Non-homogeneous Markov Wind Speed Time Series Model Considering Daily and Seasonal Variation Characteristics

Xie, Liao, Tai and Hu (2017)

Eveline Mathol, Catriona Mitchison, Ana-Maria Niculescu

Introduction

As the planet's energy consumption continues to soar, we are urgently in need of clean alternatives to traditional fuel sources. As a green, renewable and readily available energy source, wind power is considered a major part of the future global energy landscape. To incorporate wind power to a power grid, and preempt power cuts and surges, an accurate model of wind speed is essential. Previous efforts to model wind speed: 1) first order Markov Chain **FOMC**); 2) second order Markov Chain **(SOMC)** and 3) auto-regressive moving average **(ARMA)** fail to take into account both the daily and the seasonal variations in wind intensity. This paper proposes partitioning the wind speed time series into optimal segments; introducing an index to lessen the impact of seasonal variation and modelling the modified wind speeds as a *Non-Homogeneous Markov Chain* with a time variable to accommodate daily variation.

A Homogeneous Markov Chain Wind Model

Studies show wind variation satisfies the necessary Markov "memoryless" property: *the future evolution of wind states depends on the past only through the present*. To apply a Markov chain model, the wind speed stochastic process is segmented into discrete states, each state representing a bounded range of wind speeds. $I = \{i_1, i_2, ...\}$ is the state space, and a state transition occurs when the wind speed exceeds its bounds. Let $V = \{v_t,\ t = 0, 1, 2, ...\}$ be the stochastic wind speed time series. If the following equation

$$P\{v_{t+1} = i_{i+1} | v_1 = i_1, ..., v_t = i_t\} = P\{v_{t-1} = i_{t-1} | v_t = i_t\} \quad (1)$$

holds for any state i_t at any time t, the the states $v_t, t \in T$ are a first order Markov chain process (FOMC). The probability of a state transition at time t is the conditional probability

$$p_{ij}(t) = P\{v_{t+1} = j | v_t = i\}\ i, j \in I. \quad (2)$$

When $p_{ij}(t)$ depends only on state i and j and not on the time instant t, the chain is a homogeneous Markov chain (HMC). Most wind speed models are of this form and therefore cannot incorporate the daily time-dependant variation into their model. The transition probabilities constitute a first order transition probability matrix with elements estimated by

$$p_{ij} = \frac{f_{ij}}{\sum_{j=1}^{K} f_{ij}}, \quad (3)$$

where f_{ij} is the total number of transitions from state i to state j and K is the number of states. As a Markov chain, each element in the matrix is non-negative and rows sum up to one.

Wind Variation Characteristics

Wind speeds were examined at four sites and time homogeneity was analysed. The autocorrelation function for each site (unseen) shows a clear bump at the 24-hour mark, reflecting the cyclical nature of the wind speed time series over the day. The graph below focuses on a specific site, the Crosby site in Texas. It shows a comparison between the autocorrelation function of four different models which all clearly show the autocorrelation function decaying with the time lag, but **only** the NHMC and ARMA models correctly reflect the 24 hour cycle.

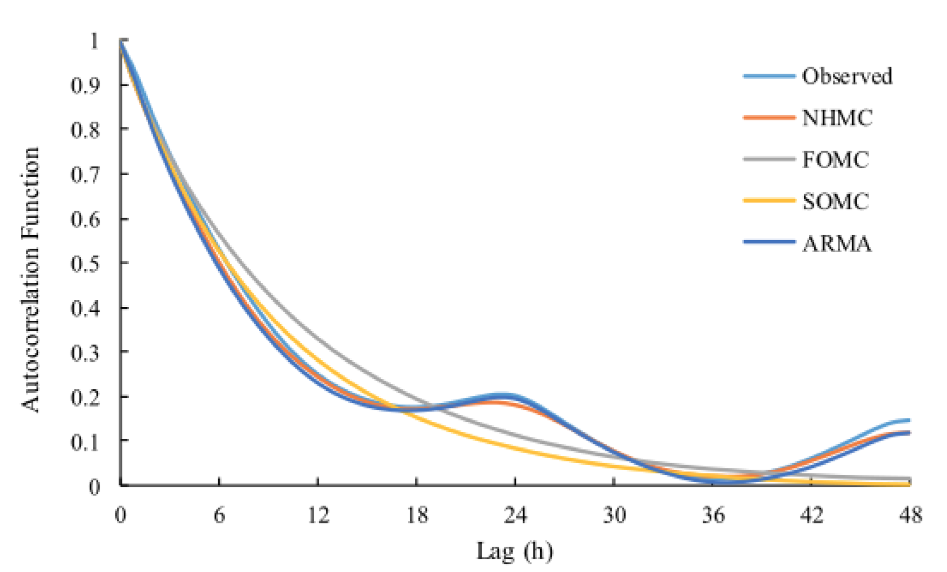

Figure 1: Autocorrelation functions of different models of wind speed time series as functions of time lags.

A Non-Homogeneous Markov Chain Wind Model

A non-homogeneous Markov chain (NHMC) wind speed model is developed to accurately characterize wind speed variation features at different time scales. This model takes both the seasonal effect and the daily variation into account. The NHMC model is illustrated in Figure 2.

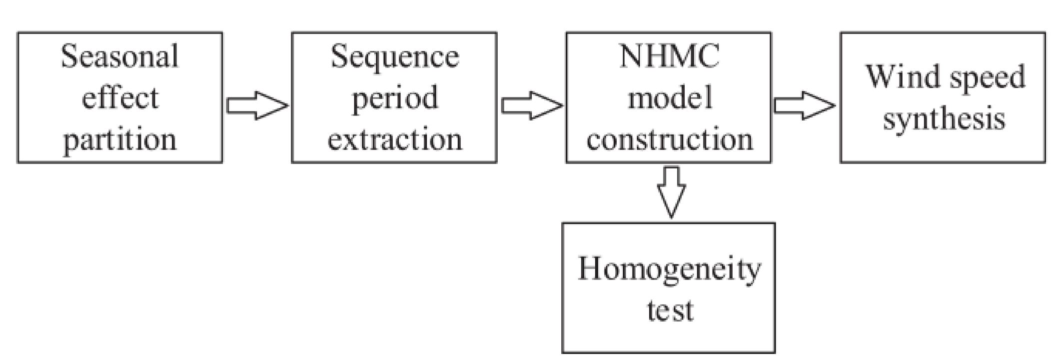

Figure 2: Non-homogeneous Markov chain wind model.

The first step of the NHMC model is to partition the wind speed sequence into several segments affected by seasonal change, which are regarded as trends in the data. For this, Fisher's optimum partition method is used [1]. The partition is done by minimizing the variance within the segment and maximizing the difference between segments.

Secondly, sequence period extraction is performed. The Fourier transform technique is applied to obtain the potential period of wind variation. The frequency corresponding to the maximum amplitude component reveals the wind speed variation [2],[3]. In the paper, all four spectra for the different sites have the peak at 11.57 μHz, which corresponds to the 24-hour period. This shows that there exists a significant daily cyclical variation feature.

After obtaining the partition from Fisher's optimum partition method, the seasonal index $S(k)$ is constructed which is used to eliminate the seasonal effect from the wind speed. The formulas below give the seasonal index $S(k)$ and the modified wind speed $v'(t)$.

$$S(k) = \frac{\bar{v}_k}{\bar{v}} \quad (4) \qquad v'(t) = \frac{v(t)}{S(k_t)}. \quad (5)$$

Thus, the seasonal index is the ratio of $\bar{v}_k$ and $\bar{v}$, the mean values of the kth segment and the whole year respectively. In (5), $v(t)$ and $v'(t)$ are the actual and modified wind speeds at time t respectively and k_t is the segment number corresponding to time instant t. Different from the HMC model, in NHMC the transition state probabilities are considered as variables related to time. Define the time-related variable $l = 0, 1, ..., R-1$ and time set

$$T_l = \{t | mod(t, R) = l\}, \quad (6)$$

where $R = 24$ represents the wind variation period. The time in T_l experiences the same periodic variation as the wind speed sequence. Transition probabilities can now be estimated by rewriting (3) as

$$p_{ij}^l = \frac{f_{ij}^l}{\sum_{t \in T_l} f_{ij}^l}, \quad (7)$$

where f_{ij}^l is the number of transitions from state i to state j for the modified wind data of T_l. These transition probabilities constitute the transition probability matrix P^l. The cumulative transition probabilities, which constitute the cumulative transition matrix P_c^l, can be calculated by adding the transition probability of the corresponding row in order:

$$p_{c,ij}^l = \sum_{k=1}^{j} p_{ik}^l. \quad (8)$$

The last step of the NHMC model is to synthesize the wind speed sequence. First, the seasonal index is calculated by formula (4) and the wind series is modified modify by using (5). Then, formulas (6)-(8) are estimated. The procedure of sampling a wind time series can then be summarized as follows.

1. Randomly generate a modified wind speed $V'(0)$ at the initial time;
2. Choose P_c^l corresponding to the current time and row i of P_c^l corresponding to the wind speed state at that time;
3. Generate a uniformly distributed variable $r \in (0, 1)$. If r is larger than the first $j-1$ cumulative probabilities but smaller than the jth one, the wind speed state is assigned at the next moment to be state j;
4. Regenerate a uniformly distributed $r' \in (0, 1)$, calculate the modified wind speed at the next moment: $V'(t) = V_{dj} + r'(V_{uj} - V_{dj})$, where V_{uj} and V_{dj} are the upper and lower bounds of state j, respectively;
5. Continue to synthesize the wind speed from step 1) until the required length is met;
6. Transform the modified wind speed into the synthetic wind speed using $V(t) = V'(t)S(k_t)$.

Homogeneity test

A homogeneity test is performed to verify the time homogeneity of wind speed time series. The null hypothesis, $\mathbf{H}_0$, of wind speed variation independent of time (homogeneous) is tested against the alternative hypothesis, $\mathbf{H}_1$, of variation that is time-dependent (non-homogeneous) according to the following χ^2 test statistic

$$\chi^2 = 2 \sum_{l=0}^{R-1} \sum_{i=1}^{K} \sum_{j=1}^{K} f_{ij}^l \left| ln \frac{p_{ij}^l}{p_{ij}} \right|. \quad (9)$$

This test corresponds to performing a Likelihood-ratio test of the form $LR = 2log \frac{L_1(x|\hat{P}_0, ..., \hat{P}_{23})}{L_0(x|\hat{P})}$, where L_1 and L_0 represent the likelihood functions under $\mathbf{H}_1$ and $\mathbf{H}_0$ respectively. $\hat{P}$ and $(\hat{P}_0, ..., \hat{P}_{23})$ represent the Likelihood Maximization Expectations (LMEs). In formula (9), the LME under $\mathbf{H}_0$ is p_{ij}, the LME under $\mathbf{H}_1$ for time l is p_{ij}^l and f_{ij}^l is the total number of transitions from state i to state j at time l.

Case Study: Performance Analysis

Data extracted from online NDAWN database [4] is used in order to illustrate the performance of NHMC. Wind speed values for four sites were recorded, but for simplicity purposes Crosby site is used as the main example. Performance of the NHMC is compared to the performance of FOMC, SOMC and ARMA models and they are evaluated in three steps: homogeneity test, performance analysis to examine fitness and state partition effect analysis.

1) Homogeneity test reveals that the χ^2 statistics calculated for each site are larger than the critical values and the $\mathbf{H}_0$ is rejected. That means that the variation of wind speed is a function of time, therefore FOMC and SOMC no longer represent a good approach.

2) Model Accuracy Analysis is performed by evaluating the following metrics:

a) Monthly averages, which illustrate the seasonal variation. Since Fisher's partition method is used, the NHMC model can capture the seasonal variation with smallest differences between synthesized and observed wind speed data.

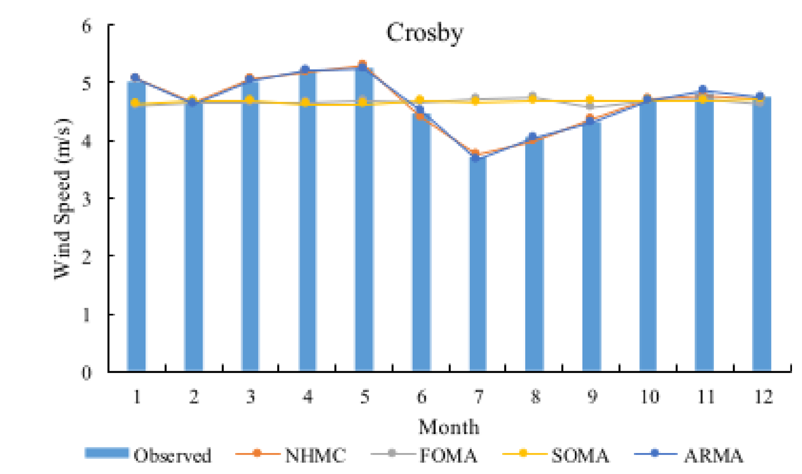

Figure 3: Monthly-average wind speed synthesized by different wind speed models.

b) Probability distributions synthesized by NHMC, unlike those synthesized by ARMA, are consistent with the observed distribution, having the largest R^2 and smallest Root Mean Squared Error ($RMSE$) and Power Density Error (PDE). Moreover, as presented in Table 1, $RMSE_{\text{CDF}}$ and PDE of the ARMA model are much larger than the values corresponding to MC models. This means that ARMA data deviates more from the observed model than MC.

Table 1: Statistics of synthetic wind speed of various models.

Site	Model	Mean Speed (m/s)	Standard deviation (m/s)	R^2	$RMSE_{\text{CDF}}$	PDE
Crosby	Observed	4.6733	2.2906			
	FOMC	4.6710	2.3239	0.999943	0.0024	0.0100
	SOMC	4.6751	2.3292	0.999924	0.0028	0.0123
	ARMA	4.6749	2.2965	0.997955	0.0148	-0.0275
	NHMC	4.6706	2.3279	0.999963	0.0020	0.0145

c) Autocorrelation function shows that the wind speed modeled by NHMC and ARMA preserve autocorrelation features, including decaying trend and relatively strong correlation between wind speeds at 24-hour interval. This aspect is also confirmed by the Power Spectral Density function, which presents a peak value at 11.57 μHz, corresponding to a 24-hour period. NHMC outperforms other models with the smallest values for $RMSE_{\text{ACF}}$.

3) Influence of State Number is a considerable factor for the accuracy of wind speed series based on MC as well as NHMC. Small state interval implies a large number of states. When the state interval increases, $RMSE_{\text{CDF}}$ and PDE increase while R^2 decreases. NHMC matches wind speed series better with small intervals.

Conclusion

The NHMC model is designed to incorporate both the seasonal and the daily wind speed variation characteristics. An optimal partition method is adopted to enable the extraction and reduction of wind speed variation components affected by seasonal change. The time-related state transition probability matrices are constructed to accurately represent the daily cyclical variation of wind speed caused by the difference of changing regularity within one day. A hypothesis test method is also introduced to verify the non-homogeneous feature of the wind speed sequence.

Case studies have demonstrated that the wind speed time series synthesized by NHMC fit well not only on the probability distribution and statistics, but also on the daily and seasonal variation features of historical wind data.

The proposed NHMC wind model can be adopted in wind farm output analysis as well as in the long-term scheduling and planning of power systems with the wind farm.

Bibliography

[1] W.D. Fisher, "On grouping for maximum homogeneity," *Amer. Stat. Assoc. J.*. vol. 53, pp. 789-798, 1958.

[2] A.C. Da Silva, A.B. Rodrigues, and M.D.G. Da Silva, "Probabilistic evaluation of long-duration voltage variations in distribution networks with wind power plants." *IET Gener. Transm. Dis.*, vol.9, no. 13, pp. 1526-1533, 2015.

[3] J. Apt. "The Spectrum of power from wind turbines," *Power Sources J.*, vol. 169, pp. 369-374, 2007.

[4] North Dakota Agricultural Weather Network (NDAWN)[Online]. Available: http://ndawn.ndsu.nodak.edu/weather-data-hourly.html