Synthetic wind speed generation for the simulation of realistic diurnal cycles

Synthetic wind speed generation for the simulation of realistic diurnal cycles

D D'Ambrosio[1], J Schoukens[1,2], T De Troyer[1], M Zivanovic[3] and M C Runacres[1]

[1] Vrije Universiteit Brussel (VUB), Department of Engineering Technology (INDI), Pleinlaan 2, 1050 Brussel, Belgium
[2] Eindhoven University of Technology, Department of Electrical Engineering, Building Flux, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
[3] Universidad Pública de Navarra, Electrical Engineering and Communication Department, Campus Arrosadia, s/n 31006 Pamplona, Spain

E-mail: daniele.dambrosio@vub.be

Abstract. Synthetic wind-speed generators can provide a detailed characterisation of the wind variability at different time scales. A keen interest in the availability of synthetic wind speeds has recently risen in wind power modelling applications. In particular, a proper simulation of the diurnal and annual variability of the wind speed is sought that can lead to a more efficient grid integration of this renewable source. This paper proposes a statistical model for generating synthetic wind speeds consistent with both the probability density function and the spectral density function of a measured wind-speed dataset and that simulates accurately its average diurnal variation. To test the proposed methodology, multiple synthetic time series are generated using three long-term wind-speed time series recorded at a meteorological site in the Netherlands. The accuracy in terms of the statistical descriptors of the generated time series and their average diurnal variation is assessed with respect to the target data. We show that the average diurnal cycles present in all the three measured time series are always reproduced accurately, and that the statistical descriptors of the target dataset are constantly matched with high accuracy. Possible advantages of the present approach in terms of power system modelling are discussed.

1. Introduction

In recent years the penetration of wind power in the electricity systems has increased considerably. As a result, a growing need to efficiently integrate the increasing share of wind energy into the grid has emerged. To meet this demand, a sound understanding of the variability of the wind resource has become crucial given that its fluctuating nature across a wide range of temporal scales drives the wind power generation [1]. In the context of grid integration, the diurnal variation of wind speed takes on a pivotal role [2, 3]. In particular, measurements at multiple sites have shown that the wind speed displays clear diurnal patterns [4] that may eventually cause misalignments between the wind power production and the electricity demand throughout the day, and thus may lead to inefficient curtailment of wind power generation [5]. Therefore, in order to perform more accurate and reliable modelling for wind power systems it has become increasingly beneficial to rely on synthetically generated wind speeds that can simulate realistic diurnal variations of the wind speed [6].

Over the years, a very diverse range of methodologies have been successfully implemented for modelling the wind speed at different temporal resolutions [7]. However, most of these methods concentrated their efforts on forecasting the exact value of the wind speed at future temporal intervals, by implementing either physical or statistical models [8]. Only few works have proposed statistical approaches for the generation of plausible wind-speed time series that conform to statistical properties of the measured wind data (so-called scenario generation) [9].

In this work we present a statistical model for the generation of synthetic wind-speed time series that are consistent with the probability density function (PDF) and the power spectral density function (PSD) of target wind-speed datasets, and can reproduce accurately the average diurnal variation (ADV) of the wind speed observed in the input data. This methodology relies on the technique for designing broadband excitation signals with a user-imposed power spectrum and a probability distribution developed by Schoukens and Dobrowiecki [10], and can be considered as a revised version of the Iterative Amplitude Adjusted Fourier Transform (IAAFT) algorithm originally developed by Schreiber and Schmitz [11, 12].

The synthetic wind speed signal is generated by iteratively performing a rank-reordering process of a sequence conforming to the target PDF followed by a signal reconstruction technique that restores the PSD of the target wind-speed dataset. This iterative algorithm converges to the desired statistical descriptors allowing to properly model the observed long-term probability distribution of the wind-speed process and its energy content across the desired range of temporal scales or frequency components. In particular, both the annual and the diurnal deterministic components of the wind speed spectrum are correctly reproduced and the nonstationarity pertaining to the diurnal temporal scale is properly recovered in the time domain.

The proposed method is tested for the generation of synthetic wind-speed data with a temporal resolution of 10 minutes by employing wind-speed measurements recorded from the meteorological mast at Cabauw in the Netherlands at three different heights. The same data are then used to validate the methodology by assessing the agreement of the synthetically generated data with the measurements in terms of their PDF, PSD, and average diurnal variation of the wind speed. With this test, we show that the proposed algorithm can generate 10-minute synthetic wind speeds with a desired probability distribution and a prescribed power spectrum whose average diurnal variation simulates accurately the one observed in the data.

2. Methodology

The proposed algorithm is explained in detail below and illustrated in the flow chart of Fig. 1. It starts with an initialization phase where two initial sequences are generated that are respectively consistent with the PSD and the PDF of the target wind-speed time series. An iterative procedure first reorders the samples conforming to the target PDF so as to match the autocorrelation of the sequence possessing the target PSD. Then, it restores the power spectrum of the target dataset by generating a new signal employing the Fourier amplitudes of the input data and the phases of the Fourier transform of the reordered sequence. Finally, a bin-averaging technique is implemented to both the synthetically generated signal and the input signal to determine and compare the average diurnal variation of the wind speed.

2.1. Initialization phase

The algorithm begins with the generation of two sequences, the first one possessing the same PSD of the target wind-speed time series and the second one consistent with the PDF of the measured wind-speed samples.

The first sequence is generated as a random-phase multisine signal with the amplitudes of its sinusoidal components determined by the Fourier amplitudes of the target dataset and randomly generated phases. Let $x(t_n)$ be the measured wind-speed time series we would like to generate synthetic data from, that consists of N samples taken at the equally-spaced time

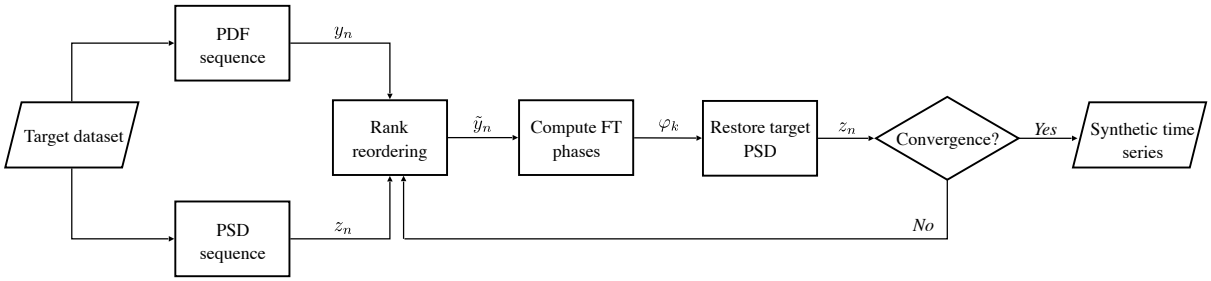

Figure 1. High-level flowchart of the proposed methodology.

instants $t_n = n\Delta t$, $n = 0, \ldots, N-1$. Its Fourier amplitudes are calculated via the discrete Fourier transform and their absolute values stored to get

$$|X(f_k)| = |\mathcal{F}\{x(t_n)\}| = \left| \sum_{n=0}^{N-1} x(t_n) e^{-2\pi j f_k n \Delta t} \right|, \tag{1}$$

that are estimated at the N discrete frequencies $f_k = k\Delta f$, $k = 0, \ldots, N-1$, which are in turn dictated by the sampling interval of the target time series, Δt, according to the definition of the frequency resolution, $\Delta f = 1/N\Delta t$. To remove the mean of the target time series, the Fourier amplitude corresponding to the zero-frequency f_0 is set to zero, $X_0 = 0$, and then the sequence $|X(f_k)|$ is modified by retaining only the values estimated at the frequencies $f_0, \ldots, f_{N/2-1}$, and by performing zero padding so as to get $|X(f_k)| = 0$ for $k = N/2, \ldots, N-1$. At this stage, the sequence possessing the target PSD is generated as a random-phase multisine signal from the frequency domain. This is achieved by performing an inverse Fourier transform on the previously modified sequence $|X(f_k)|$ multiplied by a set of random phases φ_k^{rnd} generated from the uniform distribution $\mathcal{U}[0, 2\pi)$; this yields

$$z_n = 2\Re\{\mathcal{F}^{-1}\{|X(f_k)| e^{j\varphi_k^{rnd}}\}\}. \tag{2}$$

The next step is to generate a second sequence y_n, $n = 0, ..., N-1$ that is consistent with the PDF of the target wind-speed dataset. This is obtained by performing an inverse CDF transform on a linear sequence of N samples taken on the interval $(0, 1)$, which is denoted by u_n and represents the cumulative distribution function of a uniformly distributed variable

$$u_n = \frac{1 + 2n}{2N}, \quad n = 0, ..., N-1. \tag{3}$$

This technique allows to obtain a sequence of N samples distributed with the desired probability density function by simply performing

$$y_n = F^{-1}(u_n), \tag{4}$$

where F^{-1} is the inverse target CDF. Such a technique is not restricted to cumulative distribution functions that possess analytical inverse. In cases where F^{-1} can not be analytically determined a nonlinear curve fitting can be performed to obtain a numerical inverse of the target CDF. Moreover, Eq. (4) allows to directly implement the empirical cumulative distribution function of the target wind-speed dataset if one wishes to retrieve a better match with the observed PDF without performing a numerical fit to the probability distribution of the wind-speed data.

To complete the initialization phase, the variance and the mean of the sequence consistent with the target PDF, y_n, are imposed to the signal z_n, and the absolute values of the Fourier amplitudes of this sequence, $Z_k = |\mathcal{F}\{z_n\}|$, are stored.

2.2. Iterative process

At this stage, to get one signal that is consistent with both target statistical descriptors, PDF and PSD, an iterative process is required that reorders the samples of the sequence y_n so as to match the autocorrelation of the sequence z_n. This rank-reordering iterative procedure consists of three steps. Firstly, the samples of the sequence y_n are reordered in a new sequence $\tilde{y}_n$ so that their smallest value is given the same position in this new sequence that the smallest value of z_n has in its own sequence, and so forth for all the N values. Secondly, the phases of the Fourier spectrum of this new sequence $\tilde{y}_n$ are computed by means of the Fourier transform of the signal, which yields

$$\varphi_k = \tan^{-1}\left(\frac{\Im\{\mathcal{F}\{\tilde{y}_n\}\}}{\Re\{\mathcal{F}\{\tilde{y}_n\}\}}\right), \tag{5}$$

where $\Im\{\cdot\}$ and $\Re\{\cdot\}$ indicate respectively the imaginary part and the real part of the transform. In the last step, a new multisine signal is generated by taking the real component of the inverse Fourier transform of the stored Fourier amplitudes Z_k multiplied by the last computed phases φ_k,

$$z_n = \Re\{\mathcal{F}^{-1}\{Z_k e^{j\varphi_k}\}\}. \tag{6}$$

This sequence is employed to perform the successive rank-reordering step of the samples of y_n in the next iteration, and this three-step process is iterated until both the probability density function and the power spectrum of z_n converge respectively to the PDF and the PSD of the target wind-speed dataset. This iterative process is in essence equivalent to the iterative scheme of the IAAFT algorithm. Its main difference resides in the order in which the steps are carried out. In the proposed scheme, the rank-reorder is performed first and the imposition of the target PSD is carried out as the last step; this order is reversed in the IAAFT algorithm. Therefore, the final signal delivered by the proposed algorithm is the one after the restoration of the target PSD, whereas in the IAAFT approach it is the one produced by the last rank-reorder.

The initial set of random phases φ_k^{rnd} ensures that each realization of the iterative process results in a different rearrangement of the samples of y_n in the final sequence z_n, and hence effectively delivers a different synthetic time series consistent with the same target statistical descriptors.

As for the extreme wind speeds, repeating the simulation with the same input data and with the same sampling defined by Eq. (4) results in a different synthetic signal characterized by the same extreme values. These extremes, however, appear at different time instants in the synthetic signal as a result of the different random seed drawn for the inital phases φ_k^{rnd}. Therefore, if one wishes to extend the simulated extreme winds the number of the simulated samples N is to be increased so as to increase the sampling in the tail region of the target CDF.

2.3. Post-processing phase

To estimate and compare the average diurnal variation of the wind speed, a bin-average technique is applied both to the synthetic wind-speed time series z_n and to the target dataset x_n.

This technique consists in dividing up the time series in a number of temporal bins corresponding to the number of wind-speed samples contained in a time period of 24 hours. The samples in each temporal bin are then averaged with respect to the number of full days present in the time series so as to get one averaged wind-speed sample for each temporal bin. Therefore this operation results in a sequence that has the same number of samples as the number of temporal bins in 24 hours, and shows the average variation of the wind speed in a day. This bin-averaged sequence can be defined as

$$\bar{V}_{i,\mathrm{day}} = \frac{1}{L}\sum_{l=1}^{L} x(t_{(l-1)M+i}), \tag{7}$$

where i is the index of the temporal bins, L is the number of full days in the considered time series $x(t_n)$, and M is the number of wind-speed samples in 24 hours. When applied to the generated sequence z_n, this technique will produce an average diurnal variation that suffers from a phase shift as a result of the random phases. To compensate for this shortcoming, a circular shift is performed to the samples of z_n until the best alignment with the ADV of the input data is reached.

This comparison is provided to assess how accurately the proposed algorithm can reproduce the average diurnal cycle of the wind-speed observed in the target dataset.

3. Dataset description

To generate synthetic wind-speed time series, the proposed algorithm requires a wind-speed dataset as input. The same wind-speed data record is then used to validate the algorithm in terms of the consistency of the synthetic time series with the target statistical descriptors and also in terms of the reproduced average diurnal variation.

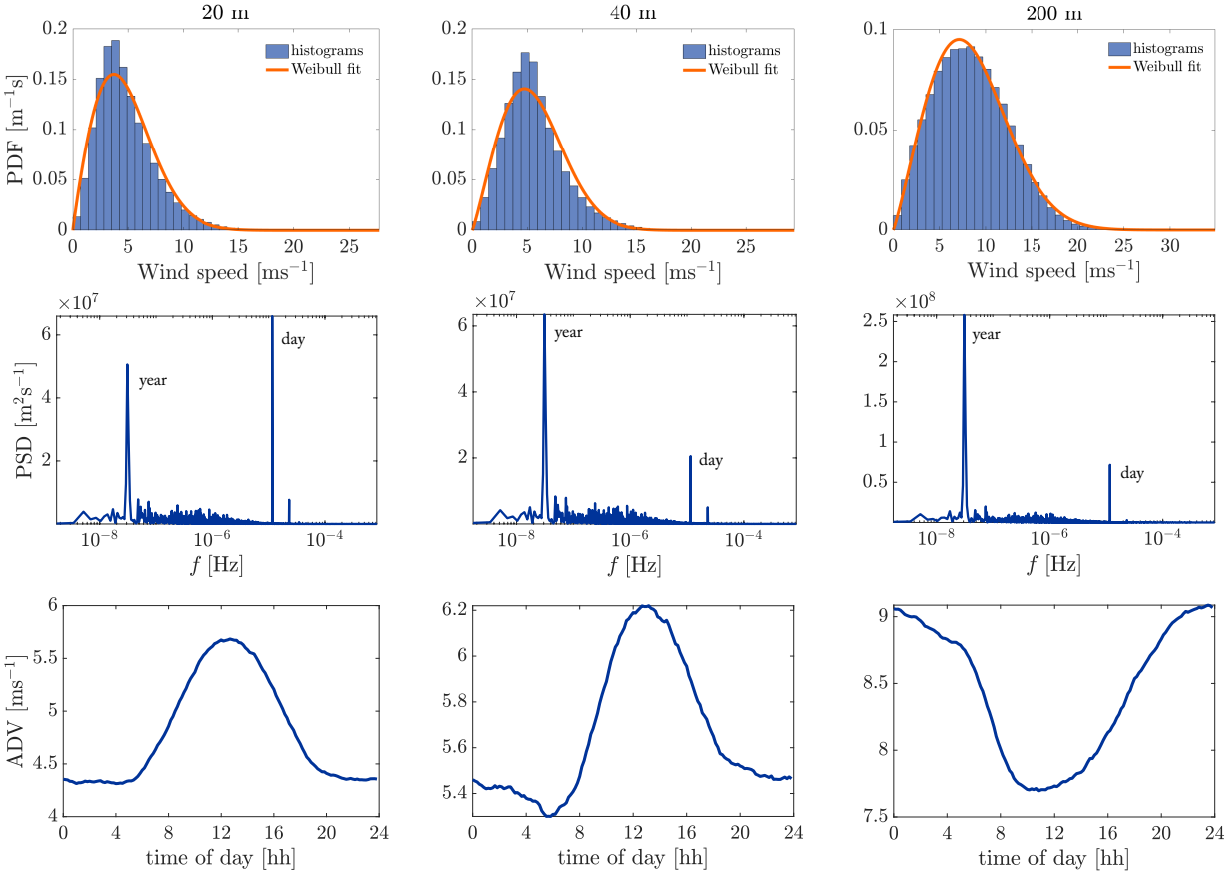

Figure 2. Data analysis results of the three 10-minute averaged wind speed time series obtained from the database of the Cabauw met mast at 20 m (left column), 40 m (center column), and 200 m (right column). The first row shows the histogram PDF along with its respective Weibull fit; second row is the PSD of the time series; third row shows the average diurnal variation of the wind speed.

These tasks are performed using the wind-speed data recorded from the 213-m tall meteorological mast at Cabauw in the western Netherlands. This met mast is owned by the Royal Netherlands Meteorological Institute (KNMI), and its wind-speed observations are collected in a database that is maintained and made available in collaboration with the Cabauw Experimental Site for Atmospheric Research (Cesar) [13]. From this database, long-term time series of 10-minute averaged wind speeds recorded from May 2000 to September 2018 are imported for three different heights, namely 20, 40, and 200 m. These 18-year long time series are processed with a zero-order hold signal reconstruction technique to replace missing and invalid data so as to obtain gap-free data series. This operation results in three time series consisting of $N = 968688$ samples of 10-minute averaged wind speed.

After the data cleaning operation, the three time series are analysed in terms of their PDF, their PSD, and their average diurnal variation (ADV) according to the definition of Eq. (7). The whole 18-year long series are used in this analysis. Fig. 2 shows the results of this analysis for the three selected heights, 20, 40, and 200 m, respectively in three different columns.

A Weibull fit to the histogram PDF is performed for each time series (first row of Fig. 2). Larger deviations of the numerical PDF can be observed around the mode of the empirical distribution both at 20 and 40 m, i.e. in the near-surface region of the atmospheric boundary layer. Also, the Weibull fit results in moderate errors when approximating the tail region of the distribution of the three time series (not shown here). Despite those deviations, the qualitative agreement with the distribution of the measured data shows that for this site the long-term mean wind speed fluctuation process can be reasonably simulated by a Weibull distribution. The PSD analysis of the three time series (second row of Fig. 2) reveals that the wind-speed observations at this site are characterised by strong annual and diurnal frequency components indicated by the highest energy peaks at $f \approx 3.17 \times 10^{-8}$ Hz, and $f \approx 1.16 \times 10^{-5}$ Hz, respectively. In addition, a clear diurnal cycle of the mean wind speed is detected in the three time series when the average diurnal variation is calculated using the technique described in Sec. 2 (third row of Fig. 2). These diurnal cycles are mainly driven by the influence of the near-surface thermal structure of the atmospheric boundary layer [14], and are reversed at 200 m due to the change of this structure with height [15].

The presence of these pronounced diurnal cycles that are detectable when investigating the daily variation of the mean wind speed makes the selected time series ideal candidates to test the proposed methodology.

4. Application

The algorithm is tested for the generation of synthetic wind-speed time series with a sampling interval or temporal resolution of 10 minutes when the 10-minute averaged wind-speed time series from Cabauw at the three selected heights are given as input. As in Section 3, the missing and invalid data of the input time series are replaced with a zero-order hold processing technique so as to obtain 18-year long gap-free time series.

For this test, multiple simulations are performed for each input time series so as to generate a large number of synthetic wind-speed time series that differ in their temporal evolution as a result of the stochastic rearrangement produced by the random phases φ_k^{rnd}. Given the deviations shown by a Weibull fit, the empirical CDF is instead used to generate the synthetic wind-speed samples (Eq. (4)). This allows to improve considerably the agreement with the PDF of the target dataset. The sampling of the CDF is always performed with the same number of samples N that are present in the input time series. Therefore, each simulation results in a 18-year long synthetic wind-speed time series.

Figures 3, 4, and 5 provide a qualitative assessment of the agreement with the statistical descriptors and the average diurnal cycle obtained in three simulations each performed for one selected height. Same temporal resolution of the input data is recovered in each of the synthetic

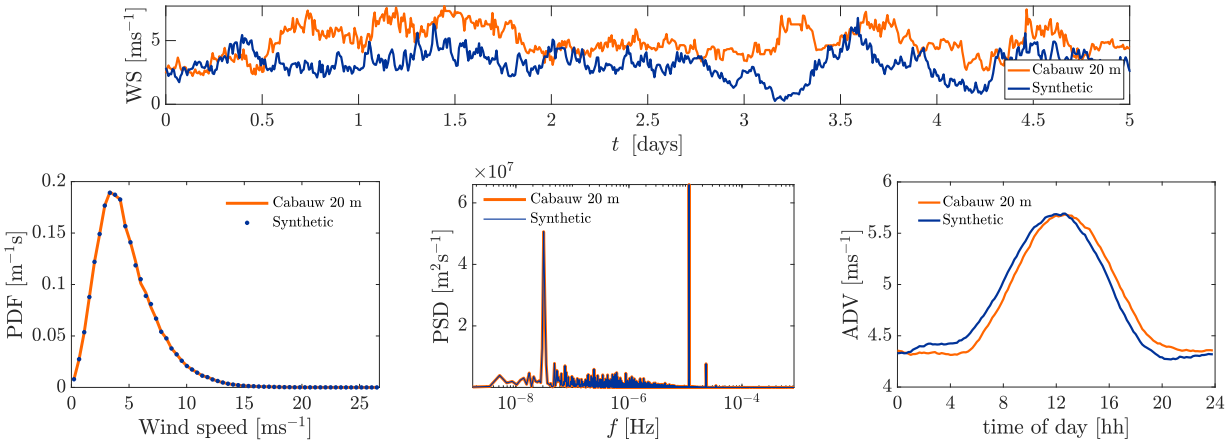

Figure 3. Simulations results for the implementation of the proposed algorithm with the 20-m input time series from Cabauw: 5-day segments of the input and the generated time series (*top*); PDF agreement between synthetic and input data (*bottom-left*); PSD agreement between synthetic and input data (*bottom-centre*); average diurnal variation comparison between synthetic and input data (*bottom-right*).

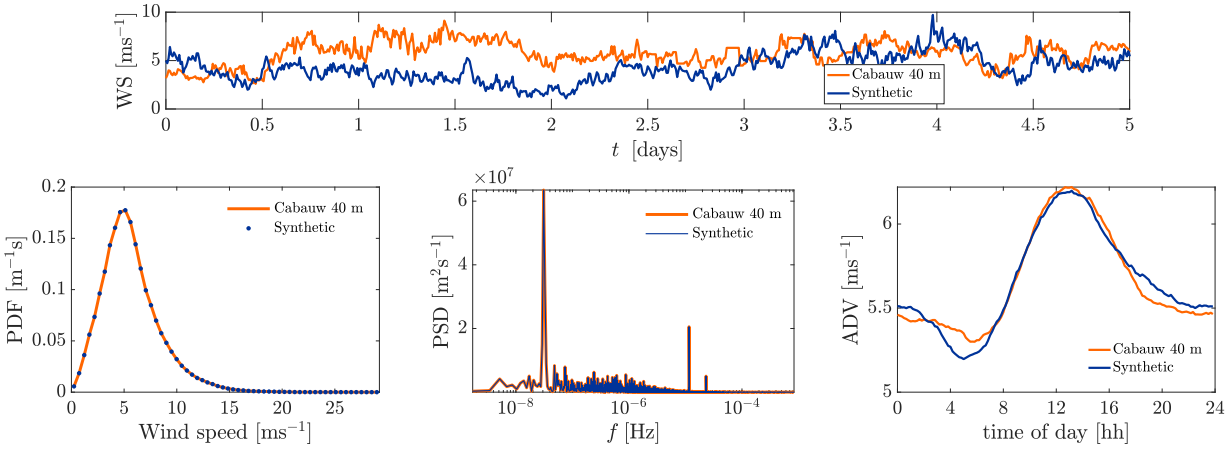

Figure 4. Simulations results for the implementation of the proposed algorithm with the 40-m input time series from Cabauw: 5-day segments of the input and the generated time series (*top*); PDF agreement between synthetic and input data (*bottom-left*); PSD agreement between synthetic and input data (*bottom-centre*); average diurnal variation comparison between synthetic and input data (*bottom-right*).

time series, as shown by the variation of the simulated wind speed in the first five days at the top of each figure. The visual comparison might suggest that the simulated wind speeds are consistently lower (20 and 40 m) or higher (200 m) than the measured wind speed. However, this only coincidentally occurs in the selected time window and the mean and variance of the target dataset are always reproduced correctly. In addition, the highly accurate match with the PDF of the respective input time series shows that the synthetic wind speeds properly model the long-term wind-speed fluctuation process of the observed data in all the three simulations (bottom-left plot). Synthetic time series also show a very accurate match with the spectral content of the target wind data in each of the three simulations. The distribution of kinetic energy of the input data is perfectly reconstructed throughout the whole range of frequencies

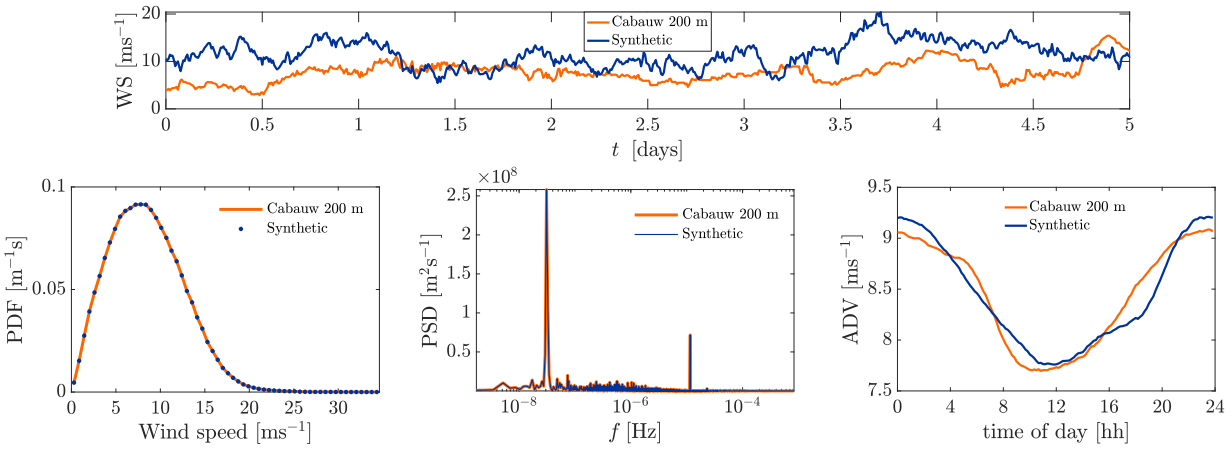

Figure 5. Simulations results for the implementation of the proposed algorithm with the 200-m input time series from Cabauw: 5-day segments of the input and the generated time series (*top*); PDF agreement between synthetic and input data (*bottom-left*); PSD agreement between synthetic and input data (*bottom-centre*); average diurnal variation comparison between synthetic and input data (*bottom-right*).

with a particular emphasis on the annual and diurnal frequency components (bottom-centre plot). Moreover, simulation results show that the generated wind speeds exhibit an average

Table 1. Simulations results given in terms of the maximum relative error of the target functions of the synthetically generated time series. Each row corresponds to a different simulation performed with the indicated input wind-speed data.

Input data Cabauw	PDF $max(\varepsilon_r)$ [%]	PSD $max(\varepsilon_r)$ [%]	ADV $max(\varepsilon_r)$ [%]
20 m			
	4.43	7.26e−8	8.36
	5.07	7.24e−8	2.50
	3.33	7.21e−8	1.98
	3.33	7.23e−8	3.05
	3.33	7.21e−8	1.59
40 m			
	0.36	3.04e−6	2.69
	0.36	3.04e−6	1.47
	0.36	3.04e−6	2.06
	0.36	3.04e−6	8.74
	0.36	3.04e−6	10.56
200 m			
	0.15	6.21e−7	4.64
	0.11	6.21e−7	3.77
	0.15	6.20e−7	3.47
	0.12	6.21e−7	3.47
	0.11	6.20e−7	3.31

diurnal variation that is in very good agreement with the one observed in the input data in all the three cases (bottom-right plot). In particular, the target amplitude of the average diurnal

cycle is always correctly reproduced in the synthetic time series. Because the PDF and the PSD are invariant under a time shift of the original time signal, it is impossible to retrieve the actual phase of the diurnal cycle from the PDF and PSD only. For that reason the synthetic cycle is aligned with the cycle of the target data by applying a proper time shift before comparing both signals.

These results can be considered representative of the multiple simulations performed for each of the three input time series as the implementation of the algorithm to the same wind speed data always results in the same level of accuracy in terms of the matching of the target statistical descriptors and reproduces accurately the average diurnal cycle of the input data. That is summarised in table 1, where the outcomes are listed for five different simulations run for each set of measurements. Each row corresponds to one simulation and the accuracy in reproducing the target functions is indicated in terms of the maximum relative errors reached by the PDF, PSD, and the ADV of the synthetic wind data with respect to the input data.

The fluctuation of the maximum error in the average diurnal variation across the different simulations is a direct consequence of the stochastic component of the algorithm. Because of the initial random phases, the rank-reordering step results in a different rearrangement of the wind-speed samples that may in turn lead to a slightly larger deviation in the amplitude of their average diurnal cycle. This is consistent with the ambiguity that the time-frequency signal analysis inevitably entails, since signals with different dynamics i.e. instantaneous amplitude and frequency can share the same power spectrum [16]. However, this maximum deviation is observed to remain constantly limited throughout the simulations, and an average diurnal cycle of approximately the same amplitude of the one present in the input data is always retrieved in the synthetic data.

5. Conclusion

A computationally fast algorithm for synthetic wind speed data has been proposed and implemented for the generation of synthetic time series with a temporal resolution of 10 minutes starting from three wind-speed time series recorded at three different heights at Cabauw in the Netherlands. The main aim of this work has been to investigate the ability of this algorithm in reproducing the average diurnal variation of the wind speed detectable in the target time series. In particular, this methodology has been chosen for its ability to match with high precision both the long-term probability distribution of the observed wind-speed process and the distribution of kinetic energy across its entire range of temporal scales, i.e. its PDF and PSD.

Multiple simulations have been performed to generate 18-year long synthetic time series of 10-minute wind speeds from each of the three measured time series that have displayed pronounced diurnal components and distinct average diurnal patterns. Simulation results have shown that the proposed algorithm can generate realistic wind speed scenarios that conform to the statistical descriptors of the target wind speed data and always show an accurate reproduction of the average diurnal cycle observed in the input wind speeds. Moreover, the non-stationary dynamics of the wind speed at the diurnal temporal scale has always been properly simulated.

Therefore, the proposed algorithm is particularly suitable for reliable wind power modelling in grid integration applications, being able to realistically simulate the variability of the wind speed at the diurnal scale. In addition, it introduces two main advantages with respect to existing methods for synthetic wind speed generation. First, its statistical approach with a stochastic component in the rearrangement of the samples allows to generate different wind speed scenarios that all possess the same statistical properties of the target dataset. Secondly, the generation is not restricted to a specific time resolution and can therefore be applied to wind-speed dataset with any desired sampling interval or averaging time.

References

[1] Draxl C, Clifton A, Hodge B M and McCaa J 2015 *Applied Energy* **151** 355–366
[2] Suomalainen K, Silva C A, Ferrão P and Connors S 2012 *Energy* **37** 41–50
[3] Carapellucci R and Giordano L 2013 *Applied Energy* **107** 364–376
[4] Dai A and Deser C 1999 *Journal of Geophysical Research: Atmospheres* **104** 31109–31125
[5] Suomalainen K, Silva C, Ferrão P and Connors S 2013 *Applied Energy* **101** 533–540
[6] Hill D C, McMillan D, Bell K R W and Infield D 2012 *IEEE Transactions on Sustainable Energy* **3** 134–141
[7] Soman S S, Zareipour H, Malik O and Mandal P 2010 A review of wind power and wind speed forecasting methods with different time horizons *North American Power Symposium 2010* pp 1–8
[8] Lei M, Shiyan L, Chuanwen J, Hongling L and Yan Z 2009 *Renewable and Sustainable Energy Reviews* **13** 915–920
[9] Chen J and Rabiti C 2017 *Energy* **120** 507–517
[10] Schoukens J and Dobrowiecki T 1998 Design of broadband excitation signals with a user imposed power spectrum and amplitude distribution *IMTC/98 Conference Proceedings. IEEE Instrumentation and Measurement Technology Conference. Where Instrumentation is Going (Cat. No.98CH36222)* vol 2 pp 1002–1005
[11] Schreiber T and Schmitz A 1996 *Physical Review Letters* **77** 635–638
[12] Schreiber T and Schmitz A 2000 *Physica D: Nonlinear Phenomena* **142** 346–382
[13] Cesar Observatory URL `http://www.cesar-observatory.nl/index.php`
[14] Barthelmie R J, Grisogono B and Pryor S C 1996 *Journal of Geophysical Research: Atmospheres* **101** 21327–21337
[15] He Y, Monahan A H and McFarlane N A 2013 *Geophysical Research Letters* **40** 3308–3314
[16] Cohen L 1995 *Time-frequency analysis* Prentice Hall signal processing series (Englewood Cliffs, N.J: Prentice Hall PTR)