A Hybrid Forecasting Method for Wind Speed

Jian Jiao

College of Economics& Management, China Three Gorges University, 443002 Yichang City, China

Abstract. Wind energy is one of the most widely used renewable energy sources. Wind power generation is uncertain because of the intermittent of wind power. To reduce the influence of wind power generation on the power system, it is necessary to forecast wind speed. This paper presents a hybrid wind speed prediction method based on Autoregressive Integrated Moving Average (ARIMA) model and Artificial Neural Network (ANN) model. In three wind speed prediction tests, the hybrid, ARIMA and ANN models are applied respectively. By analyzing the predicted results, it can be concluded that the hybrid method has better forecasting result. By analyzing the results, we can conclude that the hybrid method has better prediction effect.

1 Introduction

Wind energy is one of the most widely used renewable energy sources in the world. In 2017, global wind power generation increased by more than 17 percent to 1,120 TWh, accounting for 4.4 percent of global total power generation. In some countries, wind power is employed in a high proportion, such as Denmark, which set a new world record in 2017, with 43 percent of its electricity coming from wind. A growing number of countries, such as Germany, Ireland, Portugal, Spain, Sweden and Uruguay, are also using wind energy at more than 10 percent. At present, China is the world's largest power installed capacity, with a total installed capacity of 188 Gigawatt in 2017 [1].

In the power system, the power generation changes in real time with the change of power demand, so that the power supply and power consumption can reach a balance. Wind power, however, is heavily impacted by the weather. The uncertainty and intermittency of wind power makes it difficult to maintain this balance. Due to the irregular change of wind energy, wind power cannot guarantee the continuous and stable output.

Currently, there are two ways to address this problem: by increasing the capacity of rotating standby in the power system, the impact of wind power volatility on the power grid can be suppressed, but this will increase the operating cost of the system. Another way is making the wind speed forecasting or wind power forecasting (WPF), and dispatching the power grid timely according to the forecast results, so as to improve the economy of wind power generation.

According to the forecast duration, WPF can be subdivided into short term, medium term and long term. Data can be predicted from 30 minutes to 6 hours in the short term, 6 hours to 1 day in the medium term and 1 day to 1 week in the long term [2].

WPF model has two types: physical model and historical data based statistical model. Based on topographical features such as obstacles, rugosity and altitude, physical models usually apply numerical weather prediction (NWP) techniques to estimate wind speed [3]. However, there are limited NWP models available for wind farms. Compared with physical models, the application of statistical models has a broader application prospect. Statistical models implement the estimation by analyzing historical data and the relationship between variables. These variables generally include wind speed data, and may also have the output of the NWP model. For statistical models, there are usually two steps to complete the prediction. The first step is training. This process can optimize the model to interpret the given data as accurately as possible. By adjusting internal variables, the error between real data and model generated data is reduced. After the training, the wind speed can be predicted [4].

Scholars have conducted extensive studies on wind speed prediction, including autoregressive moving average (ARMA) model, ARIMA model and other linear models, as well as nonlinear prediction models including support vector machine (SVM) and ANN.

All of these prediction models have been widely used, but the single prediction model has its own limitations. ARMA, ARIMA and other models based on autoregressive are suitable for linear time series prediction, and the accuracy of prediction cannot be ensured for wind speed with nonlinear components. The prediction precision of the nonlinear model is higher than that of the linear model. Support vector machine needs small sample size, but its key parameters are difficult to ascertain.

Hybrid model is the developing direction to improve the accuracy of wind speed prediction. This paper presents a wind speed prediction model combining ARIMA with ANN. The hybrid model makes ARIMA to simulate the linear characteristics of wind speed time series, and ANN to simulate the nonlinear characteristics of wind speed time series, so as to increase the prediction accuracy.

2 Mathematical models

2.1 ARIMA model

Autoregressive moving average (ARMA) model is a common method to research time series problems and describe stationary stochastic processes. The data formed by wind speed over time is regarded as a random sequence, which has certain continuity [5].

The basic idea of the ARIMA model is to make non-stationary sequences into stationary ones by using multiple differences. The ARIMA model is a general form of an ARMA model. The ARIMA model is adopted when data is non-stationary. In this case, the differences in the initial steps (the process of integration) can be used once or many times, in order to eliminate non-stationary. When the difference is completed d times, d is the integration order of ARIMA [4, 6]. Then the stationary sequence is modeled by an ARMA model with p and q as parameters and the original sequence is obtained by inverse transformation. The prediction equation of ARIMA model with p, d and q as parameters can be expressed as

$$y_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \tag{1}$$

Where, y_t is the sample value of time series, φ_i and θ_j are model parameters, ε_t is white noise with independent normal distribution and satisfies

$$E(\varepsilon_t) = 0 \quad Var(\varepsilon_t) = \sigma_\varepsilon^2 > 0 \tag{2}$$

After the time series is stabilized, model recognition and order determination can be carried out with the assistance of autocorrelation function (ACF) and partial correlation function (PAC) of time series y_t.

The expression for ACF is as follows:

$$\rho_k = \frac{\sum_{i=1}^{n-k} y_i \cdot y_{i+k}}{\sum_{i=1}^{n} y_i^2}, k = 1, 2.... \tag{3}$$

The expression for PAC is as follows:

$$\alpha_{kk} = \begin{cases} \rho_1 & k = 1 \\ \dfrac{\rho_k - \sum_{j=1}^{k-1} \alpha_{k-1,j} \cdot \rho_{k-j}}{1 - \sum_{j=1}^{k-1} \alpha_{j,j} \cdot \rho_{k-j}} & k = 2, 3, ... \end{cases} \tag{4}$$

The order of the model can be preliminarily determined by ACF and PAC. Parameter identification of time series can be obtained by least square estimation, that is, the estimated parameters $\varphi_1, \varphi_2 ... \varphi_p, \theta_1, \theta_2 ... \theta_q$.

The optimal parameter model can be obtained through the AIC criterion.

After the process of identification and estimation, it is necessary to test whether the model is applicable or not. Only when the test is passed, can the model be reflected to meet the requirements and be qualified for the prediction task. Conversely, if it does not meet the requirements, the model must be modified or re-identified, and the process must be repeated until the requirements are met.

In the estimation process and the recognition process of the above models, the premise is to assume that the disturbance term is white noise. Therefore, for the existing model, $\hat{\varepsilon}_t = y_t - \hat{y}_t$ ($\hat{y}_t$ is the predicted value) needs to be tested to determine whether it belongs to the white noise sample sequence.

2.2 ANN model

The algorithm of ANN is developed in the simulation of human brain function and structure, and a series of simple processing elements are connected through some topology structures, which can effectively solve the problem of complex problems, especially nonlinear.

One of the widely used methods for predicting wind speed is ANN. The basic structure of ANN is composed of input layer, hidden layer and output layer. Layers of the ANN consist of a large number of nodes (also called neurons). Nodes are connected to each other, and each node represents a specific output function.

The learning process of neural network is composed of forward propagation and reverse propagation. In the process of forward propagation, the input information is processed layer by layer from the input layer through the hidden layer to the output layer. The state of each layer of neurons only affects the state of the next layer of neurons. If the desired output cannot be obtained at the output layer, the reverse propagation will return the error signal along the original connection path. When the error of the output layer node propagates back to the input layer, the connection weights and thresholds are adjusted so that the neural network can output the required mapping.A well-trained network has the ability to generalize, which makes predictions possible.

By training, ANN can simulate the complex nonlinear relationship [7]. The method has many virtues. For example, it is unnecessary for ANN to have calculation formulation and it has the ability of self-adaption, self-organizing and self-learning [8, 9].

Figure 1 shows a typical ANN structure with M inputs.The inputs ($\{y_{t-i1}, y_{t-i2}, \cdots, y_{t-im}\}$) are selected from the previous ω data ($\{y_{t-1}, y_{t-2}, \cdots, y_{t-\omega}\}$) . The output of hidden layer ($\{x_1, x_2, \cdots, x_n\}$) with n nodes is $\hat{y}_t$. The output is computed as follows:

$$\hat{y}_t = f_1(\sum_{i}^{n} \omega_i \cdot x_i) \tag{5}$$

$$x_i = f_2(\sum_{j}^{m} \omega_{ij} \cdot y_{t-ij}), i \in [1, n] \tag{6}$$

where f_1 and f_2 are activation functions, and ω is weight of ANN.

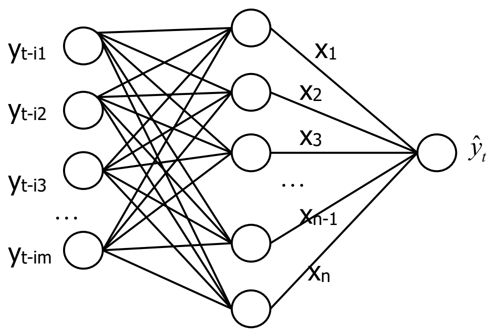

Figure 1. Typical artificial neural network structure.

2.3 Hybrid forecast method

ARIMA and ANN have different characteristics and applications. ARIMA method is suitable for the prediction of linear problems, and the prediction accuracy for nonlinear problems is low. Compared with ARIMA model, ANN model is better at solving nonlinear problems. The hybrid forecast model can predict the linear part of wind speed by ARIMA method, and the non-linear part of wind speed can be predicted by the ANN method.

The mathematical expression for the hybrid model of wind speed forecast is as follows:

$$y_t = AM_t + AN_t \tag{7}$$

Where AM_t is the linear part ARIMA, and AN_t is the nonlinear part ANN. The linearity of data is obtained by applying ARIMA model. After ARIMA model fitting, the residual only includes the nonlinear part of the data. The main process of hybrid prediction is as follows. ARIMA model is established for the linear part of data and the predicted value $AR_t^{'}$ is calculated. The residuals of the ARIMA model are obtained by comparing with observed values. Fitting the residuals by ANN model, the recasting value of the residuals $AN_t^{'}$ is obtained. The forecasted value is obtained by adding up $AR_t^{'}$ and $AN_t^{'}$. Then the forecast result for the ARIMA and ANN hybrid model becomes

$$y_t = AR_t^{'} + AN_t^{'} \tag{8}$$

2.4 Wind speed data preprocessing

Due to some uncertain factors, data loss and singularity may occur in the collected wind speed. Therefore, the existing wind speed data must be tested, interpolated and preprocessed to obtain the accurate original wind speed data.

The preprocessing steps of wind speed data are as follows: according to China's national standard wind farm wind energy resource assessment method, the wind speed and wind direction data of different height of the wind tower are analyzed and counted, and the suspicious data and missing data are listed. For suspicious data, the effective data is selected and put back to the original database. The unreasonable invalid data is preprocessed by the wind-shear index calculation method, and the missing data is finally interpolated.

3 Statistical error measures

To measure the model quantitatively, two kinds of prediction error indexes, mean square error (MSE) and mean absolute error (MAE), are applied for model evaluation and model comparison.

In mathematical statistics, the MSE refers to the expected value of the square of the difference between the estimated value of parameters and the true value of parameters. MSE is a convenient method to measure "average error". MSE can evaluate the change degree of data. The smaller the value of MSE, the more accurate the prediction model can be in describing experimental data.

MAE is the average absolute value of the deviation between all individual observed values and arithmetic mean values. The average absolute error can avoid the problem that the errors cancel each other, so it can accurately reflect the actual prediction error.

Suppose that A_t is the real wind speed value at time t, F_t is the predicted value at the same time, and error e_t is:

$$e_t = A_t - F_t \tag{9}$$

The formulae for these two metrics are shown below

$$MSE = \frac{1}{n}\sum_{t=1}^{n} e_t^2 \tag{10}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |e_t| \tag{11}$$

where n is the number of predicted values.

4 Results and discussion

Three tests were performed to verify the models. Hybrid, ARIMA and ANN methods are applied in each test and the prediction results of the three methods are compared.

It can be seen from Figure 2 (test1) that the predicted values of the three models all approximate to the real data, but the prediction error of the hybrid model is smaller. The prediction results of ARIMA and ANN models are delayed in time compared with real data.

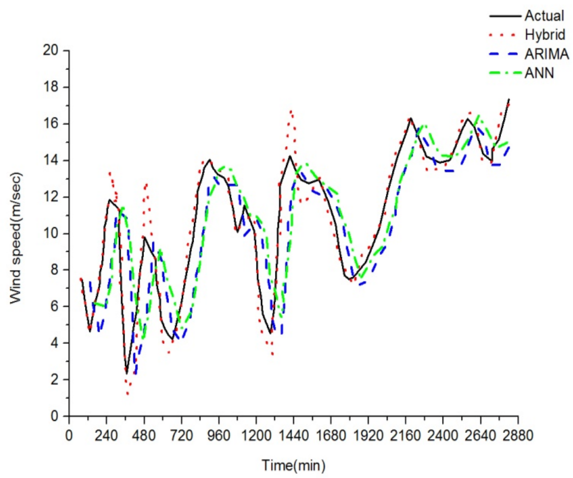

Figure 2. Wind speed forecasting test1.

In figure 3 (test2), the prediction data of all three models are consistent with the trend of real data. The forecast data of the hybrid model is very close to the real data. The difference between the forecast data of ARIMA model and the real data is small, while the prediction data of ANN model is significantly different from the real data.

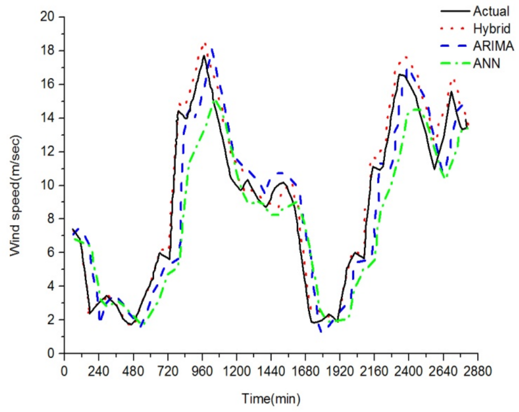

Figure 3. Wind speed forecasting test2.

It can be seen from figure 4 (test3) that the predicted data of the hybrid model is highly consistent with the real data. There is a slight delay in the data obtained from the ARIMA model, while there is a significant delay between the data predicted by the ANN model and the real data.

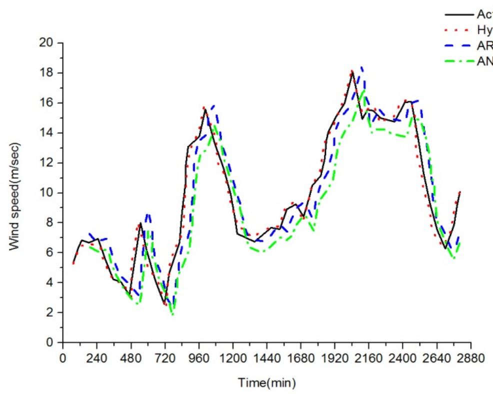

Figure 4. Wind speed forecasting test3.

To compare the three models quantitatively, the statistical errors of each model in the three tests are listed in table 1. The calculated data show that the hybrid model has the lowest value of all MAE and MSE values. That is, the hybrid model is the best way to forecast wind speed. The results show that the hybrid model has higher accuracy.

Table 1. Summary of test statistical errors.

Test	Method	MAE	MSE
Test1	ARIMA	1.3524	3.9537
	ANN	1.6532	4.9576
	hybrid	0.4677	0.3532
Test2	ARIMA	0.8976	1.6587
	ANN	0.9074	1.9544
	hybrid	0.4122	0.2361
Test3	ARIMA	0.5899	0.6325
	ANN	0.6443	0.7442
	hybrid	0.0554	0.0068

5 Conclusions

In the application of wind energy, wind speed forecast is of great significance for stable operation of power system and improving its operational efficiency.In this paper, a wind speed prediction hybrid model including ARIMA model and ANN model is proposed. Through the analysis of test data, ARIMA, ANN and hybrid models can forecast wind speed well. The prediction accuracy of different models is evaluated by MSE and MAE. Compared with ARIMA model and ANN model, the hybrid model can make better use of linear and nonlinear components to forecast wind speed, and its prediction accuracy is higher.The wind speed forecasting model proposed in this paper can also be used in power load forecasting and other forecasting fields, and provides relevant theoretical methods for similar forecasting.

References

1. Global Wind Energy Council, *Global wind report* (GWEC, Brussels, 2018)
2. A. Togelou, G. Sideratos, N. D. Hatziargyriou, IEEE Transactions on Sustainable Energy **3**, 416 (2012)
3. A. Tascikaraoglu, M. Uzunoglu, Renewable and Sustainable Energy Reviews **34**, 243 (2014)
4. E. Erdem, J. Shi, Applied Energy **88**, 1405 (2011)
5. D. C. Montgomery, C. L. Jennings, M. Kulahci, *Introduction to time series analysis and forecasting*, (John Wiley & Sons, New York, 2015)
6. C. N. Babu, B. E. Reddy, Applied Soft Computing **23**, 27 (2014)
7. D. Fadare, Applied Energy **87**, 934 (2010)
8. T. Gao, H. Li, Y.-H. Lu, ICIC Express Letters, Part B: Applications **2**, 1201 (2011)
9. P. Gomes, R. Castro, International Journal of Sustainable Energy Development **1**, 275 (2012)