A Short-Term Load Forecasting Model with a Modified Particle Swarm Optimization Algorithm and Least Squares Support Vector Machine Based on the Denoising Method of Empirical Mode Decomposition and Grey Relational Analysis

Dongxiao Niu and Shuyu Dai *

School of Economics and Management, North China Electric Power University, Beijing 102206, China; niudx@ncepu.edu.cn

* Correspondence: DaiShuyu@ncepu.edu.cn; Tel.: +86-010-61773472

Abstract: As an important part of power system planning and the basis of economic operation of power systems, the main work of power load forecasting is to predict the time distribution and spatial distribution of future power loads. The accuracy of load forecasting will directly influence the reliability of the power system. In this paper, a novel short-term Empirical Mode Decomposition-Grey Relational Analysis-Modified Particle Swarm Optimization-Least Squares Support Vector Machine (EMD-GRA-MPSO-LSSVM) load forecasting model is proposed. The model uses the de-noising method combining empirical mode decomposition and grey relational analysis to process the original load series and forecasts the processed subsequences by the algorithm of modified particle swarm optimization and least square support vector machine. Then, the final forecasting results can be obtained after reconstructing the forecasting series. This paper takes the Jibei area as an example to produce an empirical analysis for load forecasting. The model input includes the hourly load one week before the forecasting day and the daily maximum temperature, daily minimum temperature, daily average temperature, relative humidity, wind force, date type of the forecasting day. The model output is the hourly load of the forecasting day. The models of BP neural network, SVM (Support vector machine), LSSVM (Least squares support vector machine), PSO-LSSVM (Particle swarm optimization-Least squares support vector machine), MPSO-LSSVM (Modified particle swarm optimization-Least squares support vector machine), EMD-MPSO-LSSVM are selected to compare with the model of EMD-GRA-MPSO-LSSVM using the same sample. The comparison results verify that the short-term load forecasting model of EMD-GRA-MPSO-LSSVM proposed in this paper is superior to other models and has strong generalization ability and robustness. It can achieve good forecasting effect with high forecasting accuracy, providing a new idea and reference for accurate short-term load forecasting.

Keywords: short-term load forecasting; empirical mode decomposition; grey relational analysis; modified particle swarm optimization algorithm; least square support vector machine

1. Introduction

With the continuous development of the national economy, the demand for electricity is increasing day by day. The power industry plays an increasingly important role in social development. As an important part of power system planning and the basis of economic operation of power systems, the main task of power load forecasting is to predict the time distribution and spatial distribution of

future power loads. The accuracy of load forecasting will directly influence the reliability of the power system, and then influence the national economy. Short-term load forecasting is to predict the load of the next 24 h to a few days according to historical load, weather, holidays and other factors. Short-term loads present randomness, whose forecasting accuracy is sensitive to noise. Therefore, it is significant to select appropriate forecasting algorithms and denoising methods and construct a reasonable and effective forecasting model for accurate short-term load forecasting.

In recent years, with the development of artificial intelligence technology, artificial intelligence forecasting technology has been gradually replacing the classical forecasting technology in the field of load forecasting. Compared with the classical forecasting technology, artificial intelligence forecasting technology has obvious forecasting accuracy advantages. The neural network algorithm is a typical and widely used artificial intelligence forecasting technology. Artificial neural network (ANN) is a theoretical human brain neural network mathematical model which originated in the 1940s. It is connected by a large number of neurons with adjustable connection weights, and has the characteristics of large-scale parallel processing, distributed information storage, good self-learning ability and so on [1–3]. BP (Backpropagation) neural network [4,5] and RBF neural network [6–8] are common neural networks. BP neural network and RBF neural network are both nonlinear multilayer feedforward networks, but they are different in network structure, training algorithm, network resource utilization and approximation performance. Li and Huang [4] proposed a short-term power load forecasting method based on a BP neural network. The method firstly reduces the factors which affect load forecasting by using rough set theory, then takes the reduced data as input variables of the BP neural network model, and gets the forecast value by using back-propagation algorithm. Li et al. [6] proposed a dynamic RBF neural network model for load forecasting. By introducing the error discriminant function, the load forecasting error is controlled, and the prediction model is modified dynamically. In general, neural network algorithm has strong nonlinear fitting ability, which does well in nonlinear function approximation. It has simple learning rules and is easy to implement by computer. However the neural network algorithm also has some shortcomings, such as slow convergence speed, long training time, easily getting trapping into local optima, overfitting and so on [9]. Support vector machine (SVM) is another artificial intelligence forecasting technology, which is a machine learning method for small samples based on the VC dimension theory of statistical learning theory and structural risk minimization. Compared with the neural network algorithm the SVM algorithm has advantages of good robustness, strong generalization ability, fast computation speed and avoiding falling into local optima [10,11]. Least squares support vector machine (LSSVM) is an improved algorithm for the standard support vector machine using least squares value function and equality constraint [12,13]. It has faster training speed and better convergence precision [14–16]. When we use the least squares support vector machine to forecast the load, the regularization parameter and the radial basis kernel function parameter need to be set, which directly influence the forecasting accuracy. Therefore, the researchers use a variety of methods to optimize the parameters of LSSVM, such as genetic algorithm (GA) [17], ant colony optimization algorithm (ACO) [18], artificial bee colony algorithm (ABC) [19], differential evolution algorithm (DE) [20], bat algorithm(BA) [21], principal component analysis [22]. Mahjoob et al. [17] used a least-squares support vector machines approach, in combination with a hybrid genetic algorithm based optimization to forecast market clearing prices in two different power markets. The forecasting results are compared to a select variety of previously proposed methods such as MLP, ARIMA, wavelet-ARIMA, neuro-fuzzy and time series based models. Liu and Jiang [18] proposed a novel PHM algorithm based on the ACO and LSSVM algorithms. They used a time series analysis prediction method to calculate seasonal factors combined with a ACO-LSSVM algorithm to model the failure rate of an aviation device and obtained good experimental results. Yusof et al. [19] proposed a forecasting model based on an optimized Least Squares Support Vector Machine. The determination of hyper-parameters is performed using a nature-inspired algorithm, the Artificial Bee Colony. The proposed forecasting model is realized in gold price forecasting. Qiao et al. [20] built a forecasting model of natural gas daily load based on wavelet transform and Least Squares Support

Vector Machine-Differential Evolution (LSSVM-DE), which provides a practical and feasible method for predicting the time series of natural gas daily loads. The algorithms above provide some help for the optimization of LSSVM parameters, but their obvious shortcomings of low search efficiency, slow convergence speed and easily getting trapping into local optima can't be ignored. In this paper, a modified particle swarm optimization algorithm with averaging particle distance and mutation operator is proposed to optimize the parameters of least squares support vector machine. Particle swarm optimization (PSO) is a computation technique originated from the study of bird predatory behavior, proposed by Eberhart and Kennedy in 1995 [23–27]. Zhu and Wang [26] put forward the power load forecasting method between LSSVM and PSO optimization algorithm. By taking historical load data, and meteorological factors, etc. as input, they built predictive models to forecast the future power load. Sun and Yao [27] proposed a least squares support vector machine (LSSVM) model optimized by the particle swarm optimization (PSO).The phase space of the chaotic wind speed time series is reconstructed by calculating the embedding dimension and the delay time of the wind speed time series. Particle swarm optimization algorithm has the advantages of less adjustment parameters, high search efficiency and fast convergence speed, and its global search ability will be greatly enhanced after modified by averaging particle distance and mutation operator, which can effectively avoid falling into local optimum.

In order to forecast the short-term load which is sensitive to external factors more accurately, some researchers use time series decomposition techniques to decompose the original load sequence into several subsequences. The subsequences are predicted separately and then restructured to be the final forecasting series. The forecasting accuracy is greatly improved in this way. Wavelet decomposition [28–33] and classical model decomposition [34–39] are two commonly used time series decomposition methods. Xia et al. [32] gave a method which is based on the wavelet decomposition and the neural network to predict the short-time load. After decomposing the load sequence into subsequences on different scales by using wavelet transform, forecasting and restructuring the subsequences, then the final forecasting results can be obtained. Wang and Li [39] proposed a forecasting method based on empirical mode decomposition (EMD) and least squares support vector machine (LSSVM). Considering the power characteristics, unit efficiency and the operation condition of the generators, the one-month ahead forecasted output power of the wind power plant can be obtained. The wavelet decomposition has low resolution and is sensitive to the selection of wavelet bases, but the empirical mode decomposition has high resolution and the characteristic of self-adaptation with no need to select complex wavelet bases. Therefore, EMD is selected to decompose the load series in this paper. Considering that the short-term load has some noise, we combine Empirical Mode Decomposition with Grey Relational Analysis [40–42] to process the original load series for de-noising, which will further improve the accuracy of load forecasting.

In this paper, a hybrid EMD-GRA-MPSO-LSSVM model with high accuracy for short-term load forecasting is proposed. The main content and structure of this paper are as follows: the second section introduces the forecasting algorithm—MPSO-LSSVM—using a particle swarm optimization algorithm modified by averaging particle distance and mutation operator to optimize the parameters of the least squares support vector machine. The third section describes the de-noising method of EMD-GRA, using EMD to decompose the original load series, then removing the noise subsequently by GRA. The fourth section introduces the forecasting process based on the hybrid EMD-GRA-MPSO-LSSVM model. The fifth section gives an empirical analysis to verify the practicability, reliability and validity of the EMD-GRA-MPSO-LSSVM model. The sixth section summarizes the work.

2. Forecasting Algorithm

2.1. LSSVM

The least squares support vector machine (LSSVM) is an improved version of the standard support vector machine algorithm first proposed by Suykens et al. It can transform quadratic program

problems into linear problems through least squares value functions and equality constraints. It has faster training speed and better convergence precision [14,16].

When we use LSSVM to solve the regression optimization problem, the loss function in the optimization objective function is the quadratic term of the error. The constraint condition is the equality constraint, as shown in Equation (1):

$$\min(\frac{1}{2}\|\omega\|^2 + \frac{1}{2}c\sum_{i=1}^{n}\xi_i^2)$$

$$\text{s.t.} y_i = \omega^T \varphi(x_i) + b + \xi_i \tag{1}$$

where ω is the weight vector of high dimensional feature space. $\|\omega\|^2$ controls the complexity of the model. c is the regularization parameter. ξ_i is the error. b is the bias constant. Introducing the Lagrange multiplier, Equation (1) can be converted into:

$$\min L = \frac{1}{2}\|\omega\|^2 + \frac{1}{2}c\sum_{i=1}^{n}\xi_i^2 - \sum_{i=1}^{n}\alpha_i(\omega^T\varphi(x_i) + b + \xi_i - y_i) \tag{2}$$

According to the condition of Karush-Kuhn-Tucker, what we can obtain is as follows:

$$\begin{cases} \frac{\partial L}{\partial \omega} = 0 \rightarrow \omega = \sum_{i=1}^{n} \alpha_i \varphi(x_i) \\ \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{n} \alpha_i = 0 \\ \frac{\partial L}{\partial \xi_i} = 0 \rightarrow \alpha_i = c\xi_i, i = 1, 2, \ldots, n \\ \frac{\partial L}{\partial \alpha_i} = 0 \rightarrow \omega^T \varphi(x_i) + b + \xi_i - y_i = 0, i = 1, 2, \ldots, n \end{cases} \tag{3}$$

After eliminating ω and ξ, the linear system of equations are described as follows:

$$\begin{bmatrix} 0 & 1 & \cdots & 1 \\ 1 & K(x_1, x_1) + \frac{1}{c} & \cdots & K(x_1, x_n) \\ \cdots & \cdots & \cdots & \cdots \\ 1 & K(x_n, x_1) & \cdots & K(x_n, x_n) + \frac{1}{c} \end{bmatrix} \begin{bmatrix} b \\ \alpha_1 \\ \cdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} 0 \\ y_i \\ \cdots \\ y_n \end{bmatrix} \tag{4}$$

where $K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$ is the kernel function that meets the condition of Mercer. α_i and b are obtained by least square method, and the nonlinear forecasting model is shown in Equation (5):

$$y = \sum_{i=1}^{n} \alpha_i K(x, x_i) + b \tag{5}$$

2.2. MPSO

2.2.1. Standard PSO

Particle swarm optimization (PSO) is a swarm intelligence optimization algorithm based on iterative computation, originated from the study of bird predatory behavior, first proposed by Kennedy and Eberhart [23]. The PSO algorithm initializes a group of random particles in a solvable space and each particle represents a potential optimal solution to the optimization problem. The characteristics of particles contain location, velocity and fitness value. Fitness value calculated by fitness function is the index for evaluating the advantages and disadvantages of particles. In each iteration, the particles update their velocity and next position by tracking the individual extremum position and the global extremum position.

Suppose that n particles form a population in a d dimensional search space. The i particle is represented as a d dimensional vector $X_i = (x_{i1}, x_{i2}, \cdots, x_{id})^T$, $i = 1, 2, ..., n$, that is the position of the i particle in the d dimension search space.

The velocity of the i particle is $V_i = (v_{i1}, v_{i2}, \cdots, v_{id})^T$. The current best location of the i particle is $P_i = (p_{i1}, p_{i2}, \cdots, p_{id})^T$. The current best location of the whole population is $P_g = (p_{g1}, p_{g2}, \cdots, p_{gd})^T$. In each iteration, the particles update their velocity and position by Equations (6) and (7):

$$V_{id}^{k+1} = \omega V_{id}^k + c_1 r_1 (P_{id}^k - X_{id}^k) + c_2 r_2 (P_{gd}^k - X_{id}^k) \tag{6}$$

$$X_{id}^{k+1} = X_{id}^k + V_{id}^{k+1} \tag{7}$$

where k is the current iteration number. ω is the inertia weight. c_1 and c_2 are acceleration factors. r_1 and r_2 are random numbers within [0, 1] and the same random numbers are used in every dimension.

2.2.2. Modified PSO

For complex optimization problems, the standard particle swarm optimization algorithm is prone to premature convergence. In order to solve this problem, averaging particle distance and mutation operator are introduced to guide the selection of initial population and judging particle premature convergence for avoiding falling into local optimum [24,43].

The initial particle swarm selection is random. Ideally, the finite particles should be uniformly distributed over the entire solution space. However, the ideal state is difficult to achieve. So averaging particle distance is applied to increase the probability of searching the global optimal solution and avoid trapping into a local optimum. Averaging particle distance is defined as follows:

$$D(t) = \frac{1}{mL} \sum_{i=1}^{m} \sqrt{\sum_{d=1}^{n} (x_{id} - \overline{x_d})^2} \tag{8}$$

where $D(t)$ is the discrete degree of particle distribution in the population. L is the maximum diagonal length of the search space. m is the number of particles. n is the dimension of the solution space. x_{id} represents the d dimensional coordinate value of the i particle. $\overline{x_d}$ represents the mean value of the d dimensional coordinate values of all particles.

The standard particle swarm optimization algorithm converges fast in the early stage of operation and slows down later. It's easily trapped in a local optimum and loses the ability of population evolution. The particle location determines its fitness, so the current state of the population and the premature convergence of the particles can be judged according to the overall variation of the fitness of all particles in the population. Suppose the current fitness of the i particle is f_i. The current average fitness of the population is $\overline{f}$. The fitness variance of the population is defined as follows:

$$\sigma^2 = \sum_{i=1}^{m} \left(\frac{f_i - \overline{f}}{f} \right)^2 \tag{9}$$

where m is the number of the particles in the population. f is the normalized scaling factor for limiting σ^2, which can be calculated by Equation (10):

$$f = \begin{cases} \max \left| f_i - \overline{f} \right|, & \max \left| f_i - \overline{f} \right| > 1 \\ 1, & \max \left| f_i - \overline{f} \right| \leq 1 \end{cases} \tag{10}$$

The fitness variance reflects the discrete degree of particles in the population. The smaller the σ^2 is, the smaller the discrete degree of particles is. Otherwise, the larger the discrete degree is.

With the increasing numbers of iterations, $D(t)$ and σ^2 are getting smaller. When $D(t) < \alpha$ and $\sigma^2 < \beta$ (α, β belong to a certain threshold), the algorithm is considered to get into the late search stage

and the population appears to prematurely converge. We retain the current optimal location of the particle swarm and redistribute the particle solution space to guide the particles to jump out of the local optimum.

2.3. MPSO-LSSVM

When we use the least squares support vector machine with RBF kernel function to forecast the load, the regularization parameter c and the radial basis kernel function parameter g need to be set, which directly influences the forecasting accuracy [26,27]. In this paper, the modified particle swarm optimization algorithm is proposed to optimize the parameters of least squares support vector machine. The optimization steps are as follows:

(1) Set the parameters of the PSO and initialize the particle swarm.
(2) Calculate the fitness value of each particle and find the current individual extreme position and global extreme position.
(3) Calculate the averaging particle distance and mutation operator of the population and judge whether the particles fall into the premature convergence state. If the particles fall into local optimum, redistribute the particle solution space to guide the particles to jump out of local optimum.
(4) Update the velocity and position of the particles to generate new species. Calculate the fitness values and compare them with historical optimal value. Update the individual extreme position and the global extreme position and cycle this process until the number of iterations is reached. Then, the optimization results can be obtained.
(5) Forecast the load by LSSVM based on optimal parameters.

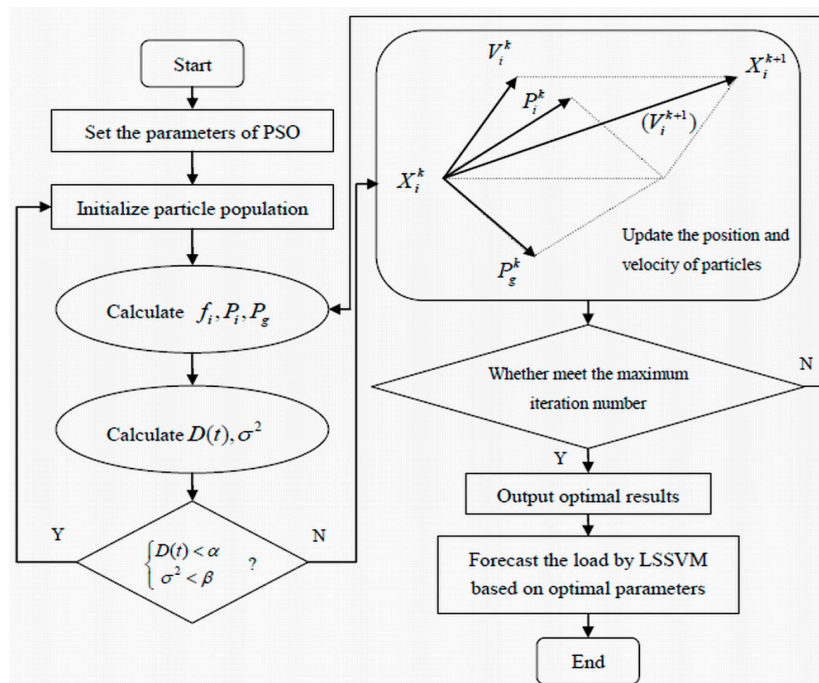The process of LSSVM parameters optimization based on MPSO is described in Figure 1.

Figure 1. The flow chart of LSSVM parameters optimization based on MPSO.

3. De-Noising Method

3.1. EMD

Empirical Mode Decomposition is the core algorithm of Hilbert-Huang transform (HHT), proposed by Huang et al. in 1998 [34]. EMD is a new adaptive signal time-frequency processing method. It can decompose complex signals into finite intrinsic mode functions (IMF) according to the time scale characteristics of data. The decomposed IMF components contain local characteristic signals with different time scales of the original signal. The IMF must meet the conditions that the difference between the extremum number and the zero number of the signal is 0 or 1 and the average value of the upper envelope and the lower envelope of the signal is zero [37–39]. For a given signal, the EMD steps are as follows:

(1) Find all the local extreme of $X(t)$ and fit the upper envelope $f_a(t)$ and the lower envelope $f_b(t)$ of $X(t)$ by cubic spline function.
(2) Calculate the average value $f_m(t)$ of the upper envelope and lower envelope.
(3) Calculate the difference between $X(t)$ and $f_m(t)$, $E(t) = X(t) - f_m(t)$.
(4) Repeat the steps (1) ~(3) regarding $E(t)$ as the original series. When the mean envelope $f_m(t)$ tends to zero, the first IMF component $imf_1(t)$ is obtained.
(5) Let $X_1(t) = X(t) - imf_1(t)$. Repeat the steps (1)~(4) regarding $X_1(t)$ as the original series, then the second IMF component $imf_2(t)$ is obtained. Repeat this process until the difference function is a constant function or a monotone function. Then, the original series can be described as Equation (11):

$$X(t) = \sum_{i=1}^{n} imf_i(t) + X_n(t) \tag{11}$$

3.2. GRA

Grey relational analysis (GRA) is a multivariate statistical analysis method, which completes the comparison of the geometric relation of the statistical data in the time series of the system through the quantitative analysis of the dynamic development trend to obtain the grey relational degrees between the reference series and the comparison series [40,41]. The larger the grey relational degree is, the closer the development direction and speed of the series are, so does the relationship between the series. The core of GRA is to calculate the grey relational degree, and the specific calculation steps are as follows:

(1) Determine the analysis series.

Set the reference series:

$$X_0 = (x_0(1), x_0(2), \cdots, x_0(n)) \tag{12}$$

Set the comparison series:

$$\begin{cases} X_1 = (x_1(1), x_1(2), \cdots, x_1(n)) \\ X_2 = (x_2(1), x_2(2), \cdots, x_2(n)) \\ \qquad \cdots \\ X_m = (x_m(1), x_m(2), \cdots, x_m(n)) \end{cases} \tag{13}$$

(2) Standardize the data.

$$y_i(k) = \frac{x_i(k)}{\frac{1}{n} \sum_{k=1}^{n} x_i(k)} \tag{14}$$

$$k = 1, 2, \cdots, n, i = 0, 1, 2, \cdots, m$$

(3)   Calculate the relevant coefficient.

The correlation coefficient between $y_j(k)$ and $y_0(k)$ is shown in Equation (15):

$$\xi_j(k) = \frac{\min\limits_{j}\min\limits_{k}\left|y_0(k) - y_j(k)\right| + \rho\max\limits_{j}\max\limits_{k}\left|y_0(k) - y_j(k)\right|}{\left|y_0(k) - y_j(k)\right| + \rho\max\limits_{j}\max\limits_{k}\left|y_0(k) - y_j(k)\right|}$$

$$\rho \in (0,1), k = 1, 2, \cdots, n, j = 1, 2, \cdots, m$$

(15)

(4)   Calculate the relational degree.

The grey relational degree between X_j and X_0 is shown in Equation (16):

$$r_j = \frac{1}{n}\sum_{k=1}^{n}\xi_j(k)$$

$$k = 1, 2, \cdots, n, j = 1, 2, \cdots, m$$

(16)

(5)   Rank the relational degree.

Rank the comparison series according to the grey relational degree. The larger the grey relational degree is, the more consistent the variation trends between the comparison series and the reference series are.

3.3. EMD-GRA

In the load forecasting, the original historical load data often contain noise pollution, which makes it difficult to accurately analyze the actual load change rules. So, the original load data need to be processed for de-noising. EMD-GRA is a data processing model combining empirical mode decomposition and grey relational analysis for de-noising. Its core idea is to decompose the non-stationary signal to obtain a set of stationary intrinsic mode functions and eliminate useless components in the signal [37]. The basic steps of EMD-GRA de-noising method are as follows:

(1)   Decompose the original time series by EMD and obtain finite intrinsic mode functions.
(2)   Calculate the grey relational degree between each IMF and the original series and rank the relational degree.
(3)   Eliminate the IMF with the lowest relational degree and reconstruct the rest IMFs. Then, get the de-noising time series.

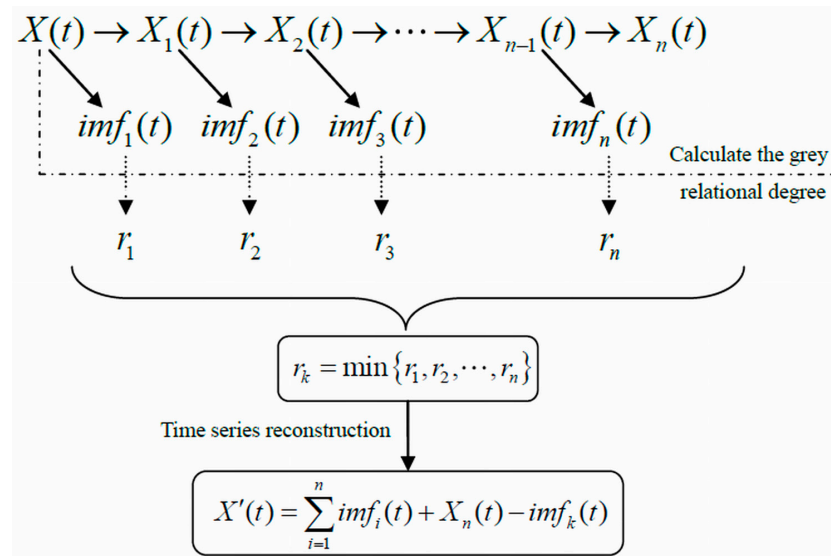The process of EMD-GRA de-noising is described in Figure 2.

$$X(t) \rightarrow X_1(t) \rightarrow X_2(t) \rightarrow \cdots \rightarrow X_{n-1}(t) \rightarrow X_n(t)$$

$$imf_1(t) \quad imf_2(t) \quad imf_3(t) \qquad imf_n(t)$$

Calculate the grey relational degree

$$r_1 \qquad r_2 \qquad r_3 \qquad r_n$$

$$r_k = \min\{r_1, r_2, \cdots, r_n\}$$

Time series reconstruction

$$X'(t) = \sum_{i=1}^{n} imf_i(t) + X_n(t) - imf_k(t)$$

Figure 2. The flow chart of EMD-GRA de-noising.

4. The Forecasting Model of EMD-GRA-MPSO-LSSVM

The accuracy of short term load forecasting is influenced by many factors, such as: temperature, humidity, wind force and holidays. In order to forecast the next 24 h load more accurately, the load forecasting model of Emd-GRA-MPSO-LSSVM considering weather and date type is proposed in this paper. The forecasting steps are as follows:

(1)   Data collecting and processing

Collecting sample data, including historical load, daily maximum temperature, daily minimum temperature, daily average temperature, relative humidity, wind force, date type, etc. Then, process the data. Normalize the weather data and use 1 and 0 to represent holidays and work days respectively for day type.

(2)   De-noising processing

Decompose the original time series by EMD. Calculate the grey relational degree between each IMF and the original series and rank the relational degree. Then, eliminate the IMF with the lowest relational degree.

(3)   Load forecasting

Use the MPSO-LSSVM model to forecast the IMFs which are processed by the EMD-GRA de-noising method and reconstruct the forecasting series. Then, get the final forecasting results.
The model of EMD-GRA-MPSO-LSSVM is shown in Figure 3.

5. Empirical Analysis

5.1. Sample Selection

The Jibei area is located in the north of Hebei Province, China, and contains Tangshan City, Langfang City, Chengde City, Zhangjiakou City and Qinhuangdao City. In this paper, we select the hourly load, daily maximum temperature, daily minimum temperature, daily average temperature, relative humidity, wind force, date type of the Jibei area from 1 January to 14 March 2016 as the training sample to forecast the hourly load in 15 March. The model input includes the hourly load one week

before the forecasting day and the daily maximum temperature, daily minimum temperature, daily average temperature, relative humidity, wind force, date type of the forecasting day where the weather data take the average value of the five cities in Jibei area. The model output is the hourly load of the forecasting day.
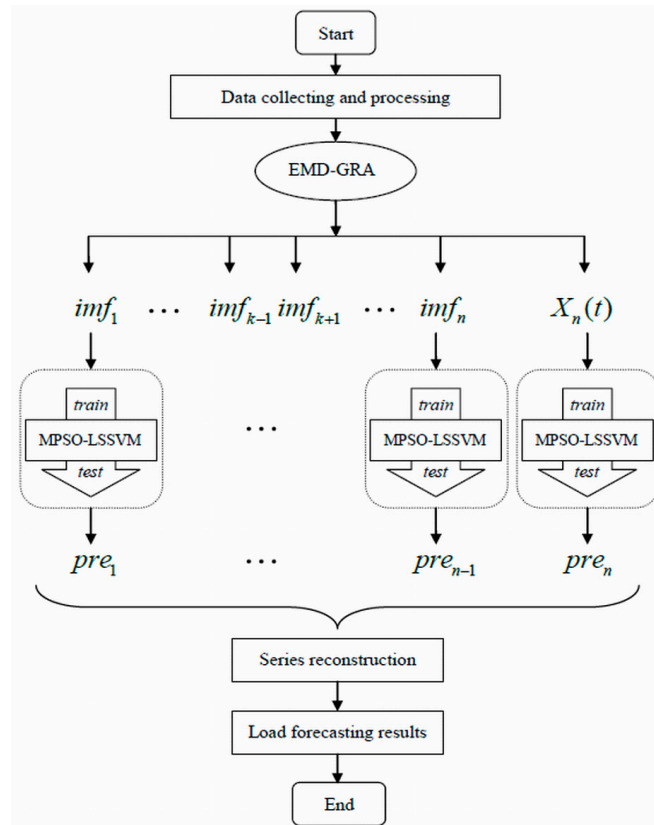

Figure 3. The model of EMD-GRA-MPSO-LSSVM.

5.2. Load Forecasting Based on EMD-GRA-MPSO-LSSVM

Decompose the original historical load series by EMD. Input the signal series containing 1800 h load from 1 January to 15 March 2016 in Jibei area into the EMD model and obtain 9 IMFs and 1 residual. The decomposition results are shown in Figure 4.

Calculate the grey relational degree between each IMF and the original series and rank the relational degree. The rank results are shown in Figure 5.

As shown in Figure 5, the grey relational degree between the fourth IMF and the original series is minimum. Eliminate the fourth IMF and use the MPSO-LSSVM model to forecast the rest IMFs and the residual. The main parameters of MPSO-LSSVM are shown in Table 1.

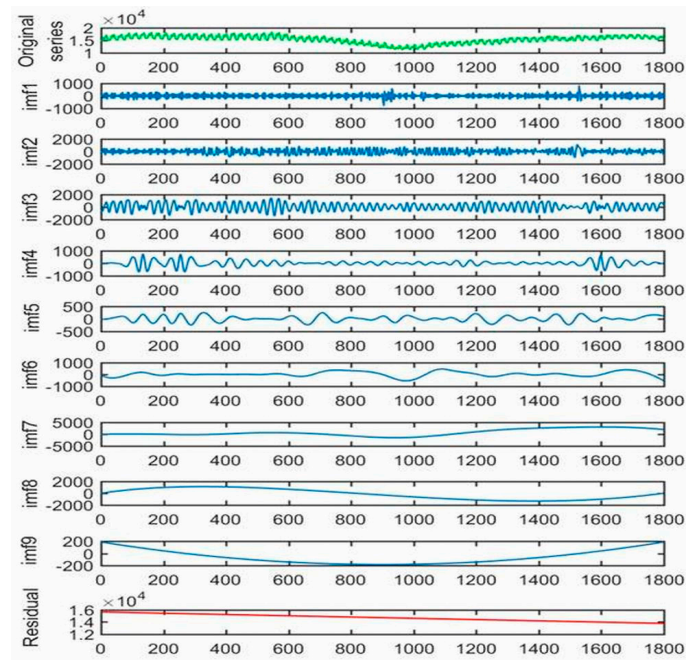The forecasting results of each series are shown in Figure 6.
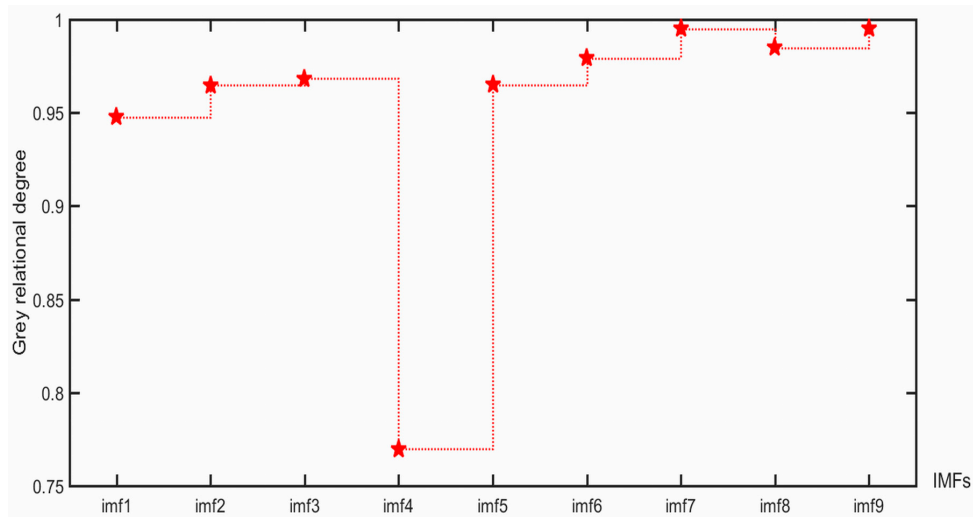
Figure 4. EMD IMFs.

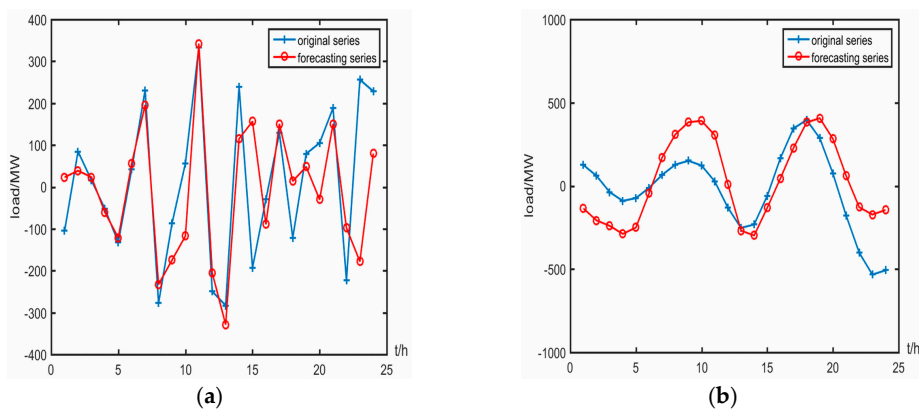**Figure 5.** The grey relational degree ranking.

Figure 6. *Cont.*
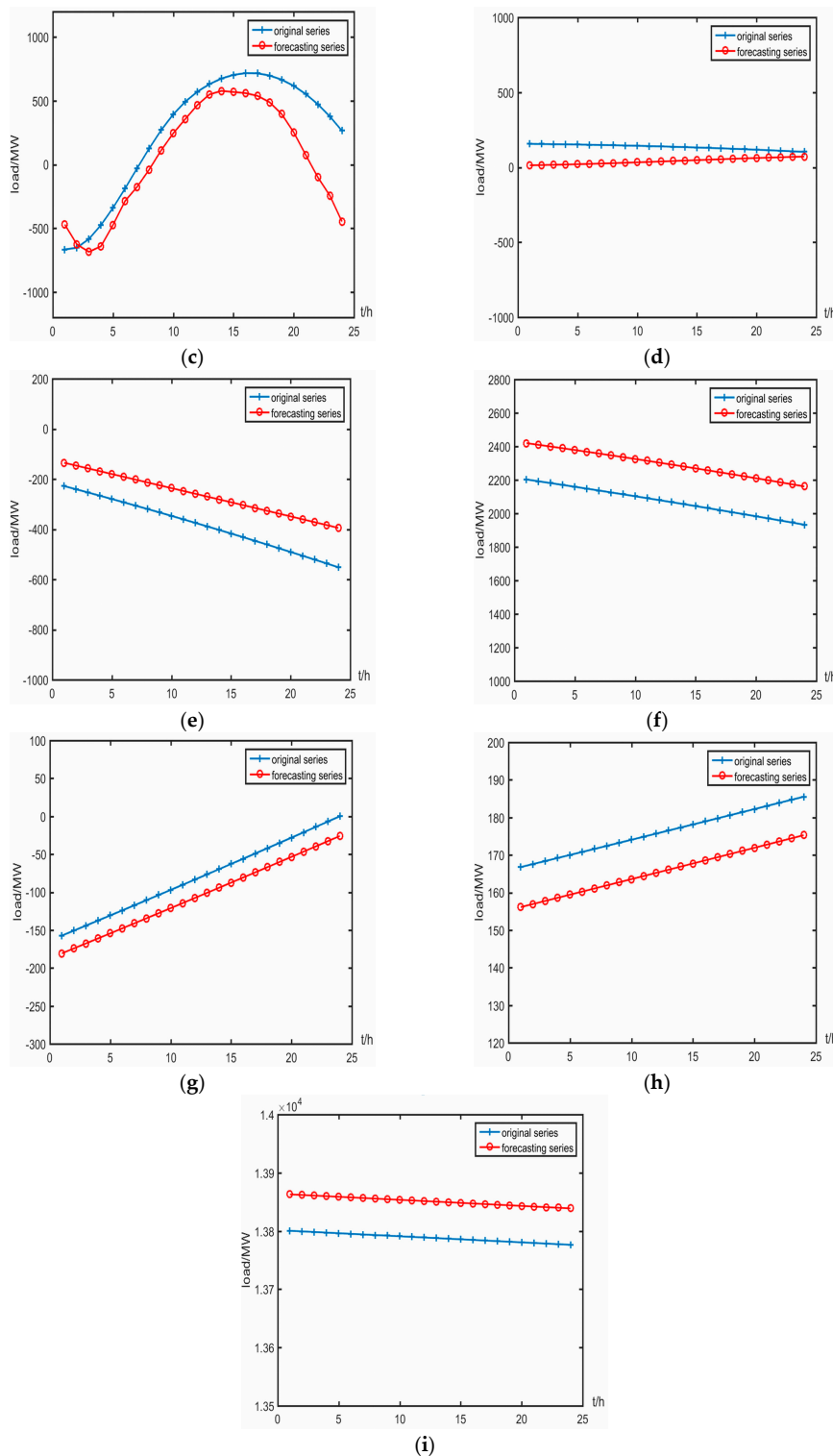
Figure 6. The forecasting results of IMFs. (**a**) IMF1; (**b**) IMF2; (**c**) IMF3; (**d**) IMF5; (**e**) IMF6; (**f**) IMF7; (**g**) IMF8; (**h**) IMF9; (**i**) Residual.

Reconstruct the forecasting series to get the final forecasting results of the hourly load in 15 March and the relative error shown in Figure 7.

As shown in Figure 7, when we use the EMD-GRA-MPSO-LSSVM model to forecast the hourly load in 15 March in the Jibei area, the forecasting curve fits very well and the forecasting precision is very high. The relative error of each forecasting point is not more than 3%.

Table 1. The main parameters of MPSO-LSSVM.

| Parameters	Values	Parameters	Values
The search range of the regularization parameter	[0.1, 150]	The acceleration coeff	2
The search range of the radial basis kernel function parameter	[0.1, 10]	Maximum Iteration number	100
Initial population size	20	The threshold of averaging particle distance	0.001
The range of the inertia weight	[0.4, 0.9]	The threshold of the fitness variance	0.01

(a)

(b)
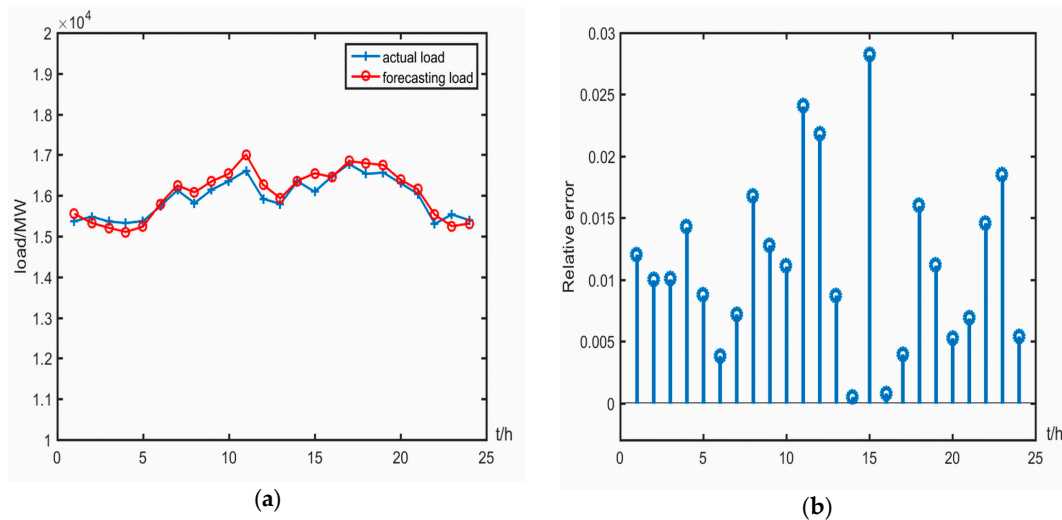
Figure 7. The final forecasting results and the relative error. (**a**) The final forecasting results of the hourly load; (**b**) The relative error.

5.3. Model Comparison

In order to verify the superiority and effectiveness of the EMD-GRA-MPSO-LSSVM model, the BP neural network, SVM, LSSVM, PSO-LSSVM, MPSO-LSSVM, EMD-MPSO-LSSVM models are selected to compare with the EMD-GRA-MPSO-LSSVM model using the same sample. The comparison of forecasting results, the relative error and the boxplot of relative error for different models are shown in Figures 8–10, respectively.
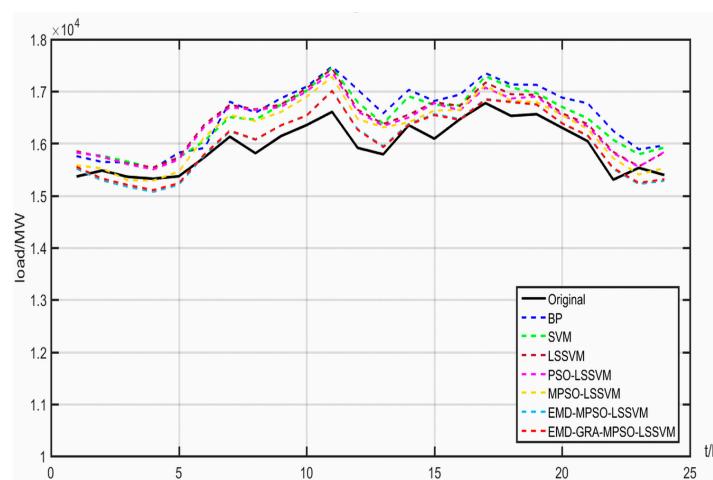

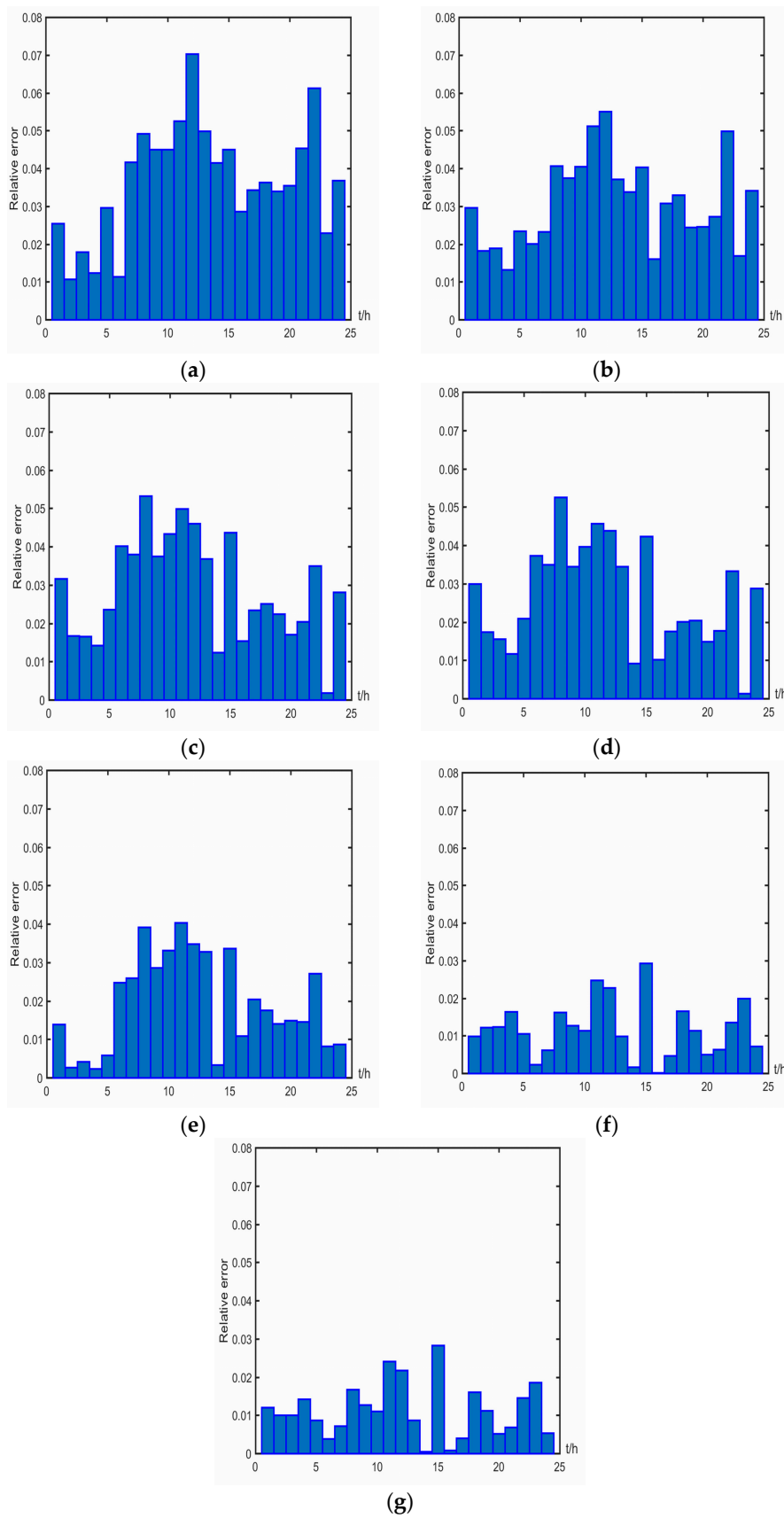**Figure 8.** Comparison of forecasting results for each model.

Figure 9. The relative error of forecasting results for each model. (**a**) BP; (**b**) SVM; (**c**) LSSVM; (**d**) PSO-LSSVM; (**e**) MPSO-LSSVM; (**f**) EMD-MPSO-LSSVM; (**g**) EMD-GRA-MPSO-LSSVM.
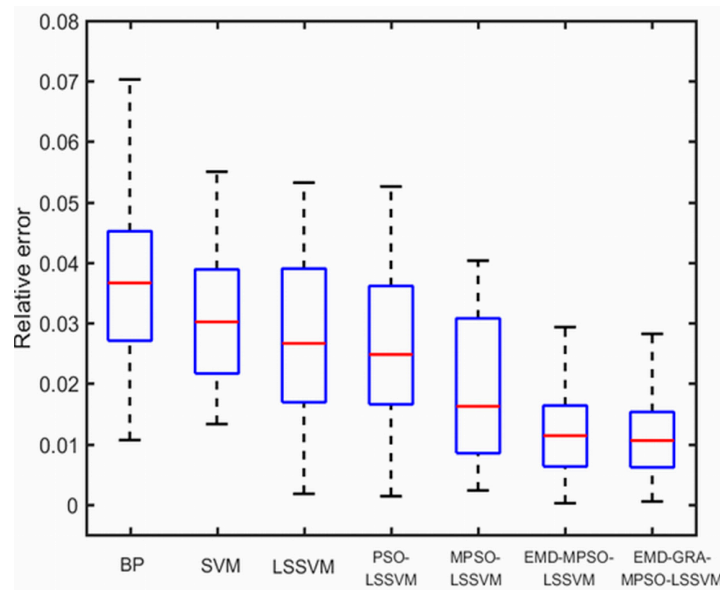
Figure 10. The boxplot of relative error for different models.

The fitting results of the forecasting curves for each model are presented in Figure 8. The forecasting curve of the EMD-GRA-MPSO-LSSVM model fits best. The relative errors of the seven models for the hourly load forecasting in 15 March 2016 in Jibei area are shown in Figure 9. The boxplot in Figure 10 shows the five statistics of the relative error of each forecasting model that is the minimum, first quartile, the median, third quartile and the maximum. As we can see from Figure 9 to Figure 10, the relative error of the EMD-GRA-MPSO-LSSVM model for load forecasting is minimum, followed by the EMD-MPSO-LSSVM model and the BP neural network model has the maximum relative error. The relative errors of all models are not more than 8%.

In order to evaluate the forecasting performance of each model more accurately, the mean absolute percentage error *MAPE*, root mean square error *RMSE*, mean absolute error *MAE* and nonlinear function goodness of fit R^2 are used to compare the forecasting accuracy of each model in this paper. The calculating equations of these indexes are shown as follows:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{17}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{18}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{19}$$

$$R^2 = 1 - \sqrt{\frac{\sum\limits_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum\limits_{i=1}^{n}y_i{}^2}} \tag{20}$$

The calculated results of the indexes for each model are shown in Table 2.

Table 2. The calculation results of the index.

Model	*MAPE* (%)	*RMSE* (100 MW)	*MAE* (100 MW)	R^2 (%)
BP	3.68	6.36	5.89	96.02
SVM	3.08	5.26	4.93	96.70
LSSVM	2.88	5.07	4.61	96.82
PSO-LSSVM	2.65	4.72	4.22	97.04
MPSO-LSSVM	1.92	3.66	3.09	97.71
EMD-MPSO-LSSVM	1.18	2.21	1.88	98.62
EMD-GRA-MPSO-LSSVM	1.14	2.13	1.81	98.67

As shown in Table 2, the MAPE, RMSE and MAE of EMD-GRA-MPSO-LSSVM model are the smallest and the goodness of fit is the best of all models, reaching 98.62%. The MAPE, RMSE and MAE of BP model are the largest and the goodness of fit is the worst of all models, reaching 96.02%.

The LSSVM model optimized by the modified particle swarm optimization algorithm is better than the LSSVM model optimized by the standard particle swarm optimization algorithm, the LSSVM model and the SVM model for load forecasting. The MPSO-LSSVM model improved by empirical mode decomposition is significantly better than the MPSO-LSSVM model, but slightly inferior to the EMD-GRA-MPSO-LSSVM model. In general, the evaluation results of the four indexes for the seven models tend to be consistent. The forecasting accuracy of each model is ranked as follows: EMD-GRA-MPSO-LSSVM > EMD-MPSO-LSSVM > MPSO-LSSVM > PSO-LSSVM > LSSVM > SVM > BP.

5.4. Supplementary Experiment

In order to fully prove the reliability of the above experimental results, the BP neural network, SVM, LSSVM, PSO-LSSVM, MPSO-LSSVM, EMD-MPSO-LSSVM, EMD-GRA-MPSO-LSSVM models are used to forecast the hourly load for 16 consecutive days from 13 to 28 September 2016. The mean absolute percentage errors of the hourly load of each forecasting day for different models are calculated and the calculation results are shown in Table 3.

Table 3. The mean absolute percentage errors for different models.

Date	BP	SVM	LSSVM	PSO-LSSVM	MPSO-LSSVM	EMD-MPSO-LSSVM	EMD-GRA-MPSO-LSSVM
13 September 2016	3.17%	2.82%	1.11%	0.99%	0.79%	0.66%	0.61%
14 September 2016	5.07%	4.03%	2.91%	1.91%	1.62%	1.40%	1.31%
15 September 2016	5.99%	4.88%	2.90%	2.15%	1.53%	1.38%	1.28%
16 September 2016	4.57%	3.90%	3.11%	2.03%	1.50%	1.40%	1.29%
17 September 2016	3.98%	3.35%	3.00%	1.87%	1.76%	1.61%	1.32%
18 September 2016	4.12%	3.50%	3.11%	2.39%	2.24%	1.12%	1.03%
19 September 2016	2.32%	2.23%	1.90%	1.86%	1.66%	1.53%	1.28%
20 September 2016	3.04%	2.85%	2.68%	2.53%	2.01%	1.91%	1.87%
21 September 2016	3.03%	2.76%	2.37%	2.14%	1.80%	1.41%	1.05%
22 September 2016	5.12%	3.11%	2.94%	2.33%	2.16%	2.00%	1.55%
23 September 2016	3.19%	2.91%	2.14%	1.99%	1.48%	0.78%	0.76%
24 September 2016	4.99%	3.99%	3.15%	1.65%	1.45%	1.17%	1.02%
25 September 2016	5.12%	4.70%	3.26%	2.52%	1.90%	1.55%	1.53%
26 September 2016	4.96%	4.21%	2.79%	2.45%	2.02%	0.61%	0.59%
27 September 2016	5.79%	5.47%	4.62%	2.71%	2.12%	1.06%	0.95%
28 September 2016	5.30%	5.05%	3.75%	1.64%	1.57%	1.45%	1.34%

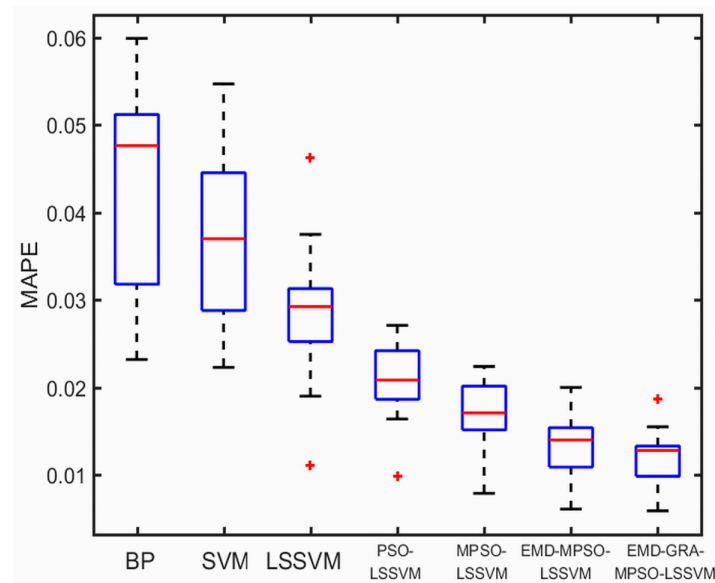The boxplot of MAPE for different models are shown in Figure 11.

Figure 11. The boxplot of MAPE for different models.

As shown in Table 3 and Figure 11, by means of the error analysis for the 384 prediction points of the newly added 16 sets of experiments, the mean absolute percentage errors for the hourly load of each forecasting day of the EMD-GRA-MPSO-LSSVM model are the smallest. The ranking of the prediction accuracy for the seven models in the 16 sets of experiments is consistent with the test result of the case in 15 March (EMD-GRA-MPSO-LSSVM > EMD-MPSO-LSSVM > MPSO-LSSVM > PSO-LSSVM > LSSVM > SVM > BP). It is proved that the EMD-GRA-MPSO-LSSVM model proposed in this paper is reliable and can achieve good prediction effect in short-term load forecasting.

Through the contrast and analysis, we can find that:

(1)   The fitting effect for short-term nonlinear time series of the combined model is obviously better than that of single model.
(2)   The modified optimization algorithm has the advantage over the original optimization algorithm in parameter optimization.
(3)   The method of time series decomposition can effectively improve the forecasting accuracy of the forecasting model.
(4)   The forecasting effect of the time series processed by de-noising method is superior to that of the original time series.

As a complex multiple combination forecasting model, EMD-GRA-MPSO-LSSVM can realize the complementary advantages of different algorithms. The EMD-GRA-MPSO-LSSVM model proposed in this paper eliminates the chaos of the original data by decomposing and denoising the nonstationary time series, which makes the time series more regular. The model realizes the reasonable choice of the parameters by modifying the optimization algorithm, which improves the forecasting accuracy. The model of EMD-GRA-MPSO-LSSVM combines a variety of forecasting methods to make up for the shortcomings of other models, and shows strong generalization ability and robustness. Through the analysis above, it is proved that the EMD-GRA-MPSO-LSSVM short-term load forecasting model proposed in this paper is practical and effective.

6. Conclusions

In this paper, a novel EMD-GRA-MPSO-LSSVM short-term load forecasting model is proposed. First of all, we process the original load data for denoising by EMD-GRA, using empirical mode decomposition to decompose the original load series and removing the noise subsequences by

grey relational analysis. Then, we forecast the subsequences using the modified particle swarm optimization algorithm and least square support vector machine (MPSO-LSSVM) model. Finally, we reconstruct the forecasting series and get the final forecasting results. This paper takes the Jibei area of China as an example to perform an empirical analysis for load forecasting considering historical load, weather conditions, holidays and other factors. And the models of BP neural network, SVM, LSSVM, PSO-LSSVM, MPSO-LSSVM, EMD-MPSO-LSSVM are selected to compare with the model of EMD-GRA-MPSO-LSSVM using the same sample. The comparison results verify that the EMD-GRA-MPSO-LSSVM short-term load forecasting model proposed in this paper has strong generalization ability and robustness and it can achieve good forecasting effect with high forecasting accuracy.

Author Contributions: In this research activity, all the authors were involved in the data collection and preprocessing phase, model constructing, empirical research, results analysis and discussion, and manuscript preparation. All authors have approved the submitted manuscript.

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

References

1. Velasco, L.C.P.; Villezas, C.R.; Palahang, P.N.C.; Dagaang, J.A.A. Next day electric load forecasting using Artificial Neural Networks. In Proceedings of the International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), Puerto Princesa, Philippines, 12–16 November 2014.

2. Kowm, D.; Kim, M.; Hong, C.; Cho, S. Artificial neural network based short term load forecasting. *Int. J. Smart Home* **2014**, *8*, 145–150. [CrossRef]

3. Hernández, L.; Baladrón, C.; Aguiar, J.M.; Calavia, L.; Carro, B.; Sánchez-Esguevillas, A.; Pérez, F.; Fernández, A.; Lloret, J. Artificial Neural Network for Short-Term Load Forecasting in Distribution Systems. *Energies* **2014**, *7*, 1576–1598. [CrossRef]

4. Li, L.J.; Huang, W. A Short-Term Power Load Forecasting Method Based on BP Neural Network. *Appl. Mech. Mater.* **2014**, *494–495*, 1647–1650. [CrossRef]

5. Bin, H.; Zu, Y.X.; Zhang, C. A Forecasting Method of Short-Term Electric Power Load Based on BP Neural Network. *Appl. Mech. Mater.* **2014**, *538*, 247–250. [CrossRef]

6. Li, C.; Liang, Z.Q.; Chen, M.Y. An Improved RBF Neural Network for Short Term Load Forecasting. *Adv. Mater. Res.* **2014**, *1008–1009*, 709–713. [CrossRef]

7. Ni, J.; Jin, G. The Short Term Load Forecasting of RBF Neural Network Power System Based on Fuzzy Control. In Proceedings of the International Conference on Electronics, Network and Computer Engineering, Yinchuan, China, 13–14 August 2016.

8. Dong, J.; Xia, C.; Zhang, W. A New Model for Short-Term Power System Load Forecasting Using Wavelet Transform Fuzzy RBF Neural Network. *Lect. Notes. Electr. Eng.* **2014**, *277*, 99–106.

9. Suka, M.; Oeda, S.; Ichimura, T.; Yoshida, K.; Takezawa, J. Advantages and Disadvantages of Neural Networks for Predicting Clinical Outcomes. In Proceedings of the International Multiconference of Engineers and Computer Scientists, Hong Kong, China, 21–23 March 2007.

10. Yang, J.F.; Cheng, H.Z. Application of SVM to power system short-term load forecast. *Electr. Power. Autom. Equip.* **2004**, *24*, 30–32.

11. Hong, X.; Wei, J.P.; Liu, H.L. Parameter Selection and Optimization Method of SVM Model for Short-term Load Forecasting. *Proc. CSEE* **2006**, *26*, 17–22.

12. De Giorgi, M.G.; Malvoni, M.; Congedo, P.M. Comparison of strategies for multi-step ahead photovoltaic power forecasting models based on hybrid group method of data handling networks and least square support vector machine. *Energy* **2016**, *107*, 360–373. [CrossRef]

13. Kong, X.; Liu, X.; Shi, R.; Lee, K.Y. Wind speed prediction using reduced support vector machines with feature selection. *Neurocomputing* **2015**, *169*, 449–456. [CrossRef]

14. Li, X.; Wang, X.; Zheng, Y.H.; Li, L.X.; Zhou, L.D.; Sheng, X.K. Short-Term Wind Power Forecasting Based on Least-Square Support Vector Machine (LSSVM). *Appl. Mech. Mater.* **2013**, *448*, 1825–1828. [CrossRef]

15. Zhang, J.; Lin, Y.; Lu, P. Short-Term Electricity Load Forecasting Based on ICA and LSSVM. In Proceedings of the Computational Intelligence and Software Engineering (CiSE 2009), Wuhan, China, 11–13 December 2009.

16. Sun, W.; Yang, C. Research of Least Square Support Vector Machine Based on Chaotic Time Series in Power Load Forecasting Model. In Proceedings of the International Conference on Neural Information Processing, Hong Kong, China, 3–6 October 2006.

17. Mahjoob, M.J.; Abdollahzade, M.; Zarringhalam, R. GA based optimized LS-SVM forecasting of short term electricity price in competitive power markets. In Proceedings of the Industrial Electronics and Applications, Singapore, 3–5 June 2008.

18. Liu, H.Y.; Jiang, Z.J. Research on Failure Prediction Technology Based on Time Series Analysis and ACO-LSSVM. *Comput. Mod.* **2013**, *1*, 219–222. [CrossRef]

19. Yusof, Y.; Kamaruddin, S.S.; Husni, H.; Ku-Mahamud, K.R.; Mustaffa, Z. Forecasting Model Based on LSSVM and ABC for Natural Resource Commodity. *Int. J. Comput. Theor. Eng.* **2013**, *5*, 906–909. [CrossRef]

20. Qiao, W.; Chen, B.; Wu, S.; Ma, J. A forecasting model of natural gas daily load based on wavelet transform and LSSVM—DE. *Nat. Gas Ind.* **2014**, *34*, 118–124.

21. Wu, Q.; Peng, C. Wind power grid connected capacity prediction using LSSVM optimized by the bat algorithm. *Energies* **2015**, *8*, 14346–14360. [CrossRef]

22. Malvoni, M.; De Giorgi, M.G.; Congedo, P.M. Photovoltaic forecast based on hybrid PCA–LSSVM using dimensionality reduced data. *Neurocomputing* **2016**, *211*, 72–83. [CrossRef]

23. Yang, W.; Li, Q. Survey on Particle Swarm Optimization Algorithm. *Eng. Sci.* **2004**, *6*, 87–94.

24. Wang, D.; Zeng, P.; Li, L.; Wang, K. A Novel Particle Swarm Optimization Algorithm. In Proceedings of the International Conference on Software Engineering and Service Sciences, Beijing, China, 16–18 July 2010.

25. Bai, J.; Zhang, X.; Xue, P.; Wang, J. Parameters Optimization and Application of v-Support Vector Machine Based on Particle Swarm Optimization Algorithm. In Proceedings of the International Conference on Computing, Measurement, Control and Sensor Network, Taiyuan, China, 7–9 July 2012.

26. Zhu, L.; Wang, Z. Research on Power Load Forecasting Based on PSO-LSSVM. *Sci. Mos.* **2014**, *2*, 11–15.

27. Sun, B.; Yao, H.T. The short-term wind speed forecast analysis based on the PSO-LSSVM predict model. *Power Syst. Prot. Control.* **2012**, *40*, 85–89.

28. Soltani, S. On the use of the wavelet decomposition for time series prediction. *Neurocomputing* **2002**, *48*, 267–277. [CrossRef]

29. Mabrouk, A.B.; Abdallah, N.B.; Dhifaoui, Z. Wavelet decomposition and autoregressive model for time series prediction. *Appl. Math. Comput.* **2008**, *199*, 334–340. [CrossRef]

30. 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]

31. Guo, H.; Liu, Y.; Liang, H.; Gao, X. An Application on Time Series Clustering Based on Wavelet Decomposition and Denoising. In Proceedings of the 4th International Conference on Natural Computation, Jinan, China, 18–20 October 2008.

32. Xia, C.; Lei, B.; Rao, C.; He, Z. Research on short-term load forecasting model based on wavelet decomposition and neural network. In Proceedings of the International Conference on Natural Computation, Shanghai, China, 26–28 July 2011.

33. De Giorgi, M.G.; Campilongo, S.; Ficarella, A.; Congedo, P.M. Comparison between wind power prediction models based on wavelet decomposition with Least-Squares Support Vector Machine (LS-SVM) and Artificial Neural Network (ANN). *Energies* **2014**, *7*, 5251–5272. [CrossRef]

34. Lima, E.R.D.; Andrade, A.O.; Pons, J.L.; Kyberd, P.; Nasuto, S.J. Empirical mode decomposition: A novel technique for the study of tremor time series. *Med. Biol. Eng. Comput.* **2006**, *44*, 569–582. [CrossRef] [PubMed]

35. Uygur, K.; Bayramoglu, I.; Nazikoglu, A.; Yilmaz, M.; Bayazit, Y.; Muftuoglu, S. Using empirical mode decomposition to correlate paleoclimatic time-series. *Nat. Hazards Earth Syst. Sci.* **2007**, *7*, 299–307.

36. Pachori, R.B.; Hewson, D.; Snoussi, H.; Duchene, J. Postural time-series analysis using Empirical Mode Decomposition and second-order difference plots. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, Taiwan, 19–24 April 2009.

37. Wei, L.Y. A hybrid ANFIS model based on empirical mode decomposition for stock time series forecasting. *Appl. Soft Comput.* **2016**, *42*, 368–376. [CrossRef]

38. Xing, H.; Wang, H.; Liu, Y.; Lu, J.; Jiang, W. Multi-scale modeling method based on EMD-LSSVM and its application. *Infrared Laser Eng.* **2013**, *42*, 1737–1742.

39. Wang, X.; Li, H. One-Month Ahead Prediction of Wind Speed and Output Power Based on EMD and LSSVM. In Proceedings of the 2009 International Conference on Energy and Environment Technology, Guilin, China, 16–18 October 2009.

40. Feng, N.D.; Yan, N.J.; Huang, N.P. A Novel De-noise Method Based on the Grey Relational Analysis. In Proceedings of the IEEE International Conference on Information Acquisition, Weihai, China, 20–23 August 2006.

41. Uzunhisarcikli, E.; Latifoglu, F.; Koza, T.; Kaya, M.G. Doppler ultrasound signal denoising based on Grey relational analysis technique. In Proceedings of the 2009 14th National Biomedical Engineering Meeting, Izmir, Turkey, 20–22 May 2009.

42. Xuxiang, L.U.; Yiming, S.U.; Jiateng, W.U.; Luping, L.I. Fault Diagnosis on Lubrication State of Journal Bearings Based on EMD and Grey Relational Degree. *J. Chin. Soc. Power Eng.* **2016**, *36*, 42–47.

43. Chi, Y.; Sun, F.; Jiang, L.; Yu, C. An efficient population diversity measure for improved particle swarm optimization algorithm. In Proceedings of the 2012 6th IEEE International Conference Intelligent Systems, Sofia, Bulgaria, 6–8 September 2012.