Reliability measures of second order semi-Markov chain applied to wind energy production

Guglielmo D’Amico
Dipartimento di Farmacia, Università ‘G. D’Annunzio’ di Chieti-Pescara, 66013 Chieti, Italy

Filippo Petroni
Dipartimento di Scienze Economiche ed Aziendali, Università degli studi di Cagliari, 09123 Cagliari, Italy

Flavio Prattico
Dipartimento di Ingegneria Industriale e dell’Informazione e di Economia, Università degli studi dell’Aquila, 67100 L’Aquila, Italy

Abstract. In this paper we consider the problem of wind energy production by using a second order semi-Markov chain in state and duration as a model of wind speed. The model used in this paper is based on our previous work where we have showed the ability of second order semi-Markov process in reproducing statistical features of wind speed. Here we briefly present the mathematical model and describe the data and technical characteristics of a commercial wind turbine (Aircon HAWT-10kW). We show how, by using our model, it is possible to compute some of the main dependability measures such as reliability, availability and maintainability functions. We compare, by means of Monte Carlo simulations, the results of the model with real energy production obtained from data available in the Lastem station (Italy) and sampled every 10 minutes. The computation of the dependability measures is a crucial point in the planning and development of a wind farm. Through our model, we show how the values of this quantity can be obtained both analytically and computationally.

Keywords: semi-Markov chains; synthetic time series; dependability analysis
1. Introduction

Wind is one of the most important renewable energy sources. Wind energy is produced by converting the kinetic energy of wind into electrical energy by means of a generator. For this reason, it is important to dispose of an efficient stochastic model for wind speed changes.

First and second order Markov chain models have been extensively used in wind speed modelling and synthetic time series generation, see e.g. Shamshad et al. [1], Nfaoui et al. [2], Youcef et al. [3], Shain et al. [4] and Torre et al. [5].

The Markovian assumption has, especially in the modeling of wind speed, several flaws. In discrete time Markov chain models, waiting times in a state before making a transition to another state are geometrically distributed. Therefore Markov chains impose artificial assumptions on the structure of the data that, very often, are inappropriate. This leads to a great simplification of the model which is unable to reproduce correctly the statistical properties of the real wind speed process.

Semi-Markov chains do not have this constraint, because the waiting time distribution functions in the states can be of any type and this allow the data to speak for themselves without any restriction. For this reason, semi-Markov chains have been extensively applied to different fields [6, 7, 8, 9, 10, 11, 12, 13].

D’Amico et al. [14] was the first paper where semi-Markov chains were applied in the modelling of wind speed. In that paper were proposed first and second order semi-Markov models with the aim of generate synthetic wind speed data. It was shown that the semi-Markov models performs better than the Markov chain model in reproducing the statistical properties of wind speed data. In particular, the model recognized as being more suitable is the second order semi-Markov model in state and duration.

In this paper we show how to compute dependability measures as availability, reliability and maintainability functions for the second order semi-Markov chain in state and duration. These indicators give important information on the feasibility of the investment in a wind farm by giving the possibility to quantify the uncertainty in the wind energy production.

Another important aspect is related to the location of the wind farm. In fact, today, many wind farms are built offshore for different reasons: the wind speed is more powerful and constant due to the absence of obstacles and visual, environmental and acoustic impact is cut down. The maintenance cost, instead, is higher than the onshore wind farm. A good stochastic model can help the planning of preventive maintenance suggesting when is suitable to execute the maintenance operation. This is possible by analyzing what happens after a particular transition between two different states that is lasted for a certain time period.

The results presented here are new and generalize some of the results obtained for semi-Markov chain of order one (see Barbu and Limnios [15] and Blasi et al. [16]). The model generalizes also Markov chains and renewal models.

We apply our model to a real case of energy production. For this reason, we choose
a commercial wind turbine, a 10 kW Aricon HAWT assumed to be installed at the
station of L.S.I -Lastem which is situated in Italy.

The paper is organized as follows. Section 2 presents some definitions and notation
on the second order semi-Markov chain in state and duration. Section 3 describes the
database and the technical characteristics of the commercial wind turbine. Section 4
shows the way in which it is possible to compute the dependability measures via kernel
transformations and the value computed on the real data and on the synthetic data
are compared. In the last section some concluding remarks and possible extensions are
presented.

2. The second order semi-Markov chains in state and duration

Higher order semi-Markov processes were introduced by Limnios and Oprisan [17]. The
dependence on the past was given only through past states. These models were recently
generalized by D’Amico et al. [14] that proposed second order semi-Markov chain in
state and duration.

Let us consider a finite set of states \( E = \{1, 2, \ldots, S\} \) in which the system can be
into and a complete probability space \((\Omega, F, P)\) on which we define the following random
variables:

\[
J_n : \Omega \rightarrow E, \quad T_n : \Omega \rightarrow \mathbb{N}.
\]

They denote the state occupied at the \( n \)-th transition and the time of the \( n \)-th transition, respectively. To be more concrete, by \( J_n \) we denote the wind speed at the \( n \)-th transition and by \( T_n \) the time of the \( n \)-th transition of the wind speed process.

We assume that

\[
\mathbb{P}[J_{n+1} = j, T_{n+1} - T_n = t | J_s, T_s, J_n = k, J_{n-1} = i, T_n - T_{n-1} = x, 0 \leq s \leq n] = \mathbb{P}[J_{n+1} = j, T_{n+1} - T_n = t | J_n = k, J_{n-1} = i, T_n - T_{n-1} = x] := xq_{i,k,j}(t).
\]

Relation (2) asserts that, the knowledge of the values \( J_n, J_{n-1}, T_n - T_{n-1} \) suffices
to give the conditional distribution of the couple \( J_{n+1}, T_{n+1} - T_n \) whatever the values of
the past variables might be.

The conditional probabilities \( xq_{i,k,j}(t) \) are stored in a matrix of functions \( q = (xq_{i,k,j}(t)) \) called the second order kernel (in state and duration). The element \( xq_{i,k,j}(t) \) represents the probability that next wind speed will be \( j \) at time \( t \) given that the current wind speed is \( k \), the previous wind speed state was \( i \) and the duration in wind speed \( i \) before of reaching wind speed \( k \) was equal to \( x \) units of time.

We can define the cumulated second order kernel probabilities:

\[
xQ_{i,k,j}(t) := \mathbb{P}[J_{n+1} = j, T_{n+1} - T_n \leq t | J_n = k, J_{n-1} = i, T_n - T_{n-1} = x] = \sum_{s=1}^{t} xq_{i,k,j}(s).
\]

The process \( \{J_n\} \) is a second order Markov chain with state space \( E \) and transition
probability matrix \( xP = xQ(\infty) \). We shall refer to it as the embedded Markov chain.
The conditional cumulative distribution functions of the waiting time in each state, given the state previously occupied and the duration of occupancy are defined as

\[ xG_{i,k,j}(t) = \mathbb{P}[T_{n+1} - T_n \leq t | J_n = k, J_{n-1} = i, T_n - T_{n-1} = x] \]

\[ = \frac{1}{xP_{i,k,j}} \sum_{s=1}^{t} xq_{i,k,j}(s) \cdot 1_{\{xP_{i,k,j} \neq 0\}} + 1_{\{xP_{i,k,j} = 0\}} \]  

(4)

Denote by \( N(t) = \sup\{n : T_n \leq t \} \) \( \forall t \in \mathbb{N} \) the number of transitions up to time \( t \). The second order semi-Markov chain in state and duration can be defined as

\[ Z(t) = (Z^1(t), Z^2(t)) = (J_{N(t)-1}, J_{N(t)}) \]  

(5)

If we define, \( \forall i, k, j \in E, \) and \( t \in \mathbb{N} \), the semi-Markov transition probabilities by:

\[ x\phi_{i,k,h,j}(t) := \mathbb{P}[J_{N(t)} = j, J_{N(t)-1} = h | J_0 = k, J_{-1} = i, T_0 = 0, T_0 - T_{-1} = x], \]

(6)

it is possible to prove that they verify the following system of equations:

\[ x\phi_{i,k,h,j}(t) = 1_{\{i = h, k = j\}} \left( 1 - \sum_{j \in E} xQ_{i,k,j}(t) \right) + \sum_{r \in E} \sum_{s=1}^{t} xq_{i,k,r}(s) s\phi_{k,r,h,j}(t-s). \]

(7)

For this model more general duration dependent transition probabilities than (7) have been obtained in D’Amico et al. [14].

3. Database and commercial wind turbine

As in our previous work [14] we used a free database of wind speed sampled in a weather station situated in Italy. The station processes the speed every 10 minute in a time interval ranging from 25/10/2006 to 28/06/2011. During the 10 minutes are performed 31 sampling which are then averaged in the time interval. In this work, we use the sampled data that represents the average of the modulus of the wind speed \( (m/s) \) without considering a specific direction. This database is composed of about 230000 wind speed measures ranging from 0 to 16 m/s.

In order to apply our semi-Markov model, we discretize wind speed into 8 states (see Table 1) chosen to cover all the wind speed distribution. This choice is done by considering a trade off between accuracy of the description of the wind speed distribution and the number of parameters to be estimated. An increase in the number of states better describes the process but requires a larger dataset to get reliable estimates and it could also be not necessary for the accuracy needed in forecasting future wind speeds. On the third column of Table 1 we report the number of times the recorded wind speed was in state \( i \in \{1, \ldots, 7\} \). As it is possible to see the number of occupancies of state 7 is small compared to all other states and wind speed exceeds 8m/s in very few cases. We stress that the discretization should be chosen according to the database to be used.
We apply our model to a real case of energy production. For this reason we choose a commercial wind turbine, a 10 kW Aricon HAWT with a power curve given in Figure 1. The power curve of a wind turbine represents how much energy it produces as a function of the wind speed. In this case, see Figure 1, there is a cut in speed at 2 m/s, instead the wind turbine produces energy almost linearly from 3 to 10 m/s, then, with increasing wind speed the energy production remains constant until the cut out speed, in which the wind turbine is stopped for structural reason. Then the power curve acts as a filter for the wind speed. In the database used for our analysis the wind speed does never exceed 16 m/s and it is seldom over 8 m/s, this is why the discretization is performed according to Table 1 and the wind never reached the cut out speed.

Through this power curve we can know how much energy is produced as a function of the wind speed at a given time.

| State | Wind speed range m/s | Frequency |
|-------|-----------------------|-----------|
| 1     | 0 to 1                | 1135      |
| 2     | 2                     | 55407     |
| 3     | 3                     | 92659     |
| 4     | 4                     | 49347     |
| 5     | 5                     | 18031     |
| 6     | 6                     | 7397      |
| 7     | 7                     | 3698      |
| 8     | >7                    | 1874      |

Table 1. Wind speed discretization

Figure 1. Power curve of the 10 kW Aricon HAWT wind turbine.
4. Reliability theory for the second order semi-Markov chain in state and duration

In this section, following the research line in Barbu and Limnios [15] and in Blasi, Janssen and Manca [16], we define and compute reliability measures for the second order semi-Markov chain in state and duration. Let $E$ be partitioned into sets $U$ and $D$, so that:

$$E = U \cup D, \emptyset = U \cap D, U \neq \emptyset, U \neq E.$$ 

The subset $U$ contains all ‘Up’ states in which the system is working and subset $D$ all ‘Down’ states in which the system is not working well or has failed. In the wind speed model the Up states are those for which the wind speed is sufficiently high to allow the production of energy or not excessive high such that the turbine should be turned off.

In the following we present both the typical indicators used in reliability theory and also their application. In order to verify the validity of our model, we compare the behaviour of these indicators for real and synthetic data. The indicators of the synthetic data are computed averaging over 500 different trajectories generated through Monte Carlo simulations based on the second order semi-Markov model in state and duration. The number of trajectories is chosen to have stable results.

The three indicators that we evaluate are:

(i) the point wise availability function $A$ giving the probability that the system is working on at time $t$ whatever happens on $(0, t]$.

In our model we denote this function by

$$xA_{i,k}(t) = \mathbb{P}[Z^2(t) \in U|J_{N(0)} = k, J_{N(0)-1} = i, T_{N(0)} - T_{N(0)-1} = x],$$

where $Z^2(t) = J_N(t)$, see relation (5). The availability $xA_{i,k}(t)$ gives the probability that at time $t$ the wind turbine produces energy given that at time zero the wind speed entered state $k$ coming from state $i$ with a duration equal to $x$.

(ii) the reliability function $R$ giving the probability that the system was always working from time 0 to time $t$:

$$xR_{i,k}(t) = \mathbb{P}[Z^2(u) \in U:\forall u \in (0, t]|J_{N(0)} = k, J_{N(0)-1} = i, T_{N(0)} - T_{N(0)-1} = x]$$

The reliability $xR_{i,k}(t)$ gives the probability that the wind turbine will always produce energy from time zero up to time $t$ given that at time zero the wind speed entered state $k$ coming from state $i$ with a duration equal to $x$.

(iii) the maintainability function $M$ giving the probability that the system will leave the set $D$ within the time $t$ being in $D$ at time 0:

$$xM_{i,k}(t) = 1 - \mathbb{P}[Z^2(u) \in D, \forall u \in (0, t]|J_{N(0)} = k, J_{N(0)-1} = i, T_{N(0)} - T_{N(0)-1} = x].$$
The maintainability $xM_{i,k}(t)$ gives the probability that the turbine will produce energy at least once from time zero up to time $t$ given that at time zero the wind speed entered state $k$ coming from state $i$ with a duration equal to $x$.

These three probabilities can be computed in the following way if the process is a second-order semi-Markov chain in state and duration of cumulated kernel $Q = (xQ_{i,k,j}(t))$.

The three indicators, computed and showed in the following figures, are plotted by varying the initial state $i$, the current state $k$ and the sojourn time $x$. The numeric choice of each parameter is given only for graphical reasons, in order to show the maximum number of curves without overlaps. As numeric indicator to compare the gap between the curves we compute the mean square error (MSE) between the indicator applied to the real data and the 500 simulated trajectories. The mean square error is defined as follows:

$$xMSE_{i,k}(t) = \sqrt{\frac{1}{500} \sum_{h=1}^{500} (xI_{i,k}^{real}(t) - xI_{i,k}^{h}(t))^2}$$

where $xI_{i,k}^{real}(t)$ stands for the indicators $xA_{i,k}(t)$, $xR_{i,k}(t)$ or $xM_{i,k}(t)$ estimated form real data and $xI_{i,k}^{h}(t)$ the indicators $xA_{i,k}(t)$, $xR_{i,k}(t)$ or $xM_{i,k}(t)$ estimated from each synthetic trajectory.

(i) the point wise availability function $xA_{i,k}(t)$:

to compute these probabilities it is sufficient to use the following formula:

$$xA_{i,k}(t) = \sum_{h \in E} \sum_{j \in U} x\phi_{i,k;h,j}(t).$$

Figure 2. Comparison of availability functions for real and simulated data.
Reliability measures of second order SMC applied to wind energy production

|        | 10 min | 50 min | 100 min |
|--------|--------|--------|---------|
| $3MSE_{4,3}$ | 0.0216 | 0.0679 | 0.0727  |
| $1MSE_{4,3}$ | 0.0095 | 0.0700 | 0.0876  |
| $1MSE_{2,3}$ | 0.0063 | 0.0180 | 0.0288  |
| $3MSE_{2,3}$ | 0.0106 | 0.0135 | 0.0140  |

Table 2. Mean square error between the curves of the availability applied to real and synthetic data.

In Figure 2 the availability functions of the real and synthetic data are compared. The comparison is made by varying the sojourn time and the starting state. Particularly, in Figure 2 the availability function for two different initial states $i$ is plotted maintaining constant the current state $k$ for two different sojourn times $x$. In Table 2 we show the MSE for each of the curves of Figure 2.

(ii) the reliability function $xR_{i,k}(t)$:
to compute these probabilities, we will now work with another cumulated kernel $\hat{Q} = (x\hat{Q}_{i,k;j}(t))$ for which all the states of the subset $D$ are changed into absorbing states by considering $\forall i \in E$ the following transformation:

$$x\hat{p}_{i,k;j} = \left\{ \begin{array}{ll} xP_{i,k;j} & \text{if } k \in U, \forall j \in E \\ 1 & \text{if } k \in D, j = k \\ 0 & \text{if } k \in D, j \neq k \end{array} \right.$$ (13)

$xR_{i,k}(t)$ is given by solving the evolution equation of a second-order semi-Markov chain in state and duration but now with the cumulated kernel $x\hat{Q}_{i,k;j}(t) = x\hat{p}_{i,k;j} \cdot xG_{i,k;j}(t)$.

The related formula will be:

$$xR_{i,k}(t) = \sum_{h \in U} \sum_{j \in U} x\hat{\phi}_{i,k;h,j}(t)$$ (14)

where $x\hat{\phi}_{i,k;h,j}(t)$ are the transition probabilities of the process with all the states in $D$ that are absorbing, i.e. with cumulated kernel $\hat{Q}$.

Figure 3 shows the reliability functions for real data compared with the simulated ones. The plotting procedure is the same as for the previous figure: we maintain constant the current state $k$ and we vary the initial state $i$ and the sojourn time $x$. The theoretical trend of the reliability function is confirmed, the probability decreases at the increasing of the time interval. A numerical comparison is given in Table 3 in which the MSE of the four curves of Figure 3 is computed.

(iii) the maintainability function $xM_{i,k}(t)$:
to compute these probabilities we will now work with another cumulated kernel $\hat{Q} = (x\hat{Q}_{i,k;j}(t))$ for which all the states of the subset $U$ are changed into absorbing states by
Reliability measures of second order SMC applied to wind energy production

Figure 3. Comparison of reliability functions for real and simulated data

|        | 10 min | 50 min | 100 min |
|--------|--------|--------|---------|
| $1MSE_{4,3}$ | 0.0095 | 0.0198 | 0.0310  |
| $2MSE_{4,3}$ | 0.0216 | 0.0571 | 0.0627  |
| $1MSE_{2,3}$ | 0.0063 | 0.0087 | 0.0116  |
| $2MSE_{2,3}$ | 0.0016 | 0.0097 | 0.0121  |

Table 3. Mean square error between the curves of the reliability applied to real and synthetic data

considering the following transformation:

$$x\hat{p}_{i,k;j} = \begin{cases} 
    xP_{i,k;j} & \text{if } k \in D, \forall j \in E \\
    1 & \text{if } k \in U, j = k \\
    0 & \text{if } k \in U, j \neq k
\end{cases}$$

(15)

$xM_{i,k}(t)$ is given by solving the evolution equation of a second-order semi-Markov chain in state and duration but now with the cumulated kernel $x\tilde{Q}_{i,k;j}(t) = x\tilde{p}_{i,k;j} \cdot xG_{i,k;j}(t)$.

The related formula for the maintainability function will be:

$$xM_{i,k}(t) = \sum_{h \in E} \sum_{j \in U} x\tilde{\phi}_{i,k;h,j}(t)$$

(16)

where $x\tilde{\phi}_{i,k;h,j}(t)$ is the transition probability of the process with all the states in $U$ that are absorbing, i.e. with cumulated kernel $\tilde{Q}$.

The maintainability function is plotted in Figure 4. As the previous figures, this one shows the comparison of the maintainability for the real and simulated data varying
the initial state and sojourn time and maintaining constant the current state. In Table 4 the MSE of the maintainability function for the curves of Figure 4 is computed.

It is possible to note that all the indicators plotted above (availability, reliability and maintainability) depend strongly on the initial and current states and that there is also a great dependence on the sojourn time $x$. In fact, all the probabilities have different values also if only the sojourn time $x$ is changed keeping constant initial states $i$ and final state $k$. For example, from Figure 3 it is possible to see that, in general

$$1R_{4,3}(s) > 3R_{4,3}(s) \forall s \in [0, 100],$$

and in particular $1R_{4,3}(40) = 0.582$ and $3R_{4,3}(40) = 0.473$. This reveals that it is important to dispose of a model that is able to distinguish between these different situations which are determined only from a different duration of permanence in the initial state $i$ before making a transition to the current state $k$. Models based on Markov chains or classical semi-Markov chain are unable to capture this important effect that our second order semi-Markov chain in state and duration reproduces according to the real data.
Reliability measures of second order SMC applied to wind energy production

For each of the indicator we have computed the mean square error between real and synthetic data at different time distance from time zeros. In all cases, the difference between real and simulated data increases with time. The data we are using are 10 minutes sampled, this implies that our model works well for very short time scales and its performance becomes worse for longer time scale. To get better results on a longer time horizon (from more than 1 hour to days) one should use data sampled less frequently but for a longer period. Then, given that the transition probabilities in our model are estimated for high frequency date it is intended for forecasting the probabilities of reliability measures at very short time scale (from 10 minutes to 1 hour).

5. Conclusion

In our previous work, we presented new stochastic models for the generation of synthetic wind speed data. In this work, instead, we compute, for the first time, typical indicators in reliability theory for wind speed phenomenon by using a second order semi-Markov model in state and duration. In order to check the validity of the presented model, we have compared the behaviour of these indicators for real data and data simulated by means of Monte Carlo simulations. To do this, we applied our model to a real case of energy production, filtering real and simulated data by the power curve of a commercial wind turbine. To compare results from real and simulated date we have computed the mean square error between real and synthetic data for each of the indicator. The results show that the proposed model is able to reproduce the behaviour of real data by exhibiting the dependence of the reliability indicators on past visited states and on the length of the sojourn times. This shows that semi-Markov approach is more suitable than simpler Markov chain models. From our results we can also say that our model applied to 10 minutes sampled data works well if one considers short time scale (from 10 minute to 1 hour) and its performance decreases with time.

The indications provided by the model are of importance for assessing the suitability of a location for the wind farm installation as well as for the planning of a preventive maintenance policy.

References

[1] A. Shamshad, M.A. Bawadi, W.M.W. Wan Hussin, T.A. Majid, S.A.M. Sanusi, First and second order Markov chain models for synthetic generation of wind speed time series, *Energy* **30** (2005) 693-708.

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

[3] F. Youcef Ettoumi, H. Sauvageot, A.-E.-H. Adane, Statistical bivariate modeling of wind using first-order Markov chain and Weibull distribution, *Renewable Energy* **28** (2003) 1787-1802.

[4] A.D. Sahin, Z. Sen, First-order Markov chain approach to wind speed modeling, *Journal of Wind Engineering & Industrial Aerodynamics* **89** (2001) 2639.

[5] M.C. Torre MC, P. Poggi, A. Louche, Markovian model for studying wind speed time series in Corsica, *International Journal of Renewable Energy Engineering* **3**(2) (2001) 3119.
[6] G. D’Amico, J. Janssen, and R. Manca, Semi-Markov Reliability Models with Recurrence Times and Credit Rating Applications, *Journal of Applied Mathematics and Decision Sciences* Article ID 625712, (2009) 17 pages.

[7] G. D’Amico, J. Janssen, and R. Manca, European and American options: The semi-Markov case, *Physica A* 388 (2009) 3181-3194.

[8] G. D’Amico, M. Guillen and R. Manca, Full backward nonhomogeneous semi-Markov processes for disability insurance models: a Catalunya real data application, *Insurance: Mathematics and Economics* 45 (2009) 173-179.

[9] J. Janssen and R. Manca, Semi-Markov Risk Models for Finance, Insurance and Reliability, *Springer* (2007).

[10] N. Limnios and G. Oprişan, Semi-Markov Processes and Reliability, Birkhäuser, Boston (2001).

[11] G. D’Amico, F. Petroni, A semi-Markov model with memory for price changes, *Journal of Statistical Mechanics* P12009 (2011).

[12] G. D’Amico, F. Petroni, A semi-Markov model for price returns, *Physica A* 391 (2012) 4867-4876.

[13] G. D’Amico, F. Petroni, Weighted-indexed semi-Markov models for modeling financial returns, *Journal of Statistical Mechanics* P07015 (2012).

[14] G. D’Amico, F. Petroni and F. Prattico, First and second order semi-Markov chains for wind speed modeling, *Physica A* (2012), DOI physa.2012.11.022.

[15] V. Barbu, M. Boussemart and N. Limnios, Discrete Time Semi-Markov Model for Reliability and Survival Analysis, *Communications in Statistics, Theory and Methods* 33 (2004) 2833-2868.

[16] A. Blasi, J. Janssen, R. Manca, Numerical treatment of homogeneous and non-homogeneous semi-Markov reliability models. *Communications in Statistics, Theory and Methods* 33 (2004) 697-714.

[17] N. Limnios, G. Oprişan, An introduction to Semi-Markov Processes with Application to Reliability, In D.N. Shanbhag and C.R. Rao, eds., *Handbook of Statistics* 21 (2003) 515-556.