The 2019 World Congress on
Advances in Structural Engineering and Mechanics (ASEM19)
Jeju Island, Korea, September 17 - 21, 2019

Short Term Wind Speed Prediction Based on EMD and ARIMA

*Wu Shengzhi[1)] and Zhang Quan[2)]

[1), 2)] Key Laboratory of Mechanics on Disaster and Environment in Western China, The Ministry of Education,
School of Civil Engineering and Mechanics, Lanzhou University, Lanzhou, 730000, P. R. China
[1)] wusz@lzu.edu.cn

ABSTRACT

This paper presents a hybrid wind speed prediction modeling based on empirical mode decomposition (EMD) and auto-regressive integrated moving average (ARIMA). The air motion in atmospheric boundary layer is often turbulent, irregular and highly nonlinear, therefore a practicable description of wind speed is of importance in engineering application. In this study, we firstly extend traditional ARIMA model to multi-step ARIMA model for further application. Then, the EMD technique is employed to reduce the influence of noise within original wind data and obtain a more stable wind data sequences for ARIMA model. The whole hybrid process consists of three steps： data preprocessing, data grouping and forecasting. (1) The original non-linear and non-stationary wind speed sequences are decomposed into a series of intrinsic mode functions (IMFs) by using EMD technique. (2) The IMFs are divided into: high frequency, intermediate frequency, low frequency components according to their run test. (3) ARIMA is used to predict three sub- sequences and superposed to produce the final predicted wind speed. The numerical results indicate that the predicted values are in good agreement with the experimental measurements, and more accurate and efficient than the results obtained by ARIMA without EMD.

1. INTRODUCTION

Sandstorm is one aspect of a far wider windblown-sand / dust environmental problem, which is still questionable. Accurate wind speed forecasts are necessary to predict and control sandstorms. There are a lot of forecasting models of wind speed and power generation developed in literature (Bao, 2018; Kavasseri, Seetharaman, 2019). In general, these models can be grouped into two categories, namely, (1) numerical weather prediction based models, (2) statistical models - as stochastic time series models. Time series models employ a general class of models such as auto-regressive moving average (ARMA) or auto-regressive integrated moving average (ARIMA) (Box, et al, 2008). However, recent studies suggest that the ARMA and ARIMA models have a feature of time delay because the air motion is a non-stationary

---

[1)] Professor
[2)] Graduate Student

and nonlinear physical process in nature (He, et al, 2018). Huang et al. (1998) proposed empirical mode decomposition (EMD) to decompose any complicated data set into a set of simple components that are defined as intrinsic mode functions (IMFs). This decomposition method is adaptive, and highly efficient, thus widely used in engineering.

The aim of this paper is to develop a hybrid system based on ARIMA and EMD and to analyze the effectiveness of the proposed model in the short term wind speed forecasting for the selected specific wind speed records in field observation. The reminder of this paper is organized as follows. The main mathematical modeling in time series analysis is introduced in section 2, one is relating to the autoregressive integrated moving average, and the other is for the empirical mode decomposition. The performance and evidence of above technique are provided in section 3, including the short term wind speed predication and comparison with field observation results. The conclusions are given in section 4.

2. MATHEMATICAL MODELING

2.1 ARIMA model formulation

Let $\{x_t\}$ $(t=1,2,3,\cdots)$ represents the time series of average wind speed. Then, an ARIMA formulation for the series can be described by

$$\varphi(B)(1-B)^d x_t = \theta(B)\varepsilon_t \tag{1}$$

where $\varepsilon_t \sim N(0,\sigma^2)$ is white noise series, d is the differencing parameter, $B(x_t)=x_{t-1}$ is the backshift operator, φ, θ are polynomial functions of the backshift operator B given by

$$\varphi(B)=1-\varphi_1 B-\varphi_2 B^2-\cdots-\varphi_p B^p \tag{2}$$

$$\theta(B)=1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q \tag{3}$$

Therefore, the ARIMA model is completely described by p parameters $\varphi_1,\varphi_2,\cdots,\varphi_p$ in Eq.(2), q parameters $\theta_1,\theta_2,\cdots,\theta_q$ in Eq.(3) and the differencing parameter d.

When the structure of the model $\mathrm{ARIMA}(p,d,q)$ is defined, the estimation of parameters in the model can be performed by various techniques. Here, we choose the least AIC (Akaike's Information Criterion),

$$\mathrm{AIC}=-2\ln L+2p \tag{4}$$

where p is the number of parameters in the model and $\ln L$ is the natural logarithm of the maximum likelihood.

2.2 EMD method performance

As described by Huang et al (1998), the sifting process of EMD components can be briefly summarized in the following outline. First, we find all extreme points on the original time series x_t; then, connect all local maxima by a cubic spline curve as the upper envelop $x_{\max}(t)$, all local minima as the lower envelop $x_{\min}(t)$. Their mean is expressed as $m_1(t)$,

$$m_1(t) = \frac{x_{\max}(t) + x_{\min}(t)}{2} \tag{5}$$

The difference between the original data and the mean value $m_1(t)$ is the first component $h_1(t)$, which may be an intrinsic mode function component, nor intrinsic mode function component. If it does not meet the IMF's requirement, we shall see $h_1(t)$ as the original time series, and repeat above procedures k times until meeting the given criterion (Huang, et al, 1998) :

$$\mathrm{SD} = \sum_{t=0}^{T} \frac{[h_{1(k-1)}(t) - h_{1k}(t)]^2}{h_{1(k-1)}^2(t)} \tag{6}$$

A typical value for SD can be set between 0.2 and 0.3.

Therefore, the first intrinsic mode function (IMF) can be obtained as

$$c_1(t) = h_{1k}(t) \tag{7}$$

Then, we subtract $c_1(t)$ from the original data by

$$x(t) - c_1(t) = r_1(t) \tag{8}$$

The residue $r_1(t)$ will be seen as the new data and subjected to the same decomposition process as described above. This procedure can be repeated on all the subsequent $r_i(t)$, and the results appear as shown below

$$r_1(t) - c_2(t) = r_2(t),\ \cdots,\ r_{n-1}(t) - c_n(t) = r_n(t) \tag{9}$$

The decomposition process can be stopped by any of the following criteria: either when the component $c_n(t)$ or the residue $r_n(t)$ is less than a predetermined tolerance, or when the residue $r_n(t)$ becomes a monotonic function.

By summing up equations (8) and (9), we finally obtain

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t) \tag{10}$$

Thus, we have a decomposition of the data into n-empirical modes $c_i(t)\ (i = 1,2,\cdots,n)$, and a residue $r_n(t)$ which can be either the mean trend or a constant.

2.3 Hybrid system of ARIMA and EMD

In order to improve the accuracy of wind speed forecast, we propose a hybrid model by combining the ARIMA model with EMD model. First, pre-process the original wind data collected using EMD method. Next, use the new EMD data instead of the original data in ARIMA modeling. Then, investigate and analyze the numerical results.

3. MAIN RESULTS

To illustrate the time series analysis, we shall use a set of wind speed data collected in a field observation station in Qintu Lake, China. The sampling frequency is 50Hz. The original horizontal wind speed is displayed in Figure 1. Clearly, the data are quite complicated with many local extreme points but no zero crossings, for the time series represents all positive numbers. Figure 2 shows the autocorrelation coefficients of wind speed. It is found that the original speed data are non-stationary , but in an order difference after , they are stationary sequences.

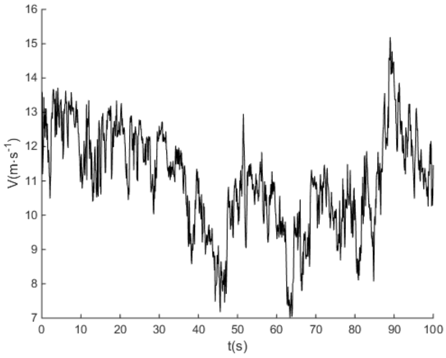

Figure 1. Original wind speed with time from a field observation

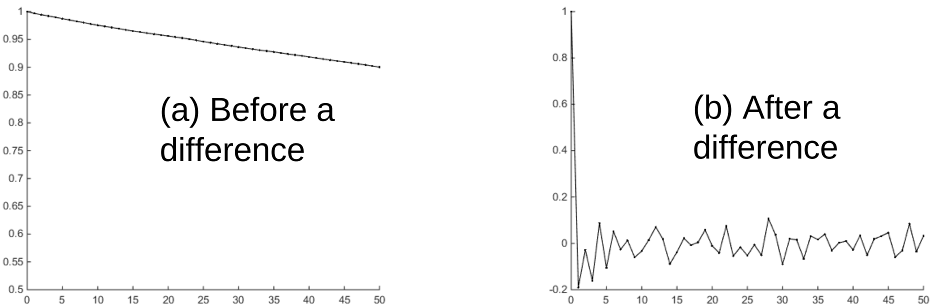

Figure 2. Autocorrelation coefficients of wind speed before and after one order difference

Now , let $\{x_t\}$ $(t=1,2,3,\cdots)$ be original wind speed , $\{x_t'\}$ $(t=1,2,3,\cdots)$ be new wind speed after an order difference. According to AIC criterion, $\mathrm{ARIMA}(p,d,q)$ is defined as $\mathrm{ARIMA}(6,1,0)$, that is $p=6, d=1, q=0$ in ARIMA model.

Assumed the training samples are from the wind speed data in first 40 seconds, we can find the equation that define ARIMA process

$$\begin{aligned}\hat{x}(t) &= 0.7885x(t-1)+0.0873x(t-2)-0.0715x(t-3)-0.1862x(t-4)\\ &\quad -0.1044x(t-5)+0.1138x(t-6)+\varepsilon_t\end{aligned} \tag{11}$$

By using equation (11), we can obtain the predicted wind speed for times 40s through 60s.

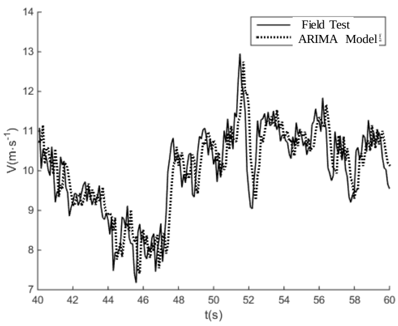

Figure 3. Comparsion of predicted wind speeds with field measured values

It is clear that the process trend are quite familiar to each other, bur there exists a distinct time delay in ARIMA model.

In a similar manner, a multi-step wind speed prediction formula will come in handy as follows

$$\begin{aligned}\hat{x}(t+1) &= 0.7885x(t)+0.0873x(t-1)-0.0715x(t-2)-0.1862x(t-3)\\ &\quad -0.1044x(t-4)+0.1138x(t-5)+\varepsilon_t\end{aligned} \tag{12}$$

$$\begin{aligned}\hat{x}(t+2) &= 0.7885x(t+1)+0.0873x(t)-0.0715x(t-1)-0.1862x(t-2)\\ &\quad -0.1044x(t-3)+0.1138x(t-4)+\varepsilon_t\end{aligned} \tag{13}$$

An example is given in Figure 4 to show 1-step and 3-step wind speed predictions to equation (11) , (13).

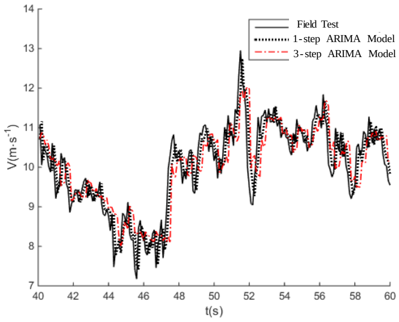

Figure 4. Comparsion of multi-step predicted wind speeds with field measured values

Note that the maximum error for the forecast value is about 12.75% for 3-step prediction model. The suggested order of multi-step is 3 in this case.

Finally, we shall give a typical example to demonstrate the use of hybrid model of ARIMA with EMD components.

Firstly, the original wind data records in Figure 1 can be decomposed into 7 components as displayed in Figure 5.

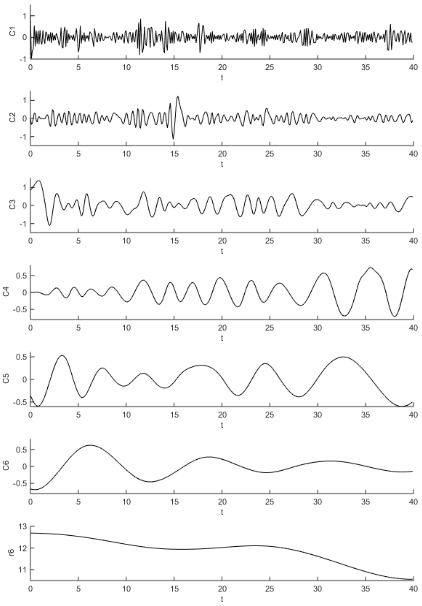

Figure 5. The resulting empirical mode decomposition components c_1-c_6, and r_6 from the wind data. Notice that r_6 is not an IMF, it is the trend.

Then, According to the run test, c_1-c_4 are high frequency components, c_5-c_6 are medium frequency components, r_6 is low frequency components. Therefore, three prediction values corresponding to above three groups of IMFs' components are shown in Fig. 6.

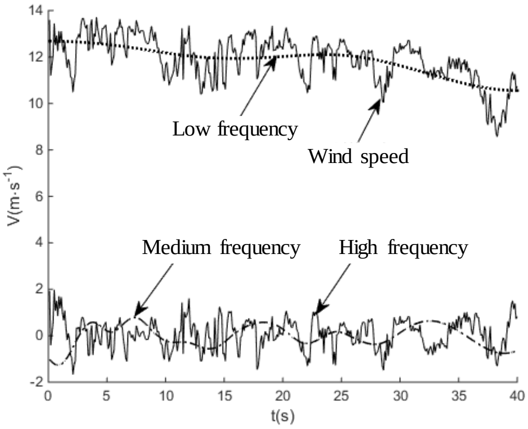

Figure 6. The reconstruction of three groups of decomposed components based on their frequencies

Now, we sum up the high frequency, intermediate frequency, low frequency components, and obtain the final predicted wind speed values by ARIMA with EMD as illustrated in Fig.7.

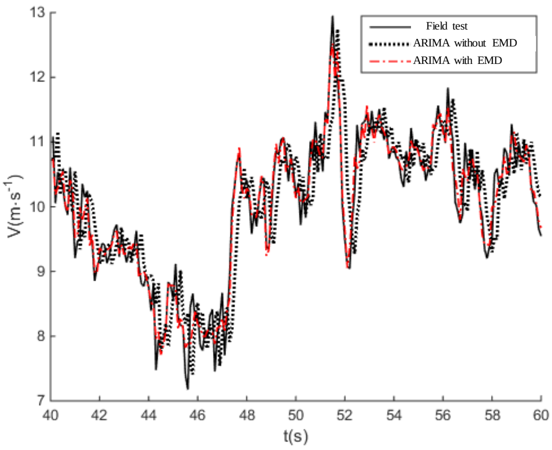

Figure 7. Numerical wind speed versus experimental results

We can see that the accuracy of hybrid model is much better than that of pure ARIMA model, and the time delay disappears.

4. CONCLUDING REMARKS

This paper has developed an extended ARIMA model with EMD components. Some examples are given to illustrate the use of the proposed modeling. The numerical results are in good agreement with field measurements. It is found the proposed technique is suitable for the short term wind speed prediction, and may overcome the shortcoming of time delay on traditional ARIMA model.

ACKNOWLEDGMENTS

This work has been supported by National Natural Science Foundation of China under Grant No. 11472121.

REFERENCES

Bao Yong. (2018), A general result on the estimation bias of ARMA models [J]. *Journal of Statistical Planning and Inference*, 197: 107-125.

Box, G. E. P., Jenkins, G. M. Reinsel, G.C. (2008), *Time series analysis: forecasting and control.* Fourth edition. John Wiley & Sons, Inc.

He, Q. Q., Wang, J. Z., Lu, H. Y. (2018) A hybrid system for short-term wind speed forecasting, *Applied Energy*, 226: 756–771.

Huang, E. Zheng Shen, Steven Long, R. et al. (1998), The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis [J]. *Proc. R. Soc. Lond. A*, 454: 903–995.

Kavasseri, R., Seetharaman, K. (2019), Day-ahead wind speed forecasting using *f*-ARIMA models. *Renewable Energy*, 34: 1388-1393.