A new Markov-chain-related statistical approach for modelling synthetic wind power time series

2015 New J. Phys. 17 055001

(http://iopscience.iop.org/1367-2630/17/5/055001)

IOP Publishing New J. Phys. 17 (2015) 055001 doi:10.1088/1367-2630/17/5/055001

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics
Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

CrossMark

OPEN ACCESS

RECEIVED
1 September 2014

REVISED
9 March 2015

ACCEPTED FOR PUBLICATION
30 March 2015

PUBLISHED
6 May 2015

PAPER

A new Markov-chain-related statistical approach for modelling synthetic wind power time series

T Pesch[1], S Schröders[2], H J Allelein[2] and J F Hake[1]

[1] Forschungszentrum Jülich, Institute of Energy and Climate Research—Systems Analysis and Technology Evaluation, D-52425 Jülich, Germany

[2] RWTH Aachen University, Institute for Reactor Safety and Reactor Technology, Kackertstraße 9, D-52072 Aachen, Germany

E-mail: t.pesch@fz-juelich.de

Keywords: renewable energy, Markov chain, stochastic modelling, wind power, energy system model, Markov process, synthetic time series

Abstract

The integration of rising shares of volatile wind power in the generation mix is a major challenge for the future energy system. To address the uncertainties involved in wind power generation, models analysing and simulating the stochastic nature of this energy source are becoming increasingly important. One statistical approach that has been frequently used in the literature is the Markov chain approach. Recently, the method was identified as being of limited use for generating wind time series with time steps shorter than 15–40 min as it is not capable of reproducing the autocorrelation characteristics accurately. This paper presents a new Markov-chain-related statistical approach that is capable of solving this problem by introducing a variable second lag. Furthermore, additional features are presented that allow for the further adjustment of the generated synthetic time series. The influences of the model parameter settings are examined by meaningful parameter variations. The suitability of the approach is demonstrated by an application analysis with the example of the wind feed-in in Germany. It shows that—in contrast to conventional Markov chain approaches—the generated synthetic time series do not systematically underestimate the required storage capacity to balance wind power fluctuation.

1. Introduction

The integration of rising shares of fluctuating renewables, such as wind and photovoltaic, is a major challenge for the future energy system (Turner 1999, Marris 2008). In particular, high propotions of wind in the generation mix lead to growing uncertainties in the system since the actual feed-in of wind power can change drastically within a very short period of time. It is therefore increasingly important to address these uncertainties in energy system models. One approach for undertaking this task is the stochastic modelling of the feed-in from wind converters.

This paper presents a new Markov-chain-related approach which is suitable for generating synthetic time series for wind power feed-in that have statistical characteristics comparable to those of real ex post time series. The synthetic time series can be used to analyse a multitude of different situations and scenarios for the feed-in from wind power–which may also include situations that have not yet occurred in historical data. This is, for example, useful to examine the robustness of an electricity system derived by energy system models such as that described in (Pesch *et al* 2014). Synthetic time series can also be used as input for further approaches to address uncertainty in energy system models, e.g. scenario trees or Monte–Carlo simulations (Negra *et al* 2008).

We focus on intermediate time scales and analyse time series with a sampling time of 15 min. This time scale is particularly important for the operation of the European electricity system as it is the shortest trading period on the European electricity market (EPEX Spot SE 2014). Extremely fast fluctuations also exist on time scales as short as seconds due to the turbulent character of wind energy. These effects are not the subject of this paper but are discussed in detail in (Milan *et al* 2013).

In contrast to other publications, our purpose is not to model the wind speed or wind power feed-in related to single wind turbines or wind parks. Instead, we consider the overall feed-in of all installed onshore wind turbines in a whole market region. For the application analysis, the model was trained with original ex post data of the wind power feed-in in Germany.

The paper is organized as follows: section 2 first gives an overview of different possible approaches for the modelling of wind time series. Afterwards, the basic principles of Markov chains are explained. For a deeper understanding of the characteristics of wind time series, their deterministic and stochastic components are discussed in the following. A basic approach for removing deterministic components is presented, which aims at obtaining a stationary time series that can be used for Markov chain models. Exisiting Markov chain approaches are then analysed with regard to their suitability for reproducing the stochastic characteristics of wind time series. Special attention is paid to why conventional Markov chains are not able to reflect these characteristics sufficiently.

In section 3, the procedure for modelling the new Markov-chain-related approach is explained. Of particular interest is the description of the main difference in our approach compared to conventional second-order Markov chains. Furthermore, two new features are introduced: a healing mechanism to avoid dead ends and a running average filter to adjust the short-term characteristics of the time series.

In section 4, the effects on the synthetic time series of varying the different parameters of the presented approach are shown. Subsequently, a suitable variant for wind power feed-in time series for Germany is presented. The quality of the approach is determined by comparing the original and the synthetic time series in terms of statistical characteristics such as autocorrelation, probability distribution and the distribution of increments on different time scales. Finally, an application analysis shows that–in contrast to conventional Markov chains–the generated synthetic time series do not systematically underestimate the required storage capacity to balance wind power fluctuation.

The main conclusions drawn from the analysis are finally summarized in section 5.

2. Approaches for the modelling of wind speed and wind power time series

This section gives an overview of the different approaches for modelling wind speed and wind power time series. Section 2.1 presents the general classification into persistence models, physical models and statistical models. As the approach pursued in this paper is based on a Markov chain, the basic principles of these chains are explained in section 2.2. An understanding of the stochastic properties of the original wind time series is necessary to develop a model which generates realistic wind power time series. For that reason, section 2.3 analyses a given wind power time series in terms of deterministic and stochastic components and characteristic trends and fluctuations. Furthermore, a basic approach for removing deterministic components of the time series is presented. Section 2.4 gives an overview of existing Markov chain approaches for modelling wind speed and wind power time series, showing that these approaches have difficulty in reproducing the stochastic behaviour of wind correctly.

2.1. Overview of the different methods

The rising share of volatile wind power in the present generation mix has led to increased interest in methods for the modelling and especially the forecasting of wind. As a result, a variety of different approaches has been developed and applied to wind time series. Overviews of the different approaches with the focus on forecasting techniques can be found in Lei *et al* (2009), Soman *et al* (2010), Wang *et al* (2011), Foley *et al* (2012).

The suitability of a method is closely linked to the purpose of the model. This especially accounts for the desired time horizon in the case of forecasting wind. Although time scale classifications are rather vague in this context and the boundaries between the respective classes often overlap, a distinction can be made between very short-term (seconds to 30 min), short-term (30 min to 6 h), medium-term (6 h to 1 d) and long-term (1 d to 1 week or more) forecasting techniques (Soman *et al* 2010). In the case of generating synthetic time series, the method should account for the proper description of both short-term and long-term characteristics.

Besides some new techniques and combinations of model types (hybrid structures), most of the common methods can be assigned to one of the following classes (Soman *et al* 2010):

- *Persistence models*—also called naive predictors—assume that the current state is maintained. This method is obviously only suitable for very short-term forecasts and may serve as a benchmark for other forecast methods.
- *Physical methods* such as numerical weather predictors are based on meteorological data and a detailed description of the atmospheric physics. These methods are especially employed for the long-term forecasting of wind.

- *Statistical methods* only need the historical data of a process to estimate the future development. The different methods cover a wide range from short-term to long-term forecasting. The Markov chain approach can be assigned to this category.

2.2. Basics of the Markov chain approach

A Markov chain models a stochastic process, in which a state changes between discrete time steps. In a Markov chain approach, it is assumed that the probability of a certain state depends on one or more past states. It is a probabilistic forecasting method, as it does not just estimate a point forecast but also provides estimates of the probability distribution associated with it (Carpinone *et al* 2010). The method only requires empirical ex post data so that a restrictive assumption on the shape of a wind power probability distribution is not necessary (Carpinone *et al* 2010). This approach is applicable for time series with any probability distribution. A necessary requirement for the ex post time series is weak stationarity. This means that the probability of making a transition from one given state to another is time-independent (Castro Sayas and Allan 1996).

For the application of the Markov chain, the amplitude range must be broken down into a defined number of discrete states. The number of equidistant amplitude intervals n for the state variables $\{s_1, s_2, \ldots, s_n\}$ is a calibration parameter of the Markov chain (Carpinone *et al* 2010, Carapellucci and Giordano 2013).

Another important parameter is the order of the Markov chain. It gives the number of past states that influence the probability of the present state (Shamshad *et al* 2005). In a first-order Markov chain, the present state only depends on the previous state.

Let $X(t)$ be a stochastic process with the discrete state space $S = \{1, 2,.., n\}$ and the discrete time steps t. The transition probability from a state i to a state j is defined as:

$$p_{ij} = p\left\{X(t) = s_j \,|X(t-1) = s_i\right\} \quad \forall i, j \in \{1, \ldots, n\}. \tag{1}$$

These transition probabilities form the elements of the transition matrix P.

For n states, the first-order transition matrix is an $n \times n$ matrix,

$$P = \begin{pmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \cdots & p_{nn} \end{pmatrix}. \tag{2}$$

The problem with Markov chains of higher order is the increasing complexity, as a K-order Markov chain needs a n^{K+1} transition matrix (Brokish and Kirtley 2009). The advantage of higher-order Markov chains is better forecasting performance, as they reduce the prediction error (Carpinone *et al* 2010). Therefore, a compromise between prediction accuracy and computing time (and particularly memory requirements) needs to be found.

The basic concept of the modelling approach presented in this paper is related to a second-order Markov chain. A conventional second-order Markov chain uses the two previous states j (first lag) and k (second lag) to forecast the present state i:

$$\begin{aligned} p_{ijk} = p\Big\{X(t) = s_i \Big| X(t-1) \\ = s_j \wedge X(t-2) = s_k\Big\} \quad \forall i, j, k \in \{1, \ldots, n\}. \end{aligned} \tag{3}$$

The resulting $n \times n \times n$ transition matrix P is shown below:

$$P = \begin{pmatrix} p_{111} & p_{112} & \cdots & p_{11n} \\ p_{121} & p_{122} & \cdots & p_{12n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1n1} & p_{1n2} & \cdots & p_{1nn} \\ p_{211} & p_{212} & \cdots & p_{21n} \\ p_{221} & p_{222} & \cdots & p_{22n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{nn1} & p_{nn2} & \cdots & p_{nnn} \end{pmatrix}. \tag{4}$$

Ex post experimental time series are used to determine the matrix parameters by counting the number of state transitions m to state i under the condition that state j and state k took place and relating them to the sum of all state transitions to state i:

$$p_{ijk} = \frac{m_{ijk}}{\sum_{j=1}^{n} \sum_{k=1}^{n} m_{ijk}} \quad \forall i, j, k \in \{1, \ldots, n\}. \tag{5}$$

The transition probabilities satisfy the following properties:

$$0 \leqq p_{ijk} \leqq 1 \quad \forall i, j, k \in \{1, \ldots, n\}, \tag{6}$$

$$\sum_{j=1}^{n} \sum_{k=1}^{n} p_{ijk} = 1 \quad \forall i \in \{1, \ldots, n\}. \tag{7}$$

The current state of the process is represented by the corresponding row of the matrix, while each column corresponds to one of the n possible states at the next time step (Carpinone *et al* 2010).

2.3. Deterministic and stochastic characteristics of wind power time series

In order to evaluate the suitability of existing Markov chain approaches for reproducing the stochastic nature of wind, it is necessary to analyse the stochastic characteristics of wind time series. For this purpose, the seasonal and daily periodicities of wind power time series are discussed in section 2.3.1. In section 2.3.2, we present an approach for removing these periodicities with the objective of obtaining a stationary time series. After testing the resulting time series for stationarity in section 2.3.3, its characteristic trends and fluctuations are analysed in section 2.3.4.

2.3.1. Deterministic trends and periodicies in wind power time series

As a data basis for the wind power time series, we use the aggregated wind feed-in time series of all wind turbines connected to the German electricity grid. The four German transmission system operators provide data for their transmission zone on a 15 min basis (TenneT TSO GmbH 2014, Amprion GmbH 2014, 50Hertz Transmission GmbH 2014, TransnetBW GmbH 2014). The observation period covers 59 months (August 2009–May 2014) and comprises 166,464 time steps. The considered wind power time series has a positive trend, as the installed capacity of wind converters increased steadily during the observation period.

Furthermore, the time series is characterized by daily as well as annual periodicities. Generally, wind speed and consequently also wind power is higher in winter than in summer. The daily periodicities differ during the year, resulting in wind power peaks in the afternoon during winter and peaks in the evening during summer.

2.3.2. Conditioning of the original wind power time series

The positive trend and also the seasonal periodicities represent deterministic components of the time series. These deterministic components must be removed to be able to analyse the stochastic component of the wind power time series. Since periodicities and trends cause instationarity of the data, this deseasonalization step is also necessary for the application of a Markov chain.

In order to eliminate the trend, the feed-in time series was divided by the respective actual installed capacity (Fraunhofer IWES 2014) to derive the specific feed-in per installed capacity $\mathrm{kW}_{\mathrm{installed}}$ (figure 2).

The next step in obtaining a deseasonalized time series is to identify and eliminate those parts of the time series caused by periodicities. For this purpose, a basic approach was chosen based on arithmetic means allocated to a certain month and a certain time. First, the average over all specific feed-in values was determined. Second, the mean value was identified for each quarter-hour in a month that was subsequently divided by the overall average value. In this way, a seasonality factor was created, which–due to seasonal or daily fluctuations—is higher than 1 in high feed-in times and lower than 1 in low feed-in times. In this way, $12 \times 96 = 1152$ different factors for the twelve months and 96 quarter-hours each day were obtained. Consequently, this approach implies that every day of a month has the same properties. This simplification is especially necessary to obtain a sufficient number of values for average determination. The seasonality factors in figure 1 reflect the above-mentioned seasonal and daily periodicities. The values of the specific feed-in time series are divided by the corresponding factor to obtain a time series that only contains the non-deterministic residuals .

2.3.3. Stationarity of the conditioned wind power time series

The extent to which the conditioned time series fulfils the requirement of weak stationarity is examined by conducting a stationarity test. For this purpose, the moving average and the moving standard deviation of the original time series are compared to those of the deseasonalized time series (figure 3). The window size of the moving average filter was set to the length of a whole year ($4 \times 8760 = 35040$ values). It can be observed that both the moving average and the moving standard deviation of the original time series are much more time-dependent than those of the conditioned time series. In particular, the strong positive trend at the end of the time series can be eliminated by data conditioning.

Nevertheless, the obtained deseasonalized time series does not completely fulfil the requirement of weak stationarity since the moving average curves are not entirely levelled. This is due to the limitation of the available data basis and the occurrence of strong and weak wind years. Therefore, we refer to the definition of wide-sense stationarity as in Thomann and Barfield (1988). Castro Sayas and Allan (1996) point out that a wind time series

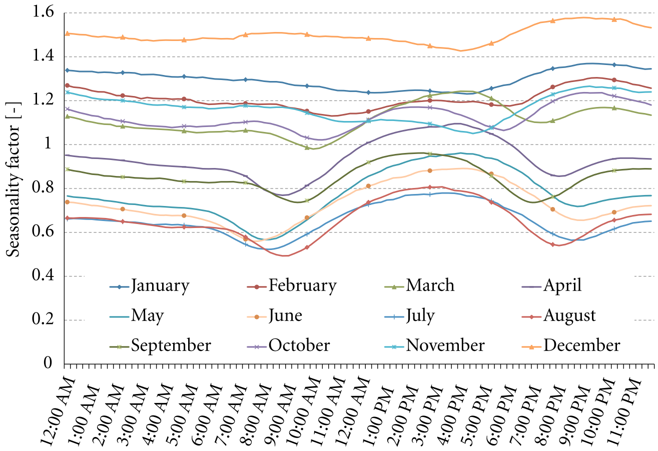

Figure 1. Seasonality factor plotted against month and time of day. While high wind speeds were observed during summer afternoons and winter nights, the wind speed was lower during summer mornings/evenings and winter middays. Generally, the wind speed was higher in winter than in summer.

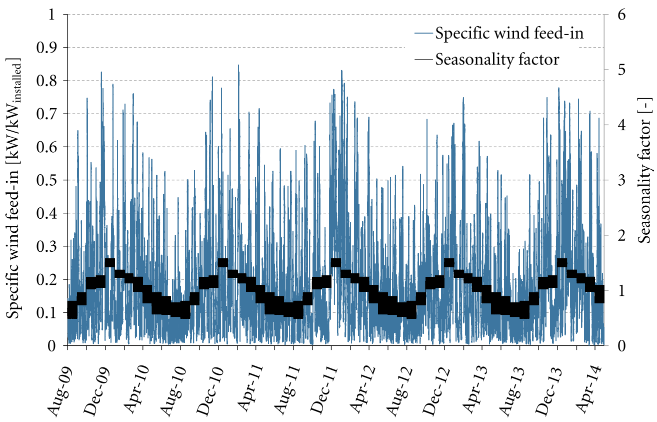

Figure 2. Original time series and seasonality factor. The values of the specific feed-in time series must be divided by the corresponding seasonality factor to obtain a time series that only contains the non-deterministic residuals.

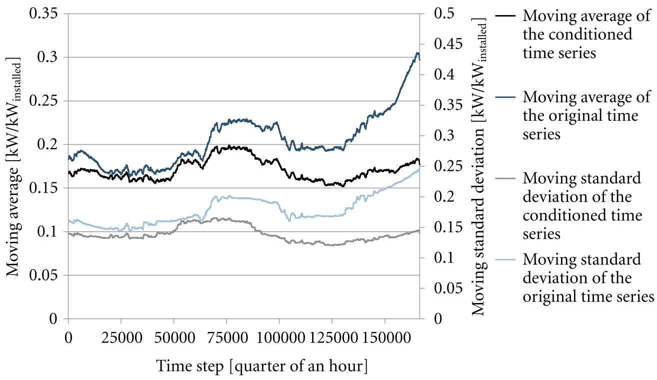

Figure 3. Moving average and moving standard deviation of the original and the deseasonalized time series. Both the moving average as well as the moving standard deviation of the deseasonalized time series are much less time-dependent than those of the original time series. Since the positive trends were eliminated, we assume wide-sense stationarity of the conditioned time series.

can be treated as wide-sense stationary as long as there is no trend in the data. Peiyuan *et al* (2009) and Jones and Lorenz (1986) also assume wide-sense stationarity of their considered wind time series which they use for a Markov chain model. Since the moving average and the moving standard deviation of the conditioned time series show no trend, we assume that the criteria for wide-sense stationarity are fulfilled and that we can use the time series for a Markov-chain-related approach.

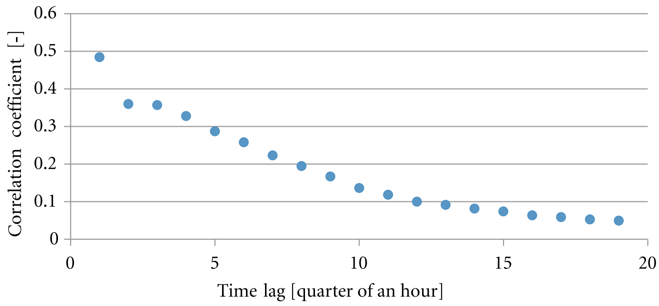

Figure 4. Correlation coefficients between the original time series of increments and those shifted by τ time steps. Since correlations remain significant for several previous time steps, it is not sufficient to merely consider the one or two previous values for anticipating trends correctly.

2.3.4. Characteristic trends and fluctuations of wind power time series

For the modelling of the stochastic behaviour of wind, a deeper understanding of the characteristic trends, fluctuations and time scales of the time series is necessary. Of special interest is the question of the extent to which the further development of a wind time series depends on the values at previous time steps and especially the changes between previous time steps. By this means, for example, positive trends in the time series indicate that a further increase from one time step to the next is more likely when there were increases the time steps before.

To analyse this assumption, a time series based on the incremental changes between two adjacent time steps was constructed. Subsequently, this time series of increments was delayed by τ time steps so that the correlation coefficients between the original time series of increments and the delayed time series of increments can be calculated. These correlation coefficients are given in figure 4 as a function of the time lag τ.

The figure shows that the correlation coefficients of the shifted time series are high at low time lags and decline with an increasing time lag. Nevertheless, the correlation coefficients of the further shifted time series of increments still have significant correlation coefficients. In the case of the feed-in time series introduced, the time series of increments shifted by up to $\tau = 12$ (equivalent to three hours) have correlation coefficients above 0.1. Further shifted time series of increments show a correlation coefficient below 0.1. It therefore can be assumed that they no longer have any significant influence on the present increment.

The results show that trends can be identified in the time series that are maintained for several quarter-hours. We interpret this effect as upcoming weather fronts and typical cycles of weather regions. This finding serves as the first evidence that on this time scale it is not sufficient to merely consider the one or two previous values for anticipating trends correctly. However, the wind power time series of other countries most likely have different characteristic time scales due to different geological conditions and different weather conditions.

2.4. Existing Markov chain models for wind speed and wind power time series

Markov chain models have been frequently used for the generation of synthetic wind speed and wind power time series (Carpinone *et al* 2010). The early approaches mostly differed in the order of the Markov chain, the transition matrix sizes, and the number and length of time steps. Jones and Lorenz (1986), Sahin and Sen (2001), Nfaoui *et al* (2004) and Hocaoglu *et al* (2008) used first-order Markov chains with different transition matrix sizes (Brokish and Kirtley 2009). Kaminsky *et al* (1991), Shamshad *et al* (2005) and Carpinone *et al* (2010) compared the use of first- and second-order Markov chains for the synthetic generation of wind speed time series, showing that better results are obtained with a second-order Markov chain (Brokish and Kirtley 2009).

Extended Markov chain approaches have been pursued by Castro Sayas and Allan (1996) and Negra *et al* (2008) using a birth-and-death Markov chain. Papaefthymiou and Klockl (2008) and Tong *et al* (2012) proposed a Markov chain Monte–Carlo method for the generation of synthetic wind power time series. Peiyuan *et al* (2009) developed a Bayesian Markov model to generate wind power data using Bayesian inference to model the uncertainty of the transition matrix.

A review of existing Markov chain models shows that better results are obtained using a second-order Markov chain rather than a first-order Markov chain. The analysis also reveals the superiority of models that directly generate wind power time series compared to models that estimate wind power by generating wind speed time series. The reason is that the multitude of parameters which influence the feed-in of a wind turbine, such as the terrain characteristics, cannot be neglected. It is advantageous that these effects are already included in the historical wind power feed-in time series.

The existing approaches have in common that the autocorrelation properties and the different time scales of wind time series are not given sufficient consideration. Brokish and Kirtley (2009) observed that 'autocorrelation plots of wind speeds generated by Markov chains are often inaccurate'. They also show that time series generated by Markov chains underestimate the required storage capacity to balance the wind power fluctuations, as the short time step behaviour cannot be modelled properly. They therefore suggest that Markov chains should not be used for time steps shorter than 15–40 min.

The reason for this underestimation of required storage capacity is the lack of modelling of the trend character, which results in time series that alternate too frequently between rising and falling. As discussed in section 2.3.4, a trend in wind power time series can last for a certain time. Consequently, the probability that wind power rises from one state to the next can be higher when there was previously a positive trend, e.g. during the last three hours. Since only the previous state influences the next state in first-order Markov chains, the previous trend cannot be considered. This is the reason why second-order Markov chains already produce better results. Nevertheless, on time scales shorter than 15–40 min it is still not sufficient to take merely the two previous time steps into account.

3. Modelling of the new Markov-chain-related approach

In this chapter, the procedure for modelling the new Markov-chain-related approach will be explained. The two main steps, determination of the transition matrix (section 3.1) and generation of the synthetic time series (section 3.2), are outlined. Furthermore, two new features are introduced: the healing mechanism to avoid dead ends and the running average filter to smooth the time series.

3.1. Determination of the transition matrix

To ensure wide-sense stationarity of the input time series, deterministic trends and seasonalities need to be removed. Therefore, the positive trend caused by a growing installed capacity and also the daily and seasonal periodicities were removed as explained in section 2.3.2.

The approach pursued in this paper is related to a second-order Markov chain. As shown in section 2.3.4, it would be preferable to consider approximately the last 12 time steps in the case of the wind time series introduced since significant correlations are given. A higher order Markov chain does not bring further benefits because the influence on the present state of the values more than 12 time steps ago can be neglected. However, a Markov chain with such a high order is hard to model, because of the high order transition matrix and the resulting high memory requirements.

Consequently, we also only consider two previous values that influence the present value to keep the memory requirements at an acceptable level. The difference to conventional second-order Markov chains is a variable second lag, which is normally set at the state before the first lag. In this way, a more distant second state can be used to determine the present state. Consequently, the trend behaviour of the wind power time series can be reproduced without overloading the computing capacities.

As with conventional Markov chains, the values of the conditioned specific wind power feed-in time series need to be discretized and assigned to the corresponding state. As explained in section 2.2, the transition matrix is determined by counting the number of state transitions to state i under the condition that state j and state k took place at the corresponding points in time.

State j belongs to the first lag and is therefore always set to the state before the present state i. In contrast to conventional second-order Markov chains, state k is set to a variable previous step L and is not fixed to the state two steps before state i:

$$p_{ijk} = p\left\{X(t) = s_i \,|X(t-1) = s_j \wedge X(t-L) = s_k\right\} \quad \forall i, j, k \in \{1, \ldots, n\}. \tag{8}$$

To derive the transition probability matrix, the specific number of state transitions needs to be divided by the sum of all state transitions to state i for all different combinations of i, j and k.

3.2. Generation of synthetic time series

The Markov chain approach requires initial values for the synthetic time series. In our approach, the second lag L determines the number of values that need to be set as a seed for the synthetic time series. The first values of the deseasonalized input time series can, for example, be used for the start of the synthetic time series.

The rest of the time series is generated step by step based on the values given so far, the cumulative transition probability matrix and a random generator. The generated random number is equally distributed between 0 and 1. Based on the given states j and k, state i is chosen by increasing i until the corresponding cumulative transition probability is greater than the random number. A random sample obeying the given distribution is thus generated in a straightforward manner.

If too many states and/or a too distant second lag is chosen, it is possible that certain (conditional) state transitions never occurred in the historical data. In this case, the transition probability would be equal to 0,

Table 1. Overview of the chosen parameters for the conducted parameter variation.

Section	Examined parameter	Number of states	Second lag	Filter length
4.1.1	Number of states	25, 50, 100, 200	2	0 (no filter)
4.1.2	Second lag	100	2, 5, 10, 20	0 (no filter)
4.1.3	Filter length	100	10	0, 2, 5, 10

which violates the condition given in equation (7). In the process of generating synthetic time series, this situation would lead to a dead end since no further state transition is possible. For this reason, we developed a healing mechanism for unknown state transitions to allow for a higher number of states.

We allow the model to choose the transition probabilities of a 'nearby' situation that occurred in the historical data. Therefore, the looked-up state for the second lag is replaced by the next state directed toward the state of the first lag as long as the transition probability is no longer equal to 0. If this procedure is not successful, the current state i is chosen based on the previous time step plus a random function that allows it to jump into the adjacent state. In this case, the transition probability is equally distributed amongst the current and the adjacent states. This methodology successfully avoids dead ends in the generation of synthetic time series. In any case, it should be ensured that these interventions are only applied extremely rarely since this practice falsifies the results slightly. In the following examinations, these interventions never occurred more than 400 times during a simulation of 166,464 time steps and only in settings with many states ($n > 50$) and a distant second lag ($L > 5$).

The next step is to transform the amplitude discrete values to amplitude continuous values. To avoid the same value always accounting for a certain state, an equal distribution function generates a random value within the interval boundaries of the corresponding state.

This practice can lead to a flickering in the time series, which is rather unusual for wind power time series. Therefore, we subsequently apply a running average filter on the amplitude continuous time series to smooth the time series again (Oppenheim and Schafer 1989). This allows for a more precise adjustment of the generated time series. The window size of the running average filter can be seen as another parameter of the model.

Finally, the seasonalities are restamped onto the generated time series. This is achieved by multiplying the synthetic time series by the seasonality factors. However, some of the generated specific feed-in values can have a value above 1. This problem occurs if an already high value has been multiplied by a seasonality factor above 1. For this reason, only the time series without seasonalities are compared and evaluated in the following.

4. Analysis of the synthetic time series

In this chapter, the suitability of the new Markov-chain-related approach is examined by comparing the original and synthetic time series in terms of statistical characteristics. By means of a parameter variation, the influences of the model parameters 'number of states', 'time lag' and 'window size of the running average filter' on the properties of the synthetic time series are examined in section 4.1. In section 4.2, the process of deriving good model parameters and an exemplary case with well-fitted parameters are presented. Finally, the model is applied in a hypothetical application in section 4.3. The analysis shows that the presented approach is suitable for correctly dimensioning the required storage capacity of an energy system supplied solely by wind power.

4.1. Effects of varying the parameters of the new Markov-chain-related approach

To analyse the influence of the different parameters on the results, two parameters were fixed and one parameter was varied for each case. Table 1 gives an overview of the parameter settings, which only serve as examples. The conclusions drawn are valid for all other cases as well.

4.1.1. Effects of varying the number of states

For the analysis of the effect of varying the number of states, the second lag L was fixed as 2 and no running average filter was applied. This setting represents a conventional second-order Markov chain.

The chosen number of states has a major influence on the resulting time series. As described in section 3.2, the number of states is limited due to the limited input data. However, if too few states are applied, the resulting time series almost takes the form of a step function with deviations in every state due to the random function. Accordingly, the random function that chooses a random value within the state interval still has a major effect on the time series. With the increasing number of states, the amplitude range of each state decreases and therefore also the effect of the random function. This effect can be explained based on the distribution of the incremental variations $t - \tau$, which is given for two different time scales ($\tau = 1$ and $\tau = 10$) and different numbers of states ($n = 25$, $n = 50$, $n = 100$ and $n = 200$) in figure 5.

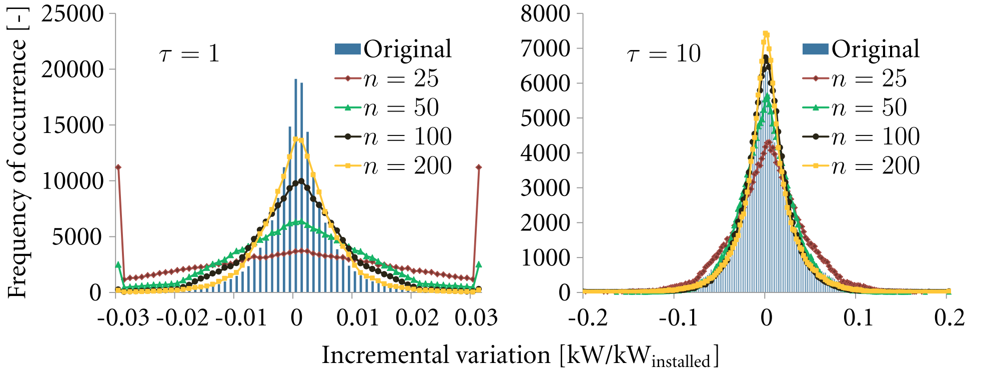

Figure 5. Distribution of the incremental variations $t - \tau$ of the original and the synthetic time series for different time scales and different numbers of states. With an increasing number of states, the distribution function of the incremental variations becomes peaked and less heavy-tailed.

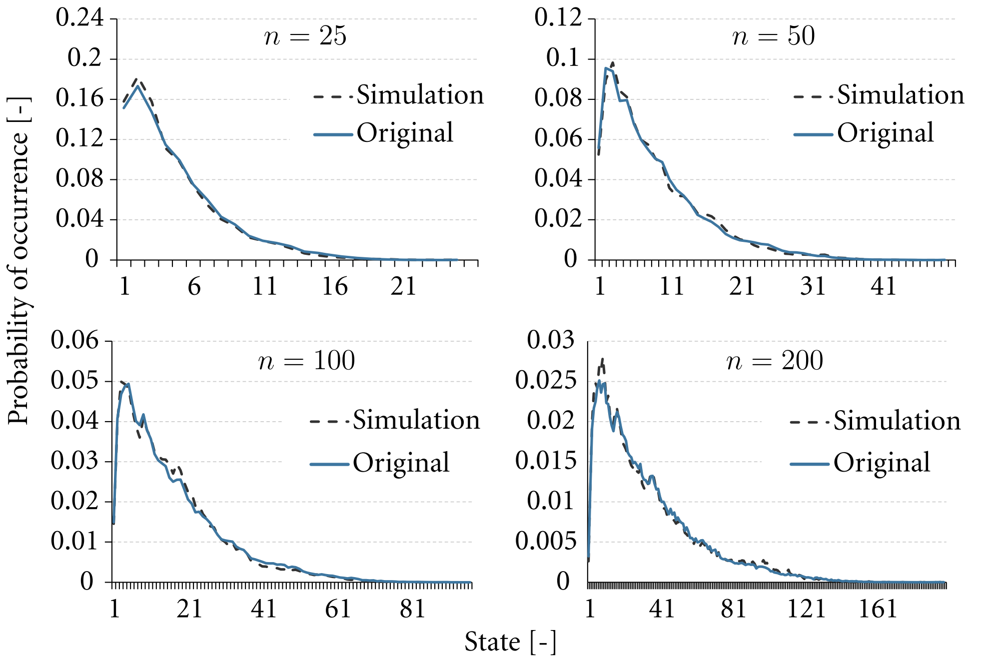

Figure 6. Probability mass function of the synthetic time series and the original time series with different numbers of states n. It is a general feature of conventional Markov chains that they reproduce the probability distribution function. This also applies to the new approach, whereas the level of detail and the accuracy increase with the number of states.

For 25 states, for example, the distribution of the incremental variations between adjacent states ($\tau = 1$) is much too flat and heavy-tailed compared to the distribution of the original ex post time series. In this way, small changes in the synthetic time series occur too rarely and larger changes much too often. With an increasing number of states, the distribution function of the incremental variations becomes peaked and closer to the distribution of the original time series. In any case, with the given second lag ($L = 2$) and without any filtering, 200 states are still not sufficient to reproduce the distribution function of increments of the ex post time series. This dependency is also valid for incremental variations with a longer time lag, for example $\tau = 10$, whereas the deviation from the incremental distribution of the original time series decreases.

The probability distribution that reflects the probability of occurrence of each state is given for all cases in figure 6. The synthetic time series are compared to the original time series that has been discretized into the corresponding number of states in each case. Of special importance is the observation that the general feature of conventional Markov chains of reproducing the probability distribution is maintained by the new approach. This accounts for any number of states. However, the level of detail and the accuracy increase with the number of states. As expected, the distribution of wind power feed-in has the shape of a Weibull distribution (Justus *et al* 1976).

A very important characteristic of a time series is the autocorrelation behaviour. The autocorrelation function (ACF) of the synthetic time series should ideally be similar to the ACF of the original ex post time series.

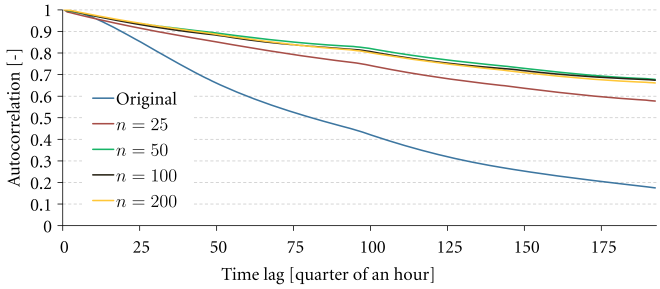

Figure 7. Autocorrelation function of the original and the synthetic time series with different numbers of states. If a sufficient number of states is applied ($n > 50$), no more effects on the autocorrelation properties as a function of the number of states can be identified. With the second lag set to $L = 2$, all synthetic time series show autocorrelations that are too high compared to the original time series.

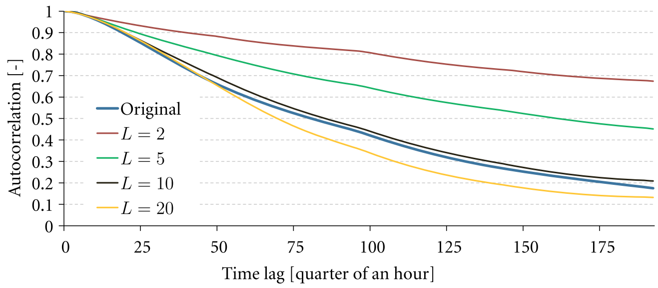

Figure 8. Autocorrelation function of the original and the synthetic time series with different second lag distances L. A more distant second lag leads to decreasing autocorrelations, which is very useful to adjust the level of fluctuation in the synthetic time series.

In general, with an increasing time lag the autocorrelations decrease in all cases (figure 7). If a sufficient number of states is applied ($n > 50$), no more effects on the autocorrelation properties as a function of the number of states can be identified. All synthetic time series show autocorrelations that are too high compared to the ex post time series. This is especially attributed to the choice of the second lag ($L = 2$), which is used in conventional second-order Markov chains. For this reason, the influence of varying the second lag will be discussed in the following section.

4.1.2. Effects of varying the second lag

The possibility of varying the second lag is the major feature of the approach presented in this paper. For the analysis of the effect of varying the second lag, no running filter was applied and the number of states was, as an example, fixed at 100. However, the presented dependencies apply for any number of states.

The possibility of influencing the shaping of trends and fluctuations is shown based on the ACFs in figure 8. It can be observed that a more distant second lag leads to decreasing autocorrelations. The effect leads to less uniformity and therefore more fluctuation in the shape of the resulting time series. The decreasing uniformity is also caused by longer-lasting trends that can be reflected much better with a more distant second lag than $L = 2$. The length of typical trends therefore depends on the distance of the second lag.

In the case of the given time series for Germany, time series with $L = 10$ to $L = 12$ fit the original time series best. This observation indicates that the characteristic dynamics of the overall wind feed-in in Germany can be better determined based on the feed-in a quarter-hour and three hours ago ($L = 12$) rather than, for example, a quarter-hour and only half an hour ago ($L = 2$).

The choice of the second lag is therefore very useful to adjust the autocorrelation properties of the synthetic time series. A well-fitted second lag automatically leads to a good shaping of trends and fluctuations in the synthetic time series.

4.1.3. Effects of varying the window size of the running average filter

In order to analyse the effect of varying the window size M of the running average filter, the number of states was, as an example, fixed at 100 and the second lag at $L = 10$. The effects on the shapes of the synthetic time series are illustrated in figure 9.

The synthetic time series without any filtering shows the typical flickering induced by the random function that distributes the values within the state interval. The running average filter smooths this flickering. The

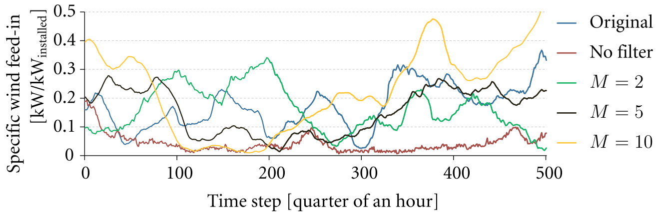

Figure 9. Original and synthetic time series with different filter lengths. The running average filter smooths the flickering resulting from the random function. In the case of the wind power time series introduced, a filter length of $M = 5$ fits the original time series best.

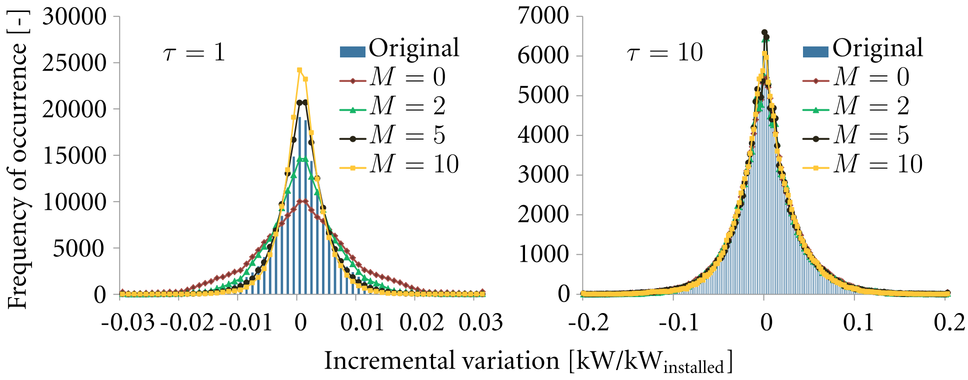

Figure 10. Distribution of the incremental variations of the original and the synthetic time series for different time scales with different filter lengths. Increasing the filter length leads to a peaked distribution of the increments between two adjacent states. The distribution of increments for higher time scales (e.g. $\tau > 10$) remains unaffected.

broader the window size of the filter, the stronger is the effect on the time series. A qualitative comparison with the ex post time series indicates that a filter length of approximately 4–5 time steps produces the best results. However, a filter with too broad a filter length such as 10 time steps smooths the time series too much.

The effect of filtering the synthetic time series can also be explained very well on the basis of the distribution of the incremental variations $t - \tau$, which is given for the time scales $\tau = 1$ and $\tau = 10$ in figure 10. Increasing the filter length leads to a peaked distribution of the increments between two adjacent states. This means that the very short-term fluctuations of the synthetic time series can be adjusted very well by varying the length of the running average filter. According to the qualitative observation of the time series, a filter length of approximately 4–5 fits the distribution of the original ex post time series best. The long-term characteristics of the shape of the generated time series are not affected by the filtering. The distributions of the incremental variations with a time lag $\tau = 10$ (figure 10, right) are hardly affected at all.

4.2. Comparision of original and synthetic time series for a well-fitted case

In this section, an exemplary case with a parameter setting is presented that is well-suited for the wind power feed-in in Germany. Besides this combination of parameters, there are a multitude of other parameter settings that generate suitable results.

The basic procedure for identifying appropriate parameters is as follows: first, the number of states n has to be determined, which–restricted by the size of the data basis –should be large enough for the probability mass function to reach a sufficient level of accuracy. Second, the distance of the second lag L needs to be chosen such that the autocorrelation properties fit those of the original time series. Finally, the short-term fluctuations can be adjusted by the running average filter. Therefore, the filter length M needs to be chosen such that the distribution of increments of the synthetic time series fits that of the original time series.

The parameters in this exemplary case were chosen according to table 2.

The shapes of the original and synthetic time series are compared in figure 11. It can be seen that it is hard to draw a distinction between the characteristics of the original and synthetically generated time series. The synthetic time series shows the same typical fluctuations and also longer-lasting trends with comparable gradients of the ramps.

Table 2. Well-suited parameters for the wind power feed-in in Germany.

Number of states	Second lag	Filter length
65	12	5

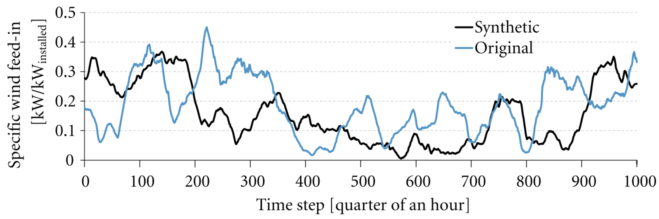

Figure 11. Original time series versus synthetic time series. The visible characteristics of the synthetic time series such as ramps and fluctuations are comparable to those of the original time series.

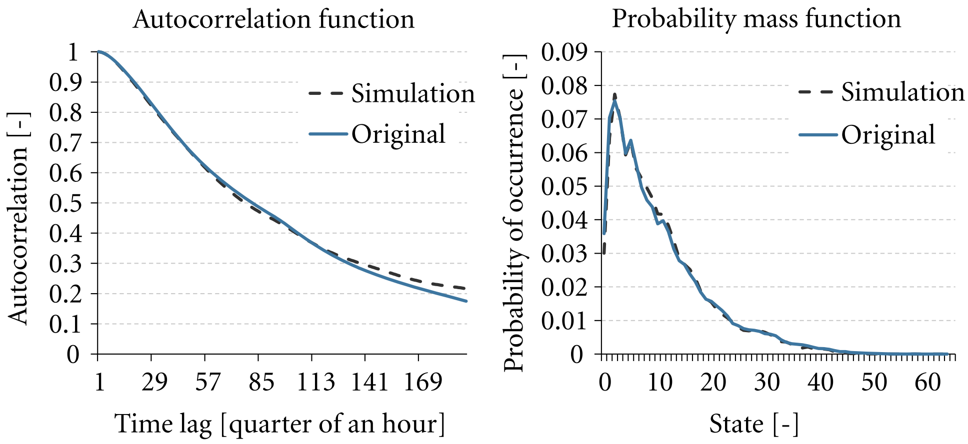

Figure 12. Autocorrelation function and probability mass function of original and well-fitted synthetic time series. The synthetic time series reproduces both characteristics of the original time series very well.

The comparability also applies to the ACF and the probability mass function of the states (figure 12). Once again, the advantageous feature of conventional Makov chains of reproducing the probability mass function is maintained by the new approach.

Of special importance for the dynamical characteristics are the distributions of the incremental variations $t - \tau$. The distributions are given in figure 13, distinguished by the different time scales $\tau = 1$, $\tau = 5$, $\tau = 10$, $\tau = 20$ and $\tau = 96$.

In the left plot in figure 13, special emphasis is put on the central peaks, whereas the tails of the distributions are cut off. For a better visualization of the tails, a semi-logarithmic scale is applied in the right plot in figure 13. Although slight variations occur, the incremental characteristics of the synthetic time series match those of the original time series very well for the different time scales.

The statistical parameters are given in table 3. Since the generation of the synthetic time series is a stochastic process, slight variations occur for every simulation. For this simulation, a minor deviation can be observed for the maximum value of the time series. This is due to the extreme rareness of the appearance of such high values as observed in the original ex post time series (according to the probability distribution given in figure 12). In total, the statistical parameters of the synthetic time series are comparable of those of the original time series.

4.3. Application for storage dimensioning

Brokish and Kirtley (2009) argue in their paper 'Pitfalls of modelling wind power using Markov chains' that synthetic wind power time series generated by a Markov chain underestimate the necessary storage capacity to balance the wind power fluctuations.

They modelled a hypothetical situation in which an isolated energy system with a constant load is only supplied by wind power. Therefore, storage is necessary to meet constant demand by fluctuating wind power.

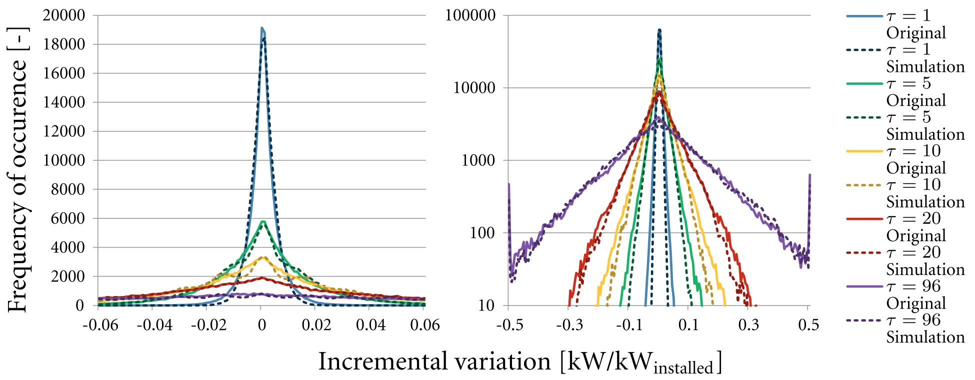

Figure 13. Distributions of the incremental variations $t - \tau$ of the original and the well-fitted synthetic time series with special emphasis put on the central peaks (left, linear scale) and on the tails (right, logarithmic-linear scale). For all time scales, comparable distributions can be observed.

Table 3. Statistical parameters of the original and the synthetically generated time series. The values confirm that both time series have comparable statistical characteristics.

Parameter	Original	Synthetic	Difference
Maximum	1.01897	0.80471	0.21427
Minimum	0.00090	0.00147	−0.00056
Mean	0.17083	0.16814	0.00269
Median	0.13255	0.13402	−0.00147
Variance	0.01966	0.01772	0.00193
Skewness	1.32178	1.24178	0.08000
Kurtosis	1.67450	1.35504	0.31946

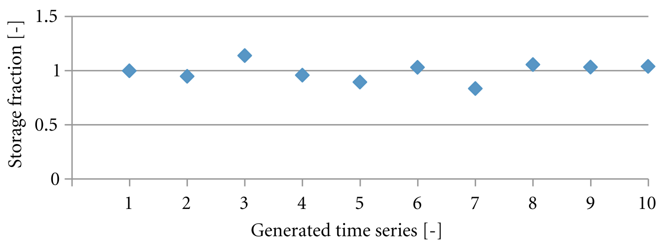

Figure 14. Storage fraction for ten randomly generated time series. As the values fluctuate around one, there is no systematic underestimation of required storage capacity.

Brokish and Kirtley (2009) estimated the required storage size using real and synthetic data, generated with a first-order Markov chain. The generated time series always underestimated the necessary storage capacity, as the first-order Markov chain cannot reproduce the stochastic properties of wind correctly.

We want to show that the wind power time series generated with the proposed Markov-chain-related approach will not lead to a systematic underestimation of the necessary storage capacity. As in the previous section, the basis of the analysis are the time series without seasonal component. We estimated the required storage size in a similar way to Brokish and Kirtley (2009). The load is assumed to be constant and is set to the value of the average wind power generation. Surplus energy is always stored and unfulfilled demand is always supplied by storage. In this way, the necessary storage capacity can be derived by subtracting the minimum storage level from the maximum storage level. This storage dimensioning was conducted for the original time series as well as for ten generated time series. The synthetic storage sizes were then divided by the real storage size, so that a storage fraction is defined, whereby a value of one means a perfect estimate.

The storage fraction for the ten generated time series is shown in figure 14. As the values fluctuate around one, there is no systematic underestimation of the required storage capacity. Therefore this example shows that the proposed Markov-chain-related approach is suitable for reproducing the stochastic behaviour of wind power time series.

5. Conclusion

The rising share of wind power feed-in leads to growing uncertainties in the electricity system. The stochastic modelling of wind is one approach to address these uncertainties in energy system models. Amongst different statistical methods, the Markov chain approach is one technique that has been frequently used in the literature for the forecasting of wind and the generation of synthetic time series.

Recently, it has been shown that conventional Markov chains are of limited use for generating wind time series with time steps shorter than 15–40 min. The reason is that more than the previous one or two time steps have significant correlations with the next time step on this time scale. Neither the level of fluctuation nor characteristic trends in the time series can thus be shaped correctly. This leads, for example, to a systematic underestimation of the size of storage used for balancing wind power simulated by conventional Markov chains.

Therefore we developed a new Markov-chain-related model. The major difference to existing Markov chain approaches is the introduction of a variable second lag, which is normally fixed at the state previous to the first lag. We also included a healing mechanism that avoids dead ends in the generation of the time series and that increases the robustness of the model. Additionally, a running average filter was introduced into Markov chain modelling that allows for the better adjustment of the very short-term characteristics of the generated time series.

The focus was placed on the generation of long-term synthetic time series with a sampling time of 15 min. The data basis covered the overall feed-in of a whole country with the example of Germany. Since weak stationarity of the input data is necessary for Markov chains, we presented a basic approach to deseasonalize the ex post time series based on arithmetic means allocated to a certain month and a certain time. After showing that the conditioned time series fulfils the criteria for wide-sense stationarity, we applied the data for the new model.

In the analysis of the generated synthetic time series, we first conducted a parameter variation to analyse the influence of the parameter settings on the properties of the synthetic time series.

The variation in the number of states to discretize the amplitude of the time series showed that a rising number of states increases the level of detail and accuracy of the probability distribution. Furthermore, the distribution of incremental variations becomes more peaked and closer to the distribution of the original wind power time series. Too few states lead to an inappropriate shape of the synthetic time series, whereas the number of states is limited due to the limited training data.

The variation of the second lag revealed that a more distant second lag leads to decreasing autocorrelations, which is very useful for the adjustment of the dynamics in the synthetic time series.

The introduction of a smoothing running average filter turned out to be a very useful method for the further adjustment of the synthetic time series. The increase of the filter length leads to a peaked distribution of the incremental variations between adjacent states. It is therefore a practical way to modify the very short-term characteristics of the synthetic time series without affecting the long-term characteristics.

Afterwards, a case with well-fitted parameters for the wind feed-in in Germany was presented as an example. The analysis revealed that the generated synthetic time series have statistical characteristics comparable to those of original ex post time series. It turned out that the second lag $L = 2$ of conventional Markov chains is not the optimal choice for wind power feed-in on a 15 min basis in Germany. The dynamics, such as the steepness of ramps and the level of fluctuation in the time series, were much better reproduced with a second lag of e.g. $L = 12$. The basis of comparision was the shapes of the time series, the autocorrelation functions, the probability mass functions, the distributions of the increments for different time scales as well as statistical parameters such as the variance.

Finally, we used the model in a hypothetical application for storage dimensioning. Therefore, an isolated system with constant load was supplied solely by wind power. On the basis of the original and the synthetic time series, the required storage capacity to balance fluctuating wind power with constant demand was simulated. It was shown that—in contrast to conventional Markov chain approaches— the generated synthetic time series do not systematically underestimate the required storage capacity. The new approach is therefore suitable for generating synthetic time series for application in energy system models or in other methods addressing uncertainty such as Monte–Carlo simulations.

Acknowledgments

TP gratefully acknowledges support from the Helmholtz Association (grant no. VH-NG-1025) and the Virtual Institute Power to Gas and Heat funded by MIWF NRW.

References

50 Hertz Transmission GmbH 2014 Extrapolated actual wind power feed-in (www.50hertz.com/en/Grid-Data/Wind-power/Archive-Wind-power)
Amprion GmbH 2014 Wind feed-in (www.amprion.net/en/wind-feed-in)
Brokish K and Kirtley J 2009 Pitfalls of modelling wind power using Markov chains *Power Systems Conf. and Exposition, 2009 PSCE '09 IEEE/PES* pp 1–6
Carapellucci R and Giordano L 2013 A new approach for synthetically generating wind speeds: a comparison with the Markov chains method *Energy* **49** 298–305
Carpinone A, Langella R, Testa A and Giorgio M 2010 Very short-term probabilistic wind power forecasting based on Markov chain models *IEEE 11th Int. Conf. on Probabilistic Methods Applied to Power Systems (PMAPS), 2010* pp 107–12
Castro Sayas F and Allan R N 1996 Generation availability assessment of wind farms *IEE Proc. Generation, Transmission and Distribution* vol 143, pp 507–18
EPEX Spot SE 2014 Intraday market with delivery on the German TSO zone (www.epexspot.com/en/product-info/Intraday/germany)
Foley A M, Leahy P G, Marvuglia A and McKeogh E J 2012 Current methods and advances in forecasting of wind power generation *Renew. Energy* **37** 1–8
Fraunhofer IWES 2014 Installed wind turbine capacity in Germany (www.windmonitor.de)
Hocaoglu F O, Gerek O N and Kurban M 2008 The effect of Markov chain state size for synthetic wind speed generation *Proc. 10th Int. Conf. on Probabilistic Methods Applied to Power Systems, 2008 PMAPS '08* pp 1–4 (http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=4912620)
Jones D I and Lorenz M H 1986 An application of a Markov chain noise model to wind generator simulation *Math. Comput. Simul.* **28** 391–402
Justus C G, Hargraves W R and Yalcin A 1976 Nationwide assessment of potential output from wind-powered generators *J. Appl. Meteorol.* **15** 673–8
Kaminsky F C, Kirchhoff R H, Syu C Y and Manwell J F 1991 A comparison of alternative approaches for the synthetic generation of a wind speed time series *J. Solar Energy Eng.* **113** 280–9
Lei M, Shiyan L, Chuanwen J, Hongling L and Yan Z 2009 A review on the forecasting of wind speed and generated power *Renewable Sustainable Energy Rev.* **13** 915–20
Marris E 2008 Energy: upgrading the grid *Nature* **454** 570–3
Milan P, Wächter M and Peinke J 2013 Turbulent character of wind energy *Phys. Rev. Lett.* **110** 138701
Negra N B, Holmstrøm O, Bak-Jensen B and Sørensen P 2008 Model of a synthetic wind speed time series generator *Wind Energy* **11** 193–209
Nfaoui H, Essiarab H and Sayigh A A M 2004 A stochastic Markov chain model for simulating wind speed time series at Tangiers, Morocco *Renew. Energy* **29** 1407–18
Oppenheim A V and Schafer R W 1989 *Discrete-Time Signal Processing* (Englewood Cliffs, NJ: Prentice-Hall) pp 311–2
Papaefthymiou G and Klockl B 2008 MCMC for wind power simulation energy conversion *IEEE Trans.* **23** 234–40
Peiyuan C, Berthelsen K K, Bak-Jensen B and Zhe C 2009 Markov model of wind power time series using Bayesian inference of transition matrix *35th Annual Conf. of IEEE Industrial Electronics, 2009 IECON '09* pp 627–32
Pesch T, Allelein H J and Hake J F 2014 Impacts of the transformation of the German energy system on the transmission grid *Eur. Phys. J. Spec. Top.* **223** 2561–75
Sahin A D and Sen Z 2001 First-order Markov chain approach to wind speed modelling *J. Wind Eng. Ind. Aerodyn.* **89** 263–9
Shamshad A, Bawadi M A, Wan Hussin W M A, Majid T A and Sanusi S A M 2005 First and second order Markov chain models for synthetic generation of wind speed time series *Energy* **30** 693–708
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 Symp. (NAPS), 2010* pp 1–8
TenneT TSO GmbH 2014 Actual and forecast wind energy feed-in (www.tennettso.de/site/en/Transparency/publications/network-figures/actual-and-forecast-wind-energy-feed-in)
Thomann G C and Barfield M J 1988 The time variation of wind speeds and windfarm power output in Kansas energy conversion *IEEE Trans.* **3** 44–9
Tong W, Xiaomeng A, Weixing L, Jinyu W and Luo W 2012 Markov chain Monte–Carlo method for the modelling of wind power time series *IEEE Innovative Smart Grid Technologies–Asia (ISGT Asia), 2012* pp 1–6
TransnetBW GmbH 2014 Wind Infeed (www.transnetbw.de/en/key-figures/renewable-energies/wind-infeed)
Turner J A 1999 A realizable renewable energy future *Science* **285** 687–9
Wang X, Guo P and Huang X 2011 A review of wind power forecasting models *Energy Procedia* **12** 770–8