

Article

# Application of a Hybrid Model Based on Echo State Network and Improved Particle Swarm Optimization in PM<sub>2.5</sub> Concentration Forecasting: A Case Study of Beijing, China

# Xinghan Xu<sup>1,\*</sup> and Weijie Ren<sup>2,\*</sup>

- <sup>1</sup> Department of Environmental Engineering, Kyoto University, Kyoto 615-8540, Japan
- <sup>2</sup> Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024, China
- \* Correspondence: xu.xinghan.57u@st.kyoto-u.ac.jp (X.X.); renweijie@mail.dlut.edu.cn (W.R.); Tel.: +86-0411-84706002-2607 (W.R.)

Received: 11 May 2019; Accepted: 30 May 2019; Published: 31 May 2019



**Abstract:** With the acceleration of urbanization, there is an increasing trend of heavy pollution. PM<sub>2.5</sub>, also known as fine particulate matter, refers to particles in the atmosphere with a diameter of less than or equal to 2.5 microns. PM<sub>2.5</sub> has a serious impact on human life, a sustainable city, national economic development, and so on. How to forecast the PM<sub>2.5</sub> concentration accurately, and then formulate a scientific air pollution prevention and monitoring program is of great significance. This paper proposes a hybrid model based on echo state network (ESN) and an improved particle swarm optimization (IPSO) algorithm for the Beijing air pollution problem, and provides a method for PM<sub>2.5</sub> concentration forecasting. Firstly, the PSO algorithm is improved to speed up the search performance. Secondly, the optimal subset of the original data is selected by the convergence cross-mapping (CCM) method. Thirdly, the phase space reconstruction (PSR) process is combined with the forecasting model, and some parameters are optimized by the IPSO. Finally, the optimal variable subset is used to predict PM<sub>2.5</sub> concentration. The 11-dimensional air quality data in Beijing from January 1 to December 31, 2016 are analyzed by the proposed method. The experimental results show that the hybrid method is superior to other comparative models in several evaluation indicators, both in one-step and multi-step forecasting of PM<sub>2.5</sub> time series. The hybrid model has good application prospects in air quality forecasting and monitoring.

Keywords: air quality; PSO; ESN; hybrid model; PM<sub>2.5</sub> forecasting; sustainable development

# 1. Introduction

With the advancement of industrialization, air pollution has become increasingly serious, and the problems of air quality have attracted more and more public attention. Beijing-Tianjin-Hebei area, the Yangtze River Delta, and the Pearl River Delta have been included in key air pollution monitoring areas. Air quality is a reflection of the degree of air pollution and is judged by the pollutant concentration in the air [1]. Air pollutants include soot, total suspended particulate matter, inhalable particulate matter ( $PM_{10}$ ), fine particulate matter ( $PM_{2.5}$ ), CO, NO<sub>2</sub>, O<sub>3</sub>, SO<sub>2</sub>, volatile organic compounds, etc., where  $PM_{2.5}$  refers to particles in the atmosphere with a diameter less than or equal to 2.5 microns, which can enter the lungs [2]. Although  $PM_{2.5}$  is only a small component in the earth's atmosphere and has a small particle size, it contains a large amount of toxic and harmful substances, has a long residence time in the atmosphere, and has a long transportation distance, which has an important impact on air quality and visibility, and has direct or indirect effects on human health and plant growth [3]. Forecasting of  $PM_{2.5}$ 



concentration is of great significance for building sustainable cities. PM<sub>2.5</sub> forecasting contributes to environmental improvement and to implement air quality citizen science, i.e., all those studies by which citizens could be more directly involved in participatory environment monitoring and sustainable cities. Therefore, it is necessary to monitor and forecast the PM<sub>2.5</sub> concentration in real time. In order to strengthen the management of air quality, the Ministry of Environmental Protection in China has released a daily report of the Air Quality Index (AQI) since 2012 [4].

Air pollution in North China is serious, and environmental issues are receiving more and more attention. Particularly, the air pollution problem in Beijing area has attracted the most attention. It is important for China's sustainable development to monitor and forecast of the  $PM_{2.5}$  concentration in Beijing area, with Beijing's unique political, economic, and cultural status. China's Environmental Status Bulletin issued by the Ministry of Environmental Protection shows that although the air quality of  $PM_{2.5}$  in the Beijing-Tianjin-Hebei metropolitan region has improved compared with previous years, it is still the most polluted area in China, and the number of cities in the Beijing-Tianjin-Hebei region accounts for more than five of the 10 cities with relatively poor air quality in 74 cities each year. Therefore, the Beijing-Tianjin-Hebei urban agglomeration, as the representative area, is chosen to forecast the  $PM_{2.5}$  concentration data in the Beijing area in order to provide an effective method for regional monitoring and control.

At present, researchers at China and abroad have conducted a lot of research on the generation, dissemination, and forecasting of  $PM_{2.5}$ . Among them, the model of air quality forecasting develops rapidly which can be roughly divided into two categories: the numerical forecasting model and the statistical forecasting model [5,6]. Numerical based forecasting models are based on different chemical mechanisms and chemical kinetic equations which is a kind of chemical transport model. However, due to the large number of parameters in the model, the application is difficult, the forecasting accuracy is not very high, and the calculation amount is large [5,7]. Statistical-based predictive models do not rely on chemical mechanisms and chemical kinetic equations, and can be forecasted by analyzing their regularity [6,8]. The accuracy of statistical forecasting is higher than that of numerical forecasting, and the statistical forecasting model is simple, fast, and low cost.

The method of air quality forecasting based on machine learning belongs to statistical forecasting and has become the main research direction of PM<sub>2.5</sub> forecasting [9]. These methods, such as backpropagation neural networks (BPNN) [10], support vector regression (SVR) [11], radial basis function neural networks [12], echo state networks (ESNs) [5], and other machine learning models, have been widely used in air quality forecasting. Although the machine learning model has achieved significant results in air quality forecasting, there are still some problems, such as manual adjustment of parameters, input variable redundancy, and so on [13]. Therefore, more and more demands make us develop more effective AQI series modeling methods. With the development of artificial intelligence technology, intelligent optimization algorithms are gradually welcomed in air quality forecasting [14].

In recent years, some researchers have applied intelligent optimization algorithms to the modeling process of time series forecasting [15–17]. They use various intelligent optimization algorithms to select input variables, optimize model structure and adjust parameters which have attracted more and more attention from researchers of different backgrounds. In 2016, Niu et al [15] proposed a new short-term forecasting for the PM<sub>2.5</sub> concentration based on complete ensemble empirical mode decomposition method and the grey wolf optimization (GWO) algorithm. In 2017, Sun et al [16] proposed a hybrid model based on principal component analysis and the cuckoo search (CS) optimized Least Squares Support Vector Machine for air quality forecasting and monitoring. In 2018, Li et al [17] established a forecasting model of atmospheric PM<sub>2.5</sub> and nitrogen dioxide (NO<sub>2</sub>) concentration based on SVR. The Quantum PSO algorithm is used to select the optimal parameters that affect the performance of the SVR model. In 2019, Zhu et al [11] focused on modeling and forecasting atmospheric pollution data, and proposed a two-step hybrid model of NO<sub>2</sub> and SO<sub>2</sub> forecasting based on data preprocessing and intelligent optimization algorithms (CS and GWO). In the above popular variants, the intelligent optimization algorithm and the machine learning model have achieved good results in forecasting AQI

time series. It has a significant effect on improving forecasting accuracy and efficiency of the model and system.

Aiming at the problems in PM<sub>2.5</sub> forecasting, this paper proposes a hybrid method based on ESN and classical PSO [18] to improve the forecasting accuracy of the PM<sub>2.5</sub> concentration. First of all, this paper improves the classical PSO algorithm, which improves the search ability of the algorithm and maintains a good balance between exploration and development. Then, the CCM [19] method is used to analyze the correlation of the original data, and the relevant variables are retained. After that, the improved PSO algorithm is used to optimize the embedding dimension and delay time of the PSR [20] process, and the hyper-parameters of the ESN model. Finally, the optimized PSR method is used to map the optimal variable subset to the high dimensional space, and the optimized ESN model is used to predict the PM<sub>2.5</sub> concentration. The establishment of the hybrid forecasting model might provide a basis for the government to formulate environmental protection policies, help the society in sustainable development, and provide suggestions for citizens to go out for activities and health care.

# 2. Methods

### 2.1. Echo States Networks

Recurrent neural networks [21] have a rich nonlinear dynamics mechanism, but the training process is mostly based on the gradient principle, resulting in suboptimal speed and local optimal problems. In addition, the output of the recurrent neural network is a function of time. The training process of recursive connection weights in the network is related to the output, so it is difficult to guarantee the stability of the network. To solve these problems, Jaeger proposed an improved recurrent neural network reserve pool construction [22]. The internal unit of the ESN reserve pool is a simulated neuron, which is a linear or sigmoid function. In 2004, Jaeger et al [23] proposed a method for learning nonlinear systems, ESN model, which caused a strong reaction from the academic community, making ESN one of the research hotspots in the field of time series prediction.

The ESN is a three-layer recursive neural structure that includes an input layer, a hidden layer, and an output layer, as shown in Figure 1. In ESN, the hidden layer, also known as the reserve pool, contains hundreds of sparsely recursively connected neurons. Prior to network training, the random input initialization of the reserved pool and the random initialization of the internal connection weights are no longer changed given the network parameters. All we need is to get the connection weight between the reserve pool and the output layer through training.



Figure 1. The structure of echo state network (ESN).

The state equation of ESN is:

$$x(t) = \phi(W_{in}u(t) + W_x x(t-1) + W_{back} y(t-1) + b_x).$$
(1)

The output equation of ESN is:

$$y(t) = w^T x(t) + b, (2)$$

where,  $u(t) \in \mathbb{R}^{M \times 1}$  is the *M*-dimensional input vector,  $y(t) \in \mathbb{R}^{M \times 1}$  is the *M*-dimensional output vector,  $b_x \in R^{N \times 1}$  is the input layer offset,  $b \in R^{N \times 1}$  is the output layer offset,  $x(t) \in R^{N \times 1}$  is the internal state of the reserve pool at time *t*, which is obtained from the input u(t) at time *t* and the state x(t-1) at time (t-1).  $\phi(\cdot)$  indicates the activation function. The elements in the input weight matrix  $W_{in} \in \mathbb{R}^{N \times N}$ take values in the interval  $[\gamma, \gamma]$  and  $\gamma$  is positive real numbers less than one, called input transform coefficients. The sparsity range of the connection weight matrix  $W_x \in \mathbb{R}^{N \times N}$  in the reserve pool is [1%, 10%], that the less the connection weight, the smaller the sparsity. In general, the spectral radius of the connection weight matrix  $W_x$  is less than 1, in order to ensure the stability of the network. As time t goes on, the impact of x(0) is getting smaller and smaller, which reflects the state of the network with state forgetting.  $W_{back} \in \mathbb{R}^{N \times M}$  is the output feedback matrix, which feeds back the output from the previous moment to the current time state. However, due to the particularity of time series prediction, the target output y(t-1) at the previous moment is also the input u(t) of the current moment, so  $W_{hack}y(t-1)$  can also be merged into  $W_{in}u(t)$ . In order to simplify the representation, the neuron input in the reserve pool is fixed to 1, and only connected to the output, not connected to the neurons in the input pool and the reserve pool, besides  $b_x$  and b can be merged into matrices  $W_x$ and *w*. Then the upper formula can be abbreviated for:

$$x(t) = \phi(W_{in}u(t) + W_x x(t-1)),$$
(3)

$$y(t) = w^T x(t), (4)$$

where there is a linear relationship with it, so it can be solved by linear regression. Compared with the traditional recurrent neural network, ESN solves the problem that the model structure is difficult to determine, simplifies the process of solving unknown parameters, and makes up for the shortcomings of traditional RNNs, such as slow convergence speed and ease of falling into the local optimum [24]. The structure of the sparse reserve pool ensures that the network has rich dynamic characteristics, good generalization ability, and predictive performance.

#### 2.2. Particle Swarm Optimization

#### 2.2.1. Basic Method

The PSO algorithm was first proposed in 1995 by American social psychologist James Kennedy and electrical engineer Russell Eberhart [18]. The basic idea of the PSO algorithm is to simulate the foraging process of bird populations and to model their social behavior. Researchers have found that birds often change direction, disperse, and suddenly gather during flight. Their behavior is unpredictable, but the entire team is always consistent and maintains the most appropriate distance between individuals that no collisions will occur [25]. Through the study of the behavior of similar biological groups, it is found that there is a social information sharing mechanism in the biological group, which provides favorable conditions for the evolution of the group and is also the basis for the formation of the PSO. The most important feature of the algorithm is that it is easy to implement, with few control parameters and fast calculation speed. Like other evolutionary algorithms, PSO also achieves the search for optimal solutions in complex spaces through cooperation and competition among individuals. In general, the problems that genetic algorithms can solve can basically be solved by PSO, and the efficiency may be higher.

If the overall size is *N*, the position of particle  $i(i = 1, 2, \dots, N)$  in the search space can be expressed as  $x = (x_1, x_2, \dots, x_D)$ , and the flight speed can be expressed as  $v = (v_1, v_2, \dots, v_D)$ . Then in the generation (t + 1), the particle *i* update its position and speed according to the following formula:

$$v_{id}(t+1) = \omega v_{id}(t) + c_1 r_1 (p_{id}(t) - x_{id}(t)) + c_2 r_2 (p_{gd}(t) - x_{id}(t)),$$
(5)

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1),$$
(6)

where  $d = 1, \dots, D$ ,  $p_{id}$  is the optimal position vector searched by the particle *i*, which is called the individual history optimal.  $p_{gd}$  is the optimal position vector searched by all the particles, which is called the group history is the best.  $\omega$  is the coefficient that maintains the original speed, which is called inertia weight.  $c_1$  and  $c_2$  are constant called learning factor, which reflects the information exchange between particles.  $c_1$  is the weight coefficient of the particle tracking its own optimal value, which means that the particle knows itself while  $c_2$  is the weighting coefficient of the particle tracking population optimal value, indicating the particle's understanding of the whole group.  $r_1$  and  $r_2$  are random numbers that is uniformly distributed within the interval [0, 1].

The update formula of speed consists of three parts: the first part is the a priori velocity of the particle, which represents the current state of the particle, its role is to balance the global and local search capabilities; the second part is the "individual cognition" part of the particle, which represents the particle's understanding of its behavior, and its R. OLE is to make the particle have sufficient global search ability, and the third part is the "social cognition" part of the particle, which is the "social cognition" part of the particle. Information sharing and cooperation between subsystems. These three parts together determine the search ability of particles in the search space. From the perspective of social psychology, the PSO algorithm can be described as the process of seeking consensus, individuals often remember their beliefs, and consider the beliefs of their peers. When individuals better understand the beliefs of their peers, they adjust themselves adaptively. The individual historical optimal position  $p_{id}$  of the particle *i* is updated by the following formula:

$$p_{id}(t+1) = \begin{cases} x_{id}\{t+1, & if x_{id}\{t+1 \le p_{id}\{t \\ p_{id}\{t, & if x_{id}\{t+1 > p_{id}\{t \end{cases}.$$
(7)

The global historical optimal position  $p_{gd}$  of the entire group is updated by the following formula:

$$p_{gd}(t+1) = \min(p_{id}(t+1)), i = 1, \cdots, N.$$
 (8)

#### 2.2.2. Algorithm Improvement

PSO is widely used in various optimization problems because of its superior performance, but at the same time it is easy to fall into local optimum. This paper makes some improvements to the PSO algorithm:

(1) Since the initial position of the particles is randomly generated, it is difficult to ensure the quality of the initial particles in the classic PSO. In some cases, the distribution of particles in space is not uniform. Therefore, after a period of iteration, the particles may convergence prematurely, causing adverse effects. The theory of good point [26] set was originally proposed by Hua Luogeng, which is an effective method for uniformly selecting points. Compared with the method of random initialization, the points can be more evenly distributed in the search space. Therefore, this paper applies the good point set to the PSO algorithm to generate the initial population.

(2) Combining chaotic mapping with PSO algorithm improves the ergodicity and randomness of the new solution. Based on the randomness and ergodicity of chaotic maps, Logistic mapping [27] can be used to prevent populations from falling into local optimum. This paper uses the current  $p_{gd}$  as an initial condition to generate a chaotic logistic sequence. Then select the point that produces the smallest target value in the Logistic sequence to randomly replace one solution in the current population. The chaotic mapping method can make the population generate some points away from the local optimal solution so that the inertial particles escape the local optimal solution and quickly search for the global optimal solution.

The improved PSO algorithm is named IPSO, it inherits the advantages of the original PSO algorithm, pays more attention to the discussion of the particle initialization process, adopts certain strategies to overcome the problems of local optimization and premature convergence, and improves

the search ability of particles in space. The pseudo-code of the improved algorithm is shown in Algorithm 1.

Algorithm 1 IPSO algorithm

Input: Population size, Maximum iteration, Objective function, Dimensionality of decision variables 1: Initialize swarm by the theory of good point 2: Evaluate each particle with objective function 3: Select the best position of particles as  $p_{gd}$  from the swarm 4: Set the position of each particle as *p<sub>id</sub>* 5: T = 1; 6: while  $T \leq MaxIt do$ 7: Adaptive adjust the inertia weight  $\omega$  and learning factors  $c_1$  and  $c_2$  [28] 8: for each particle do 9: Update particle information on the basis of the velocity and position updating formula 10: Evaluate each particle with objective function 11: Update the individual optimal position  $p_{id}$ 12: end for 13: Update the population optimal position  $p_{gd}$ 14: Generate a set of logistic sequences based on current  $p_{gd}$ , and random substitution of a solution in the current population by selecting the point with the smallest objective value in the generated Logistic sequence 15: T++ 16: end while Output: Optimal value

#### 3. Data Preprocessing

## 3.1. Data Description

Beijing [29], located in the North China Plain (39.56°N, 116.20°E), is the capital of the People's Republic of China and a modern international city. The total area of Beijing is 16,410.54 square kilometers. The eastern region of Beijing is adjacent to Tianjin, and the rest is adjacent to Hebei. The climate is typical of the semi-humid continental monsoon climate in the north temperate zone. Due to the large amount of coal burning and steel-making in the Beijing-Tianjin-Hebei metropolitan region for power generation, Beijing has suffered serious air pollution problems, which are not conducive to the sustainable city.

In order to better study the air pollution situation in the Beijing area, Beijing air data, from 2016, was used as sample data. The main reason for choosing Beijing as a research location is that a large part of Beijing and China are experiencing long-term air pollution, which is greatly affected by air pollutants and has a long history of pollution, and the research analysis of pollutants are very representative [30,31]. The main pollutants are fine particles, especially  $PM_{2.5}$ . Epidemiological evidence suggests that exposure to  $PM_{2.5}$  can lead to lung disease, severe respiratory and cardiovascular disease, and even death [2]. There are many reasons that may lead to lung diseases and even lung cancer, such as air pollution, cigarette smoking, and radon emissions [32]. Our study only deals with the effects of  $PM_{2.5}$  on the lung. Through effective forecasting of  $PM_{2.5}$  concentration, citizens can reduce outdoor activities during high pollution periods, thereby reducing the risk of diseases. If the proposed model has better prediction performance for  $PM_{2.5}$  data in Beijing, the proposed model will have strong practical value. The hybrid model proposed in this paper has conducted in-depth research on the prediction of  $PM_{2.5}$  data in Beijing.

The data used in this paper is the hourly air quality data and meteorological data observed from January 1, 2016, to December 31, 2016, in Beijing. The data include hourly averages of  $PM_{2.5}$ ,  $PM_{10}$ ,  $SO_2$ ,  $NO_2$ ,  $O_3$ , and CO, as well as hourly averages of temperature (T), pressure (P), humidity (H), wind speed (WS), and wind direction (WD). There are 11-dimensional variables in total, and the sample length included 8759 sets of data.

This article divides the sample data into two parts: the training set and the test set. The first 6570 sets of data (about 75% of the total data) are used as training data, and the remaining 25% are used as test data. Before predicting the data, the original data is mapped to the high-dimensional space by PSR, and then the input variable is selected by the IPSO algorithm. Finally, the restructured variables are input into the ESN model for prediction.

## 3.2. Causality Analysis

In 2012, Sugihara et al. proposed a convergent cross mapping (CCM) method in Science [19], which attracted extensive attention from Chinese and foreign specialists. Based on the nonlinear state space reconstruction, the nonlinear causality between the two systems is analyzed. The basic idea is that if there is a causal relationship between the systems, it is reasonable to believe that the system *X* contains the evolution information of the system *Y*. By analyzing the correlation between the system *X* and the system *Y* to reconstruct the manifold, the causal relationship between the systems can be found.

Hypothesis X(t) and Y(t) are two time series generated by projecting the system M into a one-dimensional space. For the time series X(t) and Y(t), the embedded dimension of the reconstructed manifold is *em*, the delay time is *tau*, and the reconstructed state space is as follows:

$$X(t) = [X(t), X(t - tau), \cdots, X[t - (em - 1)tau]],$$
(9)

$$\mathbf{Y}(t) = [\mathbf{Y}(t), \mathbf{Y}(t - tau), \cdots, \mathbf{Y}[t - (em - 1)tau]].$$
(10)

According to the state space reconstruction theory, the reconstructed manifolds *X*, *Y*, and system M are differentially isomorphic. Search system *X* for *em* nearest neighbor of the sample  $\{X(i,k)\} = \{X(i,1), X(i,2), \dots, X(i,m)\}$ , then map it to manifold *Y*, the corresponding sample point is  $\{Y(i,k)\}$ , and the estimated value of Y(i) can be calculated as:

$$\hat{Y}(i) = \sum_{k=1}^{em} w_k Y(i,k),$$
(11)

where:

$$u_k = \exp\left(-\frac{\left\|\mathbf{X}(i) - \mathbf{X}(i,k)\right\|}{\left\|\mathbf{X}(i) - \mathbf{X}(i,1)\right\|}\right),\tag{12}$$

$$w_k = u_k / \sum_{k=1}^m u_k \tag{13}$$

where  $\|\cdot\|$  represents the Euclidean distance between the samples. Definite  $\hat{Y}(t)$  is the cross-mapping Y(t) from manifold X to Y, and the correlation coefficient is calculated as follows:

$$r = \frac{\sum_{i=1}^{L} \left( Y(i) - \overline{Y}(i) \right) \left( \hat{Y}(i) - \overline{\hat{Y}}(i) \right)}{\sqrt{\sum_{i=1}^{L} \left( Y(i) - \overline{Y}(i) \right)^2 \sum_{i=1}^{L} \left( \hat{Y}(i) - \overline{\hat{Y}}(i) \right)^2}}.$$
(14)

As the sample volume *L* increases,  $\hat{Y}(t)$  gradually converges to Y(t), and the correlation coefficient finally converges to [0, 1], indicating the causal relationship between system *Y* and system *X*.

A causal analysis was performed on the 11-dimensional variables of the data selected in this paper, which include the hourly average concentration of  $PM_{2.5}$ ,  $PM_{10}$ ,  $SO_2$ ,  $NO_2$ ,  $O_3$ , and CO, and the hourly average values of pressure (P), temperature (T), humidity (H), wind speed (WS), and wind direction (WD), and  $PM_{2.5}$  was used as the target variable. The correlation between the other 10-dimensional

variables and  $PM_{2.5}$  was analyzed by the convergence cross-mapping method. The relevant values obtained are shown in Table 1.

**Table 1.** Correlation results between other variables ( $PM_{10}$ : inhalable particulate matter, T: temperature, P: pressure, H: humidity, WS: wind speed, WD: wind direction) and particles in the atmosphere with a diameter of less than or equal to 2.5 microns ( $PM_{2.5}$ ).

Variables	PM <sub>10</sub>	NO <sub>2</sub>	СО	<b>O</b> <sub>3</sub>	SO <sub>2</sub>	Т	Р	Н	WS	WD
CCM_value	0.780	0.385	0.778	0.392	0.452	0.068	0.011	0.196	0.146	0.203
Sorting	1	5	2	4	3	9	10	7	8	6

Variables are sorted by relevance, and the dimension of the input data is gradually increased. The prediction results based on related variables are shown in Figure 2. The prediction errors are shown in Table 2.



Figure 2. Prediction results based on related variables.

Table 2. Prediction error of different dimensional variables.

Dimension	1	2	3	4	5	6	7	8	9	10
Root mean square error (RMSE)	76.24	77.70	75.64	80.26	76.80	74.96	76.12	78.16	85.70	82.17

According to the predicted results in Figure 2 and Table 2, after CCM causality analysis, when the predicted variable is 6 dimensions, the predicted error reaches the minimum value, that is, 74.96. That is to say, the optimal subset of variables are PM<sub>2.5</sub>, PM<sub>10</sub>, SO<sub>2</sub>, NO<sub>2</sub>, O<sub>3</sub>, and CO. Further calculations are then performed on the selected best subset.

#### 3.3. Phase Space Reconstruction

Chaos refers to a deterministic evolution process that exhibits an irregular process, similar to a random process [33]. Although it is a system described by a deterministic theory, long-term evolutionary behavior is characterized by uncertainty, unrepeatability, and unpredictability. This is a deterministic system in a seemingly random sequence. Most time series in the real world have chaotic characteristics, such as hydrological sequences, meteorological sequences, and air pollutant sequences [34]. Since chaotic sequences are a comprehensive reaction of many physical factors, it contains traces of all variables involved in motion. Because the chaotic sequence must be extended to higher dimensional phase space, the system information contained in the time series can be fully revealed, namely PSR of chaotic sequences [35].

PSR theory is of great significance in chaotic time series analysis, which was proposed by American physicists Packard and Farmer in the 1980s [20]. Since then, Takens, a Dutch mathematician,

has proved that its main purpose is to output through the system [36]. The time series is used to construct a set of coordinate components that characterize the dynamics of the original system, thereby approximating the chaotic attractors of the system. And the evolution process of each component implies all the information about the system. According to the proof of Takens, as long as the delay time  $\tau$  and the embedding dimension *m* are reasonably selected, the final orbit of evolution can be recovered in the embedded dimension space, that is, the chaotic attractor is calculated. That is to say, on the trajectory in the reconstructed high-dimensional space, the reconstructed phase space maintains the same performance as the prime mover system, and the phase space of the delay time reconstruction maintains the geometric structure of the original system and the system has equivalent dynamic characteristics.

For a chaotic time series  $[x_1[t], x_2[t], \dots, x_i[t], \dots, x_d[t]], t = 1, 2, \dots, n$ , where *d* is the sequence dimension, *n* is the sequence length. If you can properly choose the delay time  $\tau = (\tau_1, \tau_2, \dots, \tau_d)$  and embedding dimension  $m = (m_1, m_2, \dots, m_d)$ , the phase space can be reconstructed as followed [37]:

$$X(t) = \begin{bmatrix} x_1(t), & \cdots, & x_1(t - (m_1 - 1)\tau_1), \\ \cdots, & x_d(t), & \cdots, & x_d(t - (m_d - 1)\tau_d) \end{bmatrix}.$$
(15)

Through the PSR process, the reconstructed data  $X \in \mathbb{R}^{n \times \dim}$  can be obtained, where the length *n* of the reconstructed chaotic sequence is the same as the length of the original sequence and dim  $= \sum_{i=1}^{d} m_i \tau_i$  represents the dimension of the reconstructed sequence. In the process of PSR, the delay time  $\tau$  and embedding dimension *m* have a great influence on the quality of the reconstructed phase space, so it is necessary to select the delay time  $\tau$  and embedding dimension *m* appropriately. At present, the more commonly used methods for determining the delay time and embedding dimension are window embedding method, average displacement method, mutual information method, false neighbor method, and Grassberger–Proccacia (G-P) algorithm [38].

#### 4. IPSO-PSR-ESN Hybrid Method

This section will introduce the  $PM_{2.5}$  concentration prediction model based on the hybrid model of IPSO, PSR, and ESN in detail. The basic structure of the model is shown in Figure 3. Aiming at the problems existing in the prediction of  $PM_{2.5}$  chaotic sequence, this paper proposes a hybrid model named IPSO-PSR-ESN. Firstly, the CCM method is used to select the original time series, and the original data is sorted according to the correlation with the  $PM_{2.5}$  data, and the variable subset ( $PM_{2.5}$ ,  $PM_{10}$ , CO, SO<sub>2</sub>, NO<sub>2</sub>, and O<sub>3</sub>) is obtained when the forecasting error is minimized. Secondly, the PSR theory is used to reconstruct the selected optimal subset, and the time series is extended to the higher-dimensional phase space so that the information contained in the dynamic system is fully revealed. Finally, the reconstructed data is predicted using the ESN model. The embedded dimension and delay time in the PSR process and multiple parameters of the ESN model are optimized simultaneously using the improved IPSO algorithm.

In theory, when all variables are projected from the same system, each component will fully expand the attractor. Therefore, in the above basic extension, it will definitely lead to redundancy. Traditional methods for determining the delay time  $\tau$  and embedding dimension *m* consider PSR from the perspective of information theory or entropy, which is separate from subsequent modeling, which can be very time consuming for time series prediction and calculated PSR parameters. Not necessarily suitable for time series prediction. In addition, if the chaotic time series contains noise, the reconstruction parameters may also change. In practical applications, the optimal PSR parameters are closely related to the application scenarios of the time series. Therefore, how to reconstruct multi-variable time series in a targeted manner is of great significance for dynamic system modeling. The heuristic algorithm can be used to optimize the delay time  $\tau$  and embedding dimension *m*, and the data reconstruction process is combined with the modeling process. Finding the optimal value of the delay time  $\tau$  and embedding dimension *m* and the data reconstruction process is combined with the modeling process.

dimension *m* in the predictive modeling process may provide a new way of thinking about revealing dynamic systems.



Figure 3. Flow chart of the IPSO-PSR-ESN model.

The reserve pool is the core part of the ESN, and its parameters and structure have a great impact on the performance of the ESN [39]. According to the characteristics of different prediction objects, designing a suitable reserve pool structure is the primary problem in ESN modeling. The main hyper-parameters in the reserve pool include the size of the reserve pool, the spectral radius of the internal connection matrix, the sparsity, the leakage rate, and the input transform coefficients, etc. [40]. When the ESN predicts data of different characteristics, the hyper-parameter settings are often different. At present, these hyper-parameters do not have fixed values or calculation methods, and need to be specifically analyzed for specific problems. The commonly used selection method is based on empirical selection or trial and error method. These methods have great contingency and will affect the modeling effect. These methods have great contingency and will affect the modeling effect. The heuristic algorithm can also be used to automatically optimize the five reserve pool hyper-parameters of the ESN, and improve the prediction effect of the ESN model by obtaining the global optimal value.

Therefore, the IPSO algorithm is used to optimize the delay time  $\tau$  and embedding dimension m in the PSR process and the hyper-parameters in the reserve pool (the size of the reserve pool, the spectral radius of the internal connection matrix, the sparsity, the leakage rate, and the input transform coefficients). The six selected optimal subsets are predicted, and the 17-dimensional variables are optimized. The objective function is set to the error function of the ESN training process.

Then the IPSO-PSR-ESN hybrid model is established. The flow chart of the hybrid model is shown in Figure 3. The steps of our algorithm are as follows:

Step 1: Select variables by inputting the original data, using the CCM method for causal analysis, and selecting the best subset.

Step 2: Input the selected subset to the hybrid model. Use the IPSO to optimize 17-dimensional decision variables include the parameters of PSR and the hyper-parameters in the ESN model.

Step 3: Initialize the IPSO population. Set the parameters such as population size, maximum iteration number, and initialize the velocity of each particle to 0, as well as the position of each particle by using the good point set theory to make the particles more evenly distributed in the decision space.

Step 4: Substitute the particle position into the objective function and calculate the target vector. Step 5: The initial position of each particle ( $p_{id}$ ) is the initial particle position, and the optimal

value from  $p_{id}$  is the initial global optimality of the particle ( $p_{gd}$ ).

Step 6: While the maximum number of iterations is not reached:

(1) Calculate the inertia weight w and the learning factor  $c_1$  and  $c_2$  in the velocity update formula as [28]:

$$w = wMax - (wMax - wMin) \times (it/MaxIt), \tag{16}$$

$$c_1 = (c_{12} - c_{11}) \times (it/MaxIt) + c_{11}, \tag{17}$$

$$c_2 = (c_{22} - c_{21}) \times (it/MaxIt) + c_{21}, \tag{18}$$

where *it* is the current number of iterations, *MaxIt* is the maximum number of iterations, and *wMax* as well as *wMin* are the maximum and minimum values of the inertia weight *w*, respectively. *w* decreases as the number of iterations increases. In the early stage of the algorithm, the larger the *w* is, the more difficult it is to fall into the local optimum while in the later stage, the smaller the *w* is, the faster the algorithm converges and the more stable the convergence is.  $c_{11} = c_{22} = 2.5$  and  $c_{12} = c_{21} = 0.5$ . The ideal result can be obtained according to  $c_1$  decreasing  $c_2$  increments.

(2) Update the velocity and position of each particle using the speed and position update formula while limiting its value to a certain range.

(3) Adopt a mutation strategy to satisfy the mutation conditions of the particles to avoid local optimum.

(4) Substitute the particle position into the objective function and calculate the target vector.

(5) Evaluate particles of the new generation, update the  $p_{id}$  of each particle and the  $p_{gd}$  of the population.

(6) Create a logistics chaotic sequence based  $p_{gd}$ , and then randomly replacing the position of one particle in the population using the point with the smallest target value in the chaotic sequence.

End while

Step 7: After the iteration is completed, the final  $p_{gd}$  is the required parameters of PSR and the hyper-parameters of the ESN model.

Step 8: The process of PSR. Map selected subsets of variables to high dimensional variable space based on optimized delay time  $\tau$  and embedded dimension *m*.

Step 9: Divide the reconstructed sequence into a training data set and a test data set.

Step 10: Substitute the optimized hyper-parameter into the ESN and train the ESN model through the training set.

Step 11: Substitute the test set into the trained model to obtain the index of test accuracy.

#### 5. Experimental Results

All the experiments in this paper were completed in the experimental simulation environment of a Windows 7 operating system, equipped with 3.70 GHz, Intel(R) core i3-4170m CPU, and 12 GB RAM memory. All simulation experiments were carried out with Matlab R2018B (The MathWorks, Inc., Natick, MA, USA). In each simulation experiment, the program parameters were identical, and each simulation experiment group was repeated 10 times.

#### 5.1. Evaluation Indicator

In recent years, many error evaluation indicators have been widely used in related literature, but there is no recognized general standard method. Therefore, this paper uses multiple common error criteria to evaluate the effectiveness of the proposed hybrid model. In this paper, four common prediction criteria (RMSE, MAE, SMAPE, and CR) are used to evaluate the prediction accuracy:

(1) Root mean square error (RMSE), which represents the degree of dispersion between the predicted series and the actual series.

RMSE = 
$$\sqrt{\frac{\sum\limits_{t=1}^{p} [\hat{y}(t) - y(t)]^2}{(p-1)}}$$
 (19)

(2) Mean absolute error (MAE), which is a quantity used to measure how close forecasts or predictions are to the eventual outcomes in statistics.

$$MAE = \frac{1}{p} \sum_{t=1}^{p} \left| \hat{y}|t| - y|t| \right|$$
(20)

(3) Symmetric mean absolute percentage error (SMAPE), which measures the size of the error in percentage terms, and reflects the relative difference between the predicted series and the actual series.

SMAPE = 
$$\frac{1}{p} \sum_{t=1}^{p} \frac{|y(t) - \hat{y}(t)|}{\frac{1}{2}|y(t) + \hat{y}(t)|} \times 100\%$$
 (21)

(4) Correlation coefficient (CR), which is a statistical criterion to reflect the degree of close correlation between variables.

$$CR = \frac{cov(Y, Y)}{\sqrt{D(Y)}\sqrt{D(\overline{Y})}}$$
(22)

where  $\hat{y}(t)$ , y(t), and  $\overline{y}(t)$  are the predicted value, observation, and the average of observation of the time series, respectively, *p* represents the number of samples,  $cov(\cdot)$  represents the covariance,

 $D(\cdot)$  represents the variance, and  $\hat{Y}$  as well as Y represent the predicted sequence and the observed sequence, respectively.

In the above evaluation indicators, the smaller the evaluation index values of RMSE, MAE, and SMAPE, the better the prediction effect of the model. CR = 1 indicates linear correlation, CR = 0 indicates no correlation. When  $CR \in (0, 1)$ , it indicates that there is a correlation between the two sequences, and the larger the *CR*, the stronger the linear correlation.

#### 5.2. Models Comparison

To enlighten possible causal relationships, the relations between air pollutants and meteorological variables are assessed here by the CCM method. Thus, we have obtained the relevant value data and optimal variable subset among air pollutants which are showed in Tables 1 and 2 and Figure 2. The models that were proposed in this study were used to test the optimal variable subset to evaluate their performance in forecasting the PM<sub>2.5</sub> concentration. According to the CCM causal analysis, when the predictor is six-dimensional, the prediction error reaches the minimum value, and the optimal subset of variables can be determined. Then the PSR process is performed on the selected optimal subset. In order to improve the overall forecast accuracy, this paper firstly combines the PSR process with the forecasting model as a whole to avoid the problems of parameter selection in isolation. And then according to the different conditions such as the forecasting step size, the phase space reconstruct parameters of the best subset is dynamically selected by the IPSO algorithm. It is obvious that the data series is mapping into high-dimensional spaces. The calculated delay time  $\tau$  and embedding dimension *m* of the Beijing data in the one-step prediction and the ten-step prediction are shown in Table 3. It can be seen from the table that the delay time  $\tau = [3, 2, 2, 2, 1, 1]$  and the embedding dimension m = [2, 1, 2, 2, 10, 3] of the optimal subset obtained by optimization in the one-step prediction as well as the delay time  $\tau = [5, 6, 5, 6, 3, 7]$  and the embedding dimension m = [8, 8, 7, 8, 10, 7] the optimal subset obtained by optimization in the ten-step prediction.

Variables		PM <sub>2.5</sub>	PM <sub>10</sub>	SO <sub>2</sub>	NO <sub>2</sub>	O <sub>3</sub>	СО
(t + 1)	Embedding	3	2	2	2	1	1
	Time delay	2	1	2	2	10	3
(t + 10)	Embedding	5	6	5	6	3	7
	Time delay	8	8	7	8	10	7

Table 3. Phase space reconstruction parameters of Beijing dataset.

After reconstruction of the data optimal time series subset by the optimized PSR process, the system information contained in the time series is adequately revealed, then it uses the optimized forecasting model to forecast the  $PM_{2.5}$  concentration sequence. In IPSO algorithm, the parameter settings including iterations (MAXTI), population size (POP), and dimension of decision space (DIM) are listed as follows: MAXIT = 100, POP = 40, and DIM = 17. In this paper, the IPSO algorithm uses the above parameter settings in each simulation experiments to ensure the validity and fairness of all experiments.

In order to verify the effectiveness of the proposed algorithm, the original model (ESN) [23], single-hidden layer feedforward network (SLFN) [41], extreme learning machine (ELM) [42], back propagation neural network (BPNN) [43], the least squares support vector machine (LSSVM) [44], and the long and short time memory (LSTM) [45] are selected as the benchmark models. The first 75% of the data is selected as the training set, and the last 25% of the data is the simulated test set.

In order to evaluate the effectiveness of the proposed method comprehensively and objectively, single-step and multi-step (10-step) forecasting experiments of  $PM_{2.5}$  concentration were carried out respectively. The one-step forecasting results of Beijing  $PM_{2.5}$  data are shown in Table 4, and the ten-step forecasting results are shown in Table 5. For example in Table 4, the RMSE, MAE, SMAPE, and CR of  $PM_{2.5}$  concentration of the proposed IPSO-PSR-ESN model for one-step ahead forecasting are

9.4961, 5.7938, 0.1029, and 0.9945, respectively. It is concluded that the two tables verify the superior performance of the proposed model compared with other state-of-the-art models.

Method	RMSE	MAE	SMAPE	CR
ESN	13.4753	9.8373	0.1396	0.9893
SLFN	16.2844	10.1191	0.1276	0.9839
ELM	18.6669	10.0054	0.1285	0.9812
BP	15.9821	11.2058	0.1481	0.9798
LSSVM	13.0054	10.2599	0.1358	0.9825
LSTM	33.5302	23.8995	0.34083	0.93754
IPSO-PSR-ESN	9.4961	5.7938	0.1029	0.9946

**Table 4.** Prediction errors of Beijing PM<sub>2.5</sub> time series based on different methods (one-step ahead).

Table 5. Prediction errors of Beijing PM<sub>2.5</sub> time series based on different methods (10-step ahead).

Method	RMSE	MAE	SMAPE	CR
ESN	74.9663	50.8694	0.6187	0.6496
SLFN	84.3316	54.9802	0.6190	0.5050
ELM	85.3812	55.4798	0.6587	0.5125
BP	82.5792	52.9125	0.6227	0.5589
LSSVM	88.5819	57.7009	0.6491	0.4518
LSTM	69.6015	50.5132	0.6130	0.72956
IPSO-PSR-ESN	62.0824	42.2215	0.51233	0.74062

As can be seen from the Tables 4 and 5, the forecast accuracies including RMSE, MAE, SMAPE, and CR are illustrated in tables where the smallest or largest value in each row is marked in boldface. From the results displayed in the Figure 7 and Tables 4 and 5, it is obvious that the developed IPSO-PSR-ESN model owns superior forecasting ability compared to other comparative models in this paper. In addition, the values of RMSE, MAE, and SMAPE of the proposed model are all smallest among all the comparison models, and the value of CR of the proposed model is largest among all the models, which further confirms that the proposed model has the best forecasting performance. The smallest error evaluation index (RMSE, MAE, SMAPE) of the proposed hybrid model shows that the forecasting data and the actual data are pretty close and the effect of model fitting is the best. From Tables 4 and 5, the correlation coefficient of  $PM_{2.5}$  concentration reaches 0.9946 and 0.74062 for one-step forecasting and ten-step forecasting, respectively, which is the maximum value in the comparative models. Therefore, it can be concluded that the proposed model is suitable for the PM<sub>2.5</sub> concentration forecasting. The proposed hybrid model has achieved good forecasting results in the one step and the ten-step prediction. The single-step forecasting and error curve of the Beijing  $PM_{2.5}$  series generated by the IPSO-PSR-ESN model are shown in Figure 4. The five hyper-parameters of the ESN are shown in Figure 5. At the same time, Figure 6 shows the prediction and error curve for the ten-step prediction.

A graph showing the prediction error of different models as a function of the number of predicted steps is given in Figure 7. It can be seen from the error curve and various precision indicators that the LSSVM model can obtain good prediction results when the number of prediction steps is small, but when the number of prediction steps increases, the prediction effect of the LSSVM model will drastically deteriorate. The predictive effect of the step prediction is the worst of the selected models. LSTM is a deep learning model that has been used in time series prediction in recent years. The LSTM model shows good performance when the number of prediction steps is small. The prediction effect of the model is not ideal when the number of prediction steps is small. The prediction effect of other comparison models is between the two models. By comparing the prediction effects of the IPSO-PSR-ESN model and other different models, the prediction accuracy of the hybrid model proposed in this paper has achieved the best prediction results in the 1~10-step prediction, which proves that the optimization part of the IPSO algorithm can effectively improve the prediction accuracy, which avoids the randomness of the parameter settings in the model, and combines the PSR process

with the ESN model to dynamically select the PSR parameters and the ESN model hyper-parameters according to the specific prediction requirements. In addition, the proposed IPSO-PSR-ESN model makes the optimal forecast accuracies in the different forecasting step size. All the experiment results can demonstrate the superiority of the proposed hybrid forecasting model. For these seven forecasting models, the best forecasting results are from the IPSO-PSR-ESN model by comparing the performance validations. Through integrating the PSR process and ESN model, and optimizing the parameters of reconstruction process and the hyper-parameters in the ESN model by IPSO algorithm, the model can obtain accurate forecast results. Thus, it is considerable to use the IPSO-PSR-ESN model to forecast atmospheric pollutants in environmental governance and prevention, and provide some basis for the government regulation.



Figure 4. One-step ahead forecasting of Beijing PM<sub>2.5</sub> in 2016.



Figure 5. ESN parameters curve with iteration (one-step ahead).



Figure 6. Ten-step ahead forecasting of Beijing PM<sub>2.5</sub> in 2016.



Figure 7. Curve charts of forecasting errors of different methods with variation of steps.

# 6. Conclusions

 $PM_{2.5}$  is the main source of air pollution, and forecasting the  $PM_{2.5}$  concentration is important for sustainable development and public health. In order to improve the forecast accuracy, this paper developed a novel hybrid model, i.e., IPSO-PSR-ESN model, for  $PM_{2.5}$  concentration forecasting. This paper collected the real-world time series of Beijing in 2016, and filtered the original dataset through CCM to obtain the optimal variable subset. Then, the IPSO algorithm was used to optimize the PSR process and ESN parameters. Finally, based on the trained hybrid model, PM2.5 concentration prediction was achieved. Based on the experimental results and comparative analysis, the following four conclusions can be obtained: (1) the proposed hybrid model outperforms other comparison models in the  $PM_{2.5}$  concentration forecasting; (2) the proposed error correction model can significantly improve the prediction ability of initial forecasting model (ESN); (3) the proposed IPSO algorithm has a positive effect on the forecasting performance of ESN model. Therefore, it can be concluded that (4) the proposed hybrid model can precisely forecast the  $PM_{2.5}$  concentration. Accurate  $PM_{2.5}$  forecasting enables government to take actions that reduce the severity of episodes of high levels of  $PM_{2.5}$ , like promoting the transformation and upgrading of high-pollution enterprises, and encouraging citizens to take public transports instead of driving. Moreover, predictions also enable individuals to take protective actions that limit their own exposure to high levels of  $PM_{2.5}$ , such as reducing outdoor activities and staying indoors as much as possible. Although the hybrid model performs well in  $PM_{2.5}$  forecasting, it fails to consider the potential factors affecting air quality in extreme conditions, such as radon emissions and other pollutants, in which the situation is very complicated and needs to be solved in future studies. The accuracy and prediction horizon of  $PM_{2.5}$ concentration series makes its forecasting become a very difficult task. We hope to apply some of the latest methods in the field of artificial intelligence, such as deep learning, multi-task learning, and so on, analyze more factors that affect  $PM_{2.5}$  concentration, for instance the radon radiations. In addition, we expect to extend the forecast time step in future research to achieve medium- and long-term forecasts. Through the interdisciplinary integration, it provides the basis for the government's air pollution control and the sustainable development of society.

**Author Contributions:** Conceptualization, X.X. and W.R.; methodology, X.X. and W.R.; data curation, X.X. and W.R.; writing—original draft preparation, X.X. and W.R.; writing—review and editing, X.X. and W.R.

Funding: This research was funded by the National Natural Science Foundation of China (61773087).

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

# References

- Huang, W.; Wang, H.; Wei, Y. Endogenous or Exogenous? Examining Trans-Boundary Air Pollution by Using the Air Quality Index (AQI): A Case Study of 30 Provinces and Autonomous Regions in China. *Sustainability* 2018, 10, 4220. [CrossRef]
- 2. Yuan, G.; Yang, W. Evaluating China's Air Pollution Control Policy with Extended AQI Indicator System: Example of the Beijing-Tianjin-Hebei Region. *Sustainability* **2019**, *11*, 939. [CrossRef]
- 3. Yang, Z.; Wang, J. A new air quality monitoring and early warning system: Air quality assessment and air pollutant concentration prediction. *Environ. Res.* **2017**, *158*, 105–117. [CrossRef]
- 4. Ministry of Environmental Protection of the People's Republic of China. *Technical Regulation on Ambient Air Quality Index (on Trial): HJ 633-2012;* China Environmental Science Press: Beijing, China, 2012.
- Xu, Z.; Xia, X.; Liu, X.; Qian, Z. Combining DMSP/OLS Nighttime Light with Echo State Network for Prediction of Daily PM<sub>2.5</sub> Average Concentrations in Shanghai, China. *Atmosphere* 2015, *6*, 1507–1520. [CrossRef]
- 6. Perez, P.; Gramsch, E. Forecasting hourly PM<sub>2.5</sub> in Santiago de Chile with emphasis on night episodes. *Atmos. Environ.* **2016**, *124*, 22–27. [CrossRef]
- Saide, P.E.; Carmichael, G.R.; Spak, S.N.; Gallardo, L.; Osses, A.E.; Mena-Carrasco, M.A.; Pagowski, M. Forecasting urban PM<sub>10</sub> and PM<sub>2.5</sub> pollution episodes in very stable nocturnal conditions and complex terrain using WRF–Chem CO tracer model. *Atmos. Environ.* 2011, 45, 2769–2780. [CrossRef]
- 8. Neal, L.; Agnew, P.; Moseley, S.; Ordonez, C.; Savage, N.; Tilbee, M. Application of a statistical post-processing technique to a gridded, operational, air quality forecast. *Atmos. Environ.* **2014**, *98*, 385–393. [CrossRef]
- 9. Fernando, H.; Mammarella, M.; Grandoni, G.; Fedele, P.; Di Marco, R.; Dimitrova, R.; Hyde, P. Forecasting PM<sub>10</sub> in metropolitan areas: Efficacy of neural networks. *Environ. Pollut.* **2012**, *163*, 62–67. [CrossRef]
- 10. Li, R.; Jin, Y. The early-warning system based on hybrid optimization algorithm and fuzzy synthetic evaluation model. *Inf. Sci.* **2018**, *435*, 296–319. [CrossRef]
- Zhu, S.; Qiu, X.; Yin, Y.; Fang, M.; Liu, X.; Zhao, X.; Shi, Y. Two-step-hybrid model based on data preprocessing and intelligent optimization algorithms (CS and GWO) for NO<sub>2</sub> and SO<sub>2</sub> forecasting. *Atmos. Pollut.* 2019. [CrossRef]
- 12. Lu, W.-Z.; Wang, W.-J.; Wang, X.-K.; Yan, S.-H.; Lam, J.C.; Lu, W.-Z.J. Potential assessment of a neural network model with PCA/RBF approach for forecasting pollutant trends in Mong Kok urban air, Hong Kong. *Environ. Res.* **2004**, *96*, 79–87. [CrossRef]

- 13. Zhang, J.; Zhan, Z.-H.; Lin, Y.; Chen, N.; Gong, Y.-J.; Zhong, J.-H.; Chung, H.S.H.; Li, Y.; Shi, Y.-H. Evolutionary Computation Meets Machine Learning: A Survey. *IEEE Comput. Intell. Mag.* **2011**, *6*, 68–75. [CrossRef]
- 14. Xu, Y.; Yang, W.; Wang, J. Air quality early-warning system for cities in China. *Atmos. Environ.* **2017**, 148, 239–257. [CrossRef]
- 15. Niu, M.; Wang, Y.; Sun, S.; Li, Y. A novel hybrid decomposition-and-ensemble model based on CEEMD and GWO for short-term PM<sub>2.5</sub> concentration forecasting. *Atmos. Environ.* **2016**, 134, 168–180. [CrossRef]
- 16. Sun, W.; Sun, J. Daily PM<sub>2.5</sub> concentration prediction based on principal component analysis and LSSVM optimized by cuckoo search algorithm. *J. Environ. Manag.* **2017**, *188*, 144–152. [CrossRef] [PubMed]
- 17. Li, X.; Luo, A.; Li, J.; Li, Y. Air Pollutant Concentration Forecast Based on Support Vector Regression and Quantum-Behaved Particle Swarm Optimization. *Environ. Model. Assess.* **2018**, *24*, 205–222. [CrossRef]
- Kennedy, J.; Eberhart, R.C. Particle swarm optimization. In Proceedings of the IEEE International Conference on Neural Networks, Perth, Australia, 27 November–1 December 1995; IEEE: Piscataway, NJ, USA; pp. 1942–1948.
- 19. Sugihara, G.; May, R.; Ye, H.; Hsieh, C.-H.; Deyle, E.; Fogarty, M.; Munch, S. Detecting Causality in Complex Ecosystems. *Science* **2012**, *338*, 496–500. [CrossRef]
- 20. Packard, N.H.; Crutchfield, J.P.; Farmer, J.D.; Shaw, R.S. Geometry from a time series. *Phys. Rev. Lett.* **1980**, 45, 712. [CrossRef]
- 21. Funahashi, K.-I.; Nakamura, Y. Approximation of dynamical systems by continuous time recurrent neural networks. *Neural Netw.* **1993**, *6*, 801–806. [CrossRef]
- 22. Jaeger, H. The "echo state" Approach to Analysing and Training Recurrent Neural Networks-with an Erratum Note. Bonn, Germany: German National Research Center for Information Technology GMD Technical Report. Available online: https://www.researchgate.net/profile/Herbert\_Jaeger3/publication/215385037\_The\_echo\_state\_approach\_to\_analysing\_and\_training\_recurrent\_neural\_networks-with\_an\_erratum\_note\T1\textquoteright/links/566a003508ae62b05f027be3/The-echo-state-approach-to-analysing\_and-training-recurrent-neural-networks-with-an-erratum-note.pdf (accessed on 1 May 2019).
- 23. Jaeger, H. Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication. *Science* 2004, 304, 78–80. [CrossRef] [PubMed]
- 24. Wang, J.; Wu, C.; Niu, T. A Novel System for Wind Speed Forecasting Based on Multi-Objective Optimization and Echo State Network. *Sustainability* **2019**, *11*, 526. [CrossRef]
- 25. Dai, Q.; Liu, J.; Wei, Q. Optimal Photovoltaic/Battery Energy Storage/Electric Vehicle Charging Station Design Based on Multi-Agent Particle Swarm Optimization Algorithm. *Sustainability* **2019**, *11*, 1973. [CrossRef]
- 26. Hua, L.; Wang, Y. *The Application of Number Theory in the Approximate Analysis*, 1st ed.; Science Press: Beijing, China, 1978; pp. 83–85.
- 27. Wu, G.C.; Baleanu, D. Discrete fractional logistic map and its chaos. *Nonlinear Dynam.* **2014**, *75*, 283–287. [CrossRef]
- 28. Ratnaweera, A.; Halgamuge, S.K.; Watson, H.C. Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients. *IEEE Trans. Evol. Comput.* **2004**, *8*, 240–255. [CrossRef]
- 29. He, K.; Yang, F.; Ma, Y.; Zhang, Q.; Yao, X.; Chan, C.K.; Cadle, S.; Chan, T.; Mulawa, P. The characteristics of PM<sub>2.5</sub> in Beijing, China. *Atmos. Environ.* **2001**, *35*, 4959–4970. [CrossRef]
- 30. Hu, L.; Liu, J.; He, Z. Self-Adaptive Revised Land Use Regression Models for Estimating PM<sub>2.5</sub> Concentrations in Beijing, China. *Sustainability* **2016**, *8*, 786. [CrossRef]
- Wang, F.; Peng, Y.; Jiang, C. Influence of Road Patterns on PM<sub>2.5</sub> Concentrations and the Available Solutions: The Case of Beijing City, China. *Sustainability* 2017, *9*, 217. [CrossRef]
- 32. Radon and Cancer. Available online: https://www.cancer.org/cancer/cancer-causes/radiation-exposure/ radon.html (accessed on 23 May 2019).
- 33. Farmer, J.D.; Sidorowich, J.J. Predicting chaotic time series. *Phys. Rev. Lett.* **1987**, *59*, 845–848. [CrossRef] [PubMed]
- Han, M.; Feng, S.; Chen, C.L.P.; Xu, M.; Qiu, T. Structured Manifold Broad Learning System: A manifold perspective for large-scale chaotic time series analysis and prediction. *IEEE Trans. Knowl. Data Eng.* 2019, accepted. [CrossRef]
- 35. Erem, B.; Orellana, R.M.; Hyde, D.E.; Peters, J.M.; Duffy, F.H.; Stovicek, P.; Warfield, S.K.; MacLeod, R.S.; Tadmor, G.; Brooks, D.H. Extensions to a manifold learning framework for time-series analysis on dynamic manifolds in bioelectric signals. *Physcial E* **2016**, *93*, 042218. [CrossRef]

- 36. Takens, F. *Detecting Strange Attractors in Turbulence;* Springer Nature: Berlin, Germany, 1981; Volume 898, pp. 366–381.
- 37. Kim, H.S.; Eykholt, R.; Salas, J.D. Nonlinear dynamics, delay times, and embedding windows. *Phys. Nonlinear Phenom.* **1999**, *127*, 48–60. [CrossRef]
- 38. Lv, J.; Lu, J.; Chen, S. *Chaotic Time Series Analysis and its Application*, 1st ed.; Wuhan University Press: Wuhan, China, 2002; pp. 57–71.
- 39. Bianchi, F.M.; Scardapane, S.; Uncini, A.; Rizzi, A.; Sadeghian, A. Prediction of telephone calls load using Echo State Network with exogenous variables. *Neural Netw.* **2015**, *71*, 204–213. [CrossRef]
- 40. Fiers, M.A.A.; Van Vaerenbergh, T.; Wyffels, F.; Verstraeten, D.; Schrauwen, B.; Dambre, J.; Bienstman, P. Nanophotonic Reservoir Computing with Photonic Crystal Cavities to Generate Periodic Patterns. *IEEE Trans. Neural Netw. Learn. Syst.* **2014**, *25*, 344–355. [CrossRef] [PubMed]
- 41. Sanger, T.D. Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural Netw.* **1989**, *2*, 459–473. [CrossRef]
- 42. Huang, G.-B.; Zhou, H.; Ding, X.; Zhang, R. Extreme Learning Machine for Regression and Multiclass Classification. *IEEE Trans. Syst. Man Cybern. Part B* (*Cybern.*) **2012**, *42*, 513–529. [CrossRef]
- 43. Hecht-Nielsen, R. *Neural Networks for Perception*, 1st ed.; Academic Press: Pittsburgh, PA, USA, 1992; pp. 65–93.
- 44. Adankon, M.M.; Cheriet, M. Model selection for the LS-SVM. Application to handwriting recognition. *Pattern Recognit.* **2009**, *42*, 3264–3270. [CrossRef]
- 45. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).