

Article

## Support Vector Regression Model Based on Empirical Mode Decomposition and Auto Regression for Electric Load Forecasting

Guo-Feng Fan<sup>1</sup>, Shan Qing<sup>1,\*</sup>, Hua Wang<sup>1</sup>, Wei-Chiang Hong<sup>2</sup> and Hong-Juan Li<sup>1</sup>

<sup>1</sup> Engineering Research Center of Metallurgical Energy Conservation and Emission Reduction, Ministry of Education, Kunming University of Science and Technology, Kunming 650093, China; E-Mails: guofengtongzhi@163.com (G.-F.F.); wanghua65@163.com (H.W.); fxzwlihongjuan@163.com (H.-J.L.)

<sup>2</sup> Department of Information Management, Oriental Institute of Technology/58 Sec. 2, Sichuan Rd., Panchiao, Taipei 220, Taiwan; E-Mail: samuelsonhong@gmail.com

\* Author to whom correspondence should be addressed; E-Mail: yanls22@163.com; Tel.: +86-1388-855-2395; Fax: +86-0871-6515-3405.

Received: 28 November 2012; in revised form: 2 February 2013 / Accepted: 25 March 2013 / Published: 2 April 2013

---

**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 a SVR model hybridized with the empirical mode decomposition (EMD) method and auto regression (AR) for electric load forecasting. The electric load data of the New South Wales (Australia) market are employed for comparing the forecasting performances of different forecasting models. The results confirm the validity of the idea that the proposed model can simultaneously provide forecasting with good accuracy and interpretability.

**Keywords:** electric load prediction; support vector regression; empirical mode decomposition auto regression

---

### 1. Introduction

Electric energy is an unstored resource, thus, electric load forecasting plays a vital role in the management of the daily operations of a power utility, such as energy transfer scheduling, unit

commitment, and load dispatch. 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 load forecasting accuracy. 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] proposed an improved singular spectral analysis method for short-term load forecasting (STLF) for the Iranian electricity market; and Kumar and Jain [5] applied three time series models—Grey-Markov model, Grey-Model with rolling mechanism, and singular spectrum analysis—to forecast the consumption of conventional energy in India. By employing artificial neural networks, references [6–9] proposed several useful short-term load forecasting models. By hybridizing the popular method and evolutionary algorithm, the authors of [10–13] demonstrated further performance improvements which could be made for energy forecasting. Though these methods can yield a significant proven forecasting accuracy improvement in some cases, they have usually focused on the improvement of the accuracy without paying special attention to the interpretability. Recently, expert systems, mainly developed by means of linguistic fuzzy rule-based systems, allow us to deal with the system modeling with good interpretability [14]. However, these models have strong dependency on an expert and often cannot generate good accuracy. Therefore, combination models, based on the popular methods, expert systems and other techniques, are proposed to satisfy both high accurate level and interpretability.

Based on the advantages in statistical learning capacity to handle high dimensional data, the SVR (support vector regression) model, especially suitable for small sample size learning, has become a popular algorithm for many forecasting problems [15–17]. As a disadvantage of an SVR method, it is easily trapped into a local optimum during the nonlinear optimization process of the three parameters, in the meanwhile, its robustness and sparsity are also lacking satisfactory levels. On the other hand, empirical mode decomposition (EMD) and auto regression (AR), a fast, easy and reliable unsupervised clustering algorithm, has been successfully applied to many fields, such as communication, society, economy, engineering, and has achieved good effects [18–20]. Particularly, the EMD method can effectively extract the components of the basic mode from nonlinear or non-stationary time series [21], *i.e.*, the original complex time series can be transferred into a series of single and apparent components. It can effectively reduce the interactions among lots of singular values and improve the forecasting performance of a single kernel function. Thus, it is useful to employ suitable kernel functions for forecasting the medium-and-long-term tendencies of the time series.

In this paper, we present a new hybrid model with clear human-understandable knowledge on training data to achieve a satisfactory level of forecasting accuracy. The principal idea is hybridizing EMD with SVR and AR, namely creating the EMDSVRAR model, to receive better solutions. The proposed EMDSVRAR model has the capability of smoothing and reducing the noise (inherited from EMD), the capability of filtering dataset and improving forecasting performance (inherited from SVR), and the in capability of effectively forecasting the future tendencies of data (inherited from AR). The forecasting outputs of an unseen example by using the hybrid method are described in the following section.

To show the applicability and superiority of the proposed algorithm, half-hourly electric load data (48 data points per day) from New South Wales (Australia) with two different sample sizes are

employed to compare the forecasting performances among the proposed model and other four alternative models, namely the PSO-BP model (BP neural network trained by a particle swarm optimization algorithm), SVR model, PSO-SVR model (optimal combination of SVR parameters determined by a PSO algorithm), and the AFCM model (an adaptive fuzzy combination model based on a self-organizing map and support vector regression). This study also suggests that researchers and practitioners should carefully consider the nature and intention in using these electric load data while neural networks, statistical methods, and other hybrid models are being determined to be the critical management tools in electricity markets. The experimental results indicate that this proposed EMDSVRAR model has the following advantages: (1) simultaneously satisfies the need for high levels of accuracy and interpretability; (2) the proposed model can tolerate more redundant information than the original SVR model, thus, it has better generalization ability.

The rest of this paper is organized as follows: in Section 2, the EMDSVRAR forecasting model is introduced and the main steps of the model are given. In Section 3, the data description and the research design are outlined. The numerical results and comparisons are presented and discussed in Section 4. A brief conclusion of this paper and the future research are provided in Section 5.

## 2. Support Vector Regression with Empirical Mode Decomposition

### 2.1. Empirical Mode Decomposition (EMD)

The EMD method is based on the simple assumption that any signal consists of different simple intrinsic modes of oscillations. 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. Each mode should be independent of the others. In this way, each signal could be decomposed into a number of intrinsic mode functions (IMFs), each of which should satisfy the following two definitions [22]:

- a In the whole data set, the number of extreme and the number of zero-crossings should either equal or differ to each other at most by one.
- b At any point, the mean value of the envelope defined by local maxima and the envelope defined by the local minima is zero.

An IMF represents a simple oscillatory mode compared with the simple harmonic function. With the definition, any signal  $x(t)$  can be decomposed as following steps:

- 1 Identify all local extremes, and then connect all the local maxima by a cubic spline line as the upper envelope.
- 2 Repeat the procedure for the local minima to produce the lower envelope. The upper and lower envelopes should cover all the data among them.
- 3 The mean of upper and low envelope value is designated as  $m_1$ , and the difference between the signal  $x(t)$  and  $m_1$  is the first component,  $h_1$ , as shown in Equation (1):

$$h_1 = x(t) - m_1 \quad (1)$$

Generally speaking,  $h_1$  will not necessarily meet the requirements of the IMF, because  $h_1$  is not a standard IMF. It needs to be determined for  $k$  times until the mean envelope tends to zero. Then, the

first intrinsic mode function  $c_1$  is introduced, which stands for the most high-frequency component of the original data sequence. At this point, the data could be represented as Equation (2):

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

where  $h_{1k}$  is the datum after  $k$  times siftings.  $h_{1(k-1)}$  stands for the data after  $k-1$  times sifting. Standard deviation ( $SD$ ) is used to determine whether the results of each filter component meet the IMF or not.  $SD$  is defined as Equation (3):

$$SD = \sum_{k=1}^T \frac{|h_{1(k-1)}(t) - h_{1k}(t)|^2}{h_{1(k-1)}^2(t)} \quad (3)$$

where  $T$  is the length of the data.

The value of standard deviation  $SD$  is limited in the range of 0.2 to 0.3, which means when  $0.2 < SD < 0.3$ , the decomposition process can be finished. The consideration for this standard is that it should not only ensure  $h_k(t)$  to meet the IMF requirements, but also control the decomposition times. Therefore, in this way, the IMF components could retain amplitude modulation information in the original signal.

- 4 When  $h_{1k}$  has met the basic requirements of  $SD$ , based on the condition of  $c_1 = h_{1k}$ , the signal  $x(t)$  of the first IMF component  $c_1$  can be obtained directly, and a new series  $r_1$  could be achieved after deleting the high frequency components. This relationship could be expressed as Equation (4):

$$r_1 = x(t) - c_1 \quad (4)$$

The new sequence is treated as the original data and repeats the steps (1) to (3) processes. The second intrinsic mode function  $c_2$  could be obtained.

- 5 Repeat previous steps (1) to (4) until the  $r_n$  can not be decomposed into the IMF. The sequence  $r_n$  is called the remainder of the original data  $x(t)$ .  $r_n$  is a monotonic sequence, it can indicate the overall trend of the raw data  $x(t)$  or mean, and it is usually referred as the so-called trend items. It is of clear physical significance. The process is expressed as Equations (5) and (6):

$$r_1 = x(t) - c_1, r_2 = r_1 - c_2, \dots, r_n = r_{n-1} - c_n \quad (5)$$

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

The original data can be expressed as the IMF component and remainder.

## 2.2. Support Vector Regression

The notions of SVMs for the case of regression are introduced briefly. Given a data set of  $N$  elements  $\{(X_i, y_i), i=1, 2, \dots, N\}$ , where  $X_i$  is the  $i$ -th element in  $n$ -dimensional space, *i.e.*,  $X_i = [x_{1i}, \dots, x_{ni}] \in \mathfrak{R}^n$ , and  $y_i \in \mathfrak{R}$  is the actual value corresponding to  $X_i$ . A non-linear mapping  $(\cdot)$ :  $\mathfrak{R}^n \rightarrow \mathfrak{R}^{n_h}$  is defined to map the training (input) data  $X_i$  into the so-called high dimensional feature space (which may have infinite dimensions),  $\mathfrak{R}^{n_h}$  (Figure 1a,b). 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 (7):

$$f(X) = W^T \varphi(X) + b \tag{7}$$

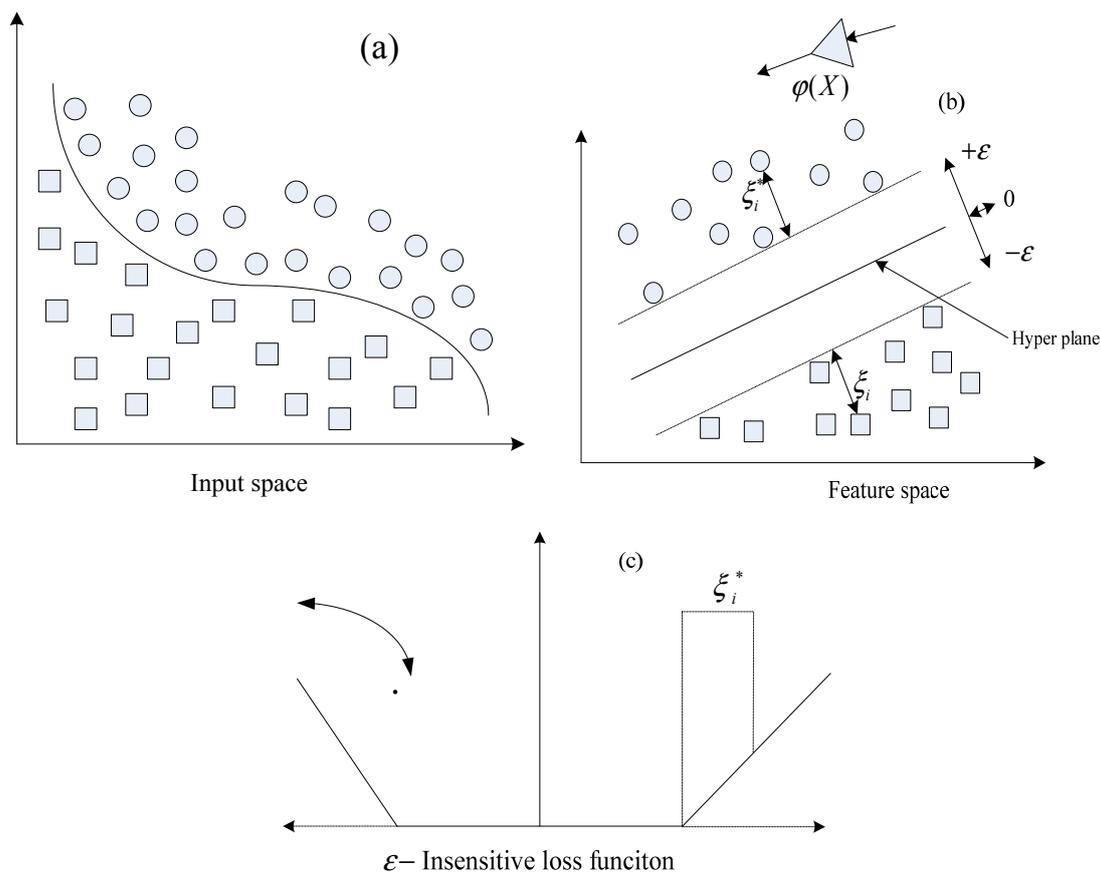
where  $f(X)$  denotes the forecasting values; the coefficients  $W$  ( $W \in \mathfrak{R}^{n_h}$ ) and  $b$  ( $b \in \mathfrak{R}$ ) are adjustable. As mentioned above, the SVM method aims at minimizing the empirical risk, shown as Equation (8):

$$R_{emp}(f) = \frac{1}{N} \sum_{i=1}^N \Theta_{\epsilon}(y_i, W^T \varphi(X_i) + b) \tag{8}$$

where  $\Theta_{\epsilon}(y, f(x))$  is the  $\epsilon$ -insensitive loss function (indicated as a thick line in Figure 1c) and defined as Equation (9):

$$\Theta_{\epsilon}(Y, f(X)) = \begin{cases} |f(X) - Y| - \epsilon, & \text{if } |f(X) - Y| \geq \epsilon \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

**Figure 1.** Transformation process illustration of a SVR model. (a) Input space; (b) Feature space; (c)  $\epsilon$ -insensitive loss function.



In addition,  $\Theta_{\epsilon}(Y, f(X))$  is employed to find out an optimum hyperplane on the high dimensional feature space (Figure 1b) to maximize the distance separating the training data into two subsets. Thus, the SVR focuses on finding the optimum hyper plane and minimizing the training error between the

training data and the  $\varepsilon$ -insensitive loss function. Then, the SVR minimizes the overall errors, shown as Equation (10):

$$\text{Min}_{W, b, \xi^*, \xi} R_\varepsilon(W, \xi^*, \xi) = \frac{1}{2} W^T W + C \sum_{i=1}^N (\xi_i^* + \xi_i) \quad (10)$$

with the constraints:

$$\begin{aligned} Y_i - W^T \varphi(X_i) - b &\leq \varepsilon + \xi_i^*, & i = 1, 2, \dots, N \\ -Y_i + W^T \varphi(X_i) + b &\leq \varepsilon + \xi_i, & i = 1, 2, \dots, N \\ \xi_i^* &\geq 0, & i = 1, 2, \dots, N \\ \xi_i &\geq 0, & i = 1, 2, \dots, N \end{aligned} \quad (11)$$

The first term of Equation (10), 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  $\varepsilon$ -insensitive loss function.  $C$  is the parameter to trade off these two terms. Training errors above  $\varepsilon$  are denoted as  $\xi_i^*$ , whereas training errors below  $-\varepsilon$  are denoted as  $\xi_i$  (Figure 1b).

After the quadratic optimization problem with inequality constraints is solved, the parameter vector  $w$  in Equation (7) is obtained as Equation (12):

$$W = \sum_{i=1}^N (\beta_i^* - \beta_i) \varphi(X_i) \quad (12)$$

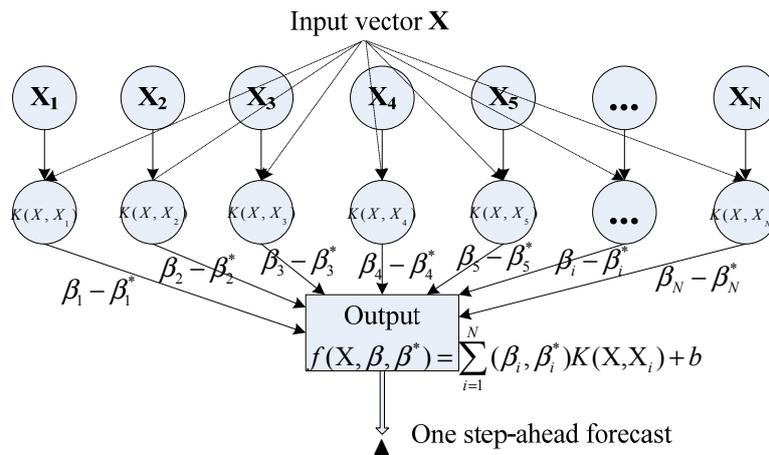
where  $\xi_i^*, \xi_i$  are obtained by solving a quadratic program and are the Lagrangian multipliers. Finally, the SVR regression function is obtained as Equation (13) in the dual space:

$$f(X) = \sum_{i=1}^N (\beta_i^* - \beta_i) K(X_i, X) + b \quad (13)$$

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 [23] 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  $\sigma$ :  $K(X_i, X_j) = \exp(-0.5 \|X_i - X_j\|^2 / \sigma^2)$  and the polynomial kernel with an order of  $d$  and constants  $a_1$  and  $a_2$ :  $K(X_i, X_j) = (a_1 X_i X_j + a_2)^d$ . 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. The forecasting process of a SVR model is illustrated in Figure 2.

Figure 2. The forecasting process of a SVR model.



2.3. AR Model

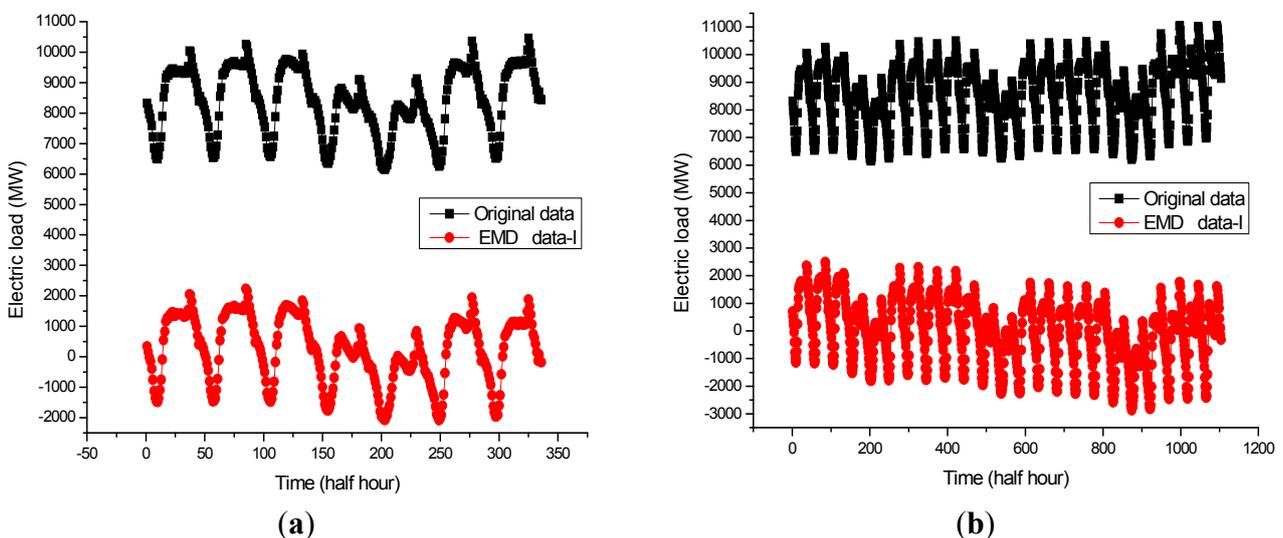
Equation (14) expresses a  $p$ -step autoregressive model, referring as  $AR(p)$  model [24]. Stationary time series  $\{X_t\}$  that meet the model  $AR(p)$  is called the  $AR(p)$  sequence. That  $a = (a_1, a_2, \dots, a_p)^T$  is named as the regression coefficients of the  $AR(p)$  model:

$$X_t = \sum_{j=1}^p a_j X_{t-j} + \varepsilon_t, t \in Z \tag{14}$$

3. Numerical Examples

In the first experiment, the proposed model is trained by electric load from New South Wales (Australia) from 2 May 2007 to 7 May 2007, and testing electric load is from 8 May 2007. The employed electric load data is on a half-hourly basis (*i.e.*, 48 data points per day). The data size contains only 7 days, to differ from the other example with more sample data, this example is so-called the small sample size data, and illustrated in Figure 3a.

Figure 3. (a) Half-hourly electric load in New South Wales from 2 May 2007 to 8 May 2007; (b) Half-hourly electric load in New South Wales from 2 May 2007 to 24 May 2007.

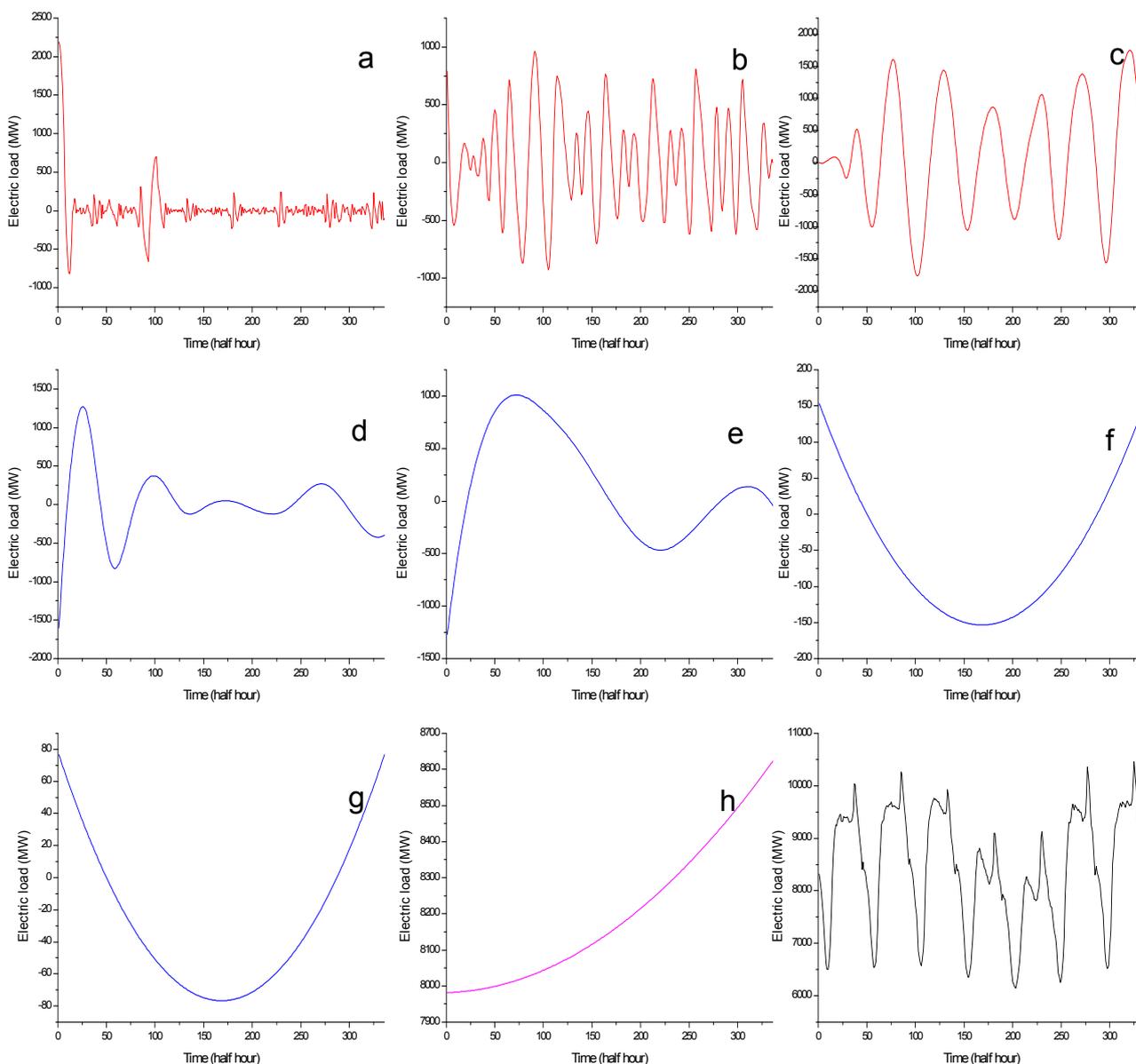


Too large training sets should avoid overtraining during the learning process of the SVR model. Therefore, the second experiment with 23 days (1104 data points from 2 May 2007 to 24 May 2007) is modeled by using part of all the training samples as training set. This example is so-called the large sample size data, and illustrated in Figure 3b.

### 3.1. Results after EMD

After being decomposed by EMD, the data can be divided into eight groups, which are shown in Figure 4a–h and the last group (Figure 4h) is a trend term (remainders). The so-called high frequency item is obtained by adding the preceding seven groups. From Figures 3a,b, the trend of the high frequency item is the same as original data, and its the structure is more regular, *i.e.*, it is more stable. Then, the high frequency item (data-I) and the remainders (data-II) have good effect of regression by the SVR and AR, respectively, and will be described as follow.

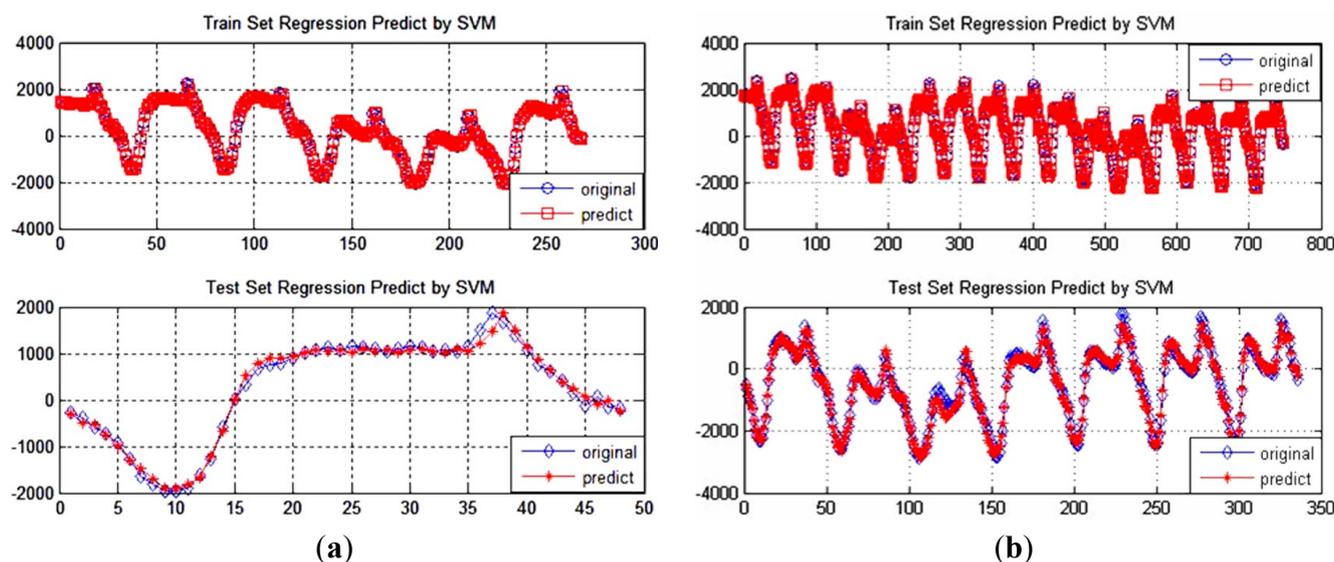
**Figure 4.** For ease of prevention, the graphs (a–h) show our section of plots at different IMFs for the small sample size.



3.2. Forecasting Using SVR for Data I (The High Frequency Item)

Firstly, for both small sample and large sample data, the high-frequency item is simultaneously employed for SVR modeling, and the better performances of the training and testing (forecasting) sets are shown in Figure 5a,b, respectively. The correlation coefficients of training effects are 0.9912 and 0.9901, respectively, of the forecast effects are 0.9875 and 0.9887, accordingly. This implies that the decomposition is helpful to improve the forecasting accuracy. The parameters of a SVR model for data I are shown in Table 1.

**Figure 5.** Comparison of the data-I and the forecasted electric load of train and test by the SVR model for the small sample and large sample data: (a) One-day ahead prediction of 8 May 2007 are performed by the model; (b) One-week ahead prediction from 18 May 2007 to 24 May 2007 are performed by the model.



**Table 1.** The SVR’s parameters for data-I and data-II.

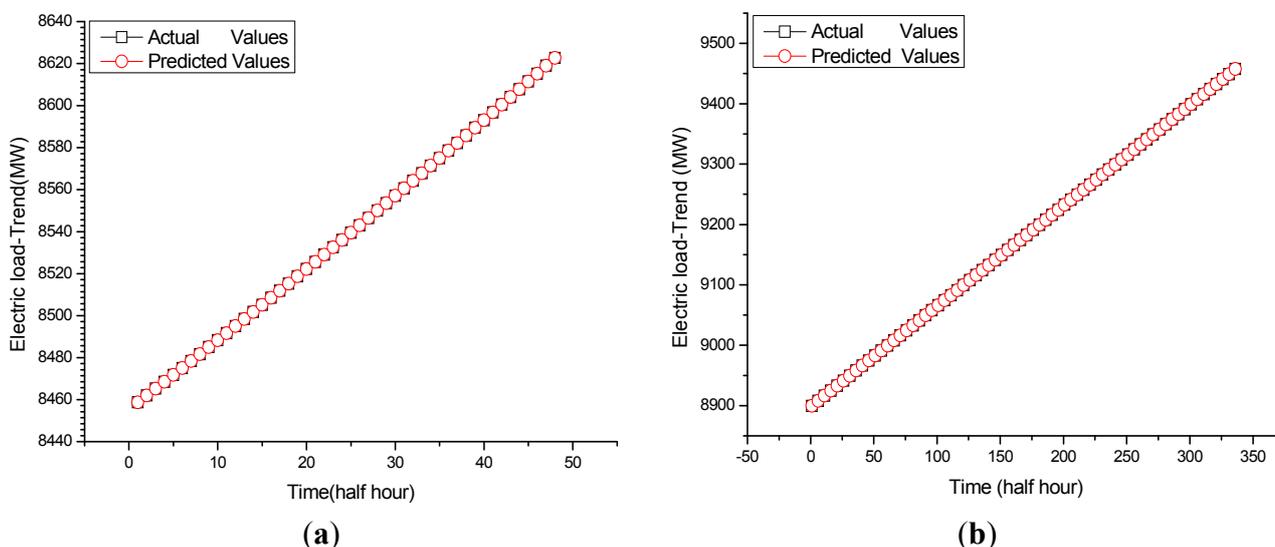
Sample size	m	$\sigma$	C	$\epsilon$	Testing MAPE
The high frequency item (data-I)	20	0.1	100	0.0061	9.85
The remainders (data-II)	20	0.35	181	0.0034	5.1

3.3. Forecasting Using AR for Data II (The Remainders)

Then, according to the geometric decay of the correlation coefficient and partial correlation coefficients fourth-order truncation for data II (the remainders), it can be regarded as AR (4) model. The parameters of a SVR model for data II are shown in Table 1.

As shown in Figure 6a,b, the remainders, for both small sample and large sample data, almost are in a straight line. The good forecasting results are shown in Table 2, and the errors have reached the level of  $10^{-7}$  for the small or large amount of data. It has demonstrated the superiority of the AR model.

**Figure 6.** Comparison of the data-II and the forecasted electric load by the AR model for the two experiments: (a) One-day ahead prediction of 8 May 2007 performed by the model; (b) One-week ahead prediction from 18 May 2007 to 24 May 2007 performed by the model.



**Table 2.** Summary of results of the AR forecasting model for data-II.

Remainders	MAE	Equation
The small sample size	$6.5567 \times 10^{-7}$	$x_n = 8417.298 + 1.013245x_{n-1} + 0.490278x_{n-2} - 0.011731x_{n-3} - 0.491839x_{n-4}$
The large sample size	$1.8454 \times 10^{-7}$	$x_n = 8546.869 + 1.000046x_{n-1} + 0.499957x_{n-2} - 5.18 \times 10^{-5}x_{n-3} - 0.499951x_{n-4}$

### 4. Result and Analysis

This section focuses on the efficiency of the proposed model with respect to computational accuracy and interpretability. To consider the small sample size modeling ability of the SVR model and conduct fair comparisons, we perform a real case experiment with relatively small sample size in the first experiment. The next experiment with 1104 datapoints is focused on illustrating the relationship between sample size and accuracy.

#### 4.1. Parameter Settings of the Employed Forecasting Models

As mentioned by Taylor [25], and to be based on the same comparison condition with Wang *et al.* [26], some parameter settings of the employed forecasting models are set as followings. For the PSO-BP model, we use 90 percent of all training samples as the training set, and the rest as the evaluation set. The parameters used in the PSO-BP are as follows: (i) The first set related to BP neural network: input layer dimension  $indim = 2$ , hidden layer dimension  $hiddennum = 3$ , output layer dimension  $outdim = 1$ ; (ii) The second set related to PSO: maximum iteration number  $itmax = 300$ , number of particles  $N = 40$ , length of particle  $D = 3$ , weight  $c_1 = c_2 = 2$ .

Because the PSO-SVR model embeds the construction and prediction algorithm of SVR in the fitness value iteration step of PSO, it will take a long time to train the PSO-SVR using the full training dataset.

For the above reason, we draw a small part of all training samples as training set, and the rest as evaluation set. The parameters used in the PSO are as follows: For small sample size: maximum iteration number  $itmax = 50$ , number of particles  $N = 20$ , length of particle  $D = 3$ , weight  $c_1 = c_2 = 2$ . For large sample size: maximum iteration number  $itmax = 20$ , number of particles  $N = 5$ , length of particle  $D = 3$ , weight  $c_1 = c_2 = 2$ .

#### 4.2. Forecasting Evaluation Methods

For the purpose of evaluating the forecasting capability, we examine the forecasting accuracy by calculating three different statistical metrics, the root mean square error (RMSE), the mean absolute error (MAE) and the mean absolute percentage error (MAPE). The definitions of RMSE, MAE and MAPE are expressed as Equations (15–17):

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - A_i)^2}{n}} \quad (15)$$

$$MAE = \sqrt{\frac{\sum_{i=1}^n |P_i - A_i|}{n}} \quad (16)$$

$$MAPE = \sqrt{\frac{\sum_{i=1}^n \left| \frac{P_i - A_i}{A_i} \right|}{n}} * 100 \quad (17)$$

Where  $P_i$  and  $A_i$  are the  $i$ -th predicted and actual values respectively, and  $n$  is the total number of predictions.

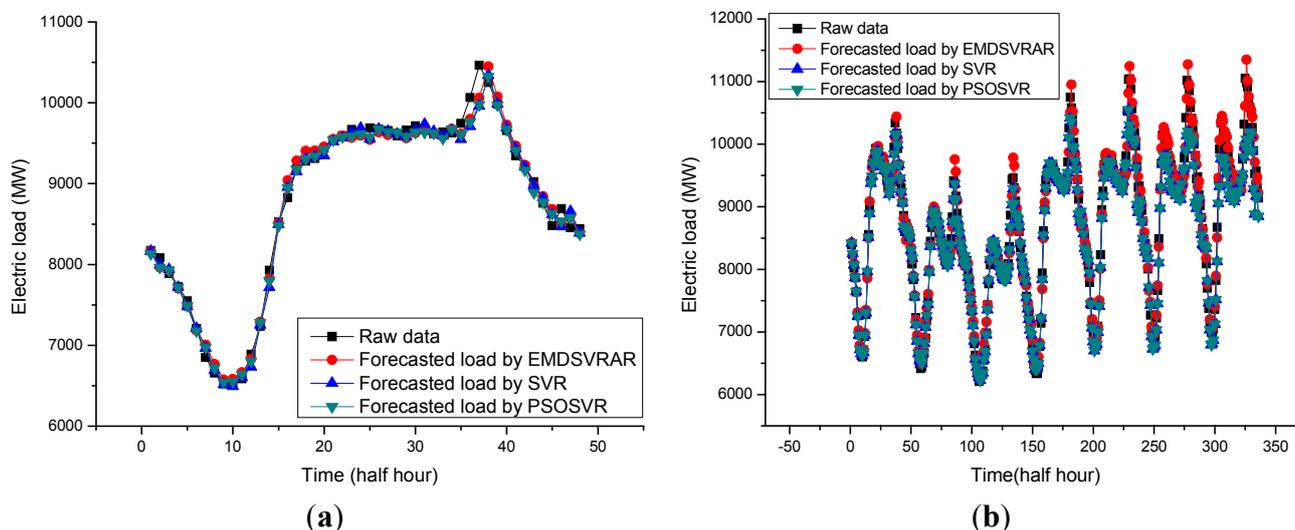
#### 4.3. Empirical Results and Analysis

For the first experiment, the forecasting results (the electric load on 8 May 2007) of the original SVR model, the PSO-SVR model and the proposed EMDSVRAR model are shown in Figure 7a. Notice that the forecasting curve of the proposed EMDSVRAR model fits better than other alternative models.

The second experiment shows the one-week-ahead forecasting for the large sample size data. The peak load values of testing set are bigger than that of training set shown in Figure 3b. The detailed forecasted results of this experiment are shown in Figure 7b. It indicates that the results obtained from the EMDSVRAR model fits the peak load values exceptionally well. In other words, the EMDSVRAR model has better generalization ability than the three comparison models.

The forecasting results from these models are summarized in Table 3. The proposed EMDSVRAR model is compared with four alternative models. It is found 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.

**Figure 7.** Comparison of the original data and the forecasted electric load by the EMDSVRAR Model, the SVR model and the PSO-SVR model for (a) the small sample size (One-day ahead prediction of 8 May 8, 2007 are performed by the models); (b) the large sample size (One-week ahead prediction from May 18, 2007 May 24, 2007 are performed by the models).



The proposed model shows the higher forecasting accuracy in terms of three different statistical metrics. In view of the model effectiveness and efficiency on the whole, we can conclude that the proposed model is quite competitive against four comparison models, the PSO-BP, SVR, PSO-SVR, and AFCM models. In other words, the hybrid model leads to better accuracy and statistical interpretation.

**Table 3.** Summary of results of the forecasting models.

Algorithm	MAPE	RMSE	MAE	Running Time(s)
For the first experiment (small sample size)				
Original SVR	11.6955	145.865	10.9181	180.4
PSO-SVR	11.4189	145.685	10.6739	165.2
PSO-BP	10.9094	142.261	10.1429	159.9
AFCM <sup>[24]</sup>	9.9524	125.323	9.2588	75.3
EMDSVRAR	9.8595	117.159	9.0967	80.7
For the second experiment (large sample size)				
Original SVR	12.8765	181.617	12.0528	116.8
PSO-SVR	13.503	271.429	13.0739	192.7
PSO-BP	12.2384	175.235	11.3555	163.1
AFCM <sup>[26]</sup>	11.1019	158.754	10.4385	160.4
EMDSVRAR	5.100	134.201	9.8215	162.0

Several observations can also be noticed from the results. Firstly, from the comparisons among these models, we point out that the proposed model outperforms other alternative models. Secondly, the EMDSVRAR 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. Finally, since the proposed model generates good results with good accuracy and interpretability, it is robust and effective as shown in Table 3. Overall, the proposed model provides a very powerful tool to implement easily for forecasting electric load.

Furthermore, to verify the significance of the accuracy improvement of the EMDSVRAR model, the forecasting accuracy comparison among original SVR, PSO-SVR, PSO-BP, AFCM, and EMDSVRAR models is conducted by a statistical test, namely a Wilcoxon signed-rank test, at the 0.025 and 0.05 significance levels in one-tail-tests. The test results are shown in Table 4. Clearly, the proposed EMDSVRAR model has statistical significance (under a significant level 0.05) among the other alternative models, particularly comparing with original SVR, PSO-SVR, PSO-BP, and AFCM models.

**Table 4.** Wilcoxon signed-rank test.

Compared models	Wilcoxon signed-rank test	
	$\alpha = 0.025; W = 4$	$\alpha = 0.05; W = 6$
EMD-SVR-AR vs. original SVR	8	3 <sup>a</sup>
EMD-SVR-AR vs. PSO-SVR	6	2 <sup>a</sup>
EMD-SVR-AR vs. PSO-BP	6	2 <sup>a</sup>
EMD-SVR-AR vs. AFCM	6	2 <sup>a</sup>

<sup>a</sup> denotes that the EMDSVRAR model significantly outperforms other alternative models.

## 5. Conclusions

The proposed model achieves superiority and outperforms the original SVR model while forecasting based on the unbalanced data. In addition, the goal of the training model is not to learn an exact representation of the training set itself, but rather to set up a statistical model that generalizes better forecasting values for the new inputs. In practical applications of a SVR model, if the SVR model is overtrained to some sub-classes with overwhelming size, it memorizes the training data and gives poor generalization of other sub-classes with small size. The EMD term of the proposed EMDSVRAR model has been employed in the present research, details of which have discussed in the above section.

The interest in applying the EMD forecast systems arises from the fact that those systems consider both accuracy and comprehensibility of the forecast result simultaneously. To this end, a combined model has been proposed and its effectiveness in forecasting the electric load data has been compared with three other alternative models. In this study, various data characteristics of electric load are identified where the proposed model performs better than the other algorithms in terms of its forecasting capability. Based on the obtained experimental results, we conclude that the proposed EMDSVRAR model algorithm can generate not only human-understandable rules, but also better forecasting accuracy levels. Our proposed model also outperforms other alternative models in terms of interpretability, forecasting accuracy and generalization ability, which are especially true for forecasting with unbalanced data and very complex systems.

## Acknowledgments

This work was supported by National Natural Science Foundation of China under Contract (No. 51064015) to which the authors are greatly obliged, and National Science Council, Taiwan (NSC 100-2628-H-161-001-MY4; NSC 101-2410-H-161-001).

## 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.
2. Bianco, V.; Manca, O.; Nardini, S. Electricity consumption forecasting in Italy using linear regression models. *Energy* **2009**, *34*, 1413–1421.
3. Zhou, P.; Ang, B.W.; Poh, K.L. A trigonometric grey prediction approach to forecasting electricity demand. *Energy* **2006**, *31*, 2839–2847.
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.
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.
6. Topalli, A.K.; Erkmén, I. A hybrid learning for neural networks applied to short term load forecasting. *Neurocomputing* **2003**, *51*, 495–500.
7. 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.
8. 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.
9. Topalli, A.K.; Cellura, M.; Erkmén, I.; Topalli, I. Intelligent short-term load forecasting in Turkey. *Electr. Power Energy Syst.* **2006**, *28*, 437–447.
10. 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.
11. Hong, W.C. Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm. *Energy* **2011**, *36*, 5568–5578.
12. Hong, W.C. Application of chaotic ant swarm optimization in electric load forecasting. *Energy Policy* **2010**, *38*, 5830–5839.
13. Pai, P.F.; Hong, W.C. Support vector machines with simulated annealing algorithms in electricity load forecasting. *Energy Convers. Manag.* **2005**, *46*, 2669–2688.
14. Yongli, Z.; Hogg, B.W.; Zhang, W.Q.; Gao, S.; Yang, Y.H. Hybrid expert system for aiding dispatchers on bulk power systems restoration. *Int. J. Electr. Power Energy Syst.* **1994**, *16*, 259–268.
15. Basak, D.; Pal, S.; Patranabis, D.C. Support vector regression. *Neural Inf. Process. Lett. Rev.* **2007**, *11*, 203–224.
16. Burges, C.J.C. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.* **1998**, *2*, 121–167.
17. Schölkopf, B.; Smola, A.; Williamson, R.C.; Bartlett, P.L. New support vector algorithms. *Neural Comput.* **2000**, *12*, 1207–1245.
18. Meng, Q.; Peng, Y. A new local linear prediction model for chaotic time series. *Phys. Lett. A* **2007**, *370*, 465–470.
19. Xie, J.X.; Cheng, C.T. A new direct multi-step ahead prediction model based on EMD and chaos analysis. *J. Autom.* **2008**, *34*, 684–689.

20. 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**, doi:10.1155/2012/720849.
21. Bhusana, P.; Chris, T. Improving prediction of exchange rates using differential EMD. *Expert Syst. Appl.* **2013**, *40*, 377–384.
22. Huang, N.E.; Shen, Z. A new view of nonlinear water waves: The Hilbert spectrum. *Rev. Fluid Mech.* **1999**, *31*, 417–457.
23. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer-Verlag: New York, NY, USA, 1995.
24. Gao, J. Asymptotic properties of some estimators for partly linear stationary autoregressive models. *Commun. Statist. Theory Meth.* **1995**, *24*, 2011–2026.
25. Taylor, J.W. Short-term load forecasting with exponentially weighted methods. *IEEE Trans. Power Syst.* **2012**, *27*, 458–464.
26. 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.

© 2013 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).