

Article

A New Perspective on Formation of Haze-Fog: The Fuzzy Cognitive Map and Its Approaches to Data Mining

Zhen Peng ^{1,*} and Lifeng Wu ²¹ Information Management Department, Beijing Institute of Petrochemical Technology, Beijing 102617, China² College of Information Engineering, Capital Normal University, Beijing 100048, China; wooleef@gmail.com

* Correspondence: zhenpeng@bipt.edu.cn; Tel.: +86-10-6022-8013

Academic Editor: Benjamin T. Hazen

Received: 9 January 2017; Accepted: 23 February 2017; Published: 27 February 2017

Abstract: Haze-fog has seriously hindered the sustainable development of the ecological environment and caused great harm to the physical and mental health of residents in China. Therefore, it is important to probe the formation of haze-fog for its early warning and prevention. The formation of haze-fog is, in fact, a fuzzy nonlinear process. The formation of haze-fog is such a complex process that it is difficult to simulate its dynamic evolution using traditional methods, mainly because of the lack of their consideration of the nonlinear relationships. It is, therefore, essential to explore new perspectives on the formation of haze-fog. In this work, previous research on haze-fog formation is summarized first. Second, a new perspective is proposed on the application of fuzzy cognitive map to the formation of haze-fog. Third, a data mining method based on the genetic algorithm is used to discover the causality values of a fuzzy cognitive map (FCM) for haze-fog formation. Finally, simulation results are obtained through an experiment using the fuzzy cognitive map and its data mining method for the formation of haze-fog. The validity of this approach is determined by definition of a simple rule and the Kappa values. Thus, this research not only provides a new idea using FCM modeling the formation of haze-fog, but also uses an effective method of FCM for solving the nonlinear dynamics of the haze-fog formation.

Keywords: formation of haze-fog; pollutants; meteorological conditions; fuzzy cognitive map; data mining; nonlinear dynamics

1. Introduction

Presently, the formation of haze-fog in China, especially in North China, exhibits a complexity that reflects two areas of concern. One is the existence of pollutants. Air pollution is changing from, originally, the single-type to multiple variations. Secondary pollution [1,2], formed by a variety of pollutants reacting with each other, is becoming increasingly prominent. The second area of concern is related to meteorological conditions, whose changes may impact the formation of haze-fog in different ways. First, global warming makes the local “extreme weather” more variable, and disturbance due to El Niño increases the likelihood of haze-fog formation [3]. Second, the Three-North Shelter Forest Project and large scale wind power stations in Southern Inner Mongolia and Zhangbei may make a certain influence on the wind power of North China [4]. Third, the South-to-North Water Diversion Project has a few effects over the humidity of the area [5]. In addition, there are also other possible factors that may affect haze-fog like atmospheric circulation, solar radiation, traffic, industrial emissions, fires, etc. [6–10].

In China, there is high false alarm rate in haze early warning systems. For example, Beijing and other areas in North China experienced a haze-fog for a long duration lasting for up to 110 h on 27 November 2015, as shown as Figure 1. However, at that time, the heavy air pollution warning had not been started in Beijing. As an example of false alarms, the heavy “red warning” of haze-fog was given by the relevant agency for 12 a.m., 23 December 2015, but in reality, serious haze-fog occurred south of Beijing at 5 p.m. on that day. Such issues with the prediction systems affect the prevention decisions about haze-fog, and bring a significant impact on the traffic of the city, and the physical and mental health of residents. The main reason for such inaccuracies in prediction is due to the lack of research on complex relationships among the different factors in the formation of haze-fog in traditional methods. Therefore, it is difficult to simulate their nonlinear dynamic evolution using these methods. Thus, there exist inconsistencies between the emergency early warning of haze-fog and its actual occurrence.



Figure 1. Haze-fog in Beijing on 27 November 2015.

The formation of haze-fog is based on different factors [11], which mainly involve $PM_{2.5}$ -based fine pollution particles, and prevailing weather conditions. This formation can be seen as a complex system. There are complex, iterative, and cognitive relationships among pollutants, weather conditions, pollution particles and haze-fog. For example, $PM_{2.5}$ mainly involves two kinds of pollutants. One are primary particles emitted in the form of a solid state. The other are secondary particles produced by chemical reactions of the precursor pollutants, for example, NO_x , SO_2 , etc. Therefore, the formation of haze-fog can be modeled in terms of the fuzzy cognitive complex system.

FCM (fuzzy cognitive map) [12,13], as a knowledge representation of causal relationship (explicit and implicit) between the concepts and a method for system modeling, has been proposed by Kosko in 1986 and has extended ternary logic relationship in the fuzzy relationship in $[-1, 1]$. FCM is composed of the fuzzy information processing ability of fuzzy logic, the causal propagation method of cognitive map and the adaptive dynamic characteristics of neural network. FCM can also be used for the sensitivity analysis of factors, and the explanation of causal relationship, etc. This also supports a feedback mechanism and it is not like traditional methods that only draw a static conclusion. The final result may be a fixed point, limit cycle, or chaotic attractor, and FCM has been expanded in better expression of knowledge and reasoning ability than traditional methods. Thus, FCM can be used to express the complex cognitive model for haze-fog formation.

Moreover, it is well established that the evolution of weather system is a fuzzy and nonlinear dynamic process [14,15]. Haze-fog is a kind of disaster weather, whose formation is a dynamic, nonlinear evolution process. In fact, the formation of haze-fog contains a wealth of data which can be exploited for its understanding. The monitoring data (such as the emission concentration of various pollutants, the meteorological data, etc.) provide massive and valuable data samples for research on the formation of haze-fog. The FCM model can be automatically obtained by learning the data [16–19],

and it has been applied in fault diagnosis, medical prediction, pollution management, and intelligent analysis of social phenomena and so on [20–24]. Thus, training of nonlinear and dynamic haze-fog formation time series data to mine the fuzzy cognitive association values becomes an immediate scientific issue to be solved for the formation of haze-fog.

2. Previous Research on the Formation of Haze-Fog

The formation of haze-fog has been mainly studied from the following three perspectives: physical chemistry mechanism, statistical analysis, and data mining.

2.1. Physical Chemistry Mechanism in the Formation of Haze-Fog

A diffusion model is used to simulate the physical chemistry mechanism in the formation of haze-fog. This model is a numerical method, which can quantitatively simulate the emission, migration, diffusion, and chemical reaction of pollution with time and space. As the core of air quality prediction model, the diffusion model has gone through three generations [25,26].

The first generation, mainly, includes the Gauss model and the Lagrange trajectory model. For example, California Puff (CALPUFF) is an atmospheric quality evaluation and prediction system for complex terrain [27]. The Gauss diffusion model, as the core of CALPUFF, uses a large number of discrete smoke clusters to represent continuous plume dispersion, and a “snapshot” method to evaluate the concentrated contribution of a single smoke group to a receptor point.

In order to overcome the uniformity of the single grid in the first generation of models, the simulated area is divided into multiple grids, and each grid has independent emission data and meteorology data in the second generation. Based on the Eulerian mesh model [28,29], it introduces a more complicated meteorological model and parameters, and detailed nonlinear chemical reaction mechanism, where chemical transport models (CTM) are used to calculate the concentration of pollutants or air pollution based on the meteorological field and the source list. The transmission process follows the principle of mass conservation of pollutants.

In order to take into account all of these atmospheric variables into the model, USEPA (United States Environmental Protection Agency) proposed the third generation called Models-3 (Third-Generation Air Quality Modeling System)/CMAQ (Community Multi-scale Air Quality). This is a kind of atmospheric chemical transport model based on “one atmosphere” [25,26]. Each chemical mechanism in Models-3/CMAQ contains a limited number of chemical reactions. In addition, it is important that the aerosol module is involved in CCTM for the formation of haze-fog because the increase in aerosol concentration reduces the visibility and, thus, forms haze-fog [2]. Aerosol particles are the product of interactions between pollutants and meteorology factors and a general name for the solid and liquid particles suspended in the air, whose main component is PM_{2.5}.

In short, the physical chemistry mechanism applies the physical and chemical reaction of pollutants under meteorological conditions in the diffusion for prediction of the air quality in time-space. The methods exhibit superiority in simulating the transition and distribution of pollutants in time-space. However, as many complex nonlinear physical and chemical mechanisms are not taken into account, it is difficult to simulate complex physical and chemical reactions under the influence of complex meteorological field for the physical chemistry mechanism. As a result, it sometimes fails to predict the formation of haze-fog.

2.2. Statistical Analysis of the Formation of Haze-Fog

Statistical analysis is used to obtain the factors that influence the formation of haze-fog from the data of haze-fog monitored by satellite or ground monitoring equipment. This is one of the most common methods used to analyze the formation of haze-fog. There are three viewpoints related to this approach.

The first view is of meteorological conditions. Ding et al. [30] show that relative humidity is an important influencing factor using the comprehensive judgment method in distinguishing fog and

haze happening in China. Fu et al. [31] obtained the influence of wind velocity and relative humidity on haze-fog using mean and frequency calculation methods in the North China Plain. Guo et al. [32] indicate that the long-lasting fog and haze event often occur in a high pressure weather system and calm wind conditions. Zhang et al. [33] show that the dynamic effect on the haze-fog evolution is almost the same as the thermodynamic effect based on meteorological factors using a multiple linear regression model. In addition, Zhang et al. [11], Yang et al. [34], and Yang et al. [35] point out that PLAM (Parameter Linking Air-quality and Meteorology) is a “pollution weather condition” index, such as air pressure, temperature, wind power, relative humidity, stability, precipitation, etc., which can quantitatively reflect the degree of static and stable weather.

The second view is of pollutants. Jansen et al. [36] obtained that both gaseous NO_2 and SO_2 are main factors resulting in the reduction of visibility according to hourly concentrations of particulate sodium and the mixing ratios by the online MARGA ADI 2080 analyzer. Shen et al. [37] obtained that nitrate and organic compounds dominate the aerosol component during the severe haze-fog episodes and are related to secondary aerosol formation and air mass origin by statistical analysis.

The third view is of both pollution and meteorological conditions. Zhang et al. [11] analyzed the factors of haze-fog origins in China, especially in North China, and shows that the existence of heavy haze-fog is related to the high concentration of aerosol particles and continuous haze-fog is related to the prevailing calm weather in the country at present. Zhang et al. [38] focused on the changes of PM_{10} in the atmosphere and the influence of meteorological conditions on the weather with heavy haze-fog at a regional background station in the Yangtze River Delta area of China. In the study performed in Chung et al. [39], due to the increase in both water supply and emission of air pollution, the typical pattern of historical mist and haze in London is observed commonly in Korea today. Sun et al. [40] obtained the influence coefficient of $\text{PM}_{2.5}$ and meteorological conditions on haze-fog under different relative humidity and $\text{PM}_{2.5}$ concentration levels by a multiple linear regression equation of visibility in Beijing.

The main contributions of the methods are the determination of the possible factors affecting the formation of haze-fog. Haze-fog is the result of a large number of pollution particles suspended in the air and accompanying meteorological conditions. The quantitative relationships among the factors responsible for the formation of haze-fog have seldom been constructed, besides the research [40] has mainly focused on the influence coefficients in the formation of haze-fog. Moreover, the statistical methods [30–40] are more like *ex post*, linear, and static computing, rather than nonlinear models for the formation of haze-fog. Ultimately, they cannot simulate the sequential evolution of the formation of haze-fog.

2.3. Data Mining Methods for the Formation of Haze-Fog

With the aim to simulate nonlinear relationships, data mining is applied in the forecast of air quality. First, it is used to predict the concentration of pollutants [41,42], using techniques such as neural networks, SVM (support vector machine), GM (1,1) Model (Grey model), comprehensive forecast model, GA_ANN (genetic algorithm and artificial neural networks), etc.

It is, then, extended to apply in the formation of haze-fog. Meng et al. [43] constructed an MRM (multiple regression model) for calculating fog and haze intensity by using a logistic function as a haze-fog intensity function. The influencing parameters in the model are mined by fitting the measured data of haze-fog to the model in order to forecast the haze-fog intensity.

At present, the existing data mining method [43] in predicting haze-fog only mines the influencing weights of meteorological and pollutant indices on haze-fog from the available data because of the limitation of the multiple regression model. They cannot get the influencing weights among these meteorological factors or the weight among these pollutant indexes. In addition, they also do not exploit the dynamics of data for the formation of haze-fog. Therefore, this aspect needs to be further explored using complex nonlinear dynamic relationships among factors responsible for the formation of haze-fog.

2.4. Proposed Problems

The research on the formation of haze-fog is still in the early stage, especially in China. It is the lack of focus on the complicated nonlinear relationship of haze-fog formation in traditional methods. It is difficult to simulate the nonlinear dynamic evolution processes responsible for the formation of haze-fog in the existing physical chemistry mechanism and statistical analysis; therefore, their prediction and analysis on the formation of haze-fog is sometimes inaccurate.

Due to the complexity of relationships among different factors and the dynamic evolution of the formation of haze-fog, two considerations need to be taken into account. One is the complex cognitive model for the formation of haze-fog. Second is to discover the degree (fuzzy values) of the cognitive relationships in the nonlinear evolution of the formation of haze-fog. These two points can be explained as follows:

(1) The complex cognitive model for the haze-fog formation

The complex cognitive model should address certain questions about the process of haze-fog formation. For example, what are the factors involved in the formation of haze-fog? What are the relationships among the factors in the cognitive map for haze-fog formation? How do we represent the states (fuzzy value) of factors in fuzzy form? How do the states and the relationships evolve with form?

The FCM structure is similar to a recurrent artificial neural network, where concepts are represented by neurons and causal relationships by weighted links connecting the neurons. The fuzzy cognitive map shown in Figure 2 is a 4-tuple (C, W, A, f) mathematically, where:

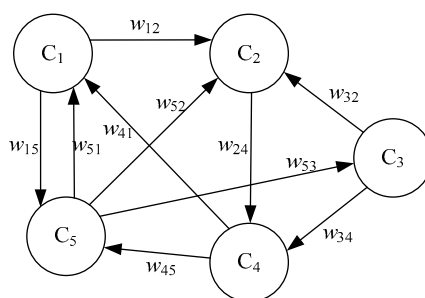


Figure 2. A simple fuzzy cognitive map.

- $C = \{C_1, C_2, \dots, C_n\}$ is the set of n nodes of a graph, which represents a set of concepts of a system, in general.
- $W: (C_i, C_j) \rightarrow w_{ij}$ is a function of $n \times n$ to a pair of concepts (C_i, C_j) taking value in the range -1 to 1 , with w_{ij} denoting a weight of directed edge from C_i to C_j , if $i \neq j$, and w_{ij} equal to zero if $i = j$. Thus, $W (n \times n) = (w_{ij})$ is a connection matrix.

There are three possible types of causal relationships between concepts:

$w_{ij} > 0$, which indicates positive causality between concepts C_i and C_j . That is, the increase (decrease) in the value of C_i leads to the increase (decrease) on the value of C_j ;

$w_{ij} < 0$, which indicates negative causality between concepts C_i and C_j . That is, the decrease (increase) in the value of C_i leads to the increase (decrease) on the value of C_j ; and

$w_{ij} = 0$, which indicates no relationship between C_i and C_j .

- $A(t) = \{A_1(t), A_2(t), \dots, A_n(t)\}$ is a sequence of concepts activation degrees at the moment t . $A(0)$ indicates the initial vector and specifies initial values of all concept nodes and $A(t)$ is a state vector at certain iteration t .

The state vector specifies current values of all concepts (nodes) in a particular iteration. The value of a given node is calculated from the preceding iteration values of nodes, which exert influence on the given node through cause–effect relationship (nodes that are connected to the given node).

- f is a transformation function, which includes recurring relationship on $t \geq 0$ between $A(t + 1)$ and $A(t)$.

$$\forall i, j \in \{1, 2, \dots, n\}, A_i(t + 1) = f \left(\sum_{\substack{i=1, \\ i \neq j}}^n w_{ji} A_j(t) \right),$$

where $A_j(t)$ is the state of cause concept j at t iteration and $A_i(t)$ is the state of effect concept i at $t + 1$ iteration, w_{ji} is a cause-effect relationship weight from C_j to C_i . The transformation function is used to confine the weighted sum to a certain range, which is usually set to $[0, 1]$. The three most commonly used transformation functions are shown in Equations (1)–(3).

- bivalent

$$f(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x > 0 \end{cases} \quad (1)$$

- trivalent

$$f(x) = \begin{cases} -1 & x \leq -0.5 \\ 0 & -0.5 < x < 0.5 \\ 1 & x \geq 0.5 \end{cases} \quad (2)$$

- logistic

$$f(x) = \frac{1}{1 + e^{-\mu x}} \quad (3)$$

There are several simulation scenarios, which are dependent on the transformation function. Applying a discrete transformation function (e.g., the bivalent or trivalent function), the simulation heads to either a fixed state vector value, which is called fixed-point, or keeps cycling between a number of fixed state vector values, which is known as a limit cycle. Using a continuous transformation function (e.g., the logistic signal function), the fixed-point and limit cycle, as well as a so called chaotic attractor, may appear.

The nodes in FCM can be expressed as concepts in the formation of haze-fog. The arc between nodes is the relationship between concepts. Each node has a time series state space. Each node or each arc has strong cognitive semantics so that it is very intuitive to express problems. The inference of FCM takes the advantages of computer in digital computing based on matrix operation. The state space of FCM is formed by the automatic transmission of the evolution function from initial condition. Thus, the dynamic behavior of haze-fog formation can be simulated through the interaction of the concepts in the cognitive network.

(2) Discovering the fuzzy cognitive relationships in the fuzzy cognitive map for haze-fog formation

The formation of haze-fog generates a lot of data. The data are complex and sequential. How to discover the fuzzy cognitive relationships from the data resources needs to be solved.

The FCM learning algorithm can obtain the association matrix of FCM from data. There are two classes of FCM learning algorithms: Hebbian-based learning and evolved-based learning. The former are Hebbian-based algorithms [17,44,45], mainly including NHL (nonlinear Hebbian learning) and AHL (active Hebbian learning). The latter are learning algorithms based on evolution theory [17,46,47], which are composed of PSO (particle swarm optimization), RCGA (real coded genetic algorithm), etc. The evolutionary learning can obtain the cause-effect relationships of FCM from the time series data. Therefore, FCM is more applicable to time series data mining, to simulate the nonlinear dynamic

evolution process in the formation of haze-fog, and to find fuzzy cognitive relationships (values) in the formation of haze-fog. FCM can be used to simulate the dynamic behavior of haze-fog formation complex cognitive model for haze-fog formation.

Being aimed at the complexity in the formation of haze, from the view of the complex cognitive relationship in the formation of haze-fog, this research is focused on mining the fuzzy cognitive relationships (values) by nonlinear time series data for the formation of haze-fog. This can also be applied in the prediction of the intensity of haze-fog and the analysis of complex relationship in the formation of haze-fog for emergency warning decision support systems.

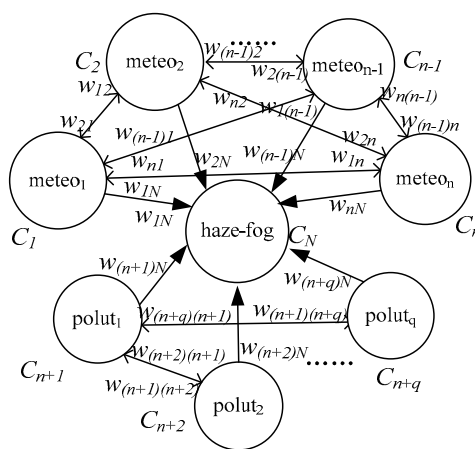
3. A New Perspective: The Fuzzy Cognitive Map for Haze-Fog Formation

3.1. The Construction of the Fuzzy Cognitive Map for Haze-Fog Formation

From the view of cognition, there is cause-and-effect relationship between the pollutants, meteorological conditions, other possible factors, and the formation of haze-fog. The factors and haze-fog can be simulated as concept nodes of a fuzzy cognitive map in the formation of haze-fog. The concept nodes $C = \{C_1, C_2, \dots, C_n\}$ are defined as a set of the concepts in the formation of haze-fog.

The concepts in the formation of haze-fog mutually affect and interact forming an organic and complex system. There are cause-effect relationships among the concepts representing pollutants. For example, the pollutants of SO_2 , NO_2 may cause “secondary pollution” of $\text{PM}_{2.5}$. In the same way, there are cause-effect relationships among the concepts representing meteorological conditions, such as temperature, may influence relative humidity. There are cause-effect relationships between the concepts representing pollutants and meteorological conditions on the formation of haze-fog. According to the cognitive relationships among them, the relationships of FCM can be constructed as Figure 3.

Then the degrees of the cause-effect relationships are described as $W = \{w_{ij} \mid w_{ij} \text{ which is the value of the arc } \langle C_i, C_j \rangle \text{ in } [-1, 1]\}$, where w_{ij} is the causality degree of the concepts C_i to C_j . A represents the state spaces of FCM in the formation of haze-fog. The concepts in the formation of haze-fog have state spaces with time series corresponding to the sequence dataset in the formation of haze-fog. Each state is a fuzzy value in $[0, 1]$. The state of the FCM in the formation of haze-fog at time t is represented as $A(t) = (\text{Polut}(t), \text{Meteo}(t), \text{Haze}(t))$. The $\text{Haze}(t)$ is a state value of haze-fog at time t . The $\text{Polut}(t) = [\text{polut}_1(t), \text{polut}_2(t), \dots, \text{polut}_q(t)]$ that is one vector representing the concentrations of pollutants at time t . The $\text{Meteo}(t) = [\text{meteo}_1(t), \text{meteo}_2(t), \dots, \text{meteo}_n(t)]$ is one state vector of the meteorological factors at time t .



*Notes: $N=n+q+1$

Figure 3. FCM in the formation of haze-fog.

3.2. The Evolution Mechanism of the Fuzzy Cognitive Map to Haze-Fog Formation

There are causal relationships of the concepts representing pollutants and meteorological conditions on the concept of haze-fog. The evolution mechanism of the formation of haze-fog follows the cognitive inference rules of FCM which can be shown in Equation (4):

$$\begin{aligned} \text{Haze}(t+1) &= f(\alpha_i \text{Haze}(t) + \beta_j \text{Polut}(t)W_p + \gamma_k \text{Meteo}(t)W_m + \theta_l), \\ t &\in \{0, 1, 2, \dots, T\}, \end{aligned} \quad (4)$$

where $\alpha_i, \beta_j, \gamma_k$, respectively, are the parameter of haze intensity, pollutants, and meteorological conditions in the formation of haze-fog, θ_l is the parameter of other possible influencing factors, f is the evolution function, such as sigmoid, and $W_p = [w_{p1}, w_{p2}, \dots, w_{pq}]^T$ and the $W_m = [w_{m1}, w_{m2}, \dots, w_{mn}]^T$. They are the relationship vectors of pollutant and meteorological conditions of the formation of haze-fog, respectively.

There are causal relationships among the concepts representing pollutants. The inference is shown as Equation (5):

$$\text{polut}_i(t+1) = f(\lambda_i \text{polut}_i(t) + \sum_{\substack{i \neq j \\ j \in S_i}} \lambda_j \text{polut}_j(t)w_{ji}) \quad (5)$$

Similarly, there are causal relationships among the concepts representing meteorological conditions. The inference is shown as Equation (6):

$$\text{meteo}_i(t+1) = f(\eta_i \text{meteo}_i(t) + \sum_{\substack{i \neq j \\ j \in S_i}} \eta_j \text{meteo}_j(t)w_{ji}), \quad (6)$$

where S_i is a set of nodes associated with C_i , w_{ji} is the casual value of the C_j to C_i , and $\lambda_i, \lambda_j, \eta_i, \eta_j$ are the state parameters in the FCM, respectively. The logistic function can be chosen as transformation function shown in Equation (3).

The state values of the concepts in the formation of haze-fog evolve with time after the initial state has been determined in accordance with the nonlinear dynamic reference rules that are represented by Equations (3)–(6), and finally may end at one of three stable states that are a fixed point, a limited cycle, and a chaotic attractor.

4. The Approach to Data Mining of the Fuzzy Cognitive Map for Haze-Fog Formation

How to discover the degrees of cause-effect is a key issue in fuzzy cognitive mapping for the formation of haze-fog. Massive sequential data is generated in the process of haze-fog formation. It can be considered as a nonlinear evolution of fuzzy cognitive map for the formation of haze-fog, and the degrees of cause-effect relationship need a data mining method of fuzzy cognitive map in order to be discovered.

Taking the genetic algorithm in evolutionary learning as an example, the work flow using the time series mining of the fuzzy cognitive map for the formation of haze-fog is shown in Figure 4. The key procedures are the initialization method, the determination of mining end condition, and optimization method. The causality value can be mined by these methods.

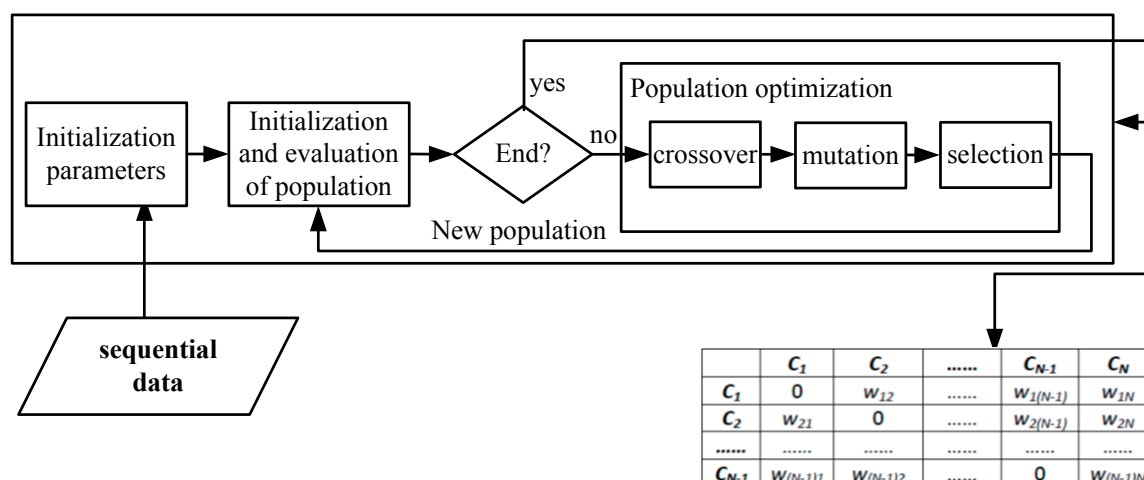


Figure 4. The genetic algorithm procedure of the fuzzy cognitive map for haze-fog formation.

The main steps are the mining of the degrees in the fuzzy cognitive map by simulating the nonlinear changes of the fuzzy cognitive map in the haze-fog formation driven by multidimensional sample data. The genetic algorithm is depicted in Table 1.

Table 1. The genetic algorithm of the fuzzy cognitive map.

Inputs: Sample Data from the Process of Haze-Fog Formation	
Step1.	Initialize parameters of genetic algorithm and the FCM within the known range.
Step2.	Generate initial population based on operator.
Step3.	Calculate fitness function according to the time series data.
Step4.	Evolve the population.
Step5.	Return to Step 3, until the fitness function is maximized (i.e., the end of mining conditions) after finite iterations.
Outputs: The Relationship Degrees in the Fuzzy Cognitive Map for Haze-Fog Formation	

Each correlation degree in the FCM can be defined as a gene. Assume that there are N genes; these degrees (weights) can be compiled as a real vector, which represents an N -dimensional chromosome, $W^{(1)} = [w_1^{(1)}, \dots, w_i^{(1)}, \dots, w_N^{(1)}]$. According to the range of each gene, created using Logistic mapping, it forms a numbers of M initial chromosomes $\{W^{(1)}, \dots, W^{(u)}, \dots, W^{(M)}\}$ for M chromosomes.

The simulation at time t can be calculated by the previous at time $t - 1$ and the evolution rules of Formulas (4)–(6). Assuming that the simulation states of the fuzzy cognitive map at time t are $A'(t) = \{A_1'(t), \dots, A_i'(t), \dots, A_v'(t)\}$ and the actual states (from the measured data) are $A(t) = \{A_1(t), \dots, A_i(t), \dots, A_v(t)\}$, the error is given as the difference between the actual value of $A(t)$ and the simulated value of $A'(t)$. This error is basis of fitness function definition, such as the one given in Equation (7). The higher the sufficiency is, the better is the fitting of time series data. When the fitness function reaches the threshold value, the process ends:

$$fitness = \frac{1}{\frac{1}{v} \sum_{i=1}^v (A_i - A_i')^2 + 1} \quad (7)$$

The individual weights are optimized through crossover and mutations in the genetic algorithm. A new individual is generated by crossover operation of two parent individuals. Mutation operation gives the individual component of the real number code. Through the crossover and the mutation, the best choice is selected from the new individuals and their parent, and it enters into the following generation groups for the purpose of optimization.

5. Experiments and Results

The meteorological data from NCEP (National Centers for Environmental Prediction) in the USA has been used [48]. The meteorological data consider temperature, pressure, relative humidity, and wind speed. The pollutant and haze-fog data have been obtained from the datacenter of the MEP (Ministry of Environment Protection) of China. The pollutant data are about PM_{2.5}, SO₂, NO₂, and CO. These daily data have been chosen from 23 October 2014 to 6 January 2015, for Beijing.

According to the study in Section 2.2 and the data characteristics, the meteorological conditions of temperature, pressure, relative humidity, and wind speed are identified as meteorological concepts of the fuzzy cognitive map for haze-fog formation. According to the study of pollutants and their characteristics summarized in Section 2.2, PM_{2.5}, SO₂, NO₂, and CO have been identified as the polluting concepts of the fuzzy cognitive map for haze-fog formation. The haze-fog has been identified as fuzzy cognitive map for understanding its formation.

Accordingly, one weight matrix has been constructed for meteorological conditions and haze-fog. Another weight matrix has been constructed for pollutants and haze-fog. The concepts included in the FCM are temperature, pressure, relative humidity, wind speed, PM_{2.5}, SO₂, NO₂, CO, and haze-fog. A genetic algorithm is chosen as a data mining approach for the FCM. The experiment has been run and the values are reported below:

- recombination method—single-point crossover;
- mutation method—random mutation;
- selection method—roulette wheel;
- probability of recombination: 0.8;
- probability of mutation: 0.5;
- population_size: 200 chromosomes;
- max_generation: 500,000;
- max_fitness: 0.9; and
- the parameters of the FCM: $\alpha_i = \beta_j = \gamma_k = \lambda_i = \lambda_j = \eta_i = \eta_j = \mu = 1, \theta_l = 0.7$

Through the implementation of experiments, using the cause-effect weights in the formation of haze-fog, the key influencing factors can be analyzed. In meteorological conditions' wind speed has a stronger influence on the haze-fog in Beijing. In the pollutants, PM_{2.5} is the main pollutant influencing the formation of haze-fog in Beijing.

In order to verify the fuzzy cognitive map for the formation of haze-fog, the intensity of haze-fog formation is represented as a fuzzy value in fuzzy cognitive map and a single rule of validity determination is formulated. The fuzzy intensity has been divided into four sections. Clear cases are in [0, 0.25), slight haze in [0.25, 0.50), mild haze in [0.50, 0.75), and severe haze in [0.75, 1]. Since the actual value is an interval such as fog, or haze and so on, while the forecast result is fuzzy value, a rule is defined for determining the validity of the FCM and its data mining:

if (the forecast value is in the interval of corresponding actual intensity)
 the number of valid forecast in the right interval plus one;
 else
 the invalid number in a wrong interval plus one.

The valid forecast points to forecast value being in the range of corresponding actual intensity. In order to further validate it over multiple times, the Kappa Index of Agreement (K) is defined to help evaluate the outputs expressed by Equation (8), where P_A is the observed consistency and P_e is the expected consistency. If all forecast results are valid, K equals 1. The larger the Kappa value, the better the consistency:

$$K = \frac{P_A - P_e}{1 - P_e} \quad (8)$$

The haze-fog has been predicted for the coming 10, 20, and 30 days from 7 January 2015. There are two experiments that are implemented. They are based on the FCM and the MRM in Meng et al. [43], respectively. The experimental results are shown in Tables 2–4. The Kappa values based on the FCM are 0.861, 0.861, and 0.814, respectively. The Kappa values based on the multiple regression model are 0.855, 0.791, and 0.717, respectively. From the Kappa values, it can be observed that these experiments obtain better results for haze-fog prediction.

Table 2. The forecast during 10 days.

Forecast Result Haze Intensity (Actual Number)	Haze Intensity	[0, 0.25)		[0.25, 0.50)		[0.50, 0.75)		[0.75, 1]	
		FCM	MRM	FCM	MRM	FCM	MRM	FCM	MRM
[0, 0.25) (4)		4	4	0	0	0	0	0	0
[0.25, 0.50) (2)		0	0	2	2	0	0	0	0
[0.50, 0.75) (1)		0	0	0	0	1	1	0	0
[0.75, 1] (3)		0	1	0	0	1	0	2	2

Table 3. The forecast during 20 days.

Forecast Result Haze Intensity (Actual Number)	Haze Intensity	[0, 0.25)		[0.25, 0.50)		[0.50, 0.75)		[0.75, 1]	
		FCM	MRM	FCM	MRM	FCM	MRM	FCM	MRM
[0, 0.25) (6)		5	5	1	1	0	0	0	0
[0.25, 0.50) (7)		0	0	7	7	0	0	0	0
[0.50, 0.75) (2)		0	0	0	0	2	2	0	0
[0.75, 1] (5)		0	1	0	0	1	1	4	3

Table 4. The forecast during 30 days.

Forecast Result Haze Intensity (Actual Number)	Haze Intensity	[0, 0.25)		[0.25, 0.50)		[0.50, 0.75)		[0.75, 1]	
		FCM	MRM	FCM	MRM	FCM	MRM	FCM	MRM
[0, 0.25) (11)		10	10	1	1	0	0	0	0
[0.25, 0.50) (10)		1	1	8	8	1	1	0	0
[0.50, 0.75) (3)		0	0	0	0	3	2	0	1
[0.75, 1] (6)		0	1	0	0	1	1	5	4

6. Conclusions

First, three types of existing methods on the formation of haze-fog have been summarized in this work.

- (1) Physical chemistry methods model the actual physics and chemical reactions of pollutants under the influence of the meteorological conditions. However, with more complex nature of reactions, the methods fail to describe and simulate the nonlinear processes involved in the haze-fog formation.
- (2) Statistical analysis methods incorporate the factors that are involved in the formation of haze-fog by using the measurements from equipment and the linear analysis of the contributing factors for the formation of haze-fog. They are important cognitive bases for the formation of the haze-fog. However, statistical analysis cannot describe the nonlinear dynamic process responsible for the formation of haze-fog.
- (3) Data mining methods can be used to discover the nonlinear relationships in the formation of haze-fog. However, at present, because of the limitation of the model such as in Meng et al. [43], not considering the correlations among the contributing factors and the dynamic changes in data, the results in existing data mining methods are unsatisfactory.

However, the following points still need to be studied further, especially under the formation conditions that are becoming increasingly complex:

- (1) Quantitatively dynamic models need to be further developed for the formation of haze-fog under increasingly complex scenarios.
- (2) The relationships among the factor concepts in the formation of haze-fog need to be well recognized and modeled.
- (3) The dynamic and nonlinear changes need to be further simulated for forecasting the formation of haze-fog.

Thus, a new view on the formation of haze-fog addressing the above issues has been proposed. It is aimed at the complex relationships in the formation of haze-fog using fuzzy cognition and data mining. The research scheme has been put forward in two sections. One is construction of the FCM and the other is the data mining approach for FCM on the formation of haze-fog. The FCM involves the evolution mechanism in its construction. The other is the construction of the FCM for the formation of haze-fog, and the data mining for the fuzzy cognitive relationships (degrees) from the dynamic sequential data.

The preliminary experimental results based on genetic algorithm indicate that the nonlinear cognitive and construction for the formation of haze-fog can get better forecast and analysis of key influencing factors. Of course, there are many problems to be improved, such as involvement of more influencing factors, parameter setting, comparison of several experiments, accuracy, operational efficiency, etc. Nevertheless, the research has proposed a kind of model and method simulating the dynamic and complex relationships among the influencing factors and haze-fog and, to some extent, demonstrated better theoretical and practical significance for the formation of haze-fog with increasingly complex meteorological conditions and pollutants. This can provide the basis for forecasting the formation of haze-fog, complex analysis of the formation of haze-fog, and provide important decision support for emergency early warning and control of haze-fog.

Acknowledgments: This research was supported by the National Natural Science Foundation of China (No. 71601022), the Natural Science Foundation of Beijing (4173074), the Project of Beijing Social Science Foundation (No. 15JDJGB028), the Natural Science Foundation of Province (No. F2014508028).

Author Contributions: Zhen Peng performed the studies on the fuzzy cognitive map and its approaches to data mining. Zhen peng and Lifeng Wu study the formation of haze-fog and implement the experiments. Zhen Peng prepared the manuscript.

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

References

1. Huang, R.J.; Zhang, Y.; Bozzetti, C.; Ho, K.F.; Cao, J.J.; Han, Y.; Daellenbach, K.R.; Slowik, J.G.; Platt, S.M.; Canonaco, F.; et al. High secondary aerosol contribution to particulate pollution during haze events in China. *Nature* **2014**, *514*, 218–222. [[CrossRef](#)] [[PubMed](#)]
2. Zhou, C.H. On-Line Numerical Research on Atmospheric Aerosols and Their Interaction with Clouds and Precipitation. Ph.D. Thesis, University of Chinese Academy of Sciences, Beijing, China, 2013.
3. Zhai, P.M.; Yu, R.; Guo, Y.J.; Li, Q.X.; Ren, X.J.; Wang, Y.Q.; Xu, W.H.; Liu, Y.J.; Ding, Y.H. The strong El Niño in 2015/2016 and its dominant impacts on global and China's climate. *Acta Meteorol. Sin.* **2016**, *74*, 309–321.
4. Hexun.com. Available online: http://yx119.blog.hexun.com/96826207_d.html (accessed on 24 November 2014).
5. Wang, W. Studies on Haze Control through Middle Route of South-to-Noah Water Diversion Project in Hebei Province. *China Water Resour.* **2014**, *2*, 11–13.
6. Ganguly, N.D.; Tzanis, C. Study of stratosphere-troposphere exchange events of ozone in India and Greece using ozonesonde ascents. *Meteorol. Appl.* **2011**, *18*, 467–474. [[CrossRef](#)]
7. Seinfeld, J.H.; Pandis, S.N.; Noone, K. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2006; p. 1595.

8. Tzanis, C. Ground-based observations of ozone at Athens, Greece during the solar eclipse of 1999. *Int. J. Remote Sens.* **2005**, *26*, 3585–3596. [[CrossRef](#)]
9. Chan, C.K.; Yao, X.H. Air pollution in megacities in China. *Atmos. Environ.* **2008**, *42*, 1–42. [[CrossRef](#)]
10. Tzanis, C.; Tsiola, E.; Efstathiou, M.; Varotsos, C. Forest fires pollution impact on the solar UV irradiance at the ground. *Fresenius Environ. Bull.* **2009**, *18*, 2151–2158.
11. Zhang, X.Y.; Sun, J.Y.; Wang, Y.Q.; Li, W.J.; Zhang, Q.; Wang, W.G.; Quan, J.N.; Cao, G.L.; Wang, J.Z.; Yang, Y.Q.; et al. Factors contributing to haze and fog in China. *Chin. Sci. Bull.* **2013**, *58*, 1178–1187. (In Chinese) [[CrossRef](#)]
12. Kosko, B. Fuzzy cognitive maps. *Int. J. Man Mach. Stud.* **1986**, *24*, 65–75. [[CrossRef](#)]
13. Papageorgiou, E.I.; Salmeron, J.L. A review of fuzzy cognitive maps research during the last decade. *IEEE Trans. Fuzzy Syst.* **2013**, *21*, 66–79. [[CrossRef](#)]
14. Lorenz, E.N. Deterministic non-periodic flow. *J. Atmos. Sci.* **1963**, *20*, 98–101. [[CrossRef](#)]
15. Du, J.; Qian, W.H. Three Revolutions in Weather Forecasting. *Adv. Meteorol. Sci. Technol.* **2014**, *4*, 13–27.
16. Papageorgiou, E.I. Learning algorithms for fuzzy cognitive maps—A review study. *IEEE Trans. Syst. Man Cybern. Part C* **2012**, *42*, 150–163. [[CrossRef](#)]
17. Peng, Z.; Wu, L.F.; Chen, Z.G. NHL and RCGA based multi-relational fuzzy cognitive map modeling for complex systems. *Appl. Sci.* **2015**, *5*, 1399–1411. [[CrossRef](#)]
18. Acampora, G.; Pedrycz, W.; Vitiello, A. A competent memetic algorithm for learning fuzzy cognitive maps. *IEEE Trans. Fuzzy Syst.* **2015**, *23*, 2397–2411. [[CrossRef](#)]
19. Stach, W.; Kurgan, L.A.; Pedrycz, W.; Reformat, M. Genetic learning of fuzzy cognitive maps. *Fuzzy Set Syst.* **2005**, *153*, 371–401. [[CrossRef](#)]
20. Peláez, C.E.; Bowles, J.B. Using fuzzy cognitive maps as a system model for failure modes and effects analysis. *Inf. Sci.* **1996**, *88*, 177–199.
21. Subramanian, J.; Karmegam, A.; Papageorgiou, E.; Papandrianos, N.; Vasukie, A. An integrated breast cancer risk assessment and management model based on fuzzy cognitive maps. *Comput. Methods Programs Biol.* **2015**, *118*, 280–297. [[CrossRef](#)] [[PubMed](#)]
22. Aju kumar, V.N.; Gandhi, M.S.; Gandhi, O.P. Identification and assessment of factors influencing human reliability in maintenance using fuzzy cognitive maps. *Qual. Reliab. Eng. Int.* **2015**, *31*, 169–181. [[CrossRef](#)]
23. Buruzs, A.; Hatwagner, M.F.; Kóczy, L.T. Expert-based method of integrated waste management systems for developing fuzzy cognitive map. In *Complex System Modelling and Control through Intelligent Soft Computations*; Zhu, Q., Azar, A.T., Eds.; Springer: Berlin, German, 2015; Volume 319, pp. 111–137.
24. Mago, V.K.; Morden, H.K.; Fritz, C.; Wu, T.; Namazi, S.; Geranmayeh, P.; Chattopadhyay, R.; Dabbaghian, V. Analyzing the impact of social factors on homelessness: A fuzzy cognitive map approach. *BMC Med. Inform. Decis. Mak.* **2013**, *13*, 859–871. [[CrossRef](#)] [[PubMed](#)]
25. Zhang, Y.J.; Zheng, M.; Cai, J. Comparison and Overview of PM_{2.5} Source Apportionment Methods. *Chin. Sci. Bull.* **2015**, *60*, 109–121. [[CrossRef](#)]
26. Tan, C.H.; Chen, X.; Zhao, T.L.; Shan, Y.P. Research Progress on the Development and Application of Air Quality Models. *Environ. Monit. Forewarning.* **2014**, *6*, 1–7.
27. Abdul-Wahab, S.; Sappurd, A.; Al-Damkhi, A. Application of California Puff (CALPUFF) model: A case study for Oman. *Clean Technol. Environ. Policy* **2010**, *23*, 177–189. [[CrossRef](#)]
28. Cheng, S.Y.; Liu, L.; Chen, D.S. Pollution abatement for improving air quality of Tangshan municipality, China: A perspective of urban-airshed carrying-capacity concept. *Int. J. Environ. Pollut.* **2010**, *42*, 5–31. [[CrossRef](#)]
29. Sickles, J.E.; Shadwick, D.S.; Kilaru, J.V.; Appel, K.W. “Transference ratios” to predict total oxidized sulfur and nitrogen deposition—Part II, modeling results. *Atmos. Environ.* **2013**, *77*, 1070–1082. [[CrossRef](#)]
30. Ding, Y.H.; Liu, Y.J. Analysis of long-term variations of fog and haze in China in recent 50 years and their relations with atmospheric humidity. *Sci. China Earth Sci.* **2014**, *57*, 36–46. [[CrossRef](#)]
31. Fu, G.Q.; Xu, W.Y.; Yang, R.F.; Li, J.B.; Zhao, C.S. The distribution and trends of fog and haze in the North China Plain over the past 30 year. *Atmos. Chem. Phys.* **2014**, *14*, 11949–11958. [[CrossRef](#)]
32. Guo, L.J.; Guo, X.L.; Fang, C.G. Observation analysis on characteristics of formation, evolution and transition of a long-lasting severe fog and haze episode in North China. *Sci. China Earth Sci.* **2015**, *58*, 329–344. [[CrossRef](#)]

33. Zhang, R.H.; Li, Q.; Zhang, R.N. Meteorological conditions for the persistent severe fog and haze event over eastern China in January 2013. *Sci. China Earth Sci.* **2014**, *57*, 26–35.
34. Yang, Y.Q.; Wang, J.Z.; Hou, Q. Research on PLAM Index Prediction Method for Air Quality in Beijing during 2008 Olympic Games. In Proceedings of the Conference of Chinese Society for Environmental Sciences, Chengdu, China, 22 August 2014.
35. Yang, Y.Q.; Wang, J.Z.; Hou, Q. A PLAM Index Forecast Method for Air Quality of Beijing in Summer. *J. Appl. Meteorol. Sci.* **2009**, *20*, 649–655.
36. Jansen, R.C.; Shi, Y.; Chen, J.M.; Hu, Y.; Xu, C.; Hong, S.; Li, J.; Zhang, M. Using hourly measurements to explore the role of secondary inorganic aerosol in PM_{2.5} during haze and fog in Hangzhou, China. *Adv. Atmos. Sci.* **2014**, *31*, 1427–1434. [[CrossRef](#)]
37. Shen, X.J.; Sun, J.Y.; Zhang, X.Y.; Zhang, Y.M.; Zhang, L.; Che, H.C.; Ma, Q.L.; Yu, X.M.; Yue, Y.; Zhang, Y.W. Characterization of submicron aerosols and effect on visibility during a severe haze-fog episode in Yangtze River Delta, China. *Atmos. Environ.* **2015**, *120*, 307–316. [[CrossRef](#)]
38. Zhang, Y.W.; Zhang, X.Y.; Zhang, Y.M. Significant Concentration Changes of Chemical Components of PM₁ in the Yangtze River Delta Area of China and the Implications for the Formation Mechanism of Heavy Haze-fog Pollution. *Sci. Total Environ.* **2015**, *538*, 7–15. [[CrossRef](#)] [[PubMed](#)]
39. Chung, Y.S.; Kim, H.S.; Yoon, M.B. Observations of Visibility and Chemical Compositions Related to fog, Mist and Haze in South Korea. *Water Air Soil Pollut.* **1999**, *111*, 139–157. [[CrossRef](#)]
40. Sun, R. Fog-haze Connecting Factors Analysis over the Beijing Region and Advance of the Standard. Master's Thesis, Nanjing University of Information Science & Technology, Nanjing, China, 2015.
41. Liu, J. Temporal-Spatial Variation as Well as Evaluation and Prediction Models of Air Pollutants in Beijing. Ph.D. Thesis, University of Science and Technology Beijing, Beijing, China, 2015.
42. Liu, D.J.; Li, L. Application Study of Comprehensive Forecasting Model Based on Entropy Weighting Method on Trend of PM_{2.5} Concentration in Guangzhou, China. *Int. J. Environ. Res. Public Health* **2015**, *12*, 7085–7099. [[CrossRef](#)] [[PubMed](#)]
43. Meng, Z.J.; Yue, X.N.; Wang, D.Z.; Yuan, Z.H. Model of Causes for Urban Fog-Haze Based on Multiple Regression Analysis. *J. Shenyang Univ. (Nat. Sci.)* **2015**, *27*, 139–142.
44. Papakostas, G.A.; Koulouriotis, D.E.; Polydoros, A.S.; Tourassis, V.D. Towards Hebbian learning of Fuzzy Cognitive Maps in pattern classification problems. *Expert Syst. Appl.* **2012**, *39*, 10620–10629. [[CrossRef](#)]
45. Papageorgiou, E.I.; Stylios, C.D.; Groumpos, P.P. Active Hebbian learning algorithm to train Fuzzy Cognitive Maps. *Int. J. Approx. Reason.* **2004**, *37*, 219–249. [[CrossRef](#)]
46. Wojciech, S.; Lukasz, K.; Witold, P. A divide and conquer method for learning large fuzzy cognitive maps. *Fuzzy Sets Syst.* **2010**, *161*, 2515–2532.
47. Oikonomou, P.; Papageorgiou, E.I. Particle Swarm Optimization Approach for Fuzzy Cognitive Maps Applied to Autism Classification. In Proceedings of the 9th IFIP International Conference on Artificial Intelligence Applications and Innovations, Paphos, Cyprus, 30 September 2013.
48. Ncep.Reanalysis.Dailyavgs. Available online: <ftp://ftp.cdc.noaa.gov/pub/Datasets/> (accessed on 7 December 2016).

