Renewable and Sustainable Energy Reviews 16 (2012) 3471–3480

Contents lists available at SciVerse ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

Evaluation of hybrid forecasting approaches for wind speed and power generation time series

Jing Shi*, Jinmei Guo, Songtao Zheng

Department of Industrial and Manufacturing Engineering, North Dakota State University, Dept 2485, PO Box 6050, Fargo, ND 58108, USA

ARTICLE INFO

Article history:
Received 5 May 2011
Accepted 19 February 2012
Available online 7 April 2012

Keywords:
Wind speed
Wind power
Hybrid forecasting
ARIMA
ANN
SVM

ABSTRACT

Forecasting of wind speed and wind power generation is indispensible for the effective operation of a wind farm, and the optimal management of its revenue and risks. Hybrid forecasting of time series data is considered to be a potentially viable alternative compared with the conventional single forecasting modeling approaches such as autoregressive integrated moving average (ARIMA), artificial neural network (ANN), and support vector machine (SVM). Hybrid forecasting typically consists of an ARIMA prediction model for the linear component of a time series and a nonlinear prediction model for the nonlinear component. In this paper, we systematically and comprehensively investigate the applicability of this methodology based on two case studies on wind speed and wind power generation, respectively. Two hybrid models, namely, ARIMA–ANN and ARIMA–SVM, are selected to compare with the single ARIMA, ANN, and SVM forecasting models. The results show that the hybrid approaches are viable options for forecasting both wind speed and wind power generation time series, but they do not always produce superior forecasting performance for all the forecasting time horizons investigated.

Contents

* Corresponding author. Tel.: +1 701 231 7119; fax: +1 701 231 7195.
E-mail address: jing.shi@ndsu.edu (J. Shi).

1364-0321/$ – see front matter
doi:10.1016/j.rser.2012.02.044

1. Introduction

Wind energy is an intermittent power source, which has been making significant penetration into the electricity markets in various countries around the world. As a result, the accurate predictions of wind speed and power generation become increasingly important [1]. Short-term forecasting of wind speed is critical to the operation of wind turbines so that dynamic control can be accomplished to increase the energy conversion efficiency and reduce the risk of overloading [2]. With the greater penetration of wind power, the forecasting of wind speed and power generation plays a pivot role in improving energy market efficiency, reducing the amount of reserves while maintaining the system security, and maximizing wind generators' revenue by optimizing their daily and/or intraday bids in the electricity markets.

A number of time series forecasting methods have been successfully applied to the short-term prediction of wind speed and power generation. Autoregressive integrated moving average (ARIMA) family is one of the most robust and widely used approach in wind forecasting [3,4]. Included in this family are autoregressive (AR), autoregressive moving average (ARMA), factional ARIMA (fARIMA), and seasonal ARIMA (sARIMA). These models can explicitly reveal the relationship between the inputs and outputs, but they are generally limited in linear forms. Artificial intelligence (AI) and machining learning (ML) approaches are also frequently adopted for wind forecasting. The AI/ML models reported in related literature include various artificial neural network (ANN) models such as back propagation (BP) and radial basis function (RBF) [5,6], support vector machine (SVM) [7], fuzzy logic [8]. In general, the AI/ML models can better handle non-linear relationship and thus are more flexible, but they describe the relationship in implicit ways and sometimes are very computationally intensive. In addition to these single model structures, two new forecasting methodologies are emerging, namely combined forecasting (or ensemble forecasting), and hybrid forecasting.

In general, combined forecasting tackles the task in two steps, with the first step being to make forecasts using multiple plausible models, and the second step being to combine these forecasts into a single forecast time series using weighting algorithms. The combined forecasting method is first proposed for wind power forecasting by Sanchez [9], in which a new prediction tool using an adaptive combination of a variety of statistical models is applied. Li and Shi [10] apply the Bayesian model averaging approach to combine the forecasts from ANN and ARIMA models. It is found that the resultant forecast can always perform better or close to the best individual models. It should be noted that the ensemble approaches can also be used to combine the forecasts from different numerical weather prediction (NWP) models in literature. For instance, Nielsen et al. [11] combine a number of meteorological forecasts originating from three various global NWP models. Also, Thordarson et al. [12] propose conditional weighted combination of wind power forecasts, and the results show that the proposed combination method outperforms the least-squares combination method for almost all prediction horizons.

On the other hand, hybrid forecasting methodology takes a different approach. It usually employs a linear model for the prediction of the linear component and a non-linear model for the non-linear component in time series. Tseng et al. [13] hybridize an sARIMA model with a neural network model to forecast the total production value of Taiwan machinery industry, and show that the sARIMA–ANN hybrid model outperforms the individual sARIMA or ANN models. Zhang [14] proposes a hybrid methodology that combines both ARIMA and ANN models to take advantage of the unique strength of ARIMA and ANN models in linear and nonlinear modeling. It is claimed that the hybrid model can improve forecasting accuracy achieved by either of the models if used separately. Similarly, Hansen and Nelson [15] compare the performance of ANN models with the hybrid ARIMA–ANN models, and indicate that the hybrid approach has the potential to produce more accurate forecasts and exhibit better robustness. Pai and Lin [16] investigate a hybrid methodology that exploits the unique characteristic of ARIMA and SVM models in forecasting stock price problems. The results show that the proposed hybrid model has lower forecasting errors compared with the single forecasting models. Gutierrez-Estrada et al. [17] propose an ARIMA–ANN hybrid model for the purpose of forecasting anchovy catch time series. The results show that the hybrid model can be an effective tool for improving the forecasting accuracy. Although the hybrid forecasting methodology has been adopted in some fields, its applications in energy related applications have been slow. To our best knowledge, the only relevant publication for wind forecasting is a recent publication by Cadenas and Rivera [18]. In that study, the ARIMA–ANN hybrid approach is adopted to forecast the wind speed for a fixed prediction horizon in different regions in Mexico. The improvement of prediction accuracy jumps 50–85% in terms of MAE and 85–99% in terms of RMSE compared with the individual ARIMA or ANN models for all the tested sites. However, the amount of improvement reported is surprising because those reported in other literature are much more modest (e.g., 1–10% according to [16]). More importantly, Taskaya-Temizel and Casey [19] indicate that hybrid approaches do not necessarily improve the forecasting performance, and sometimes they might be even slightly worse than the individual single forecasting models.

In this paper, we aim to rigorously investigate the applicability of hybrid forecasting methodology for wind time series. Two hybrid forecasting model structures, namely, ARIMA–ANN and ARIMA–SVM, are adopted, and they are compared with the individual ARIMA, ANN, and SVM forecasting models. To make the study comprehensive, our investigation adopts multiple forecasting horizons, which varying from 1-step ahead to 9-step ahead. Also, we conduct the tests on two separate types of wind time series data – the hourly wind speed time series, and the hourly wind power generation time series. The reminder of the paper is organized as follows. In Section 2, the structures and procedures of the single and hybrid forecasting models are briefly explained. In Section 3, the performance metrics and the wind time series are introduced. In Section 4, the model parameters are obtained and the performances of different models are compared, and the results are analyzed and discussed. In Section 5, conclusive remarks are made and future possible research is pointed out.

2. Methodology

2.1. ARIMA model

ARIMA models established by Box and Jenkins [20] have been widely used for the purpose of time series forecasting. An ARIMA model is linearly combined by several previous points and random errors, and the forecast is a function of the past observations and the errors. The conventional ARMA(p, q) formulation is described as,

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

where δ is a constant term, ϕ_i is the ith autoregressive coefficient, θ_j is the jth moving average coefficient, ε_t is the error term at time t, ε_{t-j} is the random error of a prior points at time $t-j$, p and q are the orders of autoregressive and moving average terms, respectively. If the time series data is not stationary, it should be differenced to become stationary. This results in an "integrated" ARMA (i.e., ARIMA) model, denoted by ARIMA(p, d, q), where d is the order of

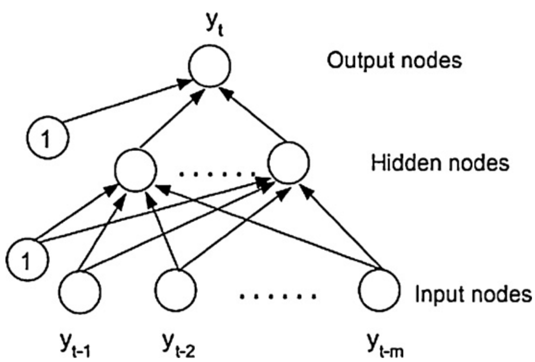

Fig. 1. Typical neural network structure.

differencing. Building an ARIMA model includes three major steps: model identification, parameter estimation, and diagnostic checking. In model identification process, one or more model candidates could be found suitable for the time series. In such case, autocorrelation function (ACF) and partial autocorrelation function (PACF) can be applied to make the first guess about the orders of the ARIMA model. However, if the models show both autoregressive and moving average nature, such method cannot identify the orders since both ACF and PACF will show exponential decay and damped sinusoid. In this case, other criteria should be adopted to determine the order of the ARIMA model. The typical criteria are Akaike's information criterion (AIC) and Bayesian information criterion (BIC). Once the model is identified, the parameters need to be estimated, and in principle the selected parameters should generate the lowest residual. This can be accomplished by using the Yule–Walker Estimation or Maximum Likelihood Estimation. A common method is to test the randomness of the residuals using Ljung–Box Statistics, and non-significant P-values indicate that the residuals are uncorrelated and the proposed model is suitable for fitting the historical data.

2.2. ANN model

Artificial neural network is also widely used for time series forecasting. There are various ANN model structures available in the literature, but the feedforward neural network has been the most widely used neural network model [21,22]. A typical feedforward neural network model used for forecasting purpose is shown in Fig. 1. The previous observations are the input of the neural network model, while the output is the forecasted value. Hidden layer stores an appropriate transfer function which is used for processing the data from the input nodes. The model can be written as:

$$y_t = w_0 + \sum_{j=1}^{Q} w_j g\left(w_{0j} + \sum_{i=1}^{P} w_{i,j} y_{t-i} \right), \tag{2}$$

where P is the number of input nodes, Q is the number of hidden nodes, g is a sigmoid transfer function. $\{w_j, j = 0, 1, \ldots, Q\}$ is a vector of weights from the hidden layer to output nodes, $\{w_{i,j}, i = 1, 2, \ldots, P, j = 1, 2, \ldots, Q\}$ are the weights from the input to hidden nodes, w_{0j} are the weights for each output between input and hidden layer.

However, the determination of the number of input and hidden nodes of the neural network architecture is non-trivial. Different architectures, e.g., numbers of input and output neurons, need to be compared to select the model with the best performance. In order to avoid overfitting problem, the available data should be divided into three parts. The first part of the data set is model training set, the second is validation set for model selection, and the last part of the available data is testing set which is used for evaluating the forecasting performance.

2.3. SVM model

Support vector machines (SVMs) are a popular approach for classification, data mining, forecasting, and other statistical analysis applications. The foundation of SVM, different from that of ANN approach, is that this approach formulates the statistical learning problem as a quadratic programming model with linear constraints. The general idea of support vector machines (SVM) for regression is to generate the regression function by applying a set of high dimensional linear functions. SVM also seeks to minimize an upper bound of the generalization error based on structural risk minimization [7,16]. Suppose $f(x)$ takes the following form,

$$f(x) = \mathbf{w}^T \phi(x) + b, \tag{3}$$

where $\phi(x)$ is the high dimensional feature space of the input vector $\mathbf{x} = (x^1, x^2, \ldots, x^K)^T$, K is the order of SVM, and $\mathbf{w}$ and b are the weight vector and bias term respectively. Least squares support vector machines (LS-SVM) are basically the SVM with only equality constraints, and the solution can be obtained by solving linear Karush–Kuhn–Tucker systems. It produces comparable results compared with the general SVM in a more efficient manner, and thus it is adopted in this study.

Suppose training data $\{(\mathbf{x}_i, y_i)\}_{i=1}^{l}$ is given, LS-SVM determines the optimal $\mathbf{w}$ and b by minimizing the following function [7],

$$\min R = \frac{1}{2}\|\mathbf{w}\|^2 + \frac{1}{2}\gamma\|\mathbf{e}\|^2 \tag{4}$$

$$\text{subject to,} \quad y_i = w^T\phi(x_i) + b + e_i, \quad i = 1, 2, \ldots, l. \tag{5}$$

where $\mathbf{e} = (e_1, e_2, \ldots, e_N)^T$, and γ is the regularization parameter that controls the trade-off between the bias and variance of LS-SVM model. By introducing Lagrangian multipliers and the Karush–Kuhn–Tucker theorem, the quadratic optimization problem can be solved and the SVM regression model for estimating y in Eq. (3) becomes

$$y = \sum_{i=1}^{l} \alpha_i k(\mathbf{x}, \mathbf{x}_i) + b, \tag{6}$$

where $k(\mathbf{x}, \mathbf{x}_i)$ is called kernel function, and α_i are the Lagrange multipliers. Gaussian kernel is recognized as a robust kernel function which generally generates satisfactory performance, and it follows the following form,

$$k_G(\mathbf{x}, \mathbf{x}_i) = \exp\left(\frac{-||\mathbf{x} - \mathbf{x}_i||^2}{\sigma^2} \right), \tag{7}$$

where $\|\cdot\|$ denotes the 2-norm, and σ is a constant determining the width of Gaussian kernel.

2.4. Hybrid forecasting

It is well documented in literature that ARIMA, ANN, and SVM have their own advantages and none of them can outperform the others universally in terms of the forecasting accuracy [19,23]. ARIMA models usually can only capture linearity information, and thus it has deficiency in the prediction of nonlinear characteristics of data. In such cases, the nonlinear modeling methods such as ANN and SVM may outperform ARIMA. In other words, the performances for different forecasting methods depend upon the linearity of given data. A hybrid methodology that is able to capture both linear and nonlinear characteristics is thus believed to be a good strategy for time series forecasting [14,22]. It is well known that the time series of wind speed and wind power are highly volatile.

This presents a major challenge for generating accurate prediction, and is an inspiring factor for one to adopt hybrid forecasting as a potential viable approach for improving the forecasting accuracies. The hybrid methodology typically employs an ARIMA model for the linear characteristics and an ANN or SVM model for the nonlinear characteristics. In general, the hybrid model can then be represented as follows:

$$y_t = L_t + N_t, \quad (8)$$

where L_t denotes the linear component, and N_t represents the nonlinear component. First of all, an ARIMA model is used to capture the linearity of the data, L_t. Then, the residuals obtained after fitting with the ARIMA model will only contain the nonlinear part of the problem. Let e_t denote the residuals from the linear fitting at time t. Hence,

$$e_t = y_t - \widehat{L_t}, \quad (9)$$

where $\widehat{L_t}$ is the forecasted value obtained by applying ARIMA methods. By applying ANN or SVM method, the nonlinear part of the model should be captured. For the case of ANN, suppose that the input for the model is n, the ANN model for the residuals will be,

$$e_t = g(e_{t-1}, e_{t-2}, \ldots, e_{t-n}) + \varepsilon_t, \quad (10)$$

where g is the function for nonlinear model determined by the ANN model, and ε_t is the random error for the model fit. Thus, we denote the forecasted residual by ANN as $\widehat{N}_t$, the forecast result for the ARIMA–ANN hybrid model becomes,

$$y_t = \widehat{L_t} + \widehat{N}_t, \quad (11)$$

In brief, the procedure for hybrid forecasting consists of the following three steps: (1) ARIMA model is constructed to capture the linear part of the problem. $\widehat{L_t}$ is obtained, and the residuals of the ARIMA model are also obtained by comparing with the observations; (2) The residuals obtained in the first step are fitted by a non-linear approach such as an ANN or SVM model. The forecasting value for the residuals, $\widehat{N}_t$, is thus obtained; (3) The forecasted values are obtained by adding the estimations for the linear and nonlinear components of the time series determined in steps 1 and 2.

3. Wind data and forecasting performance evaluation

To fairly compare the performances of the forecasting models (i.e., ARIMA, ANN, SVM, ARIMA–ANN, ARIMA–SVM), we use multiple forecasting horizons. As indicated by Fig. 2, five different forecasting horizons are included, and they are 1, 3, 5, 7, and 9-step ahead, respectively. In the case of 1-step ahead, we use the M observations to predict the 1-step ahead value. For the forecasting horizons more than 1 step, rolling forecasting mechanism is used. For instance, 3-step ahead forecasting is performed based on the available M-2 observations and the predicted values of the 2 data points which are closest to the prediction point but currently unavailable from observation. In fact, the second closest point is predicted first and the first closest point is predicted based on M-1 observations and the predicted value of the second closest point. Similarly, 5, 7 and 9-step ahead forecasts make use of the closest 4, 6, and 8 predicted data points respectively.

3.1. Wind data

As mentioned previously, this study investigates the forecasting accuracies for both wind speed and wind power generation. It should be noted that wind power generation can be indirectly predicted by converting the wind speed forecasts into wind power values via the established power curve of a wind turbine. However, this approach may introduce more randomness to the forecasting results due to the stochastic nature of the power curves [24]. As a result, we choose to directly predict wind power generation based on historical time series data in this study. As such, two separate datasets are obtained and used for wind speed and power generation forecasting, respectively. For wind speed forecasting, a 2-year hourly dataset is retrieved from a wind observation site in Colorado, USA from 2005 to 2007. The average wind speed is 4.93 m/s, and the standard deviation is 3.70 m/s. As for wind power generation, a 2-year hourly dataset is retrieved from a 1.5 MW wind turbine located in North Dakota, USA from 2005 to 2007. The average wind power generated is 300.14 kW, and the standard deviation is 254.35 kW. Figs. 1 and 2 show the time series plots of the two datasets, respectively.

3.2. Forecasting performance evaluation

Both MAE (mean absolute error) and RMSE (root mean square error) are used as the metrics of forecasting accuracy. The formulations of the two metrics are shown below:

$$\mathrm{MAE} = \frac{1}{N}\sum_{t=1}^{N}\left|o_t - f_t\right|, \quad (12)$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{t=1}^{N}(o_t - f_t)^2}{N}}, \quad (13)$$

where N is the number of forecasting periods, o_t is the observation at time t, and f_t is the forecasted value at time t.

4. Results and discussion

For both wind speed and wind power generation datasets, the model parameters need to be obtained for the single forecasting models (i.e., ARIMA, ANN, and SVM). Also, after the forecasting results of ARIMA are obtained, the residuals for the linear component are calculated by differencing the observations by the forecasting results of ARIMA. Based on the obtained residual dataset, the nonlinear methods (ANN and SVM) are utilized to describe the nonlinear pattern of the residuals. As a result, the parameters for the hybrid forecasting models (ARIMA–ANN and ARIMA–SVM) can be obtained. Then the prediction is made for a period of 744 points (hours) at various forecasting horizons based on the constructed models.

4.1. ARIMA model parameters

To find the proper ARIMA models for a time series, we should first check if the time series is stationary. Typically, the run plot and the auto correlation function (ACF) plot can be inspected to analyze the existing trend in time series [4,25]. If the data is not stationary, the differencing of time series will be needed and the order should to be determined. For our cases, the run charts of wind speed and power generation (as shown in Figs. 3 and 4 respectively) indicate no trends or seasonal patterns from either time series. The ACF graphs are also constructed for both time series data sets, and they are shown in Figs. 5 and 6. Both ACF graphs indicate an exponential decay pattern which suggests the stationarity of data. As a result, the ARMA (p, q) model structure, instead of the general ARIMA structure, is adopted for both data sets.

Bayesian information criterion (BIC) is a common tool to evaluate the goodness-of-fit of a parametric model. A major advantage of this criterion is that it can effectively address the problem of overfitting. BIC is used for ARMA model selection here, i.e., obtaining the orders for the autoregressive and moving average. The combination of orders for the autoregressive and moving average that generate

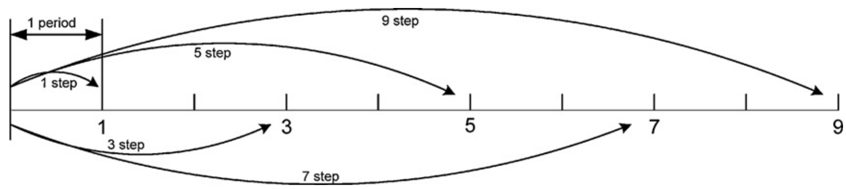

Fig. 2. Multiple forecasting horizons.

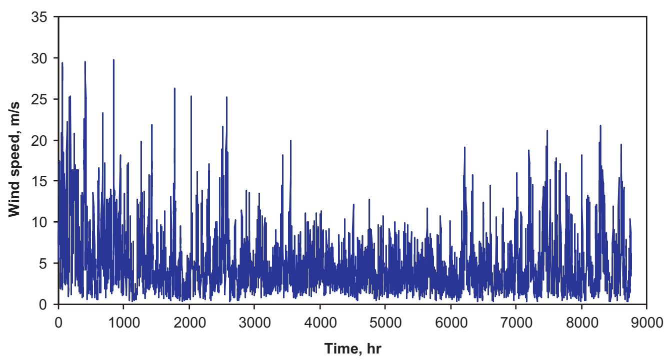

Fig. 3. Hourly wind speed time series of an observation site in Colorado from 2005 to 2007.

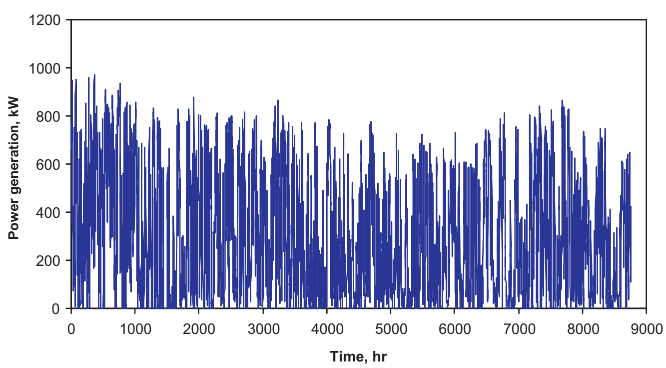

Fig. 4. Hourly wind power time series of a 1.5 MW wind turbine in North Dakota from 2005 to 2007.

the smallest BIC value will be selected. BIC can be calculated as follows,

$$\mathrm{BIC} = \ln(\hat{\sigma}_e^2) + \frac{2k\ln n}{n}, \tag{14}$$

where n is the sample size, k is the number of estimated parameters, $\hat{\sigma}_e^2$ is the error variance, which can be defined as,

$$\hat{\sigma}_e^2 = \frac{1}{n}\sum_{1}^{n}(x_i - \hat{x}_i)^2, \tag{15}$$

where x_i is the actual observation and $\hat{x}_i$ is the value obtained by the specified ARMA model.

Table 1 lists the BIC values for the wind speed dataset. It can be seen that ARMA (1,3) generates the lowest BIC value of 1.265578. As such, the ARMA (1,3) model is selected as the model for wind speed forecasting. This is then followed by parameter estimation phase. The process of parameter estimation consists of two steps: the first step is preliminary estimation, and the second step uses the results of the first step as the input for a more accurate estimation. In the preliminary estimation, Yule–Walker algorithm is applied for the autoregressive coefficients, while the moving average parameters are obtained by computing the corrected autocovariances and then applying afterwards Newton–Raphson algorithm proposed by Box and Jenkins [20]. For the second estimation of the parameters, a forecast is performed backwards with both autoregressive and moving average parameters determined by minimizing the sum of the squares of the residuals generated [26]. By applying the procedure, we obtain the value for each parameter of the ARMA(1,3) model as follows, $\delta = 4.96641$, $\phi_1 = 0.91465$, $\theta_1 = 0.02226$, $\phi_2 = 0.1576$, $\phi_3 = 0.08614$.

The identical procedure is performed on the wind power generation time series. To choose the suitable ARIMA model structure, a BIC table is constructed as shown in Table 2. It can be seen that among all the BIC values for candidate ARMA models, ARMA (1,1) shows the lowest BIC value of 8.934248. In this way, ARMA (1,1) is selected as the model for predicting wind power generation. Moreover, the parameters of the ARMA (1,1) model are: $\delta = 307.86063$, $\phi_1 = 0.91131$, $\theta_1 = 0.21857$.

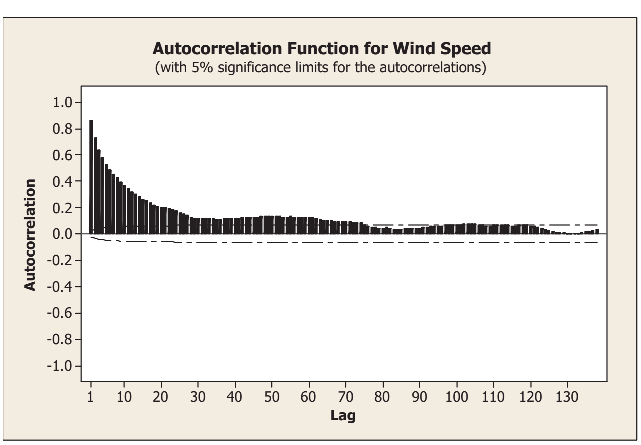

Fig. 5. Autocorrelation function (ACF) plot for wind speed time series.

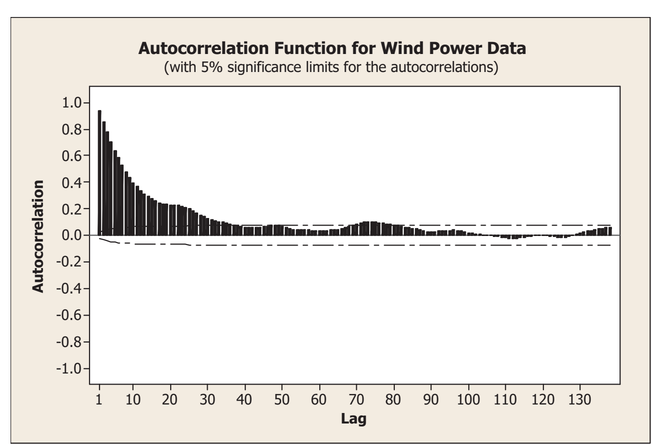

Fig. 6. Autocorrelation function (ACF) plot for wind power generation time series.

Table 1
BIC values for candidate ARMA models (wind speed).

Order	MA(0)	MA(1)	MA(2)	MA(3)	MA(4)	MA(5)
AR(0)	2.614717	2.384672	2.230869	2.120567	2.023312	1.944551
AR(1)	1.281284	1.280705	1.26991	1.265578	1.266463	1.266527
AR(2)	1.279943	1.279948	1.266539	1.266305	1.26726	1.267486
AR(3)	1.2696	1.26578	1.266791	1.267252	1.268099	1.268428
AR(4)	1.26774	1.266779	1.267745	1.266986	1.267478	1.26827
AR(5)	1.268252	1.267435	1.268072	1.267225	1.268256	1.269269

Table 2
BIC value for candidate ARMA models (wind power).

	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	11.07563	10.91435	10.75584	10.59527	10.44081	10.29859
AR 1	8.974622	8.934248	8.935247	8.935777	8.936508	8.937313
AR 2	8.935759	8.935246	8.93628	8.936807	8.937534	8.938144
AR 3	8.935502	8.935868	8.936868	8.937677	8.938314	8.939031
AR 4	8.935587	8.936523	8.937252	8.93828	8.939202	8.940067
AR 5	8.936361	8.937372	8.938201	8.939072	8.940101	8.94096

Table 3
Ljung–Box statistics for ARMA models of wind speed and wind power generation.

	Wind speed				Wind power generation			
Lag	12	24	36	48	12	24	36	48
Chi-square	18.7	31	42.1	50.1	11.7	47	62.9	73.4
DF	7	19	31	43	9	21	33	45
P-Value	0.009	0.04	0.088	0.212	0.232	0.001	0.001	0.005

Table 4
ANN model selection for wind speed data (1-step ahead).

Input	Learning rate									
	0.025	0.05	0.075	0.1	0.125	0.15	0.175	0.2	0.225	0.25
1	1.329771	1.322478	1.322478	1.322478	1.322478	1.322478	1.321627	1.321627	1.321627	1.321627
2	1.317878	1.325949	1.325949	1.325949	1.325949	1.325949	1.325949	1.325949	1.325949	1.325949
3	1.281395	1.281395	1.281395	1.281395	1.281395	1.281395	1.281395	1.262748	1.262748	1.262748
4	1.292195	1.292195	1.292195	1.292195	1.302301	1.302301	1.310347	1.310347	1.310347	1.310347
5	1.30073	1.30073	1.30073	1.30073	1.30073	1.295376	1.295376	1.295376	1.295376	1.295376
6	1.289828	1.289828	1.289828	1.289828	1.289828	1.289828	1.289828	1.289828	1.289828	1.289828
7	1.296885	1.293163	1.293163	1.293163	1.293163	1.293163	1.293163	1.293163	1.293163	1.293163
8	1.338694	1.326093	1.326093	1.326093	1.326093	1.326093	1.326093	1.286868	1.286868	1.286868

In order to test the ARMA models with the obtained parameters, we analyze the correlogram of the residuals and a global contrast is applied using Ljung–Box statistics:

$$Q = n \cdot \sum_{k=1}^{L} r_k^2(a), \tag{16}$$

where $r_k^2(a)$ is the autocorrelation coefficient of the residuals and L is the maximum delay considered. The test results are shown in Table 3. It can be seen that for both wind speed and power generation time series, the Ljung–Box statistics are smaller than the corresponding critical χ^2 values at the significance level of 0.05. As a result, the residuals of fitted models are not auto correlated based on the selected number of the lags. In other words, the model we obtained above satisfies the condition that the Ljung–Box statistic follows a χ^2 distribution, and thus they are validated and are ready to be used for forecasting.

4.2. Single ANN models

In this study, we adopt backpropagation (BP) neural network with Bayesian regularization (BR)/Levenberg-Marquardt

Table 5
ANN model selection for wind power data (1-step ahead).

Input	Learning rate									
	0.025	0.05	0.075	0.1	0.125	0.15	0.175	0.2	0.225	0.25
1	57.19957	57.19957	57.19957	57.32447	57.13274	57.13274	57.13274	56.3941	56.3941	56.97271
2	56.6886	56.39437	56.39437	56.39437	56.39437	56.67141	56.67141	56.67141	56.67141	56.67141
3	58.64867	58.64867	58.64867	58.64867	57.56069	57.56069	57.56069	57.64675	57.64675	57.64675
4	57.54116	57.54116	57.54116	57.54116	57.54116	57.54116	57.54116	56.29728	56.29728	56.29728
5	57.73545	57.73545	56.78945	55.89285	55.89285	55.89285	55.89285	55.89285	55.89285	55.89285
6	58.30632	57.46852	57.46852	57.46852	57.46852	57.46852	57.46852	57.46852	57.46852	57.46852
7	56.99647	56.99647	56.99647	56.99647	56.99647	56.99647	56.99647	56.99647	56.99647	56.99647
8	57.47547	57.47547	57.47547	57.47547	57.47547	57.47547	57.47547	57.47547	57.47547	57.47547

Table 6
Best ANN models for wind speed data.

Forecasting horizon	Learning rate	Input number	MAE
1-step	0.2	3	1.26274841
3-step	0.05	5	2.097660089
5-step	0.025	3	2.378312314
7-step	0.025	7	2.524579238
9-step	0.025	4	2.719195818

Table 7
Best ANN models for wind power data.

Forecasting horizon	Learning rate	Input number	MAE
1-step	0.1	5	55.89285343
3-step	0.025	8	113.9330078
5-step	0.05	6	155.0666085
7-step	0.1	3	182.8446967
9-step	0.05	4	207.7555537

(LM) learning rule. For the structure of BP neural network model, two parameters are tested so that the optimal combination of the parameters which generates the lowest errors for the training dataset is selected. As suggested by literature [27], the parameters are: (1) the number of input, which varies from 1 to 8 with an increment of one; and (2) learning rate, which varies from 0.025 to 0.25 with an increment of 0.025. As such, 80 combinations are tested for each forecasting horizon.

Tables 4 and 5 show the MAE results for model selection for 1-step ahead forecasting for wind speed and wind power generation, respectively. Clearly, for the wind speed data, the models with 3 inputs and 0.2–0.25 learning rate show the best performance. As for the wind power generation data, the neural network structures with 5 input neurons and a learning rate of 0.1 to 0.25 have the best forecasting performance. In both cases, the highest learning rate should be selected when multiple rates generate the same level of error in that a larger learning rate takes less time to meet the stopping criterion. In this way, the best structures for other different forecasting horizons can also be determined.

Tables 6 and 7 compile the best combinations of ANN parameters and the corresponding MAE values for wind speed and power generation, respectively. Overall, the best parameters combinations are different for different forecasting horizons, and the corresponding MAE values increases with the increase of forecasting horizon.

4.3. Single SVM models

We adopt LS-SVMlab, a Matlab toolbox developed by Katholieke Universiteit Leuven [28] for LS-SVM model building and prediction. Recall that the kernel parameter and regularization parameter are denoted as σ^2 and γ. Similar to the ANN model building process, the best combination of parameters for SVM should be obtained for each forecasting horizon. Based on the suggestions from literature [7], the σ^2 values investigated are 0.25, 1, 4, 16, and 32; the γ values investigated are 0.25, 1, 4, 16, 64, and 256; the number of input varies from 1 to 8 with an increment of one. Note that if any best combination uses the upper limit of σ^2 or γ parameter, a higher level value of that parameter is evaluated and actually the specified parameter ranges are found to fit the datasets well.

Tables 8 and 9 show the MAE values of the SVM models with different combinations of σ^2 and γ values for wind speed and power generation, respectively. Note that both tables are used for illustration purpose, and obtained when the number of input is fixed at 3. For instance, in Table 8, for 1-step ahead wind speed prediction, with 3 inputs the best SVM model has the following parameters: $\sigma^2 = 16$, $\gamma = 256$. Similarly, in Table 9, for 1-step ahead wind power prediction, with 3 inputs the best SVM model has $\sigma^2 = 4$, $\gamma = 16$. The optimal parameter combinations for other forecasting horizons are also obtained. For the purpose of brevity, for those time horizons we only compile the optimal combinations of SVM parameters and the corresponding MAE values in Tables 10 and 11 for wind speed and power generation, respectively.

4.4. ARIMA–ANN hybrid models

In the ARIMA–ANN hybrid approach, ANN models are constructed to obtain the forecasts of the residuals from ARIMA fitting. For wind speed, the ANN models are obtained for the residuals from ARMA(1,3) model. For wind power generation, the ANN models are obtained for the residuals from ARMA(1,1) model. Similar to the procedure of single ANN modeling approach described in Section 4.2, the ANN models used for residual forecasting are also tested with the number of input varying from 1 to 8 with the increment of 1, and the learning rate varying from 0.025 to 0.25 with the increment of 0.025. Therefore, 80 parameter combinations are defined for capturing the nonlinear pattern in the residuals. The best combination for forecasting residuals is thus obtained based on the MAE values for each forecasting horizon. Again, for the purpose of brevity, we only compile the best combinations and the corresponding MAE values of each forecasting horizon for wind speed and power generation in Tables 12 and 13, respectively.

Table 8
SVM model selection for wind speed time series (1-step ahead, 3 inputs).

σ^2	γ					
	0.25	1	4	16	64	256
0.25	1.463285	1.401784	1.400653	1.419658	1.459656	1.530391
1	1.366097	1.316889	1.306564	1.324735	1.368224	1.43
4	1.323191	1.29505	1.281527	1.273913	1.277639	1.289492
16	1.315371	1.292315	1.283245	1.27861	1.279212	1.275429

Table 9
SVM model selection for wind power time series (1-step ahead, 3 inputs).

σ^2	γ					
	0.25	1	4	16	64	256
0.25	61.10103	58.73013	58.46111	59.49182	61.5844	64.62953
1	58.28908	56.8384	56.18503	55.99493	56.07275	56.64293
4	57.96594	56.57461	55.85583	55.59563	55.64707	55.69221
16	59.14201	57.00141	57.52127	56.70749	56.32082	55.96689

Table 10
Best SVM model for wind speed time series.

Forecasting horizon	γ	σ^2	Input number	MAE
1-step	64	4	3	1.275429
3-step	16	16	4	2.08278
5-step	16	16	5	2.378956
7-step	16	16	5	2.567432
9-step	4	16	6	2.702614

Table 11
Best SVM model for wind power time series.

Forecasting horizon	γ	σ^2	Input number	MAE
1-step	16	4	3	55.59563
3-step	4	4	8	114.4839
5-step	4	4	6	153.6343
7-step	256	16	5	183.4561
9-step	64	16	6	203.9593

Table 12
Best ANN model for the residuals after ARMA fitting (wind speed).

Forecasting horizon	Learning rate	Input number	Overall MAE
1-step	0.125	1	1.41306492
3-step	0.025	6	2.104259
5-step	0.25	4	2.47704711
7-step	0.075	4	2.71492115
9-step	0.1	4	2.86505751

Table 13
Best ANN model for the residuals after ARMA fitting (wind power).

Forecasting horizon	Learning rate	Input number	Overall MAE
1-step	0.025	5	56.08443005
3-step	0.025	3	97.56907472
5-step	0.175	5	129.2014083
7-step	0.025	4	145.2988873
9-step	0.025	4	154.3928459

Similarly, it can be seen that the parameters combinations may not be the same for different forecasting horizons, and the corresponding MAE values for residuals increases with the increase of forecasting horizon.

4.5. *ARIMA–SVM hybrid models*

Similarly, in ARIMA–SVM approach the SVM models are constructed to forecast the residuals of the established ARIMA models. SVM models are constructed for the same residuals from ARMA(1,3) and ARMA(1,1) models for wind speed and wind power generation, respectively. The parameters and the ranges of the parameter values are selected in consistence with those in the single SVM modeling approach mentioned in Section 4.3. We provide the best combinations of SVM parameters and the corresponding MAE values for wind speed and power generation in Tables 14 and 15, respectively. Not surprisingly, it can be seen that different SVM parameters combinations are obtained for different forecasting horizons, and the optimal MAE values for residuals increases with the increase of forecasting horizon.

4.6. *Forecasting comparison*

Based on the single and hybrid models identified, the forecasts are made for hourly wind speed and power generation using multiple time horizons for a period of 31 days (i.e., 744 h). MAE and RMSE values are computed for the forecasts. The results are compiled and shown in Tables 16 and 17 for wind speed and power generation, respectively. The values in bold in the two tables indicate the smallest MAE or RMSE results among the five forecasting approaches for each forecasting time horizon. For wind speed forecasting, it can be seen from Table 16 that for all models, both MAE and RMSE values increase with the increase of forecasting time horizon. The most significant drop in performance occurs when the forecasting horizon is changed from 1-step to 3-step ahead. Further reduction in forecasting accuracy is observed as the time horizon increases from 3-step ahead to 9-step ahead, but the trends become flatter. This is in general consistent with the findings regarding the effect of time horizon on wind forecasting from literature [29,30]. It can also be observed that depending on forecasting horizon, hybrid methods or

Table 14
Best SVM model for the residuals after ARMA fitting (wind speed).

Forecasting horizon	γ	σ^2	Input number	Overall MAE
1-step	1	16	2	1.412243
3-step	1	16	5	2.106537
5-step	1	16	7	2.459093
7-step	1	16	5	2.705491
9-step	4	16	4	2.884355

Table 15
Best SVM model for the residuals after ARMA fitting (wind power).

Forecasting horizon	γ	σ^2	Input number	Overall MAE
1-step	1	16	4	56.33127
3-step	256	1	1	100.784
5-step	1	4	7	130.979
7-step	1	16	8	147.3478
9-step	1	16	8	156.1655

Table 16
MAE and RMSE for wind speed forecasting using various forecasting methods.

	ANN	ARIMA	ARIMA–ANN	SVM	ARIMA–SVM
MAE					
1-step	1.26274841	1.283049463	**1.24940572**	1.27542917	1.254313196
3-step	2.097660089	2.08811674	2.1105777	2.08278002	**2.053321793**
5-step	2.378312314	2.388163175	**2.35639349**	2.378956051	2.393676801
7-step	2.524579238	2.563882354	**2.52106251**	2.574971061	2.571131155
9-step	2.719195818	**2.70242486**	2.74326720	2.702614179	2.756355069
RMSE					
1-step	1.924871614	1.92609482	1.91340013	1.952002511	**1.905118156**
3-step	3.065580832	3.00966632	3.017293507	3.045012588	**3.009313377**
5-step	3.395435825	**3.34123076**	3.368867747	3.405734591	3.421753091
7-step	3.584713921	3.575892257	**3.55436902**	3.631997817	3.625486074
9-step	3.7880921	**3.76469715**	3.83575006	3.778017349	3.81693498

Table 17
MAE and RMSE for wind power generation forecasting using various forecasting methods.

	ANN	ARIMA	ARIMA–ANN	SVM	ARIMA–SVM
MAE					
1-step	55.89285343	58.26094625	**55.354269**	55.64706652	56.24842931
3-step	**113.933008**	118.6103271	116.5489811	114.4839392	117.3379861
5-step	155.0666085	160.0031089	159.1642621	**153.634273**	157.476753
7-step	**182.844697**	188.728031	183.9360242	183.4561242	186.0096651
9-step	207.7555537	210.2913696	210.6795646	**204.781605**	209.9555507
RMSE					
1-step	87.26140472	88.73852468	**86.7843876**	86.79886639	87.05463765
3-step	**154.578053**	155.7321938	155.8676028	157.7847937	155.6413801
5-step	**196.904888**	198.0916395	198.9799005	198.572467	198.1866068
7-step	224.2186185	225.5443257	**223.987076**	226.2824458	225.2628132
9-step	244.2465359	246.4205227	249.3001607	**243.434801**	248.5471056

ARIMA method outperform the other two single model approaches, namely, SVM and ANN methods. Nevertheless, the improvement is limited, ranging from 0.1% to 5.5%. On the other hand, the hybrid methods cannot outperform the ARIMA model for 9-step ahead forecasting according to the MAE values. This is also true for both 5-step and 9-step ahead predictions according to the RMSE values. In addition, a delicate observation can be made. When the single ARIMA model outperforms the single ANN and SVM models, the introduction of an ANN or SVM model for the non-linear part seems to deteriorate the overall forecasting performance. In this case, although the hybrid approaches consider both linear and nonlinear components, they do not necessarily generate superior forecasting performance.

According to Table 17, the forecasting accuracy also decreases with the increase of forecasting horizon for hourly wind power generation, but the pace of accuracy reduction slows down. For example, in hourly wind power data, the MAE value for 1-step ahead forecasting based on ANN method is 55.89, while those values for 3-step, 5-step, 7-step and 9-step forecasting are 113.93, 155.07, 182.84 and 207.76, respectively. Similar trends can be observed based on RMSE results. Meanwhile, the performance difference among the five forecasting methods appears to be mixed. For hourly wind power generation, the ARIMA models are not able to produce the lowest MAE or RMSE values for any of the forecasting horizons. Based on MAE values, the hybrid methodology outperforms single models only for 1-step ahead forecasting, while ANN method has the best performance for 3-step and 7-step ahead forecasting, and SVM method claims the crown for 5-step and 9-step ahead forecasting. Again, the hybrid models, which consider both linear and nonlinear performance, cannot universally outperform the single forecasting models. Furthermore, they seem to be less effective for wind power generation if compared with their relative performance in wind speed forecasting. On the other hand, for each forecasting time horizon, the performance difference is actually small – there is only around 3% discrepancy between the worst and best performing models.

Based on the above results, it can be found that the hybrid methods cannot always provide improvements compared with the single foresting models. For wind speed forecasting, the ARIMA models can outperform the ARIMA–ANN and ARIMA–SVM hybrid models. For wind power forecasting, the single ANN or SVM models can outperform the hybrid approaches. The main reason is believed to be that the hybrid methodology assumes the additive relationship between the linear and nonlinear parts for a time series. However, the components may have different types of relationships (e.g., multiplication). In such circumstances, the performance of the hybrid model may be degraded. In addition, the nonlinear assumption for the residuals may not be always guaranteed [19]. On the other hand, the hybrid approaches provide a new viable option for wind forecasting applications in that they do outperform the single forecasting models in many occasions. As a result, they should be included for consideration in searching for the best forecasting model.

5. Conclusive remarks

Hybrid forecasting methodology is a recently proposed time series forecasting approach which typically consists of an ARIMA model and a nonlinear model for the linear and nonlinear components, respectively. To comprehensively evaluate this methodology for wind forecasting, in this paper we compare it with the three major single prediction models, i.e., ARIMA, ANN, and SVM, in the prediction of hourly wind speed and power generation. Five forecasting horizons are used, and two performance metrics are computed. In the hybrid approaches, the linear component of time series is modeled by the ARIMA method, while the ANN or SVM models are adopted for fitting the nonlinear component. This generates two hybrid approaches, namely, ARIMA–ANN and ARIMA–SVM. For wind speed, it appears that the hybrid methods own the best performance when ARIMA method cannot outperform ANN and SVM methods, but they cannot further improve the performance of ARIMA model if it is already better than

ANN and SVM. Overall, the performance difference between the five single or hybrid forecasting models is less than 5.5% across all forecasting horizons investigated. For wind power generation, ARIMA models are not able to outperform either ANN or SVM models. Further improvement to the ANN or SVM models is possible by adopting the hybrid approaches, but it is limited to less than 3%. The results show that the hybrid methodology does not always outperform the individual forecasting models based on ARIMA, ANN, or SVM. As such, the argument in some literature that the hybrid methodology is always superior to single models cannot hold for wind speed or power generation forecasting. In brief, the hybrid forecasting methodology does add a viable option to the toolbox of short-term forecasting models for wind speed and power generation, but it does not always generate better performance.

References

[1] Lerner J, Grundmeyer M, Garvert M. The importance of wind forecasting. Renewable Energy Focus 2009;10(2):64–70.
[2] Riahy G, Abedi M. Short term wind speed forecasting for wind turbine applications using linear prediction method. Renewable Energy 2008;33(1): 35–40.
[3] Torres J, Garcia A, Deblas M, Defrancisco A. Forecast of hourly average wind speed with ARMA models in Navarre (Spain). Solar Energy 2005;79(1): 65–70.
[4] Liu H, Shi J, Erdem E. Prediction of wind speed time series using modified Taylor Kriging method. Energy 2010;35(12):4870–9.
[5] Kariniotakis G, Stavrakakis GS, Nogaret EF. Wind power forecasting using advanced neural network models. IEEE Transactions on Energy Conversion 1996;11(4):762–7.
[6] Alexiadis M. Short-term forecasting of wind speed and related electrical power. Solar Energy 1998;63(1):61–70.
[7] Zhou J, Shi J, Li G. Fine tuning support vector machines for short-term wind speed forecasting. Energy Conversion and Management 2011;52(4):1990–8.
[8] Kariniotakis G, Stavrakakis G, Nogaret E. A fuzzy logic and neural network based wind power forecasting model. In: Proceedings of the 1996 European union wind energy conference. 1996. p. 596–9.
[9] Sanchez I. Short-term prediction of wind energy production. International Journal of Forecasting 2003;22(1):43–50.
[10] Li G, Shi J. Bayesian adaptive combination of short-term wind speed forecasts from neural network models. Renewable Energy 2011;36(1):352–9.
[11] Nielsen H, Nielsen T, Madsen H, San Isidro Pindado M, Marti I. Optimal combination of wind power forecasts. Wind Energy 2007;10(5):471–82.
[12] Thordarson F, Madsen H, Nielsen H, Pinson P. Conditional weighted combination of wind power forecasts. Wind Energy 2010;13:751–76.
[13] Tseng FM, Yu HC, Tzeng GH. Combining neural network model with seasonal time series ARIMA model. Technological Forecasting and Social Change 2002;69(1):71–80.
[14] Zhang GP. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 2003;50:159–75.
[15] Hansen J, Nelson R. Time-series analysis with neural networks and ARIMA-neural network hybrids. Journal of Experimental & Theoretical Artificial Intelligence 2003;15(3):315–30.
[16] Pai P, Lin C. A hybrid ARIMA and support vector machines model in stock price forecasting. Omega 2005;33(6):197–505.
[17] Gutierrez-Estrada J, Silva C, Yanez E, Rodriguez N, Pulido-Calvo I. Monthly catch forecasting of anchovy Engraulis ringens in the north area of Chile: non-linear univariate approach. Fisheries Research 2007;86(2–3):188–200.
[18] Cadenas E, Rivera W. Wind speed forecasting in three different regions of Mexico using a hybrid ARIMA–ANN model. Renewable Energy 2010;35(12):2732–8.
[19] Taskaya-Temizel T, Casey M. A comparative study of autoregressive neural network hybrids. Neural Networks 2005;18(5–6):781–9.
[20] Box P, Jenkins G. Time series analysis: forecasting and control. San Francisco, CA: Holden-day Inc.; 1976.
[21] Khashei M, Bijari M. An artificial neural network (p, d, q) model for time series forecasting. Expert Systems with Applications 2010;37(1):479–89.
[22] Zhang GP, Qi M. Neural network forecasting for seasonal and trend time series. European Journal of Operational Research 2005;160(2):501–14.
[23] Chen K, Wang C. A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan. Expert Systems with Applications 2007;32(1):254–64.
[24] Shi J, Qu X, Zheng S. Short-term wind power generation forecasting: direct versus indirect ARIMA-based approaches. International Journal of Green Energy 2011;8(1):100–12.
[25] Kamal L, Jafri YZ. Time series model to simulate and forecast hourly averaged wind speed in Quetta, Pakistan. Solar Energy 1997;61(1):23–30.
[26] Brockwell PJ, Davis RA. Introduction to time series and forecasting. New York: Springer-Verlag; 2002.
[27] Li G, Shi J. On comparing three artificial neural networks for wind speed forecasting. Applied Energy 2010;87(7):2313–20.
[28] Suykens JAK, Gestel TV, Brabanter JD, Moor BD, Vandewalle J. Least squares support vector machines. 1st ed. Singapore: World Scientific; 2002.
[29] Sánchez I. Short-term prediction of wind energy production. International Journal of Forecasting 2006;22(1):43–50.
[30] Kariniotakis G, Pinson P, Siebert N, Giebel G, Barthelmi R. The state of the art in short-term prediction of wind power – from an offshore perspective. In: Proc. of 2004 SeaTechWeek. 2004. p. 1–13.