ISEIS

Journal of Environmental Informatics 32(1) 1-13 (2018)

Journal of Environmental Informatics

www.iseis.org/jei

Prediction of the Standard Deviation of Wind Speed Turbulence

G. R. Ren[1, 2], J. F. Liu[1*], J. Wan[1], Y. F. Guo[1], Q. H. Hu[3], and D. R. Yu[1*]

[1] *School of Energy Science and Engineering, Harbin Institute of Technology, Harbin, Heilongjiang, China;*
[2] *Department of Electric Power and Energy Systems, KTH Royal Institute of Technology, Stockholm, Sweden*
[3] *School of Computer Science and Technology, Tianjin University, Tianjin, China*

Received 22 December 2015; revised 31 December 2017; accepted 30 January 2018; published online 16 May 2018

ABSTRACT. Turbulent wind speed is a vital component of wind speed, corresponding to the small-scale uncertainty of wind speed. The wind power output and aerodynamic loads are significantly influenced by turbulent wind speed. The present paper concentrates on the study of turbulent wind speed prediction, which is hardly addressed in previous studies and is a novel approach to understand turbulent wind speed in advance. In the present study, turbulent wind speed is measured by turbulence standard deviation. Although turbulent wind speed itself is unpredictable, the correlation analysis proves that the turbulence standard deviation is predictable within a proper time horizon. As a result, the turbulent wind speed can be understood ahead of time by predicting turbulence standard deviation. The 10-min ahead, 30-min ahead and 60-min ahead predictions of turbulence standard deviation are provided by Support Vector Regression and Kernel Ridge Regression. Furthermore, the average wind speed is fused into the forecasting model to improve the prediction accuracy of turbulence standard deviation.

Keywords: Wind speed, Turbulence, Uncertainty, Standard deviation, Predictability, Correlation analysis

1. Introduction

With the increasing concern about energy shortage and environmental pollution problems, the development of renewable energy has attracted more and more attention all over the world. Due to the highly mature technology of wind power, wind power has been rapidly developed in recent years. The global installed capacity of wind power is 486.66 GW by the end of 2016. Compared with traditional power generation, e.g. thermal power plant, wind power output is highly uncertain due to the random fluctuation of wind speed. The high uncertainty of wind power would lead to the imbalance between power generation and load demand, which has been a major barrier for large-scale wind power penetration. In order to cope with wind power uncertainty effectively, an accurate and reliable wind speed forecasting method is desired (Abdel-Aal et al., 2009). Therefore, wind speed forecasting has attracted attention of many researchers in recent years.

In general, there are mainly two methods for wind speed forecasting: Numerical Weather Prediction (NWP) method and data-driven method. NWP method usually utilizes the atmospheric parameters (e.g., pressure, temperature and humidity) to establish a set of physical equations. Then wind speed prediction can be obtained by solving these equations. The classic NWP models include the High Resolution Limited Area Model (HIRLAM) (Gustafsson, 1993), the Weather Research and Forecasting (WRF) model (Skamarock et al., 2005), the UK Meteorological Office Mesoscale (MESO) model (Wilson et al., 1999) and so on. The advantage of NWP is its long-term forecasting horizon (Liu et al., 2012). However, the grid resolution is coarse due to its expensive computational cost (Kuik et al., 2007; Kusiak, 2009). Data-driven method utilizes statistical models to predict wind speed based on historical data. These statistical methods include autoregressive moving average model (Gomes et al., 2012), Kalman filtering (Cassola et al., 2012), artificial neural networks (Li et al., 2010), support vector regression (Santamaría-Bonfil et al., 2016) and so on. Compared with NWP, the short-term local-area wind speed forecasting is the superiority of data-driven method.

According to the power spectrum of atmospheric boundary layer (Van, 1957; Burton et al., 2011), the actual wind speed consists of two parts. The first part is average wind speed, and the other is turbulent wind speed. The outputs of NWP and data-driven methods are usually point forecasting with 15-min resolution, corresponding to the hourly average wind speed. However, the turbulent wind speed has significant impact on wind power utilization as well, including wind power output and blade aerodynamic loads.

The average wind speed and air density are considered as the principal influence factors of wind power curve in the current IEC standard (IEC, 2005). However, various studies have revealed that the wind power output is significantly influenced

* Corresponding author. Tel.: +86 13804578476; fax: 0451-86403142
E-mail address: yudaren@hit.edu.cn (D. Yu), Jinfuliu@hit.edu.cn (J. Liu)

ISSN: 1726-2135 print/1684-8799 online
 doi:10.3808/jei.201800389

by the turbulent wind speed (Sheinman and Rosen, 1992; Rosen and Sheinman, 1994; Gottschall and Peinke, 2008). With the increase of turbulence intensity, the wind power output would be overestimated at moderate wind speed and be underestimated at rated wind speed (Kaiser et al., 2007). For an offshore vertical axis wind turbine, the wind power output decreased by 23 to 42% compared with the case of no turbulence when the turbulence intensity increased from 5 to 25% (Siddiqui et al., 2015). In addition, the blade aerodynamic loads are significantly influenced by the turbulent wind speed. A study has proved that the turbulent wind speed is the major cause of the fatigue damage (Kelley et al., 2000). The fatigue loads on wind turbine blades increased due to the turbulent wind speed (Hand et al., 2003). Another study has revealed that the intense turbulent wind speed would lead to high fatigue loads on wind turbine blade although the average wind speed is low (Kim et al., 2015). The turbulent wind speed would lead to extreme load events, which increased the risk of breakdown (Lavely et al., 2011; Carpman, 2011). In summary, the impacts of turbulent wind speed are various, including the design and performance of wind turbines, the quality of power delivered to the network and its effect on consumers (Burton et al., 2011).

Considering its significant impacts, the study of turbulent wind speed is of vital significance for wind power utilization as well. Therefore, some researchers focused on the study of turbulent wind speed. The turbulence standard deviation and turbulence intensity are usually used to characterize turbulent wind speed. A study found that there exists a linear relationship between the turbulence standard deviation and the average wind speed (Welfonder, 1997). The ratio of the turbulence standard deviation to the average wind speed is defined as turbulence intensity (TI), which is an important parameter for the design and selection of wind turbines (IEC, 2005; Freudenreich, 2006). The IEC edition 2 proposes a normal turbulence model for wind turbine designing (IEC, 1998), including high turbulence level and low turbulence level. The latest IEC extends the model to three classes (IEC, 2005).

Some studies compared the IEC model with the actual TI. By studying the actual TI observations at different sites, the results revealed that the IEC model may be more conservative when the wind speed exceeds 15 m/s (Hansen and Larsen, 2005). Another study concluded that the IEC model underestimated the actual TI for high wind speed and complex surroundings (Carpman, 2011). Due to the layout of wind turbine, the wake of upstream wind turbine has a significant impact on the TI of downstream wind turbine. So the impacts of wake effect on turbulence intensity were studied (Frandse and Thøgersen, 1999; Jimenez et al., 2007; Sørensen et al., 2008; Chamorro and Fernando, 2009; Wu and Fernando, 2012).

To the best of the author's knowledge, the forecasting of turbulent wind speed is rarely addressed in previous studies. The forecasting of turbulent wind speed plays an important role for wind power utilization. Firstly, the accuracy of wind power curve is of vital importance to wind power simulation. A more accurate wind power curve can be obtained by considering the correction of turbulent wind speed (Anahua et al., 2008; Gottschall and Peinke, 2008). Secondly, the health management system of wind turbine has attracted a lot of interests in recent years. The health management system contributes to the secure and economical operation of wind turbines. As mentioned above, the blade aerodynamic loads are significantly influenced by the turbulent wind speed. If the turbulent wind speed can be understood ahead of time, effective control strategies would be adopted to reduce the aerodynamic loads. Consequently, the reliability, safety, and availability of wind turbines would be improved. Therefore, the present paper focuses on the study of turbulent wind speed prediction.

The remainder of the paper is organised as follows. Section 2 introduces the intrinsic nature of wind speed uncertainty and its description. Section 3 analyzes the predictability of turbulence standard deviation. Section 4 conducts the forecasting experiments. Section 5 concludes the paper.

2. The Intrinsic Nature of Wind Speed Uncertainty and Its Description

2.1. The Intrinsic Nature of Wind Speed Uncertainty

Wind power is proportional to the cube of wind speed. Therefore, understanding the characteristics of wind speed is critical to wind power utilization (Burton et al., 2011). Uncertainty is the intrinsic nature of wind speed, which always exists on different time scales. The wind speed spectrum proposed by Van displays clear distinctions between different time scales (Van, 1957), which is shown in Figure 1. There are three obvious peaks in Figure 1. The first peak is called as synoptic peak and the corresponding time scale is approximately 4 days. The second peak is diurnal peak and the time scale is about 24 hours. The wind speed corresponding to these time scales is fairly predictable, which is the research object of current wind speed forecasting studies. The third peak is turbulent peak on the time scale of minutes or seconds, which is known as turbulence. Turbulent wind speed fluctuates with a relatively high frequency. Furthermore, the turbulent wind speed itself is unpredictable. There is an obvious spectral gap between the diurnal peak and turbulent peak. The energy contained in the spectral gap is fairly small. As a result, the actual wind speed is considered as the superimposition of average wind speed and tur-

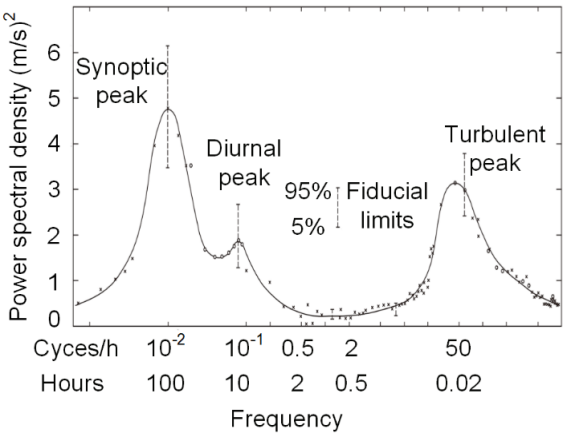

Figure 1. Wind spectrum proposed by Van der Hoven.

bulent wind speed. Both of the two components are highly uncertain, resulting in important impacts on wind power utilization. As mentioned above, the wind speed uncertainty on large time scale has been widely studied by forecasting. However, the prediction study of uncertainty on small time scale is fairly rare, which is the research object of the present study.

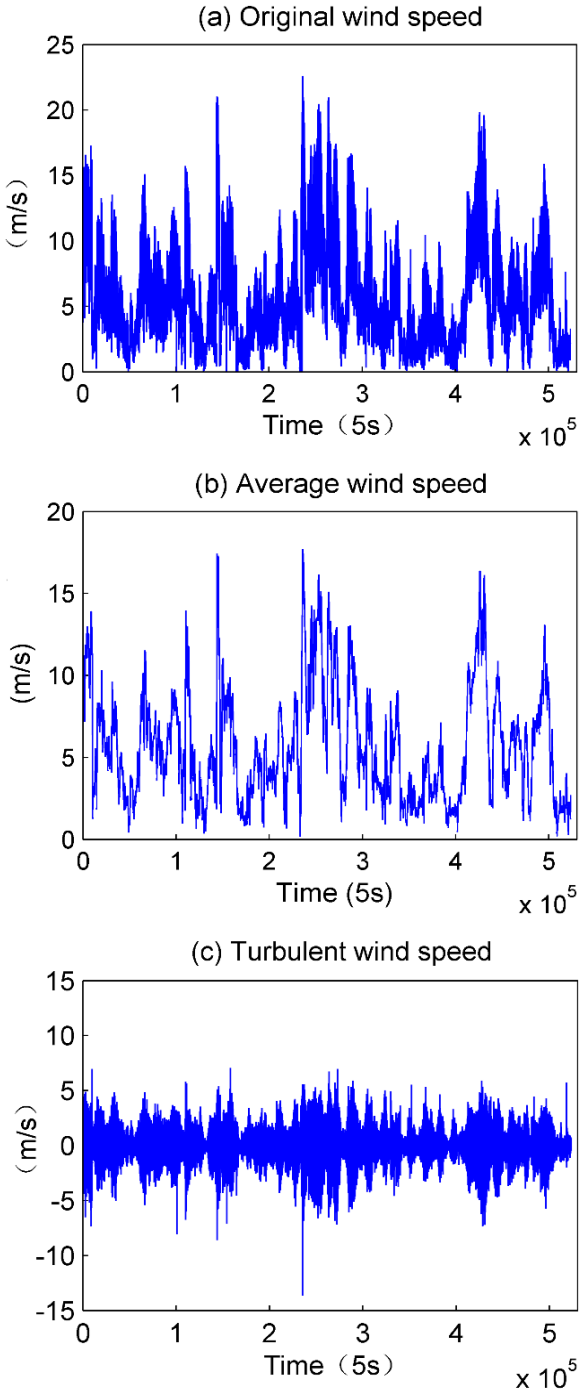

Figure 2. The decomposition results of wind speed for WF1 on August 2012.

2.2. The General Description Method of Turbulent Wind Speed

Turbulence is an extremely complex fluid phenomenon with high randomness, which is difficult to describe. The Reynolds-average method is usually used to study turbulence (Von Karman, 1948; Sagaut et al., 2006). According to Reynolds-average method, the flow variable is usually decomposed into the mean component and the fluctuating component. For the actual wind speed, decomposing it into average wind speed and turbulent wind speed is reasonable due to the existence of spectral gap.

Suppose $\{v_i\}$, $i = 1, 2, \ldots, t$ is the actual raw wind speed time series, the average wind speed $\bar{v}$ can be expressed as:

$$\bar{v} = \frac{1}{T}\sum_{i=1}^{T} v_i \tag{1}$$

where T is the time window, e.g., 30-min or 1-hour.

Then the turbulent wind speed v_i' can be isolated from the raw wind speed time series:

$$v_i' = v_i - \bar{v} \tag{2}$$

Turbulence standard deviation is usually used to measure turbulent component (Stull, 2012). Furthermore, the turbulence standard deviation is an important parameter for wind turbine designing. So the turbulence standard deviation is utilized to measure turbulent wind speed in the study. The turbulence standard deviation σ is given as:

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(v_i')^2} \tag{3}$$

Then the detailed calculations are conducted based on the wind speed data. The wind speed data studied in the present study is collected from two wind farms in China. The first wind farm is located in Inner Mongolia Autonomous Region (denoted as WF1) and the second is located in Heilongjiang province (denoted as WF2). The data is collected from August 2012 to July 2013 and the resolution is 5 s. Wavelet decomposition has been widely used to decompose wind speed (Rahmani, 2015). The raw wind speed is decomposed into a low frequency component and a high frequency component by wavelet algorithm. The low frequency component corresponds to the average wind speed and the high frequency component is turbulent wind speed. The decomposition results for wind farm 1 on August 2012 are shown in Figure 2. The relevant turbulence standard deviation is shown in Figure 3.

It is obvious that the curve profile of average wind speed in Figure 2b is highly similar to that of turbulence standard deviation in Figure 3b. The detailed comparison result is shown in Figure 4. An obvious dependency relationship between average wind speed and turbulence standard deviation is observed.

It is well known that the average wind speed is predictable. The obvious relationship implies that forecasting the turbulence standard deviation is possible. So the predictability of turbulence standard deviation is analyzed in detail in section 3.

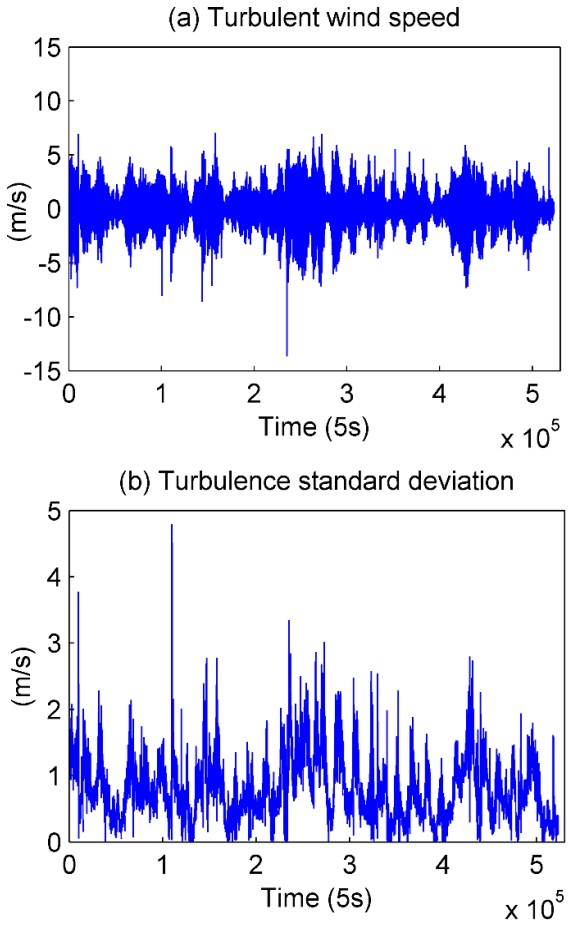

Figure 3. Comparison between turbulent wind speed and corresponding turbulence standard deviation for WF1 on August 2012.

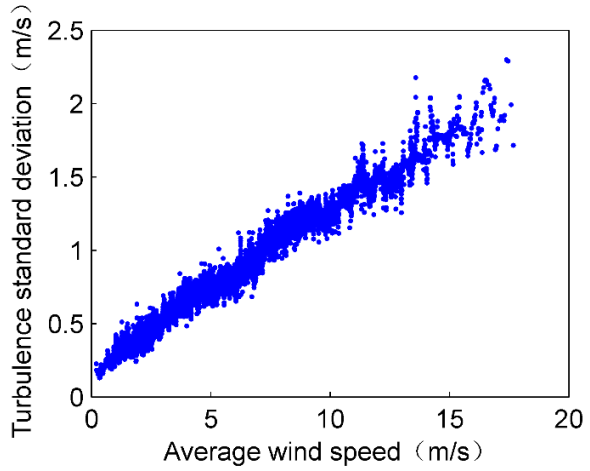

Figure 4. Relationship between turbulent standard deviation and hourly average wind speed.

3. Predictability Analysis of Turbulence Standard Deviation

In Section 2, the turbulence standard deviation is chosen to depict the wind speed uncertainty on small time scale. The preliminary analysis reveals that forecasting turbulence standard deviation is possible. Therefore, the predictability of turbulence standard deviation is further analyzed.

Data-driven forecasting model estimates future values based on historical observations. According to the input variable, the data-driven model can be classified into univariate model and multivariate model (Zheng and Kusiak, 2009). The univariate model is expressed as follows (Box et al., 2015):

$$y'(t+n) = f(y(t), y(t-1), \ldots, y(t-m+1)) \tag{4}$$

where n is the forecasting horizon (e.g., for n = 1 h, the forecasting horizon is 1 hour), $y'(t+n)$ is the forecasting value, $y(t)$, $y(t-1)$, …, $y(t-m+1)$ are the current and historical observations. The total number of inputs is m.

The multivariate forecasting model is established as follows (Box et al., 2015):

$$\begin{aligned} y'(t+n) = f(&y(t), y(t-1), \cdots, y(t-m+1) \\ &x_1(t), x_1(t-1), \cdots, x_1(t-k+1) \\ &x_2(t), x_2(t-1), \cdots, x_2(t-k+1) \\ &\cdots \\ &x_l(t), x_l(t-1), \cdots, x_l(t-k+1)) \end{aligned} \tag{5}$$

where y, x_1, x_2, …, x_l are input variables, the total number of inputs is $m + k \times l$.

The premise of forecasting based on data-driven model is that $y(t+n)$ highly depends on $y(t)$, $y(t-1)$, …, $y(t-m+1)$, $x_i(t)$, $x_i(t-1)$, …, $x_i(t-k+1)$, i = 1, 2, …, l. Autocorrelation and cross-correlation functions are useful tools to analyze the correlation between different random variables. Therefore, correlation functions are introduced to analyze the predictability of turbulence standard deviation.

3.1. Autocorrelation and Cross-correlation Functions

If the future turbulence standard deviation highly depends on its historical observations, then the turbulence standard deviation can be predicted based on the univariate forecasting model. The dependency between future values and historical observations can be measured by autocorrelation function. Suppose $x(t+\tau)$ is the time shifted signal of $x(t)$, then the autocorrelation function is defined as:

$$R_{xx}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_0^T x(t)x(t+\tau)dt \tag{6}$$

where τ is time lag.

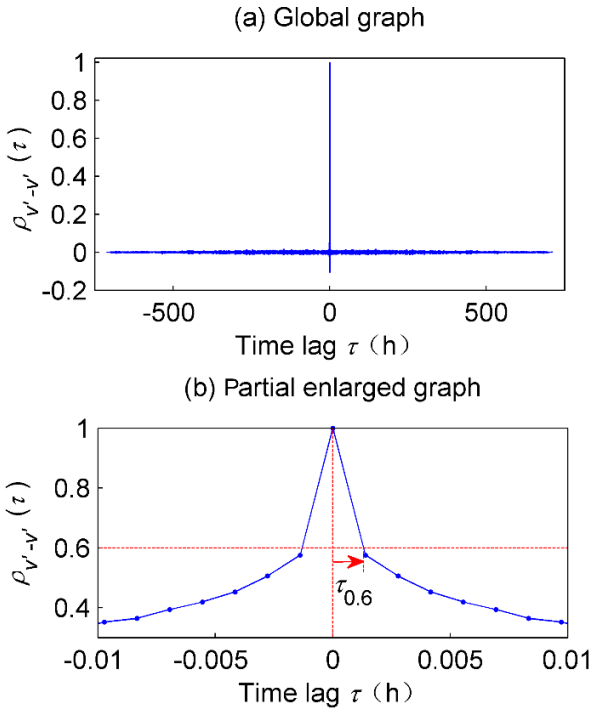

Figure 5. Predictability analysis of turbulent wind speed for WF1 on August 2012.

In practical engineering applications, the Pearson auto-correlation coefficient $\rho_{xx}(\tau)$ is usually used to measure the dependencies of signals between different times:

$$\rho_{xx}(\tau)=\frac{\sum_{t=1}^{n-\tau}(x(t)-\bar{x})(x(t+\tau)-\bar{x})}{\sum_{t=1}^{n}(x(t)-\bar{x})^2} \tag{7}$$

where $\bar{x}$ is the mean value of $\{x(t), t = 1, 2, \ldots, n\}$. $\rho_{xx}(\tau) = 1$ represents an intense correlation between $x(t)$ and $x(t + \tau)$. While $\rho_{xx}(\tau) = 0$ indicates the absence of a relation between $x(t)$ and $x(t + \tau)$. In general, $|\rho_{xx}(\tau)| \le 1$. The larger the $|\rho_{xx}(\tau)|$, the higher the correlation.

For the multivariate forecasting model, the dependencies between turbulence standard deviation and average wind speed should be measured firstly. So the cross-correlation function is introduced into the study. Suppose $\{x(t)\}$ and $\{y(t)\}$ are two given time series, then the cross-correlation function between $x(t)$ and $y(t + \tau)$ is shown as follows:

$$R_{xy}(\tau)=\lim_{T\to\infty}\frac{1}{T}\int_0^T x(t)y(t+\tau)dt \tag{8}$$

Similarly, the cross-correlation coefficient is defined as (Kisi et al., 2013; Nourani, 2015):

Table 1. The Statistical Results of Predictability Analysis

Month	*WF1*			*WF2*		
	$\tau_{v'}(0.6)$	$\tau_{\sigma}(0.6)$	$\rho_{\sigma\text{-}v'}(4)$	$\tau_{v'}(0.6)$	$\tau_{\sigma}(0.6)$	$\rho_{\sigma\text{-}v'}(4)$
1	3.0 s	5.8 h	0.951	5.8 s	1.0 h	0.909
2	3.1 s	3.2 h	0.899	7.2 s	3.2 h	0.880
3	3.5 s	2.5 h	0.861	8.0 s	1.9 h	0.892
4	4.5 s	2.6 h	0.868	8.0 s	4.1 h	0.882
5	6.0 s	0.9 h	0.767	12.5 s	2.8 h	0.839
6	4.5 s	1.4 h	0.848	12.5 s	1.3 h	0.848
7	8.5 s	0.8 h	0.784	9.0 s	1.2 h	0.874
8	4.5 s	1.2 h	0.832	14.0 s	0.6 h	0.825
9	7.5 s	1.5 h	0.832	8.0 s	1.2 h	0.863
10	4.0 s	2.2 h	0.874	5.2 s	1.4 h	0.897
11	3.5 s	2.3 h	0.840	4.1 s	3.2 h	0.943
12	3.5 s	5.8 h	0.910	6.5 s	4.1 h	0.917

* where $\tau_{v'}(0.6)$ represents the time delay when the autocorrelation coefficient of turbulent wind speed is 0.6. $\tau_{\sigma}(0.6)$ represents the time delay when the autocorrelation coefficient of turbulence standard deviation is 0.6. $\rho_{\sigma\text{-}v'}(4)$ represents the autocorrelation coefficient between turbulence standard deviation and average wind speed when the time delay is 4 h.

$$\rho_{xy}(\tau)=\frac{\sum_{t=1}^{n-\tau}(x(t)-\bar{x})(y(t+\tau)-\bar{y})}{[\sum_{t=1}^{n}(x(t)-\bar{x})^2\sum_{t=1}^{n}(y(t)-\bar{y})^2]^{1/2}} \tag{9}$$

where $\bar{x}$ is the mean value of $\{x(t), t = 1, 2, \ldots, n\}$, $\bar{y}$ is the mean value of $\{y(t), t = 1, 2, \ldots, n\}$.

The meaning of $\rho_{xy}(\tau)$ is the same as $\rho_{xx}(\tau)$. The larger the $|\rho_{xy}(\tau)|$, the higher the correlation.

3.2. Predictability Analysis of Turbulence Standard Deviation

The autocorrelation of turbulent wind speed corresponding to Figure 3 a is analyzed in the first, which is shown in Figure 5. The bottom sub-figure is the magnification of the top sub-figure of Figure 5. The autocorrelation coefficient $\rho_{v'\text{-}v'}(\tau)$ sharply decreases with the increase of the time lag τ. In general, $\rho_{xx}(\tau) \ge 0.5$ represents a significant correlation. In the present study, $\rho_{xx}(\tau) = 0.6$ is selected as threshold in the following analysis. According to the threshold, the corresponding time lag $\tau_{v'}(0.6) \approx 4.5$ s. In other words, $v'(t)$ is weakly correlated with $v'(t + \tau)$ when $\tau > 4.5$ s. The statistical results of $\tau_{v'}(0.6)$ are shown in the second and fifth columns of Table 1. $\tau_{v'}(0.6)$ ranges from 3.1 to 14.0 s. The correlation is significant only when $\tau \le 14.0$ s. Although in the case of $\tau_{v'}(0.6) \approx 14.0$ s, the time horizon is still too short for forecasting. Therefore, forecasting $v'(t)$ based on its historical data is impossible. In other words, turbulent wind speed is unpredictable.

Then the same analysis is conducted for the turbulence standard deviation corresponding to Figure 3(b). The result is shown in Figure 6. According to Figure 6, the time lag $\tau_{\sigma}(0.6) \approx 1.2$ h when $\rho_{\sigma-\sigma}(\tau) = 0.6$. The statistical results of $\tau_{\sigma}(0.6)$ are

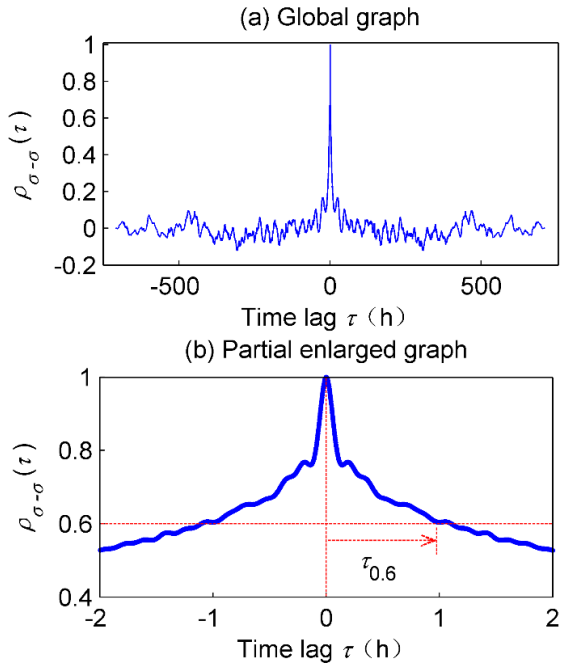

Figure 6. Predictability analysis of turbulence standard deviation for WF1 on August 2012.

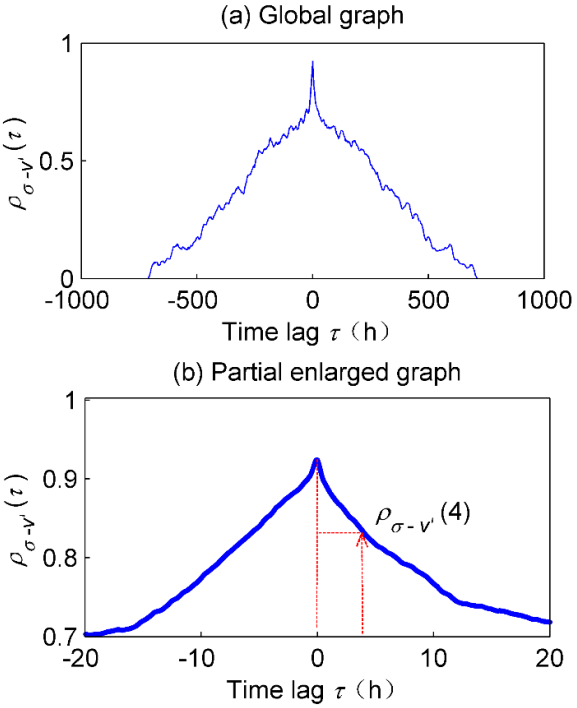

Figure 7. Cross-correlation analysis between turbulence standard deviation and average wind speed for WF1 on August 2012.

shown in the third and sixth columns of Table 1. $\tau_\sigma(0.6)$ ranges from 0.6 to 5.8 h. The correlation is significant within a long-time horizon. The time horizon is long enough for forecasting. Therefore, the turbulence standard deviation σ is predictable.

Figure 4 reveals that turbulence standard deviation is highly correlated with average wind speed. It means that average wind speed would contribute to the forecasting of turbulence standard deviation. So the cross-correlation between average wind speed and turbulence standard deviation is analyzed. Figure 7 shows the analysis results. $\rho_{\sigma-v}(\tau)$ decays slowly with the increase of time lag τ. The time lag τ is more than 24 h when $\rho_{\sigma-v}(\tau) = 0.6$. In general, the forecasting horizon of data-driven model is less than 4 h in most wind power forecasting studies. So the cross-correlation coefficient $\rho_{\sigma-v}(4)$ corresponding to $\tau = 4$ h is analyzed here, which is shown in the fourth and seventh columns of Table 1. $\rho_{\sigma-v}(4)$ ranges from 0.767 to 0.951 when the time lag τ is 4 h. So the relationship between turbulence standard deviation and average wind speed is highly significant. Average wind speed can provide valuable information for turbulence standard deviation prediction.

In conclusion, the turbulent wind speed itself is unpredictable. However, the turbulence standard deviation can be forecasted within a specific time horizon.

3.3. Physical Mechanism of the Predictability of Turbulence Standard Deviation

Wind is a phenomenon of atmospheric turbulence motion. The predictability of turbulence standard deviation can be explained by the energy spectrum of atmospheric turbulence, which is shown in Figure 8. The energy spectrum is composed of three parts: energy region, inertial sub-region and dissipation region (Panofsky, 1984; Frisch, 1996). The atmospheric turbulence field is full of a series of vortexes with different time scales. For the atmospheric turbulence motion, the energy contained in the energy region mainly originates from the larger-scale vortex corresponding to the weather process. In the inertial sub-region, the energy cannot increase or decrease. The energy is just transmitted from large-scale vortex to small-scale vortex. The transmission process abides by the law of energy conservation. Whereas in the dissipation region, energy is dissipated gradually due to viscosity. Turbulent energies at different times are interrelated with each other within a specific time scale. For wind speed, the turbulence standard deviation represents the turbulent energy. As a result, the turbulence standard deviations at different times are highly correlated with each other within a specific time horizon. Meanwhile, the energy contained in small-scale vortex (the energy represented by turbulence standard deviation) comes from that contained in large-scale vortex (the energy represented by average wind speed). So the relationship between turbulence standard deviation and average wind speed is highly significant.

4. Prediction Experiments of Turbulence Standard Deviation

Aforementioned analysis has proved that the turbulence standard deviation is predictable. So the detailed forecasting experiments are conducted based on Support Vector Regression (SVR) and the Kernel Ridge Regression (KRR) in the section.

4.1. Forecasting Model

4.1.1. Support Vector Regression (SVR)

Suppose $\{x_i, y_i\}_i = 1: N$ is a set of training samples, where x_i is the input and y_i is the associated output. The SVR function (Vapnik, 1998) can be expressed as follows:

$$f(x) = \sum_{i=1}^{n} \omega_i K(x_i, x) + b \tag{10}$$

where $K(\cdot, \cdot)$ is the kernel function.

The training process of SVR equates to the following optimization problem:

$$\min_{\omega,b,\xi,\xi^*} \frac{1}{2}\omega^T\omega + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \tag{11}$$

subject to:

$$\begin{aligned} &y_i - \left(w^T\varphi(x_i) + b\right) \le \varepsilon + \xi_i^* \\ &\left(w^T\varphi(x_i) + b\right) - y_i \le \varepsilon + \xi_i \\ &\xi_i, \xi_i^* \ge 0, i = 1, ..., n \end{aligned} \tag{12}$$

where φ is the mapping function. ξ_i and ξ_i^* are slack variables subject to the ε-insensitive loss function. C is the cost of error.

The aforementioned optimization problem can be solved by using Lagrange multiplier:

$$\begin{aligned} \max_{\alpha_i,\alpha_i^*}(&-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(x_i, x_j) \\ &-\varepsilon\sum_{i=1}^{N}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{N} y_i(\alpha_i - \alpha_i^*)) \end{aligned} \tag{13}$$

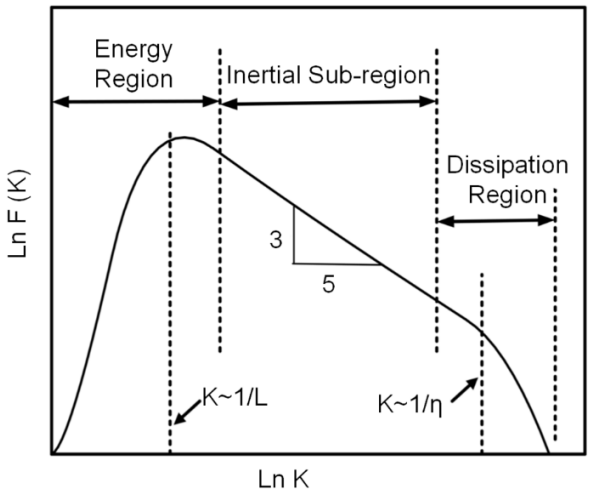

Figure 8. Energy spectrum of atmospheric turbulence.

subject to:

$$\begin{cases} \sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{cases} \tag{14}$$

where a_i and a_i^* are Lagrange multipliers.

4.1.2. Kernel Ridge Regression (KRR)

The function of kernel ridge regression is the same as SVR (Douak, 2013):

$$f(x) = \sum_{i=1}^{n} \omega_i K(x_i, x) + b \tag{15}$$

The optimization objective is:

$$\min_{\omega,b,\gamma} = \frac{1}{2}\omega^T\omega + \frac{1}{2}\gamma\sum_{i=1}^{N} e^2 \tag{16}$$

subject to:

$$\omega^T\varphi(x_i) + b + e_i = y_i \ , i = 1, 2, ..., N \tag{17}$$

where e_i is the error and γ is a regularization factor.

The optimization problem can also be transformed as:

$$\begin{aligned} &\min_{\omega,b,\gamma} \frac{1}{2}\omega^T\omega + \frac{1}{2}\gamma\sum_{i=1}^{N} e^2 \\ &-\sum_{i=1}^{N}\lambda_i\left(\omega^T\varphi(x_i) + b + e_i - y_i\right) \end{aligned} \tag{18}$$

4.2. Prediction Experiments

To evaluate the prediction performance, Mean Absolute Error (*MAE*) and Mean Square Error (*MSE*) are selected as the evaluation indexes (Purkait et al., 2008):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_{ri} - y_{fi}\right| \tag{19}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_{ri} - y_{fi}\right)^2 \tag{20}$$

where y_{ri} is the observation value, y_{fi} is the predicted value, and n is the number of test samples.

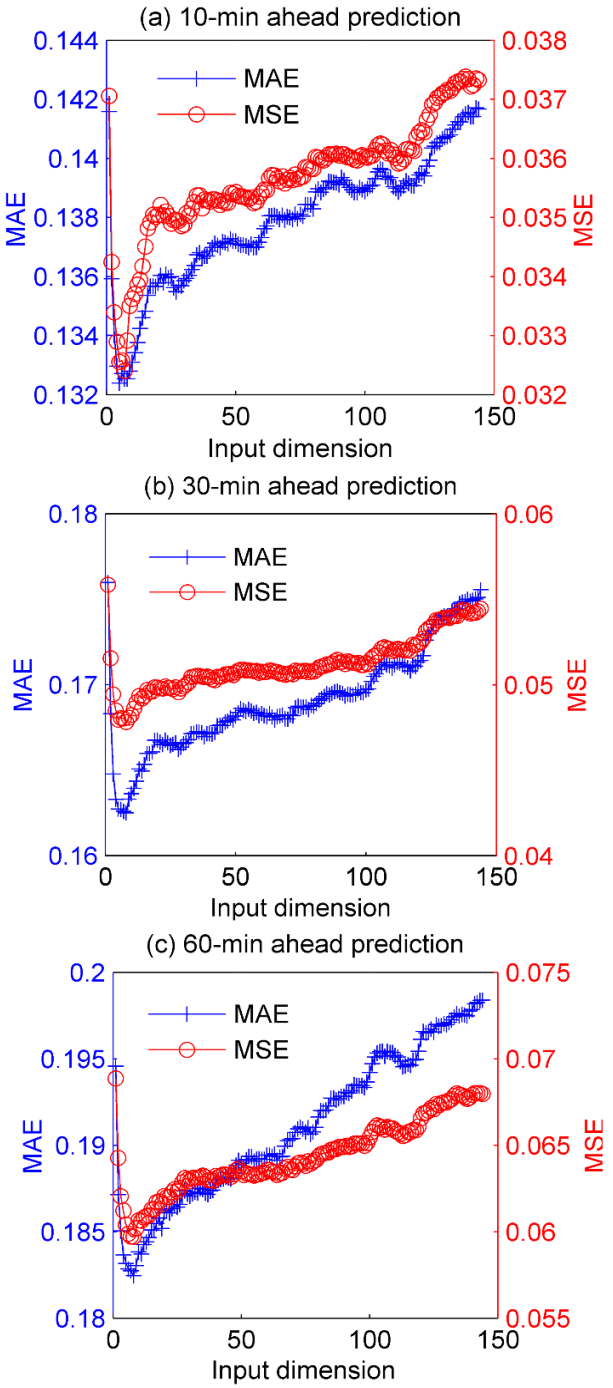

Figure 9. Selecting the input dimension of turbulence standard deviation: (a) 10-min ahead prediction; (b) 30-min ahead prediction; (c) 60-min ahead prediction.

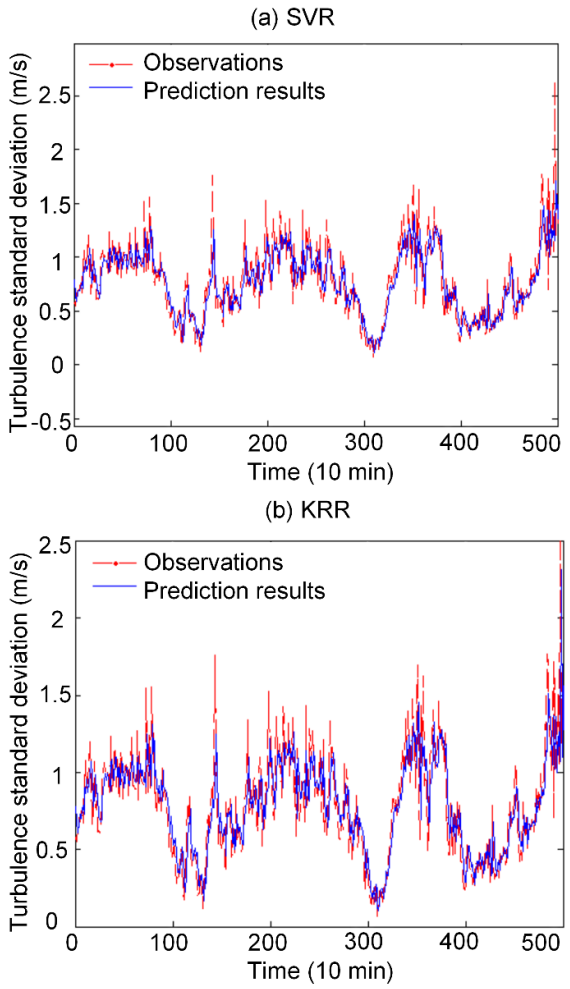

Figure 10. Actual wind speed turbulence standard deviation observations versus forecasting results (10-min ahead prediction) with different methods.

In the first, the univariate forecasting model is considered. Namely, turbulence standard deviation is predicted only based on its historical data. Each wind farm data is divided into 12 sets according to the month. In each data set, the data from the first 20 days is used for training and the rest data is used for testing. According to the results in Table 1, $\tau_\sigma(0.6)$ ranges from 0.6 to 5.8 h. So the forecasting horizon n in the present study is selected as 10-min, 30-min and 60-min, respectively. Before the experiments, the input dimension m must be determined in the first. In this study, m is selected by experiments. During the selection process, the prediction horizon n remains constant, while the input dimension m varies from 1 to 144 (the time interval is 10-min). According to the prediction results, the dimension corresponding to the minimum prediction error is selected as the input dimension m. As shown in Figure 9, the error curves firstly decrease and then increase with the increase of input dimension. According to the error curves, the optimum input dimensions for 10-min ahead prediction, 30-min ahead prediction and 60-min prediction are 5, 8 and 8, respectively.

After determining the optimum input dimension, 10-min ahead, 30-min ahead and 60-min prediction experiments are conducted based on SVR and KRR. Figures 10 ~ 12 show the prediction results. The prediction errors are presented in Tables 2 ~ 4. The prediction results demonstrated that the turbulence standard deviation indeed can be forecasted. The prediction

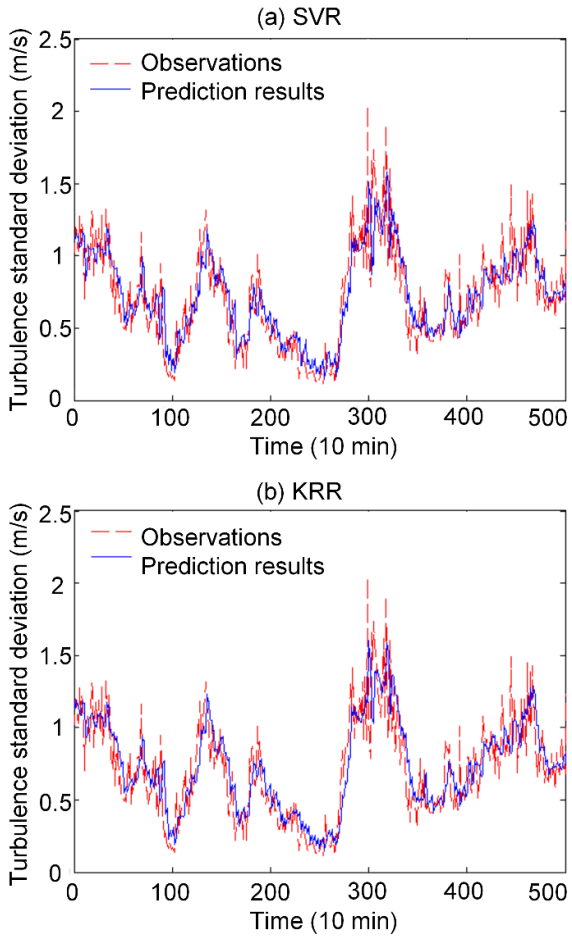

Figure 11. Actual wind speed turbulence standard deviation observations versus forecasting results (30-min ahead prediction) with different methods.

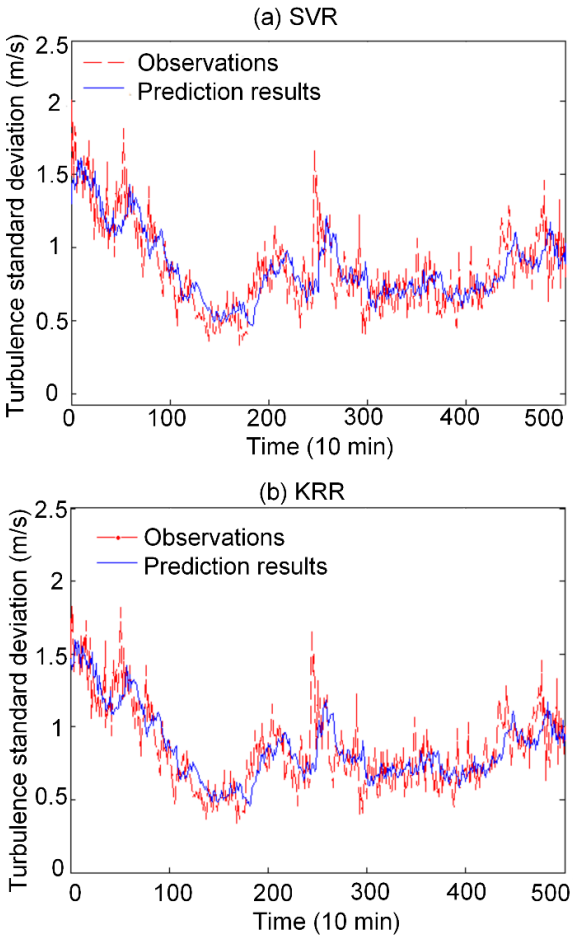

Figure 12. Actual wind speed turbulence standard deviation observations versus forecasting results (60-min ahead prediction) with different methods.

errors of SVR approximate closely to that of KRR. Furthermore, the prediction errors increase with the increase of prediction horizon.

According to section 3, the turbulence standard deviation is highly correlated with average wind speed. Therefore, the multivariate forecasting model is considered. The average wind speed is fused into the forecasting model to predict turbulence standard deviation. Aforementioned results have proved that the prediction performance of SVR is similar to KRR. So only the SVR model is considered here. Compared with the original SVR model (the univariate forecasting model), the multivariate forecasting model is denoted as the modified SVR model. The 60-min ahead prediction experiment is conducted to compare the forecasting performance of modified SVR model and original SVR model.

The input dimension of average wind speed is determined by experiment as well. In the selection process, the input dimension of turbulence standard deviation m is constant. While the input dimension of average wind speed k varies from 1 to 144 (the time interval is 10-min). Figures 13 ~ 14 show the variation curves of *MAE* and *MSE*. The red-solid line represents the forecasting error of original SVR model. The blue-dot-solid line represents the forecasting error of modified SVR model. The error curves firstly decrease and then increase with the increase of input dimension. So the optimum input dimension of average wind speed is determined. Furthermore, the forecasting errors decrease when the average wind speed is fused into the forecasting model. After determining the optimum input dimension k, the turbulence standard deviation is predicted based on modified SVR model.

Figure 15 shows the forecasting results. Obviously, the prediction results of modified SVR model are closer to the actual observations than original SVR model.

To quantify the forecasting performance, the following two indexes are defined:

$$\Delta MAE = \frac{MAE_O - MAE_M}{MAE_O} \times 100\% \tag{21}$$

Table 2. Prediction Errors for Wind Farm 1 and Wind Farm 2 (10-min ahead Prediction)

	Month		1	2	3	4	5	6	7	8	9	10	11	12
1	SVR	*MAE*	0.098	0.104	0.116	0.15	0.135	0.117	0.126	0.098	0.142	0.148	0.119	0.108
		MSE	0.022	0.023	0.029	0.051	0.045	0.034	0.044	0.021	0.041	0.047	0.029	0.024
	KRR	*MAE*	0.099	0.108	0.12	0.153	0.135	0.120	0.130	0.106	0.147	0.151	0.122	0.0115
		MSE	0.023	0.028	0.033	0.051	0.043	0.035	0.047	0.034	0.05	0.049	0.036	0.034
2	SVR	*MAE*	0.139	0.173	0.149	0.172	0.158	0.176	0.145	0.163	0.173	0.153	0.139	0.148
		MSE	0.042	0.064	0.048	0.069	0.056	0.071	0.044	0.057	0.060	0.046	0.036	0.046
	KRR	*MAE*	0.140	0.172	0.172	0.151	0.161	0.157	0.176	0.146	0.160	0.177	0.148	0.143
		MSE	0.042	0.062	0.062	0.050	0.059	0.057	0.068	0.047	0.056	0.063	0.040	0.041

Table 3. Prediction Errors for Wind Farm 1 and Wind Farm 2 (30-min ahead Prediction)

	Month		1	2	3	4	5	6	7	8	9	10	11	12
1	SVR	*MAE*	0.112	0.131	0.143	0.184	0.46	0.136	0.15	0.134	0.172	0.177	0.146	0.136
		MSE	0.028	0.043	0.047	0.07	0.06	0.043	0.059	0.05	0.069	0.063	0.046	0.044
	KRR	*MAE*	0.0118	0.147	0.151	0.189	0.162	0.142	0.164	0.154	0.183	0.178	0.145	0.144
		MSE	0.03	0.046	0.05	0.074	0.059	0.044	0.063	0.054	0.07	0.065	0.045	0.046
2	SVR	*MAE*	0.172	0.150	0.188	0.170	0.216	0.192	0.188	0.161	0.176	0.196	0.162	0.155
		MSE	0.060	0.044	0.072	0.058	0.097	0.076	0.079	0.050	0.063	0.073	0.048	0.047
	KRR	*MAE*	0.178	0.182	0.191	0.169	0.223	0.195	0.192	0.163	0.178	0.197	0.162	0.164
		MSE	0.066	0.068	0.073	0.057	0.102	0.076	0.077	0.052	0.064	0.075	0.047	0.050

Table 4. Prediction Errors for Wind Farm 1 and Wind Farm 2 (60-min ahead Prediction)

	Month		1	2	3	4	5	6	7	8	9	10	11	12
1	SVR	*MAE*	0.132	0.153	0.158	0.194	0.181	0.161	0.179	0.15	0.201	0.208	0.167	0.158
		MSE	0.036	0.051	0.054	0.082	0.076	0.06	0.075	0.052	0.087	0.089	0.059	0.058
	KRR	*MAE*	0.134	0.166	0.164	0.209	0.185	0.168	0.191	0.171	0.215	0.205	0.175	0.164
		MSE	0.036	0.06	0.059	0.087	0.082	0.068	0.086	0.094	0.125	0.091	0.078	0.069
2	SVR	*MAE*	0.204	0.182	0.196	0.192	0.245	0.213	0.208	0.184	0.207	0.231	0.189	0.176
		MSE	0.092	0.068	0.075	0.076	0.125	0.091	0.095	0.067	0.086	0.103	0.072	0.058
	KRR	*MAE*	0.216	0.193	0.208	0.207	0.251	0.227	0.221	0.189	0.214	0.221	0.183	0.182
		MSE	0.108	0.078	0.096	0.092	0.158	0.109	0.127	0.073	0.093	0.092	0.064	0.061

Table 5. Statistical Results of Δ*MAE* and Δ*MSE*

Month		1	2	3	4	5	6	7	8	9	10	11	12
1	Δ*MAE* (%)	10.40	15.56	11.90	12.50	5.87	19.07	18.37	14.08	15.93	18.66	24.52	21.94
	Δ*MSE* (%)	22.17	26.97	30.11	17.03	17.34	42.29	38.17	29.63	31.78	32.79	41.68	37.16
2	Δ*MAE* (%)	2.91	12.69	5.45	21.39	17.21	4.96	1.49	4.48	6.93	12.28	16.48	5.59
	Δ*MSE* (%)	3.88	25.35	12.43	40.08	33.41	6.88	1.90	10.04	16.64	23.52	28.02	11.45

where MAE_O is the *MAE* of original SVR model. MAE_M is the *MAE* of modified SVR model. Also:

$$\Delta MSE = \frac{MSE_O - MSE_M}{MSE_O} \times 100\% \quad (22)$$

where MSE_O is the *MSE* of original SVR model. MSE_M is the *MSE* of modified SVR model.

The positive indexes imply that the forecasting performance of modified SVR model is better than original SVR model. The statistical results of Δ*MAE* and Δ*MSE* are shown in Table 5. All the values in Table 5 are greater than 0. Therefore, the forecasting accuracy is significantly improved after fusing the average wind speed into the model.

5. Conclusions

Actual wind speed consists of average wind speed and turbulent wind speed. Both of them are characterized by significant uncertainty. The wind speed uncertainty on large time scale has been widely studied by forecasting average wind speed. However, the forecasting study of wind speed uncer-

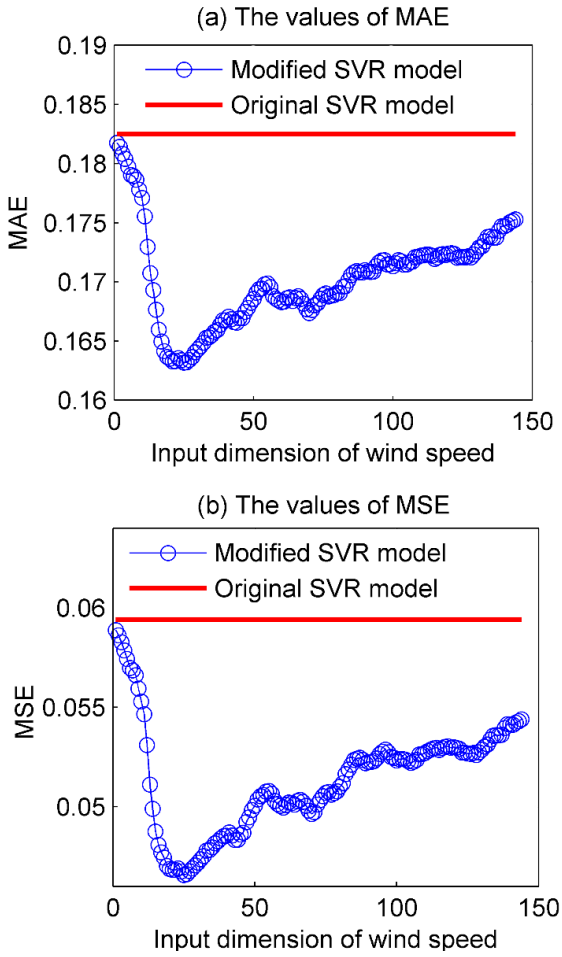

Figure 13. Selecting the input dimension of average wind speed based on WF1.

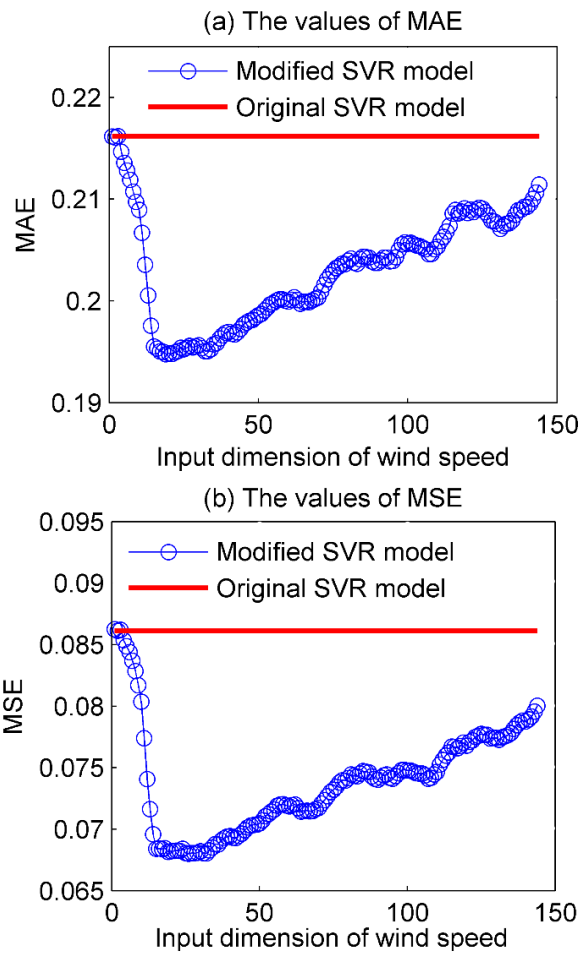

Figure 14. Selecting the input dimension of average wind speed based on WF2.

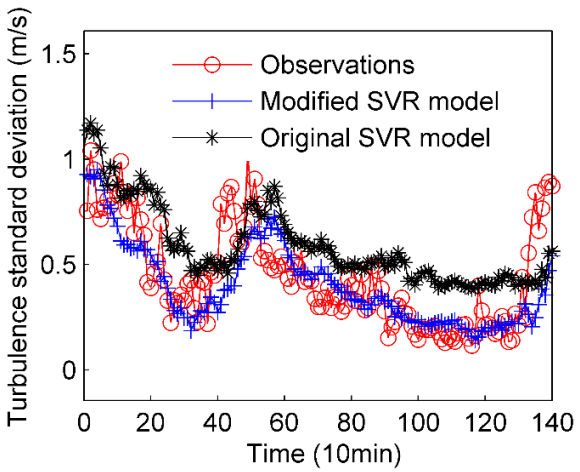

Figure 15. Prediction performance of modified SVR model (60-min ahead prediction).

tainty on small time scale (namely turbulent wind speed) is fairly rare. The turbulent wind speed is depicted by turbulence standard deviation in the present study. Then autocorrelation and cross-correlation functions are used to analysis the predictability of turbulent wind speed and turbulence standard deviation. The analysis results prove that the turbulent wind speed is unpredictable. Nevertheless, the turbulence standard deviation is predictable within a proper time horizon. After verifying the predictability of turbulence standard deviation, the 10-min ahead, 30-min ahead and 60-min ahead prediction experiments are conducted by using SVR and KRR. The forecasting results agree well with observations. The forecasting accuracies of SVR and KRR are similar. The forecasting errors increase with the increase of forecasting horizon. In the final, the average wind speed is fused into the forecasting model. The forecasting accuracy is significantly improved compared with the original model. All of these results prove that forecasting turbulence standard deviation based on data-driven model is feasible.

The applications of turbulence standard deviation prediction are not addressed in the present study. However, two possible application scenarios may be envisaged. Firstly, wind power curve is usually used to transform wind speed forecasting into wind power forecasting. Considering the impact of turbulent wind speed on wind power curve, its forecasting results

can be used to correct the wind power curve, which contributes to improve the modelling accuracy of wind power output. Secondly, turbulent wind speed is a major cause of high aerodynamic loads on wind turbine blades. Effective control strategies can be taken to reduce the aerodynamic loads according the forecasting results, which is significant for wind farm operation and management.

Acknowledgements. This work was supported by National Program on Key Basic Research Project (973 Program) of China under Grant 2012CB215201 and the China Scholarship Council.

References

Anahua, E., Barth, S., and Peinke, J. (2008). Markovian power curves for wind turbines. *Wind Energy*, 11(3), 219-232. https://doi.org/10.1002/we.243

Abdel-Aal R. E., Elhadidy M. A., and Shaahid S. M. (2009). Modeling and forecasting the mean hourly wind speed time series using GMDH-based abductive networks. *Renewable Energy*, 34(7): 1686e99. https://doi.org/10.1016/j.renene.2009.01.001

Burton, T., Jenkins, N., Sharpe, D., and Bossanyi, E. (2011). *Wind energy handbook*. John Wiley & Sons, https://doi.org/10.1002/9781119992714

Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time series analysis: forecasting and control.* John Wiley & Sons.

Chamorro, L. P. and Porté-Agel, F. (2009). A wind-tunnel investigation of wind-turbine wakes: boundary-layer turbulence effects. *Boundary-layer meteorol.*, 132(1), 129-149. https://doi.org/10.10 07/s10546-009-9380-8

Carpman, N. (2011). *Turbulence intensity in complex environments and its influence on small wind turbines.* M.Sc. Dissertation. Department of Earth Sciences, Uppsala University.

Cassola, F. and Burlando, M. (2012). Wind speed and wind energy forecast through Kalman filtering of Numerical Weather Prediction model output. *Appl. Energy*, 99, 154-166. https://doi.org/10.1016/j.apenergy.2012.03.054.

Douak, F., Melgani, F., and Benoudjit, N. (2013). Kernel ridge regression with active learning for wind speed prediction. *Appl. Energy*, 103, 328-340. https://doi.org/10.1016/j.apenergy.2012.09.055

Frisch U. (1995). *Turbulence, the legacy of A.N.* Kolmogorov. Cambridge: Cambridge University Press.

Frandsen, S. T. and Thøgersen, M. L. (1999). Integrated fatigue loading for wind turbines in wind farms by combining ambient turbulence and wakes. *Wind Eng.*, 23, 327-339.

Freudenreich K. and Argyriadis K. (2006). The new standard IEC 61400-1: 2005 and its effect on the load level of wind turbines. *Proceedings of Deutsche Wind Energie Konferenz*.

Gustafsson, N. (1993). *HIRLAM 2 final report*. HIRLAM Consortium, Norrkoeping.

Gottschall, J. and Peinke, J. (2008). How to improve the estimation of power curves for wind turbines. *Environ. Res. Lett.*, 3(1), 015005. https://doi.org/10.1088/1748-9326/3/1/015005.

Gomes, P. and Castro, R. (2012). Wind speed and wind power forecasting using statistical models: Autoregressive moving average (ARMA) and artificial neural networks (ANN). *Int. J. Sustainable Energy Dev.*, 1(1/2), https://doi.org/10.20533/ijsed.2046.3707.2012.0007.

Hand, M. M., Kelley, N. D., and Balas, M. J. (2003). Identification of Wind Turbine Response to Turbulent Inflow Structures: Preprint. *ASME/JSME 2003 4th Joint Fluids Summer Engineering Conference*, American Society of Mechanical Engineers, pp:2557-2566. https://doi.org/10.1115/FEDSM2003-45360

Hansen, K. S. and Larsen, G. C. (2005). Characterising turbulence intensity for fatigue load analysis of wind turbines. *Wind Eng.*, 29(4), 319-329. https://doi.org/10.1260/030952405774857897

International Electrotechnical Commission (IEC). (1998). *Wind turbine generator systems – Part 1: Safety requirements, Interna-tional Standard 61400–1 edition 2,* International Electrotechnical Commission.

International Electrotechnical Commission (IEC). (2005). *IEC 61400-1-Ed. 3: Wind turbines – Part 1: Design requirements*, IEC, Geneva, Switzerland.

Jimenez, A., Crespo, A., Migoya, E., and Garcia, J. (2007). Advances in large-eddy simulation of a wind turbine wake. *J. Phys.: Conf. Ser.*, 75, 012041. https://doi.org/10.1088/1742-6596/75/1/012041

Kaiser, K., Langreder, W., Hohlen, H., and Højstrup, J. (2007). Turbulence correction for power curves. *Wind Energy*, 159-162. https://doi.org/10.1007/978-3-540-33866-6_28.

Kelley, N. D., Osgood, R. M., Bialasiewicz, J. T., and Jakubowski, A. (2000). Using wavelet analysis to assess turbulence/rotor interactions. *Wind Energy*, 3(3), 121-134, https://doi.org/10.1002/we.33

Kuik, G. V., Ummels, B., and Hendriks, R. (2007). *Sustainable energy technologies*. The Netherlands: Springer, Amsterdam, 24(6), 1431-1439.

Kusiak, A., Zheng, H., and Song, Z. (2009). Wind farm power prediction: a data-mining approach. *Wind Energy*, 12(3), 275-293. https://doi.org/10.1002/we.295

Kisi, O., Shiri, J., Akbari, N., Pour, M. S., Hashemi, A., and Teimourzadeh, K. (2013). Modeling of dissolved oxygen in river water using artificial intelligence techniques. *J. Environ. Inf.*, 22(2), 92-101. https://doi.org/10.3808/jei. 201300248

Kim, S. H., Shin, H. K., Joo, Y. C., and Kim, K. H. (2015). A study of the wake effects on the wind characteristics and fatigue loads for the turbines in a wind farm. *Renewable Energy*, 74, 536-543. https://doi.org/10.1016/j.renene.2014.08.054

Li, G. and Shi, J. (2010). On comparing three artificial neural networks for wind speed forecasting. *Appl. Energy*, 87(7), 2313-2320, https://doi.org/10.1016/j.ape nergy.2009.12.013

Lavely, A. W., Vijayakumar, G., Kinzel, M. P., Brasseur, J. G., and Paterson, E. G. (2011). Space-time loadings on wind turbine blades driven by atmospheric boundary layer turbulence. *AIAA*, (2011-635). https://doi.org/10.2514/6.2011-635

Liu, H., Chen, C., Tian, H. Q., and Li, Y. F. (2012). A hybrid model for wind speed prediction using empirical mode decomposition and artificial neural networks. *Renewable Energy*, 48, 545-556. https://doi.org/10.1016/j.renene.2012.06.012

Nourani, V. (2015). Application of entropy concept for input selection of wavelet-ann based rainfall-runoff modeling. *J. of Enviro. Inf.*, 26(1), 52-70. https://doi.org/10.3808/jei.20150030 9

Panovsky, H. A. (1984). *Atmospheric Turbulence: Models and Methods for Engineering Applications*. John Wiley and Sons, New York.

Purkait, B., Kadam, S. S., and Das, S. K. (2008). Application of artificial neural network model to study arsenic contamination in groundwater of malda district, eastern India. *J. Environ. Inf.*, 12(2), 140-149. https://doi.org/10.3808/jei.200800132

Rosen, A. and Sheinman, Y. (1994). The average output power of a wind turbine in a turbulent wind. *J. Wind Eng. Ind. Aerodyn.*, 51(3), 287-302. https://doi.org/10.1016/0167-6105(94)90064-7

Rahmani, M. A. (2015). The use of statistical weather generator, hybrid data driven and system dynamics models for water resources management under climate change. *J. Environ. Inf.*, 25(1), 23-35. https://doi.org/10.3808/jei.201400285

Sheinman, Y. and Rosen, A. (1992). A dynamic model of the influence of turbulence on the power output of a wind turbine. *J. Wind Eng. Ind. Aerodyn.*, 39(1-3), 329-341. https://doi.org/10.1016/01676105(92)90557-Q

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Wang, W., and Powers, J. G. (2005). *A description of the advanced research WRF version 2 (No. NCAR/TN-468+ STR)*. National Center For Atmospheric Research Boulder Co Mesoscale and Microscale Meteorology Div.

Sagaut, P., Deck, S., and Terracol, M. (2006). Multiscale and multi-resolution approaches in turbulence. *Can. Med. Assoc. J.*, 126(12), 1420-1.

Sørensen, J. D., Frandsen, S., and Tarp-Johansen, N. J. (2008). Effective turbulence models and fatigue reliability in wind farms. *Probabilistic Eng. Mech.*, 23(4), 531-538. https://doi.org/10.1016/j.probengmech.2008.01.009

Stull, R. B. (2012). *An introduction to boundary layer meteorology (Vol. 13)*. Springer Science & Business Media.

Siddiqui, M. S., Rasheed, A., Kvamsdal, T., and Tabib, M. (2015). Effect of Turbulence Intensity on the Performance of an Offshore Vertical Axis Wind Turbine. *Energy Procedia*, 80, 312-320. https://doi.org/10.1016/j.egypro.2015.11.435

Santamaría-Bonfil, G., Reyes-Ballesteros, A., and Gershenson, C. (2016). Wind speed forecasting for wind farms: A method based on support vector regression. *Renewable Energy*, 85, 790-809. https://doi.org/10.1016/j.renene.2015.07.004

Von Karman, T. (1948). Progress in the statistical theory of turbulence. *Proceedings of the National Academy of Sciences of the United States of America*, 34(11), 530.

Van der Hoven, I. (1957). Power spectrum of horizontal wind speed in the frequency range from 0.0007 to 900 cycles per hour. *J. Meteorol.*, 14(2), 160-164. https://doi.org/10.1175/1520-0469(1957)014<0160:PSOHWS>2.0.CO;2

Vapnik, V. N. (1998). *Statistical Learning Theory*. Wiley Interscience.

Welfonder, E., Neifer, R. and Spanner, M. (1997). Development and experimental identification of dynamic models for wind turbines. *Control Eng. Pract.*, 5(1), 63-73. https://doi.org/10.1016/S0967-0661(96)00208-0

Wilson, D. R. and Ballard, S. P. (1999). A microphysically based precipitation scheme for the UK Meteorological Office Unified Model. *Q. J. R. Meteorol. Soc.*, 125(557), 1607-1636. https://doi.org/10.1002/qj.49712555707

Wu, Y. T. and Porté-Agel, F. (2012). Atmospheric turbulence effects on wind-turbine wakes: An LES study. *Energies*, 5(12), 5340-5362. https://doi.org/10.3390/en5125340

Zheng, H. and Kusiak, A. (2009). Prediction of wind farm power ramp rates: A data-mining approach. *J. Solar Energy Eng.*, 131(3), 031011. https://doi.org/10.1115/1.314