Energy Conversion and Management 52 (2011) 1990–1998

Contents lists available at ScienceDirect

Energy Conversion and Management

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

Fine tuning support vector machines for short-term wind speed forecasting

Junyi Zhou, Jing Shi *, Gong Li

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

ARTICLE INFO

Article history:
Received 8 March 2010
Accepted 21 November 2010
Available online 9 January 2011

Keywords:
Wind speed
Short-term forecasting
Least-squares support vector machines (LS-SVM)
Parameter tuning
Persistence method

ABSTRACT

Accurate forecasting of wind speed is critical to the effective harvesting of wind energy and the integration of wind power into the existing electric power grid. Least-squares support vector machines (LS-SVM), a powerful technique that is widely applied in a variety of classification and function estimation problems, carries great potential for the application of short-term wind speed forecasting. In this case, tuning the model parameters for optimal forecasting accuracy is a fundamental issue. This paper, for the first time, presents a systematic study on fine tuning of LS-SVM model parameters for one-step ahead wind speed forecasting. Three SVM kernels, namely linear, Gaussian, and polynomial kernels, are implemented. The SVM parameters considered include the training sample size, SVM order, regularization parameter, and kernel parameters. The results show that (1) the performance of LS-SVM is closely related to the dynamic characteristics of wind speed; (2) all parameters investigated greatly affect the performance of LS-SVM models; (3) under the optimal combination of parameters after fine tuning, the three kernels give comparable forecasting accuracy; (4) the performance of linear kernel is worse than the other two kernels when the training sample size or SVM order is small. In addition, LS-SVMs are compared against the persistence approach, and it is found that they can outperform the persistence model in the majority of cases.

1. Introduction

The installed wind power capacity has almost increased by 200% between 2005 and 2009, and the majority of newly added capacity resides in several nations such as the United States, Germany, and China. In particular, the US Department of Energy [1] sets a goal of 20% wind energy contribution to the US electricity supply by 2030 in a landmark report, and this can reduce the CO_2 emission due to electricity generation by 25%. On the other hand, the wind generated electricity has been plagued by the intermittent and stochastic nature of wind source, and thus it is still a less reliable source and difficult to be integrated into power grid systems. Accurate forecasting of wind speed is recognized as a critical issue for the proper and efficient utilization of wind power. For example, a 10% deviation of the expected wind speed approximately leads to a 30% deviation in the expected wind power generation because the power potential is proportional to the cubic power of the wind speed [2]. Along with more wind generation capacities being installed, good wind speed forecasting techniques become more urgent in order to improve both the efficiency of a wind power generation system [3] and the integration of wind energy with the power system [4].

Short-term prediction of wind speed can be made in the order of several days and also from minutes to hours [5]. Usually, hourly forecasts of expected winds are helpful in dispatching decision-making, daily forecasts of hourly winds are useful for the load scheduling strategy, and weekly forecasts of day-to-day winds greatly facilitate maintenance scheduling [6]. Therefore, countless efforts have been made to develop good short-term wind speed forecasting methods. Detailed reviews on these efforts can be found in [4,5]. Generally, the methods and techniques for short-term wind speed forecasting can be categorized as physical methods, conventional statistical methods, hybrid physical–statistical models, artificial intelligence and other new methods [4]. Physical methods [7] usually perform forecasting by combining multiple physical considerations. Statistical methods [8,9] make forecasting based on the observed wind speed time series. In practice, the two types of models can be utilized together where the results from physical models are factored as input variables, together with the historical data of wind speed, to train the system according to statistical theories. Besides, some new methods based on artificial intelligence techniques have also been developed and investigated such as artificial neural networks [10–13] and fuzzy logic [14,15].

In recent years, support vector machines (SVM), a novel and powerful machine learning tool, has been successfully used for time-series prediction with satisfactory prediction results in various fields [16–21]. SVM is an elegant and highly principled learning method for a feed forward network design with a single

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

0196-8904/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.enconman.2010.11.007

hidden layer of units that are usually nonlinear [22]. It is characterized by the use of a technique known as "kernel trick" to apply linear classification techniques to nonlinear classification problems. SVM can well handle high dimensional data even with a relatively small training samples and it has excellent generalization ability for complex models [23]. The SVM trained as a convex optimization problem can generate a global and usually unique solution. Capable of modeling nonlinear relations in an efficient and stable way, SVM has demonstrated good performance on regression and time-series prediction although the training time might be long for large datasets [24]. The preliminary application of SVM method for short-term wind forecasting has been reported in a few papers. Mohandes [25] applied this method to wind speed prediction and compared its performance with the multilayer perceptron (MLP) neural networks. The results indicate that SVM approach outperforms the MLP model in terms of root mean squared error (RMSE). Similarly, Sreelakshmi and Kumar [22] also indicated that SVM models could give better accuracies than the neural network models in short-term wind forecasting. Kusiak et al. [26] investigated both wind speed and power forecasting at forecast intervals of 10-min up to 1-h. It was observed that the SVM method is a better choice compared with the MLP algorithm.

The performance of SVM is affected by the SVM configuration, which includes the type of kernel function, the size of training sample, and the settings of kernel parameters. In this regard, the existing studies on using SVM for wind forecasting are very limited in that usually only one particular kernel function and a specific combination of parameters are picked and used in these studies. A systematic investigation is called for on how the SVM configuration affects its performance in short-term wind speed forecasting. The investigation can provide answers to some fundamental questions such as "which kernel function should be chosen?", and thus gives more confidence for people to apply this approach for wind energy applications. Aiming to fill the research gap, this paper studies the accuracy of one-step (1 h) ahead wind speed forecasting by least-squares support vector machines (LS-SVM), in which three types of kernel function, namely linear, polynomial, and Gaussian kernels, are compared, and moreover, the effects of training sample size, SVM order, and kernel parameters, on the forecasting accuracy are investigated.

The remainder of this paper is organized as follows. In Section 2, the principle of LS-SVM is briefly introduced. In Section 3, the analysis procedure for tuning LS-SVM parameters for optimal performance is described. In Section 4, the results are presented and analyzed. Finally, the conclusions are drawn in Section 5.

2. Principle of LS-SVM

In this study, LS-SVM is adopted due to its low computation complexity and high generalization performance [27]. SVM is a statistical learning approach that can be applied to solving problems in nonlinear classification and function estimation. Unlike the classical neural networks approach, SVM formulates the statistical learning problem as a quadratic programming with linear constraints, by the use of nonlinear kernels, high generalization ability, and sparseness of solution. However, for large-scale problems, the optimization process of SVM has high computational complexity, because of the high-dimensional matrix involved in the quadratic programming whose size is directly proportional to the training sample size. LS-SVM is a variant of the standard SVM, in which the model formulation is simplified into a linear problem without loss of its advantages. It is shown by a wide range of benchmark datasets that the generalization performance of LS-SVM is comparable to that of the standard SVM [28]. On the other hand, LS-SVM needs significantly less training effort than the standard SVM as a result of the model simplification. We briefly introduce the principle of LS-SVM in the following.

The basic principle of SVM regression is to estimate the output variable y from $\varphi(\mathbf{x})$, a high dimensional feature space of the input vector $\mathbf{x} = (x^1, x^2, \cdots, x^K)^T$, where K is the order of SVM. Thus, the general SVM regression model is

$$y = \mathbf{w}^T \varphi(\mathbf{x}) + b, \tag{1}$$

where w and b are the weight vector and bias term respectively.

Given a sample of training data $\{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$, LS-SVM determines the optimal weight vector and bias term by minimizing the following cost function, R,

$$\min_{\mathbf{w},\mathbf{e}} R(\mathbf{w}, \mathbf{e}) = \frac{1}{2}\|\mathbf{w}\|^2 + \frac{1}{2}\gamma\|\mathbf{e}\|^2, \tag{2}$$

subject to the equality constraints

$$y_i = \mathbf{w}^T \varphi(\mathbf{x}_i) + b + e_i, \quad i = 1, 2, \ldots, N \tag{3}$$

where $\mathbf{e} = (e_1, e_2, \cdots, e_N)^T$. The first part of cost function regularizes weight sizes and penalizes large weights. Therefore, the weights tend to converge to similar values in that large weights cause excessive variance and hence deteriorate the generalization ability of LS-SVM. The second part of Eq. (2) considers the regression error of all training data. The regularization parameter γ controls the trade-off between the bias and variance of LS-SVM model. Note that the LS-SVM model has equality constraints as shown in Eq. (3), rather than the inequality constraints with slack variables used in the standard SVM model. Moreover, as shown in Eq. (2), a squared loss function is considered in the objective function of LS-SVM model, while the standard SVM model has a linear combination of slack variables in its objective function. These two modifications simplify the quadratic optimization problem for the standard SVM to be linear for LS-SVM.

The Lagrangian of Eq. (2) is

$$L(\mathbf{w}, b, \mathbf{e}, \lambda) = R(\mathbf{w}, \mathbf{e}) - \sum_{i=1}^{N} \lambda_i (\mathbf{w}^T \varphi(\mathbf{x}_i) + b + e_i - y_i), \tag{4}$$

where λ_i are the Lagrange multipliers. By the Karush–Kuhn–Tucker Theorem, the conditions of optimality are

$$\begin{aligned}
\frac{\partial L}{\partial \mathbf{w}} &= 0 \rightarrow \mathbf{w} = \sum_{i=1}^{N} \lambda_i \varphi(\mathbf{x}_i) \\
\frac{\partial L}{\partial b} &= 0 \rightarrow \sum_{i=1}^{N} \lambda_i = 0 \\
\frac{\partial L}{\partial e_i} &= 0 \rightarrow \lambda_i = \gamma e_i \\
\frac{\partial L}{\partial \lambda_i} &= 0 \rightarrow \mathbf{w}^T \varphi(\mathbf{x}_i) + b + e_i - y_i = 0
\end{aligned}$$

e_i = (1/gamma)*lambda)_i

Thus, b and λ can be solved from the following set of linear equations after eliminating w and e,

$$\begin{bmatrix} 0 & \vec{1}^T \\ \vec{1} & \mathbf{K} + \gamma^{-1}\mathbf{I} \end{bmatrix} \begin{bmatrix} b \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix}, \tag{5}$$

where $\vec{1} = (1, 1, \cdots, 1)^T$, $\boldsymbol{I}$ is the identity matrix, $\mathbf{y} = (y_1, y_2, \cdots, y_N)^T$, $\mathbf{K} = (k(\mathbf{x}_i, \mathbf{x}_j))_{i,j=1}^{N}$ is the kernel matrix, and $k(\mathbf{x}_i, \mathbf{x}_j) = \phi^T(\mathbf{x}_i)\phi(\mathbf{x}_j)$ is the kernel function. As a result, given vectors x and x_i, the LS-SVM regression model for estimating y in Eq. (1) becomes

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

As mentioned previously, in this study we consider three widely-used kernel functions, namely linear, polynomial, and Gaussian kernels. They are defined as follows.

(1) Linear kernel

$$k_L(\mathbf{x},\mathbf{z}) = \mathbf{x}^{\mathrm{T}}\mathbf{z}. \tag{7}$$

(2) Polynomial kernel

$$k_P(\mathbf{x},\mathbf{z}) = (\mathbf{x}^{\mathrm{T}}\mathbf{z} + c)^d, \tag{8}$$

where c and d are the bias and degree of polynomial kernel respectively.

(3) Gaussian kernel

$$k_G(\mathbf{x},\mathbf{z}) = \exp(-\|\mathbf{x}-\mathbf{z}\|^2/\sigma^2), \tag{9}$$

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

3. Analysis procedure

We conduct the SVM tuning computation based on the hourly wind speed data of year 2002 from one wind observation site in North Dakota, USA. To reduce the impact of seasonal speed pattern on SVM forecasting, we divide the wind speed data of the entire year into four seasonal datasets, each of which consists of the data of spring, summer, fall, and winter, respectively. The tuning computation is applied to each of four datasets separately. Table 1 shows the descriptive statistics of the four seasonal datasets. It is clear that the average wind speeds in spring and winter are higher compared with summer and fall. Meanwhile, spring and winter have significantly higher maximum wind speeds, and the wind speed is more volatile in the two seasons as well.

As shown in Section 2, there are multiple parameters of LS-SVM to be tuned for obtaining optimal performance in forecasting the wind speed: (1) N, the training sample size; (2) K, the order of SVM; (3) γ, the regularization parameter; and (4) kernel parameters (i.e. c and d for polynomial kernel; and σ for Gaussian kernel). In principle, the ranges of the SVM parameters are chosen to be wide so that the common values and ranges reported in literature are well covered [16,19,25,29,30]. The analysis procedure applied in this study is illustrated in Fig. 1, where the analysis steps are described below.

3.1. Step 1: Data preprocessing

In this step, the observed wind speed series in each season are formatted into windowized wind speed vectors, $\mathbf{x}_w = (x^1, x^2, \cdots, x^K, x^{K+1})^T$, where $(x^1, x^2, \cdots, x^K)^T$ and x^{K+1} correspond to the input vector x and forecast variable y in Eq. (1) respectively. We test parameter K, the order of SVM from one to ten with an increment of one. For each value of K, the preprocessing step is run to windowize the speed series. Then the speed vectors are divided into two datasets: one is for SVM training and the other is for SVM forecasting. For each season, the testing sample contains the last 120 windowized vectors; and the training sample contains the required number of windowized vectors close to those of the testing sample. The training sample size, N, considered in the experiment is selected from the set {24, 48, 120, 240, 480, 720, 960, 1200, 1440}, which correspondingly indicates that 1, 2, 5, 10, 20, 30, 40, 50, or 60 days of wind speed data are used for training the LS-SVM model.

Table 1
Descriptive statistics of seasonal wind speed data (m/s).

	Mean	St. dev.	Minimum	Median	Maximum
Spring	7.50	3.38	0.42	7.31	21.50
Summer	6.93	2.88	0.35	6.70	16.08
Fall	7.38	2.90	0.42	7.33	17.57
Winter	7.58	3.21	0.44	7.50	26.39

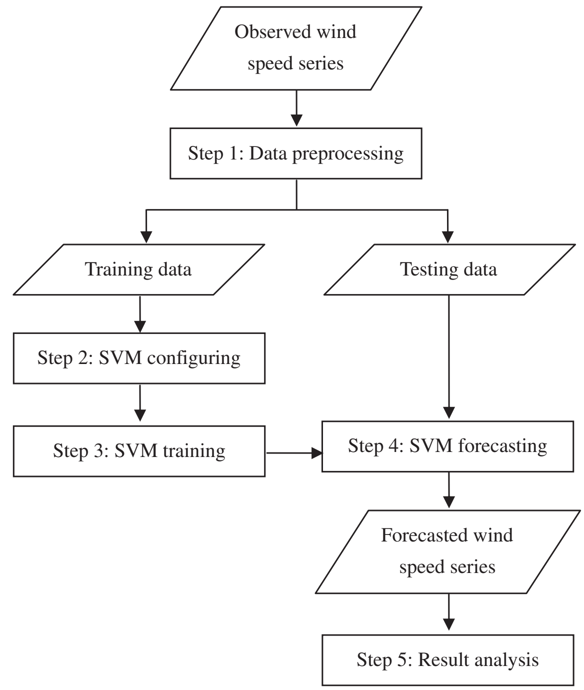

Fig. 1. Procedure of SVM tuning.

The larger the training sample size, the longer the observation history is considered in the SVM model.

3.2. Step 2: SVM configuring

In this step, the regularization parameter, γ, and kernel parameters are determined. Each parameter can be selected over a wide range such that the global effects of the parameters on forecasting performance can be revealed. We test $\gamma \in \{2^{-2}, 2^{-1}, \ldots, 2^{10}\}$. For linear kernel, there is no kernel parameter to be tested. For Gaussian kernel, We test $\sigma^2 \in \{2^{-2}, 2^{-1}, \ldots, 2^8\}$. For polynomial kernel, we test $c \in \{2^0, 2^1, \ldots, 2^{10}\}$, and $d \in \{1, 2, 3, 4\}$.

3.3. Step 3: SVM training

In this step, by solving the linear equation, Eq. (5), we obtain the values of λ_i and b, given the training sample size, SVM order, regularization parameter and kernel parameters.

3.4. Step 4: SVM forecasting

In this step, we feed λ_i and b obtained from last step into Eq. (6) on the testing sample to obtain the forecasted wind speed series.

3.5. Step 5: Result analysis

In this step, we calculate the accuracy of each forecasted wind speed series, and analyze the effect of SVM order, training sample size, regularization parameter, and kernel parameters on forecasting accuracy.

4. Results and discussion

We evaluate the forecasting accuracy by the root mean squared error (RMSE) of the forecast speed series,

$$\mathrm{RMSE} = \sqrt{(y_i - \hat{y}_i)^2/M} \tag{10}$$

Table 2
Optimal forecasting accuracy (RMSE) for three kernels on four seasonal datasets.

	Linear	Gaussian	Polynomial
Spring	1.421	1.42	1.417
Summer	1.343	1.33	1.323
Fall	0.972	0.961	0.966
Winter	0.956	0.919	0.906

Table 3
Descriptive statistics of wind speed change between two consecutive hours (m/s).

	Mean	St. dev.	Minimum	Median	Maximum	St. dev./mean
Spring	1.01	0.93	0	0.76	7.97	0.92
Summer	1.04	0.94	0	0.79	7.52	0.90
Fall	0.85	0.74	0	0.64	6.47	0.87
Winter	0.88	0.76	0	0.69	6.86	0.86

Table 4
Optimal configuration for three kernels on four seasonal datasets.

	Linear			Gaussian				Polynomial				
	N	K	γ	N	K	γ	σ^2	N	K	γ	c	d
Spring	1200	4	64	720	4	4	8	960	4	32	256	2
Summer	1200	4	64	1200	4	4	32	1200	4	32	256	2
Fall	720	5	32	720	5	4	32	960	5	16	128	2
Winter	720	5	32	1200	5	16	4	1200	5	16	128	2

where y_i and $\hat{y}_i$ are the ith observed and forecasted speed respectively; and M is the testing sample size, which is 120 in this study. RMSE is a widely-applied statistic used for time-series forecasting. In general, a smaller RMSE value indicates a smaller discrepancy between the observed and forecasted wind speed series, hence a higher forecasting accuracy.

4.1. Forecasting accuracy

Table 2 lists the highest achievable forecasting accuracy for three kernel functions in all four seasons. Note that the combination of parameters to obtain these best results may be different for each function and in each season. It shows that the forecasting accuracies of Gaussian and polynomial kernels are higher than that of linear kernel, but the difference is actually small. Moreover, all kernels perform best in winter, followed by fall, then summer, and poorest in spring. To explain this result, we resort to the characteristics of the original wind data. The examination on four seasonal datasets reveals that the difference of forecasting accuracy can be, to some extent, reflected by the temporal variation (ETV) of wind speed. The ETV values are obtained by the following procedure. First, for the wind speed time series $v_1, v_2, \ldots, v_n$, we calculate speed change at time point i, Δv_i, as the absolute difference of two consecutive hourly wind speeds, v_{i+1} and v_i, i.e., $\Delta v_i = |v_{i+1} - v_i|$, $i = 1, \ldots, n-1$. Then, the mean and the standard deviation of Δv_i are calculated. The ETV value is finally obtained as the ratio of standard deviation to mean of Δv_i. The descriptive statistics of Δv_i are summarized in Table 3. It can be seen that the mean, standard deviation (St. dev.), median, and maximum of the wind speed change for spring and summer seasons are generally higher than those for fall and winter seasons. Note that the testing dataset (5 days) for each season is also included in the computation. This indicates that the wind speed is more volatile in spring and summer, and thus more difficult to forecast. Moreover, the table shows that the ETV values in spring, summer, fall, and winter are 0.92, 0.90, 0.87, and 0.86 respectively, and this exactly matches the rank of RMSE results in Table 2. In other words, the higher the ETV, the

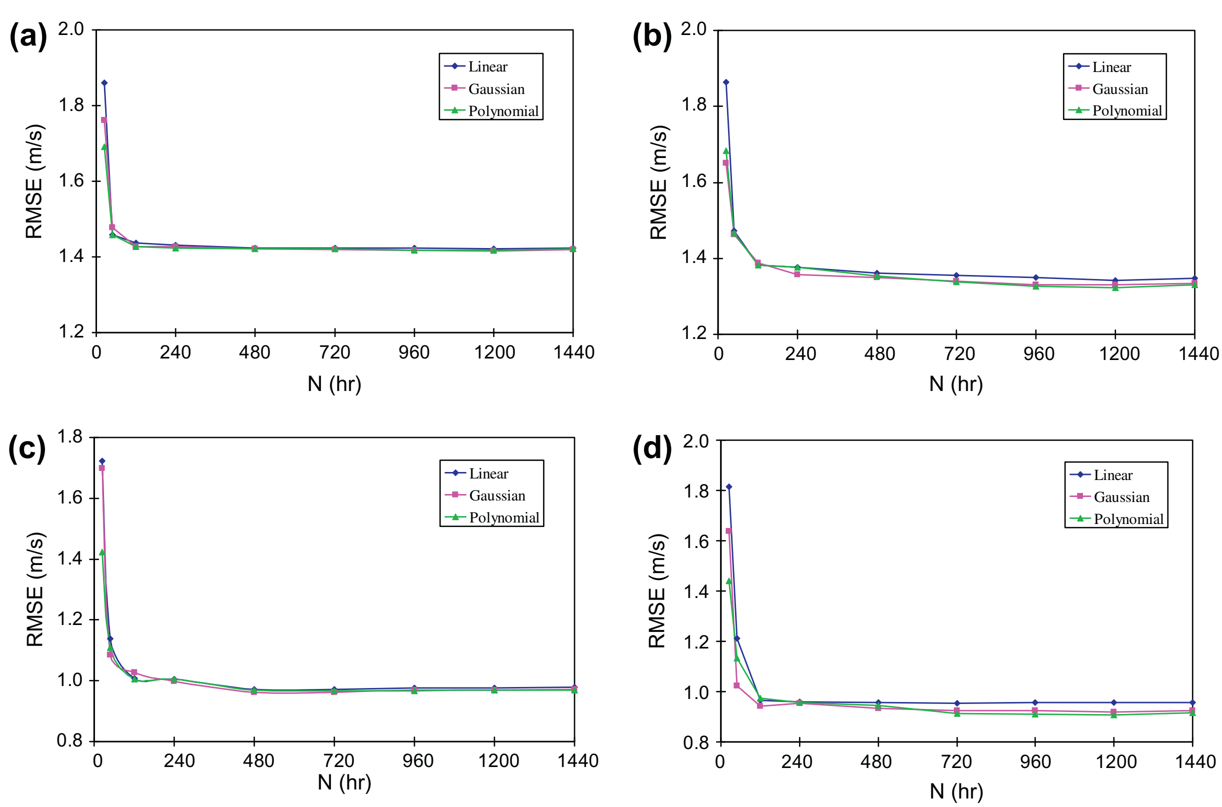

Fig. 2. Effect of training sample size on RMSE for: (a) spring, (b) summer, (c) fall, and (d) winter.

lower the forecasting accuracy. The implication is that it is not the absolute value of wind speed but the characteristics of wind speed dynamic change that affects the performance of short-term wind speed forecasting.

The corresponding settings of training sample size, SVM order, regularization parameter, and kernel parameter, under which the optimal results in Table 2 are obtained, are provided in Table 4. There are several interesting observations. First, the optimal training sample size is between 720 and 1200. Second, the optimal SVM order is either 4 or 5. Third, the optimal regularization is from 4 to 64, and the Gaussian kernel has smaller regularization parameters. Recall that the regularization parameter controls the bias-variance trade-off in the LS-SVM cost function (see Eq. (2)). Therefore, this means that Gaussian kernel matches the training dataset better but might have larger forecast variances. Finally, the optimal degree of polynomial kernel is 2 for all the seasons.

4.2. Training sample size

Fig. 2 shows how the forecasting accuracy varies with the training sample size for the seasonal wind speed data in spring, summer, fall, and winter. Note that each point in Fig. 2 corresponds to the minimum RMSE obtained among all combinations of SVM order, regularization parameter, and kernel parameters, at the specific training sample size. It can be observed from that for all the seasons, the forecasting accuracy improves significantly while the training sample size increases from 24 to 120, but stays almost unchanged while the training sample size is bigger than 480. For instance, for winter season the linear kernel generates an RMSE value of 1.815 with 24 training samples. The RMSE value reduces to 0.957 with 480 samples, and no further change is observed beyond that. More importantly, for each season more significant improvement in RMSE can be observed for linear kernel between the sample size range of 24–120, This implies that when the training sample size is small, linear kernel should be avoided.

As shown in Eq. (5), without special efficient training methods, such as that proposed by Suykens et al. [31], the basic training algorithm of LS-SVM has a computational complexity of $O(N^3)$. As a result, the efficiency of LS-SVM decreases significantly with the increase of training sample size. For this reason, although as shown in Table 1 that the optimal forecasting performance is obtained when the training sample size is between 720 and 1200, we can use a training sample size of 240–480, without losing much forecasting accuracy. Therefore, we analyze the effect of other factors in the following based on the results obtained using training samples with size of 480.

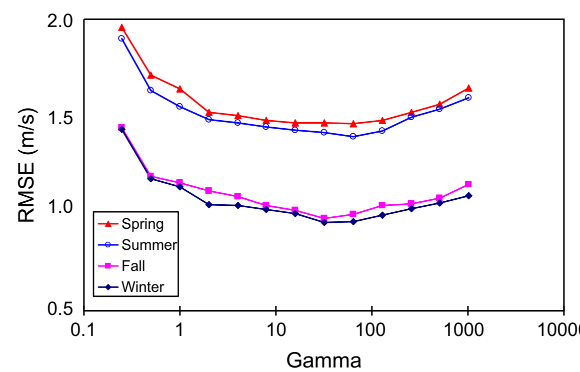

Fig. 4. Effect of regularization parameter γ (Gamma, in logarithm scale) on RMSE for the linear kernel.

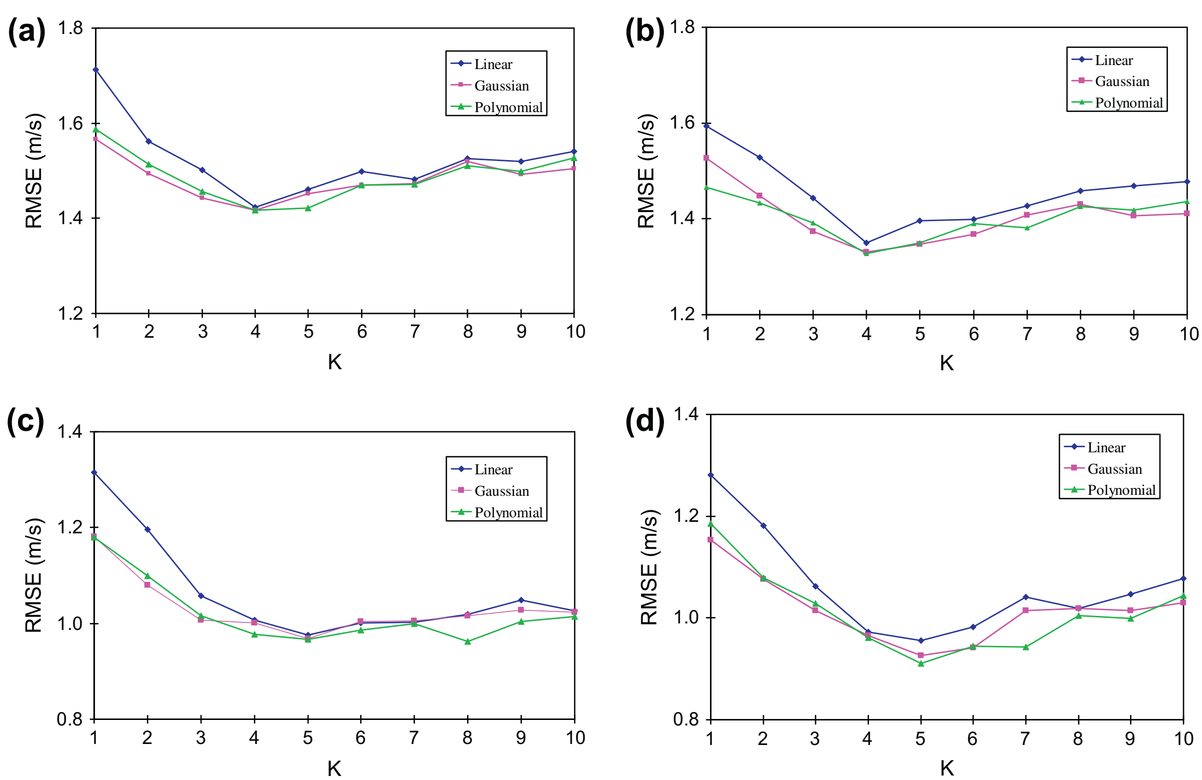

Fig. 3. Effect of SVM order on RMSE for: (a) spring, (b) summer, (c) fall, and (d) winter.

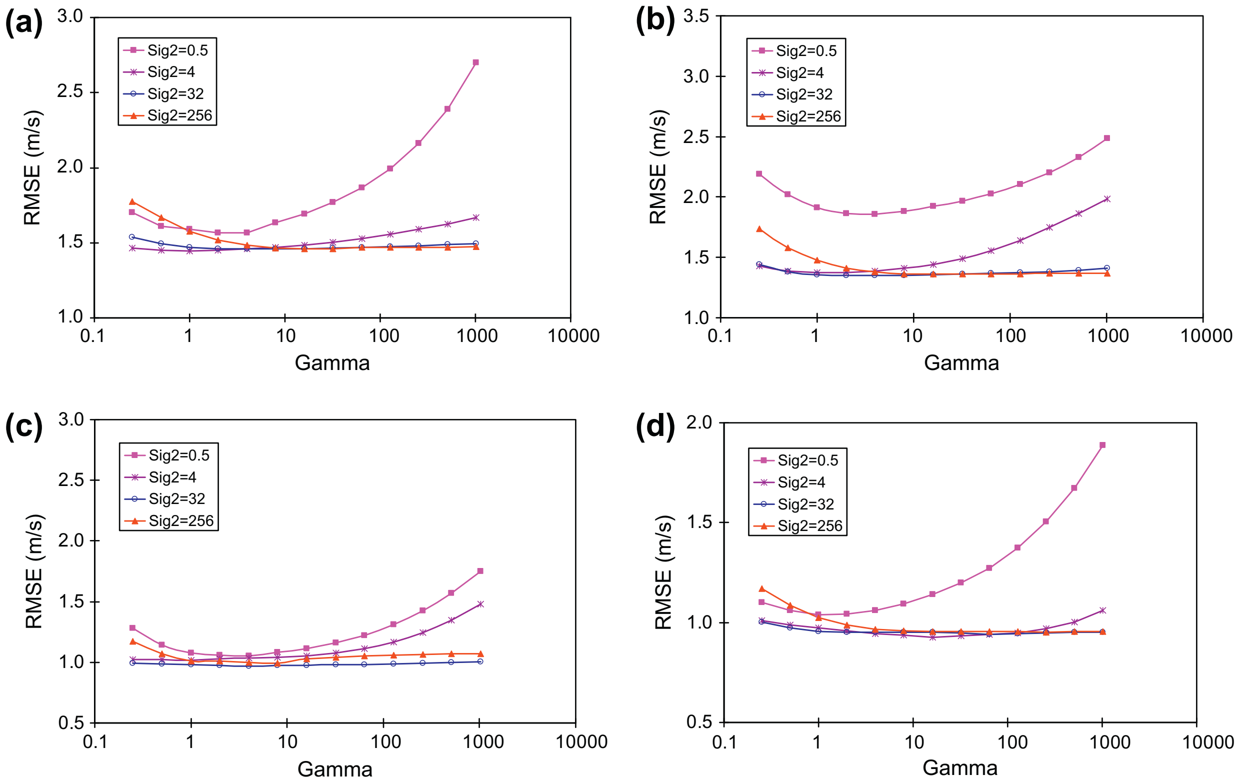

Fig. 5. Effect of regularization parameter γ (Gamma, in logarithm scale) and kernel parameter σ^2 (sig2) for the Gaussian kernel on RMSE for: (a) spring, (b) summer, (c) fall, and (d) winter.

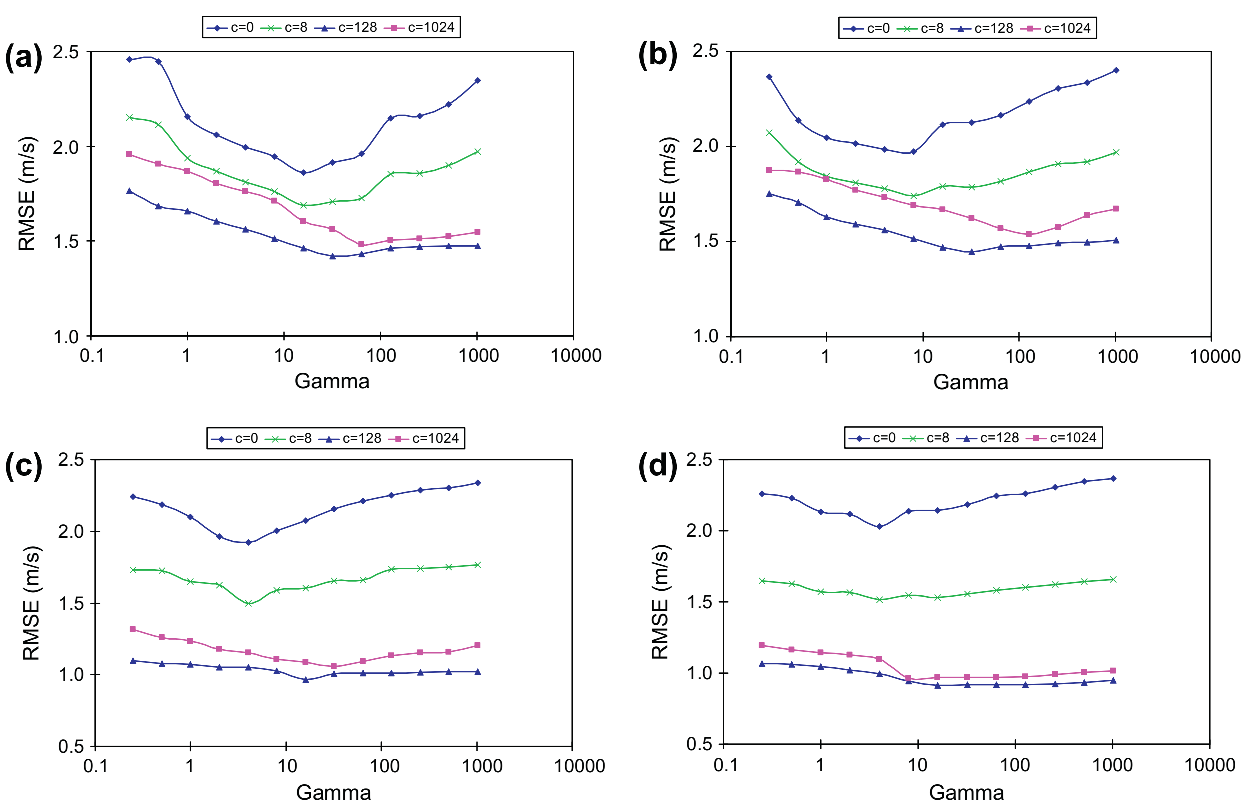

Fig. 6. Effect of regularization parameter γ (Gamma, in logarithm scale) and kernel parameter c for the polynomial kernel on RMSE for: (a) spring, (b) summer, (c) fall, and (d) winter.

4.3. SVM order

Fig. 3 shows how the forecasting accuracy varies with the SVM order. It can be observed that generally as more observed data points are used for forecasting the future wind speed, the value of RMSE first decreases significantly, but then bounces back slightly. This implies that while too few observed speed data are not enough to catch the trend of wind speed, too many observed speed data result in over fitting of wind speed trend, and hence deteriorate the forecasting accuracy. Moreover, for all the seasons and kernels, the highest forecasting accuracy is obtained when the SVM order is between 3 and 6. This is confirmed by the results in Table 2, where the optimal SVM order is shown to be either 4 or 5. Therefore, in the following analysis on the effect of regularization parameter and kernel parameters, we use the results obtained with SVM order of 5. In addition, the biggest improvement of forecasting accuracy with the increase of SVM order from 1 to 4 or 5 is always observed with linear kernel for all four seasons. For instance, such improvement in RMSE is 0.34, while those of Gaussian and polynomial kernels are around 0.21. This implies that the linear kernel is more sensitive to the change of SVM order, and it should be avoided when the SVM order is small.

4.4. Regularization parameter and kernel parameters

Fig. 4 shows how the forecasting accuracy varies with respect to the regularization parameter for the linear kernel. Note that the regularization parameter (Gamma) is plotted in logarithm scale. It can be observed that the value of RMSE first decreases significantly and then increases slowly as the regularization parameter (Gamma) increases from 0.25 to 1024. This is consistent with the results in Table 2, where the optimal regularization parameter for the linear kernel is shown to be either 32 or 64. The fact that a moderate regularization parameter performs best strengthens the function of regularization parameter in controlling the bias-variance trade-off in the LS-SVM cost function (see Eq. (2)), i.e. either large bias or variance deteriorates the performance of LS-SVM. Also, the effect of regularization parameter on forecasting performance appears to be not sensitive to the datasets in that all four curves in the figure basically have the same trend.

Fig. 5 shows how the forecasting accuracy varies with respect to the regularization parameter (Gamma, in logarithm scale) and the kernel parameter (σ^2) for the Gaussian kernel. Obviously, the kernel parameter greatly affects the RMSE value of forecasting. A difference of 50% can be observed between the best performing and the worst performing cases. It can be observed that, similar to the linear kernel, the optimal regularization parameter lies in the middle of the testing range for all the seasons. This phenomenon is more pronounced for the cases with smaller kernel parameters. This leads us to examine the interaction between the regularization parameter and the Gaussian kernel parameter. For all the seasons, the effect of regularization parameter is significant when the kernel parameter σ^2 is very small (e.g. 0.5), and the effect is weakened with a very large σ^2 (e.g. 256). However, when σ^2 is in the middle of the testing range (e.g. 32), the forecasting accuracy shows little variation as the regularization parameter increases from 0.25 to 1024, and it is close to the optimal performance. This indicates that for the Gaussian kernel, one can find a kernel parameter which makes to the LS-SVM model insensitive to the change of regularization parameter, and at the same time achieves excellent forecasting accuracy.

It is shown in Table 2 that the optimal degree (d) for polynomial SVM for all the seasons is 2. Therefore, we only analyze the effect of bias (c) based on the results obtained for the polynomial kernel with degree of 2. As shown in Fig. 6, the effect of the bias parameter is significant. Take the winter season for example, the worst performing case has an RMSE value of 2.37, while the best performing case has an RMSE value of 0.91. Meanwhile, similar to the linear and Gaussian kernels, it can be seen that the optimal regularization parameter lies in the middle of the testing range, for all the seasons. There is also a significant interaction between the regularization parameter and the bias parameter, similar to the interaction between the regularization parameter and the Gaussian kernel parameter in Fig. 5. By properly selecting the bias parameter, the performance variation due to the change of regularization parameter can be significantly reduced, and the polynomial LS-SVM

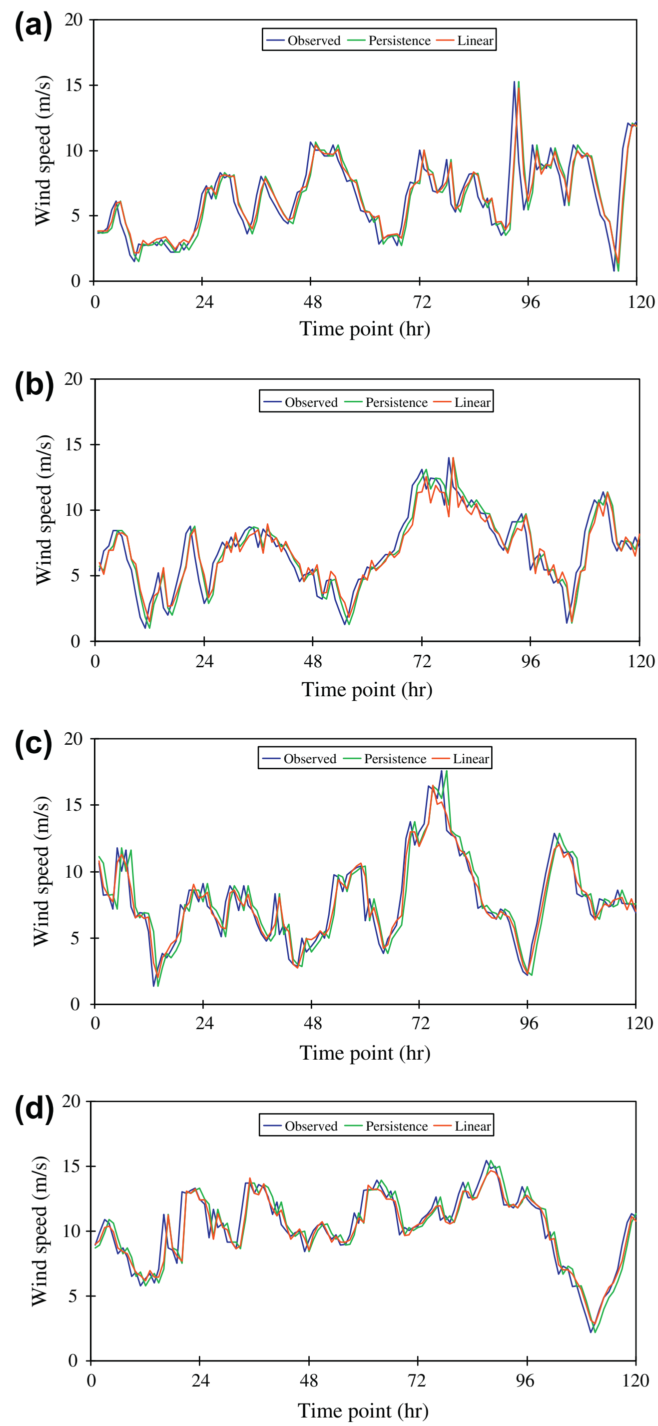

Fig. 7. Observed speed series and optimal forecasted speed series using LS-SVM with linear kernel for: (a) spring, (b) summer, (c) fall, and (d) winter.

Table 5
Forecast accuracy of the persistence approach and descriptive statistics of speed change of testing datasets.

	Statistics of speed change						RMSE
	Mean	St. dev.	Minimum	Median	Maximum	St. dev./mean	
Spring	1.16	1.18	0.01	0.82	6.17	1.01	1.650
Summer	0.96	0.74	0.00	0.77	3.61	0.76	1.209
Fall	1.19	1.04	0.00	0.88	4.63	0.87	1.584
Winter	0.90	0.82	0.00	0.73	5.51	0.92	1.216

becomes more robust. For instance, for spring season, the lowest achievable RMSE across the range of Gamma is 1.86, and the maximum variation of RMSE is about 0.6 with $c = 0$; while the values become 1.42 and 0.26, correspondingly, with $c = 128$.

4.5. Forecasted speed series and comparison with persistence approach

Although the persistence approach is the simplest method for time-series forecasting, it has excellent forecasting accuracy and great robustness in short-term, especially one-step ahead, forecasting applications. This is particularly true for wind speed forecasting, in which the volatility is usually very high. This method is widely used in practice, and often adopted as the baseline model to benchmark against when evaluating a new short-term forecasting model [5,32]. As such, it is adopted to compare with the LS-SVM models. Persistence model applies current observation at time t as the forecast on the observation at the future time $t + k$. In our case, the forecast interval $k = 1$, and the persistence model can thus be easily built by regarding current hourly wind speed as the forecast value of next hour's observation.

Fig. 7 shows the observed wind speed series, the forecasted series by LS-SVM, and the forecasted series by persistence model. Recall that the LS-SVMs with linear, Gaussian, and polynomial kernels have comparable forecasting accuracy when the optimal settings listed in Table 4 are used. For the purpose of brevity, therefore, we only include the forecasted series given by LS-SVM with the linear kernel and optimal parameters. It can be observed that the LS-SVM and persistence series show discrepancy for each season.

A comparison on the RSME values of the persistence approach in Table 5 and the corresponding values of LS-SVMs in Table 2 shows that LS-SVMs with linear, Gaussian, and polynomial kernels perform significantly better in spring, fall, and winter, but slightly poorer in summer, than the persistence approach. The advantage of LS-SVM models over persistence model is most pronounced for the fall season, where the RMSE values from LS-SVMs are around 0.96–0.97, while that from persistence model is 1.584. Similar to the difference of forecasting accuracy of LS-SVM in the four seasons, as shown in subsection 4.1, this can also be explained by the dynamic characteristics of wind speed.

Since the persistence approach has no training stage and can be applied directly to the testing dataset, here we evaluate the dynamic characteristics of wind speed in the testing datasets. The descriptive statistics of speed change are also summarized in Table 5. It shows that the temporal variation of wind speed is less significant in summer than in other seasons. More importantly, the ETV (St. dev./Mean) values of spring, summer, fall, and winter are 1.01, 0.76, 0.87, and 0.92 respectively. These values clearly show that the wind speed in summer has the lowest temporal variability. As mentioned above, the persistence approach estimates the forecasted speed as the speed of the closet previous moment. Therefore, the persistence approach gives highest forecasting accuracy in summer. On the other hand, the LS-SVM models perform better when the wind speed has higher variation in spring, fall, and winter, since the LS-SVM models can incorporate more historical information of the wind speed than the persistence model.

5. Conclusions

SVM is a powerful tool for solving nonlinear classification and function estimation problems, and has great potential for the application of wind speed forecasting. Nevertheless, like other kernel machines, SVM could give poor performance when its parameters are not tuned appropriately. In this paper, the parameters of LS-SVM are for the first time fine tuned for short-term wind speed forecasting. Three typical kernel functions, namely linear, Gaussian, and polynomial kernels, are considered for LS-SVM. The training sample size, SVM order, regularization parameter, and kernel parameters are tuned for LS-SVM using four seasonal wind speed datasets.

The results show that first of all it is not the stationary statistics but the dynamic characteristics of wind speed that determines the performance of short-term wind speed forecasting by LS-SVM. It is also found that all three kernels perform comparably in forecasting, provided that all parameters are finely tuned. However, when the training sample size and/or SVM order is small, the linear kernel produces lower forecasting accuracy and should be avoided. Furthermore, the fine-tuned LS-SVM outperforms the persistence approach in terms of forecasting accuracy for most cases, especially when the wind speed shows strong temporal variability. The effect analysis reveals that every parameter studied has a significant effect on the forecasting performance. The training sample size not only affects the forecasting efficiency, but also affects the forecasting accuracy. It is shown that there is a threshold sample size, above which the LS-SVM does not gain significantly in accuracy while lost efficiency. Moreover, the highest forecasting accuracy is obtained when the SVM order, regularization parameter, and kernel parameters are set in the middle of the testing range, indicating that extremely low or high parameters are inappropriate. For Gaussian and polynomial kernels, there is a strong interaction between the regularization parameter and kernel parameters. Thus, both types of parameters should be considered jointly rather than independently for both kernels. Finally, further research, especially the theoretical work considering the unique characteristics of wind speed time series, is needed to advance the knowledge.

References

[1] US Department of Energy, 2008. 20% Wind Energy by 2030: Increasing wind energy's contribution to US electricity supply. [Online]http://www.20percentwind.org/20percent_wind_energy_report_05-11-08_wk.pdf, Accessed in February 2010.
[2] Ackermann T, Soder L. Wind energy technology and current status: a review. Renew Sustain Energy Rev 2000;4:315–74.
[3] Monfared M, Rastegar H, Kojabadi H. A new strategy for wind speed forecasting using artificial intelligent methods. Renew Energy 2009;34:845–8.
[4] Ma L, Luan S, Jiang C, Liu H, Zhang Y. A review on the forecasting of wind speed and generated power. Renew Sustain Energy Rev 2009;13:915–20.
[5] Costa A, Crespo A, Navarro J, Lizcano G, Madsen H, Feitosa E. A review on the young history of the wind power short-term prediction. Renew Sustain Energy Rev 2008;12:1725–44.
[6] Bailey B, Stewart R. Wind forecasting for wind power stations. In: Proceedings of wind energy conversion. Edinburgh, UK; 1987. p. 265–9.
[7] Watson SJ, Landberg L, Halliday JA. Application of wind speed forecasting to the integration of wind energy into a large scale power system. In: IEEE Proc Generation, Trans Distribution 1994;141:357–62.

[8] Kamal L, Jafri YZ. Time series models to simulate and forecast hourly averaged wind speed in Quetta, Pakistan. Solar Energy 1997;61:23–32.
[9] Torres J, Garcia A, Deblas M, Defrancisco A. Forecast of hourly average wind speed with ARMA models in Navarre (Spain). Solar Energy 2005;79: 65–77.
[10] Alexiadis M. Short-term forecasting of wind speed and related electrical power. Solar Energy 1998;63:61–8.
[11] Mohandes M, Rehman S, Halawani TO. A neural networks approach for wind speed prediction. Renew Energy 1998;13:345–54.
[12] More A, Deo MC. Forecasting wind with neural networks. Mar Struct 2003;16:35–49.
[13] Li G, Shi J. On comparing three artificial neural networks for wind speed forecasting. Appl Energy 2010;87:2313–20.
[14] Kariniotakis G, Stavrakakis GS, Nogaret EF. A fuzzy logic and neural network based wind power model. In: Proceedings of the 1996 European wind energy conference, Goteborg, Sweden, 1996. p. 596–9.
[15] Wang X, Sideratos G, Hatziargyriou N, Tsoukalas LH. Wind speed forecasting for power system operational planning. In: 2004 International conference on probabilistic methods applied to power systems. 2004. p. 470–4.
[16] Trafalis BT, Adrianto I, Richman MB. Active learning with support vector machines for tornado prediction. Lect Notes Comput Sci 2007;4487: 1130–7.
[17] Hong W. Rainfall forecasting by technological machine learning models. Appl Math Comput 2008;200:41–57.
[18] Osowski S, Garanty K. Forecasting of the daily meteorological pollution using wavelets and support vector machine. Eng Appl Artificial Intell 2007;20: 745–55.
[19] Zhou J, Shi J. Performance evaluation of object localization based on active radio frequency identification technology. Comput Ind 2009;60: 669–76.
[20] Widodo A, Kim EY, Son J-D, Yang B-S, Tan ACC, Gu D-S, et al. Fault diagnosis of low speed bearing based on relevance vector machine and support vector machine. Expert Syst Appl 2009;36:7252–61.
[21] Lin GF, Chen GR, Huang P-Y, Chou YC. Support vector machine-based models for hourly reservoir inflow forecasting during typhoon-warning periods. J Hydrol 2009;372:17–29.
[22] Sreelakshmi K, Kumar PR. Performance evaluation of short term wind speed prediction techniques. Int J Comput Sci Network Security 2008;8:162–9.
[23] Belousov A, Verzakov SA, Von Frese J. A flexible classification approach with optimal generalisation performance; support vector machines. Chemom Intell Lab Syst 2002;64:15–25.
[24] Thissena U, van Brakela R, de Weijerb AP, Melssena WJ, Buydens LMC. Using support vector machines for time series prediction. Chemom Intell Lab Syst 2003;69:35–49.
[25] Mohandes M. Support vector machines for wind speed prediction. Renew Energy 2004;29:939–47.
[26] Kusiak A, Zheng H, Song Z. Short-term prediction of wind farm power: a data mining approach. IEEE Trans Energy Convers 2009;24:125–36.
[27] Suykens JAK, Vandewalle J. Least squares support vector machine classifiers. Neural Process Lett 1999;9:293–300.
[28] Gestel TV, Suykens JAK, Baesens B, Viaene S, Vanthienen J, Dedene G, et al. Benchmarking least squares support vector machine classifiers. Machine Learn 2004;54:5–32.
[29] Pai P-F, Hong W-C. Support vector machines with simulated annealing algorithms in electricity load forecasting. Energy Convers Manage 2005;46:2669–88.
[30] Li Q, Meng Q, Cai J, Yoshino H, Mochida A. Predicting hourly cooling load in the building: a comparison of support vector machine and different artificial neural networks. Energy Convers Manage 2009;50:90–5.
[31] Suykens JAK, Lukas L, Van Dooren P, De Moor B, Vandewalle J. Least squares support vector machines classifiers: a large scale algorithm. In: Proceedings of the European conference on circuit theory and design (ECCTD-99), Stresa, Italy, September 1999. p. 839–42.
[32] Landberg L, Watson SJ. Short-term prediction of local wind conditions. Boundary-layer Meteorology 1994;70:171–95.