A cyclic time-dependent Markov process to model daily patterns in wind turbine power production

Teresa Scholz\textsuperscript{a}, Vitor V. Lopes\textsuperscript{a,∗}, Ana Estanqueiro\textsuperscript{a}

\textsuperscript{a}LNEG, National Laboratory for Energy and Geology, Estrada do Paço do Lumiar, 22, 1649-038, Lisboa, Portugal

Abstract

Wind energy is becoming a top contributor to the renewable energy mix, which raises potential reliability issues for the grid due to the fluctuating nature of its source. To achieve adequate reserve commitment and to promote market participation, it is necessary to provide models that can capture daily patterns in wind power production. This paper presents a cyclic inhomogeneous Markov process, which is based on a three-dimensional state-space (wind power, speed and direction). Each time-dependent transition probability is expressed as a Bernstein polynomial. The model parameters are estimated by solving a constrained optimization problem: The objective function combines two maximum likelihood estimators, one to ensure that the Markov process long-term behavior reproduces the data accurately and another to capture daily fluctuations. A convex formulation for the overall optimization problem is presented and its applicability demonstrated through the analysis of a case-study. The proposed model is capable of reproducing the diurnal patterns of a three-year dataset collected from a wind turbine located in a mountainous region in Portugal. In addition, it is shown how to compute persistence statistics directly from the Markov process transition matrices. Based on the case-study, the power production persistence through the daily cycle is analysed and discussed.

Keywords:
Cyclic Markov process, wind power, persistence, diurnal pattern

*Corresponding author

Email addresses: teresa.scholz@lneg.pt (Teresa Scholz), vitor.lopes@lneg.pt (Vitor V. Lopes), ana.estanqueiro@lneg.pt (Ana Estanqueiro)
1. Introduction

The EC European Parliament objective to achieve 20% of the consumed energy from the renewable energy sector by 2020 introduced a serious challenge to the planning and operating of power systems. Wind energy is becoming a top contributor to the renewable energy mix due to rather high capacities and generation costs that are becoming competitive with conventional energy sources [28]. However, wind energy systems suffer from a major drawback, the fluctuating nature of their source, which affects the grid security, the power system operation and market economics. There are several tools to deal with these issues, such as the knowledge of wind power persistence and wind speed or power simulation. Persistence is related to stability properties and can provide useful information for bidding on the electricity market or to maintain reliability, e.g. by setting reserve capacity.

Wind power or speed simulation can be used to study the impact of wind generation on the power system. For this task, a sufficiently long time series of the power output from the wind plants should be used. However, real data records are commonly of short length and thus synthetic time series are generated by stochastic simulation techniques to model wind activity [16]. Shamshad et al. [23] used first and second-order Markov chain models for the generation of hourly wind speed time series. They found that a model with 12 wind speed states (1 m/s size) can capture the shape of the probability density function and preserve the properties of the observed time series. Additionally, they concluded that a second-order Markov chain produces better results. Nfaoui et al. [15] compared the limiting behavior of their Markov chain model with the data histograms gotten from hourly averaged wind speed and showed that the statistical characteristics were faithfully reproduced. Sahin and Sen [22] reported the use of a first-order Markov chain approach to simulate the wind speed, where: a) both transitions between consecutive times and within state wind speeds are sampled using an uniform distribution; and, b) extreme states are sampled with an exponential distribution. They showed that statistical parameters were preserved to a significant extent; however, second-order Markov chain models could yield improved results.

Although wind power can be computed from synthetic wind speed time series, Papaefthymiou and Klöckl [16] show that a stochastic model using wind power leads to a reduced number of states and a lower Markov chain model order. They compared a Markov chain based method for the direct
generation of wind power time series with the transformed generated wind speed. Both the autocorrelation and the probability density function of the simulated data showed a good fit. Thus, they concluded that it is better to generate wind power time series. Chen et al. [7] also modeled wind power by using different discrete Markov chain models: the basic Markov model; the Bayesian Markov model, which considers the transition matrix uncertainty; and, the birth-and-death Markov model, which only allows state transitions between immediately adjacent states. After comparing the wind power autocorrelation function, the authors find the Bayesian Markov model best. Lopes et al. [13] proposed a Markov chain model using states that combine information about wind speed, direction and power. From the transition matrix, they compute statistics, such as the stationary power distribution and persistence of power production, which show a close agreement with their empirical analogues. The model was then used for the two-dimensional stochastic modeling of wind dynamics by Raischel et al. [21]. They aim at studying the interactions between wind velocity, turbine aerodynamics and controller action using a system of coupled stochastic equations describing the co-evolution of wind power and speed. They showed that both the deterministic and stochastic terms of the equations can be extracted directly from the Markov chain model.

The knowledge of wind power production persistence provides useful information to run a wind park and to bid on the electricity market, since it provides information about the expected power steadiness. It can be seen as the average time that a system remains in a given state or a subset of states. Existent literature focuses mainly on wind speed persistence, which is used for assessing the wind power potential of a region. Persistence can be determined directly from the data [20, 19]; however, the presence of missing data leads to an underestimate of actual persistence. Alternative methods are based on wind speed duration curves [14, 10], the autocorrelation function or conditional probabilities. Koçak [11] and Cancino-Solórzano et al. [5] compare these techniques, and both conclude that wind speed duration curve yields the best results, i.e. results that follow the geographical and climatic conditions of the analyzed sites. Moreover, Cancino-Solórzano et al. [5] analyze the concept of “useful persistence”, which is the time schedule series where the wind speed is between the turbine cut-in and cut-out speed. The results gotten from this analysis coincide with the persistence classification obtained using the speed duration curves. In addition, Koçak [12] suggests a detrended fluctuation analysis to detect long-term correlations and analyze
the persistence properties of wind speed records. Sigl et al. [24], Corotis et al. [8] and Poje [19] proposed an approach based on the use of a power law or exponential probability distributions for the persistence of wind speed above and below a reference value. A Markov chain based method to derive the distribution of persistence is introduced by Anastasiou and Tsekos [1], who show its capability on wind speed data.

Most methods in literature of wind speed and power synthesis fail to represent diurnal patterns in the artificial data. However, these are relevant for energy system modeling and design, since their knowledge allows to plan and schedule better. For instance, a power production behavior that best matches demand needs smaller reserve capacity. Recently, Suomalainen et al. [26, 25] introduced a method for synthetic generation of wind speed scenarios that include daily wind patterns by sampling a probability distribution matrix based on five selected daily patterns and the mean speed of each day. Carapellucci and Giordano [6] adopt a physical-statistical approach to synthesize wind speed data and evaluate the influence of the diurnal wind speed profile on the cross-correlation between produced energy and electrical loads. The parameters of their model, such as diurnal pattern strength or peak hour of wind speed are determined through a multi-objective optimization, carried out using a genetic algorithm.

This paper introduces a cyclic time-variant Markov model of wind power, speed and direction designed to consider the daily patterns observed in the data. The model can be used to synthesize data for the three variables and is capable of reproducing the daily patterns. Moreover, it allows to compute persistence statistics depending on the time of the day. The paper is organized as follows: Section 2 introduces the proposed model as an extension of the “regular” Markov chain model, which is then used for comparison. Furthermore it is shown, how to compute the time-of-the-day dependent persistence statistics directly from the Markov model transition matrices. In section 3 the constrained convex optimization problem to get the model parameters is introduced and explained. It is applied to the analysis of a case-study based on real dataset, section 4. Since the model describes the joint statistics for wind power, speed and direction, Section 5 explains how to create synthetic time-series for these variables. Section 6 compares the synthesized data of both the time-variant and the time-invariant versions of the model. Moreover, it is shown how the persistence of power production varies through the daily cycle.
Nomenclature

$\alpha_0$ Initial state distribution at time step $t = 0$

$\beta_{\mu,j}$ Coefficients of the Bernstein polynomial modeling the transition probability $p_{i,j}(t)$

$1_A$ Unit column vector of the same size as subset $A$

$P P_0 \cdot \ldots \cdot P_{T-1}$

$A$ Subset of the state space, containing the states of interest for persistence

$S$ Set of observed state transitions

$S_z$ Set of transitions observed in the data together with the scaled time of the day $z$ at which they are observed

$\omega$ Weight of the extra transitions added to the objective function

$\pi$ Stationary distribution of a time-invariant Markov chain

$\pi^* \lim_{t \to \infty} P^t$

$\pi_r$ Stationary distribution at time $r$ of a time-variant cyclic Markov process

$\pi_r(j)$ Stationary probability, of state $j$ at time of the day $r$

$\pi_r(A)$ Vector whose elements are the stationary probabilities of the states in the set $A$ at time of the day $r$

$\tau$ Persistence

$\tau_r$ Time-dependent persistence in a cyclic Markov process

$b_{\mu,k}(z)$ $\mu$-th Bernstein basis polynomial of order $k$

$E[\cdot]$ Expected value operator

$P_t$ $t$-th step transition matrix of a Markov process

$p_{i,j}(t)$ $t$-th step transition probability of a Markov process
137 $p_{i,j}^{avg}$ Daily average probability of transition from state $s_i$ to $s_j$

138 $r_t$ Remainder of time step $t$ modulo $T$

139 $S$ Markov process state space

140 $s_i$ $i$-th state of a Markov process

141 $T$ Period of a cyclic Markov process

142 $t$ Time step of a Markov process

143 $X_t$ Markov process

144 $z$ Scaled time of the day

145 $\pi_A$ Stationary probability distribution of the states in subset $A$

146 $r$ Time of the day
2. Time-inhomogeneous Markov model

2.1. Definition

A discrete finite Markov process \(X_t \in S, t \geq 0\) is a stochastic process on a discrete finite state space \(S = \{s_0, s_1, ..., s_n\}, n \in \mathbb{N}\), whose future evolution depends only on its current state \([9]\). This Markov property is expressed mathematically by

\[
\Pr\{X_{t+1} = s_j \mid X_t = s_i \land X_t \in S \ \forall l = 0, ..., t - 1\} = \Pr\{X_{t+1} = s_j \mid X_t = s_i\}.
\]

\(\Pr\{X_{t+1} = s_j \mid X_t = s_i\}\) describes the probability of the Markov process moving to state \(s_j\) at time step \(t + 1\) given that it is in state \(s_i\) and is called the \(t\)-th step transition probability, denoted as \(p_{i,j}(t)\). Thus, for each time step \(t\) the Markov process has an associated transition probability matrix \(P_t\), a \(n\) by \(n\) matrix with entries \(P_t[i,j] = p_{i,j}(t)\) for all \(i, j \in \{0, \ldots, n\}\). Each \(P_t\) satisfies the following properties: \(p_{i,j}(t) \geq 0\) and \(\sum_j p_{i,j}(t) = 1\) \(\forall i, j \in \{0, \ldots, n\}, \forall t\). A Markov process is called cyclic with period \(T \in \mathbb{N}\), if \(T\) is the smallest number, such that \(p_{i,j}(mT + r) = p_{i,j}(r)\) for all \(m\) in \(\mathbb{N}\), \(0 \leq r < T\) \([13]\). Thus, a cyclic Markov process is described by \(T\) transition matrices \(P_r, r = 0, ..., T - 1\). The remainder of time step \(t\) modulo \(T\) will be denoted as \(r_t\) and thus \(r_t = r_{t+mT} \forall t, m \in \mathbb{N}\).

If the transition probabilities are time-independent, i.e. \(p_{i,j}(t) = p_{i,j}\), the process is called a (time-homogeneous) Markov chain and its probability matrix \(P \in \mathbb{R}^{n+1 \times n+1}\) is given by \(P[i,j] = p_{i,j}\). By analogy to the time-dependent case it holds that \(p_{i,j} \geq 0\) and \(\sum_j p_{i,j} = 1\) \(\forall i, j \in \{0, \ldots, n\}\).

2.2. Communication classes and irreducibility

2.2.1. Time-invariant Markov chain

The probability of reaching a state \(s_j\) from a state \(s_i\) in \(l\) time steps is given by \(P^l(i,j)\), i.e. the \(l\)-th power of the transition matrix \(P\). If a state \(s_j\) can be reached from a state \(s_i\) in a finite number of time steps and vice versa, i.e. \(\exists l \in \mathbb{N} \ P^l(i,j) \geq 0 \land P^l(j,i) \geq 0\), the states \(s_i\) and \(s_j\) communicate. All states that communicate with each other are said to be in the same communication class. If all states of a state space are in the same communication class, i.e. if it is possible to reach every state from any other state in a finite number of time steps, the corresponding transition matrix \(P\) is called irreducible.
2.2.2. Cyclic time-variant Markov process

A cyclic Markov process with period $T$ is described by $T$ transition matrices $P_r$, one for each time of the day $r = 0, ..., T - 1$. The probability of the process reaching state $s_j$ from state $s_i$ in $l$ time steps at time $t = 0$ is given as $$(P_0 \cdot \ldots \cdot P_{T-1}^m \cdot P_0 \cdot \ldots \cdot P_r(i,j))$$ with $l = mT + r_l$. For an arbitrary time-step $t$, the formula must be multiplied from the left with the term $P_r \cdot \ldots \cdot P_{T-1}$. Thus, the Markov process is irreducible, if the matrix $P = P_0 \cdot \ldots \cdot P_{T-1}$ is irreducible, i.e. if $\exists \iota \in \mathbb{N} \ P^\iota(i, j) \geq 0 \ \forall i, j$.

2.3. The stationary distribution

If a Markov chain is irreducible and aperiodic then the long-term statistics of a Markov chain are described by the stationary probability distribution: $\pi = \lim_{t \to \infty} \alpha_0 \cdot P^t$. The distribution is independent of the initial distribution $\alpha_0$ and satisfies the balance equation $\pi = \pi \cdot P$. By the Perron-Frobenius theorem it can be computed as the normalized eigenvector corresponding to the eigenvalue 1 of the transition matrix. In the case of the cyclic time-inhomogeneous Markov process there is also a stationary distribution $\pi_r$, for all $r < T$. It can be interpreted as the limiting distribution of the Markov process considering only the datapoints sampled at time of the day $r$. If the matrix $P$ is irreducible, i.e. if $\exists \pi^*$, such that $\pi^* = \lim_{t \to \infty} \alpha_0 \cdot P^t$ and the process is aperiodic, the stationary distribution $\pi_r$ exists and is given by $\pi^* \cdot P_0 \cdot \ldots \cdot P_{r-1}$, since it satisfies the balance equation $\pi_r = \pi^* \cdot (P_r \cdot P_{r+1} \cdot \ldots \cdot P_{T-1} \cdot P_0 \cdot P_1 \cdot \ldots \cdot P_{r-1})$.

2.4. Persistence

The persistence of a given state $s_i$ is related with the number of steps the system consecutively remains in this state. In the time-homogeneous case, it follows a geometric distribution with expected value $(1 - p_{i,i})^{-1}$ and is denoted by $\tau$. Anastasiou and Tsekos showed that it is possible to determine the expected time that a Markov chain stays consecutively inside a given subset of states using a simple closed-form expression. For example, in wind power applications, a typical subset of interest could contain all states.
corresponding to power production above a given threshold. To compute
this estimate, the states are renumbered, s.th. they can be partitioned into
two disjoint subsets: \( \mathcal{A} = \{s_\nu, \ldots, s_n\} \) containing the states of interest; and
\( \overline{\mathcal{A}} = \{s_0, \ldots, s_{\nu-1}\} \), its complement. Then, the transition matrix is rearranged
into the following block structure:

\[
P = \begin{pmatrix}
A & B \\
C & D
\end{pmatrix} = \begin{pmatrix}
p_{0,0} & \cdots & p_{0,\nu-1} & p_{0,\nu} & \cdots & p_{0,n} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
p_{\nu-1,0} & \cdots & p_{\nu-1,\nu-1} & p_{\nu-1,\nu} & \cdots & p_{\nu-1,n} \\
p_{\nu,0} & \cdots & p_{\nu,\nu-1} & p_{\nu,\nu} & \cdots & p_{\nu,n} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
p_{n,0} & \cdots & p_{n,\nu-1} & p_{n,\nu} & \cdots & p_{n,n}
\end{pmatrix},
\]

where the first and last block of rows and columns correspond to the
states in subset \( \overline{\mathcal{A}} \) and \( \mathcal{A} \), respectively. The expected value of persistence,
i.e. the expected number of time steps the Markov process consecutively
remains in the subset \( \mathcal{A} \) once it is entered, is given by:

\[
E\{\tau\} = \frac{\pi_\mathcal{A} \mathbf{1}_{\mathcal{A}}}{\pi_\mathcal{A} \mathbf{1}_{\overline{\mathcal{A}}}},
\]

where \( \pi_\mathcal{A} \) is the stationary probability distribution of the states in subset
\( \mathcal{A} \) and \( \mathbf{1}_{\mathcal{A}} \) is the unit column vector of size \((n - \nu + 1) \times 1\).

For the time-inhomogeneous case, persistence \( \tau_t \) is defined as the number of
time steps the Markov process is expected to remain in a state (set of states),
one once it is entered at time \( t \). For a cyclic Markov process, the persistence \( \tau_t \)
is equal for all \( t \) that are congruent modulo \( T \). Thus, it is only necessary
to compute the persistence for \( \tau_r \), \( r = 0, \ldots, T - 1 \). This can be achieved by
adapting the derivation of equation \([2.4]\) provided by Anastasiou and Tsekos
\([1]\), to time-variant cyclic Markov processes.

After renaming, s.th. the subset of interest is \( \mathcal{A} = \{s_\nu, \ldots, s_n\} \), the states of
each transition matrix \( P_r \) are rearranged as in equation \([2.4]\).
\[ P_r = \begin{pmatrix} A_r & B_r \\ C_r & D_r \end{pmatrix} = \begin{bmatrix} p_{0,0}(r) & \cdots & p_{0,n}(r) \\ \vdots & \ddots & \vdots \\ p_{0,0}(r) & \cdots & p_{0,n}(r) \\ p_{0,0}(r) & \cdots & p_{0,n}(r) \\ \vdots & \ddots & \vdots \\ p_{0,0}(r) & \cdots & p_{0,n}(r) \end{bmatrix}, \]

The probability of \( \tau_r \) to be equal to \( l \) is given as:

\[ Pr(\tau_r = l) = Pr(X_r \in \mathcal{A}, \ldots, X_{r+l} \in \mathcal{A}, X_{r+l+1} \notin \mathcal{A} \mid X_r \in \mathcal{A}, X_{r-1} \notin \mathcal{A}) \]

\[ = \sum_{i \in \mathcal{A}} Pr(X_r = i \mid X_r \in \mathcal{A}, X_{r-1} \notin \mathcal{A}) \cdot Pr(X_s \in \mathcal{A}, r < s \leq l, X_{r+l+1} \notin \mathcal{A} \mid X_r = i) \]

\[ = \sum_{i \in \mathcal{A}} \sum_{k \notin \mathcal{A}} \sum_{j \in \mathcal{A}} \tilde{\pi}_r(i) \cdot \mathcal{P}_{i,j}(r-1, \mathcal{A}) \cdot p_{j,k}(r+l) \tag{2} \]

with

\[ \mathcal{P}_{i,j}(r,l,\mathcal{A}) = Pr(X_{r+l} = j, X_k \in \mathcal{A}, 0 < k < l \mid X_r = i) \]

\[ = D_r \cdot \ldots \cdot D_{r+l-1} \cdot C_{r+l} \cdot 1_{\mathcal{A}} \]

and

\[ \tilde{\pi}_r(i) = Pr(X_r = i \mid X_r \in \mathcal{A}, X_{r-1} \notin \mathcal{A}) \]

\[ = \frac{\sum_{j \notin \mathcal{A}} Pr(X_r = i \mid X_{r-1} = j) \cdot Pr(X_{r-1} = j)}{\sum_{i \in \mathcal{A}} \sum_{j \notin \mathcal{A}} Pr(X_r = i \mid X_{r-1} = j) \cdot Pr(X_{r-1} = j)} \]

\[ = \frac{\sum_{j \notin \mathcal{A}} \pi_{r-1}(j)p_{j,i}(r-1)}{\sum_{k \in \mathcal{A}} \sum_{j \notin \mathcal{A}} \pi_{r-1}(j)p_{j,k}(r-1)}, \quad i \in \mathcal{A}, \tag{3} \]

where \( \pi_r(j) \) is the long term probability of occurrence (stationary probability) of state \( j \) at time of the day \( r \); also note that \( \pi_t(j) = \pi_r(j) \) for \( t = mr \), \( \forall m \in \mathbb{N} \).

Equation 3 can be rewritten in the matrix form to include all states in the subset \( \mathcal{A} \):
\[ \tilde{\pi}_r(\mathcal{A}) = \frac{\pi_{r-1}(\mathcal{A}) \cdot B_{r-1}}{\pi_{r-1}(\mathcal{A}) \cdot B_{r-1} \cdot 1_A}, \]

where \( 1_A \) is a unit vector of dimension \((n - \nu + 1) \times 1\) and \( \pi_{r-1}(\mathcal{A}) \) is a vector of dimensions \( 1 \times \nu\), whose elements are the stationary probabilities of the states in the set \( \mathcal{A} \) at time of the day \( r-1\).

Thus, equation \( \tilde{\pi} \) can be rewritten as:

\[
Pr(\tau_r = l) = \tilde{\pi}_r(\mathcal{A}) \cdot D_r \cdot \ldots \cdot D_{r+l-1} \cdot C_{r+l} \cdot 1_\mathcal{A} = \frac{\pi_{r-1}(\mathcal{A}) \cdot B_{r-1}}{\pi_{r-1}(\mathcal{A}) \cdot B_{r-1} \cdot 1_A} \cdot D_r \cdot \ldots \cdot D_{r+l-1} \cdot C_{r+l} \cdot 1_\mathcal{A}.
\]

The expected value of persistence at time \( r \) can then be derived as:

\[
E(\tau_r) = \sum_{l=1}^{\infty} l \cdot \frac{\pi_{r-1}(\mathcal{A}) \cdot B_{r-1}}{\pi_{r-1}(\mathcal{A}) \cdot B_{r-1} \cdot 1_A} \cdot D_r \cdot \ldots \cdot D_{r+l-1} \cdot C_{r+l} \cdot 1_\mathcal{A}
\]

Making use of the cyclicity of the Markov process, this can be expressed as:

\[
E(\tau_r) = \sum_{l=1}^{\infty} l \cdot \frac{\pi_{r-1}(\mathcal{A}) \cdot B_{r-1}}{\pi_{r-1}(\mathcal{A}) \cdot B_{r-1} \cdot 1_A} \cdot D_r \cdot \ldots \cdot D_{r+l-1} \cdot C_{r+l} \cdot 1_\mathcal{A} \cdot D^m \cdot D_r \cdot \ldots \cdot D_{r+l-1} \cdot C_{r+l} \cdot 1_\mathcal{A}
\]

where \( D = D_r \cdot \ldots \cdot D_T \cdot D_{T+1} \cdot \ldots \cdot D_{r-1} \) and \( l = mT + r_l \).

It can be seen that the sum converges after splitting it into \( T \) partial sums, one for each time of the day \( r \). For each partial sum, the only term not constant is the matrix power \( D^m \), which converges because all \( D \) eigenvalues are smaller than 1. The infinite sum for the expected value of persistence at time \( r \) can be approximated to an arbitrary degree of accuracy \( \epsilon \) by defining

\[
f_l = l \cdot \frac{\pi_{r-1}(\mathcal{A}) \cdot B_{r-1}}{\pi_{r-1}(\mathcal{A}) \cdot B_{r-1} \cdot 1_A} \cdot D^m \cdot D_r \cdot \ldots \cdot D_{r+r_l-1} \cdot C_{r+r_l} \cdot 1_\mathcal{A}.
\]

and successively adding \( f_l \), \( l = 0, 1, ..., L \) until the difference between two consecutive sums is smaller than \( \epsilon \), i.e. until \( |f_L| < \epsilon \).
3. Parameter estimation

3.1. Time-homogeneous Markov chain

The common approach to estimate the Markov chain transition matrix \( P \) is through the optimization of a constrained maximum likelihood function, which describes the realization probability of a given dataset [2]. For a sequence of \( M \) states \( \{X_0 = s_{i_0}, ..., X_M = s_{i_M}\} \) with \( s_{i_0}, ..., s_{i_M} \in S \) and \( i_0, ..., i_M \in \{0, ..., n\} \), its probability can be computed as \( Pr\{X_0 = s_{i_0}\} \cdot \prod_{(i,j) \in S} p_{i,j} \). Since the term \( Pr\{X_0 = s_{i_0}\} \) is constant, given a set of observed state transitions \( S \), it is possible to estimate \( P \) by maximizing the likelihood

\[
OF_1 = \prod_{(i,j) \in S} p_{i,j},
\]

where a transition is described by an ordered pair \((i, j)\) indicating the origin and the destination of the transition. In practice, instead of maximizing \( OF_1 \) with respect to the \( p_{i,j} \) variables it is preferable to minimize the negative log-likelihood function, i.e. \(-\log(OF_1)\), since it transforms the original mathematical programming problem into an equivalent one that is convex and, thus, has a unique solution [4]. The overall optimization problem is formulated as follows:

\[
\min \quad -\sum_{(i,j) \in S} \log(p_{i,j})
\]

subject to

\[
p_{i,j} \geq 0 \quad \forall i, j = 0, ..., n
\]

\[
\sum_j p_{i,j} = 1 \quad \forall i = 0, ..., n
\]

The constraints ensure non-negativity of the transition probabilities and that they sum up to 1 for each row of the transition matrix.

3.2. Cyclic time-variant Markov process

The goal of this time-variant Markov process is to get a model that accurately reproduces the long-term behavior while considering the daily patterns observed in the data. Thus, the proposed objective function combines two maximum likelihood estimators: the first term maximizes the likelihood of
the cycle-average probability; and, the second term maximizes the likelihood of the time-dependent probability. The final optimization problem is transformed into a convex one using the negative logarithm of the objective function. This section provides a detailed description of the objective function, the parametrization of the time-variant probability functions, and the constraints that must be added to the optimization problem to ensure its Markov properties.

3.2.1. Objective function

The transition probabilities are considered to be time-variant and cyclic with a period of $T$, i.e. for each time of the day $r (= 0, ..., T - 1)$ there is a different transition matrix $P_r$. In this paper, the time-dependent transition probabilities $p_{i,j}(r)$ are modeled by Bernstein polynomials. This has several advantages: a) a polynomial representation of the transition probabilities leads to a convex objective function and constraints, i.e. the optimization problem has a unique solution; b) a polynomial representation allows to decrease the number of variables in the optimization problem: for each transition, instead of $T$ variables only $k + 1$ are needed for a $k$ order polynomial; c) Bernstein polynomials are non-negative, which simplifies probability modeling, when compared to other polynomial bases; and d) they have the convex hull property, which, combined with de Casteljau algorithm, allows to easily write probability boundary conditions.

Bernstein polynomials are linear combinations of Bernstein basis polynomials $b_{\mu,k}(z)$, $z \in [0, 1]$. The $k + 1$ Bernstein basis polynomials of order $k$ are defined as:

$$b_{\mu,k}(z) = \binom{k}{\mu} z^\mu (1 - z)^{k - \mu}$$

with $\mu = 0, ..., k$ and $\binom{k}{\mu}$ the binomial coefficient. Thus, the transition probabilities $p_{i,j}(z)$ are described by

$$p_{i,j}(z) = \sum_{\mu=0}^{k} \beta_{i,j}^{\mu} b_{\mu,k}(z),$$

with $\beta_{i,j}^{\mu} \in \mathbb{R}$ and $z = \frac{r}{T}$, since the polynomial variable has to be scaled, s.th. it is between 0 and 1.
Hence, to maximize the likelihood of the time-dependent transition probabilities given the data, the objective function must consider the time of the day \( z \) when the transition happens. Therefore, the objective function introduced in (4) becomes \( \sum_{(i,j) \in S_z} \log(p_{i,j}(z)) \), where \( S_z \) is the set of observed transitions together with the time \( z \) when they happen. This objective function allows to compute the intra-cycle transition probability functions, and thus to represent the daily patterns present in the data.

A second term is added to this function, namely \( \sum_{(i,j) \in S} \log(p^{avg}_{i,j}) \), where \( S \) is the set of transitions observed in the data as defined in section 3.1 and \( p^{avg}_{i,j} \) is the cycle-average (daily) probability of transition from state \( s_i \) to \( s_j \). It can be computed as follows:

\[
p^{avg}_{i,j} = \frac{1}{1 - 0} \int_0^1 p_{i,j}(z) dz = \int_0^1 \sum_{\mu=0}^k \beta^{i,j}_{k} b_{\mu,k}(z) dz = \sum_{\mu=0}^k \beta^{i,j}_{k} \int_0^1 b_{\mu,k}(z) dz = \frac{1}{k + 1} \sum_{\mu=0}^k \beta^{i,j}_{k}
\]

This second term is the maximum likelihood estimator for the daily average probability and its addition to the objective function increases the consistency of the long-term behavior of the Markov process with the data. Therefore, the overall objective function \( OF_2 \) is given by:

\[
OF_2 = -\sum_{(i,j) \in S} \log(p^{avg}_{i,j}) - \sum_{(i,j) \in S_z} \log(p_{i,j}(z))
\]

\[
= -\sum_{(i,j) \in S} \log\left(\frac{1}{k + 1} \sum_{\mu=0}^k \beta^{i,j}_{k}\right) - \sum_{(i,j) \in S_z} \log\left(\sum_{\mu=0}^k \beta^{i,j}_{\mu} b_{\mu,k}(z)\right)
\]

and minimization is performed with respect to the coefficients \( \beta^{i,j}_{\mu} \) (model parameters).

### 3.2.2. Constraints

The estimation of the model parameters requires the transition probability functions to comply with several constraints, to ensure:

- \( C^0 \)- and \( C^1 \)-continuity at \( z = 0 \),
• row-stochasticity of the transition matrices at every time of the day $z$
and
• that the transition probability functions are non-negative and bounded
by 1.

Thus, to complete the specification of the optimization problem this section
explains all the necessary constraints required for the model parameters to
describe a cyclic Markov process.

**Periodicity**

The transition probability functions are modeled using Bernstein polyno-
mials, which are smooth, i.e. $C^\infty$-continuous functions. In general, the values
at both ends of their domain ($0$ and $1$) need not be equal. Thus, to avoid
sudden changes in the value and slope of each probability function between
the cycles, two constraints are added to ensure $C^0$ and $C^1$-continuity. An-
other reason is the arbitrariness of the cycle starting position, which affects
the position of the discontinuity if these conditions are not used.

The first constraint is $p_{i,j}(0) = p_{i,j}(1)$. Since $b_{\mu,k}(0) = \delta_{\mu,0}$ and $b_{\mu,k}(1) =
\delta_{\mu,k}$ the constraint can be reformulated as $\beta^i_j = \beta^i_k$, where $\delta$ is the Kronecker
delta. The second constraint is added to ensure $C^1$-continuity, i.e. $\frac{dp_{i,j}}{dz}(0) =
\frac{dp_{i,j}}{dz}(1)$. The first derivative of a Bernstein basis polynomial can be written
as a combination of two polynomials of lower degree:

$$\frac{db_{\mu,k}}{dz}(z) = k(b_{\mu-1,k-1}(z) - b_{\mu,k-1}(z))$$

Thus, the first derivative of a transition probability $p_{i,j}(z)$ is given by:

$$\frac{dp_{i,j}}{dz}(z) = k(\sum_{\mu=1}^{k}(\beta^i_{\mu} - \beta^i_{\mu-1})b_{\mu-1,k-1}(z) - \beta^i_k b_{k,k-1}(z))$$

Hence, using $b_{\mu,k}(0) = \delta_{\mu,0}$ and $b_{\mu,k}(1) = \delta_{\mu,k}$ as well as the first constraint
$\beta^0_k = \beta^i_j \forall i,j = 0,...,n$, the constraint $\frac{dp_{i,j}}{dz}(0) = \frac{dp_{i,j}}{dz}(1)$ reduces to the
following linear constraint:

$$\beta^i_k = \beta^i_j = 0.5(\beta^i_1 + \beta^i_{k-1})$$

**Row stochasticity of transition matrices**
To ensure row stochasticity of the time-variant transition matrices, it is
necessary to ensure that $\sum_j p_{i,j}(z) = 1$ for all $i$ and $z$. Since the Bernstein
basis polynomials of order $k$ form a partition of unity, i.e.

$$\sum_{\mu=0}^k b_{\mu,k}(z) = 1$$

the constraint can be re-written as a linear combination of the polynomial
coefficients:

$$\sum_j p_{i,j}(z) = 1 \iff \sum_j \sum_{\mu=0}^k \beta_{i,j}^{i,j} b_{\mu,k}(z) = 1$$

$$\iff \sum_j \beta_{i,j}^{i,j} = 1$$

Non-negative transition probabilities are bounded by 1
The most straightforward way to implement this constraint is to add two
inequalities for each time of the day and each transition probability $p_{i,j}$, i.e.

$$p_{i,j}(z) \geq 0 \quad \forall i, j, z$$

$$p_{i,j}(z) \leq 1 \quad \forall i, j, z$$

(5)

However, this constraint significantly increases the problem size, since it
requires $2 \cdot T \cdot n^2$ inequalities. An alternative constraint can be formulated by
using the convex hull property of the Bernstein polynomials. This constraint
makes the overall optimization problem size smaller, but is more restrictive.

Every Bernstein polynomial $\sum_{\mu=0}^k \beta_{\mu} b_{\mu,k}(z)$ always lies in the convex hull
defined by its control points $(\frac{k}{\mu}, \beta_{\mu})$, $\mu = 0, ..., k$. Thus the constraint

$$0 \leq p_{i,j}(z) \leq 1 \quad \forall i, j = 0, ..., n$$

can be reformulated in terms of the polynomial coefficients as

$$0 \leq \beta_{i,j} \leq 1 \quad \forall \mu = 0, ..., k$$

(6)

Since constraint (6) is a sufficient but not necessary condition for con-

straint (5) the reformulation leads to a more restrictive overall minimization
problem, i.e. the optimum objective function value is always higher or equal
when compared with the problem with original constraint. The convex
hull bound of Berstein polynomials can be tightened by subdivision, i.e.
by subdividing the domain in two regions and finding new control points
\( \beta_0^{i,j}(1), \ldots, \beta_k^{i,j}(1) \) and \( \beta_{k+1}^{i,j}(1), \ldots, \beta_{2k}^{i,j}(1) \) such that the function output re-
 mains unchanged. With each subdivision, the control points form a tighter
bound around the polynomial and thus the polynomial coefficients can as-
sume values in a wider range. The new control points represent the polyno-
mial restricted to the two sub-intervals \([0, z^*]\) and \([z^*, 1]\), where \(z^* \in [0, 1]\)
is the cutting point of the division. For simplicity, \(z^*\) is fixed to 0.5 in all
transition probabilities. The new control points can be determined by linear
combinations of the original control points \( \beta_0^{i,j}, \ldots, \beta_k^{i,j} \). This can be performed
efficiently using de Casteljau algorithm, which in matrix form is given as:

\[
\begin{pmatrix}
\beta_0^{i,j}(1) \\
\vdots \\
\beta_k^{i,j}(1)
\end{pmatrix} =
\begin{pmatrix}
b_{0,0}(z^*) & 0 & \cdots & 0 \\
b_{0,1}(z^*) & b_{1,1}(z^*) & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
b_{0,k}(z^*) & b_{1,k}(z^*) & \cdots & b_{k,k}(z^*)
\end{pmatrix}
\begin{pmatrix}
\beta_0^{i,j} \\
\vdots \\
\beta_k^{i,j}
\end{pmatrix} = C_l
\begin{pmatrix}
\beta_0^{i,j} \\
\vdots \\
\beta_k^{i,j}
\end{pmatrix} = C_l \cdot \beta^{ij}
\] (7)

and

\[
\begin{pmatrix}
\beta_{k+1}^{i,j}(1) \\
\vdots \\
\beta_{2k}^{i,j}(1)
\end{pmatrix} =
\begin{pmatrix}
b_{0,k}(z^*) & b_{1,k}(z^*) & \cdots & b_{k,k}(z^*) \\
0 & b_{0,k-1}(z^*) & \cdots & b_{k-1,k-1}(z^*) \\
\vdots & \vdots & \ddots & \vdots \\
0 & \cdots & \cdots & 0
\end{pmatrix}
\begin{pmatrix}
\beta_0^{i,j} \\
\vdots \\
\beta_k^{i,j}
\end{pmatrix} = C_r
\begin{pmatrix}
\beta_0^{i,j} \\
\vdots \\
\beta_k^{i,j}
\end{pmatrix} = C_r \cdot \beta^{ij}
\] (8)

The subdivision can be applied recursively to further improve the convex
bound around the polynomial. The corresponding coefficients are computed
by applying equations 7 and 8 to the left and right coefficient vectors. Defining
\( C = (C_l, C_r)^T \) and \( I_z \) as the identity matrix of dimension \(z \times z\), the
coefficients \( \beta^{ij}(w) = (\beta_0^{i,j}(w), \ldots, \beta_{2^w-1}^{i,j}(w)) \) after \( w \) subdivisions can be ob-
tained by:

\[
\beta^{ij}(w) = (C \otimes I_{2^{w-1}}) \cdot (C \otimes I_{2^{w-2}}) \cdot \ldots \cdot (C \otimes I_2) \cdot (C \otimes I_1) \cdot \beta^{ij}
\]

where \( \otimes \) denotes the Kronecker product. The number of inequalities
needed for the implementation of this constraint is \((k + 1) \cdot 2^{w+1} \cdot n^2\). Thus,
its use only makes sense if it decreases the problem size, i.e. for a number of subdivisions $\omega$ such that $(k + 1) \cdot 2^{\omega+1} \leq T$.

### 3.2.3. Problem formulation

The overall optimization problem to be solved for the estimation of the transition probability coefficients $\beta_{i,j}^{\mu}$ can be written as:

$$
\min_{\beta_{i,j}^{\mu}} - \sum_{(i,j) \in S} \log\left( \frac{1}{k+1} \sum_{\mu=0}^{k} \beta_{i,j}^{\mu} \right) - \sum_{(i,j) \in S} \log\left( \sum_{\mu=0}^{k} \beta_{i,j}^{\mu} b_{\mu,k}(z) \right) \quad (9)
$$

subject to

$$
\begin{align*}
\beta_{i,j}^{0} &= \beta_{i,j}^{1} \quad \forall i, j = 1, \ldots, n \\
\beta_{i,j}^{0} &= 0.5(\beta_{i,j}^{1} + \beta_{i,j}^{1}^{k-1}) \quad \forall i, j = 0, \ldots, n \\
\sum_{j} \beta_{i,j}^{\mu} &= 1 \quad \forall i = 0, \ldots, n; \forall \mu = 0, \ldots, k \\
\beta_{i,j}^{1}(w) &\leq 1 \quad \forall i, j = 0, \ldots, n \\
0 &\leq \beta_{i,j}^{1}(w) \quad \forall i, j = 0, \ldots, n
\end{align*}
$$

where $w$ is the number of subdivisions and $k$ is the order of the Bernstein polynomials, which have to be specified. The objective function (9) is a combination of two negative log-likelihood functions to ensure the Markov process captures both the daily patterns and the long-term behavior of the original data. The optimization is performed with respect to several constraints: constraints (10) and (11) ensure $C^0$- and $C^1$-continuity at $z = 0$. The row-stochasticity of the transition matrix is ensured by constraint (12). The last two constraints (13) and (14) bound the transition probabilities between 0 and 1.

It is expected that the objective function decreases with the polynomial order and the number of subdivisions. The parameters of the Markov chain model $\beta_{i,j}^{\mu}$ are estimated by solving the optimization problem using a rigorous numerical solver. The model was formulated making use of the casadi computation framework [3] and the optimization was performed by ipopt, a nonlinear interior-point solver [27], which ensures convergence to the global optimum in the case of convex optimization problems.
4. Application of the cyclic Markov process to wind turbine modeling

4.1. The data

The data for this study was obtained from a wind power turbine in a wind park located in a mountainous region in Portugal. The time series consists of a three-year period (2009-2011) of historical data gotten from the turbine data logger. The sampling time of 10 minutes leads to 144 samples each day. The data-set comprises three variables, wind power, speed and direction (nacelle orientation). The wind speed information was collected from the anemometer placed in the wind turbine hub. Due to confidentiality, wind power and speed data values are reported as a fraction of the rated power and the cut-out speed, respectively.

4.2. Markov chain state definition

Discrete Markov chain models require the definition of the states when applied to describe continuous variables. This work proposes to characterize the wind turbine states using three different variables: wind power, speed and direction. As such, each state is defined by all the points inside a polyhedron in three-dimensional space.

Figure 1: Representation of all data points projected into the: a) wind direction and speed plane (left); and, b) wind power and speed plane (right). Each rectangle is the projection of a state polyhedron into the two planes. Overall, they define the final state partition for the three-dimensional variable space.

Fig. 1 presents all data observations and the state partitions projected into: a) the wind direction and speed plane; and, b) the wind power and
speed plane. As expected, the observations projected into the wind power
and speed plane define the characteristic power curve of the wind turbine.
It shows the three operational regions of a wind turbine: a) below the cut-in
speed no power is produced; b) between cut-in and rated wind speed the
power increases proportionally to the cube of wind speed; c) at wind speeds
between the rated and the cut-out wind speed, the turbine control system
limits the power output to a constant value. In the wind direction and speed
plane, data is widely scattered and shows the dominant wind patterns at the
site. Three accumulation regions can be identified: one for low wind speeds,
centered on 0.25, which is the mode of the wind speed, and two defining the
dominant wind directions around 100° and 300°.

The data space is discretized unevenly to get a good resolution of the high-
slope region of the power curve. In a previous work [13], this partition was
used in a time-homogeneous Markov chain and proved to lead to an accurate
representation of the original data. The wind direction and power are divided
by an equally spaced grid leading to 12 \( \{d_1, ..., d_{12}\} \) and 20 \( \{p_1, ..., p_{20}\} \)
classes, respectively. The wind speed is divided as follows: values below the
cut-in speed define one class \( sp_1 \); between the cut-in and rated wind speed
the discretization is narrowed by selecting 10 classes \( \{sp_2, ..., sp_{11}\} \); and be-
tween the rated and cut-out wind speed discretization is widened and 4 classes
\( \{sp_{12}, ..., sp_{15}\} \) are defined. Data points with wind speed above the cut-out
wind speed are discarded. The complete state set is constructed by listing
all possible combinations of the classes of each variable. Due to physical
constraints between the variables, most of the states are empty (fig. 1(left))
and can are discarded. This reduces the number of states from 3840 to 778,
for this turbine.

4.3. Additional transitions to promote a single communication class

The solution of the optimization problem described in section 3 comprises
a set of transition matrices \( P_r, r \in \{0, ..., 143\} \). However, the constraints in
the optimization problem definition do not force the matrix \( P = P_0 \cdot \ldots \cdot P_{143} \)
to be irreducible and thus the Markov process to have a single communication
class. So, during data synthesis, the Markov process can get “trapped”
within a communication class. To induce the Markov process to have a
single communication class, additional transition counts are introduced into
the data. The goal is to add a small set of transitions to promote state
connectivity without distorting the original data. Thus, the set is composed
of transitions that connect neighboring states in the state space, since those
are the ones most likely to occur.

For a state \( s_i = (p_l, sp_p, d_q) \), its neighborhood \( V \) is defined as

\[
V(s_i) := \{(p_{l-1}, sp_{p-1}, d_{q-1}), (p_{l}, sp_p, d_q), (p_{l+1}, sp_{p+1}, d_{q+1})\}\). \hfill (15)

It should be noted that, unlike power and speed, direction is a circular variable, e.g. states \( d_0 \) and \( d_{11} \) are considered neighbors. If a neighbor state \( s_j \in V(s_i) \) is present in the dataset, a transition \((i, j)\) is added to the set of extra transitions \( S_E \). For this dataset, originally consisting of 150601 transitions, 13610 transitions are added.

The extra transitions must be considered to happen at an unknown time of the day \( z \). Thus, they can only be accounted for in the objective function term without time information, i.e. only in the time-variant part of the objective function. This directly affects the values for \( p_{avg}^{i,j} \) and, indirectly, the model parameters. Since the aim is to cause a minimal impact on the transition probabilities, the additional term is weighed by a factor \( \omega < 1 \) to directly control its influence. In this work it is fixed to 0.05. Thus, the following term is added to the objective function:

\[
-\omega \cdot \sum_{(i,j)\in S_E} \log(p_{avg}^{i,j}) = -\omega \cdot \sum_{(i,j)\in S_E} \log\left(\frac{1}{k+1}\sum_{\mu=0}^{k} \beta_{i,j}^{k+1}\right) \hfill (16)
\]

Although the use of the extra transition set does not ensure the time-variant Markov process to have a single communication class, results show a decrease of the number of communication classes from 13 to 1 in this dataset.

5. Simulation of wind power, speed and direction time series

To simulate wind power, speed and direction time series the method described by Sahin and Sen \[22\] is adapted to the cyclic time-variant Markov model as follows. First, the cumulative probability transition matrices \( P_{r}^{cum} \) with \( P_{r}^{cum}(i,j) = \sum_{k=0}^{j} p_{r,i,k}(r) \) are computed. Then an initial state \( s_i \), i.e. \( X_0 = s_i \), is randomly selected. A new datapoint \( X_{t+1} \) is generated by uniformly selecting a random number \( \epsilon \) between zero and one. The corresponding state \( s_{new} (X_{t+1} = s_{new}) \) is chosen such that the probability of reaching it from the current state \( s_i \) is bigger than \( \epsilon \), i.e. such that \( P_{r}^{cum}(i, new) \geq \epsilon \). Based on this discrete state sequence, a real value for the wind power/speed/direction variables is generated by sampling each state partition uniformly.
6. Results and discussion

6.1. Daily patterns in the data

The wind power, speed and direction time-series clearly show a daily time-dependent behavior.

![Figure 2: Two-dimensional histograms of the original time series data: speed-time (top left), direction-time (top right) and power-time (bottom left). The subfigure on the bottom right shows the p-value of the Kolmogorov-Smirnov test used to compare the wind speed distribution on the different times of the day.](image)

Figure 2 shows that, on average, the turbine does not produce power between 10 am and 3 pm. In this time interval, low wind speeds (0.1 - 0.25) are the most likely events. There are two dominant wind directions: around 100° and 300°. Moreover, they occur at specific times of the day; between 5 and 10 am, the wind typically blows from the 100° direction, the rest of the day from 300°.
To assess whether these two dominant directions might be due to summer/winter seasonality, the dataset was divided in two subsets, one covering the period from April to September and the other from October to March. The histogram analysis shows that both, summer and winter subset, have the same two dominant directions (figures not shown). Thus, it was concluded that the time-dependent pattern is not induced by this seasonality.

Figure 2 bottom-right plot shows the p-values of the Kolmogorov-Smirnov test applied to the wind speed distributions at different times of the day. The Kolmogorov-Smirnov test is a nonparametric test for the equality of continuous one-dimensional probability distributions. Thus, the high p-values around the diagonal illustrates that wind speed distributions for consecutive times of the day are similar. The same holds for wind speeds in the morning and evening, i.e. before 9am and after 4:30pm. The wind speed distribution between 10am and 3pm is clearly different.

6.2. Choice of polynomial order and number of subdivisions

The model introduced in section 3.2 has two parameters that need to be defined: $k$, the order of the Bernstein polynomials used to model the transition probabilities; and $w$, the number of subdivisions used to tighten the convex hull that bounds the polynomials. To choose proper values for these parameters, different models were computed by varying $k = 4...10$ and $w = 0...3$. For each model, synthetic data was generated following the procedure described in section 5 and compared with the real dataset.
Figure 3: Bar plots comparing the objective function value (top left), the daily average Jensen-Shannon distance between original and synthesized data (top right), the number of inequalities in the problem formulation (bottom left) and the CPU time spent in IPOPT solving the optimization problem (bottom right) of the tested models.

Figure 3 shows bar plots comparing the different models using four criteria: the objective function value, the daily average Jensen-Shannon distance between original and synthesized wind direction data, the number of inequalities in the problem formulation and the CPU time spent in IPOPT solving the optimization problem. The Jensen-Shannon distance is the square root of the Jensen-Shannon divergence $d_{js}$, which, for two discrete probability distributions $q_1$ and $q_2$ is defined as:

$$d_{js} = \frac{1}{2} \sum_i \frac{q_1(i)}{q_2(i)} \cdot q_1(i) + \frac{1}{2} \sum_i \frac{q_2(i)}{q_1(i)} \cdot q_1(i).$$

Comparing the models using the objective function value (figure 3 top left) shows a decrease of the objective function as the model order and the
number of subdivisions increase. It can be seen that the impact of the number
of subdivisions is higher for models with higher polynomial order. Moreover,
the first subdivision has the highest impact since it leads to the highest
decrease of the objective function value. The daily average of the Jensen-
Shannon distance (figure 3 top right) decreases with the polynomial order
until sixth order. The same behavior can be observed for the number of
subdivisions: until the sixth order, the Jensen-Shannon distance decreases
with the number of subdivisions. The number of inequality constraints in
the optimization problem as well as the number of CPU seconds spent in the
solver show the expected behavior (figure 3 bottom). They increase linearly
with the polynomial order and exponentially with the number of subdivisions.
Based on these observations, a basis order of 6 with 2 subdivisions was chosen
as the best trade-off between an accurate representation of the average daily
patterns and computational costs.

6.3. Capturing long-term statistics

This section compares the main statistical properties derived from the
original data with the ones derived from the data generated by the time-
variant Markov model.

Figure 4: Comparison of the probability distribution of wind power (left),
wind speed (middle) and wind direction (right) of the original with the syn-
thesized data.

Figure 4 compares the wind power (left), speed (middle) and direction
(right) distribution of the original with the synthetic data generated using
the Markov model. In general, the distributions are in close agreement. The
wind power distribution is bimodal, with the modes located at the minimum
and maximum power. It shows that the intermediary power levels are rather
rare, for instance, the states corresponding to a power production between
0.4 and 0.9 have a low probability. The wind speed distribution follows the
expected behavior, a single mode distribution with a long tail for the high wind speeds (Weibull distribution). The wind direction distribution is bi-modal with the two modes at 100 and 300 degrees, which are the prevailing wind directions at the turbine site (figure 2).

Figure 5: Two dimensional power-time histograms of the original (left) and synthetic (right) time-series data.

Figure 5 shows two plots: on the left, the empirical 2D distribution of the wind power and direction computed from the data and, on the right, the same distribution computed using the data generated by the Markov model. Its comparison shows that the model captures the joint statistics for the wind power and direction from the data. It is possible to see the two dominant directions associated with high wind power production, namely the sectors from 100 to 120 and from 290 to 320 degree. Figures 2 and 5 clearly demonstrate the capability of this Markov model to capture the combined characteristics of the wind power, speed and direction. The long-term behavior of the model is close to what is observed in the dataset.

6.4. Capturing time-dependent behavior

As shown in section 6.1, the original data clearly exhibits a time-dependent behavior. To test, if the time-dependent Markov model can capture it, synthetic data was generated and the histograms compared to the ones of the original data. Moreover, to obtain a comparison with the “regular” way of data synthesis with Markov models, data was also generated from the time-invariant Markov chain.
Figure 6: Two dimensional histograms of the synthetic time-series data, generated with the time-variant Markov model (left) and the time-invariant Markov chain (right): power-time (top), speed-time (middle) and direction-time (bottom).

The comparison of figures 2 and 6 (first column) shows that the time-variant Markov model is capable of reproducing the time-variant behavior of the data. Figure 6 (second column) presents the results of using a time-invariant Markov chain model, i.e. by using constant transition probability.
functions. As expected, each variable statistic distribution remains constant during the daily cycle.

6.5. Time-dependent persistence of production

The time-dependent Markov model allows to compute the persistence of power-production depending on the time of the day. Figure 6 shows the time-dependent persistence of power production for different power levels \((i \cdot 0.05 \cdot p_{\text{max}}, \text{for } i = 1, ..., 19)\). The persistence analysis is presented for two power production levels: a) persistence of useful power production (PUPP) defined as above \(0.15 \cdot p_{\text{max}}\), i.e. the power level corresponding to the wind speed mode at the turbine site; and, persistence for high power production (PHPP), i.e. above \(0.7 \cdot p_{\text{max}}\). It can be seen, that the higher the power level, the lower the persistence. Moreover, for all power levels, persistence is minimal between 5 and 10 am. PHPP is fairly constant throughout the day (dark line), the maximal differences are between 10 and 30 minutes, whereas PUPP reaches a maximum at around 5 pm (white line).
Figure 7: Time-dependent persistence of power production above $i \cdot 0.05 \cdot p_{\text{max}}$, for $i = 1, ..., 19$. The lines highlight the time-dependent persistence, for two conditions: a) in white, for power production above $0.15 \cdot p_{\text{max}}$ (PUPP); and, b) in black, the power production above $0.7 \cdot p_{\text{max}}$ (PHPP).

Since the data shows two different dominant directions (figure 4), figure 5 presents the persistence of power production conditioned to each dominant direction, i.e. for the direction sectors from 90° to 180° and 270° to 360°.
As expected, the persistence in both direction sectors is lower than the unconditioned persistence. For wind directions in the sector 90°-180°, all levels of power production have a minimum persistence between 80 and 100 minutes at around 1 pm. Maximum persistence is around midnight varying between 220 minutes (PUPP) and 140 minutes (PHPP). It can be seen that for power levels below 50% of maximum production the time of day dependency of persistence is very similar. For power levels above 50% persistence decreases as power production increases. However, the persistence variability with the power level is rather low, for example, maximum persistence at a level of 75% is almost 180 minutes whereas for a level above 0.05% is 200 minutes.

For wind directions in the sector 270°-360°, it shows that, for all power levels, the curves for both PUPP and PHPP are similar, i.e. their minima and maxima are located around the same time of the day. For instance, maximal persistence of production is reached at around midday. However, for this direction sector, the higher the power production level, the lower the persistence. For power production above 0.05% of maximum power the persistence is 250 minutes, persistence of production above 75% of maximum power is only 100 minutes.

Comparing with the other dominant direction, it can be seen, that they have very different persistence behavior. The maxima and minima are at different times of the day for every power level. The persistence increases as power production decreases, for all power levels in the case of the 270°-360° sector. For the 90°-180° sector it decreases only until 50% of maximum power
production. Below that, it remains approximately constant.

7. Conclusions

This paper presents an inhomogeneous Markov process to model wind power production. It is developed using states, which combine information about the wind speed, direction and power variables, using real data recorded by a wind turbine in Portugal. The joint partition of the three-dimensional variable space allows to decrease the number of the model states and, simultaneously, encodes the wind power curve into the Markov chain model. The transition probabilities are considered to be functions that depend on the time of the day and modeled as Bernstein polynomials. The estimation of the transition matrices is performed by solving a constrained convex optimization problem. Its objective function combines two log-likelihood functions with the purpose to accurately represent both the long-term behavior and the daily fluctuations seen in the original data. To evaluate the statistical properties of the estimated Markov model, synthetic time-series are generated and compared with the original data statistics. Results demonstrate that the proposed Markov model can reproduce the diurnal patterns in the data. Moreover it is demonstrated how the persistence of power production throughout the time of the day can be estimated from the Markov process transition matrices.

Acknowledgments

The authors thank the Fundação para a Ciência e a Tecnologia for financial support (SFRH/BD/86934/2012, FCOMP-01-0124-FEDER-016080 (PTDC/SENER/1141718/2009)) and GENERG, SA.
References

[1] Anastasiou, K., Tsekos, C., 1996. Persistence statistics of marine environmental parameters from markov theory, part 1: analysis in discrete time. Applied Ocean Research 18 (4), 187 – 199.

[2] Anderson, T. W., Goodman, L. A., 1957. Statistical inference about markov chains. The Annals of Mathematical Statistics 28 (1), 89–110.

[3] Andersson, J., Houska, B., 2010. Towards a Computer Algebra System with Automatic Differentiation for use with Object-Oriented modelling languages. Object-Oriented Modeling Languages.

[4] Boyd, S. P., Vandenberghe, L., 2004. Convex optimization. Cambridge Univ Pr.

[5] Cancino-Solórzano, Y., Gutiérrez-Trashorras, A. J., Xiberta-Bernat, J., 2010. Analytical methods for wind persistence: Their application in assessing the best site for a wind farm in the state of Veracruz, Mexico. Renewable Energy 35 (12), 2844 – 2852.

[6] Carapellucci, R., Giordano, L., 2013. The effect of diurnal profile and seasonal wind regime on sizing grid-connected and off-grid wind power plants. Applied Energy 107 (0), 364 – 376.

[7] Chen, P., Berthelsen, K., Bak-Jensen, B., Chen, Z., 2009. Markov model of wind power time series using Bayesian inference of transition matrix. In: Industrial Electronics, 2009. IECON ’09. 35th Annual Conference of IEEE. pp. 627 –632.

[8] Corotis, R. B., Sigl, A. B., Klein, J., 1978. Probability models of wind velocity magnitude and persistence. Solar Energy 20 (6), 483 – 493.

[9] Kemeny, J. G., Snell, J. L., 1976. Finite Markov Chains. New York : Springer-Verlag.

[10] Koçak, K., 2002. A method for determination of wind speed persistence and its application. Energy 27 (10), 967 – 973.

[11] Koçak, K., 2008. Practical ways of evaluating wind speed persistence. Energy 33 (1), 65 – 70.
[12] Koçak, K., 2009. Examination of persistence properties of wind speed records using detrended fluctuation analysis. Energy 34 (11), 1980 – 1985.

[13] Lopes, V. V., Scholz, T., Estanqueiro, A., Novais, A. Q., 2012. On the use of Markov chain models for the analysis of wind power time-series. In: Environment and Electrical Engineering (EEEIC), 2012 11th International Conference on. pp. 770 –775.

[14] Masseran, N., Razali, A., Ibrahim, K., Zin, W. W., 2012. Evaluating the wind speed persistence for several wind stations in peninsular Malaysia. Energy 37 (1), 649 – 656.

[15] Nfaoui, H., Essiarab, H., Sayigh, A., 2004. A stochastic Markov chain model for simulating wind speed time series at Tangiers, Morocco. Renewable Energy 29 (8), 1407 – 1418.

[16] Papaefthymiou, G., Klöckl, B., 2008. MCMC for wind power simulation. Energy Conversion, IEEE Transactions on 23 (1), 234 –240.

[17] Pillai, S., Suel, T., Cha, S., 2005. The perron-frobenius theorem: some of its applications. Signal Processing Magazine, IEEE 22 (2), 62–75.

[18] Platis, A., Limnios, N., Le Du, M., 1998. Dependability analysis of systems modeled by non-homogeneous Markov chains. Reliability Engineering and System Safety 61 (3), 235 – 249.

[19] Poje, D., 1992. Wind persistence in Croatia. International Journal of Climatology 12 (6), 569–586.

[20] Pryor, S., Barthelmie, R., 2002. Statistical analysis of flow characteristics in the coastal zone. Journal of Wind Engineering and Industrial Aerodynamics 90 (3), 201 – 221.

[21] Raischel, F., Scholz, T., Lopes, V. V., Lind, P. G., 2012. Uncovering wind turbine properties through two-dimensional stochastic modeling of wind dynamics. Unpublished results, arXiv:1210.7161.

[22] Sahin, A. D., Sen, Z., 2001. First-order Markov chain approach to wind speed modelling. Journal of Wind Engineering and Industrial Aerodynamics 89 (34), 263 – 269.
Shamshad, A., Bawadi, M., Hussin, W. W., Majid, T., Sanusi, S., 2005. First and second order markov chain models for synthetic generation of wind speed time series. Energy 30 (5), 693 – 708.

Sigl, A. B., Corotis, R. B., Won, D. J., 1978. Run duration analysis of surface wind speeds for wind energy application. Journal of Applied Meteorology 18, 156–166.

Suomalainen, K., Silva, C., ao, P. F., Connors, S., 2013. Wind power design in isolated energy systems: Impacts of daily wind patterns. Applied Energy 101 (0), 533 – 540.

Suomalainen, K., Silva, C., Ferro, P., Connors, S., 2012. Synthetic wind speed scenarios including diurnal effects: Implications for wind power dimensioning. Energy 37 (1), 41 – 50.

Wächter, A., Biegler, L. T., 2006. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Mathematical Programming 106, 25–57.

Wen, J., Zheng, Y., Donghan, F., 2009. A review on reliability assessment for wind power. Renewable and Sustainable Energy Reviews 13 (9), 2485 – 2494.