

Research Article

Support Vector Regression Based on Grid-Search Method for Short-Term Wind Power Forecasting

Hong Zhang,^{1,2} Lixing Chen,^{1,2} Yong Qu,³ Guo Zhao,^{1,2} and Zhenwei Guo⁴

¹ School of Electrical Engineering, Southeast University, Nanjing, Jiangsu 210096, China

² Jiangsu Key Laboratory of Smart Grid Technology and Equipment, Nanjing 210096, China

³ VLSI Lab, Nanyang Technological University, Singapore 639798

⁴ School of Information Science and Engineering, Hunan University, Changsha 410082, China

Correspondence should be addressed to Hong Zhang; hongzhangseu@gmail.com

Received 16 November 2013; Revised 18 April 2014; Accepted 23 April 2014; Published 17 June 2014

Academic Editor: Hongjie Jia

Copyright © 2014 Hong Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The purpose of this paper is to investigate the short-term wind power forecasting. STWPF is a typically complex issue, because it is affected by many factors such as wind speed, wind direction, and humidity. This paper attempts to provide a reference strategy for STWPF and to solve the problems in existence. The two main contributions of this paper are as follows. (1) In data preprocessing, each encountered problem of employed real data such as irrelevant, outliers, missing value, and noisy data has been taken into account, the corresponding reasonable processing has been given, and the input variable selection and order estimation are investigated by Partial least squares technique. (2) STWPF is investigated by multiscale support vector regression (SVR) technique, and the parameters associated with SVR are optimized based on Grid-search method. In order to investigate the performance of proposed strategy, forecasting results comparison between two different forecasting models, multiscale SVR and multilayer perceptron neural network applied for power forecasts, are presented. In addition, the error evaluation demonstrates that the multiscale SVR is a robust, precise, and effective approach.

1. Introduction

Compared to the traditional thermal power [1], wind energy is a significant aspect of renewable energy and it is getting more and more attention due to it is a renewable, inexhaustible and free source. Therefore, the corresponding wind energy forecasting becomes a critical issue for dispatch and scheduling of power systems [2]. Precise wind energy forecasting can balance and integrate the multiple volatile power sources at all levels of the transmission and distribution grid [3]. Moreover, accurate short-term wind power forecasting can reduce the problems which are caused by grid integration and energy trading [4].

Short-term wind distribution is essentially a random one although it can be described by a continuous probability distribution named Weibull distribution in a long term. It is hard to obtain the intrinsic regulation of wind speed in a short-term; thus, the soft computing can play a significant role in

short-term wind power forecasting. SVR technique has already been employed for the short energy forecasting in the existing literatures; however, many of them assume that the employed data have good quality. In fact, it is impossible to acquire the data without noise because there are many reasons that are sometimes beyond the control of human operators [5]. Generalization capability of traditional forecast approach is relatively low because the employed approach is seriously dependent on the quality of sample. In this paper, we optimize the parameters for two different kernel functions, radial basis function (RBF) or polynomial function (PF), based on the Grid-search method for SVR such that the accuracy of forecasting results is improved. The main algorithm flow is illustrated in Figure 1.

Figure 1 provides a brief illustration for the main processing step of proposed approach in this paper. The rest of this paper is organized as follows: formalization and illustration related to the dataset are investigated in Section 2. A brief

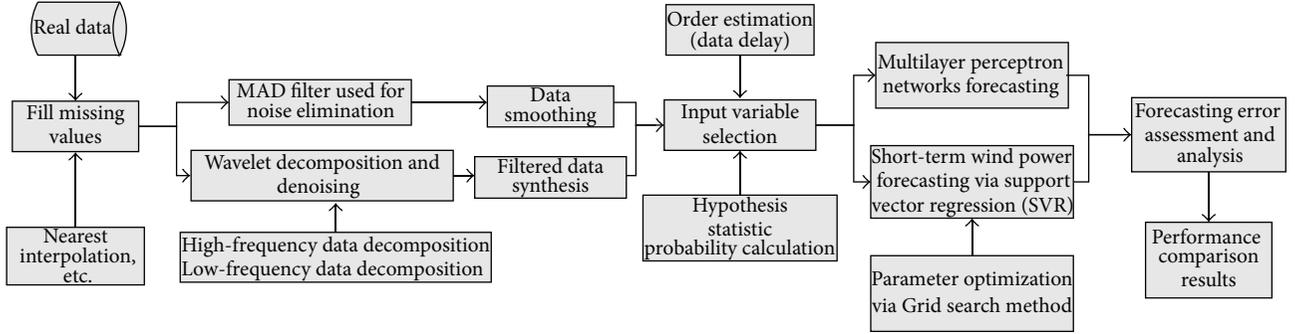


FIGURE 1: Algorithm flow.

review on wind energy forecasting is given in Section 3. Section 4 presents the detailed information related to our proposed approach including the basic employed theory. In Section 5, firstly, the missing value of employed data is filled by the data interpolation techniques. Secondly, two different filter methods (median absolute deviation (MAD) filter and wavelet decomposition and denoising) are used to eliminate the irrelevant, noisy, and outlier value. Thirdly, the input variable selection is investigated by Partial least squares (PLS) technique. Fourthly, the data order estimation is implemented through the cross-correlation methods. Fifthly, the multiscale SVR in combination with Grid-search technique is applied to forecast the short-term wind power. Finally, the performance evaluation and error analysis are given. In Section 6, the proposed results and the prospective research questions are summarized and discussed, respectively.

2. Problem Description

2.1. Data Resources. The quality of data samples plays an important role in wind power forecasting because it has a direct impact on forecasting performance. The data quality

is the fundamental issue for the data analysis; in particular, data do not exist without noise in the real application. The main objective of data analysis is to discover knowledge which will be used to solve real problem and make decisions [5]. The wind tower which is employed to collect the data contains two different heights, 30 m and 60 m. The data have been measured every 3 minutes, a total of about five days of measurement data are selected. Specifically, the employed data contain few irrelevant, corrupt, and noisy ones which must be removed and filled from the data in order to proceed with further data analysis. Taking into account the real application, because the data sampling equipment sometime encounter a temporary mistake and the associated data are recorded as zero, therefore, these data with bad quality are eliminated. Typically, the short-term wind power prediction is within 4 days and ultra-short-term one is within 4 hours. The paper mainly discussed the short-term wind power prediction by SVR techniques based on Grid-search method.

2.2. Formalization. In this paper, the short-term wind forecasting issue is formulated as a regression problem. The time series is denoted by

$$x_i(t_i) = (x_{i1}^{AWS}, x_{i2}^{AWD}, x_{i3}^{TEM}, x_{i4}^{HUM}, x_{i5}^{PRE}, x_{i6}^{AWS}, x_{i7}^{AWD}, x_{i8}^{VAR}, x_{i9}^{RWS}, x_{i10}^{RWD})^T, \quad i = 1, \dots, N, \quad (1)$$

where 10 variables indicate 30 m average wind speed, direction, temperature, humidity, and pressure and 60 m average wind speed, direction, variance, real wind speed, and real wind direction, respectively; N is a positive integer with the value more than one. In general, (1) provides all the factors in real application for short-term wind power forecasting. The primary task of this paper is to predict the output wind power $y_i(t_i)$, $i = 1, \dots, N$ at time $t = t_i$ based on the wind measurements through $x_i(t_i)$ at time $t_i, t_{i-1}, \dots, t_{i-\beta}$, $i > \beta$, $i, \beta \in N^+$, past observations.

3. Related Work

Short-term wind power forecast has attracted more and more attention in recent decades. Alessandrini et al. [4]

discussed the comparison between ECMWF EPS (Ensemble Prediction System in use at the European Centre for Medium-Range Weather Forecasts) and COSMO-LEPS (Limited-area Ensemble Prediction System developed within Consortium for Small-scale Modelling) based on two forecasting models. As a survey associated with short-term prediction in the last 30 years; Costa et al. [6] investigated the performance of two principal approaches (mathematical and physical). Kramer and Gieseke [2] applied the SVR techniques for short-term wind energy forecasting based on the real world wind power data from the National Renewable Energy Laboratory (NREL) western wind resource dataset, and Osowski and Garanty [7] utilized the SVR and wavelet decomposition method to forecast the short-term air pollution. Chen et al.

[8] proposed new learning techniques based on support vector machine (SVM) model for the power load forecasting and conducted experimental results for the short-term load forecasting. Chang [9] proposed a hybrid method which combined the orthogonal least squares (OLS) algorithm and genetic algorithm (GA) for short-term wind power forecasting based on the radial basis function (RBF) neural network. Che et al. [10] applied a two-voltage stage topology with boost converter to improve the conversion efficiency of commercial small wind grid inverter by proposed control strategy. Based on the combination of wavelet transform and neural network arithmetic, Song et al. [11] dealt with an energy management related to a hybrid power generation system such that the stability of power generation system was improved greatly. Moreover, a bibliographical survey associated with the general application of research and developments was presented by Lei et al. [12] in the fields of wind power forecasting. Li et al. [13] presented ideal subspace approximation techniques based on a chaotic time series and nonlinear Kalman filtering, and the wind speed prediction experiments were used to demonstrate the high chaotic prediction accuracy. Hoai et al. [14] optimized an empirical-statistical downscaling technique for prediction based on a feed-forward multilayer perceptron (MLP) neural network, and they gave the numerical simulation to demonstrate the robustness of proposed technology. Sánchez [15] gave a statistical forecasting system for short-term wind power prediction in 48 hours ahead based on the techniques combination with recursive methods and schemes with adaptability. An in-depth review of the current wind power generation method and advances in wind power forecasting was formulated by Foley et al. [16], and Botterud et al. [17] discussed the current development of wind power forecasting in US ISO/RTO markets and the application of state-of-the-art forecasts refers to associated forecasting. Based on the numerical and statistical models, Stathopoulos et al. [18] proposed some strategies for accurate local wind forecasts by the combination with statistical post processes.

4. Theory for Proposed Approach

Data analysis is the fundamental approach for the knowledge investigation [19]. Proper data preprocessing can eliminate the unreasonable trend of data without loss of data characteristics. Therefore, efficient techniques for automatic data preprocessing are crucial [20]. In this section, the basic theory illustration refers to data preprocessing and multiscale SVR are introduced as follows.

4.1. Missing Value. In the data preprocessing, the issues that must be considered are about the irrelevant, missing value, and noisy data. The data with poor quality will result from the poor performance of final employed approaches. In this case, the data are not complete because there are missing values which cannot be eliminated because of the requirement for time continuity as well as they may contain useful information [5]. In this paper, the missing value is filled via data interpolation by the intrinsic relationship between data; for instance, the missing value of wind power can be filled by

its associated influential factor wind speed. The nearest-neighbor interpolation (also named as proximal interpolation) is a typical method of multivariate interpolation in one or more dimensions, which can be used to approximate the value of a nongiven point based on the corresponding point of the neighborhood. Under normal circumstances, the filled data is reasonable because the value of data is not suddenly changed in the real application, and the neighboring data correlation is taken into account in nearest-neighbor interpolation method.

4.2. MAD Filter. Median absolute deviation from the median x^* for a data sequence (x_1, x_2, \dots, x_N) is more robust with good performance in the presence of multiple outliers [21, 22]. The Hampel filter (also known as MAD filter) which is used to eliminate the outlier is denoted by

$$ts < |x_i - x_i^*|, \quad s = a \cdot M, \quad i = 1, \dots, n, \quad (2)$$

where $a = 1.4826$ and $M = \text{median}\{|x_i - x_i^*|\}$ is the median value of $|x_i - x_i^*|$. t is the threshold employed to control range of convergence, and it can be estimated based on the sample standard deviation of the distribution. The MAD filter can replace the outlier-sensitive mean and standard deviation estimates with the outlier-resistant median as well as MAD from the data [23].

4.3. Wavelet Decomposition and Denoising. Unlike the MAD filter, the wavelet decomposition and denoising analysis for data is localized in both time domain and frequency domain, and it can be used to decompose the original data into high-frequency component (HFC) and low-frequency component (LFC). Typically, the HFC denotes the detailed information such as mutant value, while the LFC usually represents the generalized or stationary characteristic related to employed data. The more detailed discussion of wavelet decomposition and denoising can be founded in [24–27].

Comparing to the continuous wavelet transform, the discrete wavelet transform (DWT) is more commonly used in real application and defined by

$$C_f(j, k) = \int_{-\infty}^{\infty} f(t) \varphi_{jk}(t) dt, \quad (3)$$

$$\forall f(t) \in L^2(R), \quad j, k \in Z,$$

where \forall means for any $\varphi_{jk}(t)$ is a basic wavelet, and $\varphi_{jk}(t) = a_0^{j/2} \varphi_{jk}(a_0^j t - kb_0)$, $a_0, b_0 \in \mathbb{R}^1, a_0 \neq 0$. The above DWT is also known as Mallat algorithm [28], in this paper, the Vaidyanathan filter [29–31] is applied for the implementation of data decomposition and denoising.

4.4. Input Variable Selection. Swedish statistician named Herman Wold first introduces Partial least squares (PLS) technique which is used to find the fundamental relations between two variables (x and y); that is, a latent variable approach is employed to investigate the covariance structures between variables. Partial correlation can be used to explore

the association between pairs of random variables in the presence of other variables [32, 33], and its coefficients can be calculated between the variables and exclude the influence of other variables, then the main instruction of PLS coefficients can be derived via the following three steps.

(i) *Hypothesis.* Consider three variables, y_1 , y_2 , and x_3 ; the partial correlation coefficient $r_{12(3)}$ between y_1 and y_2 given x_3 is defined by

$$r_{12(3)} = \frac{(r_{12} - r_{13}r_{23})}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}, \quad (4)$$

where r_{ij} is the product-moment correlation coefficient between variables with subscripts i and j . The range of PLS coefficients values is $(-1, 1)$, in particular, 0 indicates no association between y_1 and y_2 .

(ii) *Statistic.* A test associated with the full correlation coefficient is used to test the original hypothesis in Step (4) under the assumption that data has an approximately normal distribution. If the $r_{12(3)}$ is the obtained partial correlation coefficient, then the appropriate t statistics is denoted as

$$t_{n-q-2} = \frac{r_{12(3)}\sqrt{n-q-2}}{1 - r_{12(3)}^2}, \quad (5)$$

where t_{n-q-2} has an approximate Student's t -distribution, and its degrees of freedom are $n - q - 2$; n is the number of observations from the computed full correlation coefficients.

(iii) *Probability Calculation.* Calculate the observation of test t -statistics as well as its corresponding values of probabilities P . If the value of probability P , which is used to test t -statistics, is less than the value of given significance level α , then the original hypothesis should not be accepted; otherwise, it is available for the test.

4.5. *Order Estimation.* Data order reflects the intrinsic relationship between past data and current data, which is derived through the autocorrelation function (ACF) for one data sequence and cross-correlation function (CCF) for two different data sequences. Autocorrelation (is also sometimes called "lagged correlation" or "serial correlation") refers to the correlation of a time series with its own past and future values, which relate to the correlation between members of a series of numbers arranged in time. CCF is a measure of similarity of two given date as a function of a time lag, and it is commonly used for the replacement of a long data by a shorter and suitable length. In the discrete domain, ACF and CCF for two real time series x_i , y_j , $i = 0, 1, \dots, M - 1$, $j = 0, 1, \dots, N - 1$ are defined by

$$r_{xx}(l) = \sum_{n=-\infty}^{\infty} x(n)x(n-l) = r_{xx}(-l), \quad l = 0, \pm 1, \pm 2, \dots, \quad (6)$$

$$r(x, y, l) = \sum_{j=\max(0, l)}^{\min(M-1+l, N-1)} x(j-l)y(j), \quad (7)$$

$$l = -(M + 1), \dots, 0, \dots, (N - 1).$$

In MATLAB, ACF and CCF are computed with the function "xcorr" which are defined by (6) and (7) in the frequency domain, respectively.

4.6. *Support Vector Regression.* SVM for regression was proposed in 1996 by Drucker et al. in [33]; this method is called support vector regression (SVR), and its basic idea is based on support vector classification, more precisely, the fact that the cost function does not take into account the training points that lie beyond the margin; thus, the SVR only depends on a subset of the training data. Analogously, least squares support vector machine (LS-SVM) which is known as another SVM has been presented by Suykens and Vandewalle [34]. Vanik-Chervonenkis theory and structural risk minimization (SRM) are the fundamental theory for the SVM [35, 36]. SVM is to investigate the intrinsic relationship between the prediction model related to wind time series and learning capability and derive the best generalization capability, if the given sample dataset is $X \times Y = \{(x_1, y_1), \dots, (x_i, y_i), \dots, (x_N, y_N)\} \in R^N \times R$.

The basic task of regression is to establish the nonlinear function $f : R^N \rightarrow R$, such that $y = f(x)$; the estimation function and loss function are, respectively, defined as

$$y = f(x) = [\omega \cdot \phi(x)] + b,$$

$$\min Q = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^N (\xi_i^* + \xi_i) \quad (8)$$

subject to $y_i - [\omega \cdot k\phi(x_i)] - b \leq \varepsilon + \xi_i^*$,

$$[\omega \cdot k\phi(x_i)] + b - y_i = \varepsilon + \xi_i,$$

$$\xi_i^*, \xi_i \geq 0, \quad i = 1, \dots, N,$$

where C is penalty factor which is employed for empirical risk and confidence range, ξ_i^* , ξ_i are relaxation factors which are used to modify the convergence speed, ε is loss function which is applied to estimate the prediction accuracy for (8), and its detailed information is given as

$$L_\varepsilon(y) = |f(x) - y| - \varepsilon, \quad \text{if } |f(x) - y| \geq \varepsilon. \quad (9)$$

Moreover, the $L_\varepsilon(y)$ is equal to zero if $|f(x) - y| < \varepsilon$. Essentially, (8) and (9) are equivalent quadratic convex optimization problems which are defined as

$$\max_{\varepsilon, \alpha_i^*} P(\alpha_i, \alpha_i^*)$$

$$= \sum_{i=1}^N y_i (\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^N (\alpha_i + \alpha_i^*)$$

$$- \frac{1}{2} \sum_{i,j=1}^N (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(x_i, x_j)$$

subject to $\sum_{i=1}^N \alpha_i = \sum_{i=1}^N \alpha_i^*$, $0 \leq \alpha_i, \alpha_i^* \leq C$, $i = 1, \dots, N$, (10)

where if the value of C is large, then the accuracy capability of data fitting will be more good in the complexity and approximation error of the control model; ε is applied to control the generalization capability and regression approximate error, α_i , α_i^* are the Lagrange multiplier, when they are not equivalent to zero and then the SVR can be used for regression prediction, and $K(x_i, x_j)$ is the kernel function which is used to simulate the inner production, which can be given as radial basis function or polynomial function. Furthermore, the regression function is $y_j = \sum_{i=1}^{N-m} (\alpha_i - \alpha_i^*) K(\bar{x}_i, \bar{x}_j) + b$; $j = m + 1, \dots, N$, the regression step l th step prediction can be denoted by

$$x_{n+l} = \sum_{i=1}^{N-m} (\alpha_i - \alpha_i^*) K(\bar{x}_i, \bar{x}_{N-m+l}) + b, \quad (11)$$

$$\bar{x}_{N-m+l} = \{x_{N-m+l}, \dots, x_{N+l-1}\}, \quad l = 1, 2, \dots, L.$$

4.7. Error Evaluation Criteria. Consider

$$\text{RMSE} = \frac{\sqrt{\sum_{t=1}^n (P_{ft} - P_{rt})^2}}{\sqrt{n}}, \quad \text{MAE} = \frac{\sum_{t=1}^n |P_{rt} - P_{ft}|}{\sqrt{n}},$$

$$\text{RMAE} = \frac{\sum_{t=1}^n |P_{rt} - P_{ft}|}{\sum_{t=1}^n |P_{rt}|}, \quad (12)$$

where MAE is the absolute average error between forecasting power P_{ft} and real power P_{rt} over the time t . n is the number of test sample; RMSE and RMAE are the root mean square error and relative average absolute error, respectively.

5. Experimental Analysis

In this section, the numerical simulation is constructed for each part of Section 4 based on the wind resources dataset which are presented in Section 2.1.

5.1. Data Interpolation and Filter. In this section, the data interpolation, MAD filter, wavelet decomposition, and denoising technique which correspond to the theory analysis in Sections 4.1–4.3 are, respectively, to handle the missing value, irrelevant, outliers, and noisy data. The variables which can be used as the forecasting factors consist of two groups, that is, 30 m average wind speed (AWS), direction (AWD), temperature (Tem), humidity (Hum), and pressure (Pre) and 60 m AWS, AWD, AWS variance, real wind speed (RWS), real wind direction (RWD), and power. The above variables are labeled in accordance with the above order as the first variable to eleventh variable, respectively. Moreover, the state trajectory of original data, filled data, and filtered data are given in Figures 2, 3, and 4, respectively.

Figures 2–4, respectively, denote the trajectory of original data, filled data, and filtered data which are derived, respectively, by nearest-neighbor interpolation (NNI), MAD filter, wavelet decomposition, and denoising techniques, where NNI is mainly focused on the relationship between 30 m

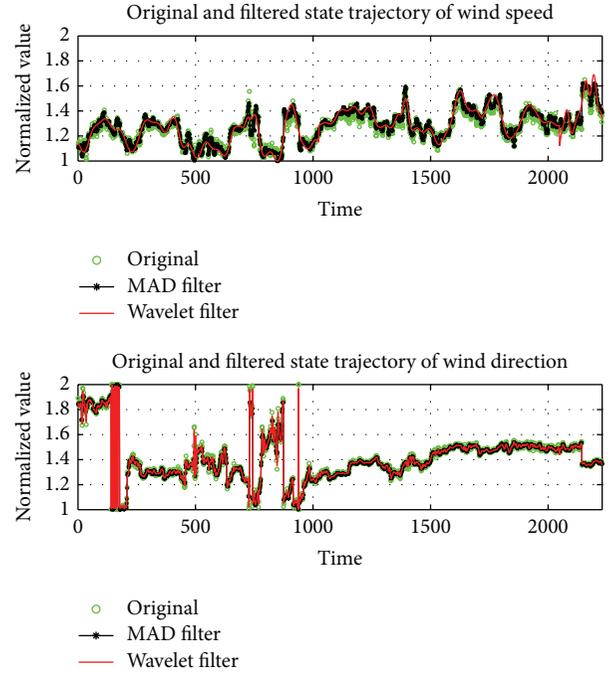


FIGURE 2: State trajectories refer to the 1st variable and 2nd variable.

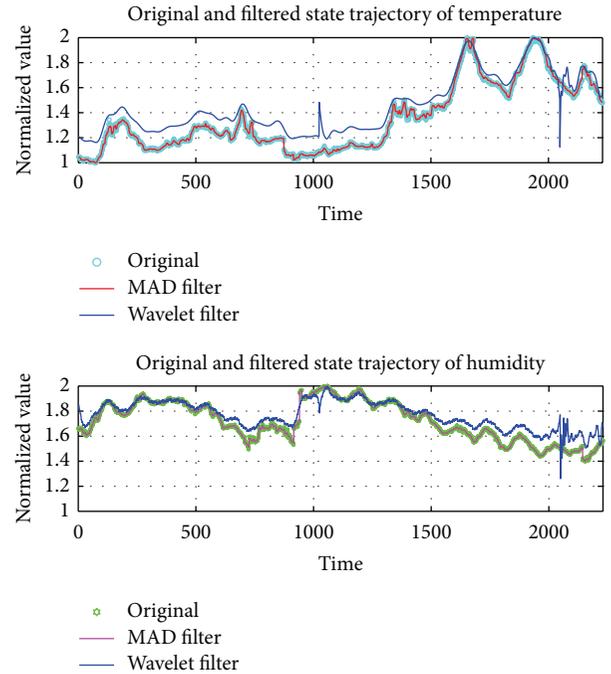


FIGURE 3: State trajectories refer to the 3rd variable and 4th variable.

AWS and power based on the discussion in Section 5.2. Because 10 samples are collected in half an hour according to sampling frequency, half of windows size of MAD filter is set as 5, and the Vaidyanathan wavelet filter is employed for the data preprocessing. Without loss of generality, other wavelet transform such as Daubechies wavelet are also accepted.

TABLE 1: PLS coefficients refer to eleven variables.

Variables	11	1	2	3	4	5
11	1.0000	-0.7644	-0.7185	-0.6779	-0.7602	-0.7214
6	-0.7595	0.9848	0.5741	0.8837	0.9815	0.5689
7	-0.7191	0.5944	0.9437	0.6413	0.5906	0.9228
8	-0.9648	0.8005	0.7235	0.7718	0.7949	0.7259
9	-0.9732	0.6883	0.6697	0.5516	0.6854	0.6705
10	-0.9998	0.7621	0.7174	0.6748	0.7579	0.7205

TABLE 2: Lower and upper bound related to Table 1.

Variables		11	1	2	3	4	5
11	UB	1.0000	-0.7434	-0.6942	-0.6508	-0.7389	-0.6973
	LB	1.0000	-0.7838	-0.7412	-0.7033	-0.7799	-0.7439
6	UB	-0.7382	0.9862	0.6058	0.8939	0.9832	0.6009
	LB	-0.7793	0.9832	0.5406	0.8726	0.9796	0.5352
7	UB	-0.6948	0.6249	0.9488	0.6690	0.6213	0.9297
	LB	-0.7417	0.5621	0.9381	0.6118	0.5580	0.9152
8	UB	-0.9613	0.8173	0.7458	0.7907	0.8121	0.7481
	LB	-0.9680	0.7824	0.6995	0.7514	0.7763	0.7021
9	UB	-0.9705	0.7130	0.6956	0.5845	0.7103	0.6964
	LB	-0.9757	0.6619	0.6420	0.5169	0.6588	0.6429
10	UB	-0.9998	0.7818	0.7402	0.7004	0.7779	0.7430
	LB	-0.9998	0.7410	0.6930	0.6475	0.7365	0.6963

From Figures 2–4, irrelevant, outliers, and noisy data of employed data are, respectively, eliminated and filled by the MAD filter and wavelet transform. The data become a relative stationary and smooth time series now, which is a necessary step for the further analysis.

5.2. *Input Variables Selection Results.* This section provides the simulation results with respect to the theory analysis of Section 4.4. The quality and quantity of samples have significant impact on the accuracy of forecasting because more data will increase the difficulty of real operation, and less data may not contain enough information for analysis. As the discussion of Section 5.1, the PLS coefficient with regard to eleven variables is computed by (4) and (5) and are given in Tables 1 and 2.

Based on the discussion of Section 4.4, all the value of “significance value” is equivalent to about 0.6936×10^{-308} , which is far less than the set value $\alpha = 0.01$; in other words, the original hypothesis should not be accepted because there are significant correlation between variables. Moreover, the lower and upper bound with confidence level of 95% are provided in Table 2, where UB represents the upper bound and LB represents the lower bound. From Tables 1 and 2, we can learn two facts: one is that all the obtained data in Table 1 are accepted because of the corresponding testing value which is presented in Table 2; the other one is that one group is enough for the wind power forecasting because of the similarity between different variables.

5.3. *Data Order Estimation and Normalization.* Based on the discussion of Section 5.2, the 30 m variables except for the pressure are selected as the input variables because there is no significant change about pressure. The data order estimation

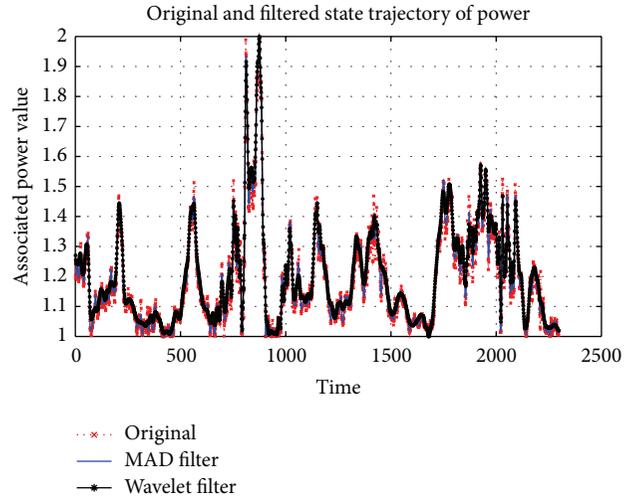


FIGURE 4: State trajectories refer to the 11th variable.

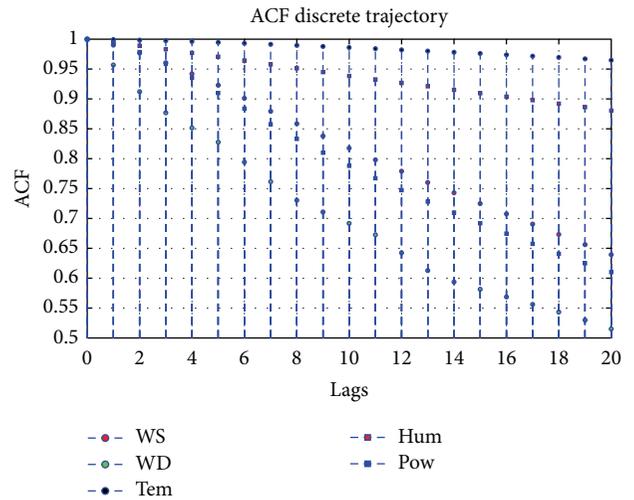


FIGURE 5: ACF discrete trajectory.

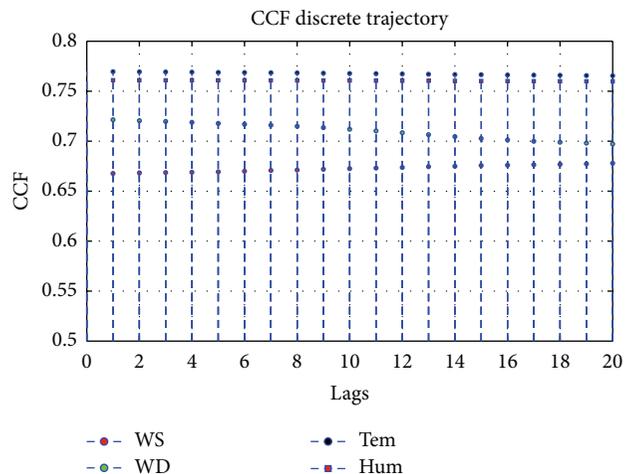


FIGURE 6: CCF discrete trajectory.

TABLE 3: Simulation results refer to multiscale SVR.

Item		BCVM	Bc	Bg	Foin	RM	RS
RBF(3)	O	0.00121642	1024	0.00195313	3465	0.00383891	0.943034
	M	0.00118972	64	0.015625	1756	0.00194981	0.972598
	W	0.00217458	0.5	1	1628	0.0211323	0.897533
PF(3)	O	0.00121642	1024	0.00195313	4365	0.00383891	0.943034
	M	0.00118972	64	0.015625	2756	0.00194981	0.972598
	W	0.00217458	0.5	1	1628	0.0211323	0.897533
RBF(6)	O	0.00117004	1024	0.00390625	4562	0.00381612	0.94281
	M	0.0010876	128	0.00390625	2810	0.00189855	0.972905
	W	0.0021321	0.5	1	1534	0.0219367	0.895725
PF(6)	O	0.00117004	1024	0.00390625	4562	0.00381612	0.94281
	M	0.0010876	128	0.00390625	2810	0.00189855	0.972905
	W	0.0021321	0.5	1	1534	0.0219367	0.895725
RBF(10)	O	0.0011804	512	0.00195313	5366	0.00381804	0.94289
	M	0.00107203	128	0.00390625	2441	0.00191227	0.972918
	W	0.00217458	0.5	1	1534	0.0219367	0.895725
PF(10)	O	0.0011804	512	0.00195313	5366	0.00381804	0.94289
	M	0.00107203	128	0.00390625	2441	0.00191227	0.972918
	W	0.00214353	0.5	1	1534	0.0219367	0.895725

O: original data; M: MAD filter; W: wavelet filter; RBF(3): number of cross-validation for testing set is 3 by RBF, similar to RBF(6) and RBF(10); PF(3): number of cross-validation for testing set is 3 by PF, similar to PF(6) and PF(10); BCVM: best cross-validation mean squared error; Bc: best c; Bg: best g; Foin: finished optimization iteration number; RM: regression mean squared error; RS: regression squared correlation coefficient.

is implemented through (6) and (7) and denoted in Figures 5 and 6, respectively.

From Figures 5 and 6, we can learn that the two past values are the most significant impact factors for the current value because the ACF value is about more than 0.9 and CCF value is more than 0.7. Thus, 2 can be set as the data order based on the data estimation by performance of ACF and CCF.

5.4. Multiscale SVR Performance. The following simulation is based on [37, 38] libsvm (2013, version 3.17, platform: MATLAB 2012b, Microsoft Visual C++ 2010), and the kernel functions are radial basis function (RBF) and polynomial function (PF). Without loss of generality, 80% and 20% of employed data (2297) are selected as the training sample and testing samples, respectively. Because the changes in wind direction are not obvious, so both sine and cosine with regard to AWD are selected as the input variables. Therefore, the number of input variables is essentially about 10 because the data order is set as 2. In order to obtain the accuracy comparison between the different variables, all the variables value are normalized at the range (1,2). Moreover, the k-fold cross-validation (K-CV) [39–42] and “Grid search” are utilized for the parameters selection. “Grid search” is to try every possible value of the parameters (c, g), and the best accuracy of the (c, g) can be derived based on the K-CV, where c and g are, respectively, penalty factor and kernel function parameters. The state trajectory of original data and forecasting data are given in Figures 7, 8, 9, and 10. Moreover, the detailed information refers to simulation results comparison between multiscale SVR and MLP neural network will be given in Tables 3 and 4.

TABLE 4: RMSE, MAE, and RMAE refer to SVR and MLP.

Item		RMSE	MAE	RMAE	Hn	Et
RBF(3)	O	1.6044	1.0931	0.1094	NA	132.122945
	M	1.0887	0.6336	0.0638	NA	110.138815
	W	3.3950	2.3570	0.2375	NA	113.548014
PF(3)	O	1.6044	1.0931	0.1094	NA	137.698291
	M	1.0887	0.6336	0.0638	NA	112.438881
	W	3.3950	2.3570	0.2375	NA	113.008574
RBF(6)	O	1.5997	1.0888	0.1090	NA	374.338732
	M	1.0743	0.5767	0.0581	NA	309.984871
	W	3.4590	2.4161	0.2435	NA	288.569125
PF(6)	O	1.5997	1.0888	0.1090	NA	346.989140
	M	1.0743	0.5767	0.0581	NA	280.287914
	W	3.4590	2.4161	0.2435	NA	284.077991
RBF(10)	O	1.6001	1.0892	0.1090	NA	624.630533
	M	1.0782	0.5872	0.0591	NA	501.316555
	W	3.4590	2.4161	0.2435	NA	368.402477
PF(10)	O	1.6001	1.0892	0.1090	NA	623.420666
	M	1.0782	0.5872	0.0591	NA	512.878352
	W	3.4590	2.4161	0.2435	NA	501.985355
MLP neural network		11.6547	9.6265	0.9631	20	97.536770

O: original data; M: MAD filter; W: wavelet filter; RBF(3): number of cross-validation for testing set is 3 by RBF, similar to RBF(6) and RBF(10); PF(3): number of cross-validation for testing set is 3 by PF, similar to PF(6) and PF(10); Hn: number of hidden layer; Et: elapsed time in seconds; RMSE: regression mean squared error for testing sample; MAE: MAE for testing sample; RMAE: RMAE for testing sample; NA: not available.

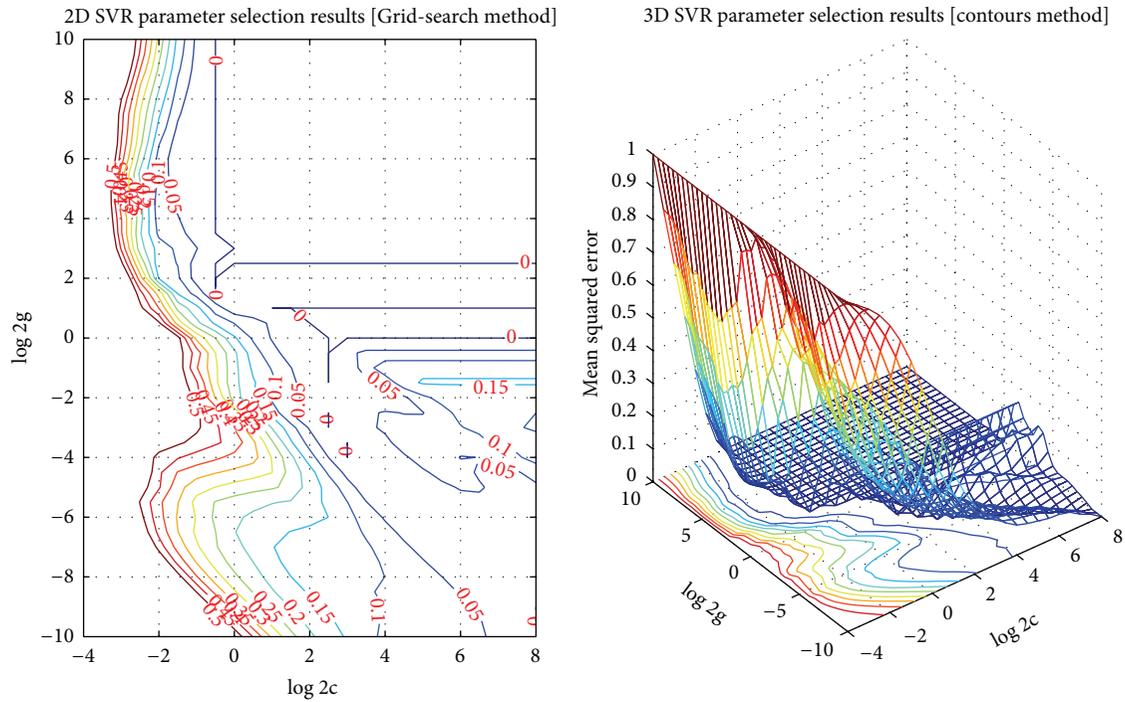


FIGURE 7: 2D and 3D parameter selection results. (Grid-search method and contours method, RBF).

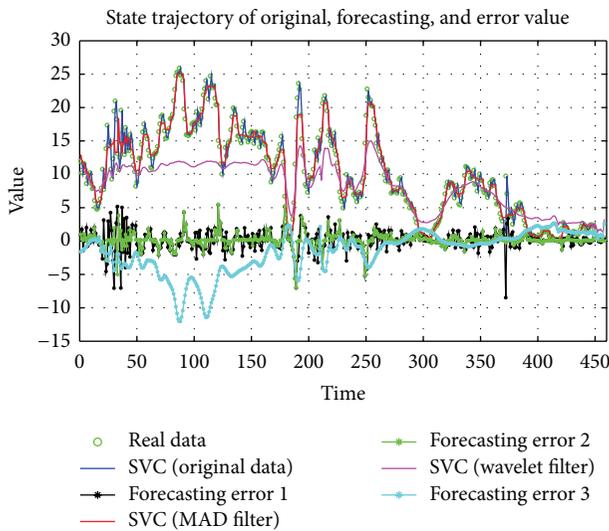


FIGURE 8: Final forecasting results via RBF.

5.5. *Error Analysis and Evaluation.* In this section, the error with respect to the forecasting results is given to evaluate the performance of multiscale SVR and MLP network of Section 5.4. Number of cross-validation for testing set is, respectively, set as 3, 6, and 10 for RBF and PF, moreover, the original data, filtered data via MAD, and wavelet are, respectively, utilized in the SVR with RBF and PF. Specifically, MLP neural network is applied to compare the performance

of SVR based on the RMSE, MAE, and RMAE which are defined in Section 4.7. In addition, the best cross-validation mean squared error, regression mean squared error, regression squared correlation coefficient, and best kernel function parameters are provided. The detailed information is given in Tables 3 and 4.

Based on Tables 3 and 4, we can learn that the performance of SVR is better than the tradition MLP neural network. Because the best performance of SVR employed MAD filter, therefore, data for MLP neural network are still filtered by MAD. In fact, SVR turns to be a robust time series forecasting method even with different parameters such as different number of cross-validation for testing set, different kernel function, and associated parameters. The simulation results denote that the SVR is an effective approach for STWPF.

6. Conclusions

In this paper, the multiscale SVR technique is applied for the short-term wind power forecasting. Firstly, we introduce a brief illustration for the main processing step and its corresponding theory analysis. Secondly, the data interpolation technique is used to fill the missing value of employed data. Thirdly, median absolute deviation filter and wavelet decomposition and denoising technique are applied to eliminate the irrelevant, noisy, and outlier value. Fourthly, Partial least squares technique, autocorrelation function, and cross-correlation function are, respectively, employed to the input

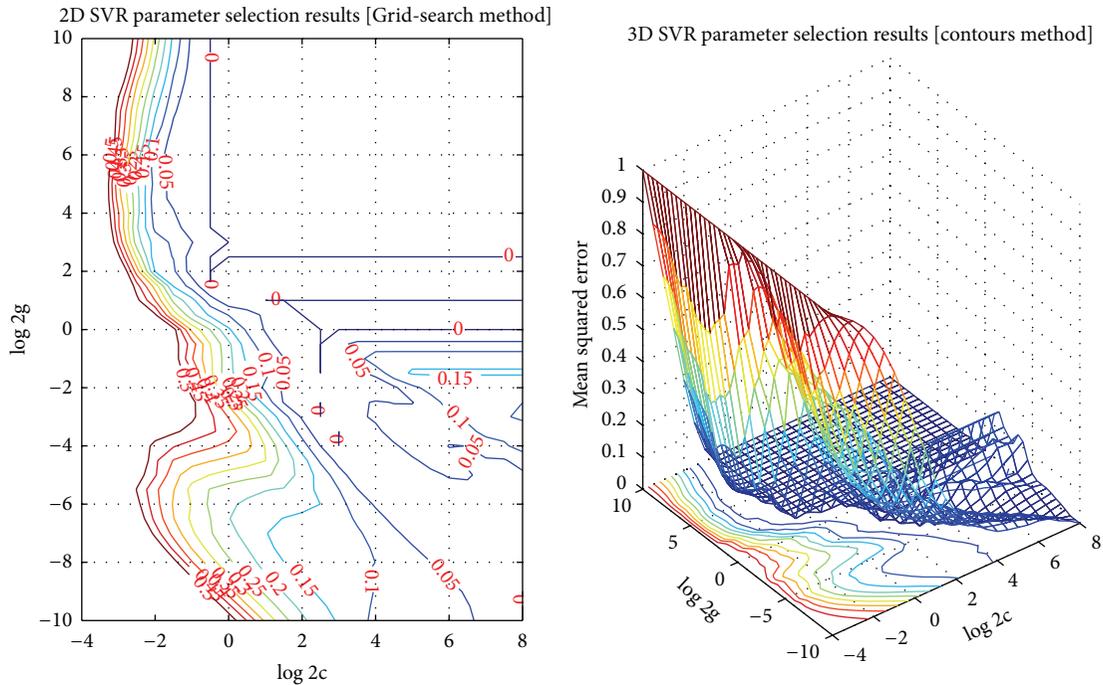


FIGURE 9: 2D and 3D parameter selection results. (Grid-search method and contours method, PF).

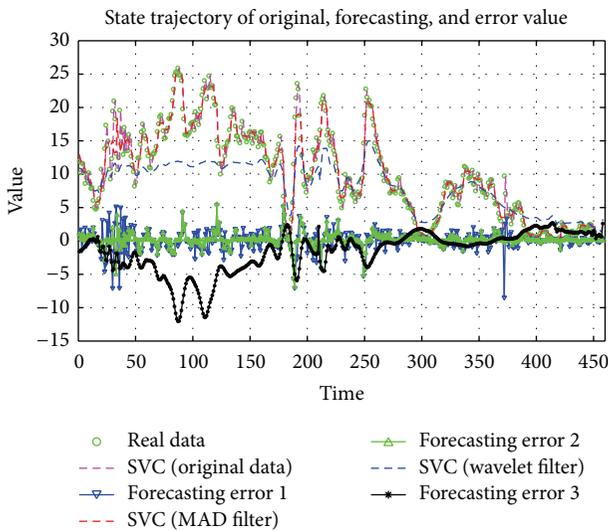


FIGURE 10: Final forecasting results via PF.

variable selection and order estimation. Fifthly, the multiscale SVR in combination with Grid-search technique is utilized to forecast the short-term wind power. Finally, the performance evaluation and error analysis are applied to evaluate the performance of multiscale SVR. Comparing to the multilayer perceptron (MLP) neural network, the performance demonstrates that SVR technique is a fast and robust time series forecasting approach. We believe that the proposed strategy

has reference value for short-term wind power forecasting and other energy consumption on the demand aspect.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors would like to thank the Professor Hongjie Jia and editor Dean who gave valuable comments by means of considerable suggestion and comments which improved highly the quality of this paper; the authors gratefully acknowledge Professor Xueliang Huang from the School of Electrical Engineering, Southeast University. This paper is partly supported by the National High Technology Research and Development Program (863 Program) (2011AA05A107).

References

- [1] M. A. Mohandes, T. O. Halawani, S. Rehman, and A. A. Hussain, "Support vector machines for wind speed prediction," *Renewable Energy*, vol. 29, no. 6, pp. 939–947, 2004.
- [2] O. Kramer and F. Gieseke, "Short-term wind energy forecasting using support vector regression," *Advances in Intelligent and Soft Computing*, vol. 87, pp. 271–280, 2011.
- [3] M. Milligan, K. Porter, E. DeMeo et al., "Wind power myths debunked," *IEEE Power and Energy Magazine*, vol. 7, no. 6, pp. 89–99, 2009.

- [4] S. Alessandrini, S. Sperati, and P. Pinson, "A comparison between the ECMWF and COSMO Ensemble Prediction Systems applied to short-term wind power forecasting on real data," *Applied Energy*, vol. 107, pp. 271–280, 2013.
- [5] F. Famili, W. M. Shen, R. Weber, and E. Simoudis, "Data pre-processing and intelligent data analysis," *Intelligent Data Analysis*, vol. 1, no. 1–4, pp. 3–23, 1997.
- [6] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Feitosa, "A review on the young history of the wind power short-term prediction," *Renewable and Sustainable Energy Reviews*, vol. 12, no. 6, pp. 1725–1744, 2008.
- [7] S. Osowski and K. Garanty, "Forecasting of the daily meteorological pollution using wavelets and support vector machine," *Engineering Applications of Artificial Intelligence*, vol. 20, no. 6, pp. 745–755, 2007.
- [8] B.-J. Chen, M.-W. Chang, and C.-J. Lin, "Load forecasting using support vector machines: a study on EUNITE competition 2001," *IEEE Transactions on Power Systems*, vol. 19, no. 4, pp. 1821–1830, 2004.
- [9] W.-Y. Chang, "An RBF neural network combined with OLS algorithm and genetic algorithm for short-term wind power forecasting," *Journal of Applied Mathematics*, vol. 2013, Article ID 971389, 9 pages, 2013.
- [10] Y. B. Che, W. Zhang, L. J. Ge, and J. J. Zhang, "A two-stage wind grid inverter with boost converter," *Journal of Applied Mathematics*, vol. 2014, Article ID 816564, 5 pages, 2014.
- [11] Y. D. Song, Q. Cao, X. Q. Du, and H. R. Karimi, "Control strategy based on wavelet transform and neural network for hybrid power system," *Journal of Applied Mathematics*, vol. 2013, Article ID 375840, 8 pages, 2013.
- [12] M. Lei, L. Shiyang, J. Chuanwen, L. Hongling, and Z. Yan, "A review on the forecasting of wind speed and generated power," *Renewable and Sustainable Energy Reviews*, vol. 13, no. 4, pp. 915–920, 2009.
- [13] T.-F. Li, W. Jia, W. Zhou, J.-K. Ge, Y.-C. Liu, and L.-Z. Yao, "Incomplete phase space reconstruction method based on subspace adaptive evolution approximation," *Journal of Applied Mathematics*, vol. 2013, Article ID 983051, 9 pages, 2013.
- [14] N. Do Hoai, K. Udo, and A. Mano, "Downscaling global weather forecast outputs using ANN for flood prediction," *Journal of Applied Mathematics*, vol. 2011, Article ID 246286, 14 pages, 2011.
- [15] I. Sánchez, "Short-term prediction of wind energy production," *International Journal of Forecasting*, vol. 22, no. 1, pp. 43–56, 2006.
- [16] A. M. Foley, P. G. Leahy, A. Marvuglia, and E. J. McKeogh, "Current methods and advances in forecasting of wind power generation," *Renewable Energy*, vol. 37, no. 1, pp. 1–8, 2012.
- [17] A. Botterud, J. Wang, V. Miranda, and R. J. Bessa, "Wind power forecasting in U.S. electricity markets," *Electricity Journal*, vol. 23, no. 3, pp. 71–82, 2010.
- [18] C. Stathopoulos, A. Kaperoni, G. Galanis, and G. Kallos, "Wind power prediction based on numerical and statistical models," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 112, pp. 25–38, 2013.
- [19] F. Famili and J. Ouyang, "Data mining: understanding data and disease modeling," in *Proceedings of the 21st IASTED International Multi-Conference on Applied Informatics*, 2003.
- [20] G. Noriega and S. Pasupathy, "Application of Kalman filtering to real-time preprocessing of geophysical data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 30, no. 5, pp. 897–910, 1992.
- [21] D. P. Allen, "A frequency domain Hampel filter for blind rejection of sinusoidal interference from electromyograms," *Journal of Neuroscience Methods*, vol. 177, no. 2, pp. 303–310, 2009.
- [22] L. Davies and U. Gather, "The identification of multiple outliers," *Journal of the American Statistical Association*, vol. 88, no. 423, pp. 782–792, 1993.
- [23] R. K. Pearson, "Outliers in process modeling and identification," *IEEE Transactions on Control Systems Technology*, vol. 10, no. 1, pp. 55–63, 2002.
- [24] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, Academic Press, 3rd edition, 2009.
- [25] R. Shalom, *Mixed Representation and Their Applications, Lecture Notes*, Technion-Israel Institute of Technology, 2011.
- [26] M. Steinbuch and M. J. G. Molengraft, "Wavelet theory and applications: a literature study," Tech. Rep., Eindhoven University of Technology, 2005.
- [27] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.
- [28] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.
- [29] P. P. Vaidyanathan, "Quadrature mirror filter banks, M-band extensions and perfect-reconstruction techniques," *IEEE ASSP Magazine*, vol. 4, no. 3, pp. 4–20, 1987.
- [30] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," Tech. Rep. 668, CIS, 1987.
- [31] J. Buckheit, S. Chen, D. Donoho, I. Johnstone, and J. D. Scargle, *Wavelab Reference Manual, Version 0. 700*, 1995.
- [32] G. E. Dallal, "Partial Correlation Coefficients," <http://www.jerrydallal.com/LHSP/partial.htm>.
- [33] H. Drucker, C. J. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support vector regression machines," *Advances in Neural Information Processing Systems*, vol. 9, pp. 155–161, 1997.
- [34] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.
- [35] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.
- [36] N. Cristianini and J. Shawe-Taylor, *An Introduction To Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, 2000.
- [37] F. Shi, X. Wang, L. Yu, and Y. Li, *30 Cases MATLAB Neural Network Analysis*, Beijing University of Aeronautics and Astronautics Press, Beijing, China, 2010, (Chinese).
- [38] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," version 3.17, 2013, <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
- [39] <http://statweb.stanford.edu/~tibs/sta306b/cvwrong.pdf>.
- [40] Y. Bengio and Y. Grandvalet, "No unbiased estimator of the variance of K-fold cross-validation," *Journal of Machine Learning Research*, vol. 5, pp. 1089–1105, 2004.

- [41] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Statistics Surveys*, vol. 4, pp. 40–79, 2010.
- [42] J. D. Rodriguez, A. Perez, and J. A. Lozano, "Sensitivity analysis of k-fold cross validation in prediction error estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 3, pp. 569–575, 2010.