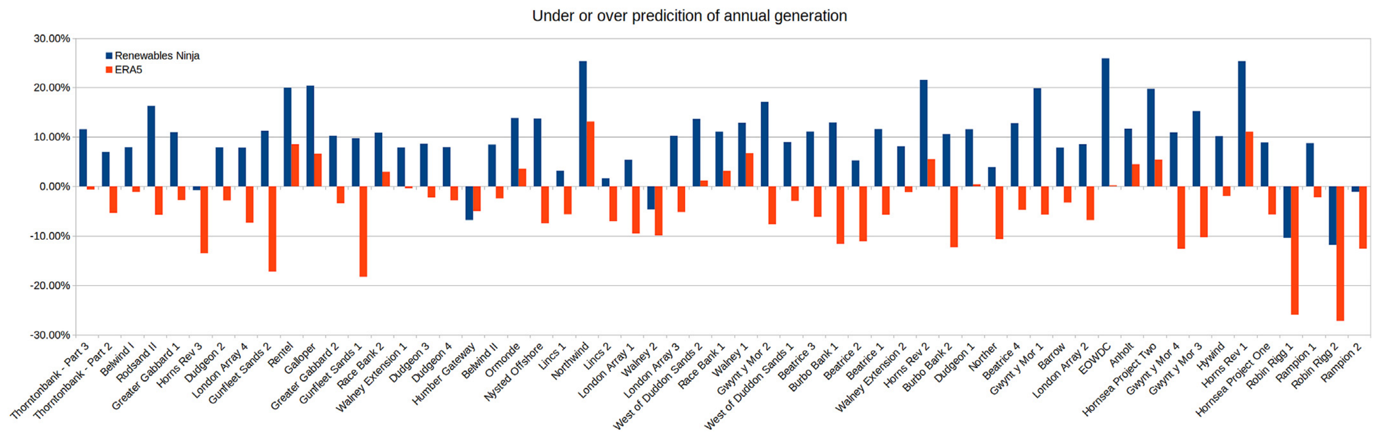

Fig. 8. The difference between the actual and predicted annual capacity factors for the 57 wind farms, presented as a percentage. A value of −10% indicates the prediction underestimates annual generation by 10%. The wind farms are presented in the same order as previous plots (our best hourly predictions on the left).

4.3. Aggregated generation predictions

In this section we present the accuracy of the aggregated hourly generation of all the wind farms. This allows our results to be more directly compared to previous studies that consider country or region wide aggregated wind generation instead of individual wind farm generation.

For each hour the capacity factors of each wind farm were averaged, weighted by the nameplate capacity of the wind farms. This was compared to the capacity-weighted average of the actual capacity factors. Where the generation data for a particular wind farm was not available for a particular time period (due to missing data or the time period was before the wind farm was commissioned), that wind farm was also excluded from the prediction. In any given hour, an average of 24 out of the 57 wind farms did not have generation data, leaving the generation of 33 wind farms to be aggregated.

For our model the RMSE of the hourly aggregated capacity factors is 0.064. This is much less than 0.17, the average RMSE for non-aggregated capacity factors, since much of the random error is reduced when averaging across many wind farms. The result is notably less than the RMSE of the Renewables Ninja prediction across the same data set, which is 0.089.

Our aggregate result is more directly comparable to the ERA5 wind generation study by Olauson [15]. The study presented the RMSE for hourly wind farm capacity factors aggregated across various regions based on ERA5 wind speeds. The lowest error (RMSE of 0.0235) corresponds to Germany, the region with the most wind farms (about 10443 according to windpower.net). The highest error (RMSE of 0.091) corresponds to a region in the north west USA, which has the least number of wind farms (about 475 wind farms). Part of this can be explained by the diminishing random error when aggregating across independent wind farms. Our aggregated result (RMSE of 0.064) compares favourably to the USA region, with much fewer wind farms.

4.4. Wind farm case study

In this section we provide a closer look at two wind farms, Thorntonbank, for which our predictions are most accurate, and Robin Rigg, the only wind farm for which our model produces worse results than the Renewables Ninja.

4.4.1. Thorntonbank wind farm

Thorntonbank wind farm is 30 km NW of the flat Belgium coastline, with 54 6.15 MW turbines with a hub height of 95 m. Fig. 9 compares the predicted hourly and daily capacity factors against the actual values and shows the predicted duration curve. Apart from a slightly underprediction of high capacity factors and overprediction of low capacity factors our model generates accurate results.

4.4.2. Robin Rigg wind farm

Robin Rigg wind farm is located within the Solway Firth, with the hilly Scottish coastline 10 km to the NW and England 13 km to the SE. It contains 58 3 MW turbines, with a hub height of 80 m.

Our model significantly underpredicts generation at Robin Rigg wind farm, as seen in Fig. 10(a). This is due to the effects of onshore topography. Robin Rigg's proximity (10 km) to a hilly region is likely to have an effect on local wind speeds that are not captured by the coarse 31 km resolution of ERA5. A wind rose from the Global Wind Atlas suggests that wind is funnelled up the bay. Fig. 11 shows that three of the four surrounding ERA5 grid points used to interpolate wind speeds to Robin Rigg's location are actually onshore. If wind speeds are not interpolated but instead taken from the location of the closest *offshore* grid point, our model performs much better. In fact, the RMSE of the hourly capacity factors, daily capacity factors and duration curve reduces by 11%, 15% and 80% respectively. This demonstrates the significant impact of onshore topography to wind energy modelling, and the improved accuracy of offshore wind farm generation predictions.

4.5. Wind farms outside Europe

In this section we validate our model against a wind farm outside of Europe. Our model for prediction wind farm generation is not based on location sensitive parameters, so can be applied globally. However, the model was created using only offshore wind farm data from the North Sea, hence validation against a wind farm outside this region is important. Unfortunately, no hourly generation traces for non-European offshore wind farms could be found, so the monthly generation of Block Island Wind Farm was sourced from the Electricity Data Browser of the US Energy Information Administration [35].

Block Island Wind Farm is the USA's only operating offshore wind farm, located in the north east. It consists of 5 turbines, each turbine is a 6 MW Haliade 150 with a hub height of 100 m, and commenced production in November 2016. Fig. 12 compares its generation to our monthly capacity factor predictions. Partial downtime events could not be accounted for, since hourly

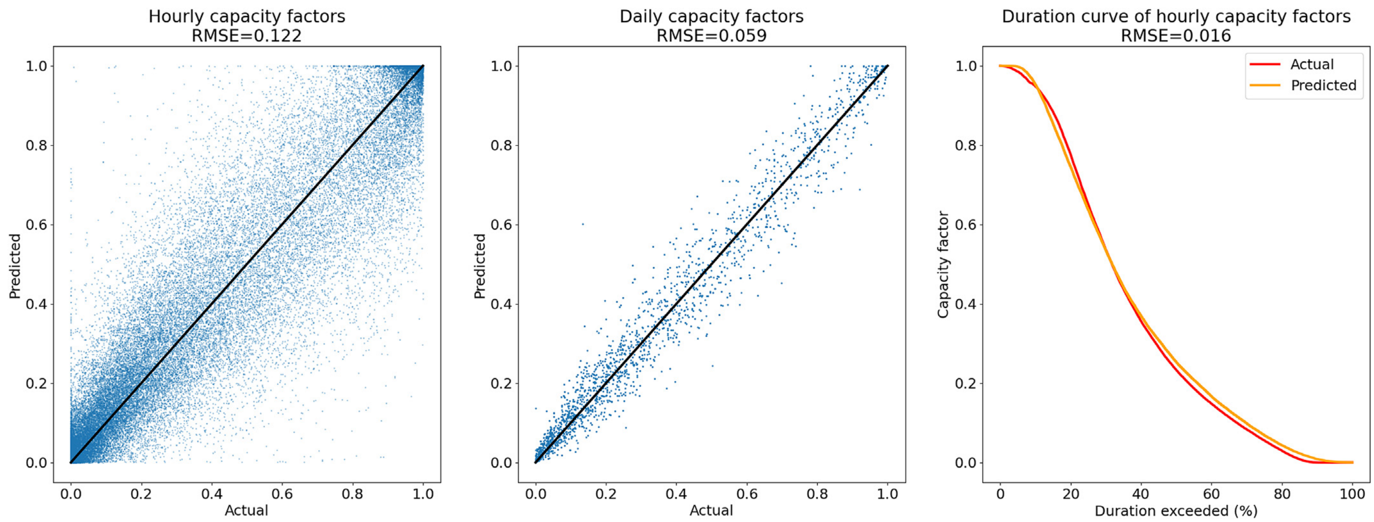

Fig. 9. Thorntonbank 3. Predicted hourly and daily capacity factors against actual values (left and middle), and predicted and actual duration curves (right).

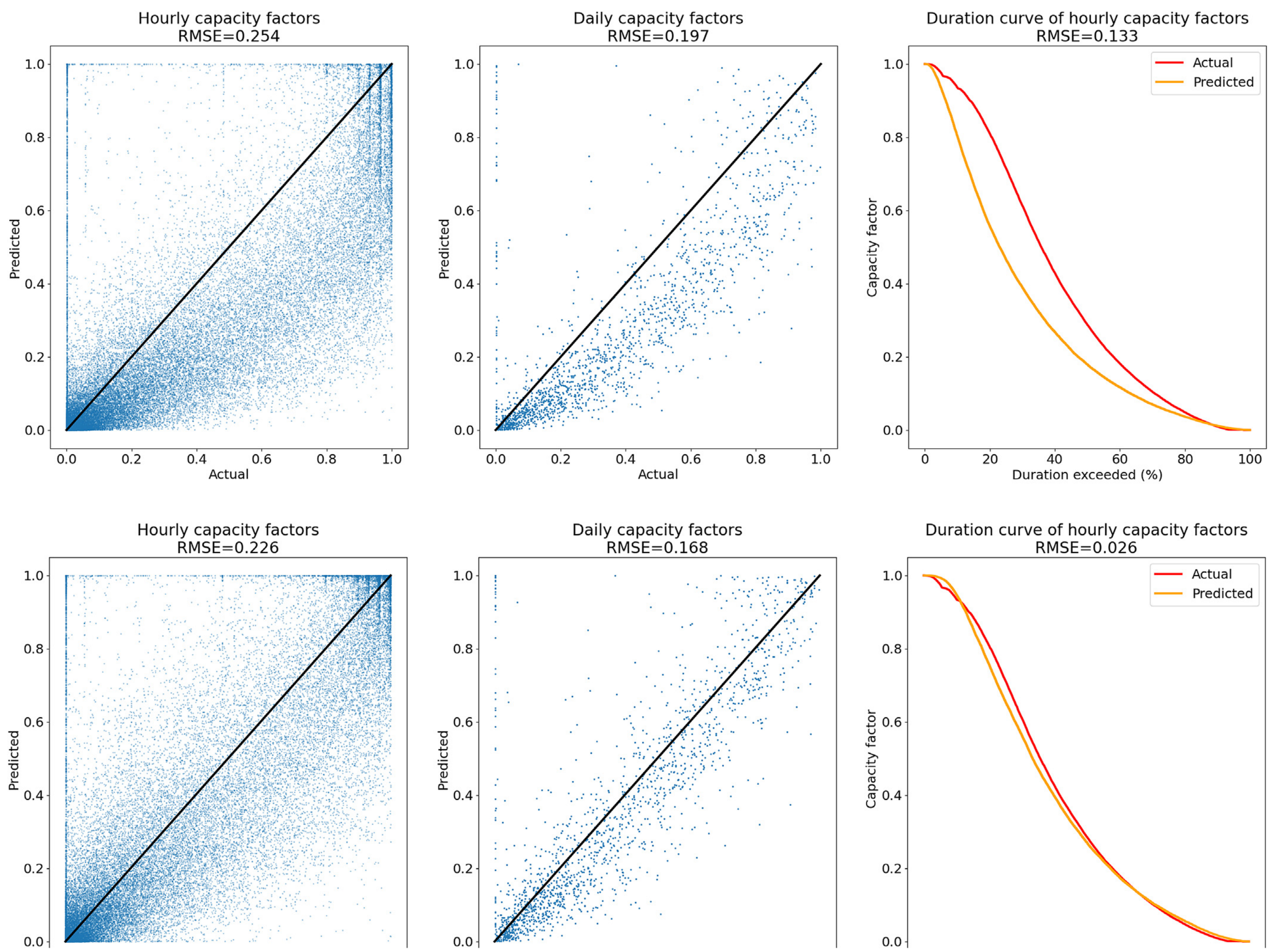

Fig. 10. Robin Rigg 2. (a) our model results when wind speeds from the four surrounding ERA5 grid points is interpolated as described in Section 3.2. (b) our model results when wind speeds are taken from the closest offshore grid point.

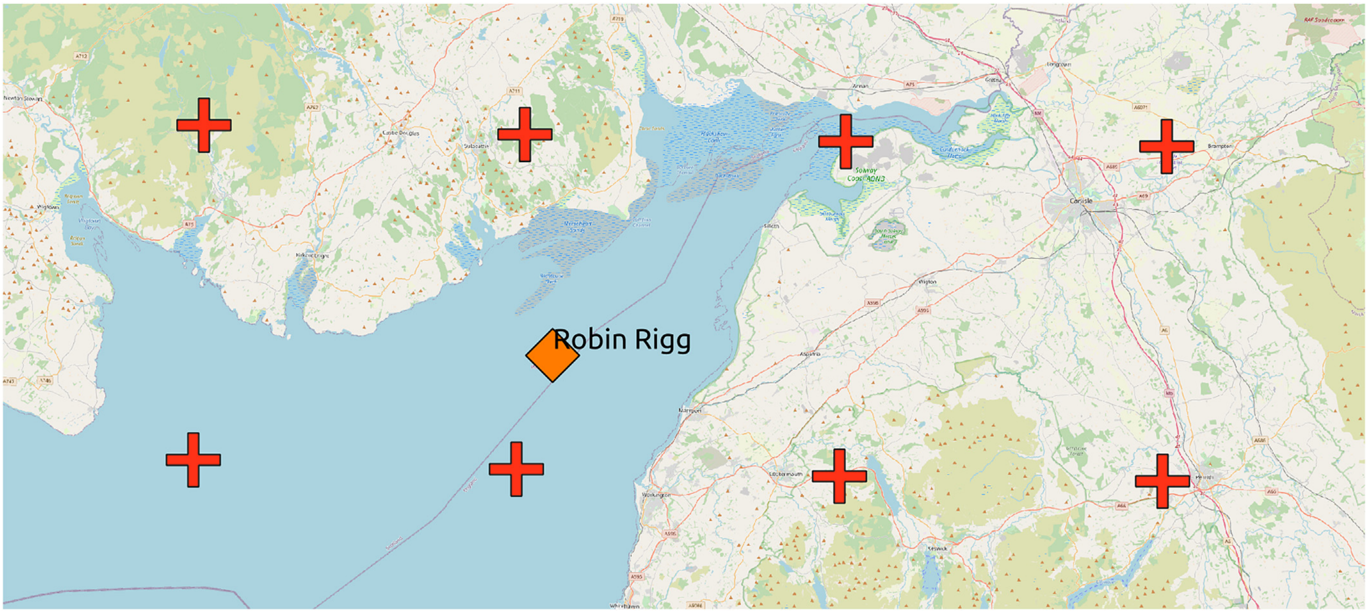

Fig. 11. Location of Robin Rigg wind farm in relation to surround ERA5 grid points.

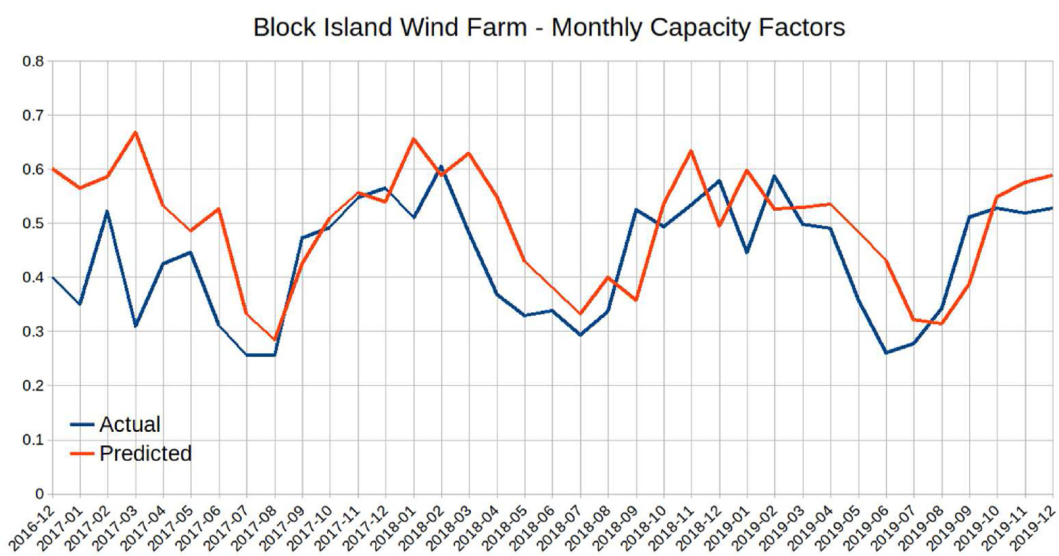

Fig. 12. Monthly capacity factors of Block Island Wind Farm.

generation data is not available, but it appears that the wind farm was not operating at full capacity for the first several months during commission, similar to Belwind II wind farm in Fig. 3. For this reason, the first 7 months of generation data was ignored.

Our model produces reasonably accurate monthly capacity factors for Block Island, with an RMSE of 0.091. Our model produces more accurate monthly capacity factors for Block Island than 8 out of 57 of the European wind farms (average RMSE of 0.065). It is also more accurate than 26 out of 57 of the monthly capacity factor predictions based on the Renewables Ninja (average RMSE of 0.091). This is an encouraging result, especially given that we were not able to identify any partial downtime (or full downtime) events in the Block Island generation data. Users should be aware, however, that this data has not been validated for the RMSE of the duration curves outside Europe and would benefit from testing against offshore data in other regions if suitable data becomes available in future.

5. Conclusion

In this work, we presented a simple model for predicting long-term, hourly offshore wind farm generation from wind speeds extracted from ERA5, a global atmospheric reanalysis. The model was tested against the hourly generation of 40 different offshore wind farms in Europe, and the monthly output for an offshore wind farm in the United States. The quality of the fit was compared to the outputs from the Renewables Ninja project, which uses a different but similar model to predict wind generation from MERRA2 wind

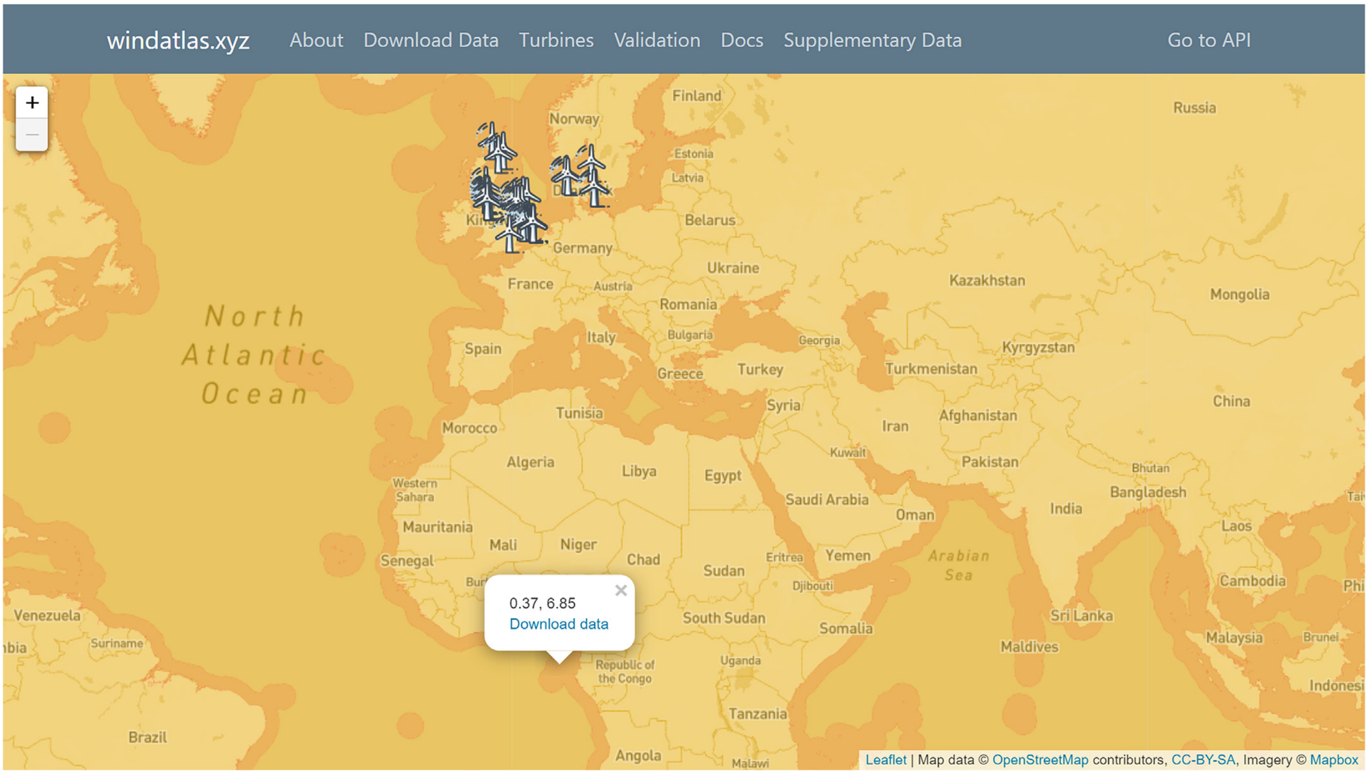

Fig. 13. Screenshot of accompanying website.

speeds. The errors in our hourly and daily capacity factors in our model were on average 10% and 18% lower respectively.

We found that analysing the generation of individual wind farms, rather than the aggregated generation of many wind farms, allows partial downtime events to be identified and accounted for in the capacity factor analysis. These partial downtime events, where the current capacity of the wind farms is limited to less than the nameplate capacity, are frequent and could be a significant source of error when analysing countrywide wind generation.

Our modelled wind generation predictions are available through an API on our offshore wind atlas website, along with 40 years of hourly ERA5 wind speeds. When the remaining ERA5 data is released in 2021, it will be possible to extend this to 71 years of hourly wind speeds. We hope this will be of use to renewable energy researchers, organisations and policymakers.

Data availability

All plots, data and analysis can be found on our accompanying website: http://windatlas.xyz/. The website provides easy access to 40 years of hourly offshore ERA5 wind speed within 200 km of the coast. It also calculates hourly wind generation predictions using a user selected power curve and the methods described in this paper. Power curves from 226 turbine models are supplied, ranging from very low wind (IEC Class IV) to high wind (IEC Class I), so that a turbine appropriate to the local wind environment can be modelled.

Fig. 13 shows an example screenshot of the website map.

Author credit statement

Liam Hayes: Methodology, Software, Validation, Writing - Original Draft **Matthew Stocks:** Conceptualization, Supervision, Writing - Review & Editing **Andrew Blakers:** Writing - Review & Editing, Funding acquisition.

Funding

This work was funded in part by the ANU Grand Challenge, Zero-carbon Energy for the Asia-Pacific and the Australian Renewable Energy Agency project, STORES-2. Responsibility for the views, information or advice expressed herein is not accepted by the Australian Government.

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.energy.2021.120603.

References

[1] GWEC-offshore-wind-2020-5.pdf [Online]. Available, https://gwec.net/wp-content/uploads/dlm_uploads/2020/08/GWEC-offshore-wind-2020-5.pdf. [Accessed 28 September 2020].
[2] New plans to make UK world leader in green energy. https://www.gov.uk/government/news/new-plans-to-make-uk-world-leader-in-green-energy. [Accessed 10 November 2020].
[3] Kroposki B. Integrating high levels of variable renewable energy into electric power systems. J. Mod. Power Syst. Clean Energy 2017;5(6):831–7. https://doi.org/10.1007/s40565-017-0339-3.
[4] Blakers A, Lu B, Stocks M. 100% renewable electricity in Australia. Energy 2017;133:471–82. https://doi.org/10.1016/j.energy.2017.05.168.
[5] Pleßmann G, Erdmann M, Hlusiak M, Breyer C. Global energy storage demand

for a 100% renewable electricity supply. Energy Procedia 2014;46:22–31. https://doi.org/10.1016/j.egypro.2014.01.154.
[6] Steinke F, Wolfrum P, Hoffmann C. Grid vs. storage in a 100% renewable Europe. Renew Energy 2013;50:826–32. https://doi.org/10.1016/j.renene.2012.07.044.
[7] Connolly D, Lund H, Mathiesen BV, Leahy M. The first step towards a 100% renewable energy-system for Ireland. Appl Energy 2011;88(2):502–7. https://doi.org/10.1016/j.apenergy.2010.03.006.
[8] Fujiwara M, et al. Introduction to the SPARC reanalysis intercomparison project (S-RIP) and overview of the reanalysis systems. Atmos Chem Phys 2017;17(2):1417–52. https://doi.org/10.5194/acp-17-1417-2017.
[9] Rienecker MM, et al. “MERRA: NASA's modern-Era retrospective analysis for research and applications. J Clim 2011;24(14):3624–48. https://doi.org/10.1175/JCLI-D-11-00015.1.
[10] Hersbach H, et al. The ERA5 global reanalysis. Q J R Meteorol Soc 2020;146(730):1999–2049. https://doi.org/10.1002/qj.3803.
[11] Kobayashi S, et al. The JRA-55 reanalysis: general specifications and basic characteristics. J. Meteorol. Soc. Jpn. Ser II 2015;93(1):5–48. https://doi.org/10.2151/jmsj.2015-001.
[12] Saha S, et al. The NCEP climate forecast system reanalysis. Bull Am Meteorol Soc 2010;91(8):1015–58. https://doi.org/10.1175/2010BAMS3001.1.
[13] Global Wind Atlas. https://globalwindatlas.info/. [Accessed 23 September 2020].
[14] Renewables.ninja. https://www.renewables.ninja/. [Accessed 28 September 2020].
[15] Olauson J. ERA5: the new champion of wind power modelling? Renew Energy 2018;126:322–31. https://doi.org/10.1016/j.renene.2018.03.056.
[16] Kubik ML, Brayshaw DJ, Coker PJ, Barlow JF. Exploring the role of reanalysis data in simulating regional wind generation variability over Northern Ireland. Renew Energy 2013;57:558–61. https://doi.org/10.1016/j.renene.2013.02.012.
[17] Huber M, Dimkova D, Hamacher T. Integration of wind and solar power in Europe: assessment of flexibility requirements. Energy 2014;69:236–46. https://doi.org/10.1016/j.energy.2014.02.109.
[18] Staffell I, Green R. How does wind farm performance decline with age? Renew Energy 2014;66:775–86. https://doi.org/10.1016/j.renene.2013.10.041.
[19] Cannon DJ, Brayshaw DJ, Methven J, Coker PJ, Lenaghan D. Using reanalysis data to quantify extreme wind power generation statistics: a 33 year case study in Great Britain. Renew Energy 2015;75:767–78. https://doi.org/10.1016/j.renene.2014.10.024.
[20] Drew D, Cannon D, Brayshaw D, Barlow J, Coker P. The impact of future offshore wind farms on wind power generation in great Britain. Resources 2015;4(1):155–71. https://doi.org/10.3390/resources4010155.
[21] Olauson J, Bergkvist M. Modelling the Swedish wind power production using MERRA reanalysis data. Renew Energy 2015;76:717–25. https://doi.org/10.1016/j.renene.2014.11.085.
[22] Olauson J, Bergkvist M. Correlation between wind power generation in the European countries. Energy 2016;114:663–70. https://doi.org/10.1016/j.energy.2016.08.036.
[23] Staffell I, Pfenninger S. Using bias-corrected reanalysis to simulate current and future wind power output. Energy 2016;114:1224–39. https://doi.org/10.1016/j.energy.2016.08.068.
[24] Bosch J, Staffell I, Hawkes AD. Temporally explicit and spatially resolved global offshore wind energy potentials. Energy 2018;163:766–81. https://doi.org/10.1016/j.energy.2018.08.153.
[25] Cradden LC, McDermott F, Zubiate L, Sweeney C, O'Malley M. A 34-year simulation of wind generation potential for Ireland and the impact of large-scale atmospheric pressure patterns. Renew Energy 2017;106:165–76. https://doi.org/10.1016/j.renene.2016.12.079.
[26] Johansson V, et al. “Value of wind power – implications from specific power. Energy 2017;126:352–60. https://doi.org/10.1016/j.energy.2017.03.038.
[27] Monforti F, Gonzalez-Aparicio I. Comparing the impact of uncertainties on technical and meteorological parameters in wind power time series modelling in the European Union. Appl Energy 2017;206:439–50. https://doi.org/10.1016/j.apenergy.2017.08.217.
[28] Bosch J, Staffell I, Hawkes AD. Temporally-explicit and spatially-resolved global onshore wind energy potentials. Energy 2017;131:207–17. https://doi.org/10.1016/j.energy.2017.05.052.
[29] Gruber K, Kloeckl C, Regner P, Baumgartner J, Schmidt J. Assessing the Global Wind Atlas and local measurements for bias correction of wind power generation simulated from MERRA-2 in Brazil [Online]. Available: ” ArXiv190413083 Stat; 2019. http://arxiv.org/abs/1904.13083. [Accessed 21 September 2020].
[30] Ren G, Wan J, Liu J, Yu D. Characterization of wind resource in China from a new perspective. Energy 2019;167:994–1010. https://doi.org/10.1016/j.energy.2018.11.032.
[31] Ren G, Wan J, Liu J, Yu D. Spatial and temporal assessments of complementarity for renewable energy resources in China. Energy 2019;177:262–75. https://doi.org/10.1016/j.energy.2019.04.023.
[32] Ryberg DS, Caglayan DG, Schmitt S, Linßen J, Stolten D, Robinius M. The future of European onshore wind energy potential: detailed distribution and simulation of advanced turbine designs. Energy 2019;182:1222–38. https://doi.org/10.1016/j.energy.2019.06.052.
[33] ENTSO-E transparency Platform. https://transparency.entsoe.eu/dashboard/show. [Accessed 28 September 2020].
[34] González-Aparicio I, et al. Simulating European wind power generation applying statistical downscaling to reanalysis data. Appl Energy 2017;199:155–68. https://doi.org/10.1016/j.apenergy.2017.04.066.
[35] US Energy Information Administration, “Electricity data browser - Block Island wind farm.”.