Renewable Energy 96 (2016) 784–791

Contents lists available at ScienceDirect

Renewable Energy

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

Restoring the missing high-frequency fluctuations in a wind power model based on reanalysis data

Jon Olauson [a, *], Hans Bergström [b], Mikael Bergkvist [a]

[a] Division of Electricity, Department of Engineering Sciences, Uppsala University, Sweden
[b] Department of Earth Sciences, Uppsala University, Sweden

ARTICLE INFO

Article history:
Received 27 February 2015
Received in revised form
16 March 2016
Accepted 3 May 2016

Keywords:
Wind power variability
Statistical modelling
Machine learning
Power spectral density
MERRA reanalysis dataset

ABSTRACT

A previously developed model based on MERRA reanalysis data underestimates the high-frequency variability and step changes of hourly, aggregated wind power generation. The goal of this work is to restore these fluctuations. Since the volatility of the high-frequency signal varies in time, machine learning techniques were employed to predict the volatility. As predictors, derivatives of the output from the original "MERRA model" as well as empirical orthogonal functions of several meteorological variables were used. A FFT-IFFT approach, including a search algorithm for finding appropriate phase angles, was taken to generate a signal that was subsequently transformed to simulated high-frequency fluctuations using the predicted volatility. When comparing to the original MERRA model, the improved model output has a power spectral density and step change distribution in much better agreement with measurements. Moreover, the non-stationarity of the high-frequency fluctuations was captured to a large degree. The filtering and noise addition however resulted in a small increase in the RMS error.

1. Introduction

The energy available in the wind varies over all time scales, from sub-seconds to decades. The shorter time scales of variability are of importance to the mechanical and electrical design of wind energy converters (WECs) but are less crucial for the power system in terms of added variability from wind power. Variations from around 5–60 min and longer are however of utmost importance to understand in order to successfully integrate wind power into the power system [1]. A larger share of wind power, with its variability and uncertainty, implies a challenge for the power system in terms of e.g. reliability [2] and increased cycling of conventional units [3].

Using meteorological models to simulate wind power generation has several advantages over upscaling historical time series or using purely statistical methods. The most important of these is perhaps the possibility to adequately model output of future wind farms, including technological improvements and deployment in new geographical areas. A model of hourly, aggregated wind power production based on MERRA reanalysis data [4] was described and evaluated in Ref. [5]. Several parameters, accounting for e.g. different types of losses and smoothing of the power curve, were tuned to give a good agreement to observations from the Swedish Transmission System Operator (TSO). Overall, the model performed well, but the power spectral density (PSD) was underestimated for frequencies above $(10\ \mathrm{h})^{-1}$, see Fig. 6d in Section 3.3. This lead to an underestimation of step changes in hourly energy production; the standard deviation of 1 h step changes was for instance 14% lower than for measurements.

The abovementioned deviances are consistent with earlier results [6–11]. The spectra of wind speeds from numerical weather prediction (NWP) models, or wind generation modelled from such time series, departs from measurements at higher frequencies. The point of departure, which often seem to be around 5–10 times coarser than the NWP resolution, can be seen as a direct measure of the model's effective or "true" resolution [7]. For modelling of wind generation, this lack of high-frequency variability leads to overly smooth profiles and an underestimation of step changes [6]. In other studies, the opposite however holds; too little smoothing and an overestimation of the aggregated high-frequency variability including step changes [12,13]. This could be a result of too high correlation of the modelled wind speed time series [12]. The abovementioned studies demonstrate the importance of validating the model output with actual (aggregated) generation in the power system of interest.

* Corresponding author.
E-mail address: jon.olauson@angstrom.uu.se (J. Olauson).

http://dx.doi.org/10.1016/j.renene.2016.05.008
0960-1481/

Earlier attempts to adjust for the missing high-frequency fluctuations in NWP models have been made [8–11], but to our knowledge always on a farm or WEC level. In Ref. [9], a correction was performed to CFSR reanalysis spectra with the objective of better estimating the 50 year extreme wind speeds. For higher frequencies, the spectra from CFSR was replaced by either spectra from measurements or an assumed slope of $-5/3$, which resulted in higher estimates of the extreme winds. In Ref. [11], observed histograms of deviations from the power curve (power vs. wind speed) from a donor site was used for probabilistic modelling of output from a WEC with more realistic fluctuations, see also [14].

The main contribution of our work is a thoroughly validated and computationally inexpensive methodology for simulating realistic high-frequency fluctuations of aggregated (power system scale) wind generation. It is also demonstrated that least square gradient boosting is a suitable technique for predicting volatility on the power system scale and that a simple search algorithm for appropriate phase angles improves the model performance considerably. Throughout this paper, when referring to the “MERRA model” we mean the model described in Ref. [5], i.e. a model that uses MERRA reanalysis data to calculate the aggregated, hourly wind power generation. With “Measurement” or “SvK” we mean actual generation measured by Svenska Kraftnät, the Swedish TSO. Hourly measurements of national wind power production from 2007 to 2012 (six years) were available. With “Improved model” we are referring to the MERRA model combined with the statistical correction described in this paper.

2. Methods

A straightforward solution to the problem would be to add stochastic noise in order to fill the gap in the PSD curve. The error in volatility of the hourly energy is however nonstationary; sometimes the MERRA model gives an accurate description of the fluctuations, but during many periods there is much more volatility in observations. We therefore developed a method to increase the high-frequency fluctuations, taking into account this nonstationary behaviour. To put it shortly, the idea was to mimic the high-frequency fluctuations in measured data and to add the simulated noise to low-pass filtered time series from the MERRA model. The methodology involves the following steps:

1) Separate hourly data from SvK and the MERRA model (M) into their low- and high-frequency components, e.g. SvK_{HF} and M_{LF}.
2) Transform SvK_{HF} to an approximately stationary time series y_{SvK} using a variance stabilising filter.
3) Find the magnitudes of the frequency domain representation of y_{SvK} using FFT (Fast Fourier Transform).
4) Generate y_{sim} of the same length as the MERRA time series, using inverse FFT with interpolated and scaled magnitudes from 3) and appropriate phase angles.
5) Transform y_{sim} to $M_{HF,sim}$ using predicted volatility.
6) The time series for the improved model is achieved as $M_{LF} + M_{HF,sim}$.

For step 5, three different models were considered for predicting the volatility, see Section 2.3. Many different sets of tuning parameters were evaluated for each of these models. In order to not get overly optimistic results, the six year dataset was partitioned into three parts: a training set (4/6 of the data), a validation set for tuning the model parameters (1/6) and a test set for determining the performance (1/6). Since we are dealing with time series with considerable autocorrelations in most of the variables, whole years of data were used in each set instead of random sampling. By using whole years, all seasons were also included in each set. Each of the six years was defined as the test set once and the training and validation years were chosen randomly before searching over the tuning parameter grids to find the fittest models.

For better readability, the continuation of this section is structured into subsections following to the numbered list above (step 1–2, step 3–4 and step 5–6 respectively).

2.1. Filtering

For the separation of frequencies a windowed sinc filter with a 5000 samples long kernel was used. This gives a fast roll-off and sufficient stopband attenuation. The cutoff frequency was set to $(10\ \text{h})^{-1}$, see discussion in Section 5. The separation of a signal into its high- and low-frequency components is shown in Fig. 1. Such as in the rest of this paper, hourly energy is expressed in per unit (p.u.), where one p.u. corresponds to the installed wind power capacity at that time.

As can be seen in Fig. 1, the volatility of SvK_{HF} varies considerably in time. We define a variability index (VI) in a similar manner as in Ref. [15]; the 24-hour moving standard deviation of a signal. Note that in this paper, the VI of several different variables are computed, but it should be clear from the context which is intended. A more stationary time series can be accomplished by dividing the signal by its variability index. This filter has been shown to effectively stabilise the variance [16]. A few weeks example of SvK_{HF}, the VI of SvK_{HF} and $y_{SvK} = SvK_{HF}/VI$ is shown in Fig. 2.

2.2. FFT-IFFT

After applying the variance stabilising filter, an approximately stationary and $N(0,1)$ distributed time series y_{SvK} resulted. The next step was to find the frequency domain representation of y_{SvK} with FFT (Fast Fourier Transform). By using interpolated magnitudes from the FFT of y_{SvK} scaled by $\sqrt{length(y_{sim})/length(y_{SvK})}$ and appropriate phase angles (see below), simulated time series y_{sim} of desired length can subsequently be generated with the inverse FFT algorithm.

In order to get real valued time series, the magnitudes and phases were made symmetrical. The DC-component was forced to be zero. In an earlier version of the method, random phase angles were used in the same manner as in e.g. Refs. [17], [18]. We however noted that the VI of y_{sim} then fluctuated quite a bit which led to a deterioration of the models' ability to predict the varying volatility. Inspired by Ref. [19], an algorithm was therefore used to search for

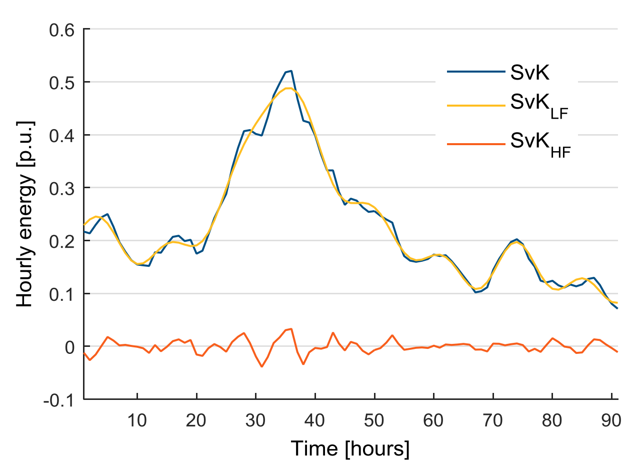

Fig. 1. Example of separation of measured hourly wind generation for the whole of Sweden (SvK) into its low- and high-frequency components (subscript LF and HF respectively).