Applied Energy 225 (2018) 209–220

Contents lists available at ScienceDirect

Applied Energy

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

Comparison of long-term wind and photovoltaic power capacity factor datasets with open-license

L. Moraes Jr.[a,*], C. Bussar[a,b], P. Stoecker[a,b], Kevin Jacqué[a,b], Mokhi Chang[a], D.U. Sauer[a,b]

[a] ISEA – Institute for Power Electronics and Eletrical Drives, RWTH-Aachen University, 52066 Aachen, Germany
[b] JARA – Jülich Aachen Research Alliance, RWTH Aachen University and Research Centre Jülich, Aachen and Jülich, Germany

HIGHLIGHTS

- Open-license datasets based on the same meteorological data diverge from each other.
- Six tests were implemented to quantify the differences.
- Deviations were mainly found at duration curves and full load hours analysis.
- Divergences found may considerably impact energy system simulation results.
- System operator's wind and PV feed-in data are not trustworthy, but the only source to compare against at national level.

ARTICLE INFO

Keywords:
Power capacity factors
Wind and photovoltaics/PV power time series
Comparing datasets based on reanalysis data
Energy systems modelling
Renewable energies integration

ABSTRACT

Investigation of pathways toward decarbonisation of energy supply systems strongly relies on integration of electricity generation from wind and photovoltaics (PV). Energy system model authors are typically not experts in creation of representative weather datasets, which are fundamental for an unbiased representation of volatile power generation within the models. The aim of this work is therefore to benchmark data quality and verify against feed-in records for datasets published from two projects: EMHIRES and Renewables.ninja; feed-in records taken from Transmission System Operators (TSO). Both projects used meteorological reanalysis data from NASA (National Aeronautics and Space Administration) and Meteosat-based datasets from CM-SAF (Satellite Application Facility on Climate Monitoring) to generate long-term hourly PV and wind power capacity factor time series. Although datasets were based on the same raw data sources, they present significant differences due to modelling of energy conversion technologies, correction and validation methods. Comparison of duration curves, full load hours, plots of hourly PV capacity factors as well as correlation analysis between datasets reveal that for PV generation EMHIRES is more similar to TSO's data, while the Ninja dataset revealed more similarity when comparing wind datasets. Results showed that even based on the same data sources, time series were strongly dependent on methods applied subsequently. Application of the datasets within energy system models therefore could present a form of hidden exogenous bias to results. System modelers, who need weather based open license data to perform energy simulations, may be aware of differences in open license datasets available.

1. Introduction

Electricity production has different levels of dependency on meteorological conditions. In the past, meteorology has already played an important role on energy, especially by forecasting: hydro power plants energy production; sea conditions at offshore operations of oil and gas; temperature of cooling water at thermal power plants; as well as, demand variations due to weather changes. In the beginning of the 1990s the term "Energy Meteorology" appeared as a new discipline. The relation between energy use and production were part of the Long-term Plan of the World Meteorological Organisation (WMO) published in 1994, which included evaluation of weather and climate implications in energy matters. Although weather influences non-critically conventional power plants, for volatile renewables it plays a major role. More recently the trend towards global massive investments in volatile renewable energies has changed the focus, requiring better understanding of fluctuating wind and PV generation [1,2].

The complex behaviour of wind and PV power production and their interaction with traditional power system components can be better understood through computer simulations. Energy system models

* Corresponding author.
E-mail addresses: lmo@isea.rwth-aachen.de, batteries@isea.rwth-aachen.de (L. Moraes).

https://doi.org/10.1016/j.apenergy.2018.04.109
Received 12 January 2018; Received in revised form 29 March 2018; Accepted 30 April 2018

represent system components and their interactions. They are used to simulate behaviour and can also optimize the operation strategies and/or investment plans. Modelling energy systems with higher shares of renewables requires further dependency on historic meteorological data which can add even more uncertainties to the problem [3,4].

Recent projects used meteorological based data to produce wind and PV power time series as in [5]. EMHIRES [6,7] and Renewables.ninja [8,9] projects published their datasets under open license. The methods applied convert wind speeds and solar radiation, derived from meteorological reanalysis and satellite datasets, generating power output by simulating wind and PV fleets within a geospatial region. This complex combination translates into a fundamental source of data for energy system simulations, because it represents the regional temporal dynamic potential for a technology. The published time series are normalized to the installed capacity in each aggregation level, giving values as hourly power capacity factors (PCFs) (from 0 to1).

Although data used from EMHIRES and Ninja originates from the same sources, the MERRA (Modern-Era Retrospective analysis for Research and Applications) dataset [10] from NASA (National Aeronautics and Space Administration), there are significant differences in data selection, pre-processing and modification, as well as in power conversion models. In Table 1, the main steps used for each project to build the wind datasets are listed. It can be seen that the downscaling methods are different; EMHIRES uses Weibull distribution based on probabilities of two data sources and Ninja interpolates using locally weighted smoothing. Wind power curves, as well as wind farm locations used in both datasets for power conversion originate from the same source: The Wind Power database [11]. EMHIRES merges these data with an internal database, which is not detailed in their publications. Ninja applies Gaussian filter to smooth and represent a wind farm composed of dispersed wind turbines.

Table 2 describes the main steps to build PV datasets. EMHIRES uses SARAH (Surface Solar Radiation Data Set – Heliosat) [13] from CM-SAF (Satellite Application Facility on Climate Monitoring) in its original resolution, applying calculation of irradiance on inclined plane based on [14]. Ninja uses both, MERRA and SARAH data, on MERRA resolution ignoring the higher resolution of SARAH, applying linear interpolation to get local values for PV farms. Both apply reconstruction for SARAH and mentioned the considerable gaps in the dataset. Ninja calculates irradiance fraction for MERRA based on [15] and calculates irradiance on inclined plane based on their own methods [8].

To calculate power conversion, EMHIRES is based on PVGIS [16]. They claim there is no complete database of PV farms and make assumptions to allocate farms within a region based on land-use assumptions. After this, they simulate with PVGIS “hourly PV potential for 2015 (given by the maximum solar energy output (watt hour) for each kilowatt of installed capacity averaged over a region)” [7]. They consider PV arrays mounted on open-rack mounting at 30° inclination south-facing.

Ninja used for power conversion the model from [17], and applies randomized panel azimuth and tilt angle orientations based on normal distribution. They used locations based on PVLog [18], PVOutput [19] and DTI [20].

Ninja and EMHIRES reported in their publications the need to perform data base completeness and gap filling, as well as making many assumptions. It is within reason that this may contribute significantly to increasing uncertainties when producing power time series. Here we listed some considerable data gaps and assumptions made: According to Ninja “The tower height was not known for 62% of farms, and so was estimated using a regression of known heights against the logarithm of turbine capacity and the date of installation. The start date was not known for 16% of farms, and so was inferred from other farms in the same country with turbines of the same capacity” [9]. They also performed their simulations for wind datasets for wind farms with capacities higher than 1 MW (82% of Europe's total) and using the 100 most popular power curves (81% of installed capacity) [9]. According to EMHIRES, wind turbine type was missing in 28% of the database [6].

There are also significant differences in time series error correction and validation methods. Ninja measures bias based on the derived power output from wind farms and applies corrections to wind speeds at national level, assuming all farms within a country experience the same bias and the power curves are correct [9]. For PV, Ninja chose to apply a continental factor, Europe-wide, to all countries. They also presented a method based on linear regression, for countries where data from TSOs (Transmission System Operators) were available, but they concluded this did not lead to overall improvements [8].

EHMIRES claims the statistical spatial downscaling applied to wind datasets improves performance capturing local effects and compensating limited spatial resolution. No further calibration was found for EMHIRES wind in the publications available [6]. EHMIRES PV datasets are calibrated with the differences of duration curves between TSOs data (corrected by annual values) and simulated data (uncalibrated) [7].

The datasets from EMHIRES and Ninja investigated in this work were validated by their authors with great care, in multiple methods as presented in the Table 3. For validation of the methods, Ninja uses in their simulations the installed capacity of every year, setting to zero the capacity of wind parks, at times when it did not exist.

As usual by open source terms of use, no guarantee of quality and accuracy is made and the user should perform a data check before proceeding. In the scope of this paper we analyse only the datasets resulting from their methods. After carrying out some simple tests we observed data from EMHIRES and Ninja showed considerable deviations from one another. The question for a potential use would be which implication and bias a dataset bears for application in an energy system model, and therefore how to interpret results.

To be able to check and compare datasets, we developed a testing scheme, to perform a standardised analysis on each set in the period from 2012 until 2014, comparing the datasets with each other and using TSOs data as reference. The dataset authors used different fleets to simulate national (or regional) power outputs and they do not contain effects of curtailment, maintenance, transmission losses, but the

Table 1
Wind power datasets main developing steps - differences in data acquisition, processing and calculations.

Step	EMHIRES [6]	Renewables.ninja [9]
Raw data selection	MERRA [10] wind speed values - grid 60 × 70 km	MERRA and MERRA-2 [10] wind speed values - grid 60 × 70 km
Wind speeds downscaling to wind farm level	Statistical spatial downscaling of hourly wind speed variations using Weibull distribution - to the specific geographic coordinates of each wind farm. Probability data extracted from Hires Dataset and Global Wind Atlas [12]	Interpolates speeds to the specific geographic coordinates of each wind farm using LOESS regression (Locally Weighted Scatterplot Smoothing)
Calculation of hub height wind speed	Vertically interpolated to the hub height using a power law profile - MERRA-derived wind speed time series at 10 and 50 m height	Extrapolates speeds to the hub height of the turbines at each site using the logarithm profile law – 2, 10 and 50 m height
Power conversion	Power curves built using as primary data the turbine database from [11] merged with an internal database	Power curves built using as primary data the turbine database from [11], which are smoothed to represent a farm of several geographically dispersed turbines, using Gaussian filter

Table 2
PV power datasets main developing steps - differences in data acquisition, processing and calculations.

Step	EMHIRES [7]	Renewables.ninja [8]
Raw data, calculations and treatment	SARAH 5 km resolution with missing values reconstructed - the global irradiance on an inclined plane calculated following [14]	MERRA 50 x 50 km - the diffuse irradiance fraction is estimated with the BRL model [15]. SARAH was taken on MERRA grid resolution with missing values reconstructed. The global irradiance on an inclined plane calculated following their methods. Temperature at 2 m taken from MERRA
Downscaling of solar radiation to farm level	Uses the 5 km^2 pixels from SARAH	Values are linearly interpolated from grid cells (MERRA resolution not considering the higher resolution of SARAH) to the given coordinates
Power conversion model	PVGIS model is used [16]. Takes into account effects of shallow-angle reflectivity, temperature and low irradiance as well as cooling of the modules by wind. Considers PV arrays mounted on open-rack mounting at 30° inclination south-facing	Power output from a given panel is calculated from using the relative PV performance model described by [17] which gives temperature-dependent panel efficiency curves

differences on results were expected to be slight.

Working with data from different sources can be quite laborious. Every source uses different file formats and even data from the same source may have differences from one file or release to another. EMHIRES used tab separated files with different headers in PV and wind files, but the same country codes. Ninja release version 1.0 uses country codes for PV and country names for wind, while for Ninja 1.1 differences are minimal, composed only of quotation marks. Data from TSOs are seldom similar with very individual file extensions and formats.

We are thankful the authors of the mentioned projects took an initiative towards open license wind and PV power data and try with this paper to collaborate on the development of this field of knowledge. With this study we aimed to analyse the datasets available in order to ascertain their validity and use them as input in our own simulations. Time series were checked for the period from year 2012 until 2014 in 5 countries. A series of tests was performed comparing datasets through correlation coefficients, plots of duration curves, PV hourly plots, number of full load hours per year, weekly capacity factors averages and seasonal ratios. At national level they were compared against data from TSOs and at NUTS-levels datasets from Ninja and EMHIRES were compared only against each other.

The next section introduces the datasets and the testing scheme developed to analyse them. Section 3 presents the results we obtained from these analyses. Section 4 shows the discussion of results and Section 5 concludes and suggests further applications.

An extension of the methods to all years and countries analysed in this paper is presented in the supplementary material.

2. Methods

2.1. Datasets analysis

Data from the EHMIRES and Ninja are available as wind and PV capacity factors time series on an hourly basis. These factors are ranged between 0 and 1, representing the instant power rate of the total installed capacity in the related geospatial region. Multiplying the installed capacity by these factors, should result in the feed-in power for the corresponding time points.

The datasets have different spatial coverages, but include mainly the European Union, Norway and Switzerland. Both projects used power plant fleets to simulate and generate their capacity factors time series. Spatial distribution and installed technologies are important factors, which they used in their models to convert wind and solar radiation into electrical power. Ninja released 3 types of wind time series simulated with different installed capacities: current, near term future and long term future. For future scenarios they included power plants that were approved, being built or planned after 2014, in the initial release and after 2016 in the release 1.1. They developed also two PV datasets based on different data sources. PV data in release 1.1 has received an extension in the time range but no changes in the method or basis data were done, that is why we show only the last PV release in Table 4. For wind, Ninja used in the second release the MERRA-2 as base and also extended the period of coverage. EMHIRES has different aggregation levels based on Eurostats NUTS[1] classification, implemented local correction in weather variables at power plant level and used the power plant fleet of 2015. After personal communication with the authors of Ninja, NUTS-level datasets were kindly made available for this analysis. The methods used to produce these datasets with higher resolution are unfortunately not available in any publication, at the time this work was performed.

TSOs publish online solar and wind energy feed-in projections based on the extrapolation of measurements made at representative reference power plants and sometimes based on weather models also. To benchmark datasets of EMHIRES and Ninja at national level we prepared a reference dataset based on TSOs wind and PV feed-in historical data. Capacity installation data was needed to produce capacity factor time series. Aiming to reach a good representation of real production, several conditions were defined and only data meeting the following pre-requisites were selected:

(a) Spatial resolution - data should match countries or all bidding zones that compose a country;
(b) Data for the same time period - data for wind and PV power should match the same coverage periods and being between 1986-01-01 00:00 (begin of EHMIRES) and 2014-12-31 23:00 (end of Ninja 1.0);
(c) Time resolution - at least 1 h;
(d) Only data for whole years were considered;
(e) Resulting time series should pass a pre-check to avoid discrepant capacity factors (negative or bigger than 1).

Wind and PV power feed-in data meeting the above mentioned requirements were only found for the time period between 2012 and 2014, and for the following countries: Czech Republic (CZ) [22], Germany (DE)[2] [23], France (FR) [24], Italy (IT) [25] and United Kingdom (UK) [26]. Data for Spain and Sweden was also found [27,28]. In Spain solar power included concentrating solar power (CSP) and was therefore not used. In Sweden the values found for PV power were deemed not to be in accordance with installed capacity data from other sources and were also not used. Resulting capacity factors were extremely low for 2012, with maximum value not exceeding 0.1.

For the same and even longer time periods, wind data was also found for Austria, Belgium, Denmark and Ireland but no satisfactory PV data records were available, perhaps because total installed capacity was rather small in those countries. Data after 2015 is more readily

[1] NUTS: Nomenclature of Territorial Units for Statistics (French: Nomenclature des unités territoriales statistiques). Level 1 is composed by major socio-economic regions and level 2 by basic regions for the application of regional policies [21].

[2] Data for Germany was obtained from Open Power System Data, which were based on TSOs.

Table 3
Validation methods, reference data and periods.

	EMHIRES [6,7]		Ninja [8,9]	
	Method	Period	Method	Period
Wind	Comparison of duration curves and boxplots with MERRRA and TSOs Comparison of annual full load hours with TSOs Comparison against TSOs' hourly time series	2015	Comparison of annual full load hours in 23 countries (different sources) Comparison of monthly full load hours in 13 countries (TSOs) Comparison of hourly capacity factors for different periods (TSOs)	2005–2014 2005–2015 2000–2014, depending on the availability in each country
PV	Comparison of annual full load hours for TSOs (14 countries), PVGIS (Photovoltaic Geographical Information System) and EMHIRES Comparison against TSOs in 14 countries, presenting analysis in the publications for duration curves (12 countries), statistical indicators (14 countries) and power generation hourly time series (3 countries)	2015	Comparison of annual full load hours for 10 countries (various sources) Comparison of weekly mean of full load hours at national level in 5 countries (TSOs); Comparison of weekly mean of full load hours for sites (8–438 per country, depending on availability) in 10 countries (Measured and TSOs)	2014

available [29,30].

In addition to power feed-in curves, electricity demand time series were found for the same countries and for the correspondent years analysed here. The total annual demand was summed from the time series and compared against other sources. The total annual power feed-in was also summed up and compared against additional sources. The results are presented in the beginning of Section 3.

In order to generate capacity factors time series, the hourly power time series was divided by the installed capacity for each technology and country. Detailed capacity installation records were only found for DE and UK, in form of time series. For the other countries capacity installation time series were generated by linearly interpolating installed capacity between the last days of two subsequent years. Installed capacity values were found deviating between sources, as shown in the supplementary material, and the average values of four sources for wind [30–33] and for PV [30,32,34,35] were calculated. The 3 years dataset generated was labelled TSOs and is composed of hourly capacity factors for wind and PV for the five countries (CZ, DE, FR, IT and UK) in the period from 2012-01-01 00:00 and 2014-12-31 23:00.

To analyse the datasets we selected the time series and named them as shown in Table 5. PV time series for Ninja were taken from release 1.1. A comparison in every country and NUTS-level was performed, in which we apply the tests further described in ****. The NUTS-1 zones chosen were: Isole – Italy (ITG), Niedersachsen – Germany (DE9), Centre-Est – France (FR7), East-Midlands – United Kingdom (UKF). For NUTS-2, we selected: Sicilia – Italy (ITG1), Hannover – Germany (DE92), Rhône-Alpes – France (FR71), Derbyshire and Nottinghamshire – United Kingdom (UKF1). Data was processed with a tool developed for this analysis, written in Python.

Before presenting the methods it is reasonable to clarify some aspects. It is expected that future energy systems will have even higher dependency on meteorology. This impacts the investment plan in energy production, flexibility capacity, as well as on decision of operation strategies. Into the methods we use the term "energy balance", which we define here, for an optimal operation of energy system with high shares of volatile renewables as: balance to ensure that energy production matches demand for every instant, also avoiding black outs. To ensure this, energy reserve and flexibility options should be available. Reserve is commonly dimensioned based on demand profile, production capacity and their dependency on meteorology. Flexibility options are dispatchable power generators, grid, demand side management and storage (see Sections 2.1.1–2.1.6).

2.1.1. Correlation analysis

There are many possible interpretations of the statistical relationship between two time series exposed by Pearson's correlation coefficient [36]. We are interested here only in the measure of similarity this coefficient represents as a standardized covariance. Correlation coefficient is placed between –1 (perfect inverse related) and +1 (perfect linear relationship), with zero meaning no correlation [36,37]. We calculated the correlation coefficient between the pairs of time series in each country and presented the result in the correlation matrices, for country level time series, and in a table of coefficients for NUTS-level datasets. For PV correlation analysis, the time points, where all capacity factors were zero, were not considered in calculation of correlation coefficients. Low correlation factors, close to zero, or inverse relationship found in negative coefficients, may indicate intra-day differences on energy production and power availability.

2.1.2. Annual duration curves

In order to be able to have an overview of the capacity factors' distribution for every year, we plotted duration curves, in which the capacity factors are shown in descending order of magnitude, for the total number of hours observed. It is possible to verify the amount of hours in which PV or wind feed-in is above a certain level. As described in [38], the duration curve for national aggregated values represents

Table 4
Datasets overview.

Source	Renewables Ninja [8,9]				EMHIRES [6,7]	
Dataset	Wind 1.0	Wind 1.1	PV-Merra 1.1	PV-Sarah 1.1	Wind	PV
Time Range	1985–2014	1980–2016	1985–2016	1985–2015	1986–2015	
Based on	MERRA-1	MERRA-2	MERRA-2	SARAH	MERRA-1	SARAH
Aggregation level	Country (NUTS -1, NUTS-2)[a]				Country, BZ[b], NUTS-1, NUTS-2	
Weather variable	Bias correction at continental and country level				Local correction at power plant level	
Onshore/Offshore	Together	Separate and together	–	–	Separate	–
Power plant fleet	2014, near and long term future	2016, near and long term future	2014	2014	2015	

[a] Datasets were kindly made available from Renewables.ninja authors for this analysis, before being published.
[b] BZ: Bidding Zone.

Table 5
Time series considered in this analysis, their labels and contents.

Source	PV		Wind	
	Label	Content	Label	Content
Renewables.ninja [8,9]	Ninja MERRA	National level from release 1.1	Ninja 1.0	National level from release 1.0 Onshore + Offshore aggregated
	Ninja SARAH	National level from release 1.1	Ninja 1.1	National level 1 On + Offshore aggregated – 2016 fleet
	–	–	Ninja 1.1 On	National level onshore only, from release 1.1
	–	–	Ninja 1.1 near	National level from release 1.1 On + Offshore aggregated – near term fleet
	–	–	Ninja 1.1 long	National level from release 1.1 On + Offshore aggregated – long term fleet
	Ninja	NUTS-1 and NUTS-2 level	Ninja	NUTS-1 and NUTS-2 level
EMHIRES [6,7]	EMHIRES	National level, NUTS-1 and NUTS-2 level	EMHIRES	National level onshore only, NUTS-1 and NUTS-2 level
TSOs [22–26,28,30–32,34,35]	TSOs		TSOs	–

Table 6
Annual energy produced from PV in DE in TWh.

	OPSD [23]	BMWi[a] Energy Data [41]	Fraunhofer ISE[b] [42]
2012	27.9	26.4	27.9
2013	29.6	31.0	29.7
2014	32.6	36.1	32.4

[a] Federal Ministry for Economic Affairs and Energy (Bundesministerium für Wirtschaft und Energie).
[b] Institute for Solar Energy Systems (Institut für Solare Energiesysteme).

Table 7
Annual energy produced from wind in DE in TWh.

	OPSD [23]	BMWi Energiedaten [41]	Fraunhofer ISE [42]
2012	45.9	50.7	45.9
2013	47.2	51.7	47.2
2014	51.1	57.4	42.6

Table 8
Correlation coefficients between Ninja and EMHIRES time series, for 2012 until 2014 in NUTS-level zones.

	NUTS-1				NUTS-2			
	ITG	DE9	FR7	UKF	ITG1	DE92	FR71	UKF1
Wind	0.61	0.59	0.61	0.58	0.96	0.95	0.91	0.84
PV	0.98	0.93	0.95	0.90	0.97	0.91	0.80	0.89

the "country wind farm" or "country PV-solar park" containing the compensation effects mainly due to spatial distribution of generation units and different energy conversion technologies/power curves. T-hese effects are also present at NUTS-1 and NUTS-2 aggregation. Once data from TSOs are close to real power feed-in, we have characteristic curves for every country to serve as reference when comparing those derived from other sources. Together with the duration curves, a verification of the most rich and poor hours was made, where the share in %, that these hours represent from the total energy yield of the year is shown.

Shapes of duration curves are decisive for planning annual energy balance, dimensioning base-load and checking requirements on flexibility capacity.

2.1.3. Annual full load hours

Full load hours represent the number of equivalent hours at rated power, which a distribution represents for a certain period of time. In this special case, the number is equal to the integral of the power duration curve, due to representation of hourly capacity factors which represents the energy yield for the observed period. Full load hours are used as an indicator for technology economic integration. It can be seen as a measure of performance, which in this case, combines natural resource potential in a spatial region with performance of generation units. 8760/8784(leap years) FLHs represent 100% of production related to the installed capacity in one year. The number of full load hours is important when dimensioning the system based on a yearly energy production perspective.

2.1.4. Weekly averages

This test analyzes week averages of capacity factor (168hs) for each dataset, comparing them in bar plots of weekly values in summer or winter. This test is used to check differences in weekly energy production and power availability. It is important when planning the medium and long term energy balance, and also by choosing flexibility/storage technology, operation strategy and dimensioning.

2.1.5. Seasonal ratios

In this analysis we divided the year in two half. One represents the warm season; the other corresponds to the cold season. The definition was performed through a statistical analysis made on historical data, based on the amount of hours of daily light combined with monthly temperature averages. For all countries analysed, about 30 years records of monthly temperature from [39] revealed that colder moths are from November until April and warmer from Mai until October. The daily sunlight hours taken from [40] revealed that highest amount of sunlight is found from April until September and lowest from October until March. We defined then, the warm season as the one from 16th of April until the 15th of October of each year. The cold season was defined from 16th of October until the 15th of April. The differences found in this analysis are important when planning the long term energy balance, choosing flexibility/storage technology, operation strategy and dimensioning.

2.1.6. PV hourly capacity factors plots

PV daily and seasonal cycles lead to higher correlation coefficients, compared to the same analysis made with wind time series. To increase capability of differentiation between datasets we performed hourly plot analysis, in which a time series is plotted showing the intensity of hourly power capacity factors in each day against the progress of the days investigated. It shows in which hours of the day there is energy production and with which intensity it is produced. It is important to see the daily and seasonal profile in one figure. Deviations between datasets may indicate impact on planning of short term energy balance, also by choosing flexibility/storage technology, operation strategy and dimensioning.

2.1.7. Future wind fleets

Additionally, using analysis taken out on the data based on current power fleets, we also performed the same tests on datasets generated based on future wind power fleets, at national level. Ninja has published in their two releases near term and long-term time series, which were generated with historic data but simulated with the inclusion of power plants currently under construction, authorized for construction, or slated for future construction. Our analysis was made only for the latest release. We performed the comparison against current power fleets (Ninja 1.1) from 2016. To have a reference point we used again the TSOs datasets. However, CZ was not included in the dataset, so we compare the other four countries.

3. Results

Reviewing the full description of the correspondent method in Section 2, may be necessary for a better understanding of the plots and tables found here in Section 3, where the results are presented.

The benchmarking of datasets requires trustable sources, to build a reference dataset. Our references in this analysis at national level were obtained from TSOs and ascertaining the precision of the data is a challenging task. Data from ENTSO-E are normally based on TSOs data, therefore data comparison was not deemed necessary. To perform a comparison, we relied on data from studies and government entities, which were open and with easy access in DE. We do not extend this comparison to other countries.

Tables 6 and 7 present the comparison of annual values in DE for PV feed-in and Wind feed-in. The values contained in the column OPSD (Open Power System Data) of all tables are those which compose TSOs datasets for DE.

3.1. Wind correlation analysis

At national level, this analysis revealed that EMHIRES and Ninja 1.0, which are both based on MERRA-1, are strongly correlated with each other for all countries, with Pearson's coefficient higher than 0.97. Conversely, they are not strongly correlated with Ninja 1.1 and Ninja 1.1 On (which are based on MERRA-2) and TSOs. The latter and Ninja 1.1 correlate strongly in all cases, with coefficients higher than 0.9, with an exception in CZ. Table 8 revealed that wind time series from Ninja and EMHIRES are well correlated for NUTS-2, but have coefficients around 0.6 for all NUTS-1 zones.

3.2. PV correlation analysis

Country level time series from EMHIRES and Ninja Sarah correlate stronger with each other (0.98), than with Ninja Merra (0.93 and 0.94). Correlation coefficients between EMHIRES and Ninja Sarah are not lower than 0.97 for all countries. In Fig. 1(b) the correlation matrix for DE for 2012–2014 is presented. In general, correlation coefficients are much higher than those for wind comparison. NUTS-1 and NUTS-2 PV datasets are strong correlated. The lowest coefficient is found in FR71, as shown in Table 8.

Correlation matrices for all countries made for two periods, 2012–2014 and 1986–2015 are found in the supplementary material.

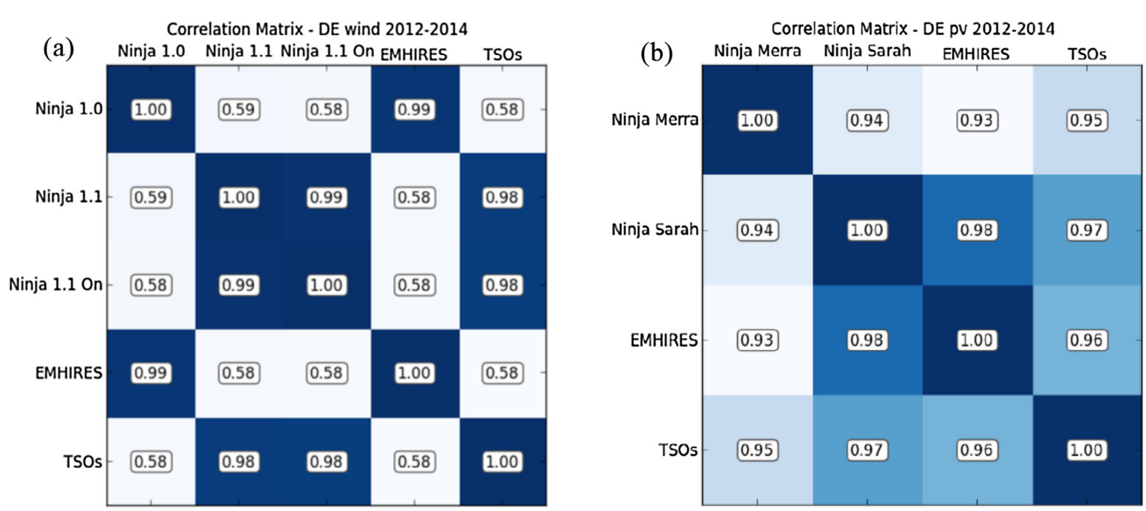

Fig. 1. Correlation matrices from 2012 until 2014 for wind (a) and PV (b) in DE.

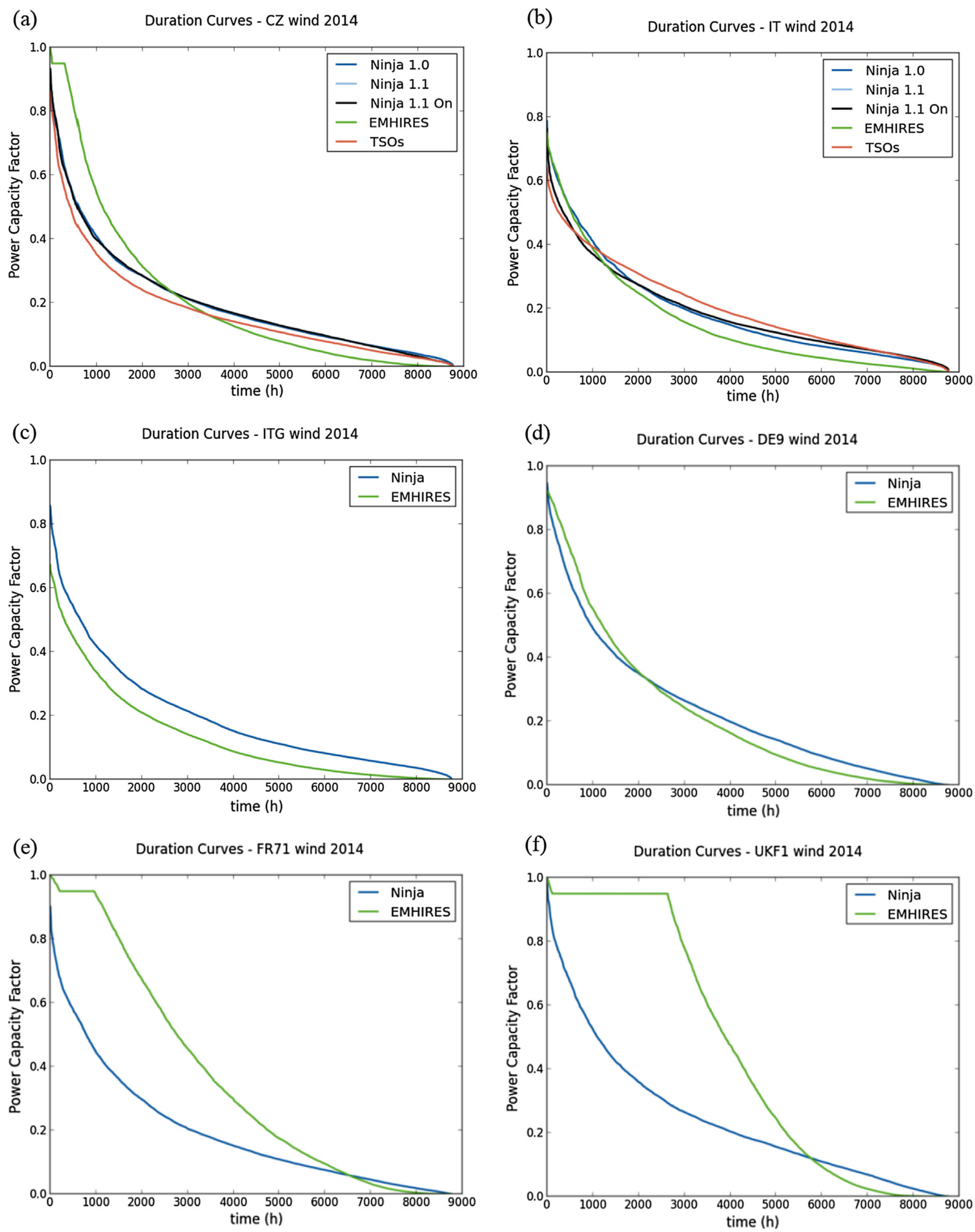

Fig. 2. Wind duration curves in 2014: at national level for CZ (a) and IT (b); at NUTS-1 level for ITG (c) and DE9 (d); and at NUTS-2 level for FR71 (e) and UKF1 (f).

3.3. Wind duration curves

This test revealed that wind duration curves were considerably different in many cases. At national level, the curves of EMHIRES were found atypical in CZ, compared to wind characteristic curves from [38] and from TSOs plots. In Ninja datasets, curve shapes and distributions were similar to TSOs. For the latter, CZ Wind 2012 and 2013 presented considerably lower values of capacity factors than those seen in 2014 and also compared with other countries.

At NUTS-1 level, more significant differences were found in FR7 and UKF, but in all zones differences were clear, as shown in for ITG and DE9 in Fig. 2(c) and (d). Duration curves from wind datasets at NUTS-2 level were those who presented the higher deviations, as can be seen in Fig. 2(e) and (f) for FR71 and UKF, respectively.

In order to quantify the differences found in duration curves at national level, a comparison was performed, where the richest and poorest hours of the distribution were calculated, as presented in Table 9. EMHIRES has higher shares of yearly energy yield in the 2000 richest hours and lower shares in the 2000 poorest hours. The richest hours of EMHIRES represent in average 57.3% while other datasets are in the range of 48.8–49.8%. Conversely, EMHIRES datasets contain only 2.7% of the poorest hours, in the same time other datasets have values between 5.3% and 5.7%. Duration curves for all countries and zones can be found in the supplementary material.

Table 9
Mean shares in % that the richest and leanest hours represent in the yearly wind energy yield for all countries and years here observed.

	Richest hours			Leanest hours		
	2000hs	1500hs	1000hs	2000hs	1500hs	1000hs
Ninja 1.0	49.5	40.6	30.0	5.7	3.6	2.0
Ninja 1.1	48.8	39.9	29.6	5.7	3.6	1.9
Ninja 1.1 On	49.6	40.8	30.3	5.5	3.5	1.9
EMHIRES	57.3	47.6	35.5	2.7	1.5	0.7
TSOs	49.8	40.8	30.3	5.3	3.3	1.8

3.4. PV duration curves

In general, at national level all PV time series presented maximum values between 0.6 and 0.8 and the distribution reached zero between 4000 and 5000 h. At NUTS-levels, Ninja dataset showed the same behaviour of national level, but for EMHIRES in all zones duration curves went over 0.8. These characteristics can be identified in Fig. 3. For FR and IT the duration curve is almost linear in all datasets. In all remaining countries and NUTS-zones duration curves are less linear.

At national level, Ninja Merra has the most linear approximate curves. EHMIRES is closer to TSOs than Ninja in all countries. TSOs maximum values in UK 2013 are far higher than others. At NUTS-levels, duration curves for Ninja and EMHIRES are quite similar. All curves are presented in the supplementary material.

3.5. Wind full load hours

Considering values of full load hours in every country and year, significant differences between the maximum and the minimum values were found. Table 10 presents these differences separated in two categories, which are showed for each country. Columns labelled "All"

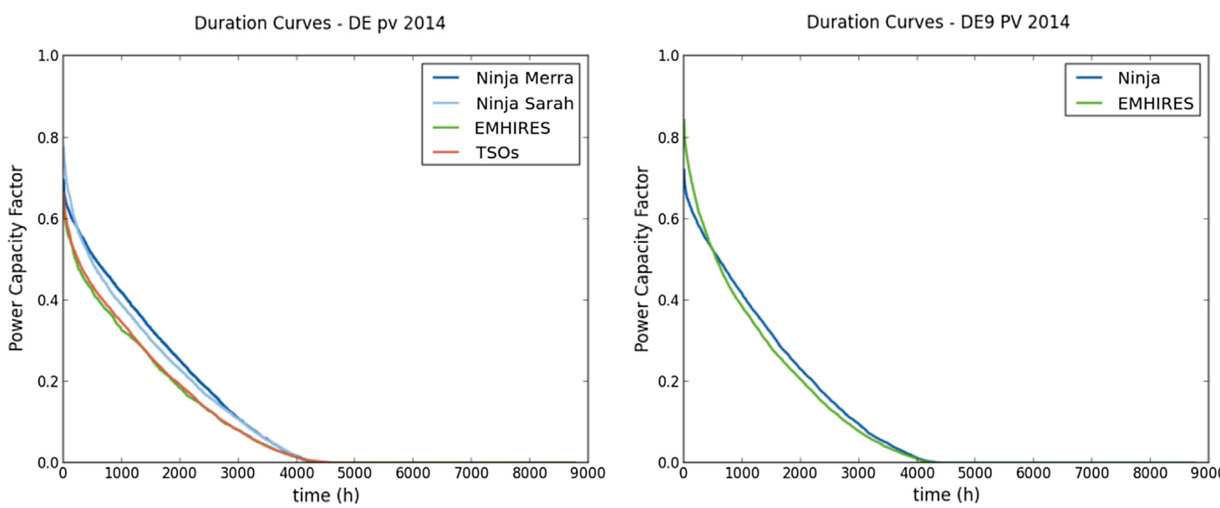

Fig. 3. PV duration curves for DE (left) and DE9 (right) in 2014.

Table 10
Maximum difference found for wind full load hours among all datasets, in every country and year - column "All". Maximum difference found for wind full load hours between Ninja 1.1 On and EMHIRES (only onshore), in every country and year - column "On".

Year	CZ		DE		FR		IT		UK	
	All	Onshore	All	Onshore	All	Onshore	All	Onshore	All	Onshore
2012	1135.8	155.5	140.5	125.2	111.5	110.5	273.1	273.1	450.7	121.5
2013	684.8	132.6	131.7	125.5	96	94.9	344.1	274.9	442.3	152.3
2014	322.4	86.2	207	142	102.9	101.9	394.7	269.7	511	131.4

Table 11
Maximum difference found for wind full load hours between EMHIRES and Ninja, in each NUTS-2 zone.

	ITG1	DE92	FR71	UKF1
2012	608.8	449.3	1502.1	2003.9
2013	608.4	455	1419	2033.3
2014	584.7	472.1	1432.5	1893.1

Table 12
PV full load hours differences between maximum and minimum values of the country in each year compared to the difference excluding TSOs, shown in the column “no TSO”.

Year	CZ		DE		FR		IT		UK	
	All	No TSO	All	No TSO	All	No TSO	All	No TSO	All	No TSO
2012	148	121.6	228	228	109	106.1	347.2	171.1	187.7	137.9
2013	197.8	197.8	299.1	299.1	151.6	151.6	231	185.8	324.3	194.8
2014	165	139.4	225.2	225.2	105.3	105.3	263	194.9	200.4	113.2

Table 13
Maximum difference found for PV full load hours between EMHIRES and Ninja, in each NUTS-2 zone.

	ITG1	DE92	FR71	UKF1
2012	95.7	58.2	341.2	61.8
2013	73.1	138.6	153	125
2014	100.8	48.1	275.3	44.4

present differences calculated, considering full load hours from all datasets. Columns labelled “Onshore” showed differences between Ninja 1.1 On and EMHIRES, which are onshore only. The differences in CZ reached 1035 h in 2012, primarily related to the doubtful values found from TSOs dataset. For all other countries the differences reached almost more than 500 h, when comparing all datasets. At NUTS-levels the differences were even higher. In Table 11, full load hours differences for NUTS-2 datasets were presented. It can be seen that there are no values lower than 449 h. In the case of UKF1, differences went even over 2000.

Table with full load hours in all countries and NUTS-zones for all datasets is found in the supplementary material.

3.6. PV full load hours

In PV datasets, numbers of full load hours presented also relatively high differences. Following the same logic used to analyse quantitatively wind full load hours, the maximum difference in each country and year was calculated. In Table 12 these differences are presented for country level. Among all datasets the maximum difference was 347 h, found for Italy in 2012. At NUTS-1 differences were lower than national level, but still with some values over 100. At NUTS-2, they were higher than NUTS-1, especially in FR71, as can be seen in Table 13.

Tables with full load hours in all regions and years are found in the supplementary material.

3.7. Weekly averages

This test was performed only with national level time series. The results showed that datasets can present considerable weekly differences. In each plot 13 weeks are showed, labelled by the first day of the week followed by its correspondent month e.g. 01-06 represents the week beginning on the 1st and finishing at the 7th of June. For each week, the power capacity factor averages in all datasets analysed are shown as bars. In Fig. 4, on the left hand side, we can identify differences in weekly averages for PV capacity factors in Italian summer of about 0.07 between Ninja Sarah and TSOs. On the right hand side, in some weeks it is possible to find considerable differences, like in week 01-06 (more than 0.1) and in week 29-06 (more than 0.25). Those differences found in DE wind are not found in every week like in IT PV, but the values are considerably higher.

3.8. Seasonal ratios

As defined in Section 2.1.5, this check tries to identify energy yield differences between warm and cold seasons, which represent 6 months of the year, when choosing different datasets. For every dataset, the energy yield of warm and the cold seasons are presented in the plot, as shares of the yearly energy yield (100%). The ratios found for PV in UK in 2014 were picked as an example and presented in Fig. 5. When comparing Ninja Merra with EMHIRES, the ratios present more than 5% difference.

3.9. PV hourly capacity factors plots

The results for this analysis reveal that datasets present different PV power profiles along the days and seasons. Data from TSOs has in all countries considerably different intensities compared to other datasets, with exception of DE. TSOs in CZ present periods with relative longer days in winter, which we doubt to be correct, and in UK the values for

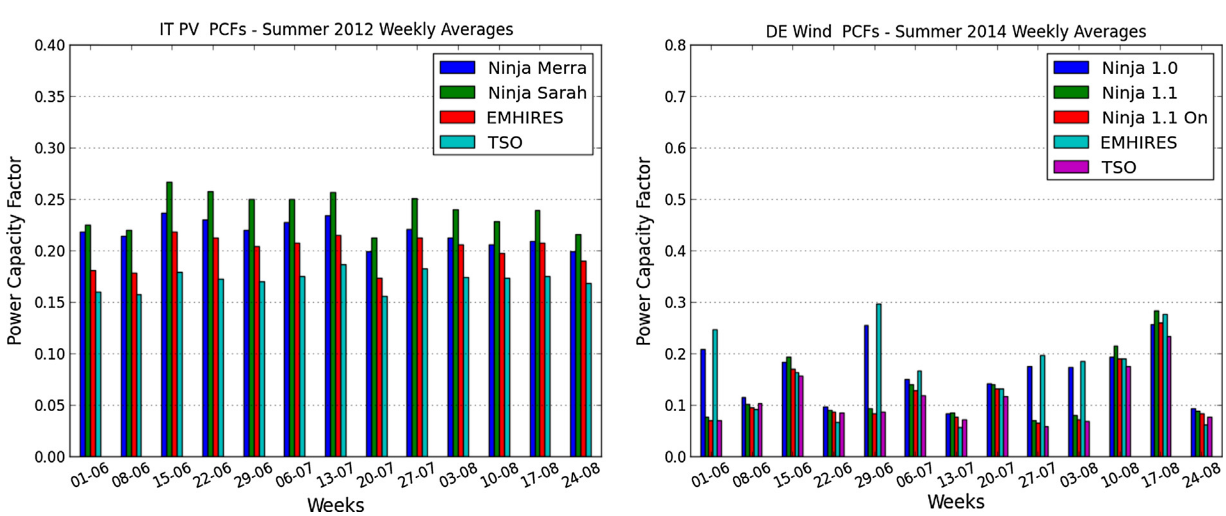

Fig. 4. Comparison of capacity factor weekly averages for PV in IT on the summer of 2012 (left) and for wind in DE on the summer of 2014 (right).

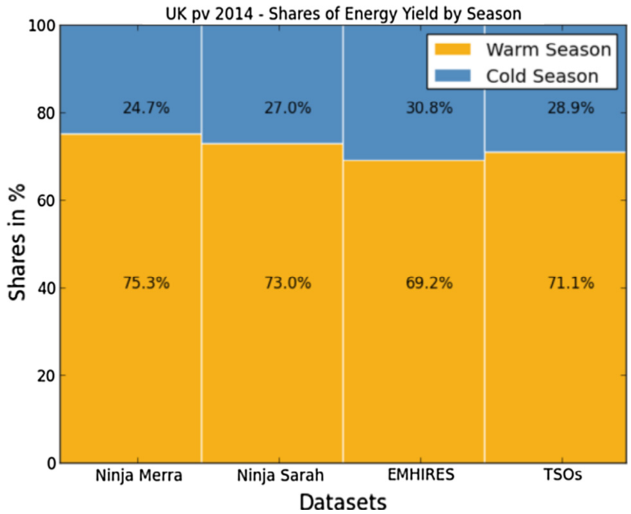

Fig. 5. Ratios of energy yield in warm and in cold seasons of 2014 for all datasets.

2013 are far higher than in the other years, as also shown in the duration curves analysis. Ninja Merra and Ninja Sarah present similar daily distribution of power, with more hours of availability than EMHIRES and TSOs. The power values in Ninja Merra are in most of the countries clearly higher. Country level plots for Ninja Sarah, Ninja Merra, TSOs and EMHIRES in IT are compared in Fig. 6. NUTS-1 and NUTS-2 plots confirm distinction between Ninja and EMHIRES PV profiles. On one side, Ninja has also more daily hours of PV power availability than EMHIRES in summer and the transition between days is smoother. On the other side, EMHIRES shows higher peak powers, different than at national level, and some periods of less power availability are identified. See supplementary material for plots from other countries analysed.

The plots present on the y axis the daily profile in hours and x axis represents yearly path in days. The colors express the intensity of the power capacity factors.

3.10. Wind future

Wind future analysis is described here with access to plots available in the supplementary material. With exception of IT, all three other countries showed more linear duration curves. Values for capacity factors were higher in almost all plots with maximum close to 1.0. This, comparing to actual fleet and TSOs, resulted in much higher full load hours: DE about 2500hs (ca. 40% higher); FR 2800–3000hs (30–37% higher); UK 3300–3450hs (15–24% higher). In IT full load hours for near term were the highest and long term always lower than actual fleets and TSOs. Correlation analysis showed that in DE, FR and IT the

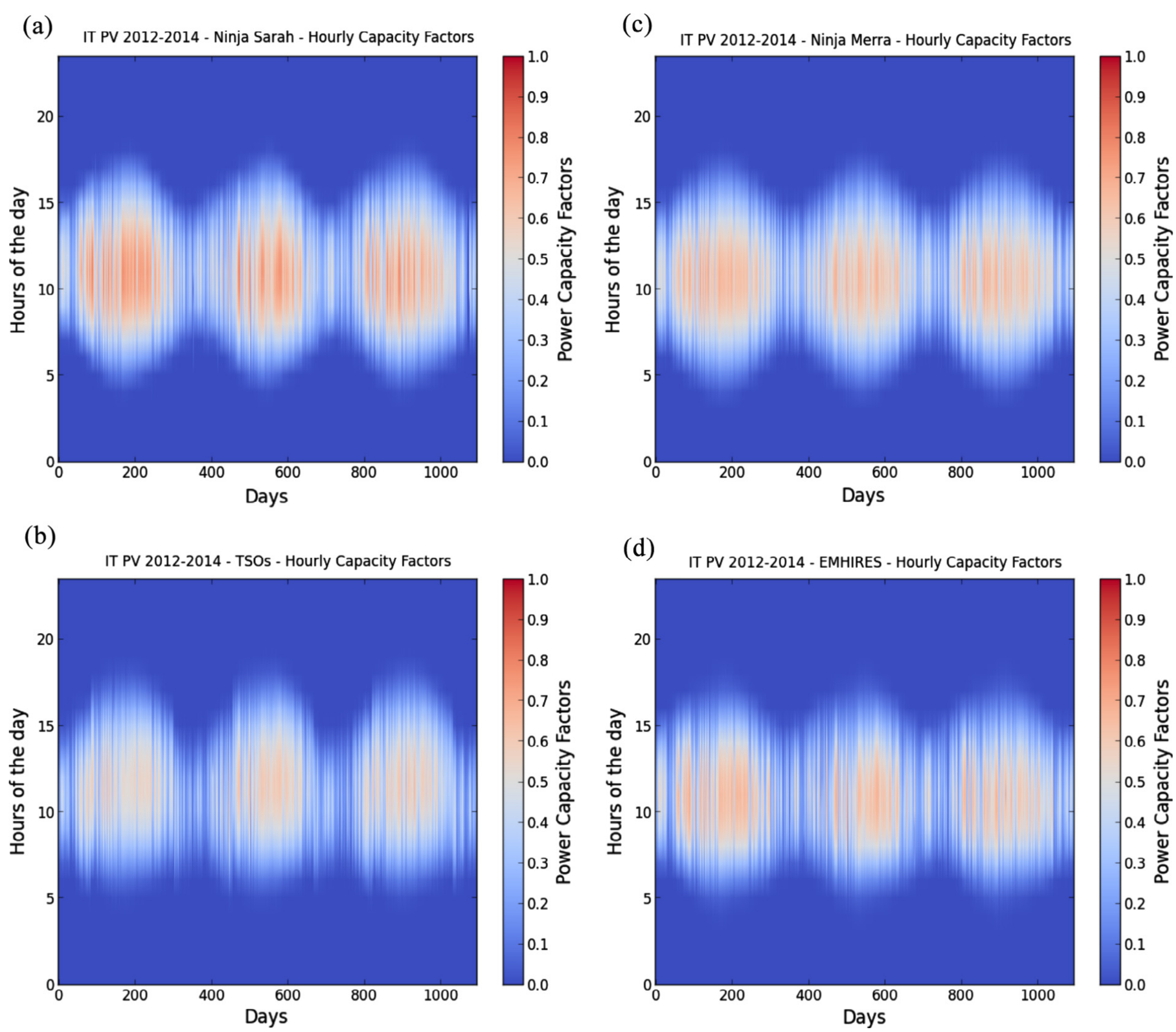

Fig. 6. Power capacity factors plots in hours of the days vs. days for IT from 2012 to 2014, in all datasets analysed.

strongest correlation lies between Ninja 1.1 and TSOs. Conversely, in IT there is weak correlation between long term and all others (from 0.34 to 0.49). To access all plots, see the supplementary material.

4. Discussion

It is relevant at this point, to highlight once again the importance of benchmarking and applying the test scheme on the datasets. These time series are long-term estimations of power production factors for different spatial resolutions represented as hourly values. The amount of information behind the capacity factor is considerably high and the datasets were built in a chain of methods, in which every step is an error source. The data sources used by their authors are result of reanalysis and satellite data, which are already estimations made through atmospheric models, commonly not composed of measurements of wind speeds and solar radiations. Also, the novelty of these open license datasets requires plausibility and quality checks, before scientist start producing analysis based on them.

The results for our analysis showed considerable differences in certain aspects among datasets. The greatest differences were found among wind data compared to PV. At national level, results indicate greater similarity between TSOs and EMHIRES for PV while TSOs and Ninja compare better in the wind datasets. Comparisons were limited to countries and years, where we found reference data from TSOs. Even though we found comparison data, in cases such as wind in CZ (2012 and 2013) and PV in UK (2013) we doubt the data accuracy, as curves (to be found in the supplementary material) expose considerable atypical values. Additionally, as presented in Tables 6 and 7, data from TSOs and studies, as well as from government sources, diverge from each other. Another limitation of data is due to the fact that there are still a considerably low number of installations of technologies like Wind in CZ and PV in CZ, FR and UK, and it causes the values to be less representative of the whole region potential. Analysis of separated onshore and offshore was not in the scope of this paper.

In EMHIRES, the results for duration curves and full load hours in CZ lead one to suppose that data contained errors. In CZ, we found periods where wind power was constant for many hours in certain periods of the year. This is not a probably occurrence for an entire country, considering the combination of randomness of wind and the wide spatial distribution of wind parks. The maximum values in CZ and DE close to 1.0, we supposed to be of extremely low probability. Just to exemplify this, one can consider the whole country of DE producing almost 100% of its wind capacity for multiple hours in the year. Studies confirm this concern, in which the maximum instant capacity factor for DE has been 0.83 in 2012, 0.81 in 2013 and 0.83 in 2014 [42]. Ninja also reported values very close to 1.0 in UK for all datasets.

Even though using the same dataset from NASA (MERRA-1) as base and preserving the similarity in the variability showed by the correlation analysis, Ninja 1.0 and EMHIRES resulted in very different duration curves and full load hours. On the other hand, the use of MERRA-2 in release 1.1 of Ninja is likely the reason of the weaker correlation when compared to release 1.0. Even with correlation factor around 0.6 they showed very similar duration curves and full load hours.

Extremely high differences found for wind duration curves at NUTS-levels, especially in French and British zones, make us concerned about using datasets before validating against other sources. This cannot be extended to PV datasets, which presented similar duration curves, very close one another, but we also recommend further validation.

Full load hours analysis indicated considerable differences between datasets. To understand the extent of these impacts on energy system planning, one can compare the case of Italy. In 2014 the country produced 267 TWh of electricity, 15 TWh produced from wind [30]. The national wind generation capacity was 8.54 GW at the end of 2013 and 8.68 GW at the end of 2014. The full load hours' difference in our analysis between Ninja 1.1 On and EMHIRES, both only onshore, was approximately 270 h, shown in Table 9. Using the wind capacity of the end of 2013 it would represent a difference of about 2 TWh on the yearly energy generation, which represents 13% of the annual wind energy production. We recommend extending this analysis, when possible, to NUTS-level time series, before using the datasets.

Correlations and duration curves were closer within each other for PV in comparison with wind data, probably because solar behaviour is more predictable, mainly due to daily and seasonal cycles. Although duration curves appear similar for PV at national level, numbers of full load hours were considerably different. This fact was also evident between EMHIRES and Ninja Sarah, which were based on the same dataset. The comparison made for wind full load hours can also be used to evaluate impacts of differences of PV full load hours. Taking into account only EMHIRES and Ninja datasets, the difference was 299 h for Germany in 2013, presented in Table 12. In this case, taking the German PV capacity at the end of 2012 (33.1 GW) [30], this difference would represent 9.89 TWh of electricity, correspondent to 1.7% of the annual PV energy produced in that year.

The differences found in weekly averages were also considerable. While for DE higher values in some sparse weeks were present, in IT smaller values in all weeks were found. In both cases differences may influence energy balance between weeks, especially when taking into account the dimensioning and/or operation strategy of medium and long-term energy storage. Also, the ratios of energy yield between the cold and the warm seasons presented differences, which need to be taken into account. The deviation between ratios can influence seasonal energy balance, impacting mainly the dimensioning and operation of long-term energy storage.

The extreme differences in the comparison of actual with future wind datasets from Ninja may be present due to the considerable additional capacity (27% near term and 33% long term), mainly offshore. Unfortunately future datasets for wind are not available for all countries and for PV they do not exist to our knowledge.

These facts show us that the chain of methods used to convert wind speeds and solar radiation into power outputs are decisive in this process, and the use of reanalysis data is promising. The lack of a trustworthy source of data for comparison makes the evaluation of data quality challenging in many countries. It is also inconclusive which dataset would be the most appropriate as both sources showed deviations at national level to what we have as reference: TSOs data. The differences in capacity factor distributions, number of full load hours and maximum peak values, as well as hourly, weekly and seasonal variations bring us the question of what kind of effects it would implicate on potential simulations results. We strongly encourage simulation analysis using different datasets in further works. Also, different problems and models may have distinct sensitivities and the comparison of effects on simulation results would enrich this analysis.

5. Conclusions

After applying the test scheme, it was clear that datasets diverge from each other in all spatial resolutions, even though some are based on the same meteorological data source. The origin of this fact is probably caused by differences in development methods compared at the introduction of this work. The main differences in methods were found at the following steps: raw data selection (PV); gap filling methods and assumptions; spatial downscaling to farm level (wind); spatial distribution of farms (PV); correction of power curves (wind); power conversion models (PV); and calibration.

From a user perspective, these differences may influence results when simulating energy systems, if compared against each other using different datasets. The main implications of these differences may appear when simulating capacity investment plans and defining operation strategies for energy balance. Not only short-term analysis may be affected, but also medium and long-term simulations. It can also be concluded that the test scheme proposed here is suitable to verify and identify differences between datasets.

To be sure which dataset represents power feed-in closer to reality, more reliable data to compare are still lacking. There is a tendency towards more availability of data after 2015, especially for PV, due to transparency portals and increasing power capacity installations worldwide. Once more data from TSOs becomes available and are composed of more dense distribution of power plants in the countries, upcoming releases of datasets from projects like EMHIRES and Ninja will have more available data for validation, probably increasing data quality. In this case, users will also have more data to compare to and be able to choose which dataset uses the better model, which will result in time series closer to our problem requirements.

Declaration of interest

None.

Acknowledgements

This work was kindly supported by the joint scholarship program from DAAD/Germany and CNPq/Brazil and the German Federal Ministry of Economic Affairs and Energy (BMWi), within the project code FKZ: 0325692). Luiz Moraes thanks Iain Staffell and Stefan Pfenninger for making Renewables.ninja NUTS-level datasets available, before publication. The authors thank the Journal reviewers and editors for the comments, which contributed considerably to enhance the quality of this work.

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.2018.04.109.

References

[1] Harper Ian. Energy meteorology & the energy mix: interview with: Vera Silva, Charles Smith and Prof. Alberto Troccoli. Available from: < http://www.wemcouncil.org/INTERVIEWS/ICEM_ENERGYMIX.mp3 > [January 15, 2018].
[2] Olsson LE. Renewable energy, technology and the environment: ENERGY-METEOROLOGY: a new discipline? Available from: < https://www.sciencedirect.com/science/article/pii/B9780080412689500110 > [March 10, 2018].
[3] Aboumahboub T. Modeling and optimization of the global electricity generation system with high shares of fluctuating renewable energy sources: Modellierung und Optimierung des globalen Stromversorgungssystems mit einem hohen Anteil an fluktuierender Einspeisung aus erneuerbaren Energiequellen. Munich, Germany; 2012.
[4] Steinke F, Wolfrum P, Hoffmann C. Grid vs. storage in a 100% renewable Europe. Renew Energy 2013:826–32.
[5] Olauson J, Bergkvist M. Modelling the Swedish wind power production using MERRA reanalysis data. Renew Energy 2015;76:717–25.
[6] Gonzalez Aparicio Iratxe, Zucker Andreas, Careri Francesco, Monforti Fabio, Huld Thomas, Badger Jake. EMHIRES dataset: Part I: wind power generation European Meteorological derived HIgh resolution RES generation time series for present and future scenarios. EUR 28171 EN.
[7] Gonzalez-Aparicio Iratxe, Huld Thomas, Careri Francesco, Monforti Fabio, Zucker Andreas. EMHIRES dataset - Part II: Solar power generation. European Meteorological derived HIgh resolution RES generation time series for present and future scenarios.: Part II: PV generation using the PVGIS model. EUR 28629 EN.
[8] Pfenninger S, Staffell I. Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data. Energy 2016;114:1251–65.
[9] Staffell I, Pfenninger S. Using bias-corrected reanalysis to simulate current and future wind power output. Energy 2016;114:1224–39.
[10] Rienecker MM, Suarez MJ, Gelaro R, Todling R, Bacmeister J, Liu E, et al. MERRA: NASA's modern-era retrospective analysis for research and applications. J Climate 2011;24(14):3624–48.
[11] Pierrot M. The wind power. Available from: < http://www.thewindpower.net > [March 20, 2018].
[12] Global Wind Atlas 2.0, a free, web-based application developed, owned and operated by the Technical University of Denmark (DTU) in partnership with the World Bank Group, utilizing data provided by Vortex, with funding provided by the Energy Sector Management Assistance Program (ESMAP). Available from: < https://globalwindatlas.info > [March 10, 2018].
[13] RichardMüller, Pfeifroth U, Träger-Chatterjee C-C, Cremer R, Trentmann J, Hollmann R. Surface solar radiation data set - heliosat (SARAH) - edition 1: EUMETSAT satellite application facility on climate monitoring (CM SAF); 2015.
[14] Suri M, Hofierka J. A new GIS-based solar radiation model and its application to photovoltaic assessments. Trans GIS 2004;8:175–90.
[15] Lauret P, Boland J, Ridley B. Bayesian statistical analysis applied to solar radiation modelling. Renew Energy 2013;49:124–7.
[16] European Commission. Photovoltaic geographical information system (PVGIS). Available from: < http://re.jrc.ec.europa.eu/pvgis.html > [February 25, 2018].
[17] Huld T, Gottschalg R, Beyer HG, Topič M. Mapping the performance of PV modules, effects of module type and data averaging. Sol Energy 2010;84(2):324–38.
[18] PV-Log: Internetportal zur Überwachung und zum Vergleich von Erträgen aus Photovoltaikanlagen. Available from: < http://www.pv-log.com/ > [March 01, 2018].
[19] PVOutput: Free service for sharing and comparing PV output data. Available from: < https://www.pvoutput.org/ > [March 01, 2018].
[20] Munzinger M, Crick F, Dayan E, Pearsall N, Martin C. Domestic photovoltaic field trials. Final technical report. Tech. Rep. URN: 06/2218, BRE and DTI. Available from: < http://webarchive.nationalarchives.gov.uk/20100104194416/http://www.berr.gov.uk/files/file36660.pdf > [March 10, 2018].
[21] Eurostat. Nomenclature of territorial units for statistics. Available from: < http://ec.europa.eu/eurostat/web/nuts > [15 Sptember 2017].
[22] CEPS. Data. Available from: www.ceps.cz/ENG/Data/Vsechna-data/Pages/default.aspx [October 10, 2017].
[23] Open power system data. Data package time series: version 2017-07-09. Available from: < https://data.open-power-system-data.org/time_series/2017-07-09/ > [January 10, 2017].
[24] RTE France. Eco2mix Téléchargement: < http://www.rte-france.com/en/eco2mix/eco2mix-telechargement-en > . Available from: < http://www.rte-france.com/en/eco2mix/eco2mix-telechargement-en > [October 17, 2017].
[25] Terna. Ex post data on the actual generation: < http://www.terna.it/it-it/sistemaelettrico/transparencyreport/generation/expostdataontheactualgeneration.aspx > . Available from: < http://www.terna.it/it-it/sistemaelettrico/transparencyreport/generation/expostdataontheactualgeneration.aspx > [October 17, 2017].
[26] National Grid. National grid data explorer > historical demand data. Available from: < http://www2.nationalgrid.com/UK/Industry-information/Electricity-transmission-operational-data/Data-Explorer/ > [October 17, 2017].
[27] Red Electrica de Espana. < https://www.esios.ree.es/es/generacion-y-consumo > . Available from: < https://www.esios.ree.es/es/generacion-y-consumo > [October 17, 2017].
[28] Svenska Kraftnät. < http://www.svk.se/en/search/?search-field=xls > . Available from: < http://www.svk.se/en/search/?search-field=xls > [October 17, 2017].
[29] Open power system data. External sources > data sources. Available from: < https://open-power-system-data.org/data-sources > [October 17, 2017].
[30] ENTSO-E. Statistical database. Available from: < https://www.entsoe.eu/data/data-portal/Pages/default.aspx > [January 02, 2018].
[31] EurObserv'ER. All Wind energy barometers: < https://www.eurobserv-er.org/category/all-wind-energy-barometers/ > . Available from: < https://www.eurobserv-er.org/category/all-wind-energy-barometers/ > [October 17, 2017].
[32] IRENA. Renewable Energy Statistics 2017. Available from: < http://www.irena.org/DocumentDownloads/Publications/IRENA_Renewable_Energy_Statistics_2017.pdf > [October 17, 2017].
[33] WindEurope. About wind > Statistics: < https://windeurope.org/about-wind/statistics/ > . Wind in Power reports from 2012 to 2015. Available from: < https://windeurope.org/about-wind/statistics/ > [September 28, 2017].
[34] EurObserv'ER. All Photovoltaic barometers: < https://www.eurobserv-er.org/category/all-photovoltaic-barometers/ > . Available from: < https://www.eurobserv-er.org/category/all-photovoltaic-barometers/ > [October 17, 2017].
[35] IEA International Energy Agency. Trends 2015 in photovoltaic applications: executive summary. Report IEA-PVPS T1-27:2015. Available from: < http://www.iea-pvps.org/index.php?id=6 > [October 17, 2017].
[36] Rodgers JL, Nicewander WA. Thirteen ways to look at the correlation coefficient. Am Stat 1988;42(1):59–66.
[37] Ruslana Dalinina. Introduction to correlation. Available from: < https://www.datascience.com/blog/introduction-to-correlation-learn-data-science-tutorials > [November 14, 2017].
[38] Sebastian Pfaffel. Wissenschaftliches Monitoring- und Evaluierungsprogramm zur Offshore-Windenergienutzung: Durchführungsphase I. [October 10, 2017].
[39] WEATHERBASE. Available from: < http://www.weatherbase.com > [March 12, 2018].
[40] timeanddate.com. Available from: < https://www.timeanddate.com > [March 12, 2018].
[41] German Ministery for Economic Affairs and Energy (BMWi). Erneuerbare Energien: Energiedaten - Tabelle 20. Available from: < http://www.bmwi.de/Redaktion/DE/Infografiken/Energie/Energiedaten/Energietraeger/energiedaten-energietraeger-25.html > [December 17, 2017].
[42] Bruno Burger. Renewable energy data: electricity production from solar and wind in Germany. Yearly reports from 2012 to 2014. Available from: < https://www.ise.fraunhofer.de/en/renewable-energy-data.html#faqitem_9-answer > [October 23, 2017].