

Article

Forecasting Fossil Fuel Energy Consumption for Power Generation Using QHSA-Based LSSVM Model

Wei Sun ^{1,*}, Yujun He ² and Hong Chang ^{3,*}

¹ School of Economics and Management, North China Electric Power University, Baoding 071003, Hebei, China

² Department of Electronic & Communication Engineering, North China Electric Power University, Baoding 071003, Hebei, China; E-Mail: heyujun@ncepubd.edu.cn

³ Key Laboratory of Advanced Control and Optimization for Chemical Processes, East China University of Science and Technology, Shanghai 200240, China

* Authors to whom correspondence should be addressed;

E-Mails: hdsunwei@ncepubd.edu.cn (W.S.); hchang@ecust.edu.cn (H.C.);

Tel.: +86-139-3024-7406 (W.S.); Fax: +86-312-752-5117 (W.S.); Tel./Fax: +86-21-6425-3463 (H.C.).

Academic Editor: Vincenzo Dovi

Received: 21 November 2014 / Accepted: 13 January 2015 / Published: 28 January 2015

Abstract: Accurate forecasting of fossil fuel energy consumption for power generation is important and fundamental for rational power energy planning in the electricity industry. The least squares support vector machine (LSSVM) is a powerful methodology for solving nonlinear forecasting issues with small samples. The key point is how to determine the appropriate parameters which have great effect on the performance of LSSVM model. In this paper, a novel hybrid quantum harmony search algorithm-based LSSVM (QHSA-LSSVM) energy forecasting model is proposed. The QHSA which combines the quantum computation theory and harmony search algorithm is applied to searching the optimal values of σ and C in LSSVM model to enhance the learning and generalization ability. The case study on annual fossil fuel energy consumption for power generation in China shows that the proposed model outperforms other four comparative models, namely regression, grey model (1, 1) (GM (1, 1)), back propagation (BP) and LSSVM, in terms of prediction accuracy and forecasting risk.

Keywords: fossil fuel energy forecasting; power generation; LSSVM; quantum harmony search algorithm (QHSA)

1. Introduction

Since the implementation of the policy of reform and opening to the outside world in 1979, China has experienced outstanding economic development, with an average annual growth rate of 10%. Due to the continuous sustainable positive economic growth rate and large scale industrialization, electricity consumption is rising quickly [1]. Electricity plays an important role in socio-economic development, which is considered as the backbone for national economy's prosperity and progress. It is estimated that the electricity demand in China will continue to grow, since the Chinese government aims to raise its GDP with an incredible speed in the next 30 years. Affected by energy resource endowments, China's power generation sector relies heavily on fossil fuel energy and its products. In 2011, national thermal power plants generated 3.8975 trillion kWh, accounting for 82.54% of the total electricity generation [2]. The fossil fuel-based generation pattern will not change in a short term. In the future, the fossil fuel consumption for power generation will inevitably increase. Accurate forecasting of fossil fuel energy consumption for power generation is therefore important and fundamental for rational energy planning formulation in electric power industry.

In energy prediction field, some traditional forecasting approaches have been adopted in the last decades, such as regression model, time series model, Grey forecasting technique and bottom up long-range energy alternatives planning system (LEAP) model. Zhang [3] and Limanond [4] applied partial least square regression model and log-linear regression model to forecast the transport energy demand in China and Thailand, respectively. Kumar [5] applied three time series models, namely Grey–Markov model, Grey–Model with rolling mechanism and singular spectrum analysis (SSA) to forecast the consumption of conventional energy in India. Amarawickrama [6] estimated the electricity demand functions for Sri Lanka using six econometric techniques. Hsu [7] proposed an improved GM (1, 1) model to predict the power demand in Taiwan. Huang [8] used a LEAP model to forecast long-term energy supply and demand and compared future energy demand and supply patterns, as well as greenhouse gas emissions in Taiwan. The conventional statistical methods usually require the assumption of the normal distribution of energy consumption data [9]. Moreover, they are inherently limited in the presence of nonlinearities in data, which partially results from the use of linear model structures or the static nonlinear function relationships. An alternative way to deal with nonlinearities is to use Neural Networks (NN), which a powerful data modeling tool that is able to capture and represent complex input/output relationships [10–13]. It is considered that the NN could perform better than the traditional models from a statistical aspect [14,15]. However, estimating the network weights requires large amounts of data, which may be very computer-intensive. Another, it always yields limited generalization capability and unpredictably large errors since the NN usually implements the empirical risk minimization (ERM) principle; and it may show slow convergence speed, arriving at local minimum and overfitting issues [16].

Support Vector Machine (SVM), pioneered by Vapnik in 1995 [17], can effectively solve the learning problems of small sample size, higher dimension and nonlinearities [18–20]. The main advantage is that the solution of SVM is global and unique since SVM implements the structural risk minimization (SRM) principle [21–23]. Moreover, the SVM often outperforms artificial neural network (ANN) in practice and it is less prone to overfitting [24]. The Least Squares Support Vector Machine (LSSVM), proposed by Suykens and Vandewalle in 1999 [25], is a variant of SVM. LSSVM adopts a least

squares linear system as a loss function instead of the quadratic program in original SVM which is time consuming in training process [26–30]. The LSSVM shows manifest advantages, such as good nonlinear fitting ability, strong generalization capability, fast computing speed, dealing with small samples, not relying on the distribution characteristics of the samples and so on [31–34]. The performance of the LSSVM model is largely dependent on the selection of the parameters. Therefore, to obtain better forecasting accuracy, how to set the parameters of the LSSVM is very crucial. So far, there is no effective way to guide the parameters selection process. The contribution of this paper is to develop a hybrid quantum harmony search algorithm-based LSSVM (QHSA-LSSVM) energy forecasting model. In our work, the QHSA is applied to determining the optimal parameters of LSSVM model in order to enhance the learning and generalization ability.

The harmony search algorithm (HSA), a novel evolutionary algorithm, is a metaheuristic technique mimicking the improvisation behavior of musicians [35–38]. The QHSA combines the quantum computation theory [39,40] and harmony search algorithm, which adopts concepts and principles of the quantum mechanism, such as quantum bit (qubit) and superposition of states [41,42] to the HSA. It can effectively improve the performance of HSA. The practical annual energy consumption data of power generation is employed to test the performance of the proposed QHSA-LSSVM forecasting model.

The remainder of this paper is structured as follows: Section 2 introduces the LSSVM model, QHSA theory, and then a hybrid QHSA-LSSVM model is proposed in detail. Section 3 shows the data source for simulation. The empirical simulation and results analysis are presented in Section 4. Finally Section 5 gives our main conclusions.

2. Methodologies

2.1. Least Squares Support Vector Machine (LSSVM)

SVM represents a relatively new computational learning method based on statistical learning theory [17]. The basic idea of SVM is that the original input data space is mapped into a higher dimensional dot product space called a feature space. The actual problem is transformed into a quadratic programming problem with inequality constraints. LSSVM is an alternate formulation of SVM regression. In LSSVM, the inequality constraints are replaced with equality constraints; and a least squares linear system is adopted as a loss function instead of the time-consuming quadratic program in original SVM [25,26,29,33,43,44]. In the following, LSSVM algorithm is described briefly.

Let $\{x_i, y_i\}_{i=1}^m$ be a given training set of m data points, where $x_i \in R^n$ is the input vector and $y_i \in R$ is the corresponding output vector. LSSVM maps the input vector x_i to a higher dimensional feature space, using a nonlinear kernel function $\phi(\cdot)$, shown as Equation (1). Through Equation (1), the nonlinear estimation in original space can be transformed into linear function estimation in feature space:

$$y(x) = w^T \phi(x) + b \quad (1)$$

where w and b denote the weights vector and the bias respectively, which should be estimated from the training data. For LSSVM function estimation, the formulation of optimization issue can be written as follows:

$$\min J(w, b, \xi_i) = \frac{1}{2} w^T w + \frac{1}{2} C \sum_{i=1}^m \xi_i^2 \tag{2}$$

$$s.t. \quad y_i = w^T \varphi(x_i) + b + \xi_i, i = 1, 2, \dots, m$$

where C is the regularization factor, and ξ_i is the slack variable which expresses the difference between the desired output and the actual output. Next, the corresponding Lagrangian function $L(w, b, \xi, a)$ is constructed to solve the above optimization problem, shown in Equation (3):

$$L(w, b, \xi, a) = \frac{1}{2} w^T w + \frac{1}{2} C \sum_{i=1}^m \xi_i^2 - \sum_{i=1}^m a_i \{ w^T \varphi(x_i) + b + \xi_i - y_i \} \tag{3}$$

where $a_i \in R$ is a Lagrangian multiplier. From the Karush–Kuhn–Tucker (KKT) conditions, the following equations must be satisfied:

$$\begin{aligned} \frac{\partial L}{\partial w} = 0 &\rightarrow w = \sum_{i=1}^m a_i \varphi(x_i) \\ \frac{\partial L}{\partial b} = 0 &\rightarrow \sum_{i=1}^m a_i = 0 \\ \frac{\partial L}{\partial \xi_i} = 0 &\rightarrow a_i = C \xi_i \\ \frac{\partial L}{\partial a_i} = 0 &\rightarrow w^T \varphi(x_i) + b + \xi_i - y_i = 0 \end{aligned} \tag{4}$$

After eliminating the variables w and ξ_i , the solution is given by the following set of linear equations, shown in Equation (5):

$$\begin{bmatrix} 0 & \vec{1}^T \\ \vec{1} & Q + C^{-1}I \end{bmatrix} \begin{bmatrix} b \\ A \end{bmatrix} = \begin{bmatrix} 0 \\ Y \end{bmatrix} \tag{5}$$

where $\vec{1} = [1, 1, \dots, 1]^T$, $Y = [y_1, y_2, \dots, y_m]^T$, $A = [a_1, a_2, \dots, a_m]^T$, and I is the identity matrix. Define $Q_{ij} = K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$, which is satisfied with Mercer’s condition. In our work, the Gaussian radial basis function (RBF) is selected as the kernel function, as is expressed in Equation (6):

$$K(x_i, x_j) = \exp\left\{-\|x_i - x_j\|^2 / 2\sigma^2\right\} \tag{6}$$

The LSSVM regression model becomes:

$$y(x) = \sum_{i=1}^m a_i K(x, x_i) + b \tag{7}$$

The kernel function width coefficient σ and the regularization factor C have manifest influence on the performance of the LSSVM model. The width coefficient σ affects the width of RBF, and the regularization factor C influences the complexity of the model and the penalty degree. In order to avoid selecting the parameters arbitrarily, the optimal values of these two parameters are determined by QHSA, which can effectively enhance the prediction accuracy.

2.2. Quantum Harmony Search Algorithm (QHSA)

The harmony search algorithm (HSA), proposed by Geem *et al.* [35], is a novel phenomenon-mimicking algorithm inspired by the improvisation process of musicians [38]. The HSA has been successfully applied to many computational optimization problems, such as the traveling salesman problem [35], continuous engineering optimization [45], the layout of pipe networks [46] and course timetabling issues [47], due to its advantages of fewer parameters, excellent effectiveness and robustness compared with other heuristic optimization algorithms [48]. However, the performance of the HSA seems to be affected by the initial harmony memory (HM) parameters and solution vectors, and it usually falls into local searching for complicated numerical optimization issues. Inspired by quantum theorem, the quantum harmony search algorithm (QHSA) is proposed in our work to solve the aforementioned problems and effectively enhance the convergence rate, generalization ability and optimization capability.

Quantum theory is considered as one of the greatest scientific achievements in the twentieth century. The concept of quantum computing is presented by Benioff [49] and Feynman [50] in the 1980s through combing quantum theory and information science. In quantum computing, the qubit is adopted to express the information; and the theory of superposition, entanglement and collapse of states is employed for information computation [39,40,42,51]. Quantum computing greatly improves the computational efficiency and has attracted widespread interest and research. QHSA merges the quantum computing theory with HSA. In QHSA, the qubit is employed to express the harmony vector in HM; and the new harmony vector is introduced to adjust the bandwidth (BW) dynamically. This algorithm can improve the global search ability and optimization speed through combining superposition of qubits. The concrete optimization procedures of QHSA are as illustrated in Figure 1.

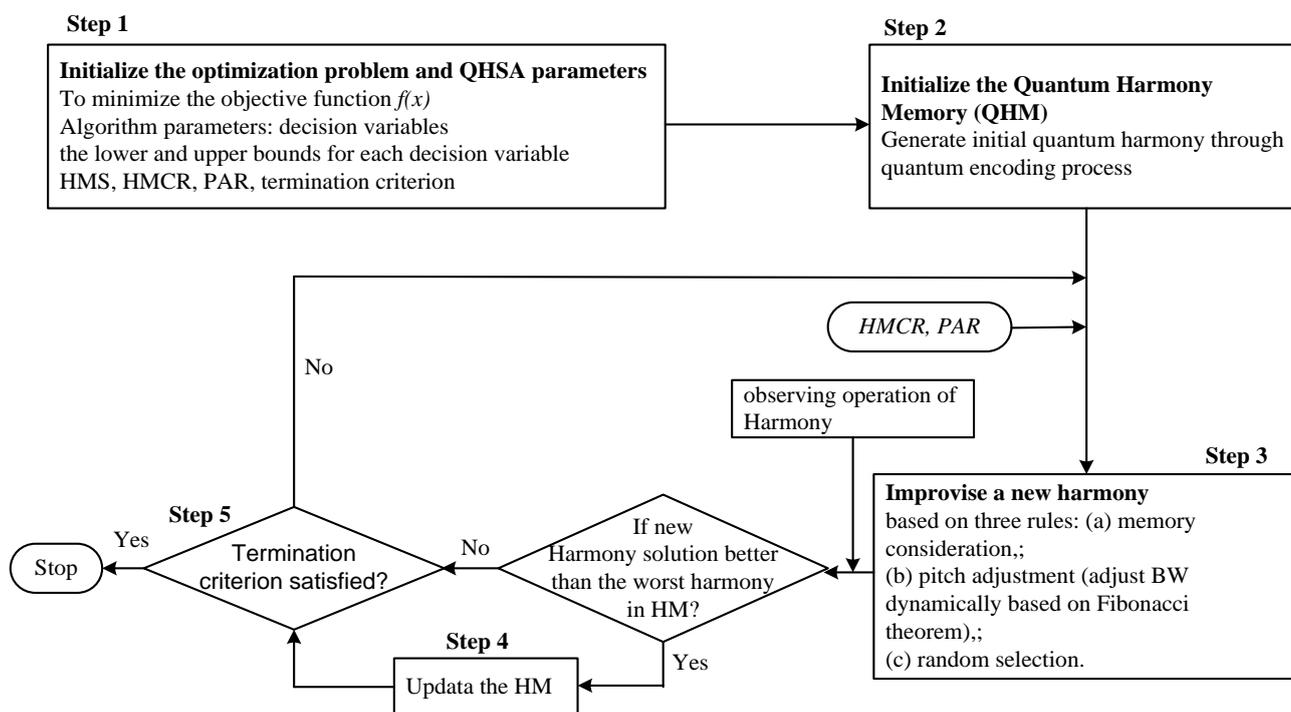


Figure 1. Quantum Harmony Search Algorithm (QHSA) optimization procedures.

Step 1. Initialize the optimization problem and algorithm parameters.

$$\text{Minimize } f(x), \text{ s.t. } x_i \in X_i \quad i = 1, 2, \dots, N$$

where $f(x)$ is the objective function; x is the set of each design variable (x_i); X_i is the set of the possible range of values for each design variable; N is the number of design variables. In addition, the HS algorithm parameters including harmony memory size (HMS), harmony memory considering rate (HMCR), pitch adjusting rate (PAR) should also be specified in this step.

Step 2. Initialize the Quantum Harmony Memory (QHM).

The HM is a location storing all the solution vectors. The HM matrix is filled with randomly generated solution vectors and sorted by the values of the objective function $f(x)$. Inspired by the concept of states superposition in quantum computing, qubit is adopted to express the harmonies in HM.

In quantum computing, the qubit is the smallest quantum piece of information, which may be in the $|1\rangle$ state, $|0\rangle$ state, or in any superposition of the two [39,40,42]. Each qubit state can be represented as the linear superposition of the two basic states, shown in Equation (8):

$$|\varphi\rangle = \alpha|0\rangle + \beta|1\rangle \tag{8}$$

where α and β is a pair of complex numbers that specify the probability amplitudes of the corresponding state; $|\alpha|^2$ and $|\beta|^2$ give the probabilities that the qubit will be in the state “0” and “1” respectively, with $|\alpha|^2 + |\beta|^2 = 1$. Compared with the single bit in classical computing, the qubits with superposition of states can carry more information and improve the computational efficiency.

In QHSA, the individual harmony can be written as Equation (9) or Equation (10):

$$q_i^t = [q_{i1}^t \quad q_{i2}^t \cdots q_{in}^t] = \begin{bmatrix} \alpha_{i1}^t & \alpha_{i2}^t & \cdots & \alpha_{in}^t \\ \beta_{i1}^t & \beta_{i2}^t & \cdots & \beta_{in}^t \end{bmatrix} \quad i = 1, 2, \dots, m \tag{9}$$

$$q_i^t = [\theta_{i1}^t | \theta_{i2}^t | \cdots | \theta_{in}^t] \tag{10}$$

where q_i^t is the i th quantum harmony individual at generation t in HM denoting a potential solution vector; α_{ij}^t and β_{ij}^t , initialized with $1/\sqrt{2}$, are the probability amplitude for q_i^t with $|\alpha_{ij}^t|^2 + |\beta_{ij}^t|^2 = 1$ ($i = 1, \dots, m$, $j = 1, \dots, n$); θ_{ij}^t is the quantum rotation angle of q_i^t which satisfies $\alpha_{ij}^t = \cos \theta_{ij}^t$, $\beta_{ij}^t = \sin \theta_{ij}^t$; m is the size of HM; n is the dimension of the problem concerned.

Step 3. Improvise a new harmony from the HM based on HMCR and pitch adjustment.

The new harmony vector is randomly selected from the historical values stored in the HM with searching probability HMCR and the possible range of values with searching probability (1-HMCR). The harmony vector is pitch-adjusted with PAR, which is shown as follows:

$$x_i = x_i \pm \text{rand}() \cdot \text{BW} \tag{11}$$

Since BW and PAR have great influence on the precision of solutions and the ability of fine-tuning [52], how to determine the suitable values of these two parameters is an important issue. A small BW value can enhance the local optimization capacity around the new harmony vector, while, a large BW can enlarge the search area of new vectors in the process of pitch adjusting and is good to escape the local optima. In classical HSA, BW is taken a fixed value and usually selected by practical

experience, not considering the effect of BW on global or local optimization. In our work, to enhance the performance of the QHSA, BW is adjusted dynamically based on Fibonacci theorem.

To use the new harmony information, we adjust BW dynamically and decrease the number of parameters chosen in the initialization process. The new harmony is adopted to calculate BW, as shown in Equation (12):

$$\begin{aligned}
 q'_{ijnew} &= q_{ijnew} + Rand(0,1) \times \left(\frac{\pi}{2} - q_{ijnew} \right), Rand(0,1) > 0.618 \\
 q'_{ijnew} &= q_{ijnew} - Rand(0,1) \times q_{ijnew}, Rand(0,1) \leq 0.618
 \end{aligned}
 \tag{12}$$

where q'_{ijnew} is the new vector after pitch adjusting; q_{ijnew} is the new vector before pitch adjusting; $Rand(0,1)$ is a uniform random number. In classical HSA, the algorithm chooses an upper neighboring value with 50% probability and lower with 50% around new vector. Inspired by Fibonacci theorem, 0.618 is chosen as the border of vector adjusting direction, shown in Equation (12) [51,53].

Step 4. Update the HM.

First, transform the new harmony vector to a solution with real values and use the objective function to get the fitness value. The quantum harmony collapses to a single state through Equation (13). The harmony observing operation is repeated during the process of updating procedure:

$$\begin{aligned}
 q^t_{ij} &\rightarrow |1\rangle, \text{ when } rand(0,1) > |\alpha_{ij}|^2, i = 1, 2, \dots, m \\
 q^t_{ij} &\rightarrow |0\rangle, \text{ others } j = 1, 2, \dots, n
 \end{aligned}
 \tag{13}$$

Second, on condition that the new harmony vector shows better fitness function than the worst harmony in the HM, the new harmony is included in the HM and the existing worst harmony is excluded from the HM.

Step 5. The Quantum Harmony Search Algorithm is terminated when there is no significant improvement in the best found solution after some predetermined number of iterations or the maximum number of iterations is reached. If not, repeat Steps 3 and 4.

2.3. QHSA Based LSSVM Model

The QHSA-based LSSVM (QHSA-LSSVM) forecasting model is described in this section. In QHSA-LSSVM, the optimization objective function is specified as the mean absolute percentage error (MAPE). The MAPE index is the most widely used accuracy measurement in forecasting. It expresses the forecasting errors from different measurement units into percentage errors on actual observations [54], shown in Equation (14):

$$\begin{aligned}
 \min f(C, \sigma) &= \min \left\{ \frac{1}{T} \sum_{t=1}^T \left| \frac{y_t - \hat{y}_t}{y_t} \right| \right\} \\
 s.t. \quad C &\in [C_{\min}, C_{\max}] \\
 \sigma &\in [\sigma_{\min}, \sigma_{\max}]
 \end{aligned}
 \tag{14}$$

where y_t is the actual value for the t th period; \hat{y}_t represents its forecasting result for the same period; C_{\min} and C_{\max} are the lower and upper bound for the regularization factor C ; σ_{\min} and σ_{\max} are the

lower and upper bound for the kernel function width coefficient σ ; T is the number of data used for the MAPE calculation. The basic idea of parameters optimization in LSSVM is to search the optimal values of C and σ through iterative algorithm. The concrete procedures of QHSA-LSSVM are described as follows:

Step 1. Initialize the optimization problem and algorithm parameters.

In our study, the MAPE serves as the fitness function to identify the suitable parameters in the LSSVM model. Algorithm parameter selection is an unavoidable issue in the intelligent optimization field process. The parameters are usually selected according to trial-and-error methods, not avoiding blindness and randomness. The uniform design technique is adopted to realize the selection of initial parameters of QHSA, which can make the combination of parameters uniformly distribute in value range space during part experiment process [55]. It can greatly reduce the number of the trials and ensure the representativeness of the test results. There are two design parameters, σ and C ; and the upper bound and lower bound should be set in this step. Moreover, HMS, HMCR and PAR will also be initialized.

Step 2. Quantum Encoding of Harmony Memory.

According to QHSA theory, each harmony vector is firstly transformed into a quantum state, namely the quantum harmony vector (QHV). For QHSA-LSSVM, C and σ are encoded into quantum harmony vectors which are formulated by a quantum bits encoding process.

Step 3. Improvise a new harmony from the HM.

A new harmony vector is generated based on three rules: memory consideration, pitch adjustment and random selection.

Step 4. Observation of Harmony and Update the HM.

Collapse the new quantum harmony vector to a single state with real values by Equation (13); and then calculate the objective function. Update the HM on condition that the new harmony vector shows better fitness than the worst harmony in the HM.

Step 5. Stopping criterion.

If there is no significant improvement in the best found solution or the maximum number of iterations is reached, the QHSA stopping criterion satisfies. The optimal values of C and σ in the LSSVM model can be obtained. Otherwise, go back to Step 2.

3. Data Sources

This section describes how to apply the QHSA to searching for the optimal values of C and σ in LSSVM and then establish the QHSA-LSSVM forecasting model for energy consumption to validate the performance of the aforementioned method. The annual data of the fossil fuel energy consumption for power generation from 1992 to 2012 were collected from China Energy Statistical Yearbooks [56]. The primary energy consumption data were converted to standard coal consumption (10^8 tce) through using conversion coefficients from the General Principles for Calculation of the Comprehensive Energy Consumption (Chinese National Principle, GB/T 2589-2008) [57]. The converted standard coal consumption data are shown in Table 1.

Table 1. Annual standard coal consumption of China between 1992 and 2012 (unit: 10^8 tce).

Year	Energy Consumption	Year	Energy Consumption	Year	Energy Consumption
1992	2.2397	1999	3.7304	2006	7.1388
1993	2.3788	2000	3.9590	2007	8.2251
1994	2.6089	2001	4.2523	2008	9.2669
1995	2.8411	2002	4.4687	2009	9.1993
1996	3.1704	2003	4.9264	2010	9.6327
1997	3.5003	2004	5.6848	2011	10.3205
1998	3.7539	2005	6.2886	2012	11.7500

4. Empirical Simulation and Results Analysis

4.1. The Selection of Comparison Models

Several comparative forecasting models for energy consumption prediction were selected so as to compare the results with the proposed QHSA-LSSVM model. It can be seen from Table 1 that the annual energy consumption series exhibits an obvious increasing trend, so the regression model and GM (1, 1) model were employed as the classical methods to capture the rising trend. The BP and LSSVM were adopted as the artificial intelligent techniques to simulate the relationship between the current data and a number of its previous values. The forecasting performance of the proposed QHSA-LSSVM model will be compared with linear regression model, GM (1, 1) model, BP model and LSSVM model in the next section.

4.2. The Network Structure and Parameters Setting

The selected data were the annual standard coal consumption of China during the period of 1992–2012. A total of 21 data points are divided into the training set (1992–2007) and the testing set (2008–2012).

In our work, according to the roll-based forecasting technique, three previous values are selected as the input variables of the BP, LSSVM and QHSA-LSSVM for each current output value. That means the inputs are $X_{n-3}, X_{n-2}, X_{n-1}$ and the corresponding output is X_n . Firstly, the top three energy data (from 1992 to 1994) were fed into BP, LSSVM and QHSA-LSSVM model, and then the first forecasting value of 1995 could be obtained. Secondly, the next roll-top three energy data (from 1993 to 1995) were employed for forecasting value of 1996. Similarly, the processes are cycling until all the forecasting values are obtained.

In LSSVM and QHSA-LSSVM model, the MAPE indicator was adopted as the optimization objective function $f(x)$ to measure the accuracy in a fitted time series value in statistics, which is expressed as Equation (14). The two parameters of σ and C for the proposed QHSA-LSSVM model could be optimized using QHSA. A flowchart of the QHSA algorithm for parameter initialization is shown in Figure 1. The details of the initial parameters are set as: HMS = 35, HMCR = 0.99, PAR = 0.6, BW = 1, lb = 0.001, ub = 250; that is, the lower bound (lb) and upper bound (ub) for σ and C are set 0.0001 and 250, respectively. In the typical LSSVM model, the default values of σ and C were set 20 and 35, respectively. And the structure of LSSVM is the same with QHSA-LSSVM. For BP network, the neuron number of hidden layer is chosen with experience. In our work, the BP model contains

three layers with three input neurons, eight hidden neurons and only one output neuron. The maximum number of training epochs (iterations) is 5000.

The BP, LSSVM and QHSA-LSSVM model are all realized in MATLAB 7.6.0 (R2008a) on Windows 7 with a 32-bit operating system. After training, the BP, LSSVM and QHSA-LSSVM can be used to forecast the future energy consumption value.

4.3. Experimental Results and Analysis

With the given data (1992–2012), about two-thirds (1992–2007) of it were used as training data to calculate the corresponding parameters of these five models, and the remaining one-third were used as testing data to validate the forecasting performance of the models. The solution parameters in regression model and GM (1, 1) model were determined using the training data. In our work, 30 independent runs were implemented in order to test the stability of model for BP, LSSVM and QHSA-LSSVM. In the proposed QHSA-LSSVM model, Gaussian radial basis function (RBF) is selected as the kernel function. According to the roll-based technique, the two parameters σ and C can be optimized step by step, and until the QHSA reaches the stopping criterion. The finally obtained optimal values for σ and C are 23.8564 and 150.00, respectively.

Due to the use of the roll-based technique, only 13 (1995–2007) data are suggested, *i.e.*, the simulation results for year 1992, 1993 and 1994 cannot be obtained. Therefore, the error evaluation is conducted from 1995 to 2012 to ensure the same comparison condition. The forecasting results and errors of regression, GM (1, 1), BP, LSSVM and QHSA-LSSVM are listed in Table 2.

Table 2. Forecasting results and errors of Regression, GM (1, 1), BP, LSSVM and QHSA-LSSVM (unit: 10^8 tce).

Year	Actual	Regression		GM (1, 1)		BP		LSSVM		QHSA-LSSVM	
		Forecast	Error (%)	Forecast	Error (%)	Forecast	Error (%)	Forecast	Error (%)	Forecast	Error (%)
1995	2.8411	2.7573	-2.9496	2.6636	-6.2490	3.6039	26.8492	3.0052	5.7759	2.9767	4.7728
1996	3.1704	3.1052	-2.0565	2.9054	-8.3600	3.6864	16.2750	3.1217	-1.5361	3.1894	0.5993
1997	3.5003	3.4531	-1.3485	3.1691	-9.4621	3.9730	13.5047	3.3180	-5.2081	3.4760	-0.6942
1998	3.7539	3.801	1.2547	3.4568	-7.9148	4.1609	10.8425	3.5915	-4.3262	3.8063	1.3959
1999	3.7304	4.149	11.2213	3.7706	1.0773	4.3371	16.2637	3.9208	5.104	3.5591	-4.5920
2000	3.9590	4.4969	13.5868	4.1129	3.8868	4.1735	5.4182	4.1392	4.5517	3.8160	-3.6120
2001	4.2523	4.8448	13.9336	4.4862	5.5015	4.9150	15.5851	4.3516	2.3352	4.4611	4.9103
2002	4.4687	5.1927	16.2016	4.8935	9.5061	5.0721	13.5038	4.5897	2.7077	4.7478	6.2457
2003	4.9264	5.5406	12.4675	5.3377	8.3494	5.2183	5.9253	4.9523	0.5257	5.0392	2.2897
2004	5.6848	5.8886	3.5850	5.8223	2.4183	6.0260	6.0027	5.4627	-3.9069	5.4183	-4.6879
2005	6.2886	6.2365	-0.8285	6.3508	0.9893	7.0366	11.8949	6.2335	-0.8762	5.9874	-4.7896
2006	7.1388	6.5844	-7.7660	6.9273	-2.9622	7.2799	1.9770	7.1550	0.2269	6.9851	-2.1530
2007	8.2251	6.9323	-15.7177	7.5562	-8.1325	8.2276	0.0302	8.0985	-1.5392	7.5299	-8.4522
2008	9.2669	7.2803	-21.4375	8.2421	-9.6503	8.9703	-3.2003	8.7695	-5.3675	8.9352	-3.5794
2009	9.1993	7.6282	-17.0783	8.9903	-0.3547	9.1812	-0.1969	8.9244	-2.9883	9.3967	2.1458
2010	9.6327	7.9762	-17.1974	9.8065	4.1871	8.8555	-8.0690	8.5952	-10.7706	9.9820	3.6262
2011	10.3205	8.3241	-19.3441	10.6967	6.4669	9.3514	-9.3901	8.2691	-19.8769	10.3205	0.0000
2012	11.7500	8.6720	-26.1958	11.6677	2.3828	9.3913	-20.0746	8.1174	-30.9157	11.5587	-1.6281

Figure 2 shows the corresponding forecasting curves of these five models in order to make a clear comparison between the proposed QHSA-LSSVM model and other four comparative models. We tested the performance of the proposed QHSA-LSSVM model with other four comparative models (regression model, GM (1, 1), BP, and LSSVM) for training data, testing data and all the sample data, respectively. The details are shown in the next sections.

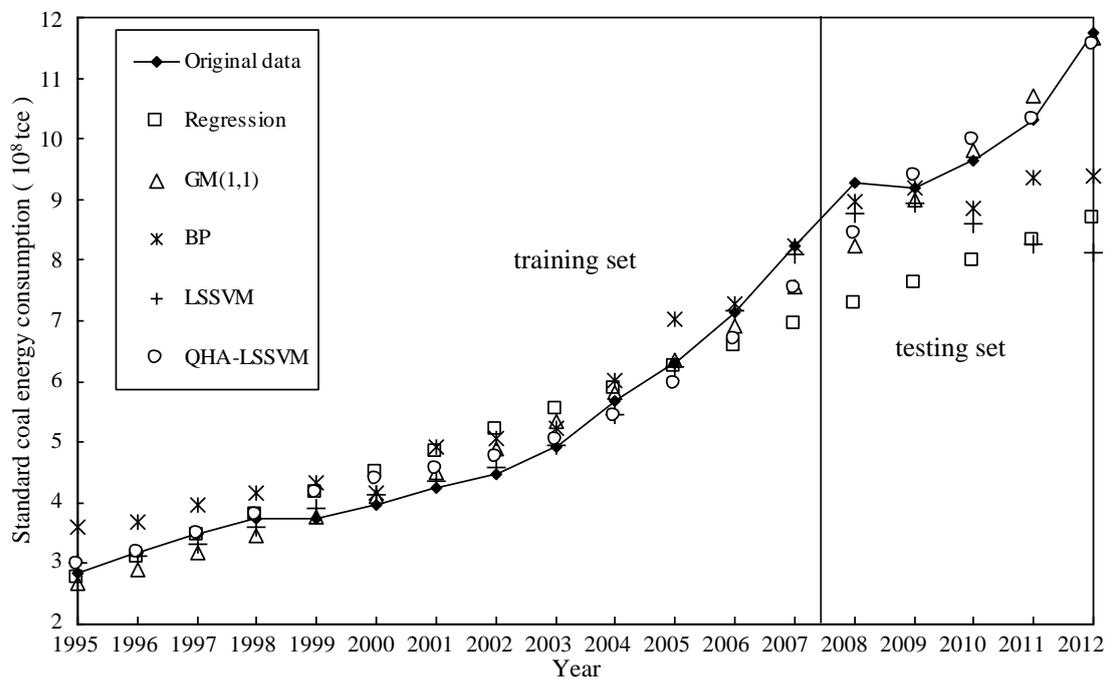


Figure 2. The curves of original data and forecasting results of five models.

4.3.1. Simulation Performance Comparison for Training Data

In our work, the simulation data points and the simulation errors for five models from 1992 to 2007 (training set) have been conducted to assess the models’ fit performance. Table 3 lists the comparison of simulation performance among five models for training set, and Figure 3 shows the corresponding error histograms for direct observation.

Table 3. The comparison of simulation performance among five models for training set.

Model	Regression	GM (1, 1)	BP	LSSVM	QHSA-LSSVM
MAPE	0.0792	0.0575	0.1108	0.0297	0.0378
RMSE	0.5375	0.3094	0.4989	0.1391	0.3314
MAE	0.4026	0.2627	0.4439	0.1226	0.2701
AAE	0.0845	0.0551	0.0932	0.0257	0.0567
MaxAPE	0.1620	0.0951	0.2685	0.0578	0.0845

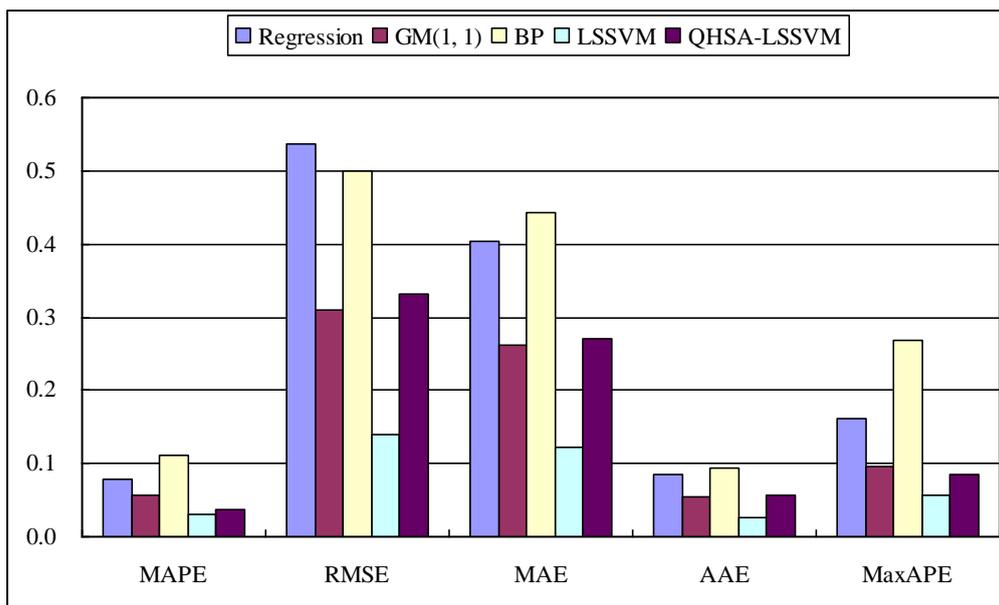


Figure 3. The error histogram for QHSA-LSSVM and other comparative models for training data.

(1) Training Data Point Analysis

In this part, the simulation data points are validated and compared among the five models. The deviations between the simulation results and the actual values are calculated, shown in Table 2. For annual data forecasting, the error range $[-5\%, +5\%]$ is considered as a satisfactory and practical error bound. For 13 data points, there are five simulation data points larger than 5%, and two points smaller than -5% for the regression model. The maximum error is 16.2016% in 2002 and the minimum relative error is -15.7177% in 2007. In the GM (1, 1) model, there are three points larger than 5% and five points lower than -5% . The corresponding maximum and minimum errors are 9.5061% and -9.4621% in 2002 and 1997, respectively. For BP, there are only two data points within the error bound, and the maximum and minimum errors are 26.8492% in 1995 and 0.0302% in 2007. In LSSVM, there is only one simulation data point larger than 5%, and only two points smaller than -5% . The maximum relative error is 5.7759% in 1995, and the minimum relative error is -5.2081% in 1997. In QHSA-LSSVM model, there are only two data points that exceeds the relative error range $[-5\%, +5\%]$, one is larger than 5% and the other is smaller than -5% . The maximum relative error is 6.2457% in 2002, and the minimum relative error is -8.4522% in 2007.

Next, the maximum absolute percentage error (MaxAPE) is compared to measure the simulation risk. The MaxAPE indicator is calculated according to Equation (15), and the results are shown in Table 3.

$$\text{MaxAPE} = \max_t \left(\left| \frac{y_t - \hat{y}_t}{y_t} \right| \right) \times 100, t = 1, 2, \dots, N \tag{15}$$

where y_t is the energy consumption value in the t th year; \hat{y}_t represents its simulating (or forecasting) result for the same period; and N is the number of data.

MaxAPE concentrate on the maximal absolute percentage error which reflects the forecasting risk of choosing one certain model. For the training set, the MaxAPE values are 0.1620, 0.0951, 0.2685, 0.0578 and 0.0845 for regression model, GM (1, 1), BP, LSSVM and QHSA-LSSVM, respectively.

Compared with other four models, the MaxAPE of LSSVM is the smaller than that of other comparative models, especially the proposed QHSA-LSSVM model. Will it really be less risky to choose LSSVM than to choose QHSA-LSSVM not only to simulate the historical training data but also to future unknown data? It is not necessary the case. We should assess the risk both in the simulation process and in prediction process to check the possible overfitting problem in models.

(2) Simulation Performance Comparison and Analysis

In this section, several common-used accuracy measures, including MAPE, root mean square error (RMSE), mean absolute error (MAE) and average absolute error (AAE), are employed to assess the simulation performance in a comprehensive aspect. These error criterion indicators are expressed as follows:

$$\text{MAPE} = \frac{1}{N} \sum_{t=1}^N \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100, t = 1, 2, \dots, N \quad (16)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2}, t = 1, 2, \dots, N \quad (17)$$

$$\text{MAE} = \frac{1}{T} \sum_{t=1}^T |y_t - \hat{y}_t| \quad (18)$$

$$\text{AAE} = \frac{1}{N} \sum_{t=1}^N \frac{|y_t - \hat{y}_t|}{\frac{1}{N} \sum_{t=1}^N y_t}, t = 1, 2, \dots, N \quad (19)$$

Table 3 provides the MAPE, RMSE, MAE and AAE values of all models, which can reflect how well the models, with estimated parameters, fit the data. The non-scaled error metric MAPE is the mean of the absolute percentage errors of forecasts, providing the errors in terms of percentage. It can avoid the problem of positive and negative errors canceling each other out. It can be seen from Table 3 that the MAPE values are 0.0792, 0.0575, 0.1108, 0.0297 and 0.0378 for the regression model, GM (1, 1), BP, LSSVM and QHSA-LSSVM. The RMSE values of regression, GM (1, 1), BP, LSSVM and QHSA-LSSVM are 0.5375, 0.3094, 0.4989, 0.1391 and 0.3314. The MAE and AAE for these five models are 0.4026, 0.2627, 0.4439, 0.1226, 0.2701 and 0.0845, 0.0551, 0.0932, 0.0257, 0.0567, respectively. For the training data, the LSSVM model shows smaller error values of MAPE, RMSE, MAE and AAE than those of QHSA-LSSVM and GM (1, 1) model. The worst two models in simulation performance are the regression model and BP neural network model. Although, the RMSE, MAE and AAE error values of GM (1, 1) model are smaller than those of QHSA-LSSVM, the values of MAPE and MaxAPE are larger than those of QHSA-LSSVM, so it is a little difficult to determine the overall simulation performance difference between GM (1, 1) and QHSA-LSSVM. The forecasting performance of these two models will be compared in the next section. It seems that the QHSA-LSSVM model for the training data shows no obvious advantages over the LSSVM and GM (1, 1) models, but the most important aim for model construction is to forecast future data accurately. Whether a model is good or not lies in the predictive ability or the extrapolation ability, not merely the simulation ability for training data. Next, the forecasting performances of these models are compared in details for testing data.

4.3.2. Forecasting Performance Comparison for Testing Data

The purpose for building a model is to apply it to extend future forecasts, so it is necessary to identify which model may prove accurate and be useful for forecasting. Table 4 lists the different error criterion indicator values of these five models for the testing set, and the corresponding error histogram is drawn in Figure 4.

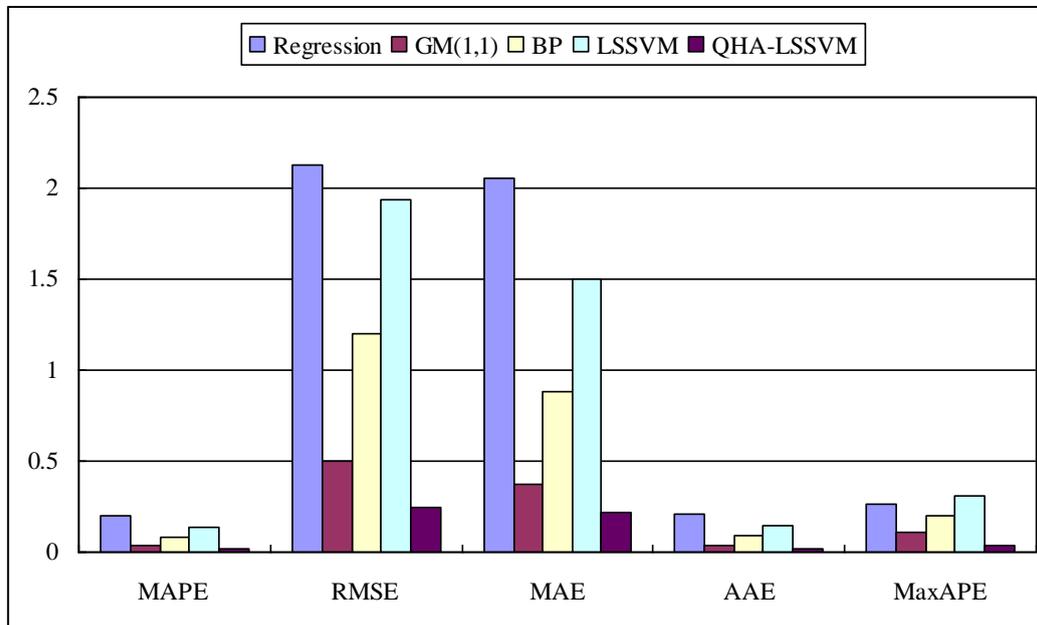


Figure 4. The error histogram for QHSA-LSSVM and other comparative models for testing data.

Table 4. The comparison of forecasting accuracy indicators among five models for testing set.

Model	Regression	GM (1, 1)	BP	LSSVM	QHSA-LSSVM
MAPE	0.2025	0.0390	0.0819	0.1398	0.0220
RMSE	2.1270	0.5045	1.1996	1.9392	0.2480
MAE	2.0578	0.3732	0.8840	1.4988	0.2139
AAE	0.2051	0.0372	0.0881	0.1494	0.0213
MaxAPE	0.2620	0.1106	0.2007	0.3092	0.0363

(1) Testing Data Point Analysis

It can be seen from Table 2 that all five forecasting data points are all within the error bound $[-5\%, +5\%]$ in the proposed QHSA-LSSVM model. There are three data points exceeding the -5% lower bound in BP. In the GM (1, 1) model, there is one point larger than $+5\%$ and one point smaller than -5% . There is only one satisfactory forecasting data point within the error bound for LSSVM and none for the regression model. The maximum error is 3.6262% in 2010 and the minimum error is 0 in 2011 for QHSA-LSSVM. The maximum errors and minimum errors for other four comparative models are larger than the proposed model. For testing set, the MaxAPE value of QHSA-LSSVM model is 0.0363, which is much smaller than that of other models. Therefore, the proposed QHSA-LSSVM model shows better forecasting ability on data point analysis.

(2) Forecasting Performance Comparison and Analysis for Testing Data

Table 4 lists the MAPE, RMSE, MAE and AAE error values of the five models. The MAPE values are 0.2025, 0.0390, 0.0819, 0.1398 and 0.0220 for the regression model, GM (1, 1), BP, LSSVM and QHSA-LSSVM, respectively. The RMSE error values are 2.1270, 0.5045, 1.1996, 1.9392 and 0.2480; the MAE values are 2.0578, 0.3732, 0.8840, 1.4988 and 0.2139; the AAE values are 0.2051, 0.0372, 0.0881, 0.1494 and 0.0213. It can be seen that the MAPE, RMSE, MAE and AAE values obtained by QHSA-LSSVM are much smaller than those obtained by the other four comparative models. It indicates that the proposed QHSA-LSSVM model shows better prediction accuracy and satisfactory forecasting performance for testing data. At the meantime, it is found that the MAPE, RMSE, MAE and AAE error values of LSSVM model are worse for the testing data than those for the training set. We can infer that the overfitting phenomenon occurs in the LSSVM model with small training samples, which may decrease the model's forecasting performance. The LSSVM model looks better because it's making absolutely small errors on the training (or learning) data, shown in Table 3, but when the LSSVM model is applied to a new dataset (the testing set), it does not perform as well on the testing data, with larger errors, shown in Table 4. The generalization ability of LSSVM model is poor. In this case, the LSSVM model overfits the training data; thus, it is not suitable for future prediction. The proposed QHSA-LSSVM model just makes good balance on training set and testing set. It has a relatively good accuracy on the learning data and the best accuracy on the testing data. Overall, the proposed QHSA-LSSVM model is suitable for both data simulation and future trend prediction. The prediction performance of GM (1, 1) for testing data is worse than the simulation performance for the training data. This model has ideal predictive effect for approximate homogenous exponential sequence; otherwise, it may result in larger forecasting error. Moreover, GM (1, 1) is not suitable for long time prediction periods. The forecasting performances of the regression model and BP for the testing data are the worst ones among these five models.

4.3.3. Overall Accuracy Comparison for All Sample Data

Besides, a comprehensive accuracy comparison of these five models for all sample data has been conducted. The values of five error criterion indicators, *i.e.*, MAPE, RMSE, MAE, AAE and MaxAPE for all sample data are listed in Table 5; and the corresponding error histogram is shown in Figure 5.

Table 5. The comparison of forecasting accuracy indicators among five models for all sample data.

Model	Regression	GM (1, 1)	BP	LSSVM	QHSA-LSSVM
MAPE	0.1134	0.0524	0.1028	0.0603	0.0334
RMSE	1.2105	0.3739	0.7613	1.0289	0.2563
MAE	0.8624	0.2934	0.5661	0.5049	0.2018
AAE	0.1385	0.0471	0.0909	0.0811	0.0324
MaxAPE	0.2620	0.1106	0.2685	0.3092	0.0845

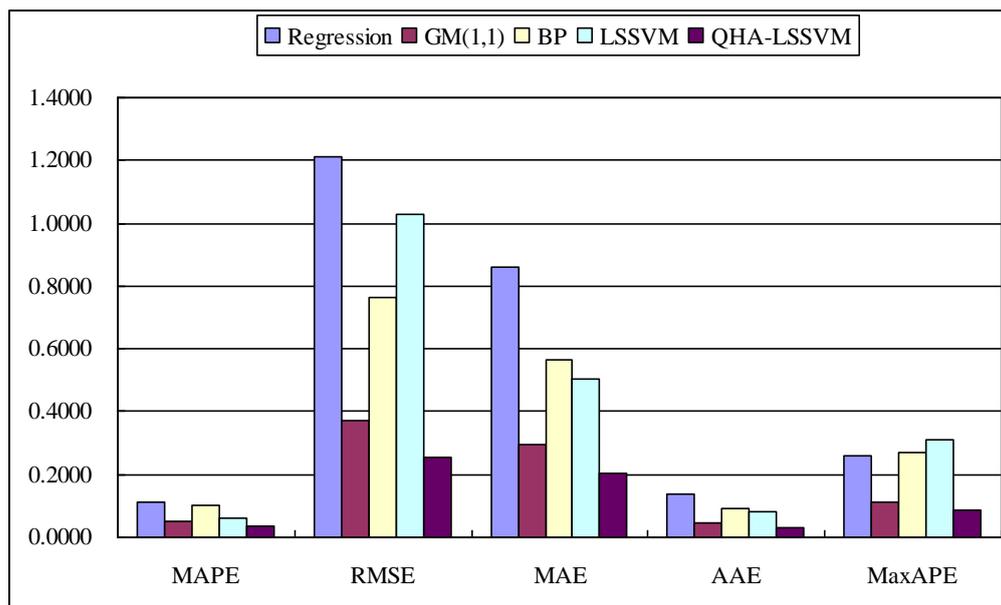


Figure 5. The error histogram for QHSA-LSSVM and other comparative models for all sample data.

As shown in Table 5, the MAPE for regression, GM (1, 1), BP, LSSVM and QHSA-LSSVM are 0.1134, 0.0524, 0.1028, 0.0603, and 0.0334. Compared with MAPE which aim is to assess average values of absolute percentage errors, RMSE is the square root of MSE; and MSE is the average value of the “total square error” which is the sum of the individual squared errors. The stated rationale for squaring each error is to remove the sign so that the “magnitude” of the errors influences the average error measure. For this indicator, large errors normally have a relatively greater influence on the “total square error” than do the smaller errors. In our work, the RMSE values of regression model, GM (1, 1), BP and LSSVM are 1.2105, 0.3739, 0.7613, 1.0289; while the RMSE of QHSA-LSSVM is only 0.2563. It means that other four comparative models may concentrate within a decreasing number of increasingly larger individual errors, thus resulting in larger MSE and RMSE values.

Similarly, MAE involves summing the magnitudes (absolute values) of the errors to obtain the “total error” and then calculating the average value of the “total error”. It is an average of the absolute errors, which summarizes performance in the way that disregards the direction of over- or under-prediction and does place emphasis on the mean signed difference. After calculating according to Equation (18), the MAE values are 0.8624, 0.2934, 0.5661, 0.5049 and 0.2018 for the regression model, GM (1, 1), BP, LSSVM and QHSA-LSSVM, respectively. This also demonstrates there are no obvious larger individual absolute errors between the actual data series and the forecasting data series in the QHSA-LSSVM model.

AAE is a more comprehensive indicator since it can assess the deviation of individual absolute errors from the average value of actual data. The total “deviation” is divided by T , thus obtaining an AAE value. For the four comparative models, the AAE values are 0.1385, 0.0471, 0.0909 and 0.0811, respectively, while the AAE value of QHSA-LSSVM is only 0.0324, which is smaller than that of the four preceding comparative models.

The MaxAPE values of regression model, GM (1, 1), BP and LSSVM model are 0.2620, 0.1106, 0.2685 and 0.3092, respectively; and that of the proposed QHSA-LSSVM is only 0.0845, which means that selecting QHSA-LSSVM has less forecasting risk.

Through errors comparison, the regression errors are almost higher than other three models (*i.e.*, GM (1, 1), BP and LSSVM). It means that the regression model is not suitable to capture the fluctuation increasing trend without constant growth rate in our work. The forecasting performance of GM (1, 1) for all sample data is better than the regression model since the errors of the GM (1, 1) model are lower; and it shows better performance for the training data than the testing data. The errors of BP are almost the highest, which indicates this model is the worst one, except for the regression model. Why does BP show such poor simulation performance not only in training data but also in testing data? The possible reasons are analyzed as follows: the network structure, especially the number of neurons in the hidden layer, has a tremendous influence on the final output. Both the number of hidden layers and the number of neurons in the input and output layers must be carefully considered. In our work, the neurons in the input layers are three, according to the rolling-based method, and the output neuron is the corresponding certain year data. The number in the hidden layer is only adjustable; and there is no theory yet to determine how many hidden neurons are reasonable for better model fitting. Therefore, the fixed number of hidden neurons is the main reason that leads to larger simulation error. Another, training neural network is a complex task using a large number of sample data, with probable long training time and local minimum. A second problem may occur even when the training data is insufficient. It indicates that BP needs a large number of samples to train the network, and is not suitable for small data samples.

For LSSVM, the errors of LSSVM are satisfactory for the training set (1992–2007), and are even smaller than those of our proposed QHSA-LSSVM model. However, the errors for the testing data become much larger, which means it has poor performance in the testing set and is not suitable for future prediction. Moreover, the errors of LSSVM for all sample data are larger than that of QHSA-LSSVM. The most possible reason is that LSSVM model can easily fall into an over-fitting problem, so the prediction ability for future trends is not satisfactory. Another is that the use of the parameters σ and C with fixed values may result in large errors.

For QHSA-LSSVM, it can be observed from Table 5 and Figure 5 that MAPE, RMSE, MAE, AAE and MaxAPE of the proposed QHSA-LSSVM model is the lowest compared with the regression model, GM (1, 1), BP and LSSVM for all the sample data. The comparison findings prove that QHSA-LSSVM performs better than the other four comparative models in terms of forecasting accuracy for the whole data. The proposed QHSA-LSSVM is also suitable for foreseeable forecasting.

In conclusion, we used MAPE, RMSE, MAE, AAE and MaxAPE to validate the forecasting performance of the five forecasting models. Through calculating the values of error criterion indicators for training data, testing data and all the sample data, the overall performance of QHSA-LSSVM is the best with the lowest error values. The results prove that the optimal parameters σ and C determined by QHSA can effectively improve the model's performance and the forecasting accuracy.

5. Conclusions

The difficulty in the LSSVM forecasting technique is how to choose suitable values for parameters σ and C since this can affect the model's learning performance and generalization ability. In order to avoid selecting the parameters randomly and reduce the adjustment process, a novel hybrid Quantum Harmony Search Algorithm-based LSSVM (QHSA-LSSVM) forecasting model is proposed for the

annual power generation standard coal consumption forecasting. The QHSA effectively combines the quantum computation and harmony search algorithm. Based on the concepts and principles of the quantum mechanism, QHSA can improve the global search ability and optimization speed through combining superposition of qubits. In LSSVM, the optimal values of the two parameters are determined through QHSA, which can solve the problem of parameters selection. To validate the performance of the proposed QHSA-LSSVM model, four other comparative models (regression model, GM (1, 1), BP and LSSVM) were employed in our work. Through calculating the error values of MAPE, RMSE, MAE, AAE and MaxAPE of the five models, the errors for QHSA-LSSVM are the smallest, not only for the testing data, but also for all the sample data. This indicates that the QHSA-LSSVM model is suitable for small sample forecasting and can effectively enhance the prediction accuracy. Moreover, the QHSA can determine the optimal parameters values of the LSSVM model, which could solve the random parameter selection problem and decrease the forecasting errors. In all, the proposed QHSA-LSSVM model is satisfactory for future data trend prediction.

Acknowledgments

This work was supported by “Philosophy and Social Science Research of Hebei Province”, “Soft Science Research Base of Hebei Province” and “the Fundamental Research Funds for the Central Universities (12MS137)”.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Lindner, S.; Liu, Z.; Guan, D.; Geng, Y.; Li, X. CO₂ emissions from China’s power sector at the provincial level: Consumption *versus* production perspective. *Renew. Sustain. Energy Rev.* **2013**, *19*, 164–172.
2. National Bureau of Statistics of China. *China Statistical Yearbook*; China Statistics Press: Beijing, China, 2012. (In Chinese)
3. Zhang, M.; Mu, H.; Li, G.; Ning, Y. Forecasting the transport energy demand based on PLSR method in China. *Energy* **2009**, *34*, 1396–1400.
4. Limanond, T.; Jomnonkwao, S.; Srikaew, A. Projection of future transport energy demand of Thailand. *Energy Policy* **2011**, *33*, 2754–2763.
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. Amarawickrama, H.A.; Hunt, L.C. Electricity demand for Sri Lanka: A time series analysis. *Energy* **2008**, *33*, 724–739.
7. Hsu, C.; Chen, C. Applications of improved grey prediction model power demand forecasting. *Energy Convers. Manag.* **2003**, *44*, 2241–2249.
8. Huang, Y.; Bor, Y.J.; Peng, C. The long-term forecast of Taiwan’s energy supply and demand: LEAP model application. *Energy Policy* **2011**, *39*, 6790–6803.

9. Yu, S.; Wei, Y.-M.; Wang, K. A PSO-GA optimal model to estimate primary energy demand of China. *Energy Policy* **2012**, *42*, 329–340.
10. Gurney, K. *An Introduction to Neural Networks*; UCL Press: London, UK, 1997.
11. Anderson, J.A. *An Introduction to Neural Networks*; MIT Press: Cambridge, MA, USA, 1995.
12. Hassoun, M.H. *Fundamentals of Artificial Neural Networks*; MIT Press: Cambridge, MA, USA, 1995.
13. Geem, Z.W.; Roper, W.E. Energy demand estimation of South Korea using artificial neural network. *Energy Policy* **2009**, *37*, 4049–4054.
14. Patterson, D.W. *Artificial Neural Networks*; Prentice Hall: Singapore, Singapore, 1996.
15. Zhang, G.; Patuwo, B.E.; Hu, M.Y. Forecasting with artificial neural networks: The state of the art. *Int. J. Forecast.* **1998**, *14*, 35–62.
16. Kecman, V. *Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models*; MIT Press: Cambridge, MA, USA, 2001.
17. Vapnik, V.N. *The Nature of Statistical Learning Theory*; Springer-Verlag: New York, NY, USA, 1999.
18. Burges, C. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.* **1998**, *2*, 121–167.
19. Cristianini, N.; Shawe-Taylor, J. *An Introduction to Support Vector Machines and other Kernel-Based Learning Methods*; Cambridge University Press: Cambridge, UK, 2000.
20. Cortes, C.; Vapnik, V. Support vector networks. *Mach. Learn.* **1995**, *3*, 273–297.
21. Vapnik, V.N. An overview of statistical learning theory. *IEEE Trans. Neural Netw.* **1999**, *10*, 988–999.
22. Vapnik, V.; Golowich, S.E.; Smola, A. Support vector method for function approximation, regression estimation, and signal processing. In *Advances in Neural Information Processing Systems 9*; MIT Press: Cambridge, MA, USA, 1997; Volume 9, pp. 281–287.
23. Suykens, J.A.K. Nonlinear modeling and support vector machines. In Proceedings of the 18th IEEE Instrumentation and Measurement Technology Conference, Budapest, Hungary, 21–23 May 2001.
24. Li, Y.B.; Li, Y. Survey on uncertainty support vector machine and its application in fault diagnosis. In Proceedings of the 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT), Chengdu, China, 9–11 July 2010.
25. Suykens, J.A.K.; Vandewalle, J. Least squares support vector machine classifiers. *Neural Process. Lett.* **1999**, *9*, 293–300.
26. Suykens, J.A.K.; van Gestel, T.; de Brabanter, J.; de Moor, B.; Vandewalle, J. *Least Squares Support Vector Machines*; World Scientific: Singapore, Singapore, 2002.
27. Suykens, J.A.K.; Lukas, L.; Vandewalle, J. Sparse approximation using least squares support vector machines. In Proceedings of the IEEE International symposium on circuits and systems, Geneva, Switzerland, 28–31 May 2000.
28. Suykens, J.A.K.; de Brabanter, J.; Lukas, L.; Vandewalle, J. Weighted least squares support vector machines-robustness and sparse approximation. *Neurocomputing* **2002**, *48*, 85–105.
29. Suykens, J.A.K.; Vandewalle, J. Recurrent least squares support vector machines. *IEEE Trans. Circuits Syst. I IEEE Trans. Fundam. Theory Appl.* **2000**, *47*, 1109–1114.

30. Suykens, J.A.K.; Lukas, L.; van Dooren, P.; de Moor, B.; Vandewalle, J. Least squares support vector machine classifiers: A large scale algorithm. In Proceedings of the European Conference on Circuit Theory and Design (ECCTD'99), Stresa, Italy, 29 August–2 September 1999.
31. Zhao, X.H.; Wang, G.; Zhao, K.K.; Tan, D.J. On-line least squares support vector machine algorithm in gas prediction. *Min. Sci. Technol.* **2009**, *19*, 194–198.
32. Espinoza, M.; Suykens, J.A.K.; de Moor, B. Fixed-size least squares support vector machines: A large Scale application in electrical load forecasting. *Comput. Manag. Sci.* **2006**, *3*, 113–129.
33. Lin, K.P.; Pai, P.F.; Lu, Y.M.; Chang, P.T. Revenue forecasting using a least-squares support vector regression model in a fuzzy environment. *Inf. Sci.* **2011**, *14*, 196–209.
34. Zhou, D. Estimation of GM (1, 1) model parameter based on LS-SVM algorithm and application in load forecasting. *Mod. Manag.* **2012**, *2*, 45–49.
35. Geem, Z.W.; Kim, J.H.; Loganathan, G.V. A new heuristic optimization algorithm: Harmony search. *Simulation* **2001**, *76*, 60–68.
36. Omran, M.G.H.; Mahdavi, M. Global-best harmony search. *Appl. Math. Comput.* **2008**, *198*, 643–656.
37. Geem, Z.W.; Tseng, C.L. New methodology, harmony search and its robustness. In Proceedings of the Late-Breaking Papers of Genetic and Evolutionary Computation Conference, New York, NY, USA, 9–13 July 2002.
38. Mahdavi, M.; Fesanghary, M.; Damangir, E. An improved harmony search algorithm for solving optimization problems. *Appl. Math. Comput.* **2007**, *188*, 1567–1579.
39. Hey, T. Quantum computing: An introduction. *Comput. Control Eng. J.* **1999**, *10*, 105–112.
40. Nielsen, M.A.; Chuang, I.L. *Quantum Computation and Quantum Information*, 1st ed.; Cambridge University Press: Cambridge, UK, 2000.
41. Han, K.H.; Kim, J.H. Quantum-inspired evolutionary algorithms with a new termination criterion, H gate, and two-phase scheme. *IEEE Trans. Evol. Comput.* **2004**, *8*, 156–169.
42. Han, K.H.; Kim, J.H. Quantum-inspired evolutionary algorithm for a class of combinatorial optimization. *IEEE Trans. Evolut. Comput.* **2002**, *6*, 580–593.
43. Esen, H.; Ozgen, F.; Esen, M. Modeling of a new solar air heater through least square support vector machine. *Expert Syst. Appl.* **2009**, *36*, 10673–10682.
44. Li, H.; Guo, S.; Zhao, H.; Su, C.; Wang, B. Annual electric load forecasting by a least squares support vector machine with a fruit fly optimization algorithm. *Energies* **2012**, *5*, 4430–4445.
45. Lee, K.S.; Geem, Z.W. A new meta-heuristic algorithm for continues engineering optimization: Harmony search theory and practice. *Comput. Method Appl. Mech. Eng.* **2005**, *194*, 3902–3933.
46. Geem, Z.W.; Kim, J.-H.; Loganathan, G.V. Harmony search optimization: Application to pipe network design. *Int. J. Model. Simul.* **2002**, *22*, 125–133.
47. Al-Betar, M.A.; Khader, A.T.; Gani, T.A. A harmony search algorithm for university course timetabling. *Ann. Oper. Res.* **2012**, *194*, 3–31.
48. Lee, K.S.; Geem, Z.W. A new structural optimization method based on the harmony search algorithm. *Comput. Struct.* **2004**, *82*, 781–798.
49. Benioff, P. The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by Turing machines. *J. Stat. Phys.* **1980**, *22*, 563–591.
50. Feynman, R.P. Simulating physics with computers. *Int. J. Theor. Phys.* **1983**, *21*, 467–488.

51. Chang, H.; Sun, W.; Gu, X. Forecasting energy CO₂ emissions using a Quantum Harmony Search Algorithm-Based DMSFE combination model. *Energies* **2012**, *6*, 1456–1477.
52. Wang, C.M.; Huang, Y.F. Self-adaptive harmony search algorithm for optimization. *Expert Syst. Appl.* **2010**, *37*, 2826–2837.
53. Chang, H.; Gu, X.S. Multi-HM adaptive harmony search algorithm and its application to continuous function optimization. *Res. J. Appl. Sci. Eng. Technol.* **2012**, *4*, 100–103.
54. Goodwin, P.; Lawton, R. On the asymmetry of the symmetric MAPE. *Int. J. Forecast.* **1999**, *15*, 405–408.
55. Fang, K.T. *Uniform Design and Uniform Design Table*; Scientific Press: Beijing, China, 1994. (In Chinese)
56. *China Energy Statistical Yearbook*; National Bureau of Statistics of China, National Development and Reform Commission: Beijing, China, 1993–2013. (In Chinese)
57. *General Principles for Calculation of the Comprehensive Energy Consumption*; GB/T 2589-2008; China Standard Press: Beijing, China, 2008. (In Chinese)

© 2015 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/4.0/>).