

RESEARCH ARTICLE

# Forecasting outpatient visits using empirical mode decomposition coupled with back-propagation artificial neural networks optimized by particle swarm optimization

Daizheng Huang\*, Zhihui Wu

Department of Biomedical Engineering, School of Preclinical Medicine, Guangxi Medical University, Nanning, Guangxi Province, China

\* [daizheng-huang@qq.com](mailto:daizheng-huang@qq.com)



## Abstract

Accurately predicting the trend of outpatient visits by mathematical modeling can help policy makers manage hospitals effectively, reasonably organize schedules for human resources and finances, and appropriately distribute hospital material resources. In this study, a hybrid method based on empirical mode decomposition and back-propagation artificial neural networks optimized by particle swarm optimization is developed to forecast outpatient visits on the basis of monthly numbers. The data outpatient visits are retrieved from January 2005 to December 2013 and first obtained as the original time series. Second, the original time series is decomposed into a finite and often small number of intrinsic mode functions by the empirical mode decomposition technique. Third, a three-layer back-propagation artificial neural network is constructed to forecast each intrinsic mode functions. To improve network performance and avoid falling into a local minimum, particle swarm optimization is employed to optimize the weights and thresholds of back-propagation artificial neural networks. Finally, the superposition of forecasting results of the intrinsic mode functions is regarded as the ultimate forecasting value. Simulation indicates that the proposed method attains a better performance index than the other four methods.

## OPEN ACCESS

**Citation:** Huang D, Wu Z (2017) Forecasting outpatient visits using empirical mode decomposition coupled with back-propagation artificial neural networks optimized by particle swarm optimization. PLoS ONE 12(2): e0172539. doi:10.1371/journal.pone.0172539

**Editor:** Yong Deng, Southwest University, CHINA

**Received:** September 6, 2016

**Accepted:** February 5, 2017

**Published:** February 21, 2017

**Copyright:** © 2017 Huang, Wu. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This work was supported by the project of Basic Ability Promotion for Young Teachers of the Guangxi Education Department (KY2016YB093).

**Competing interests:** The authors declare no competing interest.

## Introduction

Obtaining healthcare in China is currently challenging because the growth rate of healthcare agencies is far lower than the rise in patient needs. Accurately forecasting the number of outpatient visits will increase the efficiency of planning and the delivery of outpatient management. This ability can also help healthcare administrators oversee hospitals effectively, reasonably organize schedules for human resources and finances, and properly distribute hospital material resources. Therefore, forecasting the number of outpatient visits has become an important issue in public health and has motivated many researchers to establish mathematical models to realize such predictions, especially in China. Ching proposed a fuzzy time series method based on the weighted-transitional matrix, as well as the expectation method and grade-selection approach, to forecast the number of outpatient visits[1]. Meanwhile, Esmaeil Hadavandia

presented a hybrid artificial intelligence model for the development of a Mamdani-type fuzzy-rule-based system to forecast outpatient visits[2]. Decomposition and multi-local predictor fusion were proposed to predict outpatient consults for diarrhea[3].

The number of outpatient visits is a nonlinear and nonstationary time series. The forecast of this information may not be perfectly accurate if the linear method is used in such time series. Meanwhile, empirical mode decomposition(EMD) is an empirical, intuitive, direct, and self-adaptive time–frequency data analysis method introduced by Huang et al.[4]. This approach is fairly versatile in a broad range of applications on signal extraction from data generated in noisy nonlinear and nonstationary processes. Compared with Fourier transformation and wavelet transformation, it has many advantages such as good multi-resolution and wide applicability [5]. EMD technique has been successfully applied in various areas, such as signal processing[6], digital holography[7], image processing[8], detection techniques[9], and forecasting approach [10, 11]. The original morbidity data of outpatient visits can be decomposed into several sub-series by EMD. Compared with the original series, the sub-series exhibits a more apparent regularity and can be forecasted to achieve easy prediction tasks and fine results.

Though the original morbidity data of outpatient visits has been decomposed into several sub-series, the sub-series are still time series. There were many applications using the traditional statistical models like autoregressive model, moving average model and autoregressive moving average models for predicting time series in past studies. These models perform well when the data lie within the range of past observations. But they perform poorly to predict extremes and also when the data are lying just near to limits[12]. The application of artificial intelligence prediction models has been offered in recent years. One of the most widely used artificial intelligence methods in prediction is artificial neural networks (ANNs) [13]. ANNs are complex and flexible nonlinear systems inspired by biological neural networks (e.g., animal central nervous systems, particularly, the brain) and are used to estimate or approximate functions, which can depend on numerous inputs and are generally unknown[14,15]. ANN processes information by adjusting the internal relations between large connected nodes and it has strong self-learning abilities and adaptive abilities [16].The ANNs time series models can capture the historical information by nonlinear functions [17]. A highly popular and widely used ANN algorithm is the back-propagation ANN (BPANN) [18], which involves connection weights and thresholds adjusted by the backward propagation of errors. One major application area of BPANN is in forecasting, owing to its characteristics of extreme computational power, massive parallelism, and fault tolerance [19].

Network performance is affected by the initial weight and threshold of BPANN, and these parameters participate in a close relationship with the network convergence, falling into the local minimum and training time. To overcome the problem, different optimization algorithms optimize the initial weight and threshold of BPANN, including genetic algorithm [20,21], ant colony algorithm[22], simulated annealing algorithm[23] and particle swarm optimization (PSO). PSO is a type of swarm intelligence optimization algorithm for global optimization and has proven to be a competitor to GA when it comes to optimization problems[24]. Compared with other biological evolution algorithms, PSO occupies the bigger optimization ability using simple relations [25]. It is widely used in optimization because of its need for less parameter sets and its faster convergence rate and easy escape from the local optimum compared with other algorithms[26–31]. At the same time, it can perform strong parallel search and global optimization. PSO was thus selected to optimize the BPANN weights and thresholds in the present study because of its simplicity and good performance in finding desirable solutions.

This study aims to assess the forecasting accuracy of BPANN models coupled with EMD for outpatient visits. EMD is applied to decompose the original data of outpatient visits, and then different BPANN models are constructed with each sub-series. The final forecasted value is obtained by summing the forecasted value of each sub-series.

## Methods

### Empirical mode decomposition

EMD is an adaptive and efficient method to analyze nonlinear and non-stationary signals. Without any a priori knowledge, the original time series was decomposed into a sum of oscillatory functions called intrinsic mode functions (IMFs) [4,32,33]. The IMFs must meet the following conditions:

1. The number of extrema (maximum and minimum) and the number of zero-crossings must be the same or differ at most by one in the whole data set.
2. At any point, the IMF is symmetric with respect to the local zero mean.

The IMF can be extracted from the original time series through a shifting process described as follows:

*Step 1.* All the local extrema [maximum  $e_{max}(t)$  and minimum  $e_{min}(t)$ ] points of the given time series  $x(t)$  are calculated.

*Step 2.* All the local maxima points are connected as the upper envelop, and all the minima points as the lower envelop, by a cubic spline line.

*Step 3.* The mean  $m(t)$  of the upper and lower envelops are calculated as

$$m(t) = (e_{max}(t) + e_{min}(t))/2 \tag{1}$$

*Step 4.* The mean from the original time series is obtained, and the difference is defined as  $h(t)$ :

$$h(t) = x(t) - m(t) \tag{2}$$

*Step 5.*  $h(t)$  is checked and judged whether it meets the two conditions for an IMF in accordance with the stopping criterion. If the criteria are satisfied, then  $h(t)$  is denoted as the first IMF and written as  $g_1(t) = h(t)$ . Moreover,  $x(t)$  is replaced with the residue  $r(t) = x(t) - h(t)$ . If the conditions are not satisfied,  $x(t)$  is replaced with  $h(t)$  and steps 1 to 4 are repeated until  $h(t)$  meets the two conditions for an IMF.

*Step 6.* Steps 1 to 5 are repeated. The shifting procedure is then terminated until the residue becomes a constant, a monotonic function, or a function with only one maximum and one minimum from which no more IMF can be extracted.  $g_i(t) + r_m(t)$

Then, the original time series  $x(t)$  is expressed as the sum of the IMFs and a residue:

$$x(t) = \sum_{i=1}^m g_i(t) + r_m(t) \tag{3}$$

where  $m$  is the number of IMFs and  $r_m(t)$  is the final residue.

### Back-propagation artificial neural networks

ANN approach can imitate any complex and non-linear relationship through non-linear units (neurons) and has been widely used in the forecasting area. BPANN is the most extensively used ANN model. [5,34,35]. The typical topology of BPANN involves three layers: input layer, where the data are introduced to the network; hidden layer, where the data are processed; and output layer, where the results of the given input are produced [15]. A number of interconnected neuron nodes are present in the layers. The output  $H_j$  of any neuron can be represented as

$$H_j = f\left(\sum_{i=1}^n \omega_{ij}x_i - b_j\right) \tag{4}$$

where  $n$  is the number of inputs,  $x_i$  is the  $i$ th input,  $\omega_{ij}$  is the connective weight between the  $j$ th and  $i$ th neurons,  $b_j$  is the threshold,  $f$  is a nonlinear activation function, and the sigmoid function  $f(t) = 1/(1+e^{-x})$  is widely employed as the activation function. Determining the optimum weights and thresholds is usually performed by trial during the training stage as the output of the network matches the desired pattern for a specific set of inputs.

The algorithm includes training and forecast, which are detailed in the following steps:

- Step 1.* Forward propagation of input signal. The input vector propagates to the output layer after being computed by the hidden layer.
- Step 2.* Backward propagation of error signal. The error propagates backward through the original neural network if the error value does not meet the given tolerance.
- Step 3.* Weight and threshold updates. The weights and thresholds are adjusted by the backward propagation of errors until the error value meets the given tolerance. Then, the fixed structure of the BPANN model is obtained.
- Step 4.* Forecast. The trained BPANN model is used for forecast.

### Particle swarm optimization

PSO is a type of swarm intelligence optimization algorithm for global optimization proposed by Eberhart and Kennedy in 1995 and inspired by the behavior of a swarm of birds [36]. The social behavior of birds is simulated in PSO. Each bird is represented by a particle, and a collection of particles is identified as a swarm. Each particle in the swarm represents a potential solution to the optimization problem. Particles are simultaneously updated by exchanging information with one another. The basic PSO algorithm is governed by Eqs (5) and (6):

$$V_i^{k+1} = \omega V_i^k + c_1 r_1 (P_i^k - X_i^k) + c_2 r_2 (P_g^k - X_i^k) \tag{5}$$

$$X_i^{k+1} = X_i^k + V_i^{k+1} \tag{6}$$

where  $V_i$  represents the velocity of the  $i$ th particle.  $X_i$  represents the position of the  $i$ th particle and denotes a potential solution to the problem.  $P_i$  is the best solution obtained by particle  $i$  until iteration  $k$ , and  $P_g$  is the best solution obtained by all particles until iteration  $k$ .  $\omega \in [0, 1]$  is the inertia weight.  $r_1$  and  $r_2$  are independent random numbers uniformly distributed in the range of  $[0, 1]$ , whereas  $c_1$  and  $c_2$  are positive constants that are both acceleration factors, named cognitive and social parameters respectively.  $k$  represents the current iteration of the optimization process.

The speed and precision of convergence are greatly affected by the initial parameters in PSO. Therefore, selecting the reasonable the initial parameters in the algorithm is important.

1. Selecting the inertia weight. Inertia weight is the most important parameter in the PSO, and the use of an appropriate inertia weight can help achieve the balance between global search and local optimization. The global search ability of the algorithm is enhanced when the inertia weight is large. By contrast, the local search performance becomes effective and enables the search for the global optimal solution when the inertia weight is relatively small. The linear decreasing inertia weight strategy was proposed by Y. Shi in 1999[37]. The ability of global and

local searches is balanced in this method, which accelerates the convergence rate of the algorithm. The linear decreasing inertia weight is calculated as follows:

$$\omega = \omega_{start} - \frac{\omega_{start} - \omega_{end}}{t_{max}} \times t \tag{7}$$

where  $\omega_{start}$  and  $\omega_{end}$  are the initial and termination inertia weights respectively. The typical inertia weight is adapted from 0.9 to 0.3.  $t_{max}$  is the maximum iteration, and  $t$  is the current iteration. The linear decreasing inertia weight was used in this work.

2. Selecting the learning factors. The learning factors reflect the exchange information between particles. These aspects are in the range of [1, 2.5] by subjective experience.

The adaptive time-varying strategy for adjusting the learning factor was proposed by A. Ratnawecra in 2004[38]. The adjustment formula is as follows:

$$c_1 = c_{1start} + \frac{c_{1end} - c_{1start}}{t_{max}} \times t \tag{8}$$

$$c_2 = c_{2start} + \frac{c_{2end} - c_{2start}}{t_{max}} \times t \tag{9}$$

The typical number of the learning factor is that the  $c_1$  reduces from 2.5 to 0.5, whereas the  $c_2$  increases from 0.5 to 2.5. Results show that the performance is enhanced by the adaptive time-varying strategy than by the fixed learning factor. Thus, the adaptive time-varying method was used in this study.

The thresholds and weights of BPANN are encoded as a particle. The dimension  $D$  of particle swarm is as follows:

$$D = n_p + n_h + n_i \times n_h + n_h \times n_o \tag{10}$$

where  $n_p$ ,  $n_h$ , and  $n_o$  are the input, hidden, and output nodes, respectively.

The mean square error (MSE) between the prediction and raw data is considered as the fitness function used to evaluate the quality of particles.

The velocity and position of these particles with higher fitness are updated until the best particles are produced.

The flow chart of the proposed method in the paper is shown in Fig 1.

The proposed algorithm can be separated into six steps. (1) Raw time series decomposition (Fig 1). The number of outpatient visits is released from the hospital information section by month and considered as the original time series. Then, EMD is used to decompose the raw time series into a finite and often small number of IMFs plus a residue. Each IMF can represent the local characteristic time scale by itself. (2) Data normalization. By directly calling the two functions premnmx and postmnmx in MATLAB software, each IMF is normalized by the Min-Max normalization algorithm. (3) Thresholds and weights of BPANN optimization. The weights and thresholds of BPANN are mapped to and encoded as the particles. The optimal particles are found through updates of their positions and velocities. (4) BPANN training. The optimal weights and thresholds are employed as the initial weights and thresholds and adopted into the BPANN. The training data are also considered into the network. The prediction network can be obtained through training. (5) Trained BPANN testing. Each prediction result of the IMFs can be obtained by using the trained BPANN. (6) Superposition of the IMF forecast results. Finally, the superposition of forecasting results of the IMFs is regarded as the ultimate forecasting value.

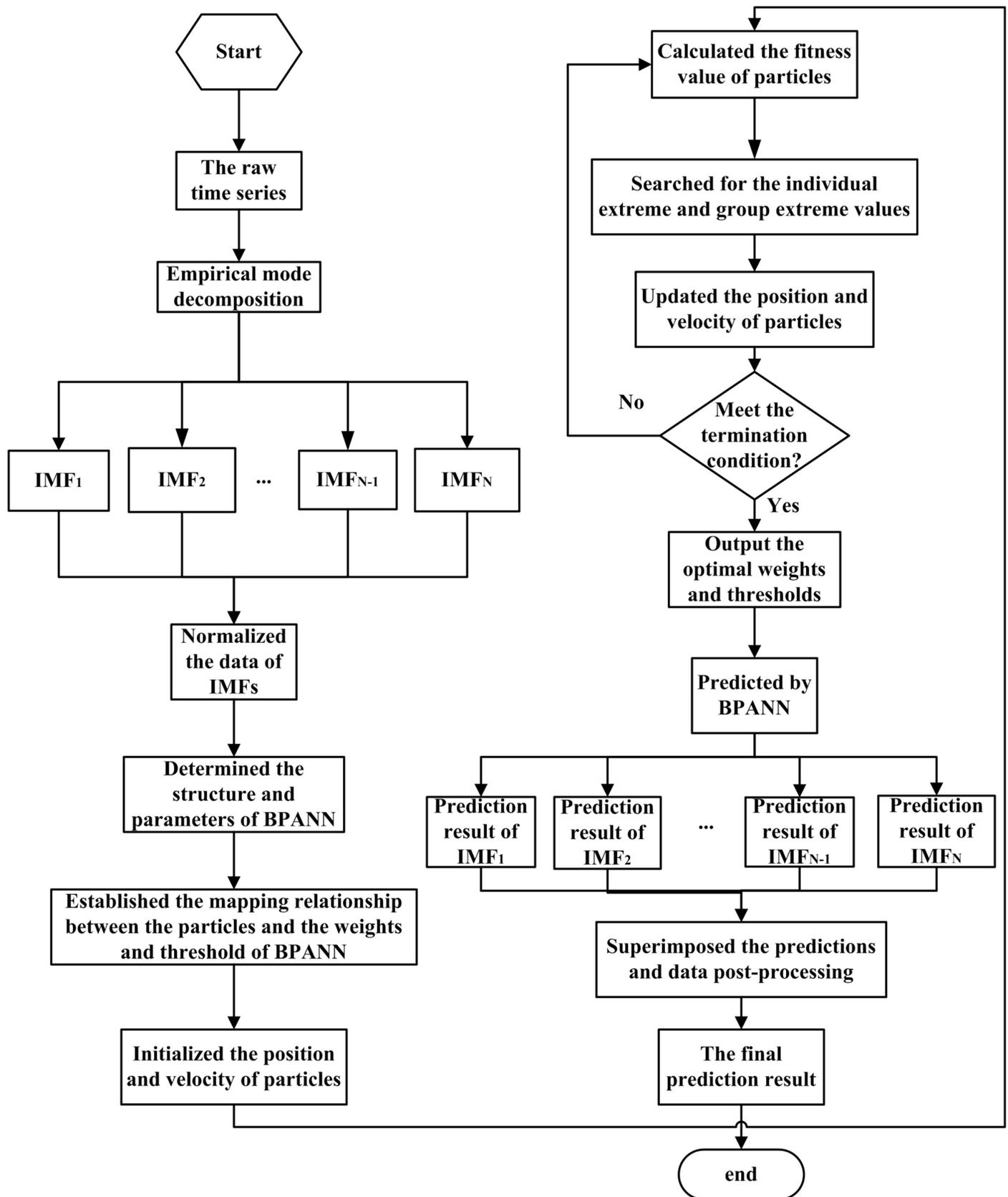


Fig 1. Flow chart of the proposed algorithm.

doi:10.1371/journal.pone.0172539.g001

## Results and discussion

### Data sources

A case will be simulated in this section. The number of outpatient visits from a hospitals in Nanning City, China was documented ([S1 Dataset](#)).

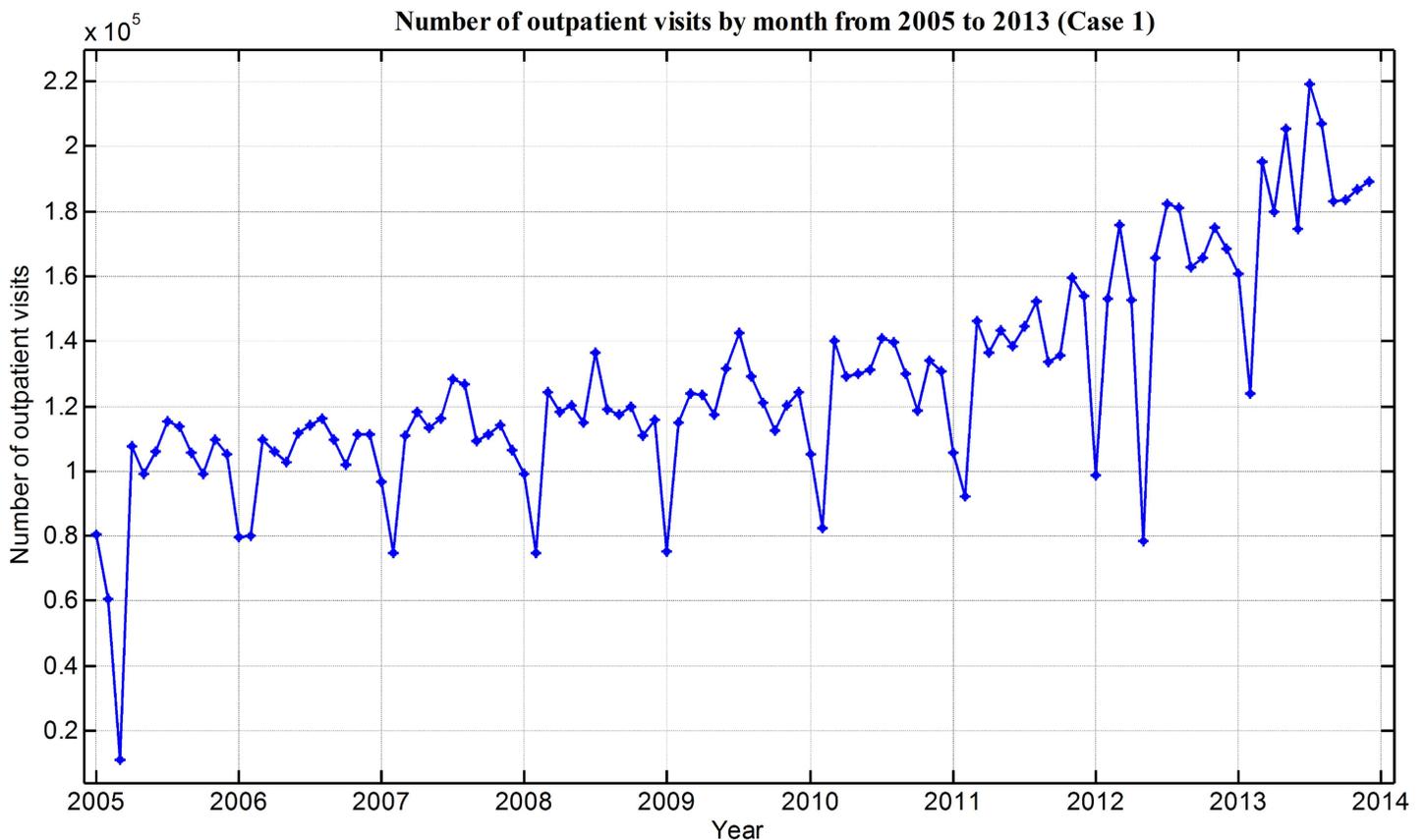
### Ethical review

The study protocol and utilization of outpatient visit data were reviewed from a hospitals in Nanning City, China, and no ethical issue was identified. Therefore, no ethical approval was required.

The number of outpatient visits from January 2005 to December 2013 are obtained as the original time series and shown in [Fig 2](#).

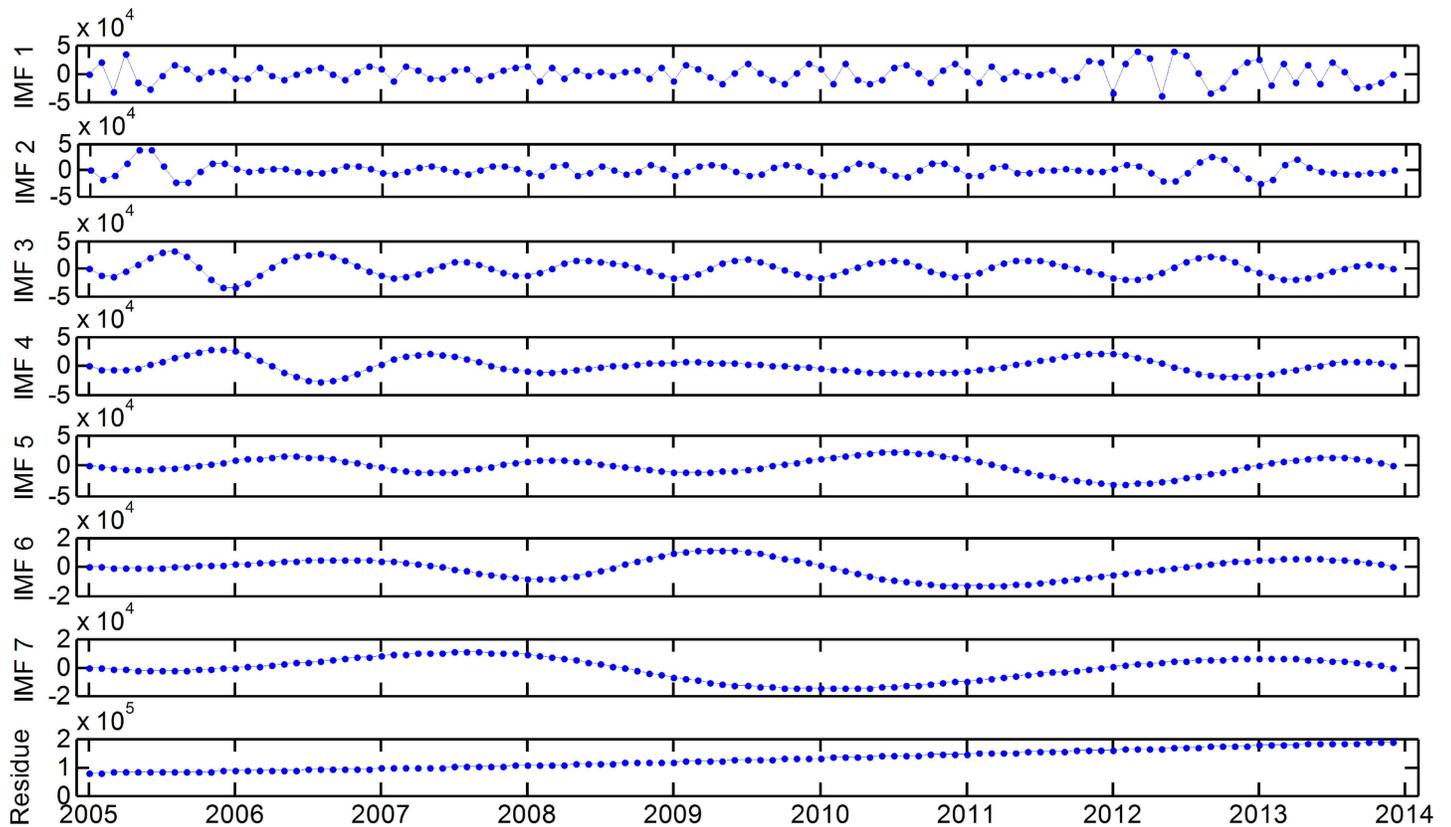
The decomposition results obtained through the EMD technique are shown in [Fig 3](#). Seven IMFs and a residue component are observed. The IMFs present the characteristics of outpatient visit fluctuations on different time scales from high frequency to low frequency. IMF1 involves an obvious periodic variability and exhibits the maximum amplitude, highest frequency, and shortest wavelength. The following IMF components decrease in amplitude and frequency and increase in wavelength. The last residue is a mode that slowly varies around the long-term average; this mode shows the overall trend of outpatient visits.

There is no mature reference method to choose appropriate network structure and neuron for building a BPANN. The optimal three input layer neurons were experimentally selected and have average square error less than 0.01. The output layer only contains one neuron



**Fig 2. Number of outpatient visits by month from 2005 to 2013.**

doi:10.1371/journal.pone.0172539.g002



**Fig 3. Decomposition of outpatient visits.**

doi:10.1371/journal.pone.0172539.g003

representing the forecast value. Some empirical formulas are used to choose hidden layer node number. Such as  $n_2 = \sqrt{n_1 + n_3} + m$ , where  $n_1$ ,  $n_2$ ,  $n_3$  and  $m$  represent the input node, hidden node, output node and a positive integer and  $m \in [1, 10]$ . After computing, the prediction is found to be most accurate when  $m = 10$  instead of other values. Therefore, a three-layer proposed model with 3 input nodes, 14 hidden nodes, and 1 output node (3–14–1) is obtained (Fig 4). The selection for parameters of BPNN is based on the literature [39]. And the parameters of BPANN are listed in Table 1.

In order to optimize the weights and thresholds of the BPANN mode, PSO is employed. Four important parameters of the PSO algorithm should be set: inertia weight, learning factors, the maximum number of iterations, and the dimensions of particle swarm. The linear decreasing inertia weight was used in this work just as previously introduced. As originally developed, the inertia weight often decreases linearly from about 0.9 to 0.4 during a run [40,41]. Here we set it from about 0.9 to 0.3. And the adaptive time-varying method was used in this study for the performance is enhanced by the adaptive time-varying strategy than by the fixed learning factor. The thresholds and weights of BPANN are encoded as a particle, 71 of dimensions will be attained based on the Eq (10). The number of iterative learning was 200. The parameters of PSO are listed in Table 2.

The raw time series obtained from January 2005 to November 2012 are adopted as training samples, and those obtained from December 2012 to December 2013 are employed as testing samples.

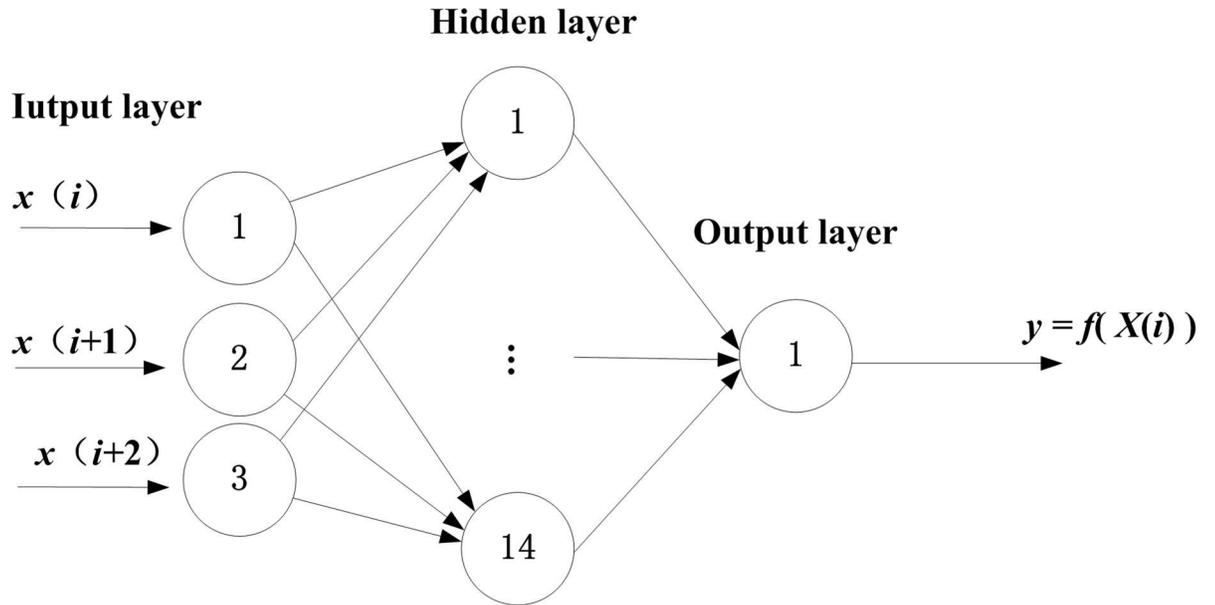


Fig 4. BPANN model.

doi:10.1371/journal.pone.0172539.g004

The weights and thresholds of BPANN are optimized by PSO by training each component. The variation of the best fitness with iteration is illustrated in Fig 5.

The result prediction and the raw data of each component are shown in Fig 6.

The ultimate forecasting value and the observed data are illustrated in Fig 7.

To compare the result prediction with other methods, a simulation is performed by PSO-BPANN directly, GA-BPANN directly, wavelet decomposition PSO-BPANN(WD-PSO-BPANN), wavelet decomposition GA-BPANN(WD-GA-ANN), and the proposed methods (abbreviated as EMD-PSO-BPANN). These methods hold the same simulation conditions.

Four main criteria are used for evaluation of level prediction and directional forecasting including the coefficient of correlation (R), root mean squared error (RMSE), mean average percentage error (MAPE), and sum of squared error (SSE). The definitions are shown as follows.

R criterion is given by

$$R = \frac{R_{xx}}{std(x)std(\hat{x})} \tag{11}$$

Table 1. Parameters of BPANN.

Parameter	Value	Parameter	Value	Parameter	Value
Input node	3	activation function of output layer	purelin	Net.trainParam.show	25
Hidden node	14	training function	trainrp	Net.trainParam.goal	0.01
Output node	1	learning function	learnsgdm	Net.trainParam.lr	0.15
Activation function of the hidden layer	tansig	Net.trainParam.epochs	1000	Net.trainParam.min_grad	1e-6

doi:10.1371/journal.pone.0172539.t001

Table 2. Parameters of PSO.

Parameter	Value	Parameter	Value	Parameter	Value
Population	40	minimum fitness	1e-30	iteration	200
Dimension	71	maximum velocity	1	minimum velocity	-1
Inertia weight	linear decrease	maximum weight	0.9	minimum weight	0.3
Learning factor	adaptive time-varying strategy	learning factor C <sub>1</sub>	0.5–2.5	learning factor C <sub>2</sub>	0.5–2.5

doi:10.1371/journal.pone.0172539.t002

RMSE criterion is defined as

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{x}_i - x_i)^2} \tag{12}$$

MAPE criterion is calculated through

$$MAPE(\%) = \frac{1}{n} \sum_{i=1}^n |\hat{x}_i - x_i| * 100 \tag{13}$$

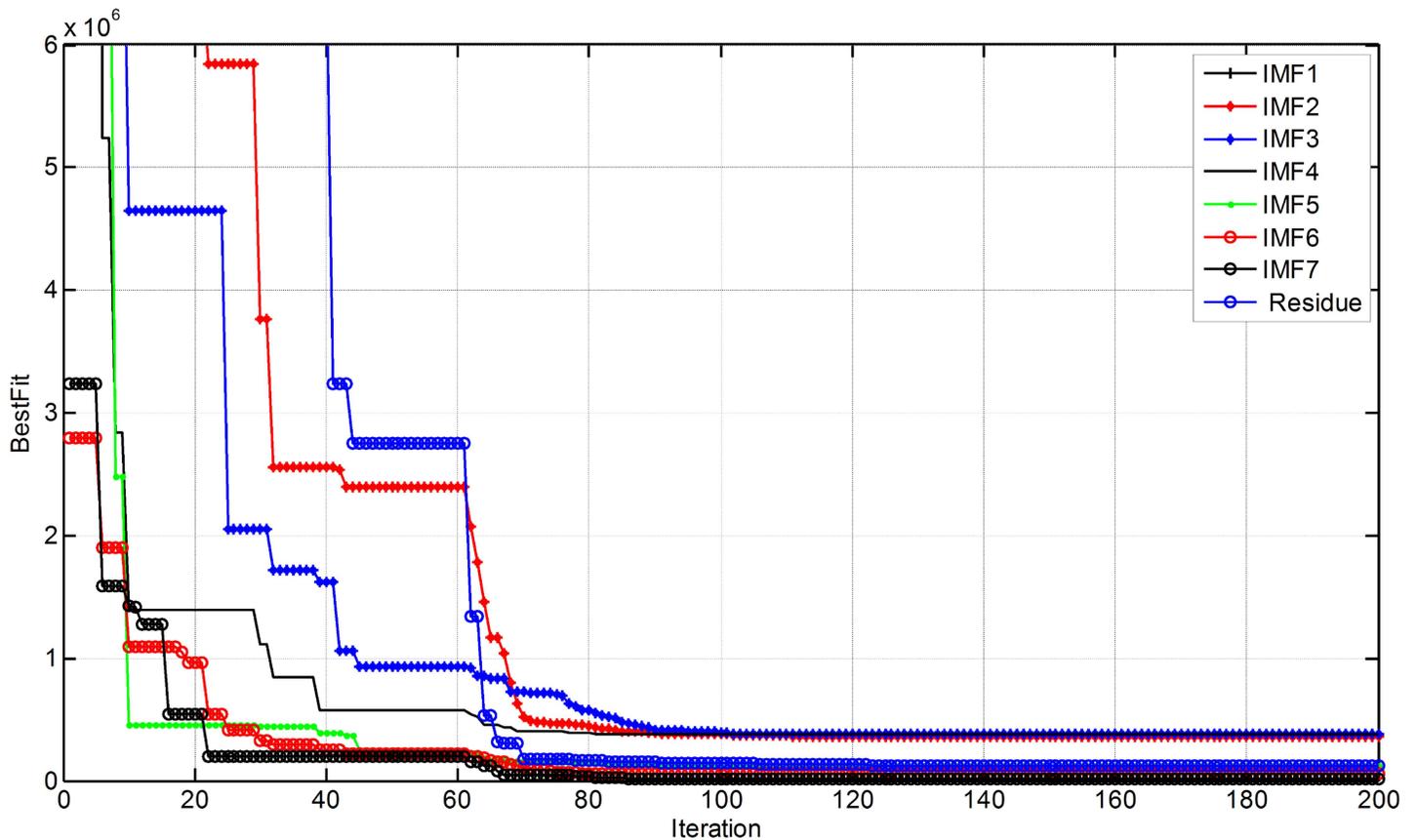
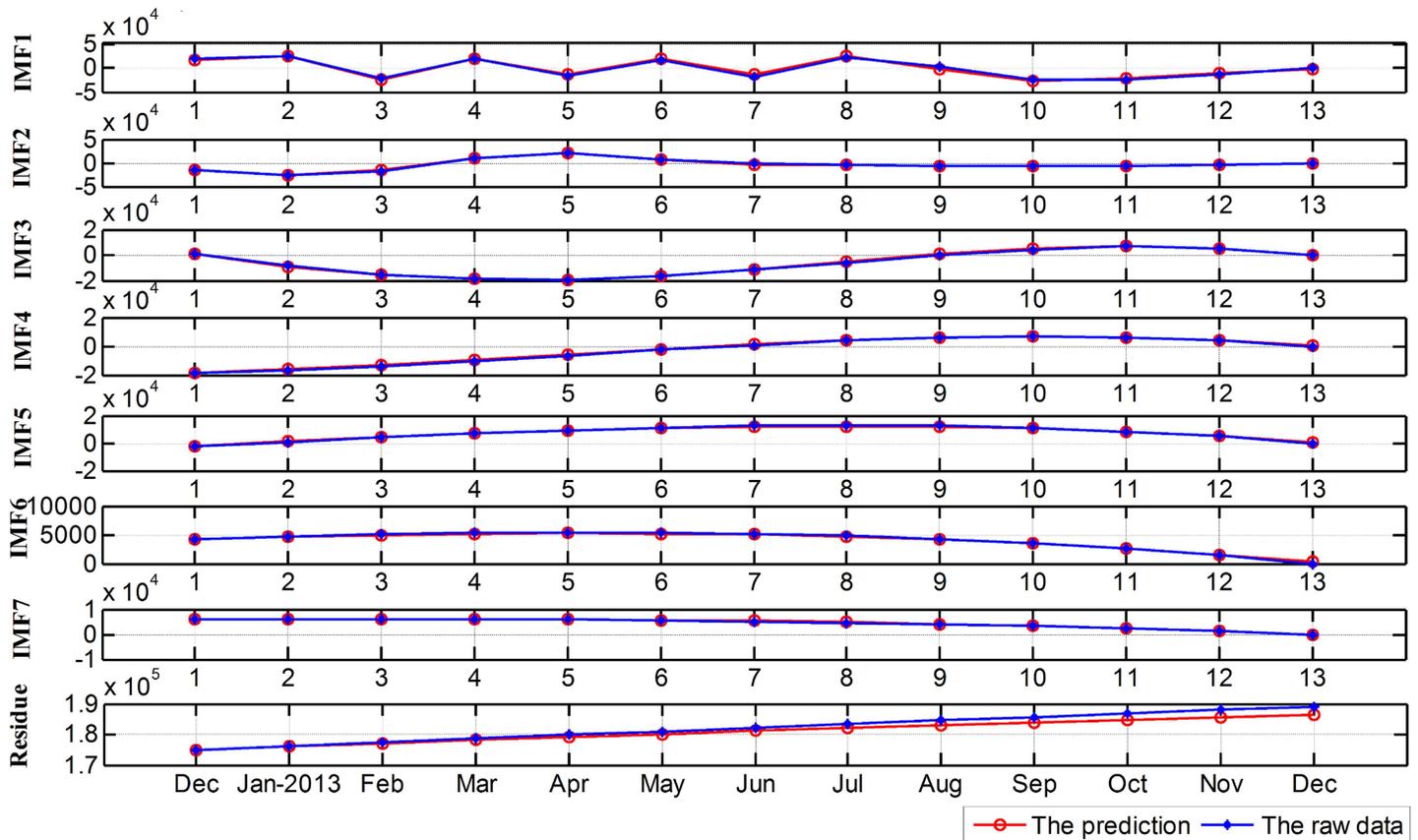


Fig 5. Variation of the best fitness with iteration.

doi:10.1371/journal.pone.0172539.g005



**Fig 6. Result prediction and raw data of each component from December 2012 to December 2013.**

doi:10.1371/journal.pone.0172539.g006

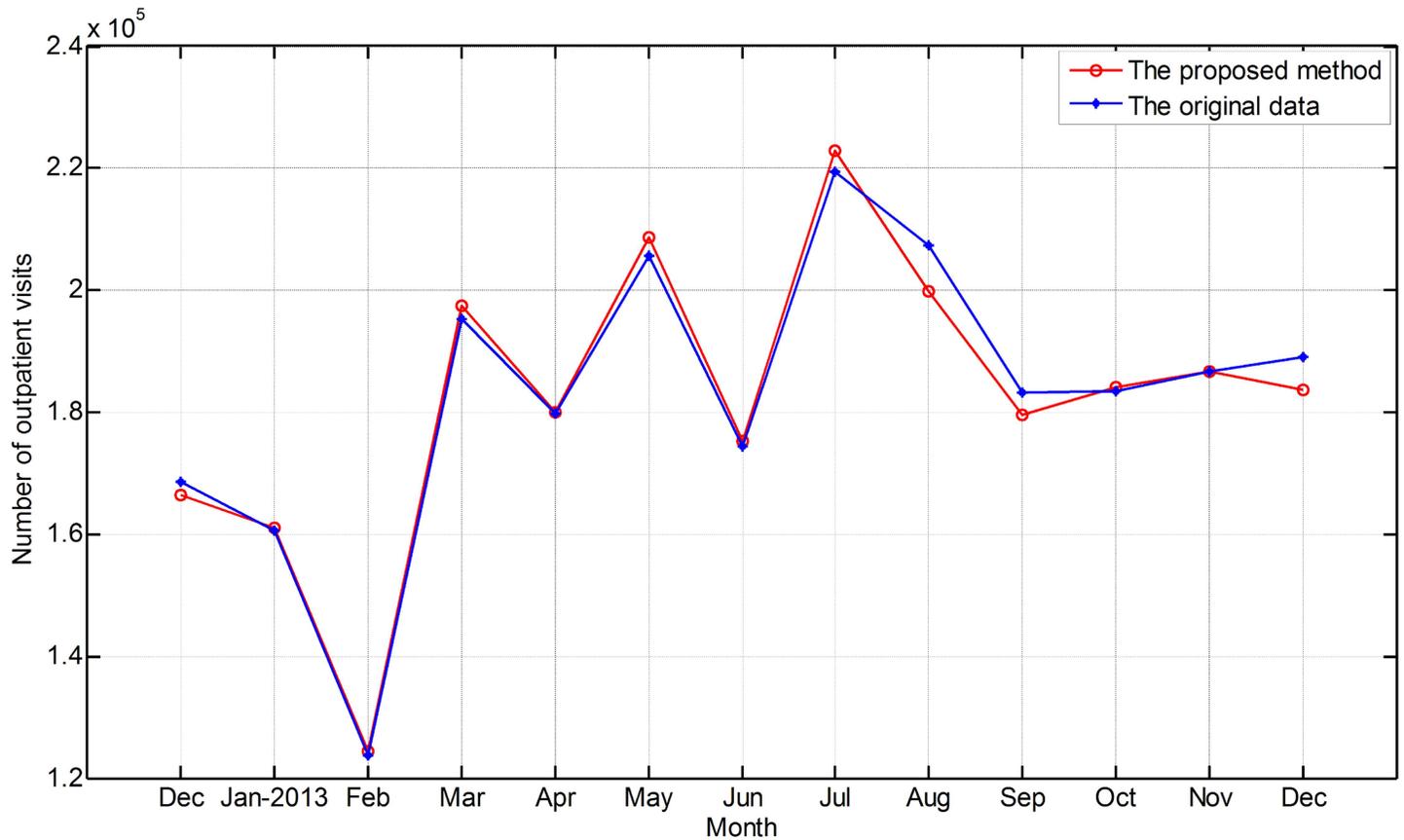
The SSE criterion is given by

$$SSE = \sum_{i=1}^n (\hat{x}_i - x_i)^2 \tag{14}$$

where  $\hat{x}_i$  is the forecasted data and  $x_i$  is the actual data.  $\bar{x}$  denotes the actual data's mean.  $n$  is the number data points considered.

The result predictions by the five methods are shown in Fig 8.

The comparison of evaluation indexes are listed in Table 3. We noted that the RMSE can provide a good measure of model performance for high flows[42] and was selected as the performance criterion of level prediction. MAPE is a relative criterion which is sensitive to the forecasting errors that occur in the low(er) magnitudes of each dataset. SSE is a measure of the discrepancy between the data and an estimation model. A small SSE indicates a tight fit of the model to the data. These three performance indexes are smaller, and the prediction result is more accurate. R has been widely used for model evaluation and was selected as the degree of collinearity criterion of level prediction [41]. The correlation coefficient is larger, and the correlation between the predicted and the observed data is improved. It can be seen from Table 8, the proposed method holds advantages over the other four methods in all the evaluation indexes except the indexes of R, thus proving the effectiveness of the prediction by the proposed approach.



**Fig 7. Prediction value and the observed data.**

doi:10.1371/journal.pone.0172539.g007

The relationship between the seasonal fluctuation index of outpatient visits and the five models prediction results is described in the followings [43].

The seasonal fluctuation index of the same month in eleven years from 2005 to 2013 can be calculated as:

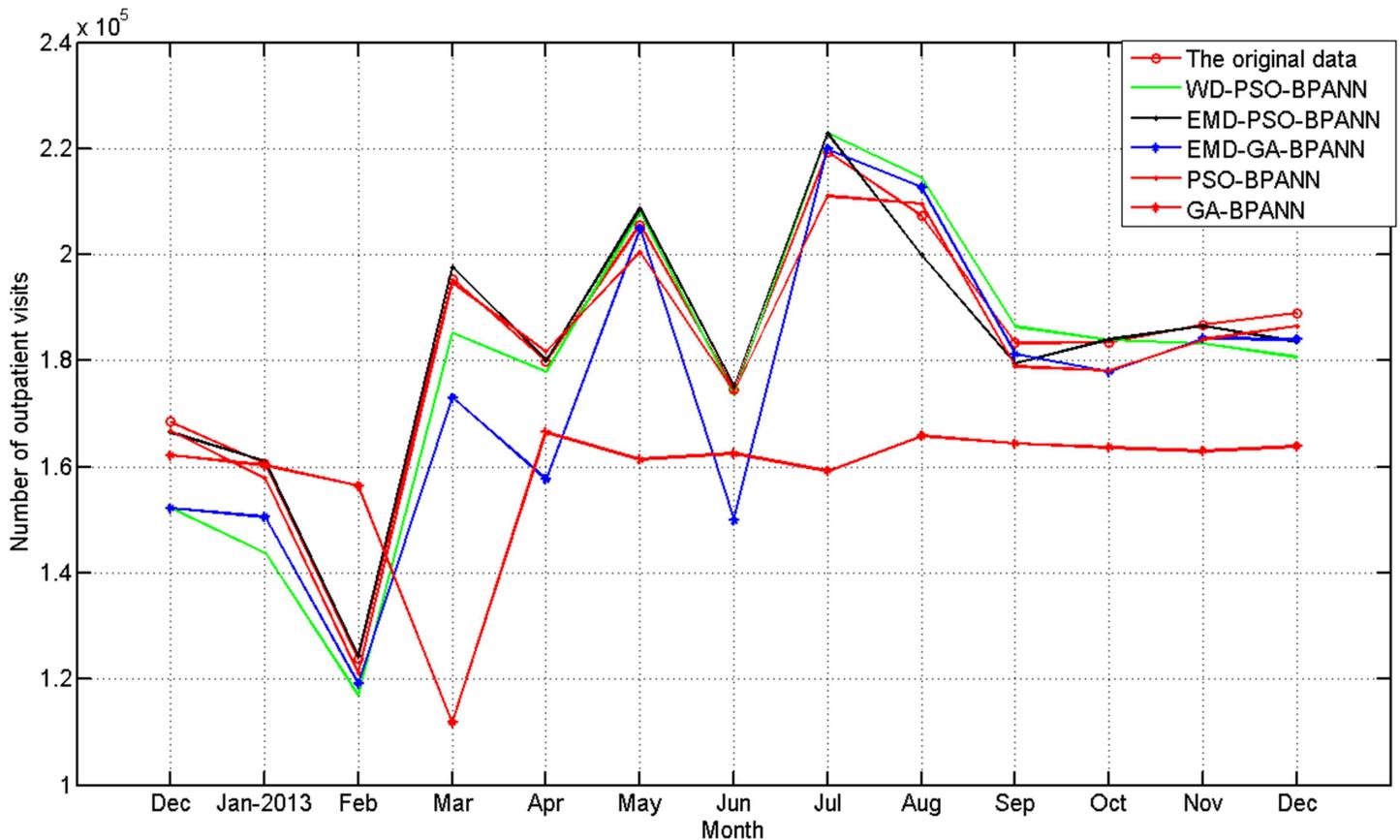
$$SFI1 = \frac{|\bar{x}_{same} - \bar{x}_{all}|}{\bar{x}_{all}} \tag{15}$$

where  $\bar{x}_{same}$  is the average outpatient visits of the same month and  $\bar{x}_{all}$  is the average outpatient visits of all of the months from 2005 to 2013. Obviously, *SFI1* indicates seasonal characteristics of the outpatient visits.

The seasonal fluctuation index of the every month from December 2012 to December 2013 is calculated as:

$$SFI2 = \frac{|x_i - \bar{x}|}{\bar{x}}, i = 1, \dots, 13 \tag{16}$$

Where  $x_i$  is the outpatient visits in each month and  $\bar{x}$  is the average outpatient visits of all of the months from December 2012 to December 2013.



**Fig 8. Comparison predictions with five methods from December 2012 to December 2013.**

doi:10.1371/journal.pone.0172539.g008

The relative error of prediction is used to measure and defined as:

$$RE_i = \frac{|\hat{x}_i - x_i|}{x_i}, i = 1, 2, \dots, n \tag{17}$$

where  $\hat{x}_i$  is the predicted value and  $x_i$  the observed values.

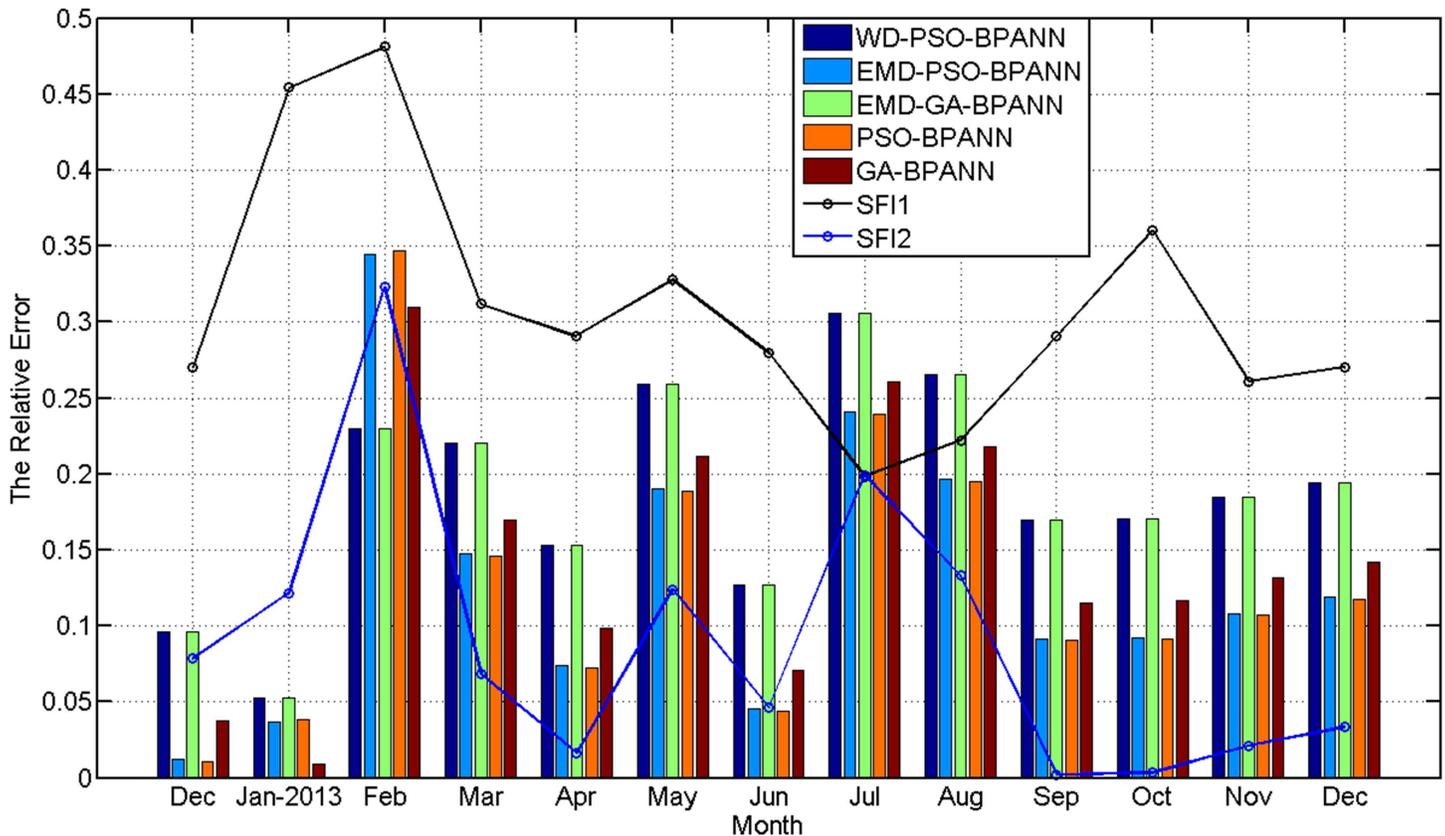
The relative error of the five prediction results and the seasonal fluctuation index of outpatient visits are shown in Fig 9.

It can be seen from Fig 8 that: 1) the curve of *SFI* indicates the outpatient visits has obvious seasonal characteristics. The outpatient visits which happens annually in January and February is relatively high with a rapid decline in July. Other months are relatively stable; 2) the greater the seasonal fluctuation index of the every month from December 2012 to December 2013, the

**Table 3. Comparison of evaluation indexes (best performers are in bold font).**

Methods	Index Performance			
	R	RMSE	MAPE	SSE
WD-PSO-BPANN	0.9759	8.1564e+3	0.13	8.6485e+8
EMD-PSO-BPANN	0.9910	<b>3.1653e+3</b>	<b>0.0241</b>	<b>1.3024e+8</b>
EMD-GA-BPANN	0.9442	12.661e+3	0.34	208.38e+8
PSO-BPANN	<b>0.9925</b>	3.7786e+3	0.11	1.8561e+8
GA-BPANN	-0.0687	36.824e+3	1.13	176.28e+8

doi:10.1371/journal.pone.0172539.t003



**Fig 9. The relationship between the seasonal fluctuation index and RE of the predictions by the five methods.** (Histograms and curves represent RE of the predictions and the seasonal fluctuation index, respectively).

doi:10.1371/journal.pone.0172539.g009

greater the relative error of the five methods. and 3) the absolute error of the WD-PSO-BPANN and EMD-GA-BPANN are larger than that of the other three methods when the outpatient visits data is stable, such as from September to October; the absolute error of the EMD-PSO-BPANN and PSO-BPANN are smaller than that of the other three methods in all the prediction months.

### Conclusion

A new forecasting method that combines EMD and BPANNs based on PSO is proposed to forecast outpatient visits. Simulation results show that this method can improve forecasting and thus help policy makers manage hospitals effectively.

### Supporting information

**S1 Dataset. Number of outpatient visits by month from 2005 to 2013.**  
(XLS)

### Acknowledgments

This work was supported by the project of Basic Ability Promotion for Young Teachers of the Guangxi Education Department (KY2016YB093). The funders held no role in the study design, data collection, and analysis, decision to publish, or manuscript preparation.

## Author Contributions

**Conceptualization:** DH.

**Data curation:** DH.

**Formal analysis:** DH.

**Funding acquisition:** DH.

**Methodology:** DH.

**Project administration:** DH.

**Software:** DH.

**Validation:** DH ZW.

**Writing – original draft:** DH ZW.

**Writing – review & editing:** DH ZW.

## References

1. Chenga CH, Wangb JW, Li CH (2008) Forecasting the number of outpatient visits using a new fuzzy time series based on weighted-transitional matrix. *EXPERT SYST APPL* 34(4):2568–2575.
2. Hadavandi E, Shavandi H, Ghanbari A (2011) An improved sales forecasting approach by the integration of genetic fuzzy systems and data clustering: Case study of printed circuit board. *EXPERT SYST APPL* 38(8):9392–9399.
3. Wang Y, Gu J, Zhou Z, Wang Z. (2015) Diarrhoea outpatient visits prediction based on time series decomposition and multi-local predictor fusion. *KNOWL-BASED SYST* 88(C):12–23.
4. Huang N E, Shen Z, Long S R, Wu M C, Shih H H, Zheng Q, et al. (1998) The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-stationary Time Series Analysis. *Proceedings: Mathematical, Physical and Engineering Sciences* 454(1971):903–995
5. Wang S, Zhang N, Wu L, Wang Y(2016) Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method. *RENEW ENERG* 94:629–636.
6. Fu K, Qu J, Chai Y, Dong Y (2014) Classification of seizure based on the time-frequency image of EEG signals using HHT and SVM. *Biomed Signal Process Control* 13(5):15–22.
7. Paturzo M, Ferraro P (2014) Multilevel bidimensional empirical mode decomposition: a new speckle reduction method in digital holography. *OPT ENG.* 14(11):139–152.
8. Zabalza J, Ren J, Wang Z, Marshall S, Wang J (2014) Singular spectrum analysis for effective feature extraction in hyperspectral imaging. *IEEE GEOSCI REMOTE S.* 11(11):1886–1890.
9. Shakya P, Kulkarni M S, Darpe A K(2014) A novel methodology for online detection of bearing health status for naturally progressing defect. *J SOUND VIB.* 333(21):5614–5629.
10. Kisi O, Latifoğlu L, Latifoğlu F(2014) Investigation of empirical mode decomposition in forecasting of hydrological time series. *WATER RESOUR MANAG.* 28(12):4045–4057.
11. Abadan S, Shabri A (2014) Hybrid empirical mode decomposition-ARIMA for forecasting price of rice. *Appl Math Sci.* 8(63): 3133–3143.
12. Sudheer C, Maheswaran R. Panigrahi B K, Mathur S (2014) A hybrid SVM-PSO model for forecasting monthly streamflow, *Neural Comput & Applic* 24:1381–1389.
13. Ilias HA, Chai XR, Abu Bakar AH, Mokhlis H (2015) Transformer incipient fault prediction using combined artificial neural network and various particle swarm optimisation techniques. *PLoS ONE* 10(6): e0129363. doi: [10.1371/journal.pone.0129363](https://doi.org/10.1371/journal.pone.0129363) PMID: [26103634](https://pubmed.ncbi.nlm.nih.gov/26103634/)
14. Warner B, Misra M(2012) Understanding neural networks as statistical tools. *AM STAT.* 50(4):284–293.
15. Grossi E, Buscema M. (2007) Introduction to artificial neural networks. *EUR J GASTROEN HEPAT.* 19(12):1046–1054.
16. Chen Y, Xu J, Yu H, Zhen Z(2016) Three-dimensional short-term prediction model of dissolved oxygen content based on PSO-BPANN algorithm coupled with kriging Interpolation. *MATH PROBL ENG*, 2016: Article ID 6564202, 10 pages.

17. Zhang X, Liu Y, Yang M, Zhang T, Young AA, Li X (2013) Comparative study of four time series methods in forecasting typhoid fever incidence in China. *PLoS ONE* 8(5):e63116. doi: [10.1371/journal.pone.0063116](https://doi.org/10.1371/journal.pone.0063116) PMID: [23650546](https://pubmed.ncbi.nlm.nih.gov/23650546/)
18. Gan R, Chen X, Yan Y, Huang D (2015) Application of a hybrid method combining grey model and back propagation artificial neural networks to forecast hepatitis B in China, *Comput Math Methods Med*. 2015: Article ID 328273, 7 pages.
19. Zhang G, Patuwo B E, Hu M Y (1998) Forecasting with artificial neural networks: The state of the art. *INT J FORECASTING* 14(1):35–62.
20. Huang DZ, Gong RX, Gong S (2015) Prediction of wind power by chaos and bp artificial neural networks approach based on genetic algorithm. *J ELECTR ENG TECHNOL*. 10(1):41–46.
21. Samanta B (2004) Gear fault detection using artificial neural networks and support vector machines with genetic algorithms. *MECH SYST SIGNAL PR*. 18(3):625–644.
22. Yu Y, Li Y, Li J (2015) Nonparametric modeling of magnetorheological elastomer base isolator based on artificial neural network optimized by ant colony algorithm. *J INTEL MAT SYST STR*. 26(14):1789–1798.
23. Alavi A H, Gandomi A H (2011) Prediction of principal ground-motion parameters using a hybrid method coupling artificial neural networks and simulated annealing. *COMPUT STRUCT*. 89(23–24):2176–2194.
24. Yu J, Wang S, Xi L (2008) Evolving artificial neural networks using an improved PSO and DPSO. *Neuro-computing*. 71(4–6):1054–1060.
25. Momeni E, Armaghani D J, Hajihassani M, Amin M (2015) Prediction of uniaxial compressive strength of rock samples using hybrid particle swarm optimization-based artificial neural networks. *Measurement*. 60(1):50–63.
26. Zhang P, Chen H, Liu X, Zhang Z (2015) An iterative multi-objective particle swarm optimization-based control vector parameterization for state constrained chemical and biochemical engineering problems. *BIOCHEM ENG J*. 103:138–151.
27. Wang Y, Shen Y, Yuan X, Yang Y (2014) Operating point optimization of auxiliary power unit based on dynamic combined cost map and particle swarm optimization. *IEEE T POWER ELECTR*. 30(12):7038–7050.
28. Fakhar M S, Kashif S A R, Saqib M A, Hassana T (2015) Non cascaded short-term hydro-thermal scheduling using fully-informed particle swarm optimization. *INT J ELEC POWER*. 73:983–990.
29. Montain M E, Blanco A M, Bandoni J A (2015) Optimal drug infusion profiles using a particle swarm optimization algorithm. *COMPUT CHEM ENG*. 82:13–24.
30. Neshat N, Amin-Naseri M R (2015) Cleaner power generation through market-driven generation expansion planning: an agent-based hybrid framework of game theory and particle swarm optimization, *J CLEAN PROD*. 105: 206–212.
31. Tsai Y L, Tien C H (2015) Approach to analytically minimize the LCD moire by image-based particle swarm optimization. *APPL OPTICS*. 54(28):41–6.
32. Ren P, Yao S, Li J, Valdes-Sosa PA, Kendrick KM (2015) Improved prediction of preterm delivery using empirical mode decomposition analysis of uterine electromyography signals. *PLoS ONE* 10(7): e0132116. doi: [10.1371/journal.pone.0132116](https://doi.org/10.1371/journal.pone.0132116) PMID: [26161639](https://pubmed.ncbi.nlm.nih.gov/26161639/)
33. Teodoro D, Lovis C (2013) Empirical mode decomposition and k-nearest embedding vectors for timely analyses of antibiotic resistance trends. *PLoS ONE* 8(4): e61180. doi: [10.1371/journal.pone.0061180](https://doi.org/10.1371/journal.pone.0061180) PMID: [23637796](https://pubmed.ncbi.nlm.nih.gov/23637796/)
34. Pi E, Mantri N, Ngai S M, Lu H, Du L (2013) BP-ANN for fitting the temperature-germination model and its application in predicting sowing time and region for bermudagrass. *PLoS ONE* 8(12): e82413. doi: [10.1371/journal.pone.0082413](https://doi.org/10.1371/journal.pone.0082413) PMID: [24349278](https://pubmed.ncbi.nlm.nih.gov/24349278/)
35. Jiang S Y, Ren Z Y, Xue K M, Li C F (2008) Application of BPANN for prediction of backward ball spinning of thin-walled tubular part with longitudinal inner ribs, *J MATER PROCESS TECH*, 196: 190–196.
36. Eberhart R, Kennedy J (1995) A new optimizer using particle swarm theory. *International Symposium on MICRO Machine and Human Science* 39–43.
37. Shi B Y, Eberhart R (1998) A modified particle swarm optimizer. *Evolutionary Computation Proceedings, IEEE World Congress on Computational Intelligence*. 69–73.
38. Ratnaweera A, Halgamuge S, Watson H C (2004) Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients. *IEEE T EVOLUT COMPUT* 8(3):240–255.
39. Gevrey M, Dimopoulos L, Lek S (2003) Review and comparison of methods to study the contribution of variables in artificial neural network models. *ECOL MODEL* 160(3):249–264.

40. Sadati N, Amraee T, Ranjbar AM (2009) A global particle swarm-based-simulated annealing optimization technique for under-voltage load shedding problem. *APPL SOFT COMPUT* 9: 652–657.
41. Wang W C, Xu D M, Chau K W, Chen S (2013) Improved annual rainfall-runoff forecasting using psosvm model based on eemd. *J HYDROINFORM* 15(4):1377–1390.
42. Karunanithi N, Grenney WJ, Whitley D, Bovee K (1994) Neural networks for river flow prediction. *J Comput Civ Eng ASCE* 8(2):201–220.
43. Gan R, Chen N, Huang D. (2016) Comparisons of forecasting for hepatitis in Guangxi Province, China by using three neural networks models. *PeerJ* 4:e2684. doi: [10.7717/peerj.2684](https://doi.org/10.7717/peerj.2684) PMID: [27843718](https://pubmed.ncbi.nlm.nih.gov/27843718/)