

Article

A Robust Weighted Combination Forecasting Method Based on Forecast Model Filtering and Adaptive Variable Weight Determination

Lianhui Li ^{1,†}, Chunyang Mu ^{2,*,†}, Shaohu Ding ^{1,*}, Zheng Wang ³, Runyang Mo ^{4,5} and Yongfeng Song ^{4,6}

Received: 12 August 2015; Accepted: 19 October 2015; Published: 31 December 2015

Academic Editor: Paul Flikkema

¹ College of Mechatronic Engineering, Beifang University of Nationalities, Yinchuan 750021, China; 2010100407@mail.nwpu.edu.cn

² State Key Laboratory of Robotics and System, Harbin Institute of Technology (HIT), Harbin 150001, China

³ State Grid Ningxia Electric Power Design Co. Ltd., Yinchuan 750001, China; zhengw1977@gmail.com

⁴ School of Management, Qingdao Technological University, Qingdao 266520, China; morunyang@126.com (R.M.); songyf_ny@126.com (Y.S.)

⁵ College of Electrical & Information Engineering, Hunan University, Changsha 410082, China

⁶ College of Electrical Engineering and Information, Sichuan University, Chengdu 610065, China

* Correspondence: mcybfmd@gmail.com (C.M.); dshaohu@gmail.com (S.D.); Tel.: +86-951-206-6853 (C.M.); +86-188-9507-9199 (S.D.)

† These authors contributed equally to this work.

Abstract: Medium-and-long-term load forecasting plays an important role in energy policy implementation and electric department investment decision. Aiming to improve the robustness and accuracy of annual electric load forecasting, a robust weighted combination load forecasting method based on forecast model filtering and adaptive variable weight determination is proposed. Similar years of selection is carried out based on the similarity between the history year and the forecast year. The forecast models are filtered to select the better ones according to their comprehensive validity degrees. To determine the adaptive variable weight of the selected forecast models, the disturbance variable is introduced into Immune Algorithm-Particle Swarm Optimization (IA-PSO) and the adaptive adjustable strategy of particle search speed is established. Based on the forecast model weight determined by improved IA-PSO, the weighted combination forecast of annual electric load is obtained. The given case study illustrates the correctness and feasibility of the proposed method.

Keywords: load forecasting; robustness; combination forecast; Markov chain; normal cloud model; immune algorithm; particle swarm optimization

1. Introduction

Nowadays, the strong smart grid (SSG) is vigorously being constructed and the renewable distributed electricity generation capacity is steadily increasing. As an important basis to ensure the security and stable operation of the electric system, electric load forecasting is playing a more and more important role in the implementations of energy policies and the investment decision-making of the electric department under this background [1–10]. However, medium-and-long-term load forecasting has non-linear characteristics caused by the influence of various factors (e.g., national policy, economic and social factors) [7–9]. It makes the medium-and-long-term load forecasting much complex and uncertain. Thus, how to improve the robustness and accuracy of annual electric load forecasting is very worthy of study.

On the one hand, the forecast accuracy has reached a high level under the given sample and condition in the existing research [1–13], but the forecast method's robustness currently becomes the bottleneck. The main reason lies in the status that the existing methods are mostly based on the error theory. By the error theory, the fact of the unknown amount of the forecast year's true load is neglected and the accurate forecast error is difficult to obtain. Chen [14,15] pointed out that the validity degree of the forecast model can be expressed by its full and average precision. In the mathematical sense, any forecast model has its inherent attributes which can be measured by its validity rather than the result error reported by Sun *et al.* [16], Chen *et al.* [17] and Jin *et al.* [18]. At the same time, the single forecast models should be filtered so that the better ones will be selected and the worse ones will be eliminated. Therefore, filtering the forecast models to select the better ones based on their comprehensive validity degrees can improve the robustness of the forecast method.

On the other hand, the general load forecast method mainly includes artificial neural network reported by Hernandez *et al.* [19] and Gofman *et al.* [20], regression analysis reported by Li *et al.* [21] and time series analysis reported by Paparoditis *et al.* [22]. The exponential smoothing method reported by Weron *et al.* [23], the gray forecast method reported by Li *et al.* [24] based on time trend extrapolation reported by Ismail *et al.* [25], the clustering forecast method reported by Kodogiannis *et al.* [26] and multiple regression analysis method reported by Hong *et al.* [27] based on the load related factor analysis cannot ensure the satisfactory result in any case. In order to make full use of the advantages and the contained information of each single forecast model, combination forecast [28–33] is an effective method. The question of how to determine the weight assignment of single forecast method is a difficult point in combination forecasting. The constant weight and the variable weight are two common weight determination ways, and the variable weight has a better adaptability. Focusing on this point, a large number of research has been carried out such as mathematical programming method reported by Ma *et al.* [34], genetic algorithm reported by Chaturvedi *et al.* [2], Bayesian method reported by Niu *et al.* [35] and neural network method reported by Hernandez *et al.* [19] and Gofman *et al.* [20]. These methods mostly have the stable performance and the accuracy meeting the application requirements, but there are still several problems such as much complexity, slow convergence or strong status dependence.

In this paper, we propose a robust weighted combination forecasting method based on forecast model filtering and adaptive variable weight determination. Firstly, the similar years are selected from the sample history years according to the similarity between the history year and the forecast year. Secondly, the forecast models are filtered based on the comprehensive validity degree which is composed of the fitted validity degree in the history year interval and the estimated forecast validity degree in the forecast year interval. Thirdly, the improved Immune Algorithm-Particle Swarm Optimization (IA-PSO), in which the disturbance variable is introduced and the adaptive adjustable strategy of particle search speed is established, is used to determine the adaptive variable weight of the selected forecast models. Lastly, the weighted combination forecast is carried out. The flowchart of the proposed method is shown in Figure 1.

2. Similar Years Selection

There are a series of factors influencing the annual load. For example, in order to reflect the influence of each factor on the load forecasting result, Zhu *et al.* [36] have investigated an Artificial Neural Network-based approach for medium-and-long-term load forecasting. In the proposed three-layer back propagation network, seven factors are selected as inputs which include Gross Domestic Product (GDP), heavy industry production, light industry production, agriculture production, primary industry, secondary industry and tertiary industry; Wang *et al.* [37] have pointed out that there are mainly eight factors affecting the annual load: area GDP, primary industry GDP ratio, secondary industry GDP ratio, tertiary-industry GDP ratio, power consumption per unit of GDP, electricity price, urban per-capita income and rural per-capita income; Lei *et al.* [38] have analyzed the variation characteristics of the annual maximum load, annual minimum load and typical daily

load based on the recent historical load data and meteorological data of Chongqing region. Then, they have studied the interrelationship between load characteristics and major influencing factors. The results show that the temperature, rainfall, holidays and festivals have a significant influence on the region power load; Liao *et al.* [39] have researched the current load characteristics of Changde region and main factors influencing load variation. Influencing extents of main influencing factors on regional load, respective proportions of these factors in the influences and the time periods influenced by these factors are analyzed, and the quantization analysis on the relation between load and influencing factors is performed. In conclusion, the factors considered by Wang *et al.* [37] are more than Zhu *et al.* [36], and the factors considered by Lei *et al.* [38] and Liao *et al.* [39] relate to the characteristics in short-term load forecasting. After comparison and summary, we choose the eight factors pointed out by Wang *et al.* [37] which are more comprehensive and reasonable than others to select the similar years in medium-and-long-term load forecasting.

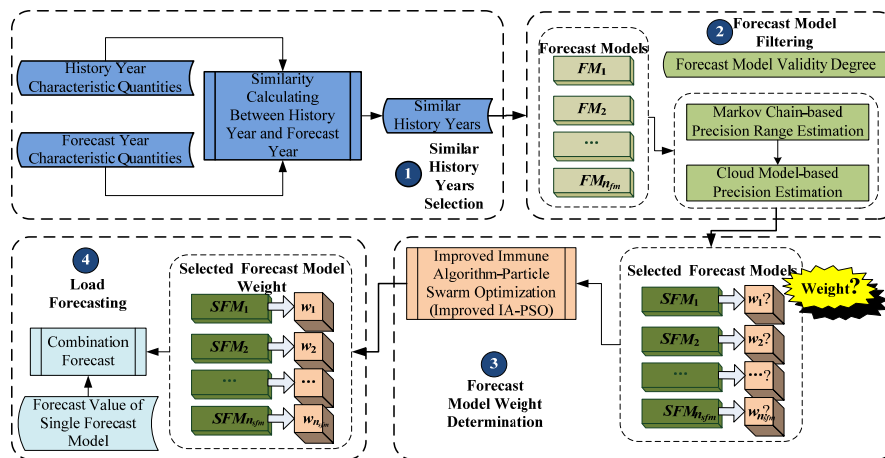


Figure 1. The flowchart of the proposed robust weighted combination forecasting method.

It is assumed that there are n_{his} history years $\{1, 2, \dots, n_{his}\}$ and n_{fo} forecast years $\{n_{his} + 1, n_{his} + 2, \dots, n_{his} + n_{fo}\}$. The year characteristic is defined as the factor affecting the annual load and the year characteristic quantity is defined as the value of year characteristic in a history year.

If the year characteristic CH_a is efficiency type, the year characteristic quantity $CHQ_{a,b}$ which means the value of the year characteristic CH_a of the history year HY_b is standardized as follows:

$$CHQ_{a,b} = \frac{CHQ_{a,b}}{\max \{CHQ_{a,1}, CHQ_{a,2}, \dots, CHQ_{a,n_{his}}\}} \quad (1)$$

where $a = 1, 2, \dots, n_{ch}$, $b = 1, 2, \dots, n_{his}$.

If the year characteristic CH_a is cost type, the year characteristic quantity $CHQ_{a,b}$ which means the value of the year characteristic CH_a of the history year HY_b is standardized as follows:

$$CHQ_{a,b} = \frac{\min \{CHQ_{a,1}, CHQ_{a,2}, \dots, CHQ_{a,n_{his}}\}}{CHQ_{a,b}} \quad (2)$$

where $a = 1, 2, \dots, n_{ch}$, $b = 1, 2, \dots, n_{his}$.

For two individuals whose characteristics have the same dimension number, distance and similarity are usually used to measure the difference between them. The distance measures the absolute distance between two individuals in space which is directly related to the position coordinates of each individual (*i.e.*, the numerical value of each characteristic dimension), but the similarity measures the angle between two individual vectors and reflects more difference in direction that in

difference in position [40–42]. Therefore, similarity is more suitable to measure the difference between a history year and a forecast year.

Here, the most common cosine similarity is chosen [41,42]. The cosine similarity between the history year HY_b and the forecast year FY_c is as follows:

$$CSI_{b,c} = \frac{\sum_{a=1}^{n_{ch}} CHQ_{a,b} \cdot CHQ_{a,c}}{\sqrt{\left(\sum_{a=1}^{n_{ch}} CHQ_{a,b}^2\right) \cdot \left(\sum_{a=1}^{n_{ch}} CHQ_{a,c}^2\right)}} \quad (3)$$

where $b = 1, 2, \dots, n_{his}$, $c = n_{his} + 1, n_{his} + 2, \dots, n_{his} + n_{fo}$.

Lastly, n_{shis} ($n_{shis} < n_{his}$) history years with the highest similarity are selected as the similar years of the forecast year FY_c .

3. Forecast Model Filtering

3.1. Forecast Model Validity Degree

As shown in Figure 2, the load sequence of similar years is $\{y'_1, y'_2, \dots, y'_{n_{shis}}\}$, and the load sequence of forecast years is $\{y_{n_{his}+1}, y_{n_{his}+2}, \dots, y_{n_{his}+n_{fo}}\}$. We assume that there are n_{fm} forecast models. By the forecast model FM_d ($d = 1, 2, \dots, n_{fm}$), $y'_{e,d}$ is the fitted value of the similar year e ($e = 1, 2, \dots, n_{shis}$), and $y_{c,d}$ is the forecast value of the forecast year c ($c = n_{his} + 1, n_{his} + 2, \dots, n_{his} + n_{fo}$).

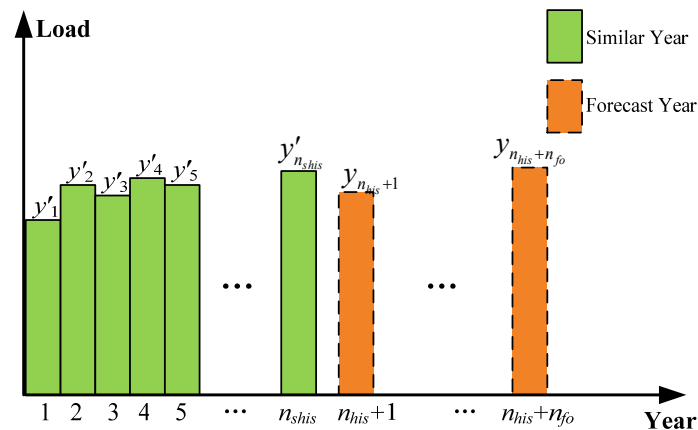


Figure 2. The history years and the forecast years.

The fitted value relative error of FM_d for the similar year e is as follows:

$$RE'_{e,d} = \frac{y'_e - y'_{e,d}}{y'_e} \quad (4)$$

The forecast value relative error of FM_d for the forecast year c is as follows:

$$RE_{c,d} = \frac{y_c - y_{c,d}}{y_c} \quad (5)$$

Then, the fitted precision of FM_d for the similar year e is as follows:

$$P'_{e,d} = \begin{cases} 1 - |RE'_{e,d}|, & 0 \leq |RE'_{e,d}| \leq 1 \\ 0, & |RE'_{e,d}| \geq 1 \end{cases} \quad (6)$$

The forecast precision of FM_d for the forecast year c is as follows:

$$P_{c,d} = \begin{cases} 1 - |RE_{c,d}|, & 0 \leq |RE_{c,d}| \leq 1 \\ 0, & |RE_{c,d}| \geq 1 \end{cases} \quad (7)$$

Lastly, the fitted validity degree of FM_d is as follows:

$$FIV_d = \text{EXP}(P'_{e,d}) \cdot (1 - \sigma(P'_{e,d})) \quad (8)$$

where $\text{EXP}(P'_{e,d}) = \frac{1}{n_{shis}} \sum_{e=1}^{n_{shis}} P'_{e,d}$ and $\sigma(P'_{e,d}) = \sqrt{\frac{1}{n_{shis}} \sum_{e=1}^{n_{shis}} (P'_{e,d} - \text{EXP}(P'_{e,d}))^2}$ are the expectation and the standard deviation of the fitted precision of FM_d for the similar year e , respectively.

The forecast validity degree of FM_d is as follows:

$$FOV_d = \text{EXP}(P_{c,d}) \cdot (1 - \sigma(P_{c,d})) \quad (9)$$

where $\text{EXP}(P_{c,d}) = \frac{1}{n_{fo}} \sum_{c=n_{his}+1}^{n_{his}+n_{fo}} P_{c,d}$, $\sigma(P_{c,d}) = \sqrt{\frac{1}{n_{fo}} \sum_{c=n_{his}+1}^{n_{his}+n_{fo}} (P_{c,d} - \text{EXP}(P_{c,d}))^2}$ are the expectation and the standard deviation of the forecast precision of FM_d for the forecast year c , respectively.

3.2. Forecast Model Precision Estimation

The validity degree of a forecast model is defined in Equation (9) which can depict its credibility and is a reflection of its inherent property [16–18]. Obviously, true value has not yet occurred and the forecast error cannot be obtained in the future forecast interval. We can only estimate the precision and the validity of a forecast model based on its inherent property. Then, the suitable forecast models are selected and the combination forecast model is put forward.

3.2.1. Markov Chain-Based Precision Range Estimation

As an inherent property, the forecast model precision is shown in the form of the fitted precision which is obtained through the virtual forecast for the multi-time load. Using the forecast model FM_d to forecast the load of the similar year e ($e = 1, 2, \dots, n_{shis}$), we can obtain the fitted precision sequence $\{P'_{1,d}, P'_{2,d}, \dots, P'_{n_{shis},d}\}$. In this sequence, the expectation and the standard deviation of the fitted precision of each similar year show the property of the forecast model. As is known, randomness and discreteness appear in the fitted precision sequence. Therefore, we can use the Markov chain transition matrix [43,44] to represent the transition rule as follows:

(1) The fitted precision distribution interval of FM_d for the similar year e is divided into n_{si} ($n_{si} \leq n_{shis}$) sub-intervals with equal distance as $S_d^1, S_d^2, \dots, S_d^{n_{si}}$, where

$$\begin{aligned} S_d^g &= [\underline{S}_d^g, \overline{S}_d^g], \underline{S}_d^1 = \min \{P'_{1,d}, P'_{2,d}, \dots, P'_{n_{shis},d}\}, \overline{S}_d^{n_{si}} = \max \{P'_{1,d}, P'_{2,d}, \dots, P'_{n_{shis},d}\}, \\ \overline{S}_d^g - \underline{S}_d^g &= \overline{S}_d^h - \underline{S}_d^h \quad s.t. \quad g \neq h, g, h = 1, 2, \dots, n_{si} \end{aligned} \quad (10)$$

Each fitted precision sub-interval can be considered as a fitted precision state.

(2) According to the fitted precision of FM_d for the similar year e , the occurrence number of the fitted precision state S_d^g is OC^g ($OC^g < n_{shis}$). That means there are OC^g times which belong to the fitted precision state S_d^g . The transition number from the fitted precision state S_d^h to S_d^g is $TRN^{h,g}$. Thus, the transition probability of FM_d from the fitted precision state S_d^h to S_d^g can be obtained as follows:

$$TP_d^{h,g} = \left\{ S_d^g = c_g \mid S_d^h = c_h \right\} = \frac{TRN^{h,g}}{OC^g} \quad (11)$$

According to Equation (11), the one-step state transition matrix of FM_d is as follows:

$$TM_d^{(1)} = \begin{bmatrix} TP_d^{1,1} & TP_d^{1,2} & \dots & TP_d^{1,n_{si}} \\ TP_d^{2,1} & TP_d^{2,2} & \dots & TP_d^{2,n_{si}} \\ \vdots & \vdots & & \vdots \\ TP_d^{n_{si},1} & TP_d^{n_{si},2} & \dots & TP_d^{n_{si},n_{si}} \end{bmatrix} \quad (12)$$

The q -step state transition matrix is as follows:

$$TM_d^{(q)} = (TM_d^{(1)})^q \quad (13)$$

(3) We construct an initial vector IV_d whose elements are the occurrence numbers of each fitted precision state of FM_d . Though multiplying the initial vector IV_d with the q -step state transition matrix $TM_d^{(q)}$, the new state matrix of FM_d is obtained as follows:

$$SM_d = IV_d \cdot TM_d^{(q)} \quad (14)$$

(4) We calculate the sum of each column vector of SM_d . If the column vector CVE_i ($i = 1, 2, \dots, n_{si}$) has the maximum sum, the forecast precision will belong to the precision state S_d^i .

3.2.2. Cloud Model-Based Precision Estimation

Due to the various factors, the fitted precision sequence $P'_d = \{P'_{1,d}, P'_{2,d}, \dots, P'_{n_{shis},d}\}$ of the forecast model FM_d in n_{shis} similar history years clearly has the property of random variables. As a result, the forecast precision is uncertain in its precision range. Therefore, we can describe this uncertainty by the expectation, entropy and hyper-entropy in the precision range and make the quantitative precision estimation based on the normal cloud model [45,46] as follows:

(1) Firstly, we construct a backward cloud generator. We can map the fitted precision sequence $P'_d = \{P'_{1,d}, P'_{2,d}, \dots, P'_{n_{shis},d}\}$ into the normal cloud model. In this normal cloud model, the input is P'_d and the output is the expectation Ex_d , the entropy En_d and the hyper-entropy He_d . The algorithm of this backward cloud generator is as follows:

$$Ex_d = \text{EXP}(P'_d) \quad (15)$$

$$En_d = \sqrt{\frac{\pi}{2}} \cdot \frac{1}{n_{shis}} \sum_{e=1}^{n_{shis}} |P'_{e,d} - Ex_d| \quad (16)$$

$$He_d = \sqrt{\frac{1}{n_{shis} - 1} \sum_{e=1}^{n_{shis}} (P'_{e,d} - Ex_d)^2 - (En_d)^2} \quad (17)$$

(2) Secondly, we construct a forward cloud generator. The input is Ex_d, En_d, He_d and the constraint is the precision range S_g^d . The algorithm using the forward cloud generator for precision estimation is as follows:

$$P_{c,d} = \text{NORM}(Ex_d, En'_d) \quad (18)$$

where $En'_d = \text{NORM}(En_d, He_d)$ is a normal random number with the expectation En_d and the variance He_d , $P_{c,d}$ is the estimated forecast precision with the expectation Ex_d and the variance En'_d in the precision range S_g^d .

3.3. Forecast Model Filtering Based on Comprehensive Validity Degree

The comprehensive validity degree of the forecast model FM_d in the whole interval $[1, n_{shis}] \cup [n_{his} + 1, n_{his} + n_{fo}]$ is as follows:

$$CV_d = \alpha \cdot FIV_d + (1 - \alpha) \cdot FOV_d \quad (19)$$

where FIV_d is the fitted validity degree of FM_d in the similar history years interval $[1, n_{shis}]$, FOV_d is the estimated forecast validity degree of FM_d in the forecast years interval $[n_{his} + 1, n_{his} + n_{fo}]$.

The empirical coefficient α ($0 \leq \alpha \leq 1$) is determined by the forecast staff based on their experiences. The bigger α is, the more important FIV_d is. Here $\alpha = 0.5$, which means that FIV_d and FOV_d are equally important.

We use the mean comprehensive validity degree of n_{fm} forecast models as the forecast model filtering threshold:

$$\overline{CV} = \sum_{d=1}^{n_{fm}} CV_d \quad (20)$$

If $CV_d \geq \overline{CV}$, the forecast model FM_d will be selected for the combination forecast, else it will be eliminated. In the applications, the filtering threshold can be adjusted according to the actual situation and forecast decision-makers' experiences.

4. Forecast Model Weight Determination and Combination Forecast

4.1. Mathematical Description

After the forecast models filtering, n_{sfm} ($n_{sfm} < n_{fm}$) better forecast models are selected from n_{fm} forecast models for combination forecast.

We assume that the weight of the selected forecast model SFM_j ($j = 1, 2, \dots, n_{sfm}$) is ω_j ($\sum_{j=1}^{n_{sfm}} \omega_j = 1, 0 \leq \omega_j \leq 1$). For the similar history year e ($e = 1, 2, \dots, n_{shis}$), the actual load is y'_e and the forecast load by SFM_j is $y''_{e,j}$, so the combination forecast load of n_{sfm} selected forecast models is as follows:

$$y'''_e = \sum_{j=1}^{n_{sfm}} \omega_j y''_{e,j} \quad (21)$$

The target is to achieve the minimum square sum of the combination forecast error, so the mathematical description is described as follows:

$$\min \sum_{e=1}^{n_{shis}} \left(\sum_{j=1}^{n_{sfm}} \omega_j y''_{e,j} - y'_e \right)^2, \quad \text{s.t.} \sum_{j=1}^{n_{sfm}} \omega_j = 1, 0 \leq \omega_j \leq 1 \quad (22)$$

4.2. Improved Immune Algorithm-Particle Swarm Optimization (Improved IA-PSO)

4.2.1. Particle Swarm Optimization (PSO)

As a kind of stochastic optimization algorithm, particle swarm optimization (PSO) [47] is developed based on the simulation of bird-group foraging behavior. To search for the optimal solution, the individuals have to cooperate and compete with each other in PSO.

There is an m -dimensional space and an initial swarm PO composed of n_{pa} particles as follows:

$$PO = \{PA_1, PA_2, \dots, PA_{n_{pa}}\} \quad (23)$$

For the particle PA_l ($l = 1, 2, \dots, n_{pa}$), its position x_l and its speed v_l are expressed by two m -dimensional vectors as follows:

$$x_l = (x_{l,1}, x_{l,1}, \dots, x_{l,m})^T \quad (24)$$

$$v_l = (v_{l,1}, v_{l,1}, \dots, v_{l,m})^T \quad (25)$$

Each particle is moving in the solution space, and its direction is determined by its speed. The speed and position of the particle is continuously updated as follows:

$$v_l^{(k+1)} = w \cdot v_l^{(k)} + LF_1 \cdot rand_1^{(k)} \cdot (pbest_l^{(k)} - x_l^{(k)}) + LF_2 \cdot rand_2^{(k)} \cdot (Gbest_l^{(k)} - x_l^{(k)}) \quad (26)$$

$$x_l^{(k+1)} = x_l^{(k)} + v_l^{(k+1)} \quad (27)$$

where $v_l^{(k)}$ and $x_l^{(k)}$ are the speed and position of the particle PA_l in the iteration $iter_k$, LF_1 and LF_2 are the learning factors, and $0 \leq rand_1^{(k)}, rand_2^{(k)} \leq 1$ are two random numbers.

- The momentum part $w \cdot v_l^{(k)}$ represents the trust in its current motion state where w is the inertia coefficient used to control the influence of the speed $v_l^{(k)}$ on the speed $v_l^{(k+1)}$. This part provides a necessary momentum which enables the particle to carry on the inertia motion based on its speed.
- The individual cognitive part $LF_1 \cdot rand_1^{(k)} \cdot (pbest_l^{(k)} - x_l^{(k)})$ represents the particle self-thinking behavior. This part encourages the particle to fly to the best position found by itself.
- The social cognitive part $LF_2 \cdot rand_2^{(k)} \cdot (Gbest_l^{(k)} - x_l^{(k)})$ represents the information sharing and cooperation of different particles. This part guides the particle to fly to the best position of the group.

Therefore, the momentum part $w \cdot v_l^{(k)}$ represents the particles' diversification; the individual cognitive part $LF_1 \cdot rand_1^{(k)} \cdot (pbest_l^{(k)} - x_l^{(k)})$ and the social cognitive part $LF_2 \cdot rand_2^{(k)} \cdot (Gbest_l^{(k)} - x_l^{(k)})$ represent the particles' centralization. The main performance of PSO is determined by the balance of the three parts.

In early evolution, PSO has the advantage of fast convergence speed and simple operation, so it can be used for solving the nonlinear, non-differentiable and multi-peak complex optimization problems. But in late evolution, PSO has a significantly slower convergence speed and reaches a poor accuracy, so it is easy to fall into the local optimum.

4.2.2. Immune Algorithm-Particle Swarm Optimization (IA-PSO)

To solve this problem, IA-PSO introduces biological immune system's specific information processing mechanism (e.g., immune memory, immune regulation and immune selection) into PSO's basic framework [48,49].

- Immune memory: The immune system keeps the antibodies opposing against the invading antigen as memory cells. If the same antigen invades again, the memory cells will be activated and produce a large number of antibodies. In IA-PSO, this idea is used to preserve the excellent particle. The best position $pbest_l^{(k)}$ searched by each particle up to now is considered as a memory cell. If the new born particles are detected not to meet the requirements, they will be replaced by the memory cells.
- Immune regulation: In IA-PSO, immune regulation is used for particle selection. If a particle has a strong affinity or a low concentration, it will be promoted. Otherwise, it will be demoted. Therefore, the particle diversification can be always kept. The selected probability [48,49] of the particle PA_l is as follows:

$$PRO_l = \chi \cdot PRO_{l1} + (1 - \chi) \cdot PRO_{l2} \quad (28)$$

In Equation (28), $PRO_{l1} = AF_l / \sum_{u=1}^{\text{all}} AF_u$ represents the selected probability determined by the affinity where AF_l is the affinity of the particle PA_l , $PRO_{l2} = CON_l^{-1} / \sum_{u=1}^{\text{all}} CON_u^{-1}$ represents the selected probability determined by the concentration where CON_l is the concentration of the particle PA_l . χ represents the weight of PRO_{l1} and $1 - \chi$ represents the weight of PRO_{l2} ($0 \leq \chi \leq 1$).

- Immune selection: In the immune system, vaccinating means to change several components of the antibody according to the vaccination. In IA-PSO, the group best position $Gbest_i^{(k)}$ up to the iteration $iter_k$ can be considered as the closest one to the optimal solution. Thus, we use several components of $Gbest_i^{(k)}$ as the vaccination to vaccinate the particles and calculate the particle fitness value for immune selection. If the particle fitness value after the vaccination is lower than its parent, the vaccination will be abolished. Otherwise, the particle will be retained.

IA-PSO, which inherits the global optimization ability of PSO and the immune information processing mechanism of IA, can improve the algorithm accuracy. But at the same time, the algorithm complexity is increased because of the introduction of the immune system.

4.2.3. Improved IA-PSO Based on Disturbance Variable

By introducing a disturbance variable and establishing the searching speed adaptive mechanism, we improve IA-PSO in this paper. Through this improvement, not only the diversity of particles can be ensured to avoid the local optimum, but also the accuracy and convergence speed can be increased.

To solve the mathematical problem described in Equation (22), the search space is set as n_{sfm} -dimensional, the particles number is n_{pa} , and the maximum iteration number is $iter_{\text{Max}}$. The position of each particle is an n_{sfm} -dimensional vector in which each dimensional represents the weight of a selected forecast model. After the iteration $iter_k$ ($iter_k = 1, 2, \dots, iter_{\text{Max}}$), the position and speed of the particle PA_l ($l = 1, 2, \dots, n_{pa}$) are as follows:

$$x_l^{(k)} = (x_{l,1}^{(k)}, x_{l,2}^{(k)}, \dots, x_{l,n_{sfm}}^{(k)})^T \quad (29)$$

$$v_l^{(k)} = (v_{l,1}^{(k)}, v_{l,2}^{(k)}, \dots, v_{l,n_{sfm}}^{(k)})^T \quad (30)$$

$pbest_l^{(k)}$, $gbest^{(k)}$, $Gbest^{(k)}$ are used to represent the best position searched by the particle PA_l in the iteration $iter_k$, the best position searched by the particle swarm in the iteration $iter_k$, the best position searched by the particle swarm up to the iteration $iter_k$, respectively.

Introducing Disturbance Variable into IA-PSO

In the production process of a new particle swarm, the particle position is updated according to Equation (27), and the step length is represented by Equation (26). The coefficients of the three parts in Equation (26) are randomly changed, but only the rules of the particle to follow the order are changed. It means that the step length only reflects the rules of the search behavior and the diversification is lacking. Group optimization should be a balance between the order following and random irrational behaving, so we introduce the disturbance term [50,51] to Equation (26). In each iteration, the particle position is updated as follows:

$$x_l^{(k)} = x_l^{(k-1)} + v_l^{(k)} + (rand_3^{(k)} - 0.5) \cdot \beta_l^{(k)} \quad (31)$$

where $0 \leq rand_3^{(k)} \leq 1$ is a random number subject to uniform distribution, $\beta_l^{(k)} > 0$ is the disturbance variable of the particle PA_l in the iteration $iter_k$. The disturbance term $(rand_3^{(k)} - 0.5) \cdot \beta_l^{(k)}$ reflects an unpredictable random behavior.

The disturbance variable $\beta_l^{(k)}$ controls the random decision-making behavior strength of the particle PA_l in the iteration $iter_k$. If $\beta_l^{(k)}$ is too big, the awareness of order following will be submerged. If $\beta_l^{(k)}$ is too small, the population diversity and global search ability will be reduced. Therefore, the disturbance variable of each particle should be adjusted in the algorithm operation according to its evolution speed. The adjustment can make the particle swarm with a good generalization ability and convergence speed in the evolution process. Thus $\beta_l^{(k)}$ is defined as follows:

$$\beta_l^{(k)} = \beta_{\min} \cdot \frac{F(pbest_l^{(k-1-\theta)}) - F(pbest_l^{(k-1-\rho)})}{F(pbest_l^{(k-1)}) - F(pbest_l^{(k-1-\theta)})} \quad (32)$$

$$\text{s.t. } pbest_l^{(k-1)} \neq pbest_l^{(k-1-\theta)} \neq pbest_l^{(k-1-\rho)}$$

$$\theta < \rho, \theta = \min\{1, 2, \dots, k\}, \rho = \min\{1, 2, \dots, k\}$$

where β_{\min} is the minimum value of the disturbance variable, $F(\cdot)$ is the fitness function, $pbest_l^{(k-1)}, pbest_l^{(k-1-\theta)}, pbest_l^{(k-1-\rho)}$ is the best position found by the particle PA_l in the iteration $iter_{k-1}, iter_{k-1-\theta}, iter_{k-1-\rho}$. When the evolution begins, β_l is bigger which means that the particle PA_l has a step length with strong randomness. After several iterations, β_l tend to β_{\min} which means that the step length randomness of the particle PA_l becomes weak.

Through the introduction of the disturbance term, Equation (31) reflects the positive and negative sides of the particle updating decision. In Equation (31), the first part $x_l^{(k-1)}$ is the original position, the second part $v_l^{(k)}$ reflects the step length of order following, and the third part $(rand_3^{(k)} - 0.5) \cdot \beta_l^{(k)}$ reflects the step length of random irrational behaving. Due to the disturbance variable, the particle position updating can be ensured and a strong search desire can be kept even if the local optimum appears when compared with Equation (26). As a result, the premature convergence problem can be overcome, the local best solution can be prevented and the algorithm accuracy can be improved.

Establishing the Adaptive Adjustable Strategy of Particle Searching Speed

In the particle searching process, the searching speed should be adaptively adjusted to accelerate the convergence based on the diversity of particles. For the excellent particles, their searching speeds should be decreased to make them quickly be close to the global best solution, and then the convergence can be accelerated. For the poor particles, their searching speed should be adjusted according to the convergence degree of the particle swarm: if the swarm individuals tend to be dispersed, the searching speed should be reduced and the swarm development ability should be enhanced to strengthen the local optimization; if the swarm individuals tend to be converged (the algorithm falls into local optimum), the searching speed should be increased and the swarm detection ability should be enhanced to effectively jump out of the local optimum and achieve the accelerated convergence [42,43].

In iteration $iter_k$, the fitness of the particle PA_l is $PF_l^{(k)}$, the fitness of the best particle is $PF_0^{(k)}$, the average fitness of the swarm is $PF_{avg}^{(k)}$, and the average fitness of the particles whose fitness are bigger than $PF_{avg}^{(k)}$ is $PF_{AVG}^{(k)}$. We use $\Delta^{(k)} = |PF_0^{(k)} - PF_{AVG}^{(k)}|$ to evaluate the swarm convergence degree in the iteration L . According to the particle fitness, the swarm is divided into three sub-swarms: $PF_l^{(k)} > PF_{AVG}^{(k)}$, $PF_{avg}^{(k)} < PF_l^{(k)} < PF_{AVG}^{(k)}$, $PF_l^{(k)} < PF_{avg}^{(k)}$. For the different sub-swarms, we take different adjustment operations to their searching speeds as follows:

$$(1) PF_l^{(k)} > PF_{AVG}^{(k)}$$

As the excellent individuals in the swarm, these particles have been relatively close to the global optimum. Their searching speeds should be reduced to prevent from jumping out of the global optimum. So the searching speed is adjusted as follows:

$$v_{\text{new},l}^{(k)} = \left(1 - 0.5 \frac{PF_l^{(k)} - PF_{\text{AVG}}^{(k)}}{|\Delta^{(k)}|} \right) \cdot v_l^{(k)} \quad (33)$$

If the particle is more excellent, it will have a lower searching speed. As a result, the local optimum ability is strengthened and the convergence is accelerated.

$$(2) PF_{\text{avg}}^{(k)} < PF_l^{(k)} < PF_{\text{AVG}}^{(k)}$$

As the general individuals of the swarm, both the local optimum ability and the global optimum ability of these particles are good. Therefore, we don't adjust their searching speeds.

$$(3) PF_l^{(k)} < PF_{\text{avg}}^{(k)}$$

These particles are the relatively poor individuals in the swarm. The searching speed is adjusted as follows:

$$v_{\text{new},l}^{(k)} = \left(1.5 - \frac{1}{1 + \eta_1 \cdot \exp(-\eta_2 \cdot \Delta^{(k)})} \right) \cdot v_l^{(k)} \quad (34)$$

where $\eta_1, \eta_2 > 0$ and η_1 is used to control the upper limit of $v_{\text{new},l}^{(k)}$. If η_1 is bigger, the upper limit of $v_{\text{new},l}^{(k)}$ will be bigger. Here we choose $\eta_1 = \eta_2 = 4$. $\Delta^{(k)} \geq 0$, so $v_{\text{new},l}^{(k)} \in [0.5 \cdot v_i^{(L)}, 1.3 \cdot v_i^{(L)}]$.

In the searching process, $v_{\text{new},l}^{(k)}$ of these particles is dynamically and adaptively adjusted according to the value of $\Delta^{(k)}$: if the individuals tend to be dispersed ($\Delta^{(k)}$ becomes bigger), $v_{\text{new},l}^{(k)}$ will be reduced and the swarm development ability will be enhanced to strengthen the local optimization; if the individuals tend to be converged ($\Delta^{(k)}$ becomes smaller), $v_{\text{new},l}^{(k)}$ will be increased and the swarm detection ability will be enhanced to effectively jump out of the local optimum.

4.3. Implementation Steps of Forecast Model Weight Determination Based on Improved IA-PSO

The flowchart of forecast model weight determination (FMWD) based on improved IA-PSO is shown in Figure 3.

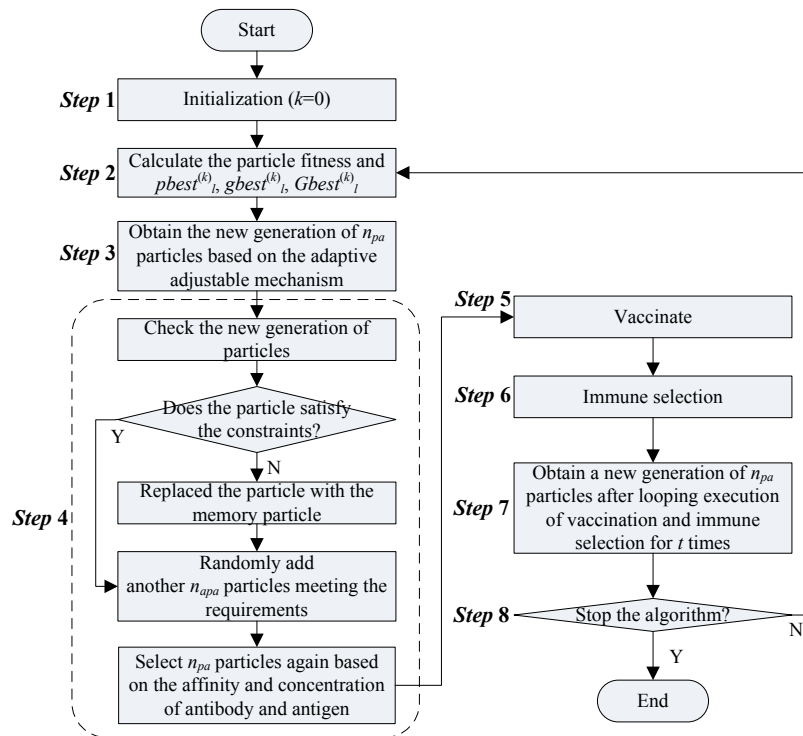


Figure 3. The flowchart of forecast model weight determination based on improved IA-PSO.

The implementation steps are as follows:

Step 1: Initialization. It is assumed that the elements of the particle position vector x all belong to the interval $[0, 1]$, the elements of the particle speed vector v all belong to the interval $[-v_{\max}, v_{\max}]$, the maximum iteration number is $iter_{\max}$ and the initial value of the iteration number k is $k = 0$. n_{pa} particles are randomly generated. The particle PA_l ($l = 1, 2, \dots, n_{pa}$) has the position $x_l^{(0)}$ and the flying speed $v_l^{(0)}$.

Step 2: Calculate $pbest_l^{(k)}$, $gbest_l^{(k)}$, $Gbest_l^{(k)}$ and the fitness of the particle PA_l . Here, the fitness of the particle PA_l can be represented by the target function in Equation (22) as follows:

$$F_l^{(k)} = \sum_{e=1}^{n_{shis}} \left(\sum_{j=1}^{n_{sfm}} x_{l,j}^{(k)} \cdot y''_{e,j} - y'_e \right)^2 \quad (35)$$

Step 3: Obtain the new generation of particles. The particle speed is adjusted based on the adaptive adjustable mechanism and the new position $x_l^{(k+1)}$ and new speed $v_l^{(k+1)}$ can be obtained according to Equations (26) and (31). The elements of the new speed $v_l^{(k+1)}$ must belong to the interval $[-v_{\max}, v_{\max}]$.

Step 4: Check the new generation of particles. If the position of a new particle is an infeasible solution (one or more elements of the new position $x_l^{(k+1)}$ don't belong to the interval $[0, 1]$), this particle will be replaced with the memory particle $pbest_l^{(k)}$.

In addition, another n_{apa} particles meeting the requirements will be randomly added. According to Equation (28), n_{pa} particles are selected from the $n_{pa} + n_{apa}$ particles based on the affinity and concentration of antibody and antigen.

Step 5: Vaccinate. One particle is randomly selected from n_{pa} new particles. Then, one element is randomly selected from the front $t - 1$ elements of $Gbest_l^{(k)}$ to exchange with the selected particle at the corresponding element. The t^{th} element of the selected particle is calculated as follows:

$$x_{l,t}^{(k+1)} = 1 - \sum_{j=1}^{t-1} x_{l,j}^{(k+1)} \quad (36)$$

Hence the vaccination is finished.

Step 6: Immune selection. The particle fitness after the vaccination is calculated. If the particle fitness value after the vaccination is lower than its parent, the vaccination will be abolished. Else, the particle will be retained.

Step 7: Looping execution of Steps 4 and 5 for r times (r times vaccination). A new generation of n_{pa} particles is obtained.

Step 8: Judge whether the algorithm should be stopped. The stopping of algorithm is usually determined by the maximum iteration number or the precision. If the algorithm meets its stopping condition, the optimization will be stopped. Else, $k = k + 1$ and go back to Step 2 to continue.

4.3.1. Weighted Combination Forecast

Based on the improved IA-PSO, the weight ω_j of the selected forecast model SFM_j ($j = 1, 2, \dots, n_{sfm}$) is obtained where $\sum_{j=1}^{n_{sfm}} \omega_j = 1, 0 \leq \omega_j \leq 1$. For the selected forecast model SFM_j , $y''_{c,j}$ is the forecast value of the forecast year c ($c = n_{his} + 1, n_{his} + 2, \dots, n_{his} + n_{fo}$). So the weighted combination forecast (WCF) value of the forecast year c is as follows:

$$y''_c = \sum_{j=1}^{n_{sfm}} \omega_j \cdot y''_{c,j} \quad (37)$$

5. Case Study

We use the proposed method to forecast a Chinese province's load of 2005 based on its loads of the history years from 1998 to 2004. The year characteristic quantity data from 1998 to 2005 is shown in Table 1.

The cosine similarity between the history years 1998–2004 and the forecast year 2005 is shown in Table 2.

Table 1. The year characteristic quantity data from 1998 to 2005 [37,52].

Year	Area GDP (10 ⁸ Yuan)	Primary Industry GDP Ratio (%)	Secondary Industry GDP Ratio (%)	Tertiary Industry GDP Ratio (%)	Power Consumption per Unit of GDP (kWh/Yuan)	Electricity Price (Yuan/kWh)	Urban per-Capita Income (Yuan)	Rural per-Capita Income (Yuan)
1998	9686.6	12.95	50.22	36.83	0.156	0.408	3005.21	1896.56
1999	9802.8	12.90	49.90	37.20	0.152	0.408	3859.86	2003.63
2000	9912.3	12.88	50.07	37.05	0.150	0.410	4663.23	2150.36
2001	10,626.6	12.83	50.06	37.11	0.148	0.412	5551.91	2340.14
2002	11,586.5	12.80	49.69	37.51	0.142	0.412	6599.24	2485.86
2003	12,955.2	12.38	50.75	36.87	0.139	0.412	7370.65	2657.93
2004	15,133.9	12.68	51.63	35.69	0.132	0.419	8245.55	3103.98
2005	17,140.8	12.79	49.62	37.59	0.125	0.444	9227.55	3391.82

Table 2. The cosine similarity between the history years 1998–2004 and the forecast year 2005.

Year	Cosine Similarity (%)
1998	98.56
1999	95.21
2000	94.66
2001	99.60
2002	97.52
2003	96.37
2004	98.63

According to forecasting decision-makers' experiences, we choose the threshold value of similar years selection as 96%. Therefore, the five history years (1998 and 2001–2004) with highest similarity are selected as the similar years of the forecast year 2005.

The power consumption of the province in 1998 and 2001–2005 is shown in Table 3.

Table 3. The power consumption of the Chinese province in 1998 and 2001–2005 (10⁸ kWh) [16,52].

Year	1998	2001	2002	2003	2004	2005
Power Consumption	437.85	557.58	628.82	725.20	833.01	946.33

The forecast values by eleven forecast models are shown in Table 4.

By the validity degree calculation method based on Markov chain and cloud model, the comprehensive validity degrees of the eleven forecast models are obtained as shown in Table 5.

The forecast model filtering threshold of the eleven forecast models is $\overline{cvd} = 81.46\%$. Therefore, the forecast models FM_4 , FM_5 , FM_8 , FM_9 , FM_{11} are selected for the combination forecast and the others are eliminated. Respectively, we use PSO, IA-PSO and improved IA-PSO for the forecast model weight determination. There are five selected forecast models SFM_1 (FM_4 : Power function model), SFM_2 (FM_5 : S-curve model), SFM_3 (FM_8 : Cubic curve model), SFM_4 (FM_9 : Artificial neural network method) and SFM_5 (FM_{11} : Grey system method). The parameters are set as follows: $n_{pa} = 100$, $n_{sfm} = 5$, $v_{\max} = 1$, $iter_{\max} = 1000$, $w = 0.6$, $LF_1 = LF_2 = 2$, $n_{apa} = 30$, $r = 25$, $\beta_{\min} = 0.001$, $\eta_1 = \eta_2 = 4$. The results of forecast model weights determination (FMWD) by PSO, IA-PSO and improved IA-PSO, which are abbreviated as FMWD-PSO, FMWD-IA-PSO, FMWD-improved-IA-PSO, are shown in Table 6.

Table 4. The forecast power consumption by the eleven forecast models (10^8 kWh).

Forecast Model	1998	2001	2002	2003	2004	2005
FM ₁ : Exponential model ($y = 780.65e^{-0.82/x}$)	343.91	636.02	662.58	680.93	694.36	704.55
FM ₂ : Logarithm model ($y = 362.13 + 188.39\ln x$)	362.09	623.33	665.32	699.65	728.73	753.90
FM ₃ : Hyperbola model ($y = 722.84 - 354.37/x$)	368.54	634.24	652.01	663.76	672.21	678.53
FM ₄ : Para-curve model ($y = 431.79 - 3.58x + 8.17x^2$)	436.88	556.81	631.53	723.67	833.29	960.30
FM ₅ : Grey system method [53]	437.94	567.04	642.01	727.00	823.23	932.08
FM ₆ : COMPERTZ model ($\ln y = 6.46 - 1.29e^{-x}$)	400.04	626.88	636.27	639.75	641.14	641.64
FM ₇ : Power function model ($y = 358.90x^{0.385}$)	358.90	612.43	667.48	716.12	759.91	800.02
FM ₈ : Cubic curve model ($y = 432.10 - 3.94x + 8.81x^2 - 0.0087x^3$)	437.04	556.81	631.56	723.84	833.17	960.04
FM ₉ : Artificial neural network method [36]	406.81	583.02	658.81	725.38	775.44	809.03
FM ₁₀ : S-curve model ($y^{-1} = 0.0015 + 0.0039e^{-x}$)	345.60	649.47	669.01	676.35	679.22	680.30
FM ₁₁ : Exponential smoothing method [54]	437.85	544.91	615.01	708.33	816.72	892.51

Table 5. The comprehensive validity degree of the eleven forecast models based on the real and forecast power consumptions in 1998 and 2001–2005.

Forecast Model	Comprehensive Validity Degree (%)
FM ₁	81.09
FM ₂	77.90
FM ₃	75.56
FM ₄	84.32
FM ₅	85.80
FM ₆	79.03
FM ₇	74.87
FM ₈	88.91
FM ₉	87.23
FM ₁₀	78.45
FM ₁₁	82.91

Table 6. The results of FMWD-PSO, FMWD-IA-PSO and FMWD-improved-IA-PSO.

Algorithm	Iteration Number	The Forecast Model Weight				
		SFM ₁	SFM ₂	SFM ₃	SFM ₄	SFM ₅
FMWD-PSO	623	0.0611	0.2120	0.3876	0.0861	0.2532
FMWD-IA-PSO	490	0.2598	0.1662	0.3343	0.0343	0.2054
FMWD-improved-IA-PSO	193	0.4105	0.0401	0.4895	0.0002	0.0597

Using the forecast model weights shown in Table 6, the results of weighted combination forecast (WCF) based on FMWD-PSO, FMWD-IA-PSO and FMWD-improved-IA-PSO, which are abbreviated as WCF-FMWD-PSO, WCF-FMWD-IA-PSO and WCF-FMWD-improved-IA-PSO, are shown in Table 7.

Table 7. The results of weighted combination forecast methods (10^8 kWh).

Weighted Combination Forecast	1998	2001	2002	2003	2004	2005
WCF-FMWD-PSO	434.82	558.22	631.93	720.71	821.93	923.03
WCF-FMWD-IA-PSO	436.28	556.97	630.82	721.19	826.19	936.41
WCF-FMWD-improved-IA-PSO	437.05	556.52	630.98	722.97	831.81	954.96

Four synthesized forecast methods reported by Kang *et al.* [55] are as follows:

(1) Equal weight method: the weights of forecast models are equal, so the combination forecast value of the forecast year c is as follows:

$$y_c = \frac{1}{n_{fm}} \sum_{d=1}^{n_{fm}} y_{c,d} \quad (38)$$

where $c = n_{his} + 1, n_{his} + 2, \dots, n_{his} + n_{fo}, d = 1, 2, \dots, n_{fm}$. It is a simple combination forecast method, and both the precision of single forecast model and the relationship between different forecast models are considered.

(2) Variance analysis method: the combination forecast value of the forecast year c is as follows:

$$y_c = \sum_{d=1}^{n_{fm}} \omega_d \cdot y_{c,d} \quad (39)$$

where $c = n_{his} + 1, n_{his} + 2, \dots, n_{his} + n_{fo}, d = 1, 2, \dots, n_{fm}$. All forecast models are independent of each other, so the variance of combination forecast can be expressed as follows:

$$Var = \sum_{d=1}^{n_{fm}} \omega_d^2 \cdot \delta_{dd} \quad (40)$$

where $c = n_{his} + 1, n_{his} + 2, \dots, n_{his} + n_{fo}, d = 1, 2, \dots, n_{fm}$, δ_{dd} is the variance of the forecast model FM_d . To obtain the minimum value of Varon ω_d , the Lagrange multiplier method is used and the weight of FM_d is defined as follows:

$$\omega_d = \frac{1}{\delta_{dd} \left(\frac{1}{\delta_{11}} + \frac{1}{\delta_{22}} + \dots + \frac{1}{\delta_{n_{fm}n_{fm}}} \right)} \quad (41)$$

(3) Optimum fitting method

The forecast model weight determination of the optimum fitting method is based on the deviations of all single forecast models and the complementarily between different forecast models. The deviations of the forecast model FM_d can be expressed as follows:

$$Dev_d = \frac{1}{2} \left(\frac{1}{n_{shis}} \left| \sum_{e=1}^{n_{shis}} (y' - y'_{e,d}) \right| + \frac{1}{n_{shis}} \sum_{e=1}^{n_{shis}} |y' - y'_{e,d}| \right) \quad (42)$$

Therefore, the weight of FM_d is defined as follows:

$$\omega_d = \frac{\max_{1 \leq d' \leq n_{fm}} Dev_{d'} + \min_{1 \leq d' \leq n_{fm}} Dev_{d'} - Dev_d}{\sum_{d=1}^{n_{fm}} \left(\max_{1 \leq d' \leq n_{fm}} Dev_{d'} + \min_{1 \leq d' \leq n_{fm}} Dev_{d'} - Dev_d \right)} \quad (43)$$

(4) Optimum forecast method

In this method, n_{fm} forecast models are used to carry out the forecasting respectively, and then these models are compared according to standard deviations, fitting goodness, correlation degree or relative error *et al.* Lastly the best one is chosen as the final forecast model.

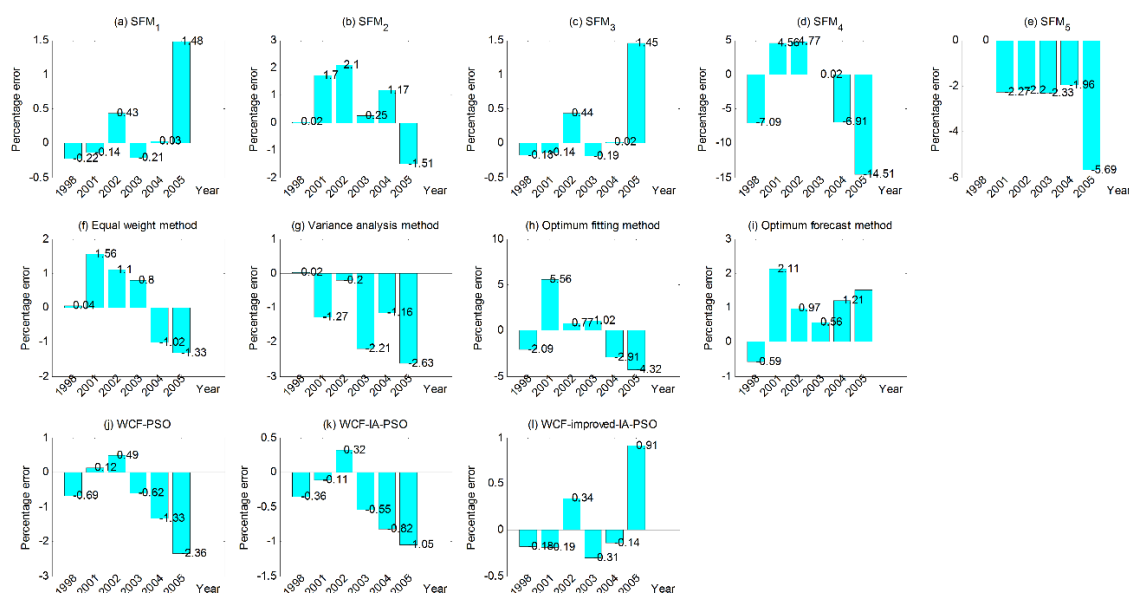
Percentage error (PE) and mean absolute percentage error (MAPE) are used as evaluating indicators to compare the proposed method (WCF-FMWD-improved-IA-PSO) to WCF-FMWD-PSO, WCF-FMWD-IA-PSO, the single forecast models (SFM_1 – SFM_5) and the four synthesized forecast methods [55] (equal weight method, variance analysis method, optimum fitting method and optimum forecast method). They are shown in Table 8.

Table 8. The percentage error of the proposed method (WCF-FMWD-improved-IA-PSO) and others.

Forecast Method	Mean Absolute Percentage Error (%)	Percentage Error (%)					
		1998	2001	2002	2003	2004	2005
SFM_1	0.42	−0.22	−0.14	0.43	−0.21	0.03	1.48
SFM_2	1.12	0.02	1.70	2.10	0.25	1.17	−1.51
SFM_3	0.40	−0.18	−0.14	0.44	−0.19	0.02	1.45
SFM_4	6.31	−7.09	4.56	4.77	0.02	−6.91	−14.51
SFM_5	2.41	0	−2.27	−2.20	−2.33	−1.96	−5.69
Equal weight method [55]	0.98	0.04	1.56	1.10	0.80	−1.02	−1.33
Variance analysis method [55]	1.25	0.02	−1.27	−0.20	−2.21	−1.16	−2.63
Optimum fitting method [55]	2.78	−2.09	5.56	0.77	1.02	−2.91	−4.32
Optimum forecast method [55]	1.16	−0.59	2.11	0.97	0.56	1.21	1.50
WCF-FMWD-PSO	0.93	−0.69	0.12	0.49	−0.62	−1.33	−2.36
WCF-FMWD-IA-PSO	0.53	−0.36	−0.11	0.32	−0.55	−0.82	−1.05
WCF-FMWD-improved-IA-PSO	0.35	−0.18	−0.19	0.34	−0.31	−0.14	0.91

From the comparison of PE and MAPE of the proposed method (WCF-FMWD-improved-IA-PSO) and others shown in Figures 4 and 5 and Table 8, we can see the follows:

- The maximum and minimum PE of the single forecast models (SFM_1 – SFM_5) are −14.51% and 0 respectively, the maximum and minimum MAPE of the single forecast models (SFM_1 – SFM_5) are 6.31% and 0.40% respectively.
- The maximum and minimum PE of the four synthesized forecast methods in Reference [55] are 5.56% and 0.02% respectively, the maximum and minimum MAPE of the four synthesized forecast methods in Reference [55] are 2.78% and 0.98% respectively.
- The maximum and minimum PE of WCF-FMWD-PSO are −2.36% and 0.12% respectively, the MAPE of WCF-FMWD-PSO is 0.93% and the iteration number is 623.
- The maximum and minimum PE of WCF-FMWD-IA-PSO are 1.05% and −0.11% respectively, the MAPE of WCF-FMWD-IA-PSO is 0.53% and the iteration number is 490.
- The maximum and minimum PE of WCF-FMWD-improved-IA-PSO are 0.91% and −0.14% respectively, the MAPE of WCF-FMWD-improved-IA-PSO is 0.35% and the iteration number is 193.

**Figure 4.** The comparison of PE of the proposed method (WCF-FMWD-improved-IA-PSO) and others.

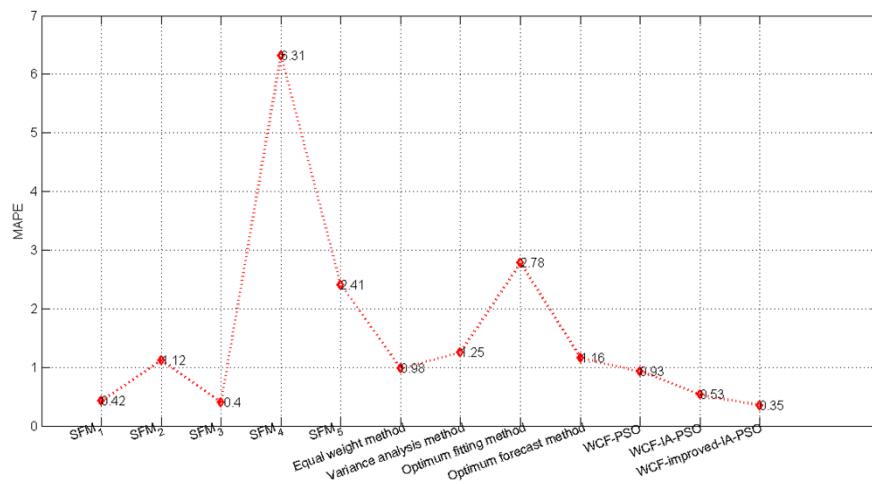


Figure 5. The comparison of MAPE of the proposed method (WCF-FMWD-improved-IA-PSO) and others.

Through the above analysis and the comparison shown in Figures 4 and 5 the proposed method WCF-FMWD-improved-IA-PSO results in an improved accuracy overall (MAPE is 0.35, maximum and minimum PE are 0.91% and -0.14%) when compared against other described methods over the chosen time period and year characteristics. In addition, the proposed method WCF-FMWD-improved-IA-PSO has the fastest convergence rate in forecast model weight determination when compared against WCF-FMWD-PSO and WCF-FMWD-IA-PSO. Therefore, using the proposed method to forecast the medium-and-long-term load is better than using other methods. The correctness and feasibility of the proposed method are proven.

6. Conclusions

In this paper, we have proposed a robust weighted combination load forecasting method WCF-FMWD-improved-IA-PSO based on forecast model filtering and adaptive variable weight determination to forecast the annual electric load. The contribution and novelty are mainly as follows:

- (1) Due to the fact that the forecast year's true load is unknown, the comprehensive validity degree of forecast model is defined by the integration of fitted value relative error and forecast value relative error, and then forecast models are filtered based on their comprehensive validity degrees.
 - The definition of validity degree can effectively overcome the inherent shortcomings of error theory. Entirely investigating the fitting level and the validity of forecast model, the comprehensive validity degree definition and the forecast model filtering method can improve the robustness of combination forecasting.
 - Revealing the transition pattern between the natural precision and validity degree, the forecast precision estimation method based on Markov chain and cloud model can provide an important basis for the subsequent weighted combination forecasting. In the forecast models' filtering, the better ones will be selected and the worse ones will be eliminated. It can also improve the robustness of combination forecasting.
- (2) The improved IA-PSO is used to determine the forecast model weight in combination forecasting. Based on the uniting of immune system's specific information processing mechanism and PSO's global convergence ability, disturbance variable and particle searching speed's adaptive adjustable strategy are introduced to improve the algorithm performance. The particles' diversity is ensured while the convergence speed is increased. It can avoid the local optimal and improve the accuracy.

As can be seen from the case study, the maximum and minimum of percentage error by the proposed method WCF-FMWD-improved-IA-PSO are 0.91%, -0.14% and the mean absolute percentage error is 0.35%, which are smaller than those by the single forecast models (SFM_1 – SFM_5), the four synthesized forecast methods [55], WCF-FMWD-PSO and WCF-FMWD-IA-PSO. These indicate that the proposed method has significant superiority over other methods in the terms of annual electric load forecasting accuracy. The iteration number of FMWD-improved-IA-PSO in the proposed method (193) is far less than the iteration numbers of FMWD-PSO and FMWD-IA-PSO (623 and 490), so its advantage that the global optimal solution is reached faster than PSO and IA-PSO is confirmed. In conclusion, the proposed method WCF-FMWD-improved-IA-PSO has a higher robustness and better accuracy, and it can meet the requirements of the annual electric load forecast and can also be applied in the forecast of related fields. In the forecast with analogous features, the proposed method WCF-FMWD-improved-IA-PSO can also be applied.

Acknowledgments: This work is supported by the National Natural Science Foundation of China (Nos. 61162005 and 61163002), External-Planned Task of State Key Laboratory of Robotics and System, Harbin Institute of Technology (HIT) (No. SKLRS-2013-MS-05), Natural Science Foundation of Ningxia (No. NZ15103), Key Scientific Research Project of Beifang University of Nationalities (No. 2015KJ08), Science & Technology Research Project of Ningxia High School (No. NGY2015149), and Introduction Talent Starting Scientific Research Project of Beifang University of Nationalities for Lianhui Li and Shaohu Ding.

Author Contributions: Lianhui Li and Chunyang Mu are the principal investigators of this work. They proposed the robust weighted combination load forecasting method and wrote the manuscript. Shaohu Ding designed the case study solution and checked the whole manuscript. Zheng Wang processed the data in the case study. Runyang Mo and Yongfeng Song did the data analysis work.

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

Nomenclature

n	number
CH	year characteristic
CHQ	year characteristic quantity
HY	history year
FY	forecast year
CSI	Cosine similarity
FM	forecast model
y'	similar year load
y	forecast year load
y'''	forecast year load by a forecast model
RE'	fitted value relative error
RE	forecast value relative error
P'	fitted precision
P	forecast precision
FIV	fitted validity degree
FOV	forecast validity degree
S	sub-interval
OC	occurrence number
TRN	transition number
TP	transition probability
TM	state transition matrix
IV	initial vector
SM	state matrix
CVE	column vector
Ex	expectation
En	entropy
He	hyper-entropy
En'	normal random number
CV	comprehensive validity degree
SFM	selected forecast model

<i>PO</i>	particle swarm
<i>PA</i>	particle
<i>x</i>	position
<i>v</i>	speed
<i>LF</i>	learning factor
<i>rand</i>	random number
<i>w</i>	inertia coefficient
<i>PRO</i>	probability
<i>AF</i>	affinity
<i>CON</i>	concentration
<i>iter</i>	iteration
<i>F</i>	fitness function
<i>PF</i>	particle fitness
<i>r</i>	vaccination times
<i>Var</i>	variance
<i>Dev</i>	deviation
Greek letters	
σ	standard deviation
α	empirical coefficient
χ	weight
β	disturbance variable
Δ	swarm convergence degree
η	coefficient used to control the upper limit
ω	forecast model's weight
Superscripts	
g	the g^{th} sub-interval
h	the h^{th} sub-interval
(h, g)	the transition from the h^{th} to the g^{th}
(k)	the k^{th} iteration
(q)	step
q	the q^{th} power
Subscripts	
<i>his</i>	history year
<i>fo</i>	forecast year
<i>shis</i>	selected history year
<i>fm</i>	forecast model
<i>sfm</i>	selected forecast model
<i>si</i>	sub-interval
<i>pa</i>	particle
<i>apa</i>	added particle
<i>avg</i>	average
<i>AVG</i>	the average of the numbers which are bigger than the global average
<i>a</i>	the a^{th} year characteristic
<i>b</i>	the b^{th} year characteristic quantity
<i>c</i>	the c^{th} forecast year
<i>d</i>	the d^{th} forecast model
<i>e</i>	the e^{th} similar year
<i>i</i>	the i^{th} column vector
<i>j</i>	the j^{th} selected forecast model
<i>l</i>	the l^{th} particle
<i>m</i>	the dimensional number of solution space
<i>t</i>	the t^{th} element

References

1. Kouhi, S.; Keynia, F. A new cascade NN based method to short-term load forecast in deregulated electricity market. *Energy Convers. Manag.* **2013**, *71*, 76–83. [CrossRef]

2. Chaturvedi, D.K.; Sinha, A.P.; Malik, O.P. Short term load forecast using fuzzy logic and wavelet transform integrated generalized neural network. *Int. J. Electr. Power Energy Syst.* **2015**, *67*, 230–237. [CrossRef]
3. Yang, Z.C. Electric load evaluation and forecast based on the elliptic orbit algorithmic model. *Int. J. Electr. Power Energy Syst.* **2012**, *42*, 560–567.
4. Bennett, C.; Stewart, R.A.; Lu, J. Autoregressive with exogenous variables and neural network short-term load forecast models for residential low voltage distribution networks. *Energies* **2014**, *7*, 2938–2960. [CrossRef]
5. Li, F.; Buxiang, Z.; Shi, C. The Medium and Long Term Load Forecasting Combined Model Considering Weight Scale Method. *TELKOMNIKA Indones. J. Electr. Eng.* **2013**, *11*, 2181–2186. [CrossRef]
6. Li, R.; Su, H.; Wang, Z. Medium-and long-term load forecasting based on heuristic least square support vector machine. *Power Syst. Technol.* **2011**, *35*, 195–199. (In Chinese)
7. Mello, P.E.; Lu, N.; Makarov, Y. An optimized autoregressive forecast error generator for wind and load uncertainty study. *Wind Energy* **2011**, *14*, 967–976. [CrossRef]
8. Li, H.; Guo, S.; Zhao, H. Annual electric load forecasting by a least squares support vector machine with a fruit fly optimization algorithm. *Energies* **2012**, *5*, 4430–4445. [CrossRef]
9. Filik, Ü.B.; Gerek, Ö.N.; Kurban, M. A novel modeling approach for hourly forecasting of long-term electric energy demand. *Energy Convers. Manag.* **2011**, *52*, 199–211. [CrossRef]
10. Suganthi, L.; Samuel, A.A. Energy models for demand forecasting—A review. *Renew. Sustain. Energy Rev.* **2012**, *16*, 1223–1240. [CrossRef]
11. Sheikh, S.K.; Unde, M.G. Short Term Load Forecasting using ANN Technique. *Int. J. Eng. Sci. Emerg. Technol.* **2012**, *1*, 97–107.
12. Wu, Y. The medium and long-term load forecasting based on improved D-S evidential theory. *Trans. China Electrotech. Soc.* **2012**, *27*, 157–162. (In Chinese)
13. Long, R.; Mao, Y.; Mao, L. A combination model for medium-and long-term load forecasting based on induced ordered weighted averaging operator and Markov chain. *Power Syst. Technol.* **2010**, *3*, 150–156. (In Chinese)
14. Chen, H.Y. *The Validity Theory of Combination Forecasting Method and Its Application*; Science Press: Beijing, China, 2008; pp. 76–109. (In Chinese)
15. Chen, H.Y. Research on combination forecasting model based on effective measure of forecasting methods. *Forecasting* **2001**, *20*, 72–73.
16. Sun, G.Q.; Yao, J.G.; Xie, Y.X. Combination forecast of medium-and-long-term load using fuzzy adaptive variable weight based on fresh degree function and forecasting availability. *Power Syst. Technol.* **2009**, *33*, 103–107. (In Chinese).
17. Chen, C.; Guo, W.; Fan, J.Z. Combined method of mid-long term load forecast based on improved forecasting effectiveness. *Relay* **2007**, *35*, 70–74. (In Chinese)
18. Jin, X.; Luo, D.S.; Sun, G.Q. Sifting and combination method of medium-and-long-term load forecasting model. *Proc. Chin. Soc. Univ. Electr. Power Syst. Autom.* **2012**, *24*, 150–156. (In Chinese)
19. Hernandez, L.; Baladrón, C.; Aguiar, J.M. Short-term load forecasting for microgrids based on artificial neural networks. *Energies* **2013**, *6*, 1385–1408. [CrossRef]
20. Gofman, A.V.; Vedernikov, A.S.; Vedernikova, E.S. Increasing the accuracy of the short-term and operational prediction of the load of a power system using an artificial neural network. *Power Technol. Eng.* **2013**, *46*, 410–415. [CrossRef]
21. Li, H.; Guo, S.; Li, C. A hybrid annual power load forecasting model based on generalized regression neural network with fruit fly optimization algorithm. *Knowledge-Based Syst.* **2013**, *37*, 378–387. [CrossRef]
22. Paparoditis, E.; Sapatinas, T. Short-term load forecasting: The similar shape functional time-series predictor. *IEEE Trans. Power Syst.* **2013**, *28*, 3818–3825. [CrossRef]
23. Weron, R.; Taylor, J. Discussion on “Electrical load forecasting by exponential smoothing with covariates”. *Appl. Stoch. Models Bus. Ind.* **2013**, *29*, 648–651. [CrossRef]
24. Li, D.C.; Chang, C.J.; Chen, C.C. Forecasting short-term electricity consumption using the adaptive grey-based approach—An Asian case. *Omega* **2012**, *40*, 767–773. [CrossRef]
25. Ismail, N.A.; King, M. Factors influencing the alignment of accounting information systems in small and medium sized Malaysian manufacturing firms. *J. Inf. Syst. Small Bus.* **2014**, *1*, 1–20.
26. Kodogiannis, V.S.; Amina, M.; Petrounias, I. A clustering-based fuzzy wavelet neural network model for short-term load forecasting. *Int. J. Neural Syst.* **2013**, *23*, 1557–1565. [CrossRef] [PubMed]

27. Hong, T.; Wang, P. Fuzzy interaction regression for short term load forecasting. *Fuzzy Optim. Decis. Mak.* **2014**, *13*, 91–103. [CrossRef]
28. Borges, C.E.; Penya, Y.K.; Fernandez, I. Evaluating combined load forecasting in large power systems and smart grids. *IEEE Trans. Ind. Inf.* **2013**, *9*, 1570–1577. [CrossRef]
29. Che, J.; Wang, J.; Wang, G. An adaptive fuzzy combination model based on self-organizing map and support vector regression for electric load forecasting. *Energy* **2012**, *37*, 657–664. [CrossRef]
30. Liu, Z.; Li, W.; Sun, W. A novel method of short-term load forecasting based on multiwavelet transform and multiple neural networks. *Neural Comput. Appl.* **2013**, *22*, 271–277. [CrossRef]
31. Wang, J.; Wang, J.; Li, Y. Techniques of applying wavelet de-noising into a combined model for short-term load forecasting. *Int. J. Electr. Power Energy Syst.* **2014**, *62*, 816–824. [CrossRef]
32. Enayatifar, R.; Sadaei, H.J.; Abdullah, A.H. Imperialist competitive algorithm combined with refined high-order weighted fuzzy time series (RHWFTS-ICA) for short term load forecasting. *Energy Convers. Manag.* **2013**, *76*, 1104–1116. [CrossRef]
33. Ko, C.N.; Lee, C.M. Short-term load forecasting using SVR (support vector regression)-based radial basis function neural network with dual extended Kalman filter. *Energy* **2013**, *49*, 413–422. [CrossRef]
34. Ma, S.; Chen, X.; Liao, Y.; Wang, G.; Ding, X.; Chen, K. The variable weight combination load forecasting based on grey model and semi-parametric regression model. In Proceedings of the TENCON 2013—2013 IEEE Region 10 Conference (31194), Xi'an, China, 22–25 October 2013; pp. 1–4.
35. Niu, D.; Shi, H.; Wu, D.D. Short-term load forecasting using bayesian neural networks learned by Hybrid Monte Carlo algorithm. *Appl. Soft Comput.* **2012**, *12*, 1822–1827. [CrossRef]
36. Zhu, J.P.; Dai, J. Optimization selection of correlative factors for long-term load prediction of electric power. *Comput. Simul.* **2008**, *5*, 226–229.
37. Wang, Q.; Wang, Y.L.; Zhang, L.Z. An approach to allocate impersonal weights of factors influencing electric power demand forecasting. *Power Syst. Technol.* **2008**, *32*, 82–86. (In Chinese)
38. Lei, S.L.; Gu, L.; Yang, J.; Liu, X.Y. Analysis of electric power load characteristics and its influencing factors in chongqing region. *Electr. Power* **2014**, *12*, 61–71. (In Chinese)
39. Liao, F.; Congying, X.U.; Yao, J.; Cai, J.; Chen, S. Load characteristics of change region and analysis on its influencing factors. *Power Syst. Technol.* **2012**, *7*, 117–125.
40. Karakoc, E.; Cherkasov, A.; Sahinalp, S.C. Distance based algorithms for small biomolecule classification and structural similarity search. *Bioinformatics* **2006**, *14*, 243–251. [CrossRef] [PubMed]
41. Ye, J. Cosine similarity measures for intuitionistic fuzzy sets and their applications. *Math. Comput. Model.* **2011**, *1*, 91–97. [CrossRef]
42. Zhu, S.; Wu, J.; Xiong, H.; Xia, G. Scaling up top-K cosine similarity search. *Data Knowl. Eng.* **2011**, *1*, 60–83. [CrossRef]
43. Keilson, J. *Markov Chain Models—Rarity and Exponentiality*; Springer Science and Business Media: Berlin, Germany, 2012.
44. Yoder, M.; Hering, A.S.; Navidi, W.C.; Larson, K. Short-term forecasting of categorical changes in wind power with Markov chain models. *Wind Energy* **2014**, *17*, 1425–1439. [CrossRef]
45. Li, D.Y.; Liu, C.Y. Study on the universality of the normal cloud model. *Eng. Sci.* **2004**, *6*, 28–34.
46. Wang, G.; Xu, C.; Li, D. Generic normal cloud model. *Inf. Sci.* **2014**, *280*, 1–15. [CrossRef]
47. Kennedy, J. Particle swarm optimization. In *Encyclopedia of Machine Learning*; Springer Publishing Company: New York, NY, USA, 2010; pp. 760–766.
48. Zhao, F.; Li, G.; Yang, C.; Abraham, A.; Liu, H. A human-computer cooperative particle swarm optimization based immune algorithm for layout design. *Neurocomputing* **2014**, *132*, 68–78. [CrossRef]
49. Fu, X.; Li, A.; Wang, L.; Ji, C. Short-term scheduling of cascade reservoirs using an immune algorithm-based particle swarm optimization. *Comput. Math. Appl.* **2011**, *6*, 2463–2471. [CrossRef]
50. Afshinmanesh, F.; Marandi, A.; Rahimi-Kian, A. A novel binary particle swarm optimization method using artificial immune system. In Proceedings of the International Conference on Computer as a Tool, EUROCON 2005, Belgrade, Serbia, 21–24 November 2005; Volume 1, pp. 217–220.
51. Wu, J.M.; Zuo, H.F.; Chen, Y. A combined forecasting method based on particle swarm optimization with immunity algorithms. *Syst. Eng. Theory Method Appl.* **2006**, *15*, 229–233. (In Chinese)
52. China National Bureau of Statistics. *China Energy Statistical Yearbook*; China Statistics Press: Beijing, China, 2011. (In Chinese)

53. Deng, J.L. Introduction to grey system theory. *J. Grey Syst.* **1989**, *1*, 1–24. (In Chinese)
54. Gardner, E.S. Exponential smoothing: The state of the art. *J. Forecast.* **1985**, *1*, 1–28. [CrossRef]
55. Kang, C.Q.; Xia, Q.; Liu, M. *Power System Load Forecasting*; China Electric Power Press: Beijing, China, 2007; pp. 73–75. (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 by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).