

ERA5: The new champion of wind power modelling?

Jon Olauson ^{a, b, *}

^a Division of Electricity, Department of Engineering Sciences, Uppsala University, Sweden

^b Integration of Renewable Energy Sources Group, KTH Royal Institute of Technology, Stockholm, Sweden



ARTICLE INFO

Article history:

Received 27 September 2017

Received in revised form

28 February 2018

Accepted 22 March 2018

Available online 26 March 2018

Keywords:

ERA5

MERRA-2

Wind power modelling

Long-term correction

Reanalyses

ABSTRACT

Output from meteorological reanalyses are used extensively in both academia and industry for modelling wind power. Recently, the first batch of the new ERA5 reanalysis was released. The main purpose of this paper is to compare the performance of ERA5 and MERRA-2 (a commonly used reanalysis today) in terms of modelling i) the aggregated wind generation in five different countries and ii) the generation for 1051 individual wind turbines in Sweden. The modelled wind power generation was compared to measurements. In conclusion, ERA5 performs better than MERRA-2 in all analysed aspects; correlations are higher, mean absolute and root mean square errors are in average around 20% lower and distributions of both hourly data and changes in hourly data are more similar to those for measurements. It is also shown that the uncertainty related to long-term correction (using one year of measurements and reanalysis data to predict the energy production during the remaining 1–5 years) is 20% lower for ERA5. In fact, using one year sample data and ERA5 gives slightly more accurate estimates than using two years of sample data and MERRA-2. Additionally, a new metric for quantifying the system size and dispersion of wind farms is proposed.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

Quoting Ref. [1], “reanalysis is the process whereby an unchanging data assimilation system is used to provide a consistent reprocessing of meteorological observations, typically spanning an extended segment of the historical data record”. The products from a reanalysis include many variables such as wind speeds, temperature and atmospheric pressure and are used extensively for modelling wind power both in academia and industry. Examples of applications include generation of long wind power time series for grid integration studies and long-term correction (LTC) of wind speed measurements conducted prior to the construction of wind farms. An important advantage of reanalyses is that these are often freely and globally available.

Over the decades, new reanalyses have gradually become more high-resolved in time and space [2]. During the last years, MERRA [3] and MERRA-2 [1] (both produced by NASA) have been very popular for modelling wind power generation [4–14] due to, e.g., the hourly temporal resolution and adequate height for wind speeds (50 m). A general conclusion from the studies cited above is

that MERRA gives good results in terms of country-wise wind power generation (relatively low errors when comparing to measurements).

During 2017–2018, a new reanalysis named ERA5 [15] will be released. Contrary to its predecessor ERA-Interim [16], ERA5 has hourly resolution and wind speed data is freely available at 100 m height, i.e. relevant for modern wind turbines (WTs). The horizontal resolution is around 31 km, which is better than ERA-Interim and MERRA. The main objective of this paper is to compare the performance of MERRA-2 and ERA5 in terms of modelling wind power, both for countries and for individual WTs. For the former, the generation in Germany, Denmark, France, Sweden and Bonneville Power Administration (BPA) in north-west USA were analysed. For simplicity, BPA is denoted a “country” in the remaining of this paper. For the analysis of individual WTs, hourly time series for more than 1000 Swedish units were obtained.

The most common approach for modelling country-wise wind power from reanalyses is to i) generate wind speed time series for each site, ii) transform wind speed to power via a power curve (PC) and iii) aggregate. Regarding point i), the reanalysis wind speeds can either be scaled so that a desired mean generation is met [9,10] or calculated from reanalysis data at different model heights [4,12] (possibly using statistical [6] or dynamical [17] downscaling). When transforming wind speed to power, some authors [4,9] use one

* Kungliga Tekniska Högskolan, SE-100 44, Stockholm, Sweden.
E-mail address: olauson@gmail.com.

generic PC while others [6,12] use farm specific ditto. It is common to alter the original PCs, e.g. using smoother “multi-turbine” power curves [4,12]. Especially for larger countries, it may be important to handle the maximum country-wise output, which can be considerably lower than the installed capacity [4,10,12]. Post-processing of the outputs at any of steps i–iii) is possible for improving the results. Examples include correction of seasonal and diurnal biases [10] and noise addition for a better representation of observed wind power variability.

The remainder of the paper is structured in the following manner. First, the two reanalyses and all measurement datasets are described. In Section 3, the methods used are presented, including a suggestion for a new metric describing the geographical dimensions of a wind power system. Subsequently the results are given and the paper is concluded with a short discussion.

2. Data

Several different datasets were retrieved for the purpose of this study. The two reanalyses are described in Section 2.1 while measurements and metadata are described in Section 2.2.

2.1. Reanalyses

In this work, the performance of MERRA-2 and ERA5 are compared. Other freely available reanalyses also exist (e.g. ERA-Interim and JRA-55), but have lower spatiotemporal resolutions and heights for wind speed data.

The first version of MERRA, the “Modern-Era Retrospective Analysis for Research and Applications”, was released in 2009. In early 2016 the production stopped and was replaced by MERRA-2 which is now updated in “near real-time”, i.e. data is available with a couple of months lag. For the purpose of modelling wind power, these two reanalyses perform similarly [12].

The first batch of the ERA5 dataset, produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), was released in July 2017 and covers the period 2010–2016. Subsequently, ERA5 will span 1950 to near real-time. Some important properties of MERRA-2 and ERA5 are shown in Table 1. For comparison, data for ERA-Interim, the predecessor of ERA5, are also given. ECMWF reanalyses are produced on reduced Gaussian grids, i.e. using different number of grid points along different latitudes and thus keeping the grid point separation in metres approximately constant. Since the data is downloaded on a user-defined regular lat/lon grid, the resolutions given in Table 1 are only approximate. Downloading with a higher resolution is possible, but this merely implies an interpolation of the data.

The spatial resolutions of ERA5 and MERRA-2 are illustrated qualitatively in Fig. 1. For comparison, mean wind speeds from the meso-scale model MIUU [18] (spatial resolution 0.5 km) are also shown. Note that data from the native model heights are shown

(50 m for MERRA-2 and 100 m for ERA-I) and that the colors thus represent different ranges of mean wind speeds. Although ERA5 gives more correct patterns than MERRA-2 e.g. along the coast, it is obvious that none of the reanalyses are capable of resolving the local variations, especially in more complex terrain. Using reanalyses directly for determining the mean wind speed at a site is therefore not recommended.

2.2. Wind power measurements and metadata

In order to model country-wise generation and to evaluate the results, the following measurements and metadata were used:

1. Coordinates, capacity and, if available, commissioning and decommissioning dates of each farm.
2. Hourly measurements of nation-wide generation.
3. Time series of the installed capacity corresponding to the measured generation. Since the wind farm database might be incomplete and that commissioning/decommissioning dates might not be available, this time series can differ from the one calculated from the wind farm data (item one in this list). The measured and modelled generation time series were therefore normalised to their corresponding capacity time series in order to get comparable per unit generation; $p_{p.u.}[t] = p_{MW}[t]/p_{inst}[t]$.

For Germany, Denmark and France, data were retrieved from the “Open Power System Data” (OPSD) database (data packages “Renewable power plants” version 2017-07-03 and “Time series” version 2017-07-09). Primary data are from various sources, for a complete list see Ref. [19]. Since coordinates for German offshore farms were lacking or seemed suspicious, these were instead obtained from wikipedia.org. The latter source was also used for country-wise installed capacities (linear interpolation of end-of-year values was employed to calculate $p_{inst}[t]$). BPA generation time series and farm metadata except coordinates were retrieved from Ref. [20]. Coordinates for BPA farms were obtained from on-line sources; thewindpower.net, wikipedia.org and farm developers' web pages. The BPA generation time series is given with five-minute resolution, so the data were averaged into hourly blocks.

Generation measurements for individual Swedish WTs were obtained from the Cesar database. This dataset can be retrieved by request from the Swedish Energy Agency and a more thorough description can be found in Ref. [21]. Only wind turbines commissioned year 2000 or later with a capacity of ≥ 1 MW and at least one year records were considered. Wind turbines for which the rated capacity changed during the measurement period and measurements taken for farms in aggregation were also excluded. In total, 1–6 year long time series for 1051 WTs remained to be analysed.

For Sweden, generation time series for individual WTs but no coordinates were available. The latter were therefore estimated from the reanalyses grid points with the highest correlation between modelled and measured generation. Since choosing the best fitting data could potentially give overly optimistic results, a test was performed. Relatively exact coordinates for ten randomly chosen WTs were found from the developers/owners web pages and reanalyses time series were computed for these sites. The corresponding correlations to measurements were in principle identical to those obtained from the best fitting time series; the largest difference for a single WT was 0.001 and the average correlations were the same within three digits precision. It was therefore concluded that estimating coordinates from the best fitting reanalyses grid points does not skew the results (neither positively nor negatively). It can be mentioned that the mean

Table 1

Information on the two reanalyses considered in this study (MERRA-2 and ERA5). For comparison, data for ERA-Interim, the predecessor of ERA5, is also shown. The spatial resolutions of the ERA reanalyses are approximate (see Section 2.1). In general, wind speed data are available at several heights or pressure levels. Only the most proper for modelling wind power is given here.

	MERRA-2	ERA-Interim	ERA5
Temporal resolution	1 h	6 h	1 h
Spatial resolution	$0.5^\circ \times 0.625^\circ$	79 km ($\approx 0.75^\circ$)	31 km ($\approx 0.28^\circ$)
Coverage	1980–	1979–	2010–2016, soon 1950–
Wind speed height	50 m	10 m	100 m

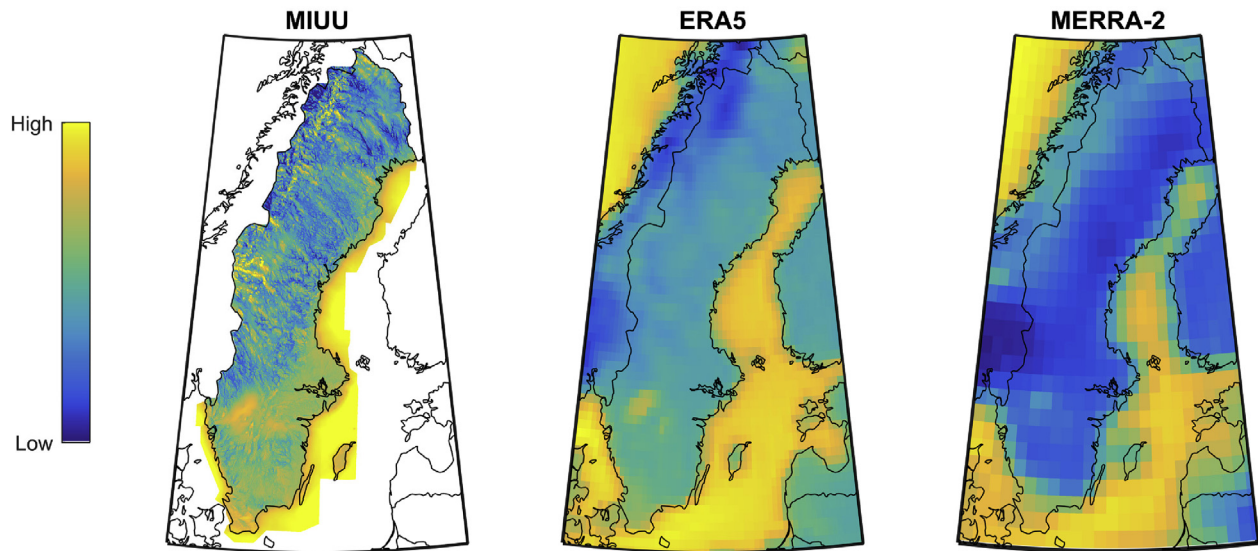


Fig. 1. Mean wind speeds for ERA5 (at 100 m above ground) and MERRA-2 (50 m). For comparison, mean wind speeds from the meso-scale model MIUU [18] with a spatial resolution of 0.5 km are also shown. No numeric color bars are given since the comparison should only be seen as qualitative (one particular color does not represent the same mean wind speed in the three panels).

distances between the correct and the estimated locations were 36 km for MERRA-2 and 15 km for ERA5.

Summary statistics of the country-wide measurements are shown in Table 2. “Capacity 1” and “Capacity 2” give the first and last samples of the capacity time series that were used to normalise measured and modelled generation, respectively. These values can sometimes, as described in the numbered list opening this section, differ. Note that for France, no commissioning and decommissioning dates were available (Capacity 2 is constant), so all farms were included in the simulations for the entire period. The “equivalent system radius” is a new metric for quantifying the system size and wind farm dispersion, which can be important for wind power variability as well as the difficulty of modelling the generation, see Section 3.4. The capacity factor (CF) is the average p.u.-generation during the whole period. Note that aggregated Swedish generation was only studied during 2014–2015 while time series for individual WTs were studied for the whole concurrent periods of measurements and ERA5 data (between one and six years during 2010–2015).

3. Methods

This section contains four parts. Firstly, the methodology for modelling hourly generation is described; secondly, the method for identifying downtime of Swedish WTs is given; thirdly, long-term correction is described and lastly a new metric for quantifying the size of a wind power system is proposed.

3.1. Modelling hourly generation

The hourly energy production for each unit (WT or farm) was modelled from reanalyses wind speeds and subsequently aggregated to country-wise time series. The model was deliberately kept relatively simple and the same parameters (or rules for determining parameters) were used for all countries. Because of this, no separate training, validation and test sets were used. No formal optimisation procedure was employed for trimming parameter values, only manual trial and error. All model parameters are described below and the chosen values are given in Table 3.

In horizontal, bi-linear interpolation of the four neighbouring grid points was employed, thus neglecting the impact from local topography. Based on earlier studies [2,6,17,21], statistical or

Table 3

Parameter values used for modelling hourly generation for countries and individual wind turbines (WTs). See Section 3.1 for parameter descriptions. No formal optimisation procedure was employed for finding parameter values, only manual trial and error.

	Countries	Individual WTs
Specific power	360 W/m ²	360 W/m ²
Multi-turbine smoothing	1.5 m/s	0.5 m/s
Internal losses	10%	15%
External loss factor	1.01 × measured maximum	1
β	1/40	1/40

Table 2

Information on country-wise wind power measurements. “Capacity 1” and “Capacity 2” give the first and last samples of the capacity time series that were used to normalise measured and modelled generation (see main text). The “equivalent system radius” is a metric describing the system size and wind farm dispersion, see Section 3.4.

	Germany	Denmark	France	Sweden	BPA
Country code	DE	DK	FR	SE	
Time period	2016	2015–2016	2015–2016	2014–2015	2014–2015
Temporal resolution	Hourly	Hourly	Hourly	Hourly	5 min
Capacity 1 (GW)	44–50	4.9–5.2	9.3–12.1	2.0–2.3	4.5–4.8
Capacity 2 (GW)	43–48	4.7–5.2	10.2	2.0–2.3	4.5–4.8
Equivalent system radius, r_{eq} (km)	329	166	376	471	91
Capacity factor	0.186	0.303	0.215	0.329	0.266

dynamical downscaling does, however, not necessarily implies better results in terms of e.g. correlations with measurements. As described below, the mean wind speed was also not taken directly from reanalysis data.

Since no information on rotor diameters were available, a power curve corresponding to a specific rating (rating over swept rotor area) of 360 W/m^2 was used for all units. The power curve was first smoothed, i.e., transformed to a “multi-turbine” power curve, using a Gaussian distribution with standard deviation 1.5 m/s [10].

There are several types of losses in a wind farm. Wake losses, mild icing, degraded blades etc. shift the power curve (a higher wind speed is required in order to give a certain output) but does not affect the maximum generation level. Downtime and some electrical losses can however lead to a lower maximum generation, at least for a country. Furthermore, reanalyses generally underestimate the diversity of wind speeds; in reality, all farms in a country will never have wind speeds above rated simultaneously and thus the generation in countries like Germany, France and Sweden will seldom exceed 0.8 p.u. In order to mimic these two different phenomena, denoted “internal” and “external” losses in Ref. [10], the incoming energy in the wind was reduced by a fixed amount and the resulting energy time series were multiplied by a factor.

As has been showed in Section 2.1, reanalyses are generally not suitable for determining the mean wind speed at a particular site. In the present study, hub heights were also not available, i.e. it was not possible to interpolate/extrapolate the wind speeds from different reanalysis model heights. The mean wind speeds were therefore determined from the estimated CF of each farm (CF_i), which was calculated as

$$CF_i = CF_{country} + \beta(\bar{u}_i - \bar{u}_{country}), \quad (1)$$

where $CF_{country}$ is the measured CF of the country in question, \bar{u}_i and $\bar{u}_{country}$ are the mean reanalyses wind speeds for the particular site and the capacity-weighted mean for all sites in the country, respectively, and β is a parameter determining how much the mean wind speeds influence the CFs. Note that \bar{u}_i and $\bar{u}_{country}$ were computed for the reanalyses model heights, i.e. 50 m for MERRA-2 and 100 m for ERA. For Denmark and Germany, separate calculations were performed for onshore and offshore farms. The wind speed time series were subsequently linearly scaled to the desired means (the mean wind speeds that corresponds to the desired CFs), i.e. assuming constant wind shears. Although this is obviously not entirely correct, using time-varying wind shears give little or no improvement of the model performance [10].

For the modelling of generation for individual Swedish WTs, internal losses were increased to 15% and the external loss factor was set to unity. The reason for not using any external losses was, of course, that the generation for a single WT can sometimes be 1 p.u. The multi-turbine smoothing parameter was lowered to 0.5 m/s since this gives a somewhat better representation of the output from individual WTs.

As a final note, it should be recognised that the method outlined above make use of the measured average CF and maximum generation, i.e. the modelled time series will have the correct averages and maxima.

3.2. Identifying downtime

When evaluating the performance of reanalyses in terms of modelling generation for individual WTs, downtime (stops due to e.g. maintenance and technical faults) will make the comparison less fair. A method for automatic detection of downtime was therefore used and these data samples were removed. The basic idea was that if the modelled generation is relatively high while the

measured generation is zero for an extended period of time, this is due to downtime.

For each WT, a generation time series was first calculated with the method given in the previous section. For periods with zero actual generation for at least five hours, the corresponding mean reanalyses generation was computed. The period was considered downtime if at least one of the following conditions was fulfilled:

1. The period is longer than one week
2. The period is at least twelve hours and the corresponding reanalysis mean is at least 0.05 p.u.
3. The corresponding reanalysis mean is at least 0.15 p.u.

In general, there is a trade-off between false positive and false negative downtime classifications. If, for example, the threshold of five hours was to be lowered to three hours, additional downtime events would be correctly captured at the expense of increasing the number of events falsely classified as downtime. The chosen parameters give balanced results; when validating with actual downtime records, there were 9% false negative and 8% false positive errors (in relation to the true number of downtime hours) [21].

After the downtime samples were removed, a new calculation of reanalysis generation was performed based on the new (higher) measured CF. Note that downtime data was only removed for the analysis of individual WTs, not for the analysis of aggregated Swedish data.

3.3. Long-term correction

Long-term correction (LTC) is routinely used for finding the best estimate of the long-term wind speed distribution at a site given a shorter (one or a few years) measurement. A LTC can also be used to determine the long-term CF given a shorter generation time series, e.g. to see whether a new farm performs according to plan or to analyse the value of a farm that is to be sold. All types of LTC methods require some sort of long, preferably at least ten years, reference time series. A relationship is established between the site measurement and the reference during the period with concurrent data. This relationship is subsequently “fed” with the whole reference time series in order to determine the long-term distribution or mean of the variable of interest.

In recent years, reanalyses wind speeds have become the de facto standard as reference time series due to their consistency and long records. Wind speed measurements from meteorological masts, generally 10 m high, can also be used but often have consistency problems due to e.g. gradually growing vegetation and changes in anemometry.

In this work, the energy errors from using short, long-term corrected generation time series of single WTs are evaluated. One randomly chosen year of measurements (sample data/training period) and reanalyses wind speeds were used for predicting the energy production for the remaining 1–5 years (out-of-sample data). Since at least two years of data is thus required, “only” 893 WTs were analysed. A simple “wind index” method was employed:

1. A one-year period was randomly chosen as sample data and the corresponding CF was calculated.
2. Generation time series for the entire measurement period were modelled from ERA5 and MERRA-2 data. The mean wind speeds were linearly scaled so that the CF during the training period matches measurements.
3. The predicted CF was taken as the average modelled generation for out-of-sample data.

Note that independent LTCs were performed for each WT, i.e.,

the training periods were not the same. Since the results to some degree depend on the random choice of training periods, ten different runs were performed.

Small errors in such an exercise is not only of academic interest; when, for instance, a farm is to be sold, the most important figure is often not the predicted generation but rather the generation that will be achieved with a given certainty, e.g. 75% or 90%. A low uncertainty is thus worth a lot in monetary terms.

3.4. Equivalent system radius

There are many possible candidates for quantifying the system size and/or dispersion of wind farms. Such metrics can be used for a general system description or as a proxy for wind power variability and include, e.g., the system dimensions in North-South and West-East [22], the system area [8] and the (weighted) mean distance between farms [23]. Although these metrics might be adequate, there are some issues:

- System dimensions and areas are the same no matter if the farms are evenly distributed or heavily concentrated in one part of the system.
- The area can be the same for various system shapes, although wind power variability differs.
- The mean distance is influenced by the level of details of the dataset, e.g. whether coordinates are given for farms or WTs. The mean distance also, in my opinion, gives a less intuitive feeling of the system dimensions.

Because of these drawbacks, a new metric was developed: the “equivalent system radius” r_{eq} . The idea is that a set of wind farms can be described by a disk of uniformly distributed wind power that has the same combined variance as the actual wind farm fleet, assuming that the correlation between generation of two farms declines in an exponential fashion with separation distance. Before proceeding with the derivation of the equation for r_{eq} , the concept is illustrated in Fig. 2.

Let $p_{tot}[t]$ denote the per unit time series of the combined generation for the actual farm distribution. The variance of $p_{tot}[t]$ is

$$\text{Var}(p_{tot}[t]) = \sum_{i=1}^N \sum_{j=1}^N P_i P_j \sigma_i \sigma_j \rho_{ij}, \quad (2)$$

where N is the number of farms, P_i is farm i 's share of the total capacity (i.e. $\sum_i P_i = 1$), σ_i is the standard deviation of $p_i[t]$ (the per unit time series for farm i) and ρ_{ij} is the correlation between the outputs for farms i and j . Assuming the same σ_i for all farms and using the common [24] correlation model

$$\rho_{ij} = e^{-d_{ij}/D}, \quad (3)$$

where d_{ij} is the distance between farms i and j and D is a constant, Eq. (2) can be rewritten as

$$\text{Var}(p_{tot}[t]) = \sigma^2 \sum_{i=1}^N \sum_{j=1}^N P_i P_j e^{-d_{ij}/D}. \quad (4)$$

We now want to find r_{eq} so that a uniform “wind power disk” with this radius will have the same variance as that obtained with Eq. (4). Let us denote the total output $p_{tot,eq}[t]$. The variance of $p_{tot,eq}[t]$ can, in analogue with Eq. (4) for the discrete case, be expressed as

$$\text{Var}(p_{tot,eq}[t]) = \sigma^2 \int_{A_{eq}} \int_{A_{eq}} e^{-d_{ij}/D} dP_i dP_j, \quad (5)$$

where A_{eq} is the area of our disk. Since $\int_A dP = 1$, we can write

$$dP = \frac{dA}{\pi r_{eq}^2}. \quad (6)$$

In polar coordinates,

$$dA = r dr d\theta \quad \text{and} \quad (7)$$

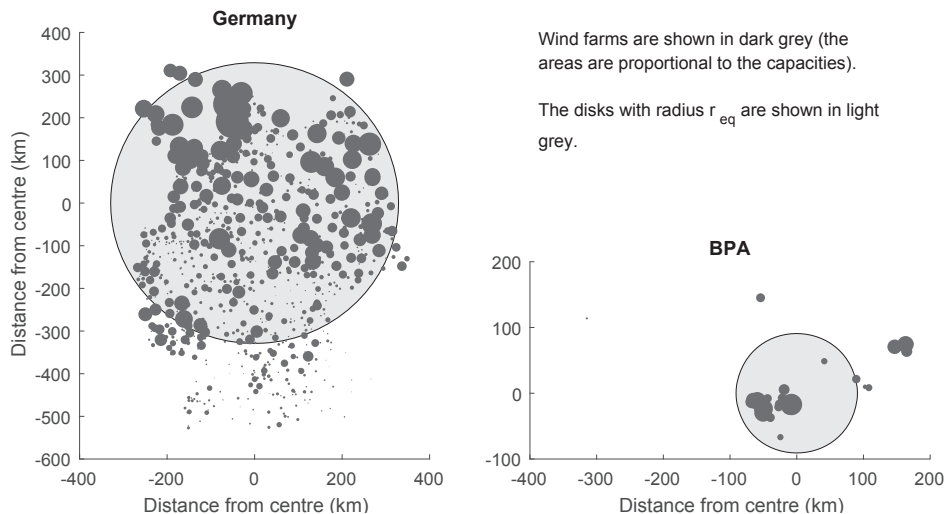


Fig. 2. Illustration of the equivalent system radius (r_{eq}) for Germany and BPA. Note that BPA farms are concentrated in a relatively small area which gives a small r_{eq} as compared to the total system dimensions. In particular, the small farm at around $(-300, 100)$ does not impact r_{eq} much. For Germany, wind turbines with the same coordinates are plotted like one farm although this is not always the case.

$$d_{ij} = \sqrt{r_i^2 + r_j^2 - 2r_i r_j \cos(\theta_i - \theta_j)} \quad (8)$$

and Eq. (5) can thus be rewritten as

$$\text{Var}(p_{\text{tot},eq}[t]) = \frac{\sigma^2}{\pi^2 r_{eq}^4} \int_{r_i=0}^{r_{eq}} \int_{r_j=0}^{r_{eq}} \int_{\theta_i=0}^{2\pi} \int_{\theta_j=0}^{2\pi} r_i r_j e^{-\sqrt{r_i^2 + r_j^2 - 2r_i r_j \cos(\theta_i - \theta_j)} / D} dr_i dr_j d\theta_i d\theta_j \quad (9)$$

Combining Eq. (4) and Eq. (9) gives us

$$\sum_{i=1}^N \sum_{j=1}^N P_i P_j e^{-d_{ij}/D} = \frac{1}{\pi^2 r_{eq}^4} \int_{r_i=0}^{r_{eq}} \int_{r_j=0}^{r_{eq}} \int_{\theta_i=0}^{2\pi} \int_{\theta_j=0}^{2\pi} r_i r_j e^{-\sqrt{r_i^2 + r_j^2 - 2r_i r_j \cos(\theta_i - \theta_j)} / D} dr_i dr_j d\theta_i d\theta_j \quad (10)$$

Equation (10) was solved numerically for r_{eq} using $D = 500$ km, a common value in the literature [24]. For practical purposes, an approximate solution of Eq. (10) suffice. The left-hand side of Eq. (10) can be computed easily, let us denote this sum x . A good polynomial approximation of r_{eq} for $0.22 \leq x \leq 1$ is given in Fig. 3, using $D = 500$ km as before. The mean absolute residual of the fit is 0.5 km.

Rather than quantifying variability by looking at the variance of generation, one might want to study the variance of changes in generation. As the correlation of changes is much lower than that for generation itself, a lower value of D must be used. Since D appears on both sides of Eq. (10), the value of r_{eq} does however not change very much. As an example, using $D = 50$ km instead of $D = 500$ km changes r_{eq} with -4% and -8% for Germany and Denmark, respectively. It should be mentioned that a lower value of D requires a higher quality of the farm coordinates since the estimated correlations become more sensitive to farm separation distances.

4. Results

This section is divided into two subsections. First an evaluation of the modelling of country-wise generation is presented and second results for individual Swedish WTs are given.

4.1. Country-wise generation

The challenges of modelling wind power generation in different countries vary with e.g. system size, installed capacity, terrain complexity and wind regimes. A large system with many WTs is generally easier to model since the errors for individual WTs are partially cancelled out; the correlations between errors reduce with separation distance. Because of this, the most important results of this paper are not the absolute errors but rather the difference in performance between MERRA-2 and ERA5.

In order to give a reference frame for the quantitative results that will soon be presented, some graphical examples (Sweden modelled with ERA5 data) are shown in Fig. 4. As an overall

judgment, the model performance for Sweden is medium good in comparison to all five evaluated countries. Panel a shows a representative 4000 sample excerpt of the time series of measured and modelled generation. In panel b, all data have been sorted from highest to lowest, forming so called duration curves. Panel c

demonstrates the distribution of changes of hourly generation. The magnitudes of these “step changes” are somewhat underestimated

by the model; the standard deviation of ERA5 step changes are around 8% lower than that for measurements (in relative terms), see Table 4. Panel d, finally, gives a scatter plot of actual versus predicted monthly CFs. For all months but one, the modelled and measured CFs agree well. For February 2014, ERA5 predicts 0.40 but the actual outcome was 0.32. A likely explanation for at least part of the difference is icing losses (which are not modelled).

Quantitative results for all countries and for both MERRA-2 and ERA5 are presented in Table 4. The metrics considered are correlations, root-mean-square errors (RMSE), mean absolute errors (MAE), relative errors in the standard deviation of one hour step changes ($\sigma_{\Delta p}$) and RMSE of the duration curves (RMSE_{dur}). Note that RMSE, MAE and RMSE_{dur} are given in absolute terms, so e.g. 2.82% corresponds to an error of 0.0282 p.u. For $\sigma_{\Delta p}$, the errors are given in relative terms; +12.8% implies that the standard deviation is 12.8% higher for the model than for measurements (e.g. 0.0338 p.u. versus 0.03 p.u.). All metrics in Table 4 are calculated for hourly data.

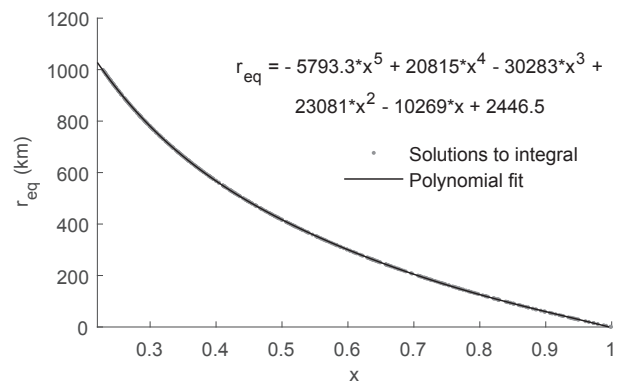


Fig. 3. Polynomial approximation of the equivalent system radius (r_{eq}) as a function of x , the left-hand side of Eq. (10). The variable x can be interpreted as the per unit variance of a country-wise time series as compared to that for individual wind turbines (WTs). The approximation is valid for $D = 500$ km, where D is the parameter determining how fast correlations decline with WT separation distance.

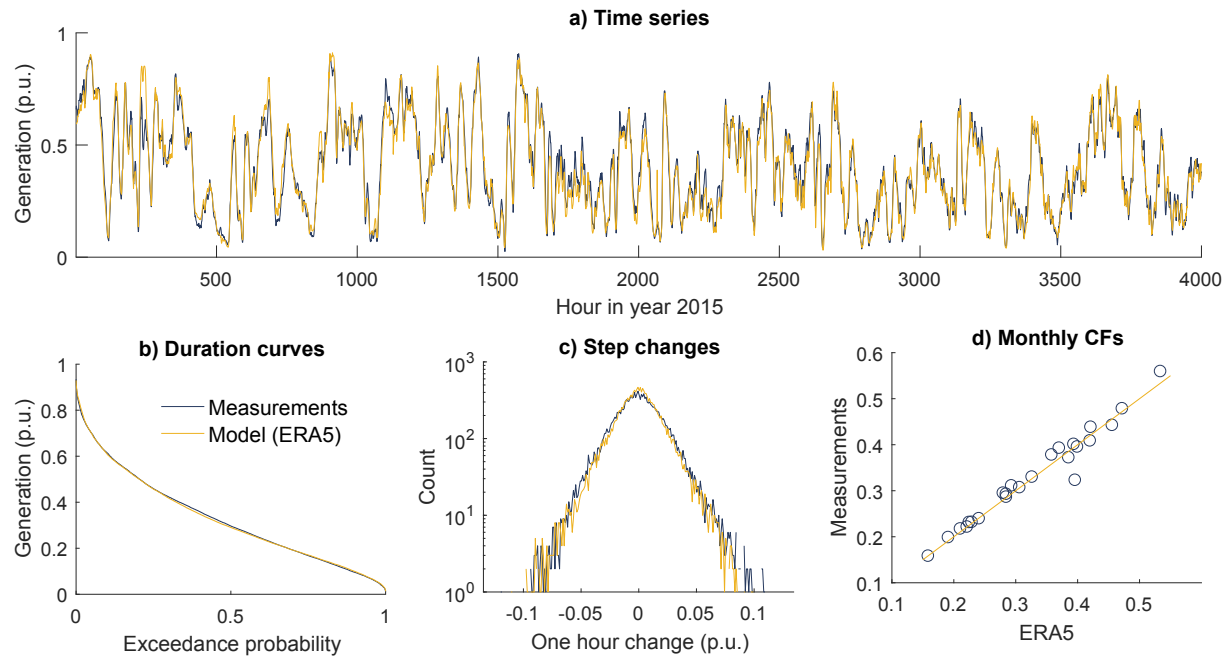


Fig. 4. Illustration of the model performance (ERA5 data used for modelling the Swedish wind power generation). Panel a shows a representative 4000 sample excerpt of the time series of measured and modelled generation. Panel b gives duration curves (generation sorted from highest to lowest) of all data. Panel c shows the distribution of one hour step changes, i.e. the changes in average power from one hour to the next. In panel d, a scatter plot of modelled and measured monthly capacity factors (CFs) is given. The straight line shows the desired one-to-one relationship.

Table 4

Quantitative results for the modelling of hourly, county-wise wind power. The metrics considered are correlations, root-mean-square errors (RMSE), mean absolute errors (MAE), relative errors in the standard deviation of one hour step changes ($\sigma_{\Delta P}$) and RMSE of the duration curves (RMSE_{dur}). Note that RMSE, MAE and RMSE_{dur} are given in absolute terms, so e.g. 2.82% corresponds to an error of 0.0282 p.u. For $\sigma_{\Delta P}$, the errors are given in relative terms; +12.8% implies that the standard deviation is 12.8% higher for the model than for measurements (e.g. 0.0338 p.u. versus 0.03 p.u.).

Country	Reanalysis	Correlation	RMSE	MAE	Error $\sigma_{\Delta P}$	RMSE _{dur}
Germany	MERRA-2	0.982	2.82%	2.06%	+12.8%	0.43%
	ERA5	0.987	2.35%	1.66%	+7.5%	0.44%
Denmark	MERRA-2	0.973	5.40%	3.75%	−11.7%	1.75%
	ERA5	0.973	5.45%	3.71%	−6.0%	1.52%
France	MERRA-2	0.975	3.49%	2.63%	+17.2%	0.79%
	ERA5	0.982	2.97%	2.25%	+14.3%	1.00%
Sweden	MERRA-2	0.950	6.10%	4.58%	−8.7%	1.27%
	ERA5	0.975	4.40%	3.21%	−7.9%	0.52%
BPA	MERRA-2	0.756	18.4%	13.3%	−10.4%	6.09%
	ERA5	0.945	9.10%	6.06%	1.5%	2.34%

For a clearer comparison of the performance of the reanalyses, Fig. 5 shows the RMSE and MAE of ERA5 normalised to those for MERRA-2 (a bar of e.g. height 0.8 thus implies that the errors are 20% lower for ERA5 than for MERRA-2). In average, the RMSE is 22% lower and the MAE is 24% lower. Large differences however exist between the five countries. For Denmark, the MERRA-2 and ERA5 results are very similar. Likely, this can be explained by the flat terrain and lack of forest; the benefit of a higher model resolution is limited for such conditions. For BPA, on the other hand, the errors are around 50% lower for ERA5. In comparison to Denmark, the equivalent system radius of BPA is much lower and the surrounding terrain is more complex. As a consequence, the generation is more volatile in the BPA area; the standard deviation of one hour step changes is for instance 0.048 p.u. as compared to 0.028 p.u. for Denmark. Taken together, these factors increase the benefits of

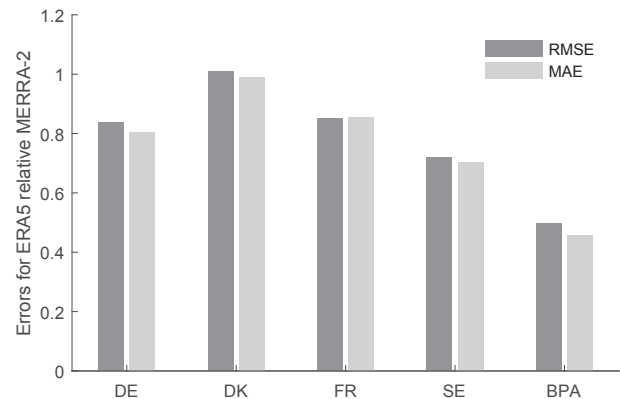


Fig. 5. Root-mean-square errors (RMSE) and mean absolute errors (MAE) for ERA5 in relation to those for MERRA-2. A bar of e.g. height 0.8 thus implies that the considered metric is 20% lower for ERA5 than for MERRA-2. In average, the RMSE is reduced with 22% and the MAE is reduced with 24% by using ERA5 data.

using a reanalysis with higher spatial resolution.

The performance of MERRA for modelling wind generation in several different European countries has been evaluated earlier [6,12]. The results in this work (correlation and RMSE) for MERRA-2 are comparable or better, which indicates that the methodology and measurement data quality are adequate. One should be careful in drawing too far-reaching conclusions on the pros and cons of the different modelling approaches and reanalyses since the studied time periods were not the same and that different WT databases were used. With that said, it is interesting to note that the performance of ERA5 seem to be better than that of “EMHIRES”, a dataset obtained by statistically downscaling MERRA data [6], for the countries studied here.

When comparing the Swedish results here to those in Ref. [10], the errors are somewhat higher and the correlation is lower. The

main explanation is that only modern WT were included in this study. These are generally located in more complex terrain (hilly forests) and are more exposed to icing than the early WT which were erected mainly in open landscape in southern Sweden. The CF for the WT studied here is around 35% higher than for those studied in Ref. [10], which tend to increase the per unit errors. The model in Ref. [10] was also more detailed, e.g. accounting for seasonal/diurnal biases and direction dependent losses. Finally, MERRA-2 seem to perform slightly worse than MERRA in Sweden [12]. A final note on the modelling of Swedish, aggregated generation is that only the average national CF was used as input to the model, i.e. not information on the measured CFs of individual WT. Using this extra information does however, surprisingly enough, not improve the results at all.

4.2. Individual WT

In this section, results are given for modelling of individual Swedish WT. As described in Section 3.2, periods classified as downtime were first removed. Table 5 gives mean and median values of correlations, RMSE, MAE and RMSE of duration curves for the 1051 analysed WT. In the last column, correlations of monthly generation are given (only months with at least 300 valid data samples were considered). Boxplots of the analysed hourly metrics are given in Fig. 6. On each box, the central mark is the median, the edges of the box represent the 25th and 75th percentiles and whiskers extend to the most extreme data. It is clear from Table 5 and Fig. 6 that ERA5 performs better than MERRA-2 in all analysed aspects. One can note that the median performance is always better than mean. This is due to very large errors for a few WT.

Neither MERRA-2 nor ERA5 is capable of reproducing the variability of individual WT; the measured fluctuations are much larger in magnitude than those obtained from reanalyses. This is not surprising since coarse models are not able to reproduce the fluctuations at a local site. In average, the standard deviation of one hour step changes are underestimated by 49% and 43% by the models based on MERRA-2 and ERA5 data, respectively. A Welch power spectral density estimate [25] can be used to illustrate the magnitude of fluctuations of different frequencies. In Fig. 7, a representative example is given, using 50 segments to reduce the noise and make the graph easier to interpret. For frequencies higher than around $(1 \text{ day})^{-1}$, the spectra for modelled generation start to deviate from that for measurements. At the highest resolvable frequency, the difference is large, in particular for MERRA-2.

Although the variability of an individual WT is underestimated, that for aggregated country-wise generation might not be. This is because the correlations between different time series are overestimated by the models and consequently the modelled generation is smoothed more by aggregation. From Table 4, the variability as quantified by the standard deviation of one hour step changes is overestimated for some countries and underestimated for some. The absolute mismatch is always lower for ERA5.

Next, energy errors from long-term correction (LTC) are evaluated. As described in Section 3.3, one randomly chosen year was used as sample data in order to predict the energy production

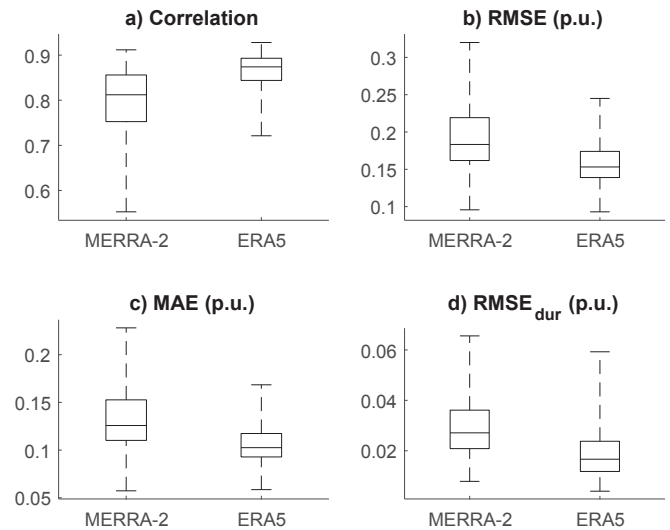


Fig. 6. Boxplots of the results for 1051 Swedish wind turbines. On each box, the central mark is the median, the edges of the box represent the 25th and 75th percentiles and whiskers extend to the most extreme data. For an explanation of the metrics, see Table 4.

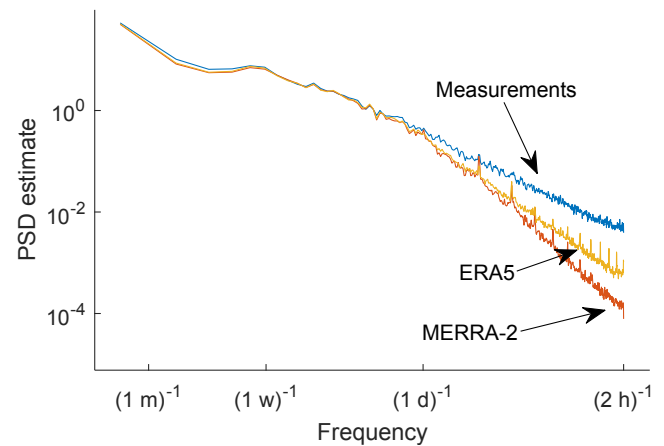


Fig. 7. Representative example of Welch power spectral density (PSD) estimates for one wind turbine (WT). It is clear that the variability of the generation for a single WT is underestimated by both reanalyses (however, somewhat less by ERA5).

during the remaining 1–5 years (out-of-sample data). Note that all results given in this section are means from ten different realisations. Results for the 893 studied WT are shown in Table 6. For comparison, results are also given for the “Naïve estimate” which is simply assuming the same capacity factor as for the sample data. All three estimates are in principle unbiased which implies that the standard deviations of the errors are similar to the RMSEs.

First, one should note that the errors are considerably lower with a proper LTC than with the Naïve estimator. This is of course not surprising, but nevertheless important. In conclusion, the

Table 5

Results for the modelling of generation for individual wind turbines. Means and medians for the 1051 analysed WT are given. For an explanation of the metrics, see Table 4.

	Correlation	RMSE	MAE	RMSE _{dur}	Monthly correlation
MERRA-2 (mean)	0.802	19.1%	13.2%	2.94%	0.922
ERA5 (mean)	0.868	15.8%	10.6%	1.92%	0.964
MERRA-2 (median)	0.812	18.3%	12.6%	2.71%	0.935
ERA5 (median)	0.874	15.3%	10.3%	1.66%	0.973

Table 6

Energy errors in long-term correction for 893 individual Swedish wind turbines (WTs). Both root-mean-square errors (RMSE) and mean absolute errors (MAE) are given. For each WT, one randomly chosen year of data was used to train a model which was subsequently used to predict the energy production during the remaining 1–5 years of out-of-sample data. The Naïve estimator assumes the same capacity factor as for the training period. The results are means from ten realisations and normalised to the measured generation.

	MAE	RMSE
Naïve estimate	7.42%	9.26%
MERRA-2	3.07%	4.00%
ERA5	2.43%	3.23%

energy errors are around 20% lower for ERA5 than for MERRA-2. If two years' worth of data are used for training the models, the errors for both reanalyses are reduced with around 18%. This means that one year of measurements long-term corrected with ERA5 actually gives slightly more accurate estimates than two years of measurements long-term corrected with MERRA-2.

In order to validate the robustness of the results, two-tailed *t*-tests were performed, testing the null hypothesis that the differences in absolute errors, i.e. $|E_{ERA5,i}| - |E_{MERRA2,i}|$, come from a (normal) distribution with mean equal to zero. Using one year sample data for both reanalyses, the mean difference is -0.64 percentage points (pp) and the 95% confidence interval is $(-0.79, -0.49)$ pp. The null hypothesis can thus be rejected on the 0.05 significance level (or any other meaningful level) and we can conclude that ERA5 performs better than MERRA-2. With one year training data for ERA5 and two years training data for MERRA-2, the corresponding figures are -0.11 ($-0.31, 0.09$) pp, i.e. the difference is not statistically significant. For verification, 10,000 samples bootstrap tests were also performed; the results were identical within the precision given above.

In Ref. [21], correlations between measurements and generation modelled from MERRA, ERA-Interim and ConWx (a dataset dynamically downscaled from ERA-Interim) were analysed. It was concluded that MERRA and ConWx performed similarly, which was somewhat surprising given the higher spatial resolution of ConWx. Similar results have however been seen also in earlier work [2]; a higher resolution does not guarantee better performance. As has already been mentioned, the WTs chosen for the current work are more often located in complex terrain and suffer more severe icing losses than Swedish WTs in general. The average correlation obtained for MERRA-2 here is thus lower than that for MERRA in Ref. [21] (0.80 versus 0.84). Despite the tougher modelling conditions, the average correlation for ERA5 is however higher (0.87).

5. Concluding discussion

In this work, MERRA-2 and the new reanalysis ERA5 were compared in terms of modelling wind power, both for countries and for individual wind turbines. The main conclusion is that ERA5 performs better in all analysed aspects. As an average for the five studied countries, the mean absolute errors in modelled hourly generation (when compared to measurements) were 24% lower for ERA5. For the small BPA system, the errors were roughly 50% lower. Correlations were also higher and the mismatches in standard deviations of one hour step changes, a common variability metric, were lower.

ERA5 also performed better in terms of modelling generation for the 1051 individual Swedish WTs that were studied. The average correlation in hourly data was 6.6% points higher and the average MAE was 20% lower. A similar reduction was seen for the long-term correction energy errors (using one year of data to predict the

energy production during the remaining period). As a matter of fact, using one year of measurements long-term corrected with ERA5 gives slightly more accurate estimates than using two years of measurements long-term corrected with MERRA-2. The latter result is however not statistically significant. This is good news for the wind industry since lower uncertainties imply reduced risks and increased project values. Apart from the better performance, ERA5 will also soon be available from 1950 to near real-time, i.e. for a longer period than most other reanalyses.

The proposed metric “equivalent system radius” has its advantages as compared to earlier metrics for quantifying the size of a wind power system. Most importantly, it is based on the smoothing of wind power fluctuations from geographical dispersion, which is often an important characteristic when the system size is to be quantified. Unfortunately, the equation for r_{eq} turned out to be quite cumbersome and had to be solved numerically. Although a polynomial approximation works fine for practical purposes, an analytic solution to the integral would be a welcome contribution.

In this work, a relatively simple and generic wind power model was chosen and data on e.g. hub heights and farm specific power curves were not available. An important conclusion is that good results (generally equally or better than in earlier work) were still obtained. By including bias-correction and trimming the model harder for a specific country, somewhat better results can however be achieved [10]. Likely, the improvement would be larger for MERRA-2 since e.g. the biases in seasonal and diurnal wind patterns are larger. In practice, the models used in both research and industry are however also relatively simple, i.e. the results presented here are representative for the benefits of using the more high-resolved ERA5.

A common application of reanalysis based wind power models is to study hypothetical, highly renewable power systems [8,11]. In a future with more offshore wind power, higher hub heights and lower specific rating, the capacity factors will be higher and the WTs will more often generate at or near rated capacity. Even for large countries, the maximum generation will thus be closer to 1 p.u. than today. Here, the maximum generation in a country was determined from measurements (around 0.8 p.u. for larger countries). A better modelling approach for future scenarios might be to add multivariate noise to the wind speed time series and simulate downtime stochastically. If a proper model is set up, the spectra, correlations, maximum generation etc. will be correct both for individual WTs and country-wise generation (at the expense of increasing the errors to some extent).

Acknowledgements

This work was conducted within the StandUP for Energy strategic research framework and was funded by Vindforsk through contract 40290-1. Energimyndigheten, ECMWF, NASA, BPA and OPSD provided data. The Matlab function for projecting latitudes and longitudes onto an azimuthal equidistant image was written by Kurt von Laven. The method for solving quadruple integrals in Matlab was inspired by Mike Hosea. Dorothée Vallot and Jonatan Samuelsson are acknowledged for their valuable comments on the equivalent system radius method. Lasse Johansson is acknowledged for the idea of approximating unknown WT coordinates from coordinates of the best fitting reanalyses time series.

References

- [1] R. Gelaro, W. McCarty, M.J. Suárez, R. Todling, A. Molod, L. Takacs, C.A. Randles, A. Darmenov, M.G. Bosilovich, R. Reichle, K. Wargan, L. Coy, R. Cullather, C. Draper, S. Akella, V. Buchard, A. Conaty, A.M. da Silva, W. Gu, G.-K. Kim, R. Koster, R. Lucchesi, D. Merkova, J.E. Nielsen, G. Partyka, S. Pawson, W. Putman, M. Rienecker, S.D. Schubert, M. Sienkiewicz, B. Zhao, The Modern-

- Era retrospective analysis for research and applications, version 2 (MERRA-2), *J. Clim.* 30 (14) (2017) 5419–5454, <https://doi.org/10.1175/JCLI-D-16-0758.1>.
- [2] S. Liléo, E. Berge, O. Undheim, R. Klinkert, R.E. Bredeisen, Long-term Correction of Wind Measurements - State-of-the-art, Guidelines and Future Work, *Tech. rep., Elforsk report 13:18*, Jan. 2013.
 - [3] M.M. Rienecker, M.J. Suarez, R. Gelaro, R. Todling, J. Bacmeister, E. Liu, M.G. Bosilovich, S.D. Schubert, L. Takacs, G.-K. Kim, S. Bloom, J. Chen, D. Collins, A. Conaty, A. da Silva, W. Gu, J. Joiner, R.D. Koster, R. Lucchesi, A. Molod, T. Owens, S. Pawson, P. Pegion, C.R. Redder, R. Reichle, F.R. Robertson, A.G. Ruddick, M. Sienkiewicz, J. Woollen, MERRA: NASA's modern-era retrospective analysis for research and applications, *J. Clim.* 24 (2011) 3624–3648, <https://doi.org/10.1175/JCLI-D-11-00015.1>.
 - [4] D.J. Cannon, D.J. Brayshaw, J. Methven, P.J. Coker, D. Lenaghan, Using reanalysis data to quantify extreme wind power generation statistics: a 33 year case study in Great Britain, *Renew. Energy* 75 (2015) 767–778, <https://doi.org/10.1016/j.renene.2014.10.024>.
 - [5] A. Cosseron, U.B. Gunturu, C.A. Schlosser, Characterization of the wind power resource in europe and its intermittency, *Energy Procedia* 40 (2013) 58–66, <https://doi.org/10.1016/j.egypro.2013.08.008>.
 - [6] I. González-Aparicio, F. Monforti, P. Volker, A. Zucker, F. Careri, T. Huld, J. Badger, Simulating European wind power generation applying statistical downscaling to reanalysis data, *Appl. Energy* 199 (2017) 155–168, <https://doi.org/10.1016/j.apenergy.2017.04.066>.
 - [7] W.L. Henson, J.G. McGowan, J.F. Manwell, Utilizing reanalysis and synthesis datasets in wind resource characterization for large-scale wind integration, *Wind Eng.* 36 (1) (2012) 97–109, <https://doi.org/10.1260/0309-524X.36.1.97>.
 - [8] M. Huber, D. Dimkova, T. Hamacher, Integration of wind and solar power in Europe: Assessment of flexibility requirements, *Energy* 69 (2014) 236–246, <https://doi.org/10.1016/j.energy.2014.02.109>.
 - [9] M.L. Kubik, D.J. Brayshaw, P.J. Coker, J.F. Barlow, Exploring the role of reanalysis data in simulating regional wind generation variability over Northern Ireland, *Renew. Energy* 57 (2013) 558–561, <https://doi.org/10.1016/j.renene.2013.02.012>.
 - [10] J. Olauson, M. Bergkvist, Modelling the Swedish wind power production using MERRA reanalysis data, *Renew. Energy* 76 (2015) 717–725, <https://doi.org/10.1016/j.renene.2014.11.085>.
 - [11] J. Olauson, M.N. Ayob, M. Bergkvist, N. Carpman, V. Castellucci, A. Goude, D. Lingfors, R. Waters, J. Widén, Net load variability in Nordic countries with a highly or fully renewable power system, *Nature Energy* 1 (2016) 16175, <https://doi.org/10.1038/nenergy.2016.175>.
 - [12] I. Staffell, S. Pfenninger, Using bias-corrected reanalysis to simulate current and future wind power output, *Energy* 114 (2016) 1224–1239, <https://doi.org/10.1016/j.energy.2016.08.068>.
 - [13] M. McPherson, T. Sotiropoulos-Michalakakos, L.D. Harvey, B. Karney, An open-access web-based tool to access global, hourly wind and solar PV generation time-series derived from the MERRA reanalysis dataset, *Energies* 10 (7) (2017) 1007, <https://doi.org/10.3390/en10071007>.
 - [14] S. Liao, W. Yao, X. Han, J. Wen, S. Cheng, Chronological operation simulation framework for regional power system under high penetration of renewable energy using meteorological data, *Appl. Energy* 203 (2017) 816–828, <https://doi.org/10.1016/j.apenergy.2017.06.086>.
 - [15] ERA5 data documentation, Available online: <https://software.ecmwf.int/wiki/display/CKB/ERA5+data+documentation> (Accessed: 2017-07-18).
 - [16] D.P. Dee, S.M. Uppala, A.J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M.A. Balsameda, G. Balsamo, P. Bauer, P. Bechtold, A.C.M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A.J. Geer, L. Haimberger, S.B. Healy, H. Hersbach, E.V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A.P. McNally, B.M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavalato, J.-N. Thépaut, F. Vitart, The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.* 137 (656) (2011), <https://doi.org/10.1002/qj.828>.
 - [17] E. Nuño, P. Maule, A. Hahmann, N. Cutululis, P. Sørensen, I. Karagali, Simulation of transcontinental wind and solar pv generation time series, *Renew. Energy* 118 (2018) 425–436, <https://doi.org/10.1016/j.renene.2017.11.039>.
 - [18] H. Bergström, Boundary-layer modelling for wind climate estimates, *Wind Eng.* 25 (5) (2001) 289–299, <https://doi.org/10.1260/030952401760177864>.
 - [19] Open Power System Data, Available online: <https://data.open-power-system-data.org> (Accessed: 2017-08-02).
 - [20] Wind generation and total load in the BPA balancing authority, Available online: <https://transmission.bpa.gov/Business/Operations/Wind/> (Accessed: 2017-08-09).
 - [21] J. Olauson, P. Edström, J. Rydén, Wind turbine performance decline in Sweden, *Wind Energy* (2017) 1–5, <https://doi.org/10.1002/we.2132>.
 - [22] H. Holttinen, Hourly wind power variations in the Nordic countries, *Wind Energy* 8 (2) (2005) 173–195, <https://doi.org/10.1002/we.144>.
 - [23] S.M. Fisher, J.T. Schoof, C.L. Lant, M.D. Therrell, The effects of geographical distribution on the reliability of wind energy, *Appl. Geogr.* 40 (2013) 83–89, <https://doi.org/10.1016/j.apgeog.2013.01.010>.
 - [24] C.M. St Martin, J.K. Lundquist, M.A. Handschy, Variability of interconnected wind plants: correlation length and its dependence on variability time scale, *Environ. Res. Lett.* 10 (4) (2015), 044004, <https://doi.org/10.1088/1748-9326/10/4/044004>.
 - [25] P.D. Welch, The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms, in: *IEEE Trans. Audio and Electroacoust.* AU-15, 1967, pp. 70–73.