Hybridizing DEMD and Quantum PSO with SVR in Electric Load Forecasting

Authors:

Li-Ling Peng, Guo-Feng Fan, Min-Liang Huang, Wei-Chiang Hong

Abstract:

Electric load forecasting is an important issue for a power utility, associated with the management of daily operations such as energy transfer scheduling, unit commitment, and load dispatch. Inspired by strong non-linear learning capability of support vector regression (SVR), this paper presents an SVR model hybridized with the differential empirical mode decomposition (DEMD) method and quantum particle swarm optimization algorithm (QPSO) for electric load forecasting. The DEMD method is employed to decompose the electric load to several detail parts associated with high frequencies (intrinsic mode function—IMF) and an approximate part associated with low frequencies. Hybridized with quantum theory to enhance particle searching performance, the so-called QPSO is used to optimize the parameters of SVR. The electric load data of the New South Wales (Sydney, Australia) market and the New York Independent System Operator (NYISO, New York, USA) are used for comparing the forecasting performances of different forecasting models. The results illustrate the validity of the idea that the proposed model can simultaneously provide forecasting with good accuracy and interpretability.

Record Type: Published Article

Submitted To: LAPSE (Living Archive for Process Systems Engineering)

Hybridizing DEMD and Quantum PSO with SVR in Electric Load Forecasting

Li-Ling Peng [1], Guo-Feng Fan [1], Min-Liang Huang [2] and Wei-Chiang Hong [3,4,*]

[1] College of Mathematics & Information Science, Ping Ding Shan University, Pingdingshan 467000, China; plling1054@163.com (L.-L.P.); guofengtongzhi@163.com (G.-F.F.)
[2] Department of Industrial Management, Oriental Institute of Technology, 58 Sec. 2, Sichuan Rd., Panchiao, New Taipei 220, Taiwan; minglianghuang2016@gmail.com
[3] School of Economics & Management, Nanjing Tech University, Nanjing 211800, China
[4] Department of Information Management, Oriental Institute of Technology, 58 Sec. 2, Sichuan Rd., Panchiao, New Taipei 220, Taiwan
[*] Correspondence: fi013@mail.oit.edu.tw; Tel.: +886-2-7738-0145 (ext. 5316); Fax: +886-277-386-310

Abstract: Electric load forecasting is an important issue for a power utility, associated with the management of daily operations such as energy transfer scheduling, unit commitment, and load dispatch. Inspired by strong non-linear learning capability of support vector regression (SVR), this paper presents an SVR model hybridized with the differential empirical mode decomposition (DEMD) method and quantum particle swarm optimization algorithm (QPSO) for electric load forecasting. The DEMD method is employed to decompose the electric load to several detail parts associated with high frequencies (intrinsic mode function—IMF) and an approximate part associated with low frequencies. Hybridized with quantum theory to enhance particle searching performance, the so-called QPSO is used to optimize the parameters of SVR. The electric load data of the New South Wales (Sydney, Australia) market and the New York Independent System Operator (NYISO, New York, USA) are used for comparing the forecasting performances of different forecasting models. The results illustrate the validity of the idea that the proposed model can simultaneously provide forecasting with good accuracy and interpretability.

Keywords: electric load forecasting; support vector regression; quantum theory; particle swarm optimization; differential empirical mode decomposition; auto regression

1. Introduction

Electric energy can not be reserved, thus, electric load forecasting plays a vital role in the daily operational management of a power utility, such as energy transfer scheduling, unit commitment, load dispatch, and so on. With the emergence of load management strategies, it is highly desirable to develop accurate, fast, simple, robust and interpretable load forecasting models for these electric utilities to achieve the purposes of higher reliability and better management [1].

In the past decades, researchers have proposed lots of methodologies to improve the load forecasting accuracy level. For example, Bianco *et al.* [2] proposed linear regression models for electricity consumption forecasting; Zhou *et al.* [3] applied a Grey prediction model for energy consumption; Afshar and Bigdeli [4] presented an improved singular spectral analysis method to predict short-term load in the Iranian power market; and Kumar and Jain [5] compared the forecasting performances among three Grey theory-based time series models to explore the consumption situation of conventional energy in India. Bianco *et al.* [6] indicate that their load model could be successfully used as an input of broader models than those of their previous paper [2]. References [7–10] proposed

several useful artificial neural networks models to conduct short-term load forecasting. The authors of [11–14] proposed hybrid models with evolutionary algorithms that demonstrated improved energy forecasting performances. These methods can achieve significant improvements in terms of forecasting accuracy, but without reasonable interpretability, particularly for ANN models. Artificial neural networks (ANNs), with mature nonlinear mapping capabilities and data processing characteristics, have achieved widely successful applications in load forecasting. Recently, expert systems with fuzzy rule-based linguistic means provided good interpretability while dealing with system modeling [15]. Various approaches and models have been proposed in the last decades in many area such as climate factors (temperature and humidity), social activities (human social activities), seasonal factors (seasonal climate change and load growth), and so on. However, these models have strong dependency on an expert and lack expected forecasting accuracy. Therefore, combination models which are based on these popular methods and other techniques can satisfy the two desired requests: high accuracy level and interpretability.

With superiority in handling high dimension nonlinear data, support vector regression (SVR) has been successfully used to solve forecasting problems in many fields, such as financial time series (stocks index and exchange rate) forecasting, tourist arrival forecasting, atmospheric science forecasting, and so on [16–18]. However, SVR methods have a significant disadvantage, in that while its three parameters are determined simultaneously during the nonlinear optimization process, the solution is easily trapped into a local optimum. In addition, it also lacks a statistically significant level of robustness. These two shortcomings are the focused topics in the SVR research field [19]. On the other hand, the empirical mode decomposition (EMD) with auto regression (AR), a reliable clustering algorithm, has been successfully used in many fields [20–22]. The EMD method is particularly powerful for extracting the components of the basic mode from nonlinear or non-stationary time series [23], *i.e.*, the original complex time series can be transferred into a series of single and apparent components. However, this method cannot deal well with the signal decomposition effects while the gradient of the time series is fluctuating. Based on the empirical decomposition mode, reference [24] proposes the differential empirical mode decomposition (DEMD) to improve the fluctuating changes problem of the original EMD method. The derived signal is obtained by several derivations of the original signal, and the fluctuating gradient is thus eliminated, so that the signal can satisfy the conditions of EMD. The new signal is then integrated into EMD to obtain each intrinsic mode function (IMF) order and the residual amount of the original signal. The differential EMD method is employed to decompose the electric load into several detailed parts with higher frequency IMF and an approximate part with lower frequencies. This can effectively reduce the unnecessary interactions among singular values and can improve the performance when a single kernel function is used in forecasting. Therefore, it is beneficial to apply a suitable kernel function to conduct time series forecasting [25]. Since 1995, many attempts have been made to improve the performance of the PSO [26–31]. Sun *et al.* [32,33] introduced quantum theory into PSO and proposed a quantum-behaved PSO (QPSO) algorithm, which is a global search algorithm to theoretically guarantee finding good optimal solutions in the search space. Compared with PSO, the iterative equation of QPSO needs no velocity vectors for particles, has fewer parameters to adjust, and can be implemented more easily. The results of experiments on widely used benchmark functions indicate that the QPSO is a promising algorithm [32,33] that exhibits better performance than the standard PSO.

In this paper, we present a new hybrid model to achieve satisfactory forecasting accuracy. The principal idea is hybridizing DEMD with QPSO, SVR and AR, namely the DEMD-QPSO-SVR-AR model, to achieve better forecasting performance. The outline of the proposed DEMD-QPSO-SVR-AR model is as follows: (1) the raw data can be divided into two parts by DEMD technology, one is the higher frequency item, the other is the residuals; (2) the higher frequency item has less redundant information than the raw data and trend information, because that information is gone to the residuals, then, QPSO is applied to optimize the parameters of SVR (*i.e.*, the so-called QPSO-SVR model), so the QPSO-SVR model is used to forecast the higher frequency, the accuracy is higher than the original SVR

model, particularly around the peak value; (3) fortunately, the residuals is monotonous and stationary, so the AR model is appropriate for forecasting the residuals; (4) the forecasting results are obtained from Steps (2) and (3). The proposed DEMD-QPSO-SVR-AR model has the capability of smoothing and reducing the noise (inherited from DEMD), the capability of filtering dataset and improving forecasting performance (inherited from SVR), and the capability of effectively forecasting the future tendencies (inherited from AR). The forecast outputs obtained by using the proposed hybrid method are described in the following sections.

To show the applicability and superiority of the proposed model, half-hourly electric load data (48 data points per day) from New South Wales (Australia) with two kind of sizes are used to compare the forecasting performances among the proposed model and other four alternative models, namely the PSO-BP model (BP neural network trained by the PSO algorithm), SVR model, PSO-SVR model (optimizing SVR parameters by the PSO algorithm), and the AFCM model (adaptive fuzzy combination model based on a self-organizing mapping and SVR). Secondly, another hourly electric load dataset (24 data points per day) from the New York Independent System Operator (NYISO, USA), also, with two kinds of sizes are used to further compare the forecasting performances of the proposed model with other three alternative models, namely the ARIMA model, BPNN model (artificial neural network trained by a back-propagation algorithm), and GA-ANN model (artificial neural network trained by a genetic algorithm). The experimental results indicate that this proposed DEMD-QPSO-SVR-AR model has the following advantages: (1) it simultaneously satisfies the need for high levels of accuracy and interpretability; (2) the proposed model can tolerate more redundant information than the SVR model, thus, it has more powerful generalization ability.

The rest of this paper is organized as follows: in Section 2, the DEMD-QPSO-SVR-AR forecasting model is introduced and the detailed illustrations of the model are also provided. In Section 3, the data description and the research design are illustrated. The numerical results and comparisons are shown in Section 4. The conclusions of this paper and the future research focuses are given in Section 5.

2. Support Vector Regression with Differential Empirical Mode Decomposition

2.1. Differential Empirical Mode Decomposition (DEMD)

The EMD method assumes that any signal consists of different simple intrinsic modes of oscillation. Each linear or non-linear mode will have the same number of extreme and zero-crossings. There is only one extreme between successive zero-crossings. In this way, each signal could be decomposed into a number of intrinsic mode functions (IMFs). With the definition, any signal $x(t)$ can be decomposed, and the corresponding flow chart is shown as Figure 1:

(1)　　Identify all local extremes.
(2)　　Repeat the procedure for the local minima to produce the lower envelope m_1.
(3)　　The difference between the signal $x(t)$ and m_1 is the first component, h_1, as shown in Equation (1):

$$h_1 = x(t) - m_1 \tag{1}$$

In general, h_1 is unnecessary to satisfy the conditions of the IMF, because h_1 is not a standard IMF, and until the mean envelope approximates zero it should be determined k times. At this point, the data could be as shown in Equation (2):

$$h_{1k} = h_{1(k-1)} - m_{1k} \tag{2}$$

where h_{1k} is the datum after k siftings. $h_{1(k-1)}$ stands for the data after shifting $k-1$ times. Standard deviation (*SD*) is defined by Equation (3):

$$SD = \sum_{k=1}^{T} \frac{\left| h_{1(k-1)}(t) - h_{1k}(t) \right|^2}{h_{1(k-1)}^2(t)} \in (0.2, 0.3) \tag{3}$$

where T is the length of the data.

(4) When h_{1k} has met the basic conditions of *SD*, based on the condition of $c_1 = h_{1k}$, and a new series r_1 could be presented as Equation (4):

$$r_1 = x_1(t) - c_1 \tag{4}$$

(5) Repeat previous steps 1 to 4 until the r_n cannot be decomposed into the IMF. The sequence r_n is called the remainder of the original data $x(t)$ as Equations (5) and (6):

$$r_1 = x_1(t) - c_1, r_2 = r_1 - c_2, ..., r_n = r_{n-1} - c_n \tag{5}$$

$$x_1(t) = \sum_{i=1}^{n} c_i + r_n \tag{6}$$

Finally, the differential EMD is proposed by Equation (7):

$$DEMD = x_n(t) - c_0(t) \tag{7}$$

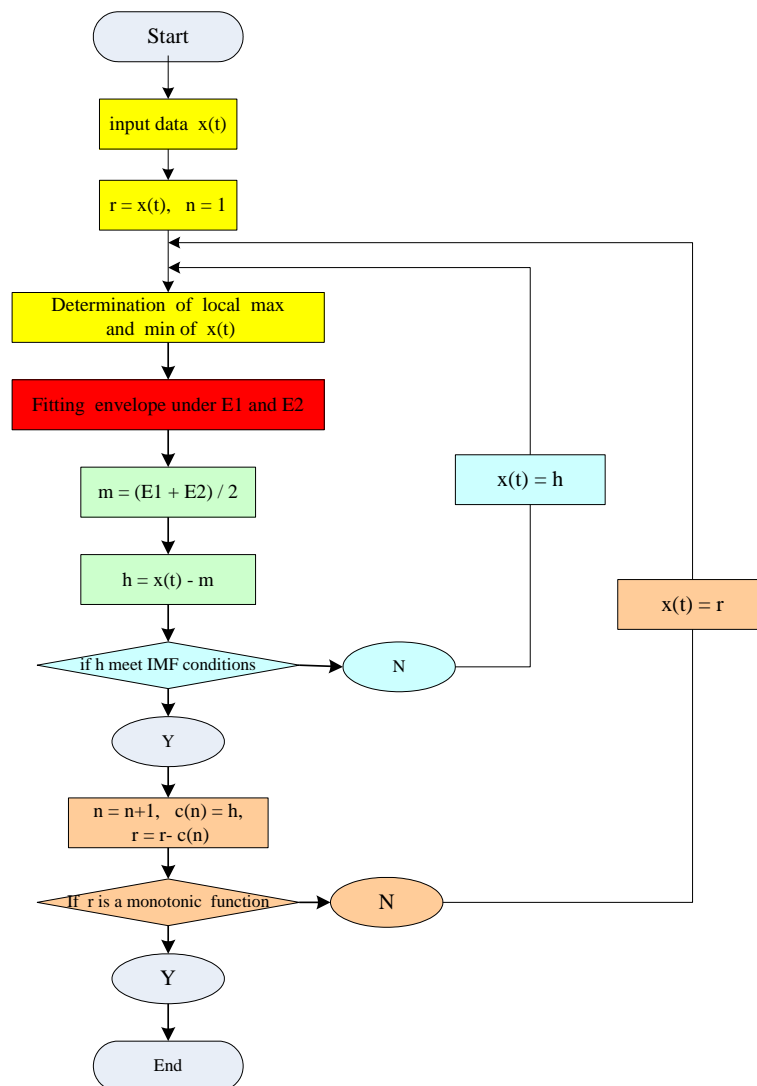where $x_n(t)$ refers to dependent variables.

Figure 1. Differential EMD algorithm flowchart.

2.2. Support Vector Regression

The notion of an SVR model is briefly introduced. Given a data set with N elements $\{(X_i, y_i), i = 1, 2, ..., N\}$, where X_i is the i-th element in n-dimensional space, *i.e.*, $X_i = [x_{1i}, ... x_{ni}] \in \Re^n$, and $y_i \in \Re$ is the actual value corresponding to X_i. A non-linear mapping $(\cdot) : \Re^n \rightarrow \Re^{n_h}$ is defined to map the training (input) data X_i into the so-called high dimensional feature space (which may have infinite dimensions), $\Re^{n_h}$. Then, in the high dimensional feature space, there theoretically exists a linear function, f, to formulate the non-linear relationship between input data and output data. Such a linear function, namely SVR function, is shown as Equation (8): align equations

$$f(X) = W^T \varphi(X) + b \tag{8}$$

where $f(X)$ denotes the forecasting values; the coefficients W ($W \in \Re^{n_h}$) and b ($b \in \Re$) are adjustable. As mentioned above, the SVM method aims at minimizing the empirical risk, shown as Equation (9):

$$R_{emp}(f) = \frac{1}{N} \sum_{i=1}^{N} \Theta_\varepsilon(y_i, W^T \phi(X_i) + b) \tag{9}$$

where $\Theta_\varepsilon(y_i, f(X))$ is the ε-insensitive loss function and defined as Equation (10):

$$\Theta_\varepsilon(y_i, f(X)) = \begin{cases} |f(X) - y| - \varepsilon & , \quad if \ |f(X) - y| \geqslant \varepsilon \\ 0 & , \quad\quad otherwise \end{cases} \tag{10}$$

In addition, $\Theta_\varepsilon(y_i, f(X))$ is employed to find out an optimum hyper-plane on the high dimensional feature space to maximize the distance separating the training data into two subsets. Thus, the SVR focuses on finding the optimum hyperplane and minimizing the training error between the training data and the ε-insensitive loss function. Then, the SVR minimizes the overall errors, shown as Equation (11):

$$\underset{W,b,\xi^*,\xi}{Min} R_\varepsilon(W, \xi^*, \xi) = \frac{1}{2} W^T W + C \sum_{i=1}^{N} (\xi_i^* + \xi_i) \tag{11}$$

with the constraints:

$$\begin{aligned} y_i - W^T \phi(X_i) - b &\leqslant \varepsilon + \xi_i^* \\ -y_i + W^T \phi(X_i) + b &\leqslant \varepsilon + \xi_i \\ \xi_i^*, \xi_i &\geqslant 0 \\ i &= 1, 2, ..., N \end{aligned} \tag{12}$$

The first term of Equation (11), employing the concept of maximizing the distance of two separated training data, is used to regularize weight sizes to penalize large weights, and to maintain regression function flatness. The second term penalizes training errors of $f(x)$ and y by using the ε-insensitive loss function. C is the parameter to trade off these two terms. Training errors above ε are denoted as ξ_i^*, whereas training errors below $-\varepsilon$ are denoted as ξ_i.

After the quadratic optimization problem with inequality constraints is solved, the parameter vector w in Equation (8) is obtained as Equation (13):

$$W = \sum_{i=1}^{N} (\beta_i^* - \beta_i) \phi(X_i) \tag{13}$$

where β_i^*, ξ_i are obtained by solving a quadratic program and are the Lagrangian multipliers. Finally, the SVR regression function is obtained as Equation (14) in the dual space:

$$f(X) = \sum_{i=1}^{N} (\beta_i^* - \beta_i) K(X_i, X) + b \tag{14}$$

where $K(X_i, X)$ is called the kernel function, and the value of the kernel equals the inner product of two vectors, X_i and X_j, in the feature space $\varphi(X_i)$ and $\varphi(X_j)$, respectively; that is, $K(X_i, X_j) = \varphi(X_i)\varphi(X_j)$. Any function that meets Mercer's condition [34] can be used as the kernel function.

There are several types of kernel function. The most used kernel functions are the Gaussian radial basis functions (RBF) with a width of σ: $K(X)_i, X_j) = \exp\left(-0.5||X_i - X_j||^2/\sigma^2\right)$ and the polynomial kernel with an order of d and constants a_1 and a_2: $K(X_i, X_j) = (a_1 X_i + a_2 X_j)^d$. However, the Gaussian RBF kernel is not only easy to implement, but also capable of non-linearly mapping the training data into an infinite dimensional space, thus, it is suitable to deal with non-linear relationship problems. Therefore, the Gaussian RBF kernel function is specified in this study.

2.3. Particle Swarm Optimization Algorithm

PSO is a heuristic global optimization algorithm broadly applied in optimization problems. PSO is developed on a very simple theoretical framework that is easily implemented with only primitive mathematical operators [26]. In PSO, a group of particles is composed of m particles in D dimension space where the position of the particle i is $X_i = (x_{i1}, x_{i2}, \ldots, x_{iD})$ and the speed is $V_i = (v_{i1}, v_{i2}, \ldots, v_{iD})$. The speed and position of each particle are changed in accordance with the following equations, Equations (15) and (16):

$$v_{id}^{j+1} = wv_{id}^j + c_1 r_1(p_{id}^j - x_{id}^j) + c_2 r_2(p_{gd}^j - x_{id}^j) \tag{15}$$

$$x_{id}^{j+1} = x_{id}^j + v_{id}^{j+1} \tag{16}$$

where $i = 1, 2, \ldots, m$; $d = 1, 2, \ldots, D$; m is the particle size; p_{id}^j is the dth dimension component of the *pbest* that is the individual optimal location of the particle i in the jth iteration; p_{gd}^j is the dth dimension component of the *gbest* that is the optimal position of all particles in the jth iteration; w is the inertia weight coefficient; c_1 and c_2 are learning factors; r_1 and r_2 are random numbers in the range [0,1].

The inertia weight w, which balances the global and local exploitation abilities of the swarm, is critical for the performance of PSO. A large inertia weight facilitates exploration but slows down particle convergence. Conversely, a small inertia weight facilitates fast particle convergence it sometimes leads to the local optimal. The most popular algorithm for controlling inertia weight is linearly decreasing inertia weight PSO [31]. The strategy of linearly decreasing inertia weight is widely used to improve the performance of PSO, but this approach has a number of drawbacks [27]. Several adaptive algorithms for tuning inertia weight have been presented [27–30]. In the present work, we propose the method of nonlinearly decreasing inertia weight to tune the value of w for further performance improvement as Equation (17):

$$w = w_{\max} - \frac{(w_{\max} - w_{\min}) \times (t-1)^2}{(t_{\max} - 1)^2} \tag{17}$$

where $w_{\max}$ and $w_{\min}$ are the maximum and minimum values of w, respectively; t is the current iteration number; and $t_{\max}$ is the maximum iteration number.

2.4. Quantum-Behaved Particle Swarm Optimization

The main disadvantage of the PSO algorithm is that global convergence is not guaranteed [35]. To address this problem, Sun *et al.* [32,33], inspired by the trajectory analysis of PSO and quantum mechanics, developed and proposed the QPSO algorithm. Particles move according to the following iterative equations, Equations (18) to (21), and the flow chart is shown as Figure 2.

$$x_{ij}(t+1) = p_{ij}(t) + \alpha \left| mbest_j(t) - x_{ij}(t) \right| \times \ln(1/u) \qquad if \ k \geqslant 0.5 \tag{18}$$

$$x_{ij}(t+1) = p_{ij}(t) - \alpha \left| mbest_j(t) - x_{ij}(t) \right| \times \ln(1/u) \qquad if \ k \leqslant 0.5 \tag{19}$$

$$mbest_j(t) = \frac{1}{M}\sum_{i=1}^{M} pbest_{ij}(t) \tag{20}$$

$$p_{ij}(t) = \phi_{ij}(t)pbest_{ij}(t) + (1 - \phi_{ij}(t))gbest_j(t) \tag{21}$$

where *mbest* is the mean best position defined as the mean of all the *pbest* positions of the population; k, u and φu are random numbers generated using a uniform probability distribution in the range [0,1]. The parameter α is called the contraction expansion coefficient, which is the only parameter in the QPSO algorithm that can be tuned to control the convergence speed of particles. In general, this parameter can be controlled by two methods: (1) fixing; or (2) varying the value of α during the search of the algorithm. In [36], setting α to a number in the range (0.5, 0.8) generates satisfied results for most benchmark functions. However, fixing the value of α is sensitive to population size and the maximum number of iterations. This problem can be overcome by using a time-varying CE coefficient. The literatures on QPSO suggest that decreasing the value of α linearly from α_1 to α_0 ($\alpha_0 < \alpha_1$) in the course of the search process makes the QPSO algorithm perform efficiently [36,37].
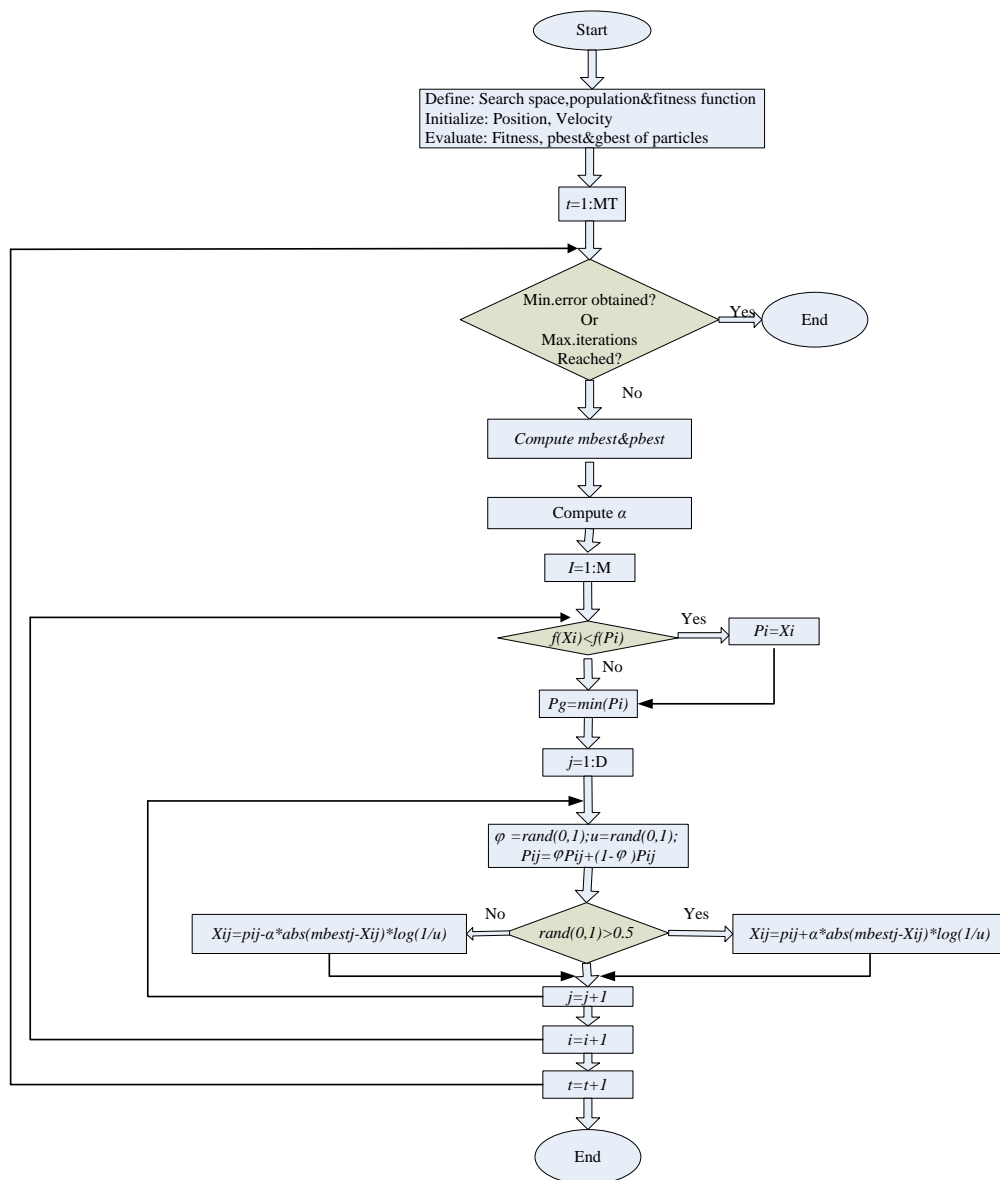

Figure 2. QPSO algorithm flowchart.

In this paper, the value of α is computed as Equation (22):

$$\alpha = \alpha_1 - \frac{(t-1) \times (\alpha_1 - \alpha_0)}{T-1} \tag{22}$$

where α_1 and α_0 are the final and initial values of α, respectively; t is the current iteration number; and T is the maximum iteration number. Previous studies on QPSO [36,37] recommend that α be linearly decreased from 1.0 to 0.5 for the algorithm to attain a generally good performance.

QPSO has already been implemented with excellent results [32] in various standard optimization problems. Moreover, the QPSO algorithm has been proven more effective than traditional algorithms in most cases [38–42]. In the current work, QPSO algorithm is utilized in SVR parameter optimization for forecasting the high frequency data, and its performance is compared with that of the classical PSO algorithm [43,44].

2.5. AR Model

Equation (23) expresses a *p*-step autoregressive model, referring as *AR(p)* model [45]. Stationary time series $\{X_t\}$ that meet the model *AR(p)* is called the *AR(p)* sequence. That $a = (a_1, a_2, \ldots, a_p)^{\mathrm{T}}$ is named as the regression coefficients of the *AR(p)* model:

$$X_t = \sum_{j=1}^{p} a_j X_{t-j} + \varepsilon_t \tag{23}$$
$$t \in Z$$

2.6. The Full Procedure of DEMD-QPSO-SVR-AR Model

The full procedure of the proposed DEMD-QPSO-SVR-AR model is briefed as follows and is illustrated in Figure 3.
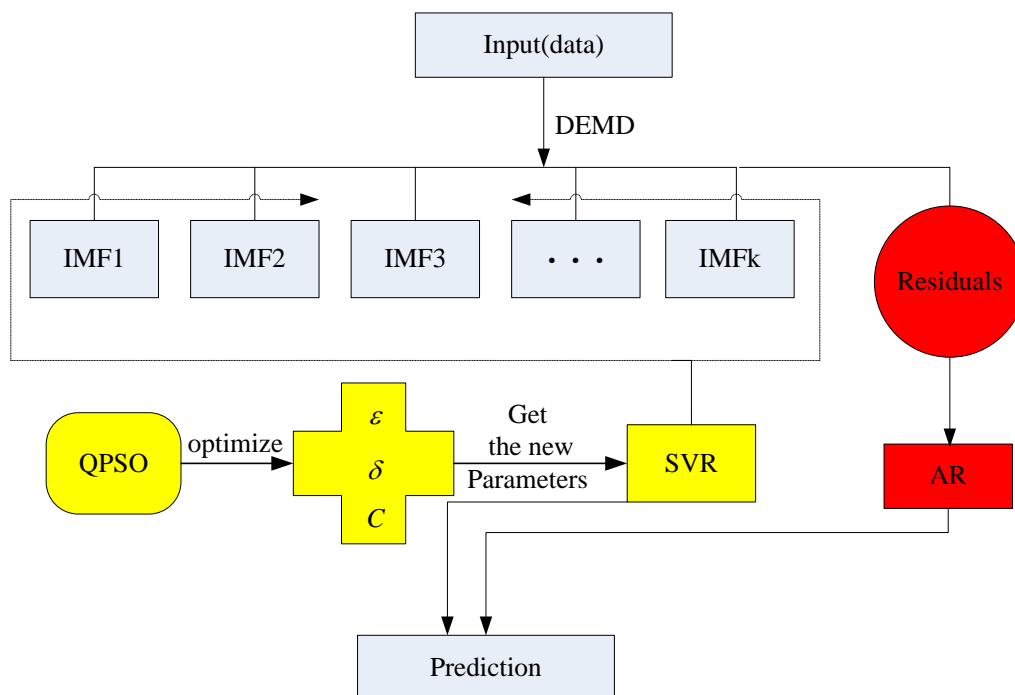

Figure 3. The full flowchart of the DEMD-QPSO-SVR-AR model flowchart.

Step 1: Decompose the input data by DEMD. Each electric load data (input data) could be decomposed into a number of intrinsic mode functions (IMFs), *i.e.*, two parts, one is the higher frequency item,

the other one is the residuals. Please refer Section 2.1 and Figure 1 to learn more about the DEMD process details.

Step 2: QPSO-SVR modeling. The SVR model is employed to forecast the high frequency item, thus, to look for most suitable parameters by QPSO, different sizes of fed-in/fed-out subsets will be set in this stage. Please refer Section 2.2 to learn in more details the SVR process. The QPSO algorithm is utilized in SVR parameter optimization for forecasting the higher frequency data, please refer Sections 2.3 and 2.4 and Figure 2 to learn more about the details of the QPSO process.

Step 3: AR modeling. The residuals item is forecasted by the AR model due to its monotonous and stationary nature. Please refer Section 2.5 to learn in more detail the processes of AR modeling. Similarly, while the new parameters have smaller MAPE values or maximum iteration is reached, the new three parameters and the corresponding objective value are the solution at this stage.

Step 4: DEMD-QPSO-SVR-AR forecasting. After receiving the forecasting values of the high frequency item and the residuals item from SVR model and AR model, respectively, the final forecasting results would be eventually obtained from the high frequency item and the residuals.

3. Numerical Examples

To illustrate the superiority of the proposed model, we use two datasets from different electricity markets, that is, the New South Wales (NSW) market in Australia (denoted as Case 1) and the New York Independent System Operator (NYISO) in the USA (Case 2). In addition, for each case, we all use two sample sizes, called small sample and large sample, respectively.

3.1. The Experimental Results of Case 1

For Case 1, firstly, electric load data obtained from 2 to 7 May 2007 is used as the training data set in the modeling process, and the testing data set is from 8 May 2007. The electric load data used are all based on a half-hourly basis (*i.e.*, 48 data points per day). The dataset containing only 7 days is called the small size sample in this paper.

Secondly, for large training sets, it should avoid overtraining during the SVR modeling process. Therefore, the second data size has 23 days (1104 data points from 2 to 24 May 2007) by employing all of the training samples as training set, *i.e.*, from 2 to 17 May 2007, and the testing data set is from 18 to 24 May 2007. This example is called the large sample size data in this paper.

(i) Results after DEMD in Case 1

As mentioned in the authors' previous paper [25], the results of the decomposition process by DEMD, can be divided into the higher frequency item (Data-I) and the residuals term (Data-II). The trend of the higher frequency item is the same as that of the original data, and the structure is more regular and stable. Thus, Data-I and Data-II both have good regression effects by the QPSO-SVR and AR, respectively.

(ii) Forecasting Using QPSO-SVR for Data-I (The Higher Frequency Item in Case 1)

After employing DEMD to reduce the non-stationarity of the data set in Case 1, QPSO with SVR can be successfully applied to reduce the performance volatility of SVR with different parameters, to perform the parameter determination in SVR modeling process.

The higher frequency item is simultaneously employed for QPSO-SVR modeling, and the better performances of the training and testing (forecasting) sets are shown in Figure 4a,b, respectively. This implies that the decomposition and optimization by QPSO is helpful to improve the forecasting accuracy. The parameters of a QPSO-SVR model for Data-I are shown in Tables 1 and 2 in which the forecasting error for the higher frequency decomposed by the DEMD and QPSO-SVR has been reduced.
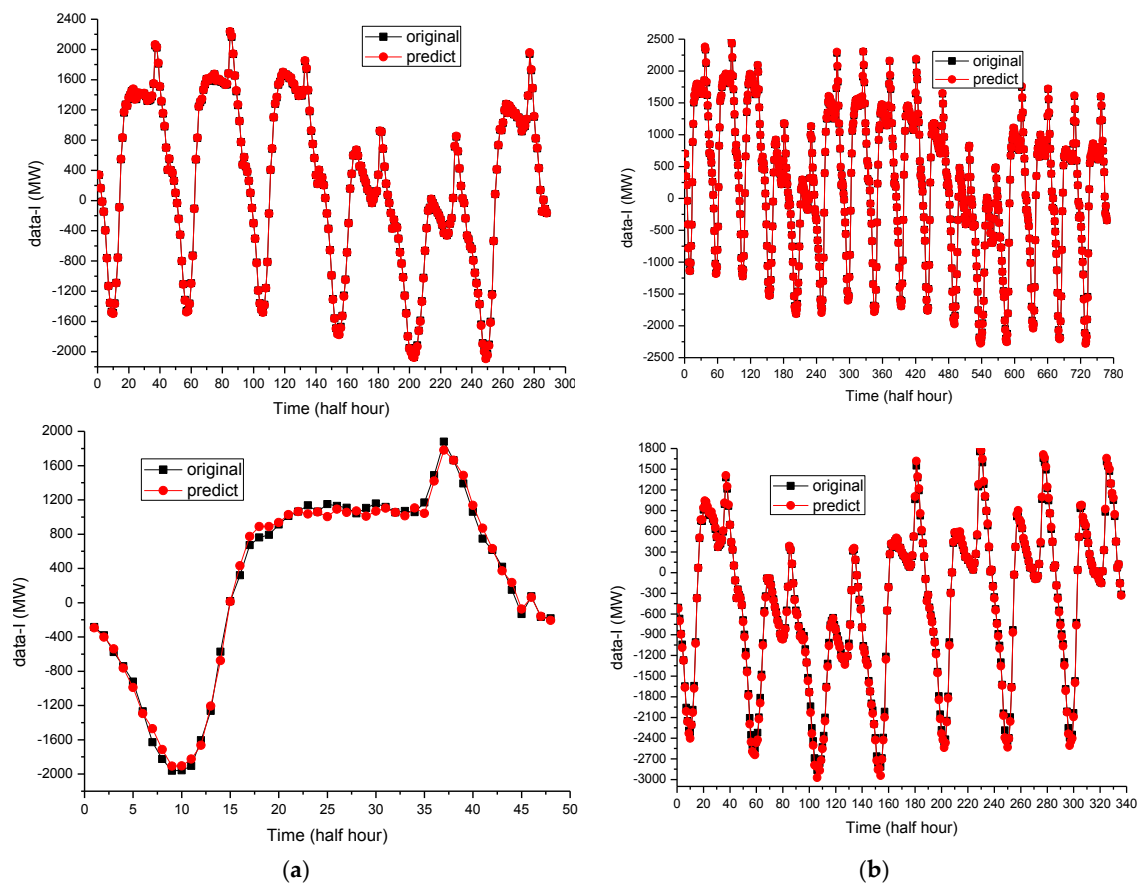
Figure 4. Comparison the forecasted electric load of train and test by the QPSO-SVR model for the data-I of sample data: (**a**) One-day ahead prediction of May 8, 2007 are performed by the model; (**b**) One-week ahead prediction from May 18, 2007 May 24, 2007 are performed by the model.

Table 1. The QPSO's parameters for SVR in Data-I.

N	C_{min}	C_{max}	σ_{min}	σ_{max}	*itmax*
30	0	200	0	200	300

Table 2. The SVR's parameters for Data-I.

Sample Size	m	σ	C	ε	Testing MAPE
The samall sample data	20	0.12	88	0.0027	9.13
The large sample data	20	0.19	107	0.0011	4.1

where *N* is number of particles, C_{min} is the minimum of C, C_{max} is the maximum of C, σ_{min} is the minimum of σ, σ_{max} is the maximum of σ, *itmax* is maximum iteration number.

(iii) Forecasting Using AR for Data-II (The Residuals in Case 1)

As mentioned in the authors' previous paper [25], the residuals are linear locally and stable, so use of the AR technique to predict Data-II is feasible. Based on the geometric decay of the correlation analysis for Data-II (the residuals), it can be denoted as the AR(4) model. The associated parameters of the AR(4) model for Data-II are indicated in Table 3. The errors almost approach the level of 10^{-5} both for the small or large amounts of data, *i.e.*, the forecasting error for Data-II by DEMD has been significantly reduced. This shows the superiority of the AR model.

Table 3. Summary of results of the AR forecasting model for data-II.

Residuals	MAE	Equation
The small sample size	9.7732×10^{-5}	$x_n = 5523.827 + 1.23x_{n-1} + 0.5726x_{n-2} + 0.0031x_{n-3} - 0.70465x_{n-4}$
The large sample size	7.5923×10^{-5}	$x_n = 5524.9 + 1.0152x_{n-1} + 0.3628x_{n-2} + 0.0019x_{n-3} - 0.6752x_{n-4}$

3.2. The Experimental Results of Case 2

For Case 2, electric load data obtained from 1 to 12 January 2015 is used as the training data set in the modeling process, and the testing data set is from 13 to 14 January 2015. These employed electric load data are all based on an hour basis (*i.e.*, 24 data points per day). The dataset contains only 14 days so it is also called the small sample in this paper.

Secondly, for large training sets, the second dataset size is 46 days (1104 data points from 1 January to 15 February 2015) by employing all of the training samples as training set, *i.e.*, from 1 January to 1 February 2015, and the testing dataset is from 2 to 15 February 2015. This example is also called the large size sample data in this paper.

(i) Results after DEMD in Case 2

As mentioned in the authors' previous paper [25], similarly, the data results of the decomposition process by DEMD can be divided into the higher frequency item (Data-I) and the residuals term (Data-II). The trend of the higher frequency item is also the same as that of the original data, and the structure is also regular and stable. Thus, Data-I and Data-II both have good regression effects by the QPSO-SVR and AR, respectively.

(ii) Forecasting Using QPSO-SVR for Data-I (The Higher Frequency Item in Case 2)

After employing DEMD to reduce the non-stationarity of the data set in Case 2, to further resolve these complex nonlinear, chaotic problems for both small sample and large sample data, the QPSO with SVR can be successfully applied to reduce the performance volatility of SVR with different parameters to perform the parameter determination in the SVR modeling process, to improve the forecasting accuracy. The higher frequency item is simultaneously employed for QPSO-SVR modeling, and the better performances of the training and testing (forecasting) sets are shown in Figure 5a,b, respectively. This implies that the decomposition and optimization by QPSO is helpful to improve the forecasting accuracy. The parameters of a QPSO-SVR model for Data-I are shown in Tables 1 and 4 in which the forecasting error for the higher frequency decomposed by the DEMD and QPSO-SVR has been reduced.

Table 4. The SVR's parameters for data-I in Case 2.

Sample Size	m	σ	C	ε	Testing MAPE
The small data	24	0.10	102	0.0029	7.19
The large data	24	0.19	113	0.0011	4.62

(iii) Forecasting Using AR for Data-II (The Residuals in Case 2)

As mentioned in the authors' previous paper [25], the residuals are linear locally and stable, so the AR technique is feasible to predict Data-II. Based on the geometric decay of the correlation analysis for Data-II (the residuals), that can also be denoted as the AR(4) model, the associated parameters of the AR(4) model for Data-II are indicated in Table 5. The errors almost approach a level of 10^{-5} both for the small or large amount of data, *i.e.*, the forecasting error for Data-II by DEMD has significantly reduced. This shows the superiority of the AR model.
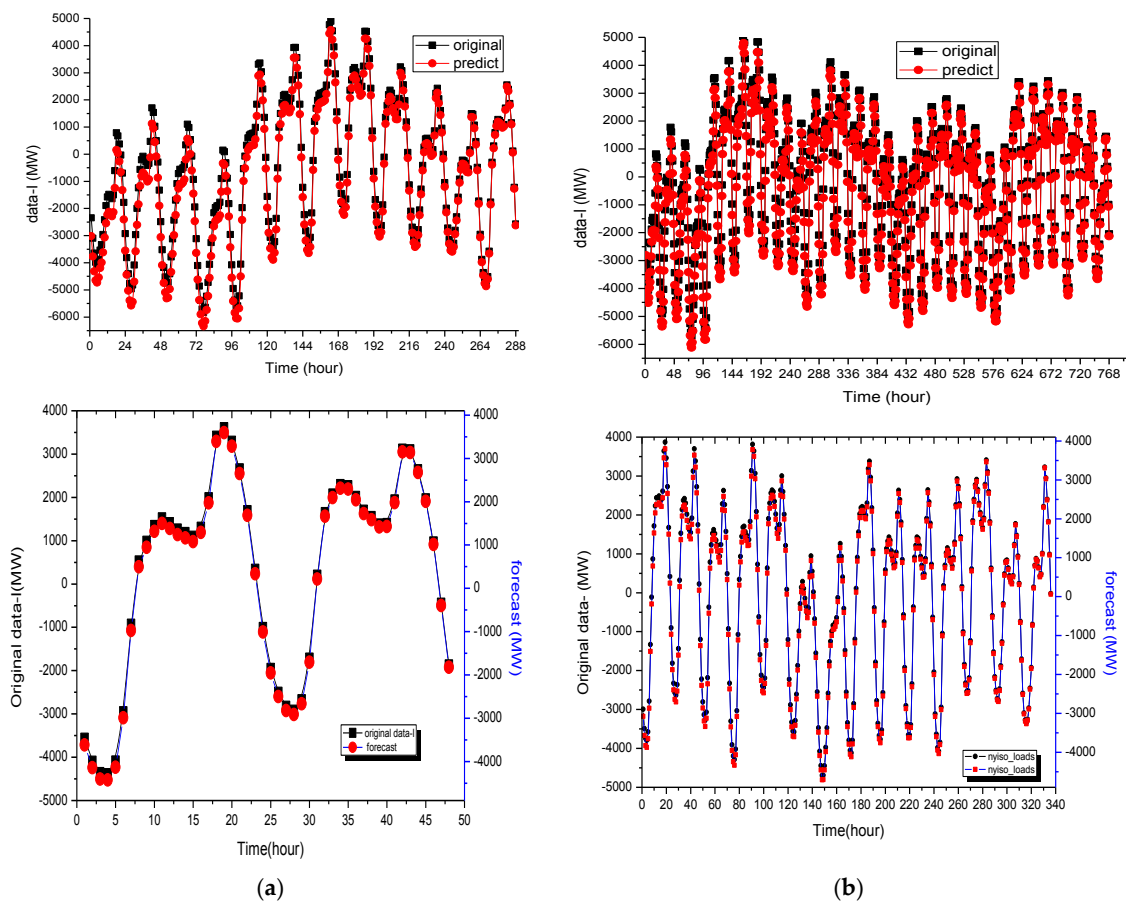
Figure 5. Comparison the forecasted electric load of training and test data by the QPSO-SVR model for the data-I of sample data in Case 2: (**a**) One-day ahead prediction from 13 to 14 January 2015 are performed by the model; (**b**) One-week ahead prediction from 2 to 15 February 2015 are performed by the model.

Table 5. Summary of results of the AR forecasting model for Data-II in Case 2.

Residuals	MAE	Equation
The small sample size	9.138×10^{-5}	$x_n = 5521.7 + 1.13x_{n-1} + 0.5676x_{n-2} + 0.021x_{n-3} - 0.845x_{n-4}$
The large sample size	6.02×10^{-5}	$x_n = 5522.7 + 0.9152x_{n-1} + 0.3978x_{n-2} + 0.0049x_{n-3} - 0.52x_{n-4}$

where x_n is the n-th electric load residual, x_{n-1} is the $(n-1)$th electric load residual similarly, *etc.*

4. Results and Analysis

This section illustrates the performance of the proposed DEMD-QPSO-SVR-AR model in terms of forecasting accuracy and interpretability. Taking into account the superiority of an SVR model for small sample size and superiority comparisons, a real case analysis with small sample size is used in the first case. The next case with 1104 data points is devoted to illustrate the relationships between two sample sizes (large size and small size) and accurate levels in forecasting.

4.1. Setting Parameters for the Proposed Forecasting Models

As indicated by Taylor [46], and according to the same conditions of the comparison with Che *et al.* [47], the settings of several parameters in the proposed forecasting models are illustrated as

follows. For the PSO-BP model, 90% of the collected samples is used to train the model, and the rest (10%) is employed to test the performance. In the PSO-BP model, these used parameters are set as follows: (i) for the BP neural network, the input layer dimension (*indim*) is set as 2; the dimension of the hidden layer (*hiddennum*) is set as 3; the dimension of the output layer (*outdim*) is set as 1; (ii) for the PSO algorithm, the maximum iteration number (*itmax*) is set as 300; the number of the searching particles, *N*, is set as 40; the length of each particle, *D*, is set as 3; weight c_1 and c_2 are set as 2.

The PSO-SVR model not only has its embedded constraints and limitations from the original SVR model, but also has huge iteration steps as a result of the requirements of the PSO algorithm. Therefore, it would be time consuming to train the PSO-SVR model while the total training set is used. For this consideration, the total training set is divided into two sub-sets, namely training subset and evaluation subset. In the PSO algorithm, the parameters used are set as follows: for the small sample, the maximum iteration number (*itmax*) is set as 50; the number of the searching particles, *N*, is set as 20; the length of each particle, *D*, is set as 3; weight c_1 and c_2 are set as 2; for the large sample, the maximum iteration number (*itmax*) is set as 20; the number of the searching particles, *N*, is set as 5; the length of each particle, *D*, is also set as 3; weight c_1 and c_2 are also set as 2.

Regarding Case 2, to be based on the same comparison conditions used in Fan *et al.* [25], the newest electric load data from NYISO is also employed for modeling, five alternative forecasting models (including the ARIMA, BPNN, GA-ANN, EMD-SVR-AR, and DEMD-SVR-AR models) are used for comparison with the proposed model. Some parameter settings of the employed forecasting models are set the same as in [25], and are briefly as follows: for the BPNN model, the node numbers of its structure are different for small sample size and large sample size; for the former one, the input layer dimension is 240, the hidden layer dimension is 12, and the output layer dimension is 48; and these values are 480, 12, 336, respectively, for the latter one. The parameters of GA-ANN model used in this case are as follows: generation number is set as 5, population size is set as 100, bit numbers are set as 50, mutation rate is set as 0.8, crossover rate is 0.05.

4.2. Evaluation Indices for Forecasting Performances

For evaluating the forecasting performances, three famous forecasting accurate level indices, RMSE (root mean square error), MAE (mean absolute error), and MAPE (mean absolute percentage error), as shown in Equations (24) to (26), are employed:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(P_i - A_i)^2}{n}} \tag{24}$$

$$MAE = \frac{\sum_{i=1}^{n}|P_i - A_i|}{n} \tag{25}$$

$$MAPE = \frac{\sum_{i=1}^{n}\left|\frac{P_i - A_i}{A_i}\right|}{n} * 100 \tag{26}$$

where P_i and A_i are the *i*-th forecasting and actual values, respectively, and *n* is the total number of forecasts.

In addition, to verify the suitability of model selection, Akaike's Information Criterion (AIC), an index of measurement for the relative quality of models for a given set of data, and Bayesian Information Criterion (BIC), also known as the Schwartz criterion, which is a criterion for model selection among a finite set of models (the model with the lowest BIC is preferred), are both taken into account to enhance the robustness of the verification. These two indices are defined as Equations (27) and (28), respectively:

$$AIC = \text{L}og(SSE) + 2q \tag{27}$$

where *SSE* is the sum of squares for errors, *q* is the number of estimated parameters:

$$BIC = \text{Log}(SSE) + qLog(n) \tag{28}$$

where *q* is the number of estimated parameters and *n* is the sample size.

4.3. Empirical Results and Analysis

For the first experiment in Case 1, the forecasting results (the electric load on 8 May 2007) of the original SVR model, the PSO-SVR model and the proposed DEMD-QPSO-SVR-AR model are shown in Figure 6a. For Case 2, the forecasting results of the ARIMA model, the BPNN model, the GA-ANN model and the proposed DEMD-QPSO-SVR-AR model are shown in Figure 7a. Based on these two figures, the forecasting curve of the proposed DEMD-QPSO-SVR-AR model seems to achieve a better fit than other alternative models for the two cases in this experiment.
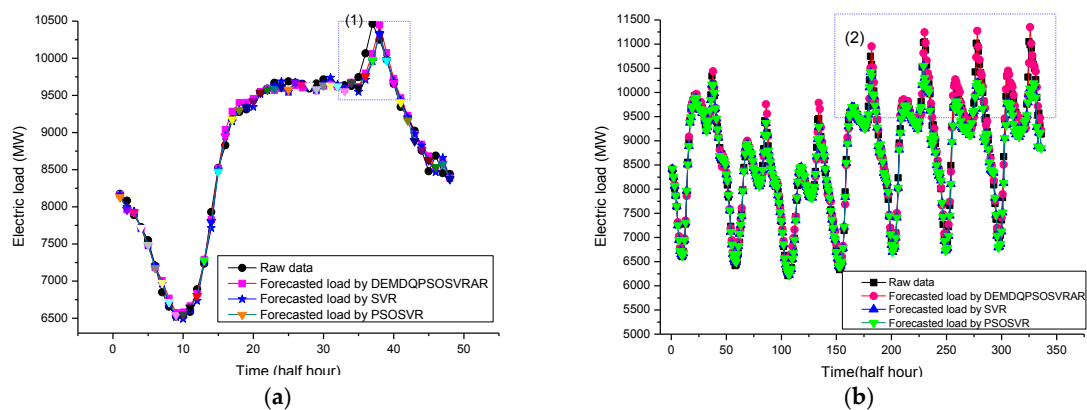

Figure 6. Comparison of the original data and the forecasted electric load by the DEMD-QPSO-SVR-AR Model, the SVR model and the PSO-SVR model for (**a**) the small sample size (One-day ahead prediction of 8 May 2007 are performed by the models); (**b**) the large sample size (One-week ahead prediction from 18 May 2007 to 24 May 2007 are performed by the models).
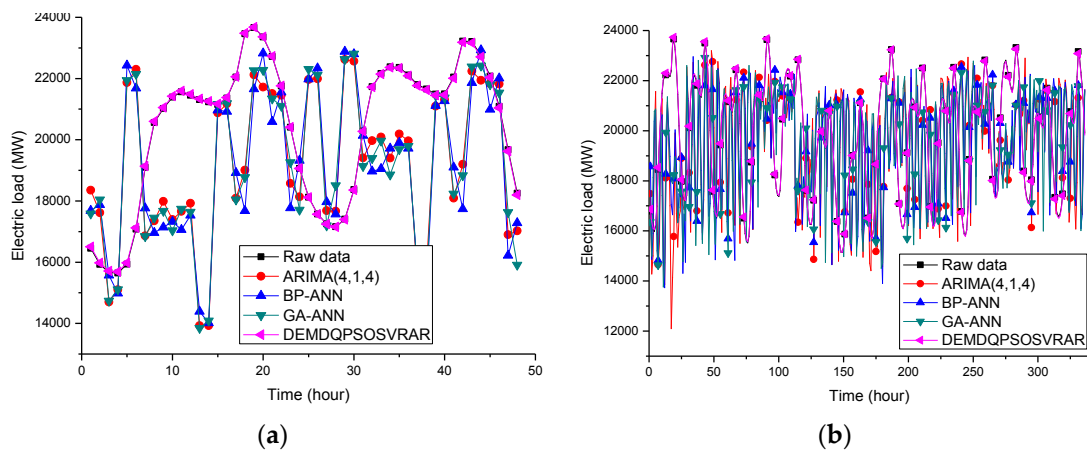

Figure 7. Comparison of the original data and the forecasted electric load by the DEMD-QPSO-SVR-AR Model, the ARIMA model, the BPNN model and the GA-ANN model for: (**a**) the small sample size (One-day ahead prediction from 13 to 14 January 2015 are performed by the models); (**b**) the large sample size (One-week ahead prediction from 2 to15 February 2015 are performed by the models).

The second experiments in Cases 1 and 2 show the large sample size data. The peak load values of the testing set are bigger than those of the training set. The detailed forecasting results

in this experiment are illustrated in Figures 6b and 7b. It also shows that the results obtained from the proposed DEMD-QPSO-SVR-AR model seem to have smaller forecasting errors than other alternative models.

Notice that for any particular sharp points in Figures 6 and 7 after extracting the direction feature of the trend by DEMD technology, these sharp points fixed in their positions represent the higher frequency characteristics of the remaining term, therefore, quantizing the particles in PSO the algorithm is very effective for dealing with this kind of fixed point characteristics. In other words, the DEMD-QPSO-SVR-AR model has better generalization ability than other alternative comparison models in both cases. Particularly in Case 1, for example, the local details for sharp points in Figure 6a,b are enlarged and are shown in Figure 8a,b, respectively. It is clear that the forecasting curve of the proposed DEMD-QPSO-SVR-AR model (red solid dots and red curve) fits more precisely than other alternative models, *i.e.*, it is superior for capturing the data change trends, including any fluctuation tendency.
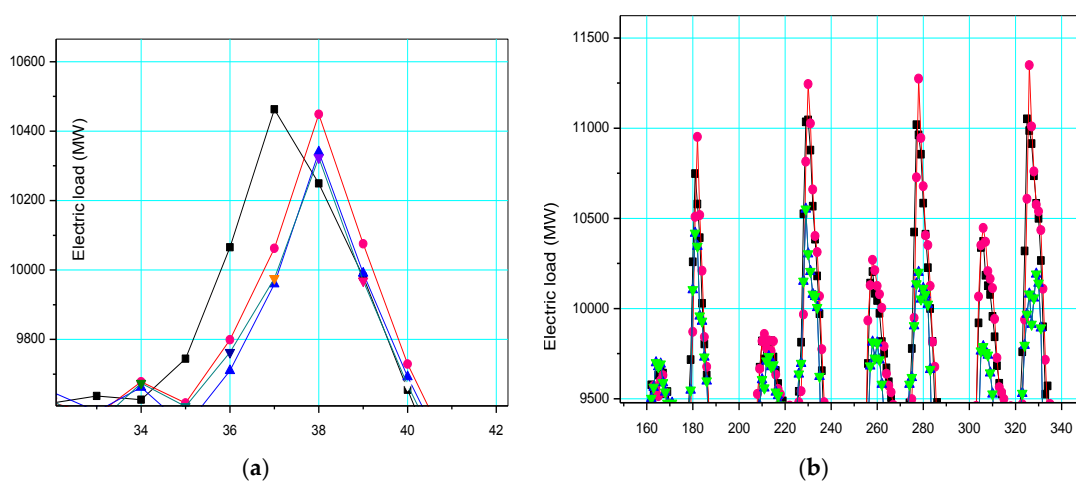

Figure 8. The local enlargement (peak) comparison of the DEMDQPSOSVRAR Model, the SVR model and the PSO-SVR model for (**a**) the small sample size (*1*); (**b**) the large sample size (*2*).

To better explain the superiority, the shape factor (*SF*), defined as Equation (29) and shown in Figure 9, is employed to illustrate the fitting effectiveness of the method, the *SF* value of the model closer to the one of the raw data, the fitness of the model is better than others. The results are shown in Table 6. It indicates that the data of *SF* from DEMD-QPSO-SVR-AR model is closer to the raw data than other models:

$$SF\,(\text{Shape Factor})\;=\;S_1(\text{Square of blue area})/S_2(\text{Square of red area}) \tag{29}$$
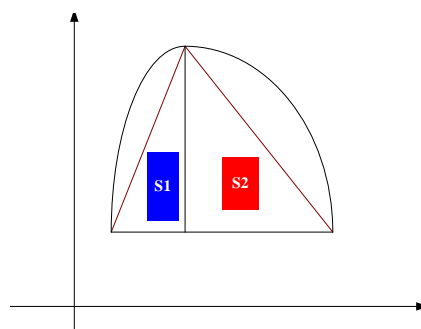

Figure 9. The definition of shape factor.

Table 6. The shape factor (*SF*) for local sharp points comparison.

Algorithms	Small Sample Size	Large Sample Size								
	SF	*SF1*	*SF2*	*SF3*	*SF4*	*SF5*	*SF6*	*SF7*	*SF8*	*MSF* *
Raw data	0.83	0.92	0.95	0.81	0.76	0.73	0.71	0.74	0.71	0.791
Original SVR	1.05	1.13	1.02	0.92	0.88	0.90	0.85	0.96	0.83	0.924
PSO-SVR	1.12	0.99	0.82	0.89	0.85	0.87	0.64	0.72	0.79	0.821
DEMD-QPSO-SVR-AR	**0.90**	**0.94**	**0.93**	**0.81**	**0.75**	**0.69**	**0.74**	**0.72**	**0.70**	**0.785**

* *MSF* = mean(*SF*).

The forecasting results in Cases 1 and 2 are summarized in Tables 7 and 8 respectively. The proposed DEMD-QPSO-SVR-AR model is compared with alternative models. It is indicated that our hybrid model outperforms all other alternatives in terms of all the evaluation criteria. One of the general observations is that the proposed model tends to fit closer to the actual value with a smaller forecasting error. This is ascribed to the fact that a well combined DEMD and QPSO can effectively capture the exact shape characteristics, which are difficult to illustrate by many other methods while data often has intertwined effects among the chaos, noise, and other unstable factors. Therefore, the unstable impact is well solved by DEMD, especially for those border points, and then, QPSO can accurately illustrate the chaotic rules, *i.e.*, achieve more satisfactory parameter solutions for an SVR model.

Table 7. Summary of results of the forecasting models in Case 1.

Algorithm	MAPE	RMSE	MAE	AIC	BIC	Running Time (s)	MSF (Data)
For the first experiment (small sample size)							
Original SVR [47]	11.6955	145.865	10.9181	112.3	111.9	180.4	0.972 (0.749)
PSO-SVR [47]	11.4189	145.685	10.6739	120.7	125.8	165.2	0.904 (0.749)
PSO-BP [47]	10.9094	142.261	10.1429	110.5	116.0	159.9	0.897 (0.749)
AFCM [23]	9.9524	125.323	9.2588	82.6	85.5	75.3	0.761 (0.749)
DEMD-QPSO-SVR-AR	9.1325	122.368	9.2201	80.9	83.1	100.7	0.756 (0.749)
For the second experiment (large sample size)							
Original SVR [47]	12.8765	181.617	12.0528	167.7	180.9	116.8	1.062 (0.830)
PSO-SVR [47]	13.503	271.429	13.0739	215.8	220.3	192.7	0.994 (0.830)
PSO-BP [47]	12.2384	175.235	11.3555	150.4	157.2	163.1	0.925 (0.830)
AFCM [23]	11.1019	158.754	10.4385	142.1	146.7	160.4	0.838 (0.830)
DEMD-QPSO-SVR-AR	4.1499	140.105	9.6258	129.0	128.1	169.0	0.826 (0.830)

In view of the model effectiveness and efficiency on the whole, we can conclude that the proposed model is quite competitive against other compared models, such as the ARIMA, BPNN, GA-ANN, PSO-BP, SVR, PSO-SVR, and AFCM models. In other words, the hybrid model leads to better accuracy and statistical interpretation.

In particular, as shown in Figure 8, our method shows higher accuracy and good flexibility in peak or inflection points, because the little redundant information could be used by statistical learning or regression models, and the level of optimization would increase. This also ensures that it could achieve more significant forecasting results due to the closer *SF* values to the raw data set. For closer insight, this can be viewed as the fact that the shape factor reflects how the electric load demand mechanism ia affected by multiple factors, *i.e.*, the shape factor reflects the change tendency in terms of ups or downs, thus, closer SF value to the raw data set can capture more precise trend changes than others, and this method no doubt can reveal the regularities for any point status.

Table 8. Summary of results of the forecasting models in Case 2.

Algorithm	MAPE	RMSE	MAE	AIC	BIC	MSF (Data)
For the first experiment (small sample size)						
ARIMA(4,1,4) [25]	45.33	320.45	25.72	278.4	285.1	1.132 (0.864)
BPNN [25]	31.76	219.43	21.69	198.5	200.2	0.955 (0.864)
GA-ANN [25]	23.89	220.96	23.55	199.3	202.3	0.947 (0.864)
EMD-SVR-AR [25]	14.31	158.11	17.44	140.7	141.9	0.873 (0.864)
DEMD-SVR-AR [25]	8.19	140.16	12.79	128.3	130.1	0.865 (0.864)
DEMD-QPSO-SVR-AR	7.48	138.89	14.44	125.4	126.7	0.859 (0.864)
For the second experiment (large sample size)						
ARIMA(4,1,4) [25]	60. 65	733.22	54.05	551.4	579.4	1.091 (0.875)
BPNN [25]	42.5	479.48	50.39	334.6	342.3	0.976 (0.875)
GA-ANN [25]	33.12	450.63	44.35	321.7	323.6	0.953 (0.875)
EMD-SVR-AR [25]	11.29	289.21	20.76	239.0	238.2	0.890 (0.875)
DEMD-SVR-AR [25]	5.37	160.58	15.82	141.4	142.2	0.884 (0.875)
DEMD-QPSO-SVR-AR	4.62	153.22	16.30	132.5	133.8	0.879 (0.875)

Several findings deserved to be noted. Firstly, based on the forecasting performance comparisons among these models, the proposed model outperforms other alternative models. Secondly, the proposed model has better generalization ability for different input patterns as shown in the second experiment. Thirdly, from the comparison between the different sample sizes of these two experiments, we conclude that the hybrid model can tolerate more redundant information and construct the model for the larger sample size data set. Fourthly, based on the calculation and comparison of *SF* in Table 6, the proposed model also receives closer *SF* values to the raw data than other alternative models. Finally, since the proposed model generates good results with good accuracy and interpretability, it is robust and effective, as shown in Tables 7 and 8 comparing the other models, namely the original SVR, PSO-SVR, PSO-BP and AFCM models. Overall, the proposed model provides a very powerful tool that is easy to implement for electric load forecasting.

Eventually, the most important issue is to verify the significance of the accuracy improvement of the proposed model. The forecasting accuracy comparisons in both cases among original SVR, PSO-SVR, PSO-BP, AFCM, ARIMA, BPNN, and GA-ANN models are conducted by a statistical test, namely a Wilcoxon signed-rank test, at the 0.025 and 0.05 significant levels in one-tail-tests. The Wilcoxon signed-rank test is a non-parametric statistical hypothesis test used when comparing two related samples, matched samples, or repeated measurements on a single sample to assess whether their population mean ranks differ (*i.e.*, it is a paired difference test). It can be used as an alternative to the paired Student's t-test, t-test for matched pairs, or the *t*-test for dependent samples when the population cannot be assumed to be normally distributed [48]. The test results are shown in Tables 9 and 10. Clearly, the outstanding forecasting results achieved by the proposed model is only significantly superior to other alternative models at a significance level of 0.05. This also implies that there are still lots of improvement efforts that can be made for hybrid quantum-behavior evolutionary SVR-based models.

Table 9. Wilcoxon signed-rank test in Case 1.

Compared Models	Wilcoxon Signed-Rank Test	
	$\alpha = 0.025; W = 4$	$\alpha = 0.05; W = 6$
DEMDQPSOSVRAR *vs.* original SVR	8	3 [a]
DEMDQPSOSVRAR *vs.* PSO-SVR	6	2 [a]
DEMDQPSOSVRAR *vs.* PSO-BP	6	2 [a]
DEMDQPSOSVRAR *vs.* AFCM	6	2 [a]

[a] denotes that the DEMD-QPSO-SVR-AR model significantly outperforms other alternative models.

Table 10. Wilcoxon signed-rank test in Case 2.

Compared Models	Wilcoxon Signed-Rank Test	
	$\alpha = 0.025; W = 4$	$\alpha = 0.05; W = 6$
DEMDQPSOSVRAR *vs.* ARIMA	6	2 [a]
DEMDQPSOSVRAR *vs.* BPNN	6	2 [a]
DEMDQPSOSVRAR *vs.* GA-ANN	6	2 [a]
DEMDQPSOSVRAR *vs.* EMD-SVR-AR	6	2 [a]
DEMDQPSOSVRAR *vs.* DEMD-SVR-AR	6	2 [a]

[a] denotes that the DEMD-QPSO-SVR-AR model significantly outperforms other alternative models.

5. Conclusions

This paper presents an SVR model hybridized with the differential empirical mode decomposition (DEMD) method and quantum particle swarm optimization algorithm (QPSO) for electric load forecasting. The experimental results indicate that the proposed model is significantly superior to the original SVR, PSO-SVR, PSO-BP, AFCM, ARIMA, BPNN, and GA-ANN models. To improve the forecasting performance (accuracy level), quantum theory is hybridized with PSO (namely the QPSO) into an SVR model to determine its suitable parameter values. Furthermore, the DEMD is employed to simultaneously consider the accuracy and comprehensibility of the forecast results. Eventually, a hybrid model (namely DEMD-QPSO-SVR-AR model) has been proposed and its electric load forecasting superiority has also been compared with other alternative models. It is also demonstrated that a well combined DEMD and QPSO can effectively capture the exact shape characteristics, which are difficult to illustrate by many other methods while data often has intertwined effects among the chaos, noise, and other unstable factors. Hence, the instability impact can be well solved by DEMD, especially for those border points, and then, QPSO can accurately illustrate the chaotic rules, thus achieving more satisfactory parameter solutions than an SVR model.

Author Contributions: Li-Ling Peng and Wei-Chiang Hong conceived and designed the experiments; Guo-Feng Fan and Min-Liang Huang performed the experiments; Li-Ling Peng and Wei-Chiang Hong analyzed the data; Li-Ling Peng wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bernard, J.T.; Bolduc, D.; Yameogo, N.D.; Rahman, S. A pseudo-panel data model of household electricity demand. *Resour. Energy Econ.* **2010**, *33*, 315–325. [CrossRef]
2. Bianco, V.; Manca, O.; Nardini, S. Electricity consumption forecasting in Italy using linear regression models. *Energy* **2009**, *34*, 1413–1421. [CrossRef]

3. Zhou, P.; Ang, B.W.; Poh, K.L. A trigonometric grey prediction approach to forecasting electricity demand. *Energy* **2006**, *31*, 2839–2847. [CrossRef]

4. Afshar, K.; Bigdeli, N. Data analysis and short term load forecasting in Iran electricity market using singular spectral analysis (SSA). *Energy* **2011**, *36*, 2620–2627. [CrossRef]

5. Kumar, U.; Jain, V.K. Time series models (Grey-Markov, Grey Model with rolling mechanism and singular spectrum analysis) to forecast energy consumption in India. *Energy* **2010**, *35*, 1709–1716. [CrossRef]

6. Bianco, V.; Scarpa, F.; Tagliafico, L.A. Long term outlook of primary energy consumption of the Italian thermoelectric sector: Impact of fuel and carbon prices. *Energy* **2015**, *87*, 153–164. [CrossRef]

7. Topalli, A.K.; Erkmen, I. A hybrid learning for neural networks applied to short term load forecasting. *Neurocomputing* **2003**, *51*, 495–500. [CrossRef]

8. Kandil, N.; Wamkeue, R.; Saad, M.; Georges, S. An efficient approach for short term load forecasting using artificial neural networks. *Int. J. Electr. Power Energy Syst.* **2006**, *28*, 525–530. [CrossRef]

9. Beccali, M.; Cellura, M.; Brano, V.L.; Marvuglia, A. Forecasting daily urban electric load profiles using artificial neural networks. *Energy Convers. Manag.* **2004**, *45*, 2879–2900. [CrossRef]

10. Topalli, A.K.; Cellura, M.; Erkmen, I.; Topalli, I. Intelligent short-term load forecasting in Turkey. *Electr. Power Energy Syst.* **2006**, *28*, 437–447. [CrossRef]

11. Pai, P.F.; Hong, W.C. Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms. *Electr. Power Syst. Res.* **2005**, *74*, 417–425. [CrossRef]

12. Hong, W.C. Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm. *Energy* **2011**, *36*, 5568–5578. [CrossRef]

13. Yesilbudak, M.; Sagiroglu, S.; Colak, I. A new approach to very short term wind speed prediction using k-nearest neighbor classification. *Energy Convers Manag.* **2013**, *69*, 77–86. [CrossRef]

14. Peng, H.; Liu, F.; Yang, X. A hybrid strategy of short term wind power prediction. *Renew. Energy* **2013**, *50*, 590–595. [CrossRef]

15. An, X.; Jiang, D.; Liu, C.; Zhao, M. Wind farm power prediction based on wavelet decomposition and chaotic time series. *Expert Syst Appl.* **2011**, *38*, 11280–11285. [CrossRef]

16. Lei, Y.; Lin, J.; He, Z.; Zuo, M.J. A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mech. Syst. Signal. Process.* **2013**, *35*, 108–126. [CrossRef]

17. Wong, P.; Xu, Q.; Vong, C.; Wong, H. Rate-dependent hysteresis modeling and control of a piezostage using online support vector machine and relevance vector machine. *IEEE Trans. Ind. Electron.* **2012**, *59*, 1988–2001. [CrossRef]

18. Wang, Z.; Liu, L. Sensitivity prediction of sensor based on relevance vector machine. *J. Inf. Comput. Sci.* **2012**, *9*, 2589–2597.

19. Hong, W.C. *Intelligent Energy Demand Forecasting*; Springer: London, UK, 2013.

20. Huang, B.; Kunoth, A. An optimization based empirical mode decomposition scheme. *J. Comput. Appl. Math.* **2013**, *240*, 174–183. [CrossRef]

21. Fan, G.; Qing, S.; Wang, H.; Shi, Z.; Hong, W.C.; Dai, L. Study on apparent kinetic prediction model of the smelting reduction based on the time series. *Math. Probl. Eng.* **2012**, *2012*. [CrossRef]

22. Bhusana, P.; Chris, T. Improving prediction of exchange rates using Differential EMD. *Expert Syst. Appl.* **2013**, *40*, 377–384.

23. An, X.; Jiang, D.; Zhao, M.; Liu, C. Short-term prediction of wind power using EMD and chaotic theory. *Commun. Nonlinear Sci. Numer. Simul.* **2012**, *17*, 1036–1042. [CrossRef]

24. Huang, N.E.; Shen, Z. A new View of nonliner water waves: The Hilbert spectrum. *Rev. Fluid Mech.* **1999**, *31*, 417–457. [CrossRef]

25. Fan, G.; Peng, L.-L.; Hong, W.-C.; Sun, F. Electric load forecasting by the SVR model with differential empirical mode decomposition and auto regression. *Neurocomputing* **2016**, *173*, 958–970. [CrossRef]

26. Wu, P.F.; Gao, L.Q.; Zou, D.X.; Li, S. An improved particle swarm optimization algorithm for reliability problems. *ISA. Trans.* **2011**, *50*, 71–81. [CrossRef] [PubMed]

27. Alfi, A. Particle swarm optimization algorithm with dynamic inertia weight for online parameter identification applied to Lorenz chaotic system. *Int. J. Innov. Comput. Inf.* **2012**, *8*, 1191–1203.

28. Alfi, A.; Modares, H. System identification and control using adaptive particle swarm optimization. *Appl. Math. Model.* **2011**, *35*, 1210–1221. [CrossRef]

29. Alfi, A. PSO with adaptive mutation and inertia weight and its application in parameter estimation of dynamic systems. *Acta Autom. Sin.* **2011**, *37*, 541–549. [CrossRef]

30. Wu, X.J.; Huang, Q.; Zhu, X.J. Thermal modeling of a solid oxide fuel cell and micro gas turbine hybrid power system based on modified LS-SVM. *Int. J. Hydrogen Energy.* **2011**, *36*, 885–892. [CrossRef]

31. Liu, X.Y.; Shao, C.; Ma, H.F.; Liu, R.X. Optimal earth pressure balance control for shield tunneling based on LS-SVM and PSO. *Autom. Constr.* **2011**, *29*, 321–327. [CrossRef]

32. Sun, J.; Feng, B.; Xu, W.B. Particle swarm optimization with particles having quantum behavior. In Proceedings of the IEEE Proceedings of Congress on Evolutionary Computation, New Orleans, LA, USA, 19–23 June 2004; pp. 325–331.

33. Sun, J.; Xu, W.B.; Feng, B. A global search strategy of quantum-behaved particle swarm optimization. In Proceedings of the Cybernetics and Intelligent Systems Proceedings of the 2004 IEEE Conference, Singapore, 4–9 May 2004; pp. 111–116.

34. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer-Verlag: New York, NY, USA, 1995.

35. Bergh, F.V.D. An Analysis of Particle Swarm Optimizers. Ph.D. Thesis, University of Pretoria, Pretoria, South Africa, 2001.

36. Sun, J.; Fang, W.; Wu, X.J.; Palade, V.; Xu, W.B. Quantum-behaved particle swarm optimization: Analysis of individual particle behavior and parameter selection. *Evol. Comput.* **2012**, *20*, 349–393. [CrossRef] [PubMed]

37. Sun, J.; Fang, W.; Palade, V.; Wu, X.J.; Xu, W.B. Quantum-behaved particle swarm optimization with gaussian distributed local attractor point. *Appl. Math. Comput.* **2011**, *218*, 3763–3775. [CrossRef]

38. Sun, T.H. Applying particle swarm optimization algorithm to roundness measurement. *Expert Syst. Appl.* **2009**, *36*, 3428–3438. [CrossRef]

39. Xi, M.; Sun, J.; Xu, W. An improved quantum-behaved particle swarm optimization algorithm with weighted mean best position. *Appl. Math. Comput.* **2008**, *205*, 751–759. [CrossRef]

40. Coelho, L.S. An efficient particle swarm approach for mixed-integer programming in reliability—Redundancy optimization applications. *Reliab. Eng. Syst. Saf.* **2009**, *94*, 830–837. [CrossRef]

41. Fang, Y.; Zhao, Y.D.; Ke, M.; Zhao, X.; Iu, H.H.; Wong, K.P. Quantum inspired particle swarm optimization for power system operations considering wind power uncertainty and carbon tax in Australia. *IEEE Trans. Ind. Inform.* **2012**, *8*, 880–888.

42. Ko, C.N.; Jau, Y.M.; Jeng, J.T. Parameter estimation of chaotic dynamical systems using quantum-behaved particle swarm optimization based on hybrid evolution. *J. Inf. Sci. Eng.* **2015**, *31*, 675–689.

43. Ko, C.N.; Chang, Y.P.; Wu, C.J. An orthogonal-array-based particle swarm optimizer with nonlinear time-varying evolution. *Appl. Math. Comput.* **2007**, *191*, 272–279. [CrossRef]

44. Ratnaweera, A.; Halgamuge, S.K.; Watson, H.C. Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients. *IEEE Transactions on Evol. Comput.* **2004**, *8*, 240–255. [CrossRef]

45. Gao, J. Asymptotic properties of some estimators for partly linear stationary autoregressive models. *Commun. Statist. Theory Meth.* **1995**, *24*, 2011–2026. [CrossRef]

46. Taylor, J.W. Short-term load forecasting with exponentially weighted methods. *IEEE Trans. Power Syst.* **2012**, *27*, 458–464. [CrossRef]

47. Che, J.; Wang, J.; Wang, G. An adaptive fuzzy combination model based on self-organizing map and support vector regression for electric load forecasting. *Energy* **2012**, *37*, 657–664. [CrossRef]

48. Lowry, R. *Concepts & Applications of Inferential Statistics*; Vassar College: New York, NY, USA, 2011.