



Simulating European wind power generation applying statistical downscaling to reanalysis data

Gonzalez-Aparicio, I.; Monforti, F.; Volker, Patrick; Zucker, A.; Careri, F.; Huld, T.; Badger, Jake

Published in:
Applied Energy

Link to article, DOI:
[10.1016/j.apenergy.2017.04.066](https://doi.org/10.1016/j.apenergy.2017.04.066)

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Gonzalez-Aparicio, I., Monforti, F., Volker, P., Zucker, A., Careri, F., Huld, T., & Badger, J. (2017). Simulating European wind power generation applying statistical downscaling to reanalysis data. *Applied Energy*, 199, 155-168. <https://doi.org/10.1016/j.apenergy.2017.04.066>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Simulating European wind power generation applying statistical downscaling to reanalysis data



I. González-Aparicio^{a,*}, F. Monforti^b, P. Volker^c, A. Zucker^a, F. Careri^a, T. Huld^b, J. Badger^c

^a European Commission, DG-Joint Research Centre, Knowledge for the Energy Union Unit, Energy, Transport and Climate Directorate, Petten, The Netherlands

^b European Commission, DG-Joint Research Centre, Energy Efficiency and Renewables Unit, Energy, Transport and Climate Directorate, Ispra, Italy

^c Technical University of Denmark, Department of Wind Energy, Denmark

HIGHLIGHTS

- Wind speed spatial resolution highly influences calculated wind power peaks and ramps.
- Reduction of wind power generation uncertainties using statistical downscaling.
- Publicly available dataset of wind power generation hourly time series at NUTS2.

ARTICLE INFO

Article history:

Received 7 December 2016

Received in revised form 24 April 2017

Accepted 26 April 2017

Available online 9 May 2017

Keywords:

Wind power generation

Hourly time series

High spatial resolution wind speed

ABSTRACT

The growing share of electricity production from solar and mainly wind resources constantly increases the stochastic nature of the power system. Modelling the high share of renewable energy sources – and in particular wind power – crucially depends on the adequate representation of the intermittency and characteristics of the wind resource which is related to the accuracy of the approach in converting wind speed data into power values. One of the main factors contributing to the uncertainty in these conversion methods is the selection of the spatial resolution. Although numerical weather prediction models can simulate wind speeds at higher spatial resolution (up to 1×1 km) than a reanalysis (generally, ranging from about 25 km to 70 km), they require high computational resources and massive storage systems: therefore, the most common alternative is to use the reanalysis data. However, local wind features could not be captured by the use of a reanalysis technique and could be translated into misinterpretations of the wind power peaks, ramping capacities, the behaviour of power prices, as well as bidding strategies for the electricity market. This study contributes to the understanding what is captured by different wind speeds spatial resolution datasets, the importance of using high resolution data for the conversion into power and the implications in power system analyses. It is proposed a methodology to increase the spatial resolution from a reanalysis. This study presents an open access renewable generation time series dataset for the EU-28 and neighbouring countries at hourly intervals and at different geographical aggregation levels (country, bidding zone and administrative territorial unit), for a 30 year period taking into account the wind generating fleet at the end of 2015.

© 2017 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

The European power sector is currently experiencing a structural transition. The goal of the European Union for Renewable Energy Sources for electricity (RES-E) to provide for at least 27% of the total energy consumption by 2030 could translate into 50% of total electricity production from renewables. The Energy Union

strategy includes the aim of the European Union to become “the number one in renewables” continuing the significant growth of RES-E experienced during the last decade [1]. However, the growing share of generation from solar and mainly, wind resources constantly increases the stochastic nature of the power system, potentially jeopardizing the security of supply. As a consequence, planning and scheduling tools for the power sector have been improved to simulate the high share of RES-E. Particular care has been given to the adequate representation of the wind intermittency to better catch wind power generation peaks and ramping capacities, which are key aspects for understanding power system

* Corresponding author at: European Commission, DG-Joint Research Centre, Knowledge for the Energy Union Unit, PO Box 2, NL-1755 ZG Petten, The Netherlands.

E-mail address: iratxe.gonzalez-aparicio@ec.europa.eu (I. González-Aparicio).

flexibility needs, as well as the behaviour of market participants in defining their bidding strategies in the different electricity markets.

However, before analysing or modelling the impact of the wind on the power system, it is important to understand the wind resource characteristics, in particular the nature of their variability. The effects of the surface heterogeneities vary depending on the local area but also on larger scales (such as meso and continental scales). At wind farm level, the wind flows are influenced by the co-occurrence of meteorological variations and turbine wake effects and their nonlinear interactions, particularly over strongly heterogeneous surfaces like coastal areas or mountainous regions [2,3]. Considering that wind farms are often built over such heterogeneous surfaces, it is important to know the significance of the spatial variability as they influence power production at larger scales (countries, bidding zones and administrative territorial unit levels).

Thus, in order to analyse the European power system, the wind speed and direction data should keep a compromise between the geographical coverage accounting for European climate zones; the time intervals and a period long enough to capture the climate variability [4]. The wind speed data should reproduce the diversity of the local effects due to the orography and wind features at hub height. Attention should also be given to technical data parameters of wind turbines (e.g. hub height and power curves), the losses of performance due to the age of the turbines and the interpolation method of the wind speed data at the hub height. All these factors crucially depend on the accuracy of the approach to convert the wind speed data into power and how the uncertainties are treated (Fig. 1). It also depends on the aggregation level of the study and the smoothness of those factors. For example, at wind farm level wake effects are the main drivers for modelling power generation and then, a wind power fluctuation parameterization is required to include in the conversion such as in [5–7]. When aggregating to a regional, bidding area or country level those effects are, to some degree, smoothed and existing studies are focused on the accuracy of wind speed data and derived wind generation by developing

simplified models ([8,9]); by correcting the wind output biases with factors derived from the transmission system operators data [10]; by estimating the factors affecting the cascade of uncertainties [11–14].

However, the cascade of the uncertainties in the whole conversion process starts in the selection of the characteristics of the primary wind speed and direction data. Generally, the trend is to use weather derived data from Numerical Weather Prediction (NWP) models or from reanalyses. The use of NWP models could perform higher spatial resolution wind speed data (generally, ranging from 1×1 km to 5×5 km) than the reanalysis (so far, ranging from 25 to 70 km) but requires high computational resources and massive storage systems and therefore, most of studies use the reanalyses [15]. A good comparison between the methods using NWP models can be found in [16]; the assessment of what a reanalysis can provide for wind power and the bias associated is published in [17] and a summary of publicly available reanalysis can be found in ([18,15]). One of the most used reanalysis for power system analysis comes from the NASA atmospheric reanalysis dataset which was generated within the Modern Era Retrospective-Analysis for Research and Applications (MERRA) project [19]; in [18] the description of the local wind climate in terrain with low complexity has shown good correlation with wind measurements at relevant heights with Pearson's correlation coefficients values around 0.85 on an hourly basis and 0.94 on a monthly basis for Nordic countries and Baltic states; in [20,21] the wind power production modelling has been studied for Sweden and Great Britain, respectively; in [22] a wind energy index for site assessment, turbine selection and local feed-in tariffs has been developed for Germany; in [23] the results of offshore wind energy resource simulations forced by different reanalysis have been compared for the Iberian Peninsula; flexibility options for systems with high renewables penetration have been studied for Ireland [24] and Europe ([25,26]); a techno-economic analysis of the effects of North African electricity import on the European power system was carried out in [27]; the cost-potentials for large onshore wind turbines in Europe has been analysed in [28] while the validation of Danish

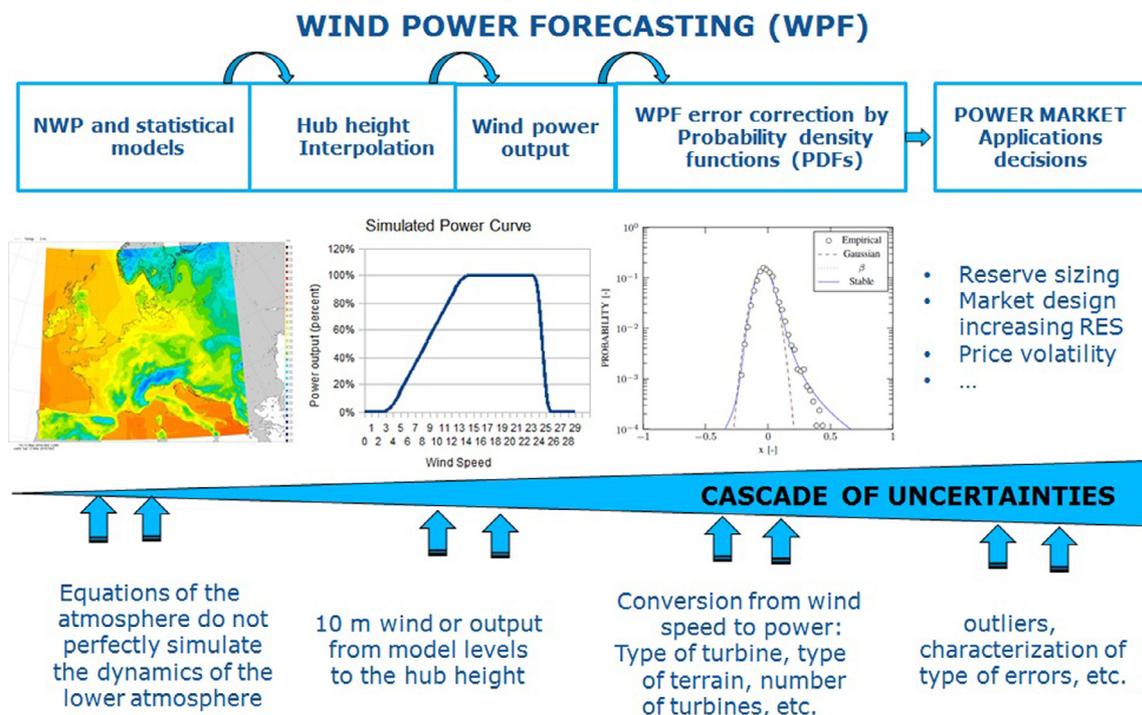


Fig. 1. General approach for the conversion of wind speed data into power and the main factors contributing to the cascade of uncertainties.

wind time series from a global renewable atlas for energy system analysis is reported in [29].

However, there is not a complete understanding concerning the adoption of different wind speeds datasets with different spatial resolutions and the implications of the results in power system models, analyses and impact assessments. Therefore, basing on the huge number of studies using reanalysis, the main novelty of this study is to show the importance of using high resolution wind speed data, derived from a reanalysis, for the conversion into power values at different aggregation levels. A methodology is proposed to increase the spatial resolution of the wind speed using data from MERRA reanalysis for Europe. Finally, the generated high resolution hourly time series has been made open-access to use in the future RES-E integration studies since there is no publicly available robust European weather derived wind power generation time series at country, bidding zone and administrative territorial unit levels (NUTS [30]). The three major publically available RES-E datasets publish annual and monthly averages of wind and solar power production (the 'Global Atlas for Renewables' [31] and the 'Global Wind Atlas' [32]) or hourly time series but only at country level (Renewables.ninja [33]).

This new methodology consists of applying a spatial statistical downscaling technique to the reanalysis wind speed data, a robust and established approach in climate modelling field [34]. By using the Global Wind Atlas data – that includes local orography and roughness features at 250×250 m spatial resolution – the approach has been applied to the hourly wind speed at each wind farm in Europe in order to capture the effect of fine-scale forcing in areas characterised by fine spatial variability of features (e.g. rugged topography, etc.) taking into account the diverse land surface and local climate conditions over the geographical European domain; therefore, the corrected wind speed has been converted into wind power generation data. For reliability and validation of the downscaled wind speed, the results have been compared with the high resolution ($12 \text{ km} \times 12 \text{ km}$) wind speed non-freely available operational forecast products released by the European Centre for Medium-Range Weather Forecast (ECMWF) [35]. The 30 years of corrected hourly wind speed data at each wind farm site in Europe have been then converted into power values taking into account the wind generating fleet in Europe at the end of 2015: then, the wind power time series have been compared with the wind power generation data provided by the Transmission System Operators (TSOs) in Europe for the year 2015.

The present study is part of the methodology applied in the EMHIREs dataset [36]: European Meteorological derived High Resolution dataset, an open access RES-E generation time series datasets developed by the European Commission, DG JRC. The wind power time series are calculated combining a highly detailed reconstructed wind farm database (WFDB) with the high spatial and temporal resolution wind speed dataset, providing a new and validated methodology. The development of the EMHIREs dataset and the overall results of the project are published in [36]; the reconstruction and sensitivities of the EU WFDB are presented in [37], while this article aims at assessing the impact of wind speed spatial resolution on power generation, contributing to the discussion of the latest tendencies in the fundamental aspects of meteorological data used for intermittent RES-E integration studies.

2. Data and tools

2.1. Data used to generate EMHIREs

The wind farms database procured from the 'thewindpower.net' [38] has been used as the primary source to define the characteris-

tics of each wind farm included in EMHIREs. The original database includes worldwide information for onshore and offshore wind farms, containing however a significant amount of gaps, inconsistencies and inaccuracies; this has been solved by means of gap filling and data homogenization. To validate the improved database the aggregated installed capacities have been compared with data from (among others) different European Transmission System Operators.

The primary source of meteorological data used in EMHIREs comes from the NASA atmospheric reanalysis MERRA dataset [19]. MERRA datasets are the output of the Goddard Earth Observing System Model v.5 (GEOS-5) and its Atmospheric Data Assimilation System (ADAS), version 5.2.0. The data streams assimilated by the GEOS-5 DAS come from radiosondes, pilot balloon winds, wind profiles, radar winds, aircraft reports, dropsondes, spectroradiometer (MODIS water vapour winds), surface land observations, surface ship and buoy observations. The native horizontal resolution is 0.66-degree longitude by 0.5-degree latitude (60×56 km in the south and 25×56 km in the north approximately in the area covered by EMHIREs: West -11° North 73° South 35° and East 40°) and it is available at 72 levels. Surface data, near surface meteorology, selected upper-air levels and vertically integrated fluxes and budgets are produced at one-hour intervals. The 30-year-period selected ranges from the 1st of January 1986 to the 31st of December 2015 from 00:30 to 23:30 Universal Time Coordinates (UTC) at hourly frequency. The variables extracted are the eastward and northward wind at 2, 10 and 50 m above displacement height (ms^{-1}).

The Global Wind Atlas [32] provides the corrections on wind speeds taking part in application of the downscaling technique of EMHIREs dataset. The GWA data-set is the only publicly available data-set that provides worldwide wind conditions that include micro-scale information. It uses the generalisation methodology [39] to include micro-scale processes (i.e. orography and roughness induced flow accelerations) to the large scale data. The use of the GWA and the downscaling could be also useful when developing scenarios of future wind power deployment if reanalysis is applied directly for this purpose, being inappropriate considering the coarse resolution.

The ECMWF dataset is used for validation of the downscaled wind speed comes from the operational forecast wind speeds at 12×12 km spatial resolution and at 100 m height, from 2012 to 2015 [40]. The new dataset represents a valuable improvement with respect to coarse reanalysis data and to the direct extrapolation of ECMWF 10-m wind, which was shown to produce a considerable degradation of energy power production with respect to observed values [40]. The new variable of ECMWF meets the need of calculating the wind speed at turbine height level, and is the result of the vertical linear interpolation from the two nearest ECMWF model levels, which are, respectively at approximately 70 and 110 m. In order to obtain hourly data, horizontal wind fields are taken from ECMWF analyses at 00:00 and 12:00 UTC, and, in the other hours, from the short-term forecasts in the range $+1$ to $+11$ h. At such a very short range, the forecasts <12 h are the analyses, since it is a combination between a short-range forecast data with observations to produce the best fit of both, so that they can be used as realistic proxies. Data cover a wide region extending from 30°N to 75°N and from 25°W to 45°E , considering both onshore and offshore grid points almost all over Europe, including Iceland. However, as the ECMWF dataset is only available for 4 years, it was not used as a primary source for generating the EMHIREs time series.

2.2. Actual wind power generation time series and statistics

The calculated wind power time series in EMHIREs are validated against the actual wind power generation outputs provided

by the TSOs for the year 2015 at country level and by bidding zone. The main source for TSOs time series is the Transparency Platform provided by the European Network of Transmission System Operators for Electricity (ENTSO-E) [41] in agreement with Regulation 543/2013 [42]. This database has been consulted last time in February 2017: in case data were not available on the ENTSO-E transparency platform (e.g. Croatia or Italian bidding zones) or contained significant amount of missing values (e.g. United Kingdom, Republic of Ireland, Cyprus), data from the corresponding TSO was preferred as a source. Regardless this, for Bulgaria, Luxembourg, Slovenia and Slovakia data were not available (Table 1). To crosscheck the level of accuracy of the ENTSO-E and national hourly time series, the sum over all hours in 2015 is compared with the annual generation reported by the same source in a different section of its web data repository. It is observed that there are mismatches for most of the countries between the total annual production reported and the sum of the hourly reported values. The ENTSO-E time-series include hours that are not registered while the total annual generation could have been metered and reported separately e.g. for support scheme payments. Moreover, according to Regulation 543/2013, all the information shall be provided for all bidding zones only in Member States with more than 1% feed-in of wind or solar power generation per year or for bidding zones with more than 5% feed-in of wind or solar power generation per year. The reasons behind possible mismatches could be different: that national borders do not always coincide with bidding zones (e.g. Denmark, United Kingdom), renewable energy produced might be outside of the interconnected system (not interconnected islands like in Greece) or generation by auto-producers might not be included in TSOs statistics. For the validation at NUTS 1 and NUTS 2 level, regional statistics have been searched but, for most of the countries, neither time series nor monthly or annual statistics yet available for 2015. Some EU Member States (e.g. the UK) will publish annual statistics are on a regional level at the time of release of this paper. In the case of Spain [43], there are monthly statistics of wind generation by region, and in the case of France [44] and Finland, the statistics available show the total annual production by regions.

In order to keep the consistency with yearly statistics the EMHIREs wind power time series have been normalised to the ENTSO-E annual production statistics reported (last column in Table 1), although the bias with the hourly time series will be implicitly in the analysis.

3. Description of the methodology

3.1. Statistical downscaling of wind speed hourly variations

MERRA hourly eastward and northward wind components at three different heights 2, 10 and 50 m are interpolated at each wind farm location to calculate the wind speed and direction, using the arctangent functions to convert angles into radians and then downscaled using a robust and established technique. This technique is based on a probabilistic approach and aimed at predicting the changes in the probability density function (pdf) of local scale wind speed conditioned to large-scale hourly wind speed predictors [45,46]. The analytical expression summarizing the methodology was developed in [47]. In this study it was applied a downscale daily wind speed time series from a meso-scale meteorological model to the wind farm level. EMHIREs is based upon the same algorithms used in [47] but the large scale time series are available at hourly level while the microscale hourly wind speed distribution is provided by the Global Wind Atlas (GWA). The downscaling technique is not applied to the offshore wind farms because they are not affected by orography and roughness. Hence the offshore

wind power generation time series are directly produced with MERRA primary data.

For each wind farm location, a proper Weibull distribution function best fitting the MERRA interpolated hourly data series is computed and parameters A_{rean} and k_{rean} determined for both the 10 m and 50 m heights. In the same locations and for the same heights, A_{micro} and k_{micro} given by Global Wind Atlas are also collected. For both probability distribution functions, the related cumulative distribution functions F_{micro} and F_{rean} are computed by the Weibull distribution properties as

$$F_x(X) = 1 - e^{-\left(\frac{X}{\lambda}\right)^k} \quad (1)$$

Each value of x_{rean} arising from the EMHIREs hourly time series (1) is then associated to the value of x_{micro} leading to equal values of F_{rean} and F_{micro} , as described in Eqs. (2) and (3)

$$F_{\text{micro}}(X_{\text{micro}}) = F_{\text{rean}}(X_{\text{rean}}) \quad (2)$$

$$1 - e^{-\left(\frac{X_{\text{micro}}}{A_{\text{micro}}}\right)^{k_{\text{micro}}}} = 1 - e^{-\left(\frac{X_{\text{rean}}}{A_{\text{rean}}}\right)^{k_{\text{rean}}}} \quad (3)$$

This leads to the following direct relation between x_{micro} and x_{rean} that has been implemented in the downscaling software

$$X_{\text{micro}} = A_{\text{micro}} \left(\frac{X_{\text{rean}}}{A_{\text{rean}}} \right)^{\frac{k_{\text{rean}}}{k_{\text{micro}}}} \quad (4)$$

To calculate the wind power generation time series, the EMHIREs wind speed time series are then vertically interpolated to the hub height of each wind farm using a logarithmic profile

$$\text{WS}_{h_2}/\text{WS}_{h_1} = (h_2/h_1)^\alpha \rightarrow \ln\left(\frac{\text{WS}_{h_2}}{\text{WS}_{h_1}}\right) = \alpha * \text{Ln}\left(\frac{h_2}{h_1}\right) \quad (5)$$

where WS_h represent wind speed at height h . Once α is identified, the same profile is used to estimate the WS at the given hub height of each wind farm (WS_{HH}). The value of α is calculated using the MERRA-derived wind speed time series at 10 and 50 m height

$$\text{WS}_{h_2} = \text{WS}_{h_1} \frac{\ln\left(\frac{h_2}{z_0}\right)}{\ln\left(\frac{h_1}{z_0}\right)} \rightarrow \alpha = \frac{\ln\left(\frac{\text{WS}_{h_2}}{\text{WS}_{h_1}}\right)}{\ln\left(\frac{h_2}{h_1}\right)} \quad (6)$$

$$\text{WS}_{\text{HH}} = \text{WS}_1 \left(\frac{\text{HH}}{h_1}\right)^\alpha \quad (7)$$

However, although a finer spatial resolution account for additional processes, in complex atmospheric conditions it adds an extra factor to the cascade of uncertainties. Therefore, to assess the degree of improvement of the downscaling technique the MERRA and EMHIREs wind speed datasets are compared with ECMWF high resolution wind speeds. This comparison allows analysing the variability and correlation of wind speeds at different spatial resolutions (i.e. MERRA at $60 \text{ km} \times 56 \text{ km}$ in the South and $25 \times 56 \text{ km}$ in the North; EMHIREs site level correction and aggregated at different levels and ECMWF at $12 \text{ km} \times 12 \text{ km}$).

3.2. Conversion into wind power generation

The 30 years of corrected wind speed at hub height are converted into power using the reconstructed wind farm database, taking into account the wind fleet in Europe at the end of 2015. The wind farms selected are in production phase (status = production) and commissioned before end 2015 (commissioning year <2016 or absent). The database completeness, gap filling and statistical approximations are based on the criteria followed by [37]. They take into account technological aspects such as the installed capacity of each wind farm, number of turbines, the turbine types,

manufacturer, hub height, swept area, power, minimum, maximum and nominal power, the geographical location, onshore/offshore and distance to the shore, commissioning year and status. In the EMHIRE dataset, one power curve is assigned for each wind farm considering its characteristics: design of the turbine and manufacturer, the swept area of the turbines, the installed power and the minimum, nominal and maximum wind power. Firstly, the power curves are built using as primary data the turbine database from [38]. Additional databases including information of power curves given by manufacturer are also used for completeness [48]. After the completeness, the missing records without power curves information are 28% and a power curve parameterization has been applied following the methodology applied in [37].

After the reconstruction, the database includes 1061 power curves from the 160 manufacturers registered in Europe at the precise hub height of the turbines; there are 16,171 wind farms located in European countries, 80 wind farms registered of which are offshore. The database produces a result that matches 95% of ENTSO-E statistical factsheet of installed capacity, while the original database matched 89%.

4. Results and discussions

4.1. Comparison of wind speed from different datasets

This section presents the statistical comparison of hourly wind speed time series at 100 m for 2012–2015 between MERRA, EMHIRE and ECMWF datasets.

The internal consistency of the three datasets is assessed by the Pearson's correlation coefficient (R) and to gauge their statistical significance, the Student's t -test is applied. For each wind farm in Europe, MERRA EMHIRE and ECMWF wind speed time series have been found to be highly correlated between each other with $R_{\text{mean}} = 0.88\text{--}0.89$, $R_{\text{min}} = 0.48\text{--}0.42$ and $R_{\text{max}} = 1.0\text{--}1.0$ once the results are aggregated at country and NUTS-2 level, respectively. In all cases, EMHIRE dataset shows a high consistency with the original data (MERRA) with a correlation greater than 0.95. According to these general results, it can be stated that the three datasets capture similar features of wind patterns on the European continental scale.

Nevertheless, a deeper analysis shows how the details of this broad picture differ when prediction skills are closely compared. Averaging the correlation by country it is observed that in all cases the $R_{\text{MERRA-ECMWF}}$ is very similar with the $R_{\text{EMHIRE-ECMWF}}$ with a difference of $R = 0.02$ (with a level of significance $p < 0.05$). This similarity can suggest that at this level of aggregation, the local effects due to the orography that both ECMWF and EMHIRE could introduce are smoothed. On the contrary, by disaggregating from country to NUTS-2, there are regions where the $R_{\text{EMHIRE-ECMWF}}$ is visibly higher than $R_{\text{MERRA-ECMWF}}$; that means EMHIRE and ECMWF both capture more variability than MERRA. These results occur in the 20% of the NUTS 2 regions, with the highest correlation of 0.943 and with a difference of $R_{\text{EMHIRE-ECMWF}} - R_{\text{MERRA-ECMWF}} = 0.20$. The NUTS 2 with such differences in the correlations are coastal areas in Spain, Germany, Greece, Romania, Portugal, Norway and United Kingdom.

Wind farm sites are extremely heterogeneous across Europe and this could indicate that EMHIRE indeed introduces more variability in the dataset although its actual added value in properly assessing the local wind effects could differ site by site. For this reason, site level data have been deeper analysed by crosschecking the variability, the dispersion and spread of the datasets. While low standard deviation indicates a dataset is closer to the mean and has lower variability, high standard deviation shows that data points are spread out over a wider range of values so the dataset is more

dispersed. This behaviour can be observed indeed in the scatter density plots of Fig. 2 and the boxplots of Fig. 3. A visual comparison of the scatter density plots indicates the variability is higher and more spread in the case of EMHIRE (wind speed corrected at site level) and in the ECMWF (wind speed extracted at 12 km \times 12 km resolution) with respect to MERRA datasets (60 km \times 56 km in the South and 25 \times 56 in the North, approximately). The scatter density plot between MERRA-EMHIRE (1) shows that the standard deviation of EMHIRE is more spread out and the range is higher than in MERRA, the cloud is shifted upwards to $x = y$ axis and distributed over the first quadrant. In the case of MERRA-ECMWF (2) the pattern is similar to MERRA-EMHIRE but lower. On the contrary, the comparison between EMHIRE and ECMWF (3) indicates that EMHIRE has slightly more spread than ECMWF but less than with MERRA. In this case, the cloud is closer to the $x = y$ axis.

The boxplots represent the average absolute deviation of the dataset, that is, the average of the absolute deviations from a central point. In this case, the central point is the median of the interquartile range. The boxplots show the difference of the mean between mean (EMHIRE-ECMWF) and mean (MERRA-ECMWF): the negative values indicate that the difference between the mean of EMHIRE and ECMWF is lower than the mean between MERRA and ECMWF showing that EMHIRE and ECMWF contribute with more variability than MERRA. It is also possible to highlight that EMHIRE simulates higher wind speed values than MERRA (negative values for most of the countries) since wind farms are typically built on sites with higher wind resources that are better captured thanks to the statistical downscaling procedure.

The statistical results obtained so far are in line with the physical behaviour of the wind speed variability at different resolutions. In coarser resolutions (MERRA) the effects of capturing fine-scale forcing – in particular in areas characterised by fine spatial variability of features such as rugged topography and very diverse land surface conditions – are smoothed and the variability of the wind speed is underestimated with respect to the dataset where the downscaling is applied. For countries with little orography (such as Belgium, Denmark, Netherlands and United Kingdom) MERRA does not account for the local roughness and can slow down the winds significantly with respect to EMHIRE and ECMWF, over predicting the wind speed and introducing less variability than EMHIRE and ECMWF. This behaviour is observed in the wind speed duration curves for those countries of the Fig. 4. Since the scope of this study is to improve the quality of the wind power time series, the improvement of the EMHIRE is analysed by comparing with the actual wind power generation time series provided by the TSOs.

4.2. Wind power generation at different aggregation levels

The simulated wind power time series derived from the EMHIRE, MERRA and ECMWF wind speed datasets are validated with the actual wind power generation at different aggregation levels using time series and statistics provided by TSOs. Although the time series provided by the TSOs are not complete, they are used to compare MERRA, EMHIRE and ECMWF datasets.

4.2.1. Statistical indicators

The first comparison is carried out using hourly time series by country and by bidding zone for 2015. Further, since the bidding zones are smaller regions than country areas (e.g. Norway, Sweden, Denmark and Italy) both aggregation levels are compared to assess the improvement of EMHIRE with respect to the MERRA-derived wind power. It can be noted that the method applied to convert wind speed into wind power generation is indeed a source of uncertainty but this is also related to the TSOs data mismatches

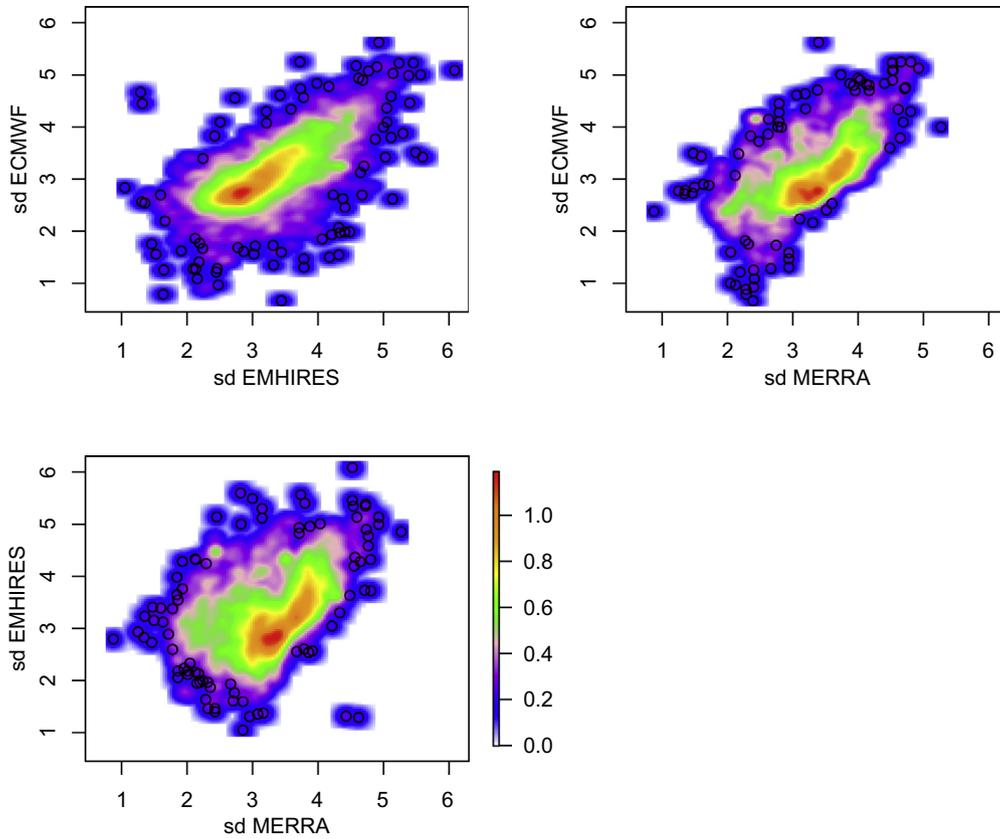


Fig. 2. Scatter density plots of the standard deviation of each hourly time series of 2012–2015 for (a) MERRA-EMHIREs, (b) ECMWF-EMHIREs and (c) MERRA-ECMWF.

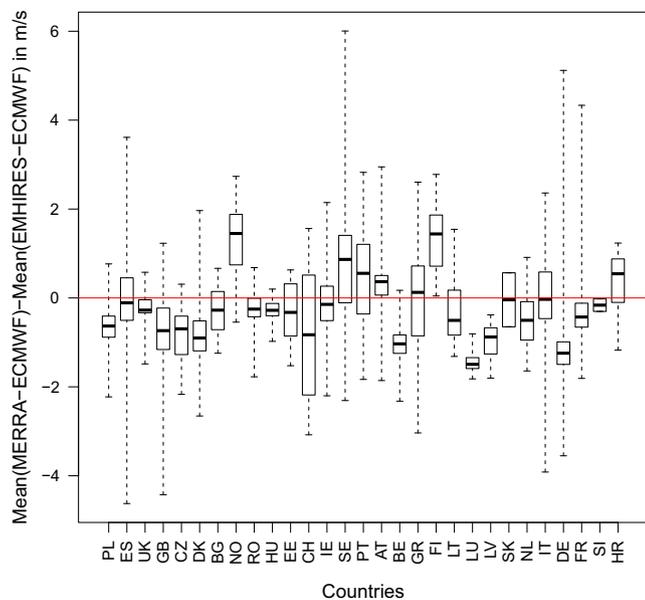


Fig. 3. Boxplots of the difference between the mean of MERRA-ECMWF and the mean of EMHIREs-ECMWF by country.

already mentioned (Table 1). ECMWF wind speed of 2015 has been also converted into wind power to be compared with the EMHIREs and MERRA statistical performance.

The overall results of the statistical performance for each country and bidding zone are summarized in Tables 2 and 3 for MERRA (a), EMHIREs (b) and ECMWF (c), respectively in the [supplemen-](#)

[tary material](#). The fractional bias (FB) measures the mean bias and indicates only systematic error which leads to an underestimation or overestimation (in the range of ± 2.0 ratio) of the TSO values. There are no significant differences between the FB associated to ECMWF, MERRA and EMHIREs by country and by bidding zone. A tendency to underestimate is found in some countries while others are overestimated, but in the three datasets the systematic error has a similar level and pattern. The Pearson's linear correlation coefficient indicates that the three datasets have good internal consistency; they range between 0.8 and 0.97 when comparing with the TSO time series, except for Cyprus (0.55), Bulgaria (0.76), Switzerland (0.56) and Croatia (0.80). Note that for those countries, the ENTSO-E time series mismatch severely with the total generation provide by the same source in the annual statistical factsheet (Table 1 last two columns). Therefore, the lower correlation between the three simulated datasets and the TSO data also may be due to inhomogeneities in the latter. Other reason of the lower correlation could be the extreme complex terrains characterised in those countries leading to greater bias of the simulations. EMHIREs shows a better correlation to the TSO time series than MERRA (in green in Table 2b) in Belgium, Germany, Denmark, Estonia, Finland, France, Hungary, Lithuania, Latvia, Netherlands and Portugal and for the bidding zones of Norway (NO4), Sweden (SW1) and Italy (SUD, SICI). EMHIREs also shows an improvement in the internal consistency. In addition, the mean error (ME), the difference between standard deviations (SD) and the root mean square error and the unbiased root mean square error (RMSE and RMSE_{ub}) are computed to gauge the simulation's accuracy. Indeed, high values of RMSE_{ub} indicate a high level of non-systematic (i.e., random) discrepancy between the simulations and the TSO data. In addition, the ability of a simulation to reproduce the "real" values is also assessed following the criteria defined by [49] consisting of: (1)

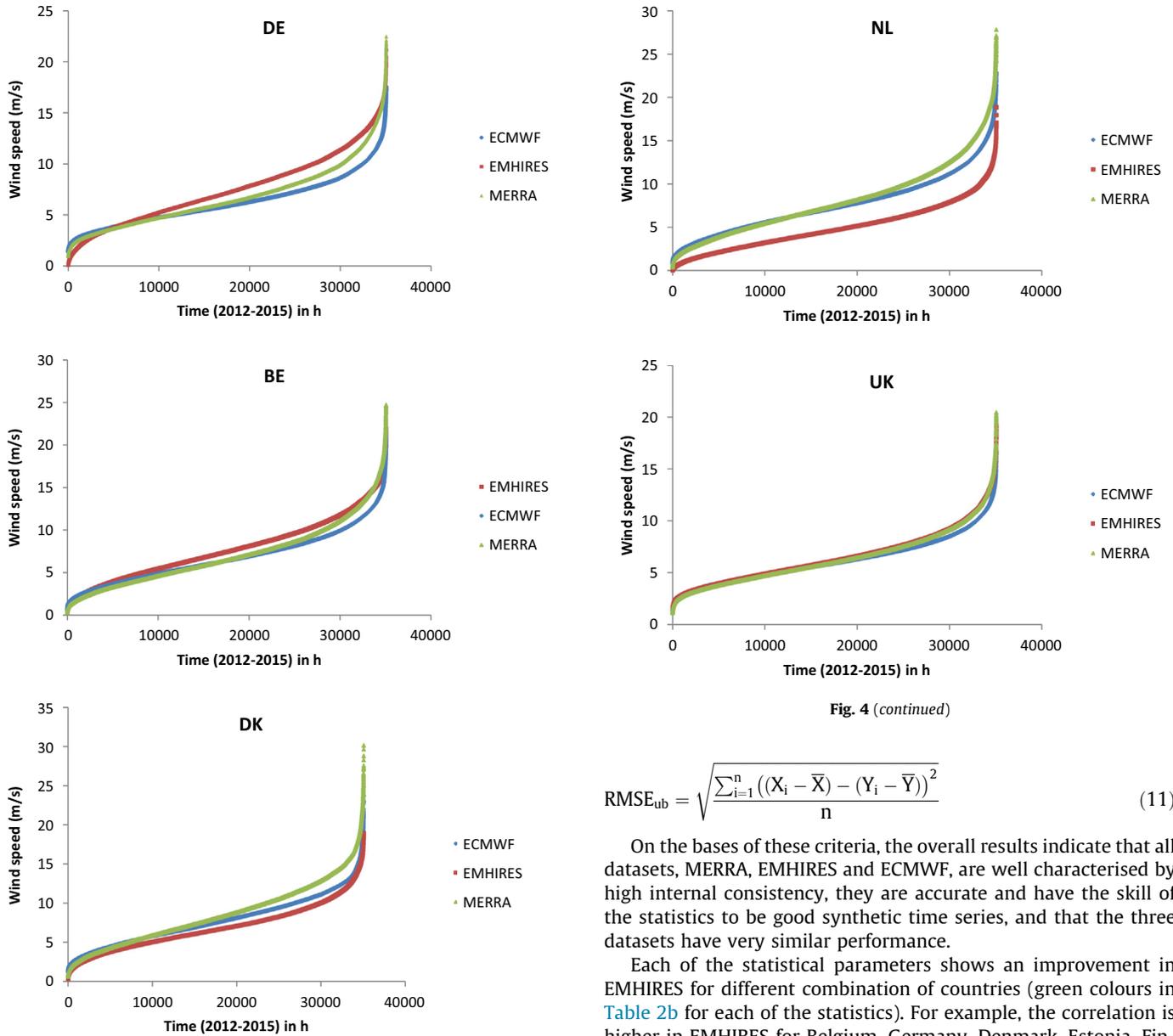


Fig. 4 (continued)

$$\text{RMSE}_{\text{ub}} = \sqrt{\frac{\sum_{i=1}^n ((X_i - \bar{X}) - (Y_i - \bar{Y}))^2}{n}} \quad (11)$$

On the bases of these criteria, the overall results indicate that all datasets, MERRA, EMHIREs and ECMWF, are well characterised by high internal consistency, they are accurate and have the skill of the statistics to be good synthetic time series, and that the three datasets have very similar performance.

Each of the statistical parameters shows an improvement in EMHIREs for different combination of countries (green colours in Table 2b for each of the statistics). For example, the correlation is higher in EMHIREs for Belgium, Germany, Denmark, Estonia, Finland, France, Hungary, Ireland, Lithuania, Latvia, Netherlands and Portugal and the standard deviations is closer to the TSO for most of the countries except from Austria, Cyprus, Finland, Croatia and Sweden. In the case of the bidding zones, the correlation coefficients hover around 0.8–0.95, decreasing in regions where terrain is extremely complex (e.g. Norway North (NO3 and NO5); Sweden (SW1) and Northern Italy (NORD and CNOR)). Over the 25 countries and 16 bidding zones analysed, all of them except for Cyprus, Czech Republic, Spain, Poland, Switzerland, and Italy-NORD bidding zone and Italy-SARD bidding zone have better internal consistency, or they are more accurate when EMHIREs series are considered with differences in the standard deviations, the ME, MSE, or RMSEub improving with respect to MERRA. Moreover, although EMHIREs is more variable and dispersed than MERRA, it has lower differences with respect with the actual values.

In the cases EMHIREs time series do not improve with respect to MERRA time series, the results have small differences and in no case EMHIREs visibly worsen the time series with respect to the TSO-time series. If the overall performance is considered at European scale (aggregated statistics) then, it is observed that EMHIREs improves in Europe except from Spain, Cyprus, Czech

Fig. 4. EMHIREs (red), MERRA (blue), ECMWF (green) 100 m wind speed duration curves for 2012–2105 for Belgium, Denmark, Netherlands and United Kingdom. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the simulated and TSO standard deviations are similar; (2) the RMSE are lower than the standard deviation and (3) the unbiased RMSE (RMSE_{ub}) which represents the accuracy of the MERRA and EMHIREs is also lower than the standard deviation.

$$\text{FB} = \frac{\sum_i (X_{i_{\text{obs}}} - Y_{i_{\text{simu}}})}{0.5 * \sum_i (X_{i_{\text{obs}}} + Y_{i_{\text{simu}}})} \quad (8)$$

$$\text{ME} = \frac{\sum X_i - Y_i}{n} \quad (9)$$

$$\text{SD} = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}} \quad (10)$$

Republic, Poland and Switzerland and in the bidding zones except from Nord Italy and Sardinia (green colours in the country code in Tables 2b and 3b). The Fig. 5 shows the overall performance for all the countries and bidding zones, taking into account weighted factors based on the installed capacity for each country and bidding zone and the total installed capacity in Europe, as reference. Thus, the size of the wind power system in each of country and zone is also measured to gauge the accuracy of the overall performance at European scale.

ECMWF derived wind power time series show a clear improvement in most of the statistical indicators for Austria, Belgium, Bulgaria, Cyprus, Germany, Denmark, Estonia, Greece, Netherlands and Poland with respect to EMHIREs. For the rest of the countries, both datasets (ECMWF and EMHIREs) are very similar in the performance of the statistics and, in all cases EMHIREs is closer to the ECMWF indicators than MERRA. By assessing the same indicators at bidding zones, the results show that EMHIREs is the dataset representing better the TSO time series. In most of the cases, EMHIREs shows an improvement with respect to MERRA while only in two bidding zones (Norway NO2 and NO3) ECMWF results have better performances. This may indicate that by increasing the aggregation levels from country to bidding zone, EMHIREs better captures the wind power variability than MERRA and ECMWF.

The Fig. 6a shows the quality of agreement between the MERRA, EMHIREs and ECMWF datasets and the TSO data in the form of Taylor diagram for Belgium, Germany, Denmark, Spain, France, Netherlands and United Kingdom. Those diagrams assess comparatively the modelled and observed data by the use of the Pearson correlation coefficient, the root mean square error and the standard deviation. The normalised standard deviation is higher for EMHIREs and ECMWF than for MERRA and the indicators representing highest resolution are closer to each other than to MERRA. The Fig. 6b shows the same indicators for several bidding zones of Italy, Norway, Sweden and Denmark. It is observed that EMHIREs is closed to ECMWF except for the case of Italy-CNOR where the Italian Alps are located. The very complex terrain may be the reason why the EMHIREs does not capture the real effects of the local climate.

4.2.2. Duration curves

Wind power duration curves for MERRA, EMHIREs, ECMWF and TSO data are shown in Figs. 7 and 8. It is observed that for the countries and bidding zones where the EMHIREs statistics are better than MERRA, also the cumulative distributions are closer to the TSO data, mainly at the highest wind power values. It is observed

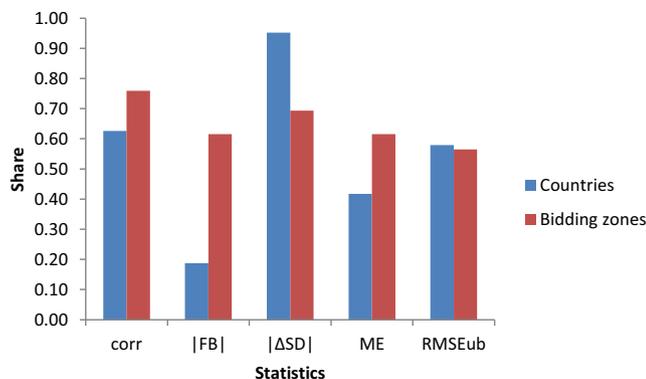


Fig. 5. Share of countries and bidding zones where EMHIREs performs better than MERRA for different metrics (based on the Tables 2 and 3). The share is computed accounting for weighted factors in each country and bidding zone with the installed capacity, considered the total European Installed capacity as the 100%.

that EMHIREs is closer to ECMWF duration curve and both are closer to TSO data than MERRA.

Assessing the extremes of the curves it is observed that EMHIREs and ECMWF reproduce higher values than MERRA in some countries with little orography (Belgium, Denmark, Netherlands and United Kingdom, mainly). The flattening of MERRA's power duration curve could be related to higher wind speeds in MERRA than in ECMWF and EMHIREs, since MERRA does not account for the local roughness that can slow down the winds significantly. Because of the higher wind speeds in MERRA rated power is reached at an earlier stage in the duration curve, as it was also observed in Fig. 4.

An example of the statistical significance (Student's *t*-test) between MERRA and EMHIREs with TSO is indicated in Table 4. The significance measured by the *p*-value indicated the probability of obtaining the result equal to or more extreme than was actually observed and is considered statistically significant when $p < 0.05$. The table includes for each pair of datasets the *t* indicator and the *p*.value, showing that in all cases the datasets follow a Student's *t*-distribution under the null hypothesis. In general, there is overestimation in the three datasets that could be due to the installed capacity considered for the power conversion, dated at the end of 2015, while the TSO data span the whole 2015. Differences may also be due to curtailment episodes, maintenances of the wind farms or caused by the uncertainty intrinsically associated with the methodology. On the other side, it has been shown in Section 2.2 that TSO data are not always consistent.

4.2.3. Time series and ramping rates

The overall statistical performance shows good results in all datasets, which means that they are able to reproduce the wind power generation with similar errors. However, EMHIREs incorporates higher variability improving the wind power time series, being closer to the ECMWF variability than MERRA.

The direct comparison of modelled and TSO duration curves and time series can provide further useful information on the suitability of EMHIREs in reproducing actual data. For instance, the Fig. 9 shows the wind power generation time series of the four datasets (ENTSO-E, MERRA, EMHIREs and ECMWF) for Belgium. It is observed that MERRA is not able to reproduce the wind power generation peaks as well as EMHIREs and ECMWF. Once again it is very likely that the difference results from the spatial resolutions of the wind speed; the coarser resolution is not able to reproduce the variability and local effects of the wind speed. Those effects are smoothed and the main consequence is the underestimation of the wind power peaks. This effect is even more pronounced when the time series are more spatially disaggregated by bidding zones; for example in Denmark (DK1), Fig. 10. In those cases the improvement of the EMHIREs is more significant and it is observed that EMHIREs and ECMWF are able to reproduce the ramps better than MERRA. Although in some cases EMHIREs overestimates the peaks of wind power generation, the statistical analysis indicate that the contribution to the uncertainty is lower than the improvement of the results.

On the physical basis, the results show that the increased power simulated by EMHIREs may be the overall effect of wind turbines being sited in favourable locations with speed up due to orographic or roughness effects, which are captured by Global Wind Atlas and then properly transferred to EMHIREs, but missed by MERRA. This effect can be observed at country and by bidding zone aggregation levels and the comparison with the wind power derived from the ECMWF high resolution wind speeds also confirmed those results at the two levels of aggregation.

In order to assess the quality of EMHIREs in capturing the sudden increase or decrease of power characterised by large positive or negative hour by hour differences, the ramp rate distribution

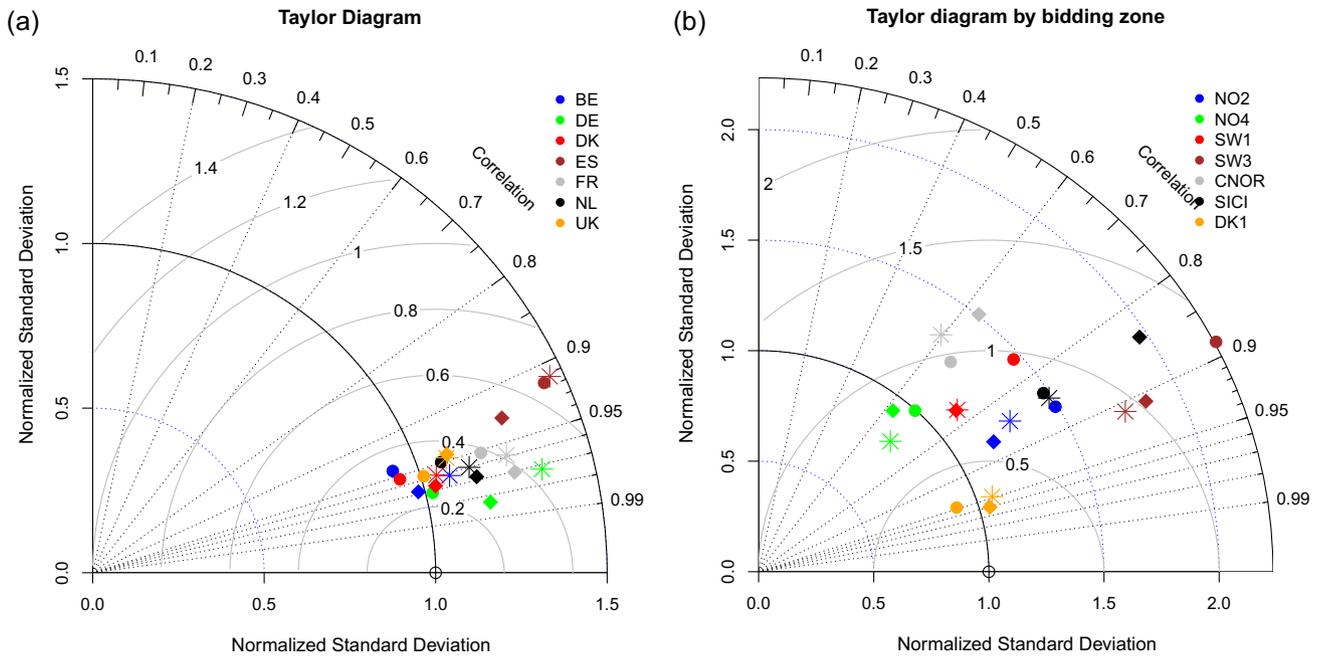


Fig. 6. Taylor diagrams of the statistical indicators by (a) countries and (b) bidding zones. The asterisk, rhomboid and circle correspond to EMHIRES, ECMWF and MERRA datasets, respectively.

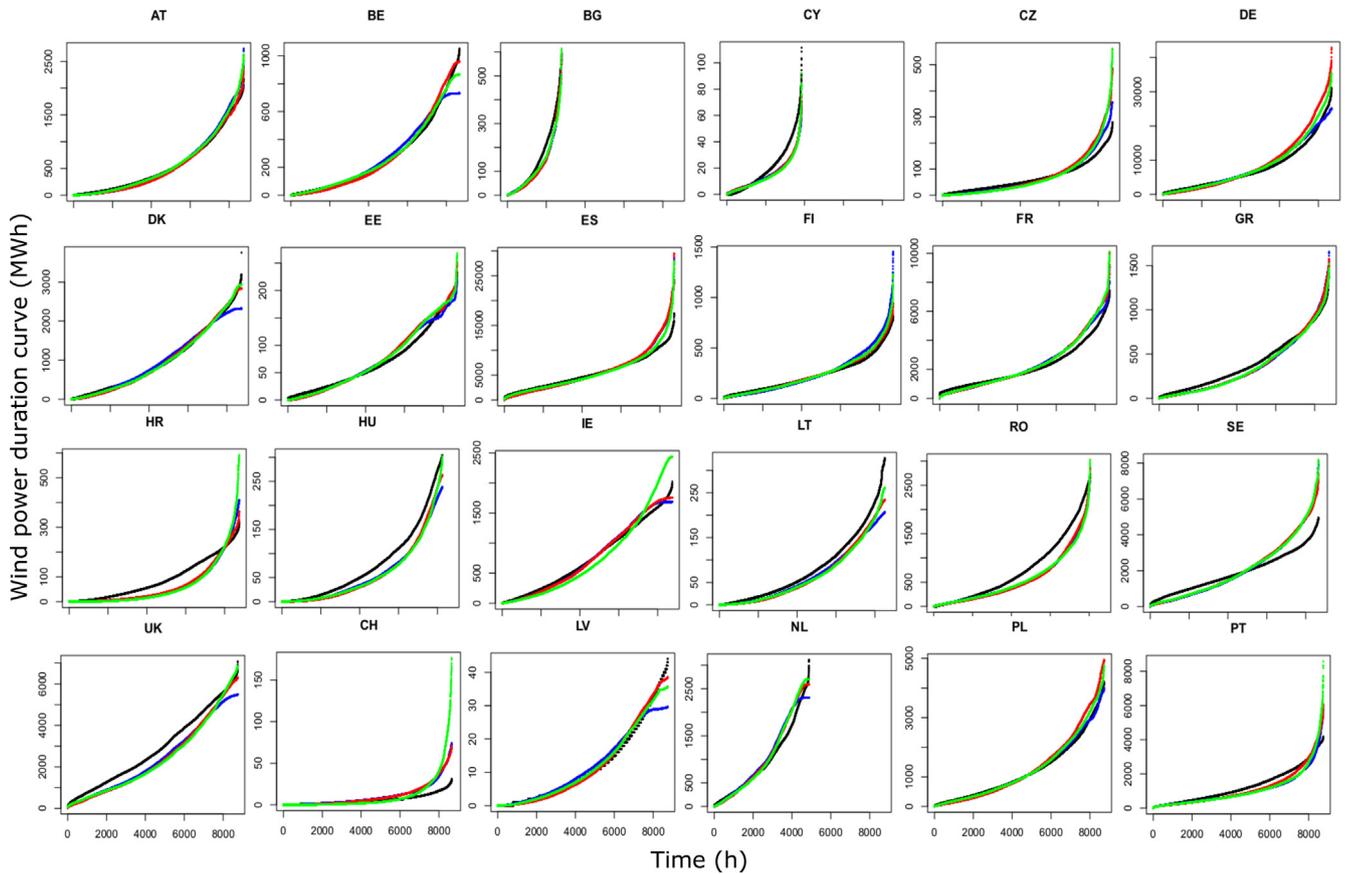


Fig. 7. EMHIRES (red), MERRA (blue), ECMWF (green) and ENTSO-E (black) wind power duration curves for 2015 by country. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

is calculated. The following plots (Fig. 11) show the frequency distribution of the ramping rates, calculated by the difference between the power production at hour (h) and at (h-1), namely

($WP_{t-(t-1)}$) in the countries with significant installed capacity (Spain, Denmark, United Kingdom and Germany). The histograms represent the TSO data, divided into 100 intervals in order to take

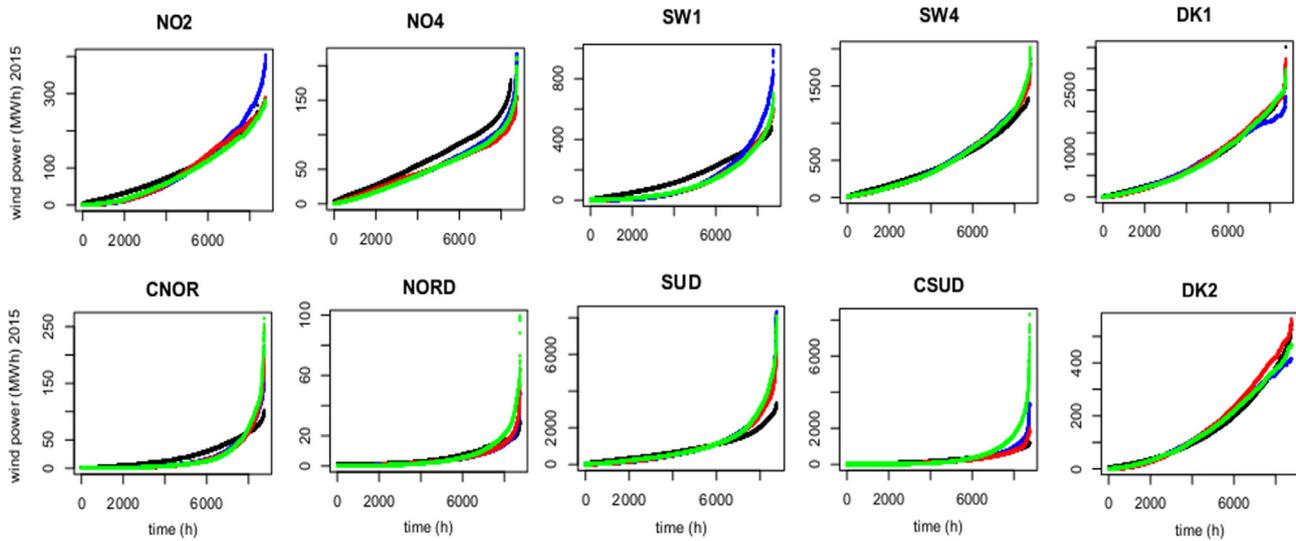


Fig. 8. EMHires (red), MERRA (blue), ECMWF (green) and TSO (black) derived wind power duration curves for 2015 selected bidding zones in Norway, Italy and Sweden. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

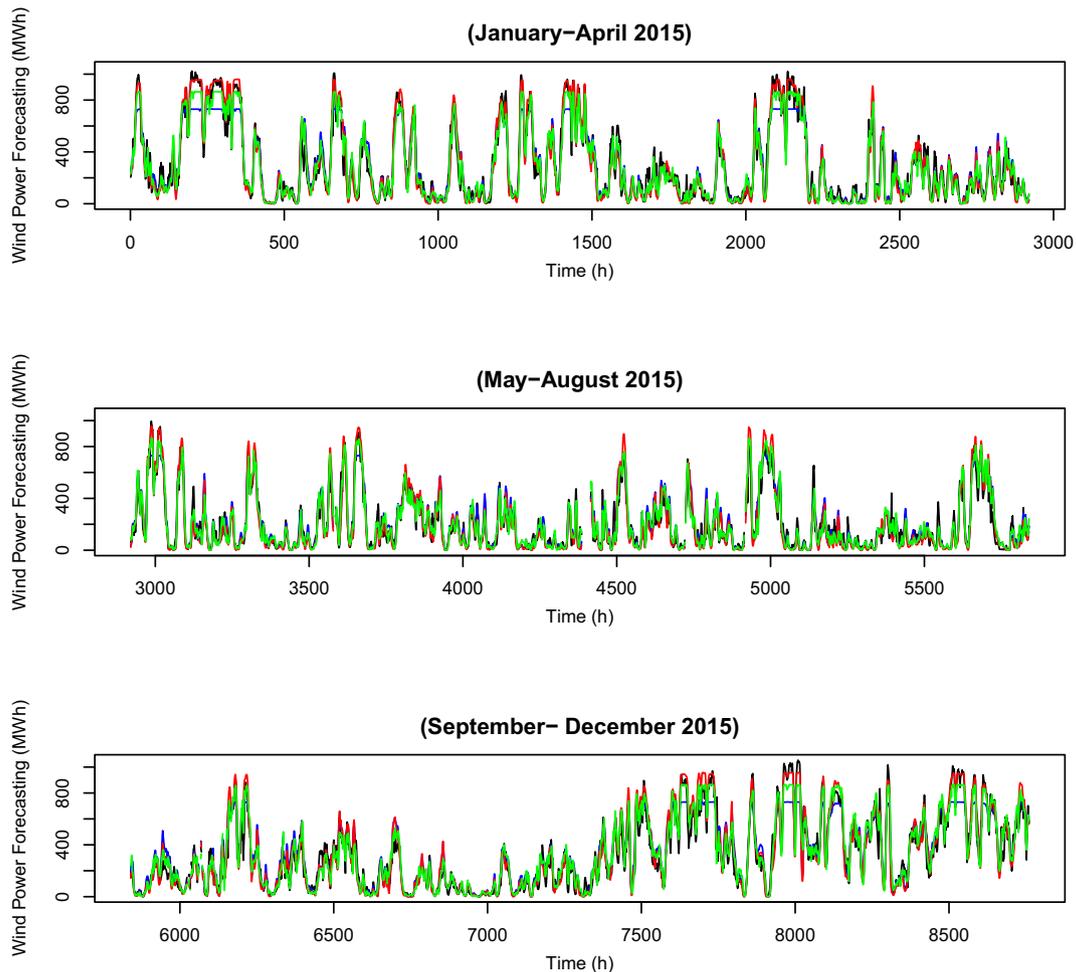


Fig. 9. EMHires (red), MERRA (blue), ECMWF (green) and TSO (black) wind power hourly time series for Belgium in 2015. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

into account the minimum and maximum power difference. The distribution curves correspond to EMHires (red) and MERRA (blue). The range of the plot is $\pm 10,000$ MWh in order to compare

the ramping rates peaks of all countries considered. In general, the distribution of MERRA is steeper in the “bell” and less prolonged in the “edges” of the distribution than EMHires. This

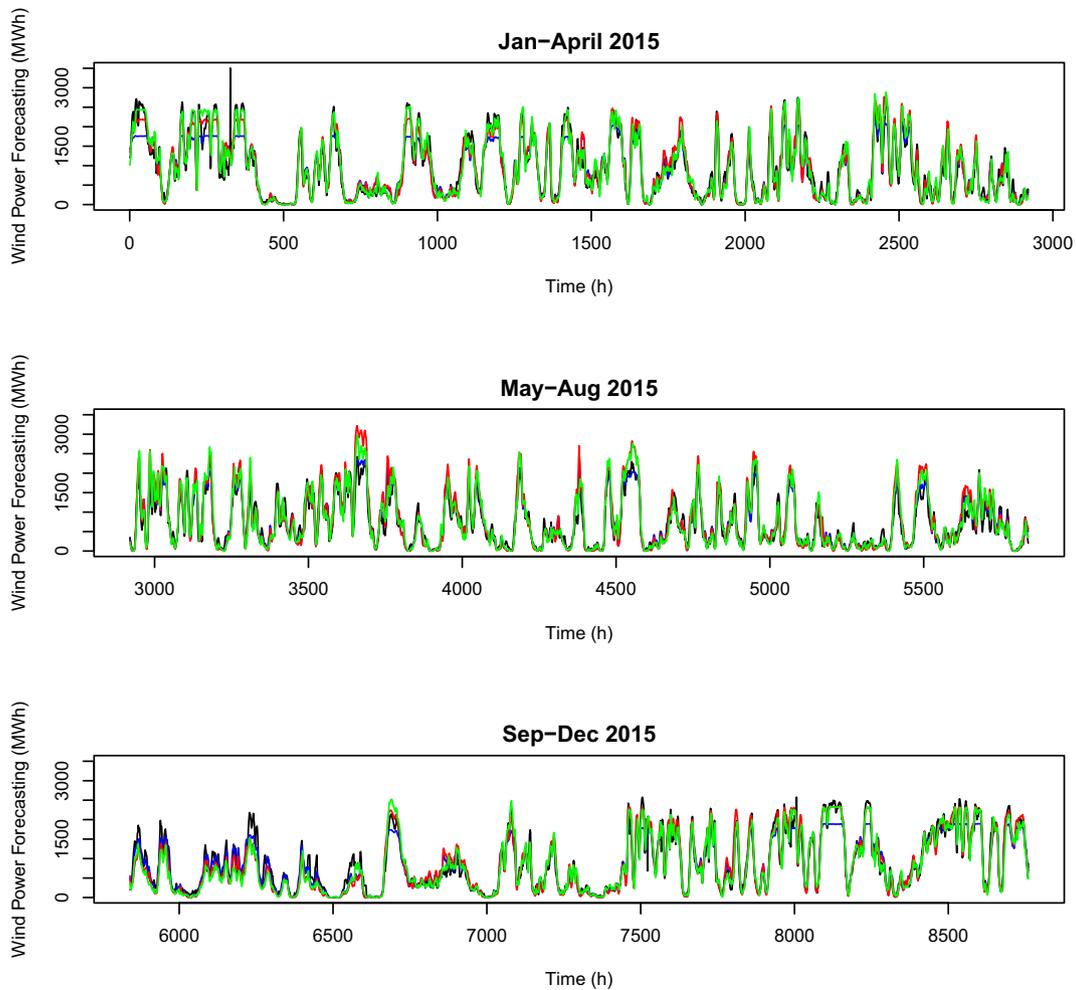


Fig. 10. EMHIREs (red), MERRA (blue), ECMWF (green) and TSO (black) wind power hourly time series for one of the bidding zones in Denmark (DK1) in 2015. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

reflects once again the underestimation of the wind speed due to smoothing, i.e. not taking into account the local effects at the sites of the wind farms. It can be seen that the ramping rates of EMHIREs are better representing the TSO data.

In the case of Denmark, the 95% coverage of measured ramp rates occur in an interval of a $WP_{t-(t-1)}$ -600 to 600 MWh. Both cases, MERRA and EMHIREs are able to capture the negative ramping rates in the 99.5% of the hours. In the case of large sudden positive increases (the 5% coverage of the most extreme ramp rates), EMHIREs improves with respect to MERRA. MERRA is not able to capture differences greater than 600 MWh while EMHIREs is able to reproduce 4 ramps that occurred in the range of 600 and 1200 MWh. On the contrary, during the 2015, there were 2 h with a positive increases of 1200 MWh and 1800 MWh respectively that neither dataset is able to capture.

In the case of Germany and Spain, EMHIREs and MERRA slightly underestimate the ramping rates with respect to TSOs time series. In Germany, the 95% coverage of the ramping rates occurs mainly between -1000 and 2000 MWh, and the datasets capture the sudden changes in a 92% of these hours. For Spain, the 95% coverage ranges between -2000 and 2000 MWh and EMHIREs captures the 97% of the hours. There is one positive maximum of 8000 MWh in Germany and $10,000$ MWh in Spain and one negative maximum of -6000 MWh and -9000 MWh in Germany and Spain, respectively. Neither EMHIREs nor MERRA are able to capture them.

EMHIREs shows a noticeable improvement with respect to MERRA in the United Kingdom. The 95% coverage ranges between -940 and 1300 MWh redistributed into three different subintervals. That is, there are 340 h with sudden decrease between -940 and -340 MWh and EMHIREs captures 50% of the hours compared with 35% captured by MERRA. There are 7278 ramping rates between -340 and 200 MWh and EMHIREs overestimates the changes in 8% of the cases while MERRA overestimates in 15%. Finally, there are 14 large changes between 700 and 1300 MWh. While MERRA is not able to capture any of these ramps, EMHIREs captures 7 cases. According to the TSO data, the positive and negative maximums occurred once at 6000 MWh and at -5000 MWh, this time neither EMHIREs nor MERRA are able to capture them.

For the four countries analysed, EMHIREs shows a clear improvement with respect to MERRA in the 95% coverage of ramping rates during 2015 and it also better captures the large negative sudden increases of wind power out of the 95% of the cases. Finally, the standard deviations and the extreme values for MERRA, EMHIREs and TSO data are quantified. Those parameters are normalised with the installed capacity of each country and shown in the Fig. 12. The standard deviations are represented by the bars and the extreme values are depicted with the black lines over the bars. It is observed that both the standard deviation and the extreme values improve in EMHIREs in a significant amount of countries analysed and when there is no improvement the result of EMHIREs is very similar to MERRA.

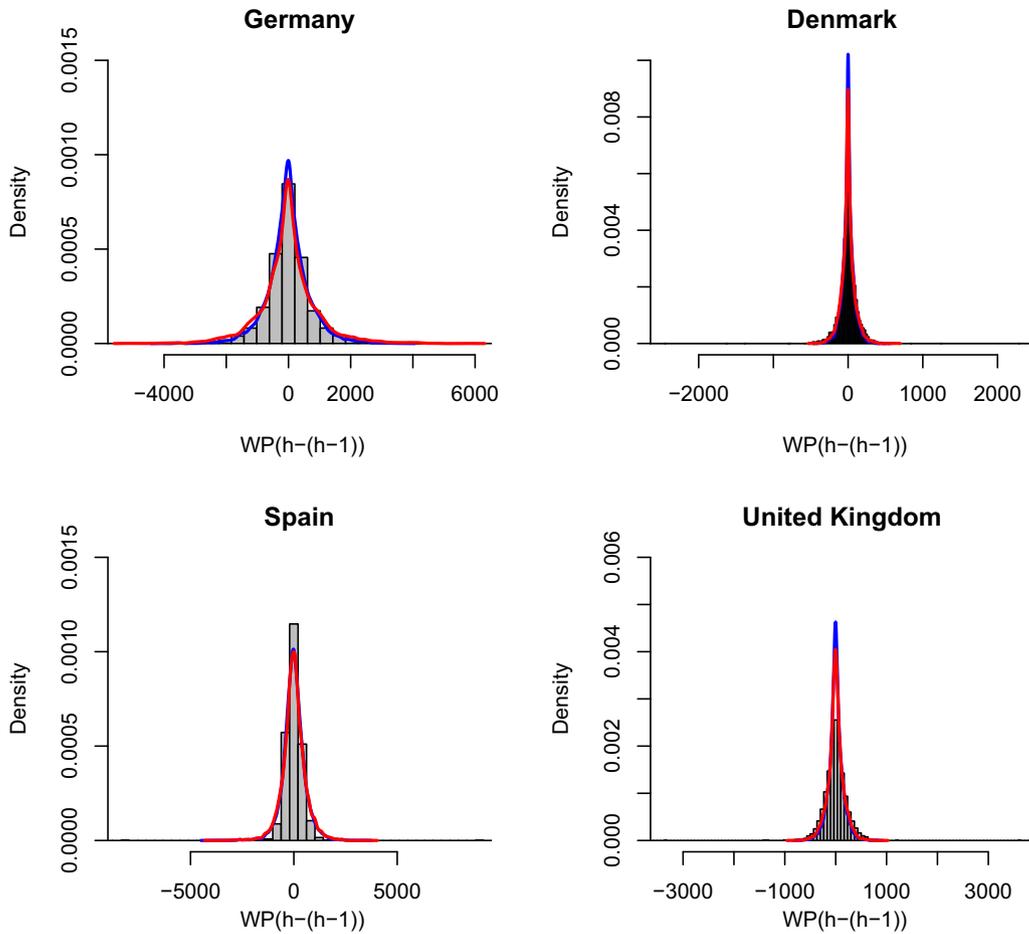


Fig. 11. Ramp rate distributions in 2015 for Denmark, Germany, Spain and United Kingdom. TSO data are represented with the grey histograms; EMHIREs (in red) and MERRA (in blue) are represented by the distribution density curves. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

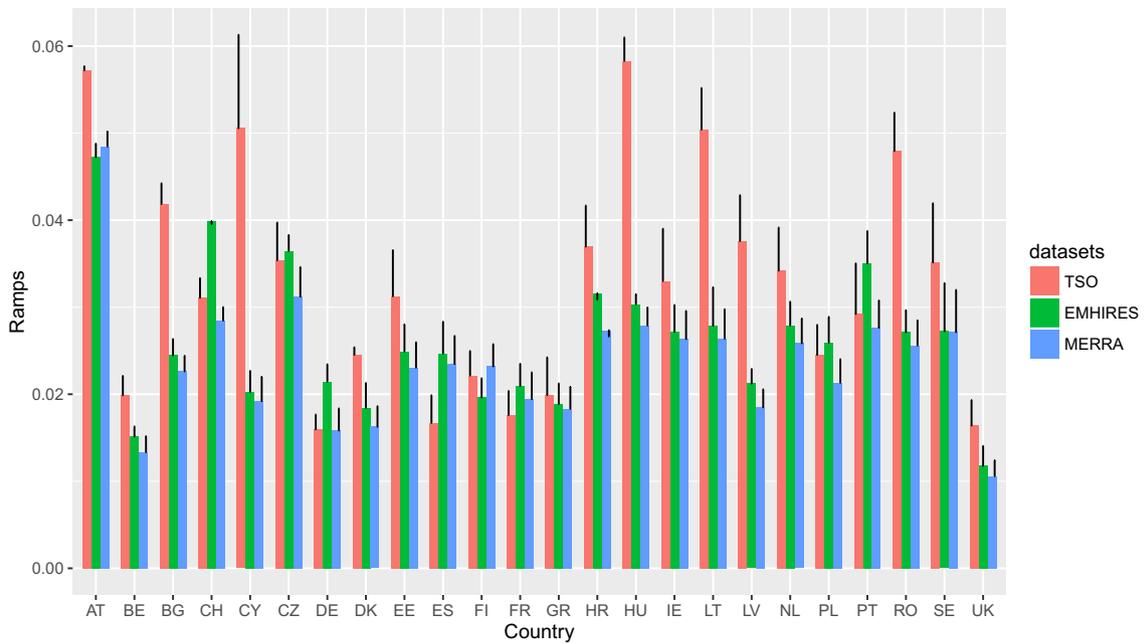


Fig. 12. Standard deviation and extremes of the normalised ramp rates for the TSO, EMHIREs and MERRA datasets for all the countries with TSO data available.

The limitation of the weather derived wind power time series appears when trying to capture the maximum of the ramping rates (1 and 100 percentiles of the ramping rates distributions). EMHIREs (and consequently MERRA) are not able to reproduce some extreme situations. Because of the limitation of the data sets it is difficult to assess if these extreme ramps have occurred because of purely meteorological events or on the contrary they were triggered by other non-meteorological effects: curtailment, outages such as maintenances, grid losses and sudden disruptions. If this were the cases, it is not surprising if the purely meteorological methodology developed here did not allow taking them into account. There could be an extra effect because the change in wind speed could happen at any time during the hour. The first version of EMHIREs is release at hourly frequency and this effect is not captured when we only have one value per hour.

4.2.4. Offshore power

The validation for the MERRA generated offshore time series also shows the skill of the statistics (Table 5). The correlations obtained when comparing with the TSO are in all cases high; 0.93, 0.83, 0.95, 0.90 and 0.91 for Belgium, Denmark, Germany, Netherlands and United-Kingdom, respectively. However, the overall statistics show that the bias is greater than in the onshore. This can be explained by the fact that offshore, turbines are mostly grouped together in wind farms: in the case of Belgium there are 3 wind farms aggregated; Denmark has 16 wind farms, Germany 13 wind farms, the Netherlands 5 wind farms, and the United Kingdom 28 wind farms. Due low turbulence levels over water wind speed reductions last over relatively long distances, which can lead to significant power losses of turbines inside a wind farm. Since the MERRA data does not include turbine induced flow reductions a bias with the observations can arise. In addition, wind farms are frequently built in coastal waters. The local wind conditions in these locations, characterised by coastal wind speed gradients, are not captured by the MERRA data.

4.2.5. Capacity factors and regional statistics

Additional comparison of the EMHIREs dataset with the TSO time series is done by calculating the capacity factors (CF); that is, the ratio between the sums of the energy produced (GWh) and the maximum theoretically possible generation (installed capacity (GW) * 8760 h (GWh)) per country. The Table 6 includes the CF coefficients for EMHIREs, MERRA, ECMWF and TSO data showing that all values are hovering between 12.8% and 69% with an average of 22.6%. While EMHIREs presented an improvement in the ramping rates of several countries, the capacity factors of the three datasets are very similar between them. The CF TSO are calculated with the total generation of the annual statistical factsheet provided by ENTSOE (Table 1, last column) and with the total generation derived from the hourly time series from ENTSOE (Table 1, penultimate column). The differences in some cases reflect the mismatches of the data provided by the TSO data. As mentioned in the data, although for keeping the consistency with yearly statistics the EMHIREs wind power time series have been normalised to the ENTSOE-E annual production statistics reported, the bias with the hourly time series are implicitly in the EMHIREs time series. That is the reason why in some cases, the CF EMHIREs differs significantly from the CF TSO annual, for example the case of Belgium.

Currently, at NUTS 2 aggregation levels, the only statistics available for benchmarking theoretical CF values are the annual total wind power production for Spain, France and Finland. Table 6 presents the comparison of the annual wind power generation by NUTS 2 for the three countries. It is observed that both MERRA and EMHIREs the total production is similar to the annual statistics

except for three NUTS2 in Spain (ES53, ES70, and ES13), three in France (FR10, FR23, FR25) and one in Finland (FI20). Although in the comparison between MERRA, EMHIREs and ECMWF the results were highly correlated between the three datasets, it would be necessary to validate the data at regional scale with actual hourly time series. Therefore, the validation by NUTS 2 region will continue once the data is released by the national TSO.

5. Conclusions

The methodology presented has been used to capture local geographical information to generate meteorologically derived wind power time series at high spatial resolution. A statistical downscaling technique has been applied to capture the effect of fine-scale forcing, in particular in areas characterised by fine spatial variability of features such as rugged topography and very diverse land surface conditions. This allows a better understanding of the wind resource at the precise location of wind farms. This study is a part of the methodology followed to develop the EMHIREs dataset, the first publically available European wind power generation dataset derived from meteorological sources that is available by country, bidding zone, NUTS-1 and NUTS-2 level.

Although the three datasets compared, EMHIREs, MERRA and ECMWF 100 m wind speeds, capture similar broad features of the wind patterns on the European continental scale, the statistical results show that in coarser resolutions (MERRA) The variability of the wind speed is underestimated with respect to the dataset where the downscaling is applied (EMHIREs) and in the ECMWF wind speeds. Although the somewhat limited availability of robust benchmarking data, the validation of EMHIREs against power system statistics and time series published by TSOs shows an improvement in most of the countries and bidding zones in the performance with respect to time series not applying an accurate spatial downscaling and the ECMWF derived products. It is shown that the increased power from EMHIREs may be the overall effect of wind turbines being sited in favourable locations with speed up due to orographic or roughness effects, which are captured by Global Wind Atlas predicted wind climate data, but not by MERRA. Also increased variability can be captured because these effects are locally a function of wind direction. This effect can be observed at country and by bidding zone aggregation levels.

EMHIREs is able to capture the variability of wind energy, in particular peaks and ramps, in a much more accurate way than previous meteorologically derived time series. The limitation of the weather derived wind power time series appears when trying to capture the maximum of the ramping rates. EMHIREs (and consequently other meteorological derived time series) are not able to reproduce some extreme situations since the methodology does not take into account effects of curtailment, outages such as maintenances and grid losses or network incidences. However, this is the only study, to the best of the authors' knowledge, trying to reproduce wind power time series at both national and regional levels covering the whole Europe with a homogeneous methodology avoiding the use of artificial or on-purpose tuned correction factors. Although it is possible to obtain higher correlation values on more limited and homogeneous areas and/or using purposely tailored additional parameters to be set *a posteriori* through data fitting, the purpose of this study was to develop an *ab initio* methodology for wind power production simulation and apply homogeneously to all Europe. The methodology has provided results ranging between good and excellent for all countries for which reliable TSOs data are available, regardless their sometimes huge geographical diversity. Using EMHIREs and therefore, the statistical downscaling technique for power system analysis will

increase the accuracy of generation adequacy assessments, renewable energy integration studies and market studies for power system flexibility options such as storage systems, electric vehicles and demand response.

Acknowledgements

We would like to thank the DG-JRC PETTEN Linux and High Performance Computing support for advising and helping in setting the entire IT infrastructure to develop EMHIREs; in particular to Steve Clements and Igor Simonovski. Acknowledge the Global Modelling and Assimilation Office (GMAO) and the GES DISC for the dissemination of MERRA. We also thank to the GWA data team to provide and support with corresponding guidelines and files to downscale the MERRA data.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.apenergy.2017.04.066>.

References

- [1] European Commission. Energy Union package – a framework strategy for a resilient energy union with a forward-looking climate change policy, European Commission COM (2015) 80; 2015.
- [2] Song-Lak K, Hoonill W. Intra-farm wind speed variability observed by nacelle anemometers in a large inland wind farm. *J Wind Eng Ind Aerodyn* 2015;147:164–75.
- [3] Sorensen P, Cutululis NA, Madsen H, Pinson P, Jensen L, Hjerrild J, et al. Modelling of power fluctuations from large offshore wind farms. *Wind Energy* 2008;11:29–43.
- [4] W. M. Organization. “WMO,” WMO. [Online]. Available: <http://www.wmo.int/pages/prog/wcp/ccl/faqs.php#q3>; 2016 [accessed 2016].
- [5] Basit A, Hansen DA, Altin M, Sorensen P, Gamst M. Wind power integration into the automatic generation control of power systems with large-scale wind power. *J Eng* 2014.
- [6] Sorensen P, Cutululis N. Wind farms’ spatial distribution effect on power system reserves requirements. In: *IEEE international symposium on industrial electronics*.
- [7] Lei Y, Mullane A, Lightbody G, Yacmini R. Modeling of the wind turbine with a doubly fed induction generator for grid integration studies. *IEEE Trans Energy Convers* 2006;21:257–64.
- [8] Karki R, Hu P, Billinton R. A simplified wind power generation model for reliability evaluation. *IEEE Trans Energy Convers* 2006;533–40.
- [9] Slootweg J, de Haan S, Polinder H. General model for representing variable speed wind turbines in power system dynamics simulations. *IEEE Trans Power Syst* 2003;18:144–51.
- [10] Staffell I, Pfenninger S. Using bias corrected reanalysis to simulate current and future wind power output. *Energy* 2016;114:12241239.
- [11] Miller L, Brunzell N, Mechem D, Gans F, Monaghan A, Vautard R, et al. Two methods for estimating limits to large-scale wind power generation. *Appl Phys Sci* 2015;112:11169–74.
- [12] Akdag S, Dinler A. A new method to estimate Weibull parameters for wind energy applications. *Energy Convers Manage* 2009;50:1761–6.
- [13] Wagner R, Antoniou I, Medersen S, Courtney M, Jorgensen H. The influence of the wind speed profile on wind turbine performance measurements. *Wind Energy* 2009;12:348–62.
- [14] Archer C, Jacobson M. Spatial and temporal distributions of U.S. winds and wind power at 80 m derived from measurements. *J Geophys Res* 2003;109.
- [15] González-Aparicio I, Zucker A. Meteorological data for RES-E integration studies – state of the review. *EUR* 27587; 10,2790/349276 European Commission; 2015.
- [16] Jie Z, Draxl C, Hopson T, Delle Monache L, Vanvyve E, Hodge B-M. Comparison of numerical weather prediction based deterministic and probabilistic wind resource assessment methods. *Appl Energy* 2015;156:528–41.
- [17] Rose S, Apt J. What a reanalysis data tell us about wind power? *Renewable Energy* 2015;83:963–9.
- [18] Liléo S, Berge E, Undheim O, Klinkert R. Long-term correction of wind measurements – state-of-the-art, guidelines and future work. *Elforsk Report* 2013;13:18.
- [19] R.L. File specification for MERRA products. GMAO office note n1. (version 2.3); 2012.
- [20] Olauson J, Bergkvist M. Modelling the Swedish wind power production using MERRA reanalysis data. *Renewable Energy* 2015;76:717–25.
- [21] Cannon D, Brayshaw D, Methven J, Coker P, Lenaghan D. Using reanalysis data to quantify extreme wind power generation statistics: a 33 year case study in Great Britain. *Renewable Energy* 2015;75:767–78.
- [22] Ritter M, Deckert L. Site assessment, turbine selection, and local feed-in tariffs through the wind energy index. *Appl Energy* 2017;185:1087–99.
- [23] Carvalho D, Rocha A, Gomez-Gesteira M, Silva Santos C. Offshore wind energy resource simulation forced by different reanalyses: comparison with observed data in the Iberian Peninsula. *Appl Energy* 2014;134:57–64.
- [24] Kubik M, Coker P, Barlow J. Increasing thermal plant flexibility in a high renewables power system. *Appl J* 2015;154:102–11.
- [25] Huber M, Dimkova D, Hamacher T. Integration of wind and solar power in Europe: assessment of flexibility requirements. *Energy* 2014;69:236–46.
- [26] Olauson J, Bergkvist M. Correlation between wind power generation in the European countries. *Energy* 2016;114:663–70.
- [27] Brancucci Martinez-Anido C, Abbate AL, Migliavacca G, Calisti R, Soranno M, Fulli G, et al. Effects of North African electricity import on the European and the Italian power systems: a techno-economic analysis. *Electric Power Syst Res* 2013;96:119–32.
- [28] McKenna R, Hollnaicher S, Ostman P, Fichtner W. Cost-potentials for large onshore wind turbines in Europe. *Energy* 2015;83:217–29.
- [29] Andresen G, Sondergaard A, Greiner M. Validation of Danish wind time series from a new global renewable energy atlas for energy system analysis. *Energy* 2015;93:1074–88.
- [30] EUROSTAT. Nomenclature of territorial units for statistics. [Online]. Available: <http://ec.europa.eu/eurostat/web/nuts/overview>.
- [31] I. R. E. Agency. Global atlas for renewable energy. International Renewable Energy Agency. [Online]. Available: <http://irena.masdar.ac.ae/>; 2015.
- [32] D. T. U. D. o. W. Energy. Global Wind Atlas. DTU Wind Energy. [Online]. Available: <http://globalwindatlas.com>; 2015.
- [33] Pfenninger S, Staffell I. Renewables.ninja. ETH Zurich and Imperial College London. [Online]. Available: <https://www.renewables.ninja/>; 2015.
- [34] G. d. m. d. Santander. Grupo de Meteorología de Santander. [Online]. Available: http://www.meteo.unican.es/es/research/statistical_downscaling; 2016.
- [35] ECMWF. European Centre for Medium-Range Weather Forecasts (ECMWF). ECMWF. [Online]. Available: <http://www.ecmwf.int/en/about>; 2010.
- [36] González-Aparicio I, Zucker A, Careri F, Monforti F, Huld T, Badger J. EMHIREs dataset. Part I: Wind power generation. European Meteorological derived High resolution RES generation time series for present and future scenarios. *EUR* 28171 EN; 2016.
- [37] Monforti F, González-Aparicio I. Building a robust and complete database of wind farms in Europe for continental wind power simulations. Data consistency and sensitivity analysis. *Appl Energy* 2017 [submitted for publication].
- [38] The Wind Power. The Wind Power. [Online]. Available: <http://www.thewindpower.net/>; 2015.
- [39] Troen I, Petersen E. European wind atlas. Brussels, Belgium: Risø National Laboratory; 1989.
- [40] Gisinger S, Mayr J, Messner J, Stauffer R. Spatial and temporal variation of wind power at hub height over Europe. *Nonlinear Process Geophys* 2013;20:305–10.
- [41] E. N. o. T. S. O. f. electricity, ENTSOE Reliable Sustainable Connected. [Online]. Available: <https://www.entsoe.eu/data/entso-e-transparency-platform/Pages/default.aspx>; 2015.
- [42] E. C. R. 543/2013. Submission and publication of data in electricity markets, European Commission; 2013.
- [43] R. E. d. Espana. Sistema de Información del operador del sistema. [Online]. Available: <https://www.esios.ree.es/es/>; 2016.
- [44] R. d. t. d. RTE. Panorama de L’électricité renouvelable en 2015. RTE, Réseau de transport d’électricité, France; 2015.
- [45] Wood A, Maurer E. Long-range experimental hydrologic forecasting for the eastern United States. *J Geophys Res* 2002;107:6–15.
- [46] Kirchmeier M, Lorenz D, Vimont D. Statistical downscaling of daily wind speed variations. *J Appl Meteorol Climatol* 2014;53:660–74.
- [47] Naudin-Aparicio. Aggregated power curve for multiple wind turbines in power system area. Roskilde: Technical University of Denmark; 2013.
- [48] I. N. L. P. Gipe. Wind Works [Online]. Available: [http://www.wind-works.org/cms/index.php?id=625&tx_ttnews\[tt_news\]=3935&cHash=39424df20f9c961ee7cd158a85cabe36](http://www.wind-works.org/cms/index.php?id=625&tx_ttnews[tt_news]=3935&cHash=39424df20f9c961ee7cd158a85cabe36); 2007.
- [49] Pielke. Mesoscale meteorological modeling. 2nd ed. San Diego: Academic Press; 2002.