An Improved Markov Chain Model and Its Application

Huizhe Yan, *Member, IAENG*, Lihua Ma, *Member, IAENG*, Xiuling Yuan, Jichao Wang, Qun Zhou*

***ABSTRACT*—An enhanced forecast model of air quality was developed based on an improved Markov chain model and unascertained c-means clustering model with grey relational weights. The grey relational degree was used to describe the importance of the number of pollution days at each step for the forecasting target. To analyze changes in air quality, uncertain characteristics, unascertained c-means clustering division, and grey weights were applied to standardize the air quality time series, and future air quality scenarios were predicted with the weighted Markov chain. The state transition-probability matrix was also generated. The improved Markov chain model optimizes uncertain c-means clustering, the Markov chain, and the traditional one-step transfer probability methods, by combining the corresponding state transition-probability matrix for prediction with historical real-time data. According to the analysis and comparison of the results forecasted by the Grey Correlation-Markov model and Markov model, the probability interval obtained from the modified Markov model has distinct advantages in improving the accuracy and scientificity of forecasting. The results should provide government agencies with more detailed information that can help them to take measures to predict monthly pollution days and control air pollution in the city.**

***Index Terms*—unascertained c-means clustering; grey relational weight; improved Markov chain; air quality forecasting.**

I. Introduction

AIR quality is a major environmental and livelihood issue[1],and air pollution has emerged as a major issue regarding public health safety in China. Several studies have developed reliable and effective forecasting methods to counter the issue of air pollution. For example, Prasad developed ANFIS models for air quality forecasting[2]. Wang and Yixin analyzed the source of surface PM2.5 concentration in the Beijing-Tianjin-Hebei region based on MODIS data and air trajectories[3]. Cui performed a correlation analysis of changes in pollutants with time[4]. Sun used principal component analysis and the least squares vector support machine (LSSVM) optimized by the cuckoo search algorithm to predict daily PM2.5 concentration[5]. Chen used geographically-weighted regression to estimate the concentration of PM2.5[6]. Liu performed a correlation analysis of factors affecting PM2.5 concentration[7]. Guo predicted AQI based on the Kalman filter fusion algorithm[8]. Zhou predicted simulations of pollutants affecting air quality by using the WRF-chem model[9]. Wu and his collaborators conducted a series of prediction studies on air quality by using the fractional cumulative grey system model[10]. These methods can predict air quality well. Similarly, as a universal prediction method, Markov is rarely used in the field of air quality prediction, because the index value of air quality is not a state, but a series of nonlinear values[11]. This study aims to introduce the unascertained classification method to preprocess the air index and solve the application premise of the Markov prediction method.

This study is based on air quality samples corresponding to Handan city. The National Environmental Protection Department issued 2013–2016 key areas and 74 cities, including Handan City, regarding air quality status. Handan City is located in Hebei province, China, and the industrial (iron, steel, coal) development model is a pillar of its economy. With the Chinese Economic Reform and Opening Up, air pollution in this city is becoming increasingly serious[12]. Therefore, it is of practical significance to produce effective predictions according to real-time data of air quality in Handan City. This study reviews the development of new methods in the current literature, and mainly focuses on the correlation analysis of the pollution characteristics in Handan City[13]. This paper forecasts air quality from another aspect. Instead of analyzing the influencing factors of air quality, it uses real-time data by mathematical method to forecast air quality probability in the near future, and provides scientific basis for relevant departments to take measures. Common methods for forecasting air pollution include multivariate statistics, Monte Carlo simulation, spectrum analysis, and Markov chain. However, these methods are quite limited. This study determined the number of pollution days per month (AQI > 150) based on air quality index (AQI) data of Handan City. The AQI is a dimensionless parameter that measures overall quality of air, the calculation process of which is to calculate the individual index of each pollutant first, such as: SO_2, NO_2, PM10, PM2.5, O_3 and CO, and to then take the maximum value to perform the calculation[9]. The paper selects the AQI of Handan City as the research

Manuscript received April 4, 2022; revised February 3, 2023.This work was was supported in part by the Hebei Social Science Development Research Project under Grant 20220202093,and the Humanities and Social Science Research Project of Higher Education in Hebei Province under Grant SQ201027.

Huizhe Yan is a doctor in School of Management Engineering and Business, Hebei University of Engineering, Handan 056038, China.Email: yanhuizhe@163.com.

Lihua Ma is an associate professor in School of Management Engineering and Business, Hebei University of Engineering, Handan 056038, China. Email: malihua2004@126.com.

Xiuling Yuan is a Postgraduate in School of Management Engineering and Business, Hebei University of Engineering, Handan 056038, China. Email: 1198470708@qq.com.

Jichao Wang is a lecturer in School of International Education,Anyang Institute of Technology, Anyang 455000,China. Email:sdwjc@126.com.

Qun Zhou is a PhD student in School of Information Resource Management,Renmin University of China,Beijing 100872,China. Phone:+861062511461;email:zhouq1105@ruc.edu.cn.(Corresponding Author)

object, and extracts corresponding data spanning from November 2013 to September 2016 from Data Center website of Ministry of Environmental Protection of the People's Republic of China (http://datacenter.mep.gov.cn/index!MenuAction.action?name=402880fb24e695b60124e6973db30011.2017). The data were collected from various sources and physical causes were investigated to research uncertainty problems. The introduction of mathematical tools is advantageous for this purpose. First, we used the grey relational measure as a quantitative method of the unascertained c-means clustering for classifying the change interval and to reflect pollution scenarios, and then adopted grey weights to standardize the air quality time series. Simultaneously, the Markov model without grey correlation weighting is used for forecasting and conducting a comparative analysis, so as to demonstrate the advanced and scientific features of the method proposed by the current study. Finally, weighted Markov chains are used to predict the changing trend of pollution[11][14][15].

II. UNASCERTAINED C-MEANS CLUSTERING

Unascertained c-means clustering is a new method for classifying samples based on the unascertained mathematical theory[11]. To take into full consideration the uncertainty of air pollution status and the rationality of interval division, the unascertained c-means clustering method is used to divide the change interval of pollution. The basic idea is to divide the dimension space sample into a class, use the unascertained c-means clustering to classify the time series of pollution days, and generate the classification standard of pollution. There are exact numbers of pollution days in a year, and the specific process is described as follows:

A. Data prepossessing

Given N known samples, there are d characteristics that affect sample classification and the observed value of Sample x_i relating to Characteristic j is x_{ij}; observed values of observation indices of the samples involve different dimensions, so standardized processing is necessary to substitute for unfeasible direct comparison; each dimension of data $\{x_{1j}, x_{2j}, \cdots, x_{Nj}\}(j = 1,2,\cdots,d)$ shall be standardized. Sample x_i can be shown as a point in d dimension feature space: $x_i = (x_{i1}, x_{i2}, \cdots, x_{id})$, $(i = 1,2,\cdots,N)$, The formula is expressed as follows:

$$y_{ij} = (x_{ij} - \min_{1\le i\le N}\{x_{ij}\}) \Big/ (\max_{1\le i\le N}\{x_{ij}\} - \min_{1\le i\le N}\{x_{ij}\}) \quad (1)$$

Then the dimension $y_i = (y_{i1}, y_{i2}, \cdots, y_{id})$ of each sample is between 0 and 1.

B. Initialization of classification

Assume that the sample set after standardization is $y = \{y_1, y_2, \cdots, y_n\}$,

where $y_i = (y_{i1}, y_{i2}, \cdots, y_{id}), i = 1,2,\cdots,n$.

Let:

$$Sum(i) = \sum_{j=1}^{d} y_{ij} \quad (2)$$

$$MA = \max_i Sum(i) \quad (3)$$

$$MI = \min_i Sum(i) \quad (4)$$

$$J_i = \frac{(C-1)(sum(i)-MI)}{MA-MI} + 1 \quad (5)$$

Make $k_i = [J_i + 0.5]$, wherein $[y]$ refers to rounding of y; and k_i means Sample y falls into class k_i. Thus, N samples are divided into C categories, that is, giving an initial classification[16]. Let $m_k^{(0)}(k = 1 \sim C)$ be the category center of initial classification C_k, and then there is $m_k^{(0)} = (m_{k1}^{(0)}, m_{k2}^{(0)}, \cdots, m_{kd}^{(0)})^T$. To quantitatively describe the contribution of d characteristics to the initial classification separately:

$$\overline{m} = \frac{1}{C}\sum_{k=1}^{C} m_k \quad (6)$$

$$\overline{m}^{(0)} = (\overline{m}_1^{(0)}, \overline{m}_2^{(0)}, \cdots, \overline{m}_C^{(0)}) \quad (7)$$

where

$$\sigma_j^{2(0)} = \frac{\alpha_i}{C}\sum_{k=1}^{C}(m_{kj}^{(0)} - \overline{m}_j^{(0)})^2, (1 \le j \le d) \quad (8)$$

Then, variance σ_j^2 reflects the dispersion degree of K category centers $m_1, m_2, \cdots, m_K$ valuing on the j th Characteristic wherein α_i is an adjustable constant, and the simplest case is α_i =1. When σ_j^2 increases, the dispersion degree from each category center to $\overline{m}$ increases accordingly, that is, j Characteristic makes more contribution to distinguish C classifications.

$$w_j^{(0)} = \sigma_j^{2(0)} \Big/ \sum_{l=1}^{d} \sigma_l^{2(0)} \quad (9)$$

Obviously, $0 \le w_j \le 1$, $\sum_{k=1}^{d} w_j = 1$, and w_j expresses the classification weight of the given classification relating to Characteristic j.

The classification weight w_j describes the following: the proportion of contribution made by Characteristic j to separate each category center into different points with respect to all of the d classification characteristics.

And when $\sigma_j^2 = 0$, the j th weight of C category centers $m_1, m_2, \cdots, m_C$ is the same, and classification J should not appear in relevant formats calculating corresponding distance. Therefore, the distance from the sample point to the category center can be expressed as a weighted distance.

$$\left\| y_i - m_k^{(0)} \right\|^2 = \sum_{l=1}^{d} w_l^{(0)} (y_{il} - m_{kl}^{(0)})^2 \quad (10)$$

C. Unascertained membership degree

Given that the bigger the weighted Euclidean distance from Sample y_i to category center Γ_k , the lower the membership degree of y_i to Γ_k classification; if $\mu_{\Gamma_k}(y_i)$ is

used as membership degree of y_i relating to Γ_k classification, Although we can hardly know the true value of $\mu_{\Gamma_k}(y_i)$, we affirm that $\mu_{\Gamma_k}(y_i)$ becomes smaller when $\|y_i - m_k\|$ becomes bigger. Therefore, we only need to determine the relative magnitudes between membership degrees. Thus, let

$$\mu_{\Gamma_k}^{(0)}(y_i) = \frac{1}{\|y_i - m_k^{(0)}\|^2 + \varepsilon} \Big/ \sum_{l=1}^{C} \frac{1}{\|y_i - m_l^{(0)}\|^2 + \varepsilon} \tag{11}$$

Here, $\varepsilon = 0.01$ is used to control the constant, and adjust the impact of $\|y_i - m_k\|$ on the membership degree. The process is completed by performing these steps.

III. Markov Chain And Grey Prediction Model Combining Markov Chain

Definition 1. The state of each system can be expressed with random variables corresponding to a certain probability, and this probability is called 'state probability'. When the system transfers from the state of one stage into the state of another stage, the probability existing in such a transitional process is termed 'transition probability'. If the transition probability is only related to the current two adjacent states—that is, the state of the next stage is only related to the current state rather than the past state—then this random transition system process of a discrete state according to discrete time is called the Markov Process.

Definition 2. The whole limited Markov Process is called the Markov Chain[14]. Motion and variation analysis of the Markov Chain mainly analyzes and researches the state and interrelation of limited Markov Processes within the chain, which can further predict the future state of the chain and facilitate decision-making. According to the composition of the Markov Chain, its process is characterized by discreteness, randomness, and non-aftereffect.

A. Grey Correlation degree and analysis of Grey Correlation Markov forecasting

Definition 3 The grey correlation degree is the measure of the closeness degree of comparison sequence and reference sequence [17]. The grey correlation degree can be expressed as:

$$r_{oi} = \frac{1}{n}\sum_{k=1}^{n} L_i(k) \tag{12}$$

In the format, r_{oi} is the correlation degree of reference sequence and i th comparison sequence, and $L_i(k)$ is correlation coefficient, such that:

$$L_i(k) = \frac{\min\limits_i \min\limits_k \left|v_k^* - v_k^i\right| + \rho \max\limits_i \max\limits_k \left|v_k^* - v_k^i\right|}{\left|v_k^* - v_k^i\right| + \rho \max\limits_i \max\limits_k \left|v_k^* - v_k^i\right|}, \tag{13}$$

According to the format, v_k^* is the k th value of reference sequence; v_k^i is the k th value of i th comparison sequence; ρ is the identification coefficient, usually valued as 0.5; $\min\limits_i \min\limits_k \left|v_k^* - v_k^i\right|$ is two-stage minimum difference; $\max\limits_i \max\limits_k \left|v_k^* - v_k^i\right|$ is two-stage maximum difference.

The traditional Markov chain prediction method studies the current state of an object and predicts the future state according to a one-step transition probability; this method has been widely used for predicting the state of an object[18]. However, the main disadvantage of this method lies in its simplicity, which is not conducive to obtaining more comprehensive information. In addition, factors affecting the air pollution process are very complicated, and the traditional one-step transfer probability method has a higher rate of error in predicting pollution scenarios, and the number of polluted days is a set of dependent random variables. Therefore, this study introduced the grey relational degree to describe the importance of the number of polluted days in each step for the forecasting target. In this case, pollution days are the initial state through the historical real-time data, combined with the corresponding state transition probability matrix. Detailed steps are provided as follows for the predicted conditions corresponding to the full and rational information:

B. .Basic realization procedures of Grey Correlation Markov forecasting

Based on the train of thought above, in order to more clearly explain the model process of this paper, the prediction flowchart is shown in Figure 1.

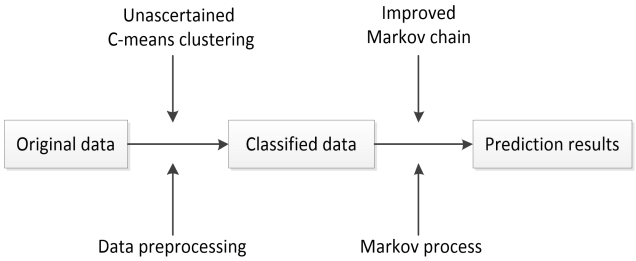

Fig. 1 The basic realization procedures of the model

From Figure 1, basic realization procedures relating to the unascertained cluster-Grey Correlation Markov Chain forecasting model are:

(1) Use the unascertained C-means cluster to produce an unascertained classification of the AQI time sequence, and generate the AQI classification standard, so as to determine the state space of the Markov chain.

(2) Determine the state of the AQI according to the classification standard generated.

(3) Use the state sequence obtained to generate the state-transition matrix of the Markov Chain of different step lengths.

(4) Determine the ideal state transition sequence p^*, obtain the standard value of the sequence; then the state is assumed to be equal to the probability of events; then $p^* = \{1/5, 1/5, 1/5, 1/5, 1/5\}$.

(5) Based on $V' = \left\{v'_{ij} \middle| v'_{ij} = v_{ij} / v_j^*\right\}$, arrange the state transition sequence and the ideal state transition sequence in an appropriate order.

(6) With the foundation of $C_i = \left\{c_{ij} \middle| c_{ij} = \left|v_j^* - v'_{ij}\right|\right\}$, calculate the difference in results between the state transition sequence of each order and the ideal state transition sequence.

(7) Calculate Grey correlation degree of state transition sequence of each stage and ideal state transition sequence by applying format (12) and (13).

(8) Based on $\bar{r}_{oi} = r_{oi} \Big/ \sum_{i=1}^{n} r_{oi}$ the grey correlation degree of each normalized order, the weight of $\bar{r}_{oi}$ is the order Markov chain.

(9)Calculate the number of days in the period with pollution state probability $P_i^{(n)}$ based on the Markov process, where i is the state, $i \in I$, n is step size, and $n = 1, \cdots, m$;

(10)Using the weighted state probability $P_i^{(n)}$ and $P_i = \sum_{i=1}^{m} \bar{r}_{oi} P_i^{(n)}$, take $\max\{P_i, i \in I\}$ corresponds i to the predicted state of the time period.

IV. Case Analysis

In this study, statistical data were obtained, and the number of pollution days per month (AQI > 150) was calculated based on AQI data from November 2016 to September 2019 in Handan City. Since air quality data are subject to the influence of different times of day and different seasons, this paper took 1 month as a unit of time to eliminate (or weaken) this influence, and the AQI data corresponding to each month were sourced from the data center website of the Ministry of Environmental Protection of the People's Republic of China. The AQI data[12][17] for April 2017 are shown in Table 1.

Table 1
April 2017 AQI value

Time	4/1	4/2	4/3	4/4	4/5	4/6	4/7	4/8	4/9	4/10
AQI	172	137	163	186	167	183	206	222	202	112
Time	4/11	4/12	4/13	4/14	4/15	4/16	4/17	4/18	4/19	4/20
AQI	179	161	203	187	167	186	157	163	112	140
Time	4/21	4/22	4/23	4/24	4/25	4/26	4/27	4/28	4/29	4/30
AQI	144	110	143	161	176	174	136	144	132	158

We can see from Table 1 that the data in Table 1 are irregular numbers with an unclear classification. For the purpose of predicting air quality, it is necessary to preprocess the original data similar to Table 1. Therefore, according to the AQI classification index, an AQI value greater than 150 is classified as moderate pollution, an AQI value greater than 200 as severe pollution and an AQI value greater than 300 as heavy pollution. Outdoor exercise time should be reduced when the AQI value indicates heavy pollution. Therefore, in this paper, number of pollution days per month (subject to AQI > 150) are shown in Chart 1: there were 20 pollution days in April 2017, and corresponding statistical data were obtained for the purposes of this study.

A. Pollution state by the unascertained C-means clustering division

In this study, we used MATLAB (Mathworks Inc., US) to analyze the unascertained c-means clustering iterative algorithm, taking $\alpha_i = 1$ and $\varepsilon = 0.01$. Refer to the relevant materials[19], input number of 5, divided into 1 to 5 states, the lowest state 1, the highest state of 5 is the most serious, and the classification results are shown in Table 3. The pollution states that are illustrated in Table 3 are classified according to the information shown in Table 2.

Table 2
Pollution rating

Status	Numerical interval
1	$x < 3$
2	$3 \le x < 9$
3	$9 \le x < 15$
4	$15 \le x < 21$
5	$x \ge 21$

Data from 35 samples covering the study period were inputted into the model, and air quality in October 2019 was predicted. The specific data are shown in Table 3.

Table 3
Handan City air quality state

Time period	11 (2016)	12	1 (2017)	2	3	4	5	6	7	8
Pollution Days	22	31	29	15	7	20	22	27	21	6
Pollution Days	22	31	29	15	7	20	22	27	21	6
state	5	5	5	4	2	4	5	5	5	2
Time period	9	10	11	12	1 (2018)	2	3	4	5	6
Pollution Days	22	24	21	26	24	21	24	17	25	3
state	5	5	5	5	5	5	5	4	5	2
Time period	7	8	9	10	11	12	1 (2019)	2	3	4
Pollution Days	4	8	4	13	22	25	31	22	12	9
state	2	3	2	3	5	5	5	5	3	3
Time period	5	6	7	8	9	10				
Pollution Days	7	4	7	3	4	9				
state	2	2	2	2	2	3				

B. State transition matrix for various steps

The state transition probability matrix was calculated from Table 2, and the results are as follows:

$$P_1 = \begin{bmatrix} 3/7 & 4/7 & 0/7 & 0/7 & 0/7 \\ 2/12 & 7/12 & 2/12 & 1/12 & 0/12 \\ 2/6 & 2/6 & 0/6 & 0/6 & 2/6 \\ 0/6 & 0/6 & 2/6 & 4/6 & 0/6 \\ 0/3 & 0/3 & 1/3 & 1/3 & 1/3 \end{bmatrix}$$

$$P_2 = \begin{bmatrix} 2/6 & 3/6 & 1/6 & 0/6 & 0/6 \\ 3/12 & 5/12 & 1/12 & 2/12 & 1/12 \\ 1/5 & 2/5 & 1/5 & 0/5 & 1/5 \\ 1/6 & 1/6 & 2/6 & 2/6 & 0/6 \\ 0/3 & 1/3 & 0/3 & 2/3 & 0/3 \end{bmatrix}$$

$$P_3 = \begin{bmatrix} 1/4 & 3/4 & 0/4 & 0/4 & 0/4 \\ 2/9 & 3/9 & 2/9 & 2/9 & 0/9 \\ 2/6 & 2/6 & 1/6 & 1/6 & 0/6 \\ 1/5 & 2/5 & 1/5 & 1/5 & 0/5 \\ 0/3 & 1/3 & 1/3 & 1/3 & 0/3 \end{bmatrix}$$

$$P_4 = \begin{bmatrix} 1/4 & 2/4 & 1/4 & 0/4 & 0/4 \\ 2/12 & 4/12 & 1/12 & 4/12 & 1/12 \\ 2/6 & 3/6 & 0/6 & 1/6 & 0/6 \\ 1/6 & 4/6 & 1/6 & 0/6 & 0/6 \\ 1/3 & 0/3 & 2/3 & 0/3 & 0/3 \end{bmatrix}$$

$$P_5 = \begin{bmatrix} 0/3 & 2/3 & 0/3 & 1/3 & 0/3 \\ 3/13 & 3/13 & 3/13 & 3/13 & 1/13 \\ 1/6 & 3/6 & 2/6 & 0/6 & 0/6 \\ 1/6 & 5/6 & 0/6 & 0/6 & 0/6 \\ 2/3 & 1/3 & 0/3 & 0/3 & 0/3 \end{bmatrix}$$

C. Results Forecast by Grey Correlation-Markov Model

Based on the Air Quality State Table (November 2016 to September 2019) and data from the state transition matrix of each step size, the results of the state transition-probability of air quality were obtained. The grey correlation grades were determined by using Equations 12 and 13, and then the grey correlation grade at each order was calculated by following Steps 7 and 8 to obtain the weight of the Markov Chain at each order. Predicted air quality status for October 2019 is shown in Table 4.

TABLE 4
2019 OCTOBER AIR QUALITY FORECAST TABLE

Starting date	Step length/a	Weights (Grey Correlation)	Transition probability State 1	State 2	State 3	State 4	State 5
9 (2019)	1	0.166	2/12	7/12	2/12	1/12	0/12
8 (2019)	2	0.171	1/3	3/6	1/6	0/6	0/6
7 (2019)	3	0.215	1/4	3/4	0/4	0/4	0/4
6 (2019)	4	0.216	1/4	2/4	1/4	0/4	0/4
5 (2019)	5	0.232	0/3	2/3	0/3	1/3	0/3
Pi Weighted summation			0.192	0.606	0.110	0.091	0.000

In Table 4, the calculation of the Grey correlation degree was analyzed using MATLAB2018. The calculation precision is 0.0001, excluding the weight and value.

D. Results Forecast by Markov Model

A comparative analysis was carried out to verify that the method provided herein is advanced, scientific and innovative. Based on the Air Quality State Table (November 2016 to September 2019) and the data from the state transition matrix of each step size, the results of the state transition-probability of air quality were obtained by applying the Markov Process (Steps 1 to 6, and Step 9)[20] . At the same time, Step 10 was replaced by summing and normalizing the state probability $P_i^{(n)}$. Predicted air quality status for October 2019 is shown in Table 5.

TABLE 5
2019 OCTOBER AIR QUALITY FORECAST TABLE

Starting date	Step length/a	Transition probability State 1	State 2	State 3	State 4	State 5
9 (2019)	1	2/12	7/12	2/12	1/12	0/12
8 (2019)	2	1/3	3/6	1/6	0/6	0/6
7 (2019)	3	1/4	3/4	0/4	0/4	0/4
6 (2019)	4	1/4	2/4	1/4	0/4	0/4
5 (2019)	5	0/3	2/3	0/3	1/3	0/3
Pi Summation		0.016	0.562	0.172	0.250	0

E. Comparative Analysis of Results Forecasted

1) As shown in Table 4, the corresponding probability Pi = 0.606 is the maximum of state 2, and is significantly larger than the values corresponding to the other four states. The results show that the number of air quality pollution days for October 2019 is 2, while Table 2 shows that the number of air pollution days for October 2019 is 3, which is correct.

2) As shown in Table 5, the corresponding probability is Pi=0.562, which is correct.

3) According to Table 5, State 3 corresponds to a probability of Pi=0.172, and State 4 corresponds to a probability of Pi=0.250, which is not obviously different from the result forecasted for State 2, while the probability corresponding to State 2 in Table 4 is apparently higher than the value corresponding to each of the rest four states. In addition, the results forecasted in Table 4 have distinct advantages (Pi=0.606>0.562). Thus, the modified Markov Model can be regarded as more scientific and effective.

V. CONCLUSIONS

To predict the number of air pollution days, AQI data from November 2016 to September 2019 in Handan City were analyzed on the basis of their quality condition research and statistics corresponding to the number of days where AQI per month was greater than 150. According to an AQI standard of more than 150, the AQI grade was divided into moderate pollution, serious pollution, and severe pollution. The prediction was improved by taking unascertained c-means clustering, grey correlation, and the Markov chain prediction model of monthly pollution. A more accurate range of predictions was obtained. According to the historical data, a more comprehensive Markov chain model based on unascertained c-means clustering with grey relational weights has been established. And according to the analysis and comparison of the results forecasted by the grey correlation-Markov model and Markov model, the probability interval (Pi=0.606>0.562) obtained from the modified Markov model has distinct advantages in improving the accuracy and scientificity of forecasting. The improved Markov chain model can predict the range more accurately and make the prediction more credible. This provides an effective and scientific theoretical basis for government agencies to take effective measures for controlling all kinds of air pollution. This paper offers novel insight into air quality supervision through this model which selects an approximate range of days instead of a specific AQI value to forecast air quality at a certain period.

DATA AVAILABILITY

The datasets used and analyzed in this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] Xie R , Wang F , Chevallier J , et al. Supply-side structural effects of air pollutant emissions in China: A comparative analysis. Structural Change and Economic Dynamics. 46, 89–95. https://doi.org/10.1016/j.strueco.2018.04.005 (2018).

[2] Prasad, K., Gorai, A. K., & Goyal, P. Development of anfis models for air quality forecasting and input optimization for reducing the computational cost and time. Atmospheric Environment. 128, 246–262. https://doi.org/10.1016/j.atmosenv.2016.01.007 (2016).

[3] Wang, Y., Jiang, H., Zhang, S., Xu, J., Lu, X., & Jin, J., et al. Estimating and source analysis of surface pm2.5 concentration in the beijing-tianjin-hebei region based on modis data and air trajectories.International Journal of Remote Sensing. 37, 4799–4817. https://doi.org/10.1080/01431161.2016.1220031 (2016).

[4] Cui, S., Ni, Y., & University, N. F. The correlation analysis of main gas pollutants and pm2.5 in an urban city. Forest Engineering. 32, 65–68 (2016).

[5] Sun, W., & Sun, J. Daily pm2.5 concentration prediction based on principal component analysis and lssvm optimized by cuckoo search algorithm. Journal of Environmental Management. 144–188. https://doi.org/10.1016/j.jenvman.2016.12.011 (2016).
[6] Chen H, Li. Q., Zhang, Y. H., Zhou, C.Y., & Wang, Z.T. Estimations of PM2.5 concentrations based on the method of geographically weighted regression. Acta ScientiaeCircumstantiae. 36, 2142–2151. https://doi.org/10.13671/j.hjkxxb.2015.0780 (2016).
[7] Wang, Y. Y., Zhang, X. X., Zhao, J. Y., & Jiang, Q. O. Temporal and spatial distribution of PM 2.5 and its relationship with vegetation coverage in Beijing during the period of 2013-2014.Ecology and Environmental Sciences. 25, 103–111 (2016).
[8] Guo, L., Jing, H., Nan, Y., & Xiu, C. Prediction of air quality index based on kalman filtering fusion algorithm. Environmental Pollution & Control. 39, 388–391 (2017).
[9] Ranran, H. E., Zhu, L., & Zhou, K. Spatial autocorrelation analysis of air quality index(aqi) in eastern china based on residuals of time series models. Acta Scientiae Circumstantiae. 7, 2459–2467 (2017).
[10] Wu L , Liu S , Yao L , et al. Grey system model with the fractional order accumulation[J]. Communications in Nonlinear Science and Numerical Simulation. 18, 1775–1785. https://doi.org/10.1016/j.cnsns.2012.11.017 (2013).
[11] Lizheng, Li, Shuming, Li, X.Y.Zhang, & L.N. Zhao.The application of uncertain measure model in the air quality appraisal of urban environment. Environmental Monitoring & Forewarning . (2013).
[12] Zhang, J., Song, X., & Office, H. M. Analysis of atmospheric environment features and impact factors of meteorological conditions in handan in 2014. Meteorological & Environmental Sciences. 9, 63–68 (2016).
[13] Zhe, W., Jing, Y., Wang, L., Wei, W., Zhang, F., & Jie, S. U.Characteristics of the severe haze episode in handan city in january,2013. Acta Scientiae Circumstantiae. 34, 1118–1124. https://doi.org/10.13671/j.hjkxxb.2014.0176 (2013).
[14] Zhao, L. L., & Xia, L. T. Improved gray markov scgm(1,1) model and its application. Journal of Hohai University. 35, 487–490. https://doi.org/10.1007/s10800-006-9244-6 (2007).
[15] Gabriel, K. R., & Neumann, J. A markov chain model for daily rainfall occurrence at tel aviv. Quarterly Journal of the Royal Meteorological Society. 88, 90–95 (1962).
[16] Cao, Q. K., Yang, R. X., & Di, L. K. Research on unascertained clusters on the gas emission of the working face. Journal of China Coal Society. 31, 337–341. https://doi.org/10.1016/S1872-2067(06)60047-8 (2006).
[17] Marwan, N., Trauth, M. H., Vuille, M., & Kurths, J. Comparing modern and pleistocene enso-like influences in nw argentina using nonlinear time series analysis methods. Climate Dynamics. 21, 317–326. https://doi.org/10.1007/s00382-003-0335-3 (2003).
[18] Fan J , Wu L , Zhang F , et al. Evaluating the effect of air pollution on global and diffuse solar radiation prediction using support vector machine modeling based on sunshine duration and air temperature[J].Renewable and Sustainable Energy Reviews. 94, 732–747. https://doi.org/10.1016/j.rser.2018.06.029 (2018).
[19] Wang, Y., Xue, Y., Tian, H., Gao, J., Chen, Y., & Zhu, C., et al. Effectiveness of temporary control measures for lowering pm 2.5, pollution in beijing and the implications. Atmospheric Environment. 157, 75–83. https://doi.org/10.1016/j.atmosenv.2017.03.017 (2017).
[20] Allagui, A., Rojas, A. E., Bonny, T., Elwakil, A. S., & Abdelkareem, M. A. Nonlinear time-series analysis of current signal in cathodic contact glow discharge electrolysis. Journal of Applied Physics. 119, 1–31. https://doi.org/10.1063/1.4952732 (2016).

Huizhe Yan was born in Xingtai City, Hebei Provence, P.R.China in 1981. She became a Member of IAENG in 2022(Member No: 321989).She is a doctor. She is the author of three books, more than 20 articles. Her research interests include issues related to multi-attribute decision making, data analysis and processing, risk analysis, etc.. She is the reviewers of the several international journals.