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