A Fusion Model for Day-Ahead Wind Speed Prediction based on the Validity of the Information

Jie Wan[1,a], Wenbo Hao[2,b], Guorui Ren[1,c],Leilei Zhao[2,d], Bingliang Xu[2,e], Chengzhi Sun[2,f],Zhigang Zhao[3,g], Chengrui Lei[4,h], Yufeng Guo[1,i],Jinfu Liu[1,J] and Daren Yu[1,k]

[1] *Department of Energy Science and Engineering*, Harbin Institute of Technology, Harbin, Heilongjiang, 150001, China
[2] Heilongjiang Electric Power Research Institute, Harbin, Heilongjiang, 150001, China
[3] Jilin Longyuan Wind Power Co.,Ltd, Changchun, Jilin, 130022, China
[4] Harbin Meteorological Bureau, Harbin, Heilongjiang, 150026, China

Abstract **— In recent years wind power is a hot topic in the field of new energy. However, due to the larger fluctuations of wind power, which is a threat to the safe operation of the power grid, large-scale wind power integration has been greatly restricted. Therefore, accurate wind speed forecast has great significance to wind power integration, especially the day-ahead wind speed forecast of 1-24 hours. For the sake of accurate wind speed forecast, the information efficiency of NWP and measured data were taken into account in this paper. Based on this, a fusion model was proposed. Experiments show that the fusion strategy can improve the accuracy of the day-ahead wind speed prediction results.**

Keywords - day-ahead wind speed prediction; information efficiency; fusion model;

I. INTRODUCTION

Wind power generation is a hot issue of new energy power industry in recent years [1-3]. However, the fluctuations of wind power make a severe impact on the safe and stable operation of the power grid. Therefore accurate prediction of wind speed is the foundation of wind power forecast. Particularly the day-ahead wind speed forecast of 1-24 hours can guarantee the grid reasonable dispatch, ensure power supply quality and help wind farms participating in bidding for wind power integration.

In recent years, many scholars began to focus on the day-ahead wind speed forecast by integrating wind numerical weather prediction and launched related research. The widely idea is using statistical models (such as neural networks, etc.) to correct the results of numerical weather prediction (NWP) in order to achieve the final forecast results. A.Vaccaroa used mesoscale numerical weather prediction, local high-precision numerical weather prediction, the measured wind speed and meteorological parameters to form a feature vector, designed Lazy learning algorithm based on k-Nearest Neighbour and taken the average of k most similar vector of history to forecast[4]. Federico combined Kalman filter and NWP for the dynamic correction of numerical weather prediction result and indicated that Kalman filter algorithm can reduce systematic errors in NWP[5]. Cai Qizhen used historical data of NWP and other related data as input, the measured data as the output to train the neural network to obtain correction model; then the date of NWP was inputted and the corrected NWP forecast results were achieved [6].

However, the current methods do not take into account the information efficiency of numerical weather prediction and measured data. The final prediction result was a simple synthesis of the prediction results of NWP and statistical method, which has the same prediction time. In fact, the prediction time of NWP and statistical method are significant different. So the forecast results is not accurate.

In order to solve the above mentioned problems, the information efficiency of NWP and measured data are taken into account in this paper. Firstly, the length of prediction step of NWP and measured data are analyzed respectively. Secondly, fusion method is selected based on the information efficiency of NWP and measured data. Finally, the finale prediction result is achieved by fusion method. Figure 1 is the sketch map of fusion model.

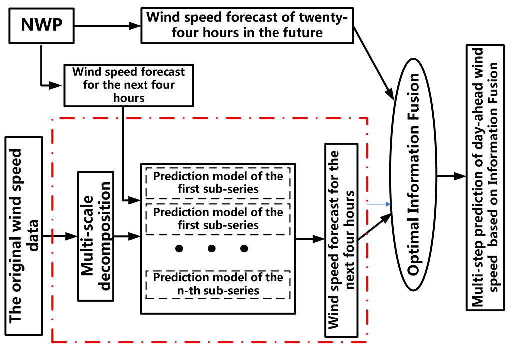

Fig.1The sketch map of fusion model

II. THE INFORMATION EFFICIENCY OF MEASURED DATA

The information efficiency of measured data need to be analyzed before setting up the statistical model. There are many factors (such as temperature, pressure, surface roughness, atmospheric circulation, etc.) have influence on wind speed and the mechanisms are complex. So the wind speed signal exhibits a strong multi-scale feature. Namely that the signal frequencies generated by different factor are different, so wind speed time series can be regarded as the superposition of different frequency signals. For the sake of analyzing the information efficiency of measured data, the following analysis steps are proposed in this paper:

Step-1: Based on the frequency domain characteristics of multi-scale wind speed signal, the original wind speed signal was decomposed into sub-sequences of different frequencies by using wavelet decomposition.

Due to decomposition level affecting the final scale signal frequency, signal is usually divided into three layers according to the previous experience. Figure 2 is original time series of the wind speed. Figure 3 is the results of three wavelet decomposition. To facilitate the latter modeling, the high-frequency section in the first sub-series is recorded as {} series, the high-frequency section in the second sub-series is recorded as {} series, the high-frequency section in the third sub-series is recorded as {}series, and the low-frequency section in the third sub-series is recorded as {} series.

Step-2: The information efficiency in each frequency-domain sequence was measured by using the autocorrelation and mutual information analysis. After the analysis of predictability, the length of prediction step in each frequency-domain sequence could be achieved according to the result of analysis. This procedure is useful to determine the number of input variables that can influence the forecast variable as all the inputs are selected from the wind speed points and the forecast horizon of each sub-series can also be analyzed through the analysis. Then the model architecture and forecast horizon of the sub-series in each frequency can be achieved.

The autocorrelation function is useful to achieve the dependencies among the data to identify the number of inputs and forecast horizons. Let {xt,t=1:n }is a time series, the autocorrelation function is:

$$\rho_k = \frac{\sum_{t=1}^{n-k}(x_t - \bar{x})(x_t + k - \bar{x})}{\sum_{t=1}^{n-k}(x_t - \bar{x})^2}$$

where $\bar{x}$ is the mean value of {xt,t=1:n }.

Figure 4 is autocorrelation function values with the change of delay number k in each frequency-domain sequence. Generally speaking, when the autocorrelation function value exceed 0.8, there is a strong correlation between the measured data. Namely that the higher the credibility of the forecast results. Studying on the low-frequency section in the third sub-series{}, we can find that the autocorrelation function decreased monotonically with the change of delay number. If we take the threshold value of 0.8, the autocorrelation length of this layer is 4 hour. Similarly, the predictable time in the sub-series{},{} and {} are 1 hour, 0.4 hour and 0.2 hour. Based on the analysis of information efficiency, it revealed that the predictability of each frequency sub-series are different. The predictability of low-frequency sub-series is stronger than that of high-frequency and secondary high-frequency sub-series.

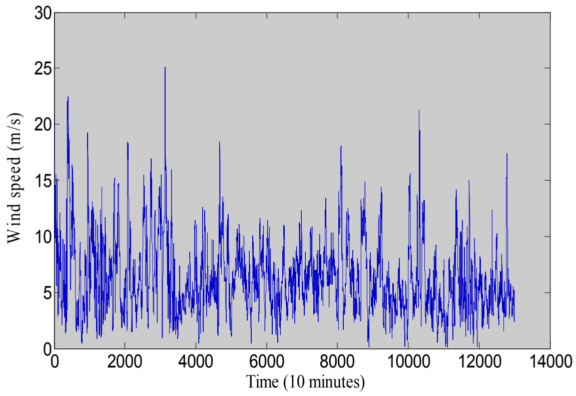

Fig.2 Original time series of the wind speed

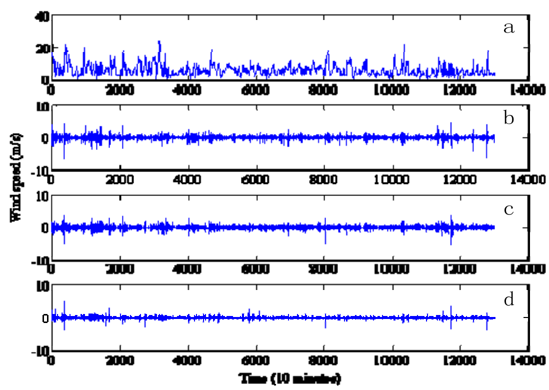

Fig.3 Decomposition results of the original series.

DOI 10.5013/IJSSST.a.17.28.13 ISSN: 1473-804x online, 1473-8031 print

Step-3: Statistical models are established in each frequency-domain sequence after the analysis of information efficiency. In this paper, support vector machine regression model (SVR) is selected to make prediction. The input of the statistical model for each layer is selected according to the correlation length of each layer. Assume it is time t, the current wind speed v(t) and the L-1 past states of wind speed are considered as the input vectors, while L is the autocorrelation length of each sub-series. Meanwhile, the predictable time of statistical model for each layer is also determined by the correlation length of each layer.

Step-4: After getting the prediction results of each sub-series, we can use wavelet reconstruction algorithm to obtain the final wind speed by multi-scale synthesis. The SVR model in each sub-series just forecast the wind speed which is in the range of predictability. For the high-frequency section, which is only responsible for the prediction result in next 0.2 hours, for other frequency sections, using the same method. When the prediction time exceeds one hour, the prediction result is only obtained from the low frequency section.

Through above analysis of information efficiency for measured data, it can be found that the prediction result of statistical methods is high credible within 4 hours. When the prediction time exceeds 4 hours, the credibility of prediction result begin to decrease.

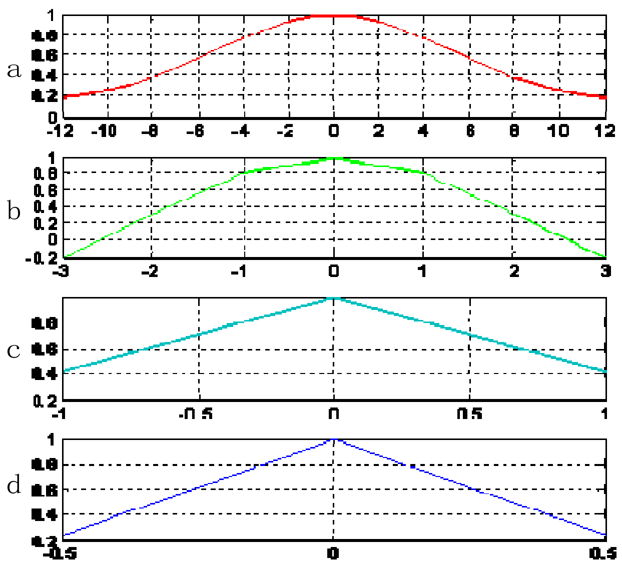

Fig.4 Autocorrelation function values with the change of delay number k in each frequency-domain sequence.
(a){} series.(b){} series.(c){} series.(d){}

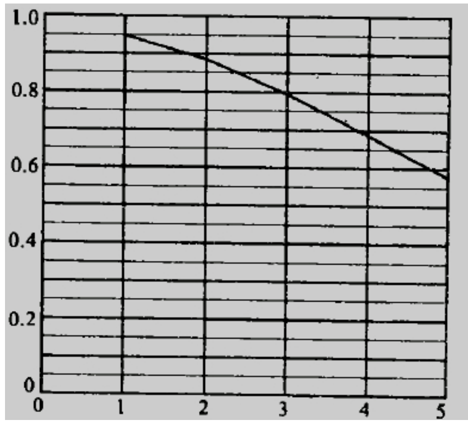

Fig.5 Anomaly correlation coefficient of Northern Hemisphere 500hPa

III. THE INFORMATION EFFICIENCY OF NWP

The basic principle of numerical weather prediction is: Using mass conservation, momentum conservation, energy conservation, and conservation of water vapor to establish the basic atmospheric equations, meanwhile turbulence models are used to close the equations. Then using the initial and boundary conditions to solve the atmospheric equations in order to forecast the future state of the atmosphere conditions. The typical forecast model is WRFa, using three nested grids, the calculated horizons are: 27km × 27km, 9km × 9km, 3km × 3km. The output of NWP system are average value of time and space for each computational grid, rather than instantaneous value. Therefore, the principle of calculation determines the prediction ability of NWP, which only contains the long-period component rather than short-cycle transient component, the length of prediction step is 1h. The effect of instantaneous forecasts is not ideal.

Many scholars studied the predictability problem of NWP. Smagorinsky's experiments showed that，with the increase of prediction time, the changes of forecast error is a nonlinear process. Pointed out that the predictability time of NWP is about 8 days [7]. Loren proved that NWP can predict atmospheric conditions within a week [8]. Many studies have fully confirmed that 2 weeks prediction results of NWP is high reliability [9]. At present, European Centre for Medium-Range Weather Forecasts (ECMWF) can provide 7-9 days prediction results of atmosphere conditions, meanwhile, National Centers for Environmental Prediction (NCEP) can provide 6—8 days prediction results [10]. Figure 5 (The vertical axis represents the correlation coefficient and the horizontal axis represents prediction time) is the T42L9 mode's anomaly correlation coefficient of Northern Hemisphere 500hPa height. From the graph we can see the anomaly correlation coefficient of the fourth day is 0.7. As a general rule, the prediction results is creditable when the anomaly correlation coefficient exceeds 0.6. Above study results show that the predictable ability of NWP is strong, the long-term prediction results of NWP is creditable [11].

When using SVR and numerical weather prediction (NWP) forecast wind speed, the average relative error and the Mean Square Error (MSE) of prediction results are shown in Figure 6 and Figure 7. It can be found that when the prediction time is in the range of four hours, the prediction error of SVR and NWP have little difference, but once the prediction time exceeds 4 hours, both the MSE and MAE of NWP are better than that of SVR. Through the analysis of information efficiency for NWP, we can find that NWP has higher predictable ability, namely that the prediction results of long-term are better than instantaneous forecast.

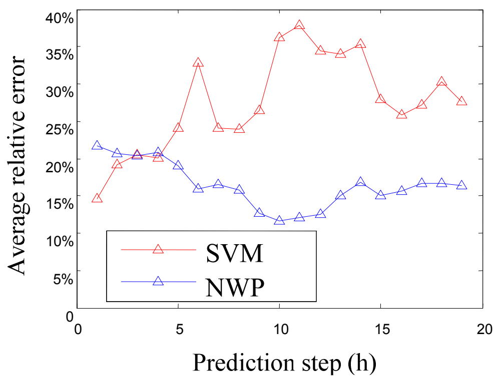

Fig.6 The average relative error curve

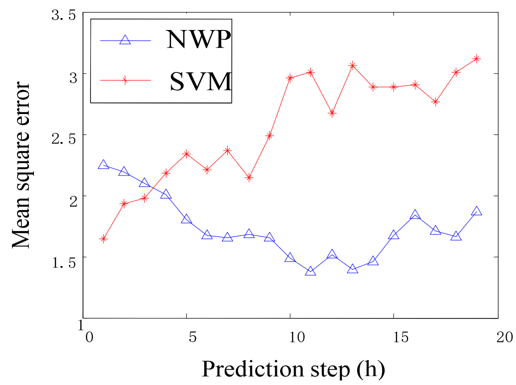

Fig.7 The mean square error curve

IV. FUSION MODEL AND EXPERIMENTAL RESULTS

Above analysis shows that: Using statistical model (which is based on measured data) to predict wind speed, when the prediction time exceeds 4 hours, the correlation function value of low frequency section is less than 0.8 and decreases with the increase of predicted time, so the credibility of prediction results begins to decrease. However, NWP has higher predictable ability, namely that the prediction results of long-term are better than instantaneous forecast. Therefore we can use the differences and complementarities between NWP and statistical model to establish fusion model. The fusion strategy is: Wind speed forecast within four hours can be obtained from NWP and historical wind speed, the specific approach is using genetic algorithms to synthesize the prediction results of NWP and statistical model. When the prediction time exceeds 4 hours, wind speed prediction results mostly rely on NWP.

In order to verify the effectiveness of the fusion model, the day-ahead wind speed prediction result of the fusion model is compared with actual wind speed, which is shown in Figure 8. As can be seen, there is a good coincidence degree between the day-ahead wind speed prediction result of the fusion model and actual wind speed.

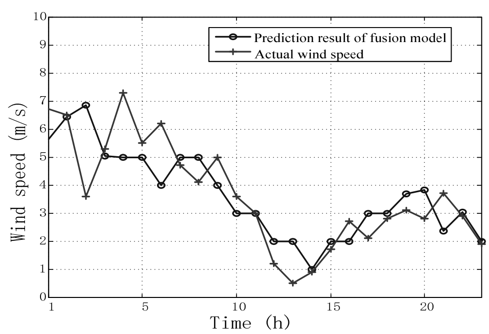

Fig.8 The time series graph of experimental results and actual wind speed

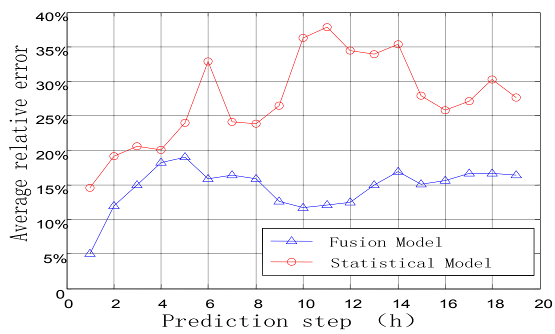

Fig.9 The average relative error variation curve for different models

To further illustrate the effectiveness of the fusion strategy, comparing the prediction result of fusion model with the result of statistical model on the same test set. The average relative error variation curve for different models is shown in Figure 9. By comparison，it can be seen in the entire test set the prediction error of fusion model is smaller than that of statistical model, which fully demonstrate the effectiveness of the fusion model and can improve the accuracy of prediction results.

V. CONCLUSION

The day-ahead wind speed forecast of 1-24 hours can guarantee the grid reasonable dispatch, ensure power supply quality and help wind farms participating in bidding for wind power integration. After considering the information efficiency of NWP and measured data, a fusion model is proposed in this paper. (1). Based on the analysis of information efficiency, it revealed that the predictable time of NWP and statistical models (which are based on measured data) are different. The short-term prediction results of statistical models is high credible, while as for NWP, the prediction results of long-term are better than instantaneous forecast. (2). The fusion modrl, which is proposed in this paper, takes full advantage of the differences and complementarity between NWP and statistical. Experiments show that the fusion model can improve the accuracy of the day-ahead wind speed prediction results.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

ACKNOWLEDGMENT

This work was supported by National Program on Key Basic Research Project (973 Program) under Grant, 2012CB215201, and by Heilongjiang Electric Power Company Science and Technology Key Projects ([2014] 190).

REFERENCES

[1] Roulston M S, Kaplan D T, Hardenberg J, et al. Using medium-range weather forcasts to improve the value of wind energy production [J]. Renewable Energy, 28(4): 585-602.2003

[2] Alboyaci B, Dursun B. Electricity restructuring in Turkey and the share of wind energy production [J]. Renewable Energy, 33(11): 2499-2505.2008

[3] Wichser C, Klink K. Low wind speed turbines and wind power potential in Minnesota, USA [J]. Renewable Energy, 33(8): 1749-1758.2008

[4] Vaccaro A, Mercogliano P, Schiano P, et al. An adaptive framework based on multi-model data fusion for one-day-ahead wind power forecasting [J]. Electric Power Systems Research, 81(3): 775-782.2011

[5] Cassola F, Burlando M. Wind speed and wind energy forecast through Kalman filtering of Numerical Weather Prediction model output [J]. Applied energy, 99: 154-166.2012

[6] Zhenqi Cai. Corrected Numerical Weather Prediction-BP Neural Network Based Wind Power Short- term Prediction Research. [D]. Zhejiang University, 2012.（in Chinese）

[7] Smagorinsky, J., Problems and promises of deterministic extened range forecasting, Bull. Amer.Meteor. Soc. 50, 286-311,1969.

[8] Lorenz E N. A study of the predictability of a 28□variable atmospheric model [J]. Tellus, 17(3): 321-333.1965

[9] Jifan Chou. Weather and climate predictability.[J] Meteorological Science and Technology Progress , 1(2): 12r14.（in Chinese）2011

[10] Zhihai Zhen. Operational technology research of 6-15 days numerical weather prediction based on predictable components [D]. Lanzhou University, 2010.（in Chinese）

[11] Zhijin Li. Analysis of spatial and temporal dependence of the actual forecast predictability [J]. Atmospheric Science, 20(3): 290-297.1996（in Chinese）