PIONEER
ACADEMIC PUBLISHING LIMITED

Journal of Progress in Engineering
and Physical Science
ISSN 2709-4006
www.pioneerpublisher.com/jpeps
Volume 2 Number 2 June 2023

Analysis and Evaluation of Global Temperature Changes and the Prediction of Future Global Temperature

Yuxing Du[1], Jun Liu[1], Jicong Yang[2] & Lu Zhang[1]

[1] School of Electrical and Electronic Engineering, Chongqing University of Technology
[2] College of Mechanical Engineering, Chongqing University of Technology
Correspondence: Lu Zhang, School of Electrical and Electronic Engineering, Chongqing University of Technology.

doi:10.56397/JPEPS.2023.06.03

Abstract

Temperatures have continued to rise in recent years, many areas have shown extreme temperatures, and global warming has been a perennial topic. We use various optimization prediction and optimization evaluation methods in this paper to determine the trends and causes of global temperature variation and make effective suggestions for them. We first constructed a first-order differential temperature model and found that the temperature rise in March 2022 was larger than that observed in the past 10 years. By constructing a prediction model (BP) based on BP neural network and a grey temperature prediction model (GAGM) optimized based on adaptive genetic algorithm, and by importing the data, we find that the average temperature in 2116 and 2145 will reach 20℃. In addition, if the correlation factor evaluation model (FAEM) is constructed in the country, and we conclude that there is a correlation between location, time and temperature. Next, we analyzed key environmental factors, such as volcanic eruptions, forest fires, and COVID-19. We imported the CO 2 concentration and temperature data into the FAEM evaluation model to obtain the results that natural disasters can lead to the temperature increase. Then nine environmental factors, including COND, NYKC, NYXH, GDP, PJRK, LJL, LZMJ, KJSP, and GJJY. Taking these factors as independent variables and temperature as the dependent variables, COND, NYXH, PJRK, LJMJ, JSL are the main causes of global temperature change and KJSP. This study provides an assessment of future global temperature changes, making recommendations for policymakers to optimize the energy mix and plant trees to effectively mitigate the global temperature rise.

Keywords: temperature variation, first-order difference model, BP neural network, GAGM

1. Introduction

Global climate change is a statistically significant change in the average state of climate around the world or a change in climate that lasts for an extended period of time (typically 30 years or more). Climate change can be caused by natural internal processes, or by external forcing, or by continuous anthropogenic changes in atmospheric composition and land use. This is due to the continuous accumulation of the

greenhouse effect, which leads to an imbalance between the energy absorbed and released by the Earth's atmospheric system, and the continuous accumulation of energy in the Earth's atmospheric system, which leads to temperature rise and global warming. And temperature is an important data point to describe global climate change.

Therefore, it is important to correctly assess the future global temperature level and explore which factors will affect global temperature. In a number of studies conducted in the early 1990s, researchers used trend analysis—specifically the Mann-Kendall (M-K) method and the Sen slope method—to study how human activities affect the Earth's natural systems and ecosystems. Many researchers in the field of environmental and atmospheric sciences already use these two methods to analyze and look for trends in data. In these studies, precipitation, temperature and water flow were considered steady. However, due to human influence on the atmosphere and hydrology, the assumption of stationarity no longer applies, which is caused by the rapid acceleration of climate change. Mann-kendall (M-K) and other methods make the data do not need to follow any distribution, is not disturbed by a few outliers, and the calculation is simple. However, it does not apply to sequences that detect multiple mutation points.

In this paper, we use the first-order differential temperature model construction, and the Excel pivot function is used to compare the temperature increase in March 2022 to the increase in the second decade from January 2013 to March 2002, determining that the temperature increase in March 2022 was greater than the past decade before the observation. Secondly, by constructing the gray-scale temperature prediction model (GAGM) optimized based on the adaptive genetic algorithm and the data-based BP neural network prediction model (BP), we used the data from 1900-2012 as the prediction data set, and the data from 2013-2022 as the prediction object, and compared it with the real data. Further conduct error test on the model, output the model conclusion with small error and high prediction accuracy, and select the model with small error is used as the prediction model for predicting the 2050 and 2100 temperature data. Finally, compare the time when the temperature concentration reaches 20℃ and output the conclusion of the judgment model.

In addition, according to the location height data of national location and degree, and then the location time of each country as the independent variable and temperature as the dependent variable, we constructed the temperature factor association evaluation model (FAEM), and calculated the relationship between the temperature and the iterative model of time and position height. Next, we analyzed the key factors affecting the environment, such as volcanic eruptions, forest fires, and COVID-19, and introduced these key factors into the FAEM evaluation model as the key influencing factors, analyzed the degree of their influence on temperature, and correlate the effects of these factors on temperature through their characteristic relationship score coefficients.

2. Experiment Datasets

This study used 239177 historical average temperature data from 1833 to 2012 in different countries from the Berkeley Earth website as experimental data.

2.1 Description of Symbol Meaning

Table 1. Description of symbol meaning

Symbols	Meaning of the symbols
$X_n = (x_1, x_2, \cdots, x_n)$	Network input vector
$T_n = (y_1, y_2, \cdots, y_n)$	Network target vector
$S_n = (s_1, s_2, \cdots, s_n)$	Intermediate layer cell input vector
$B_n = (b_1, b_2, \cdots, b_n)$	Intermediate layer unit output vector
$L_n = (l_1, l_2, \cdots, l_n)$	Output layer unit input vector
$Z_n = (z_1, z_2, \cdots, z_n)$	Output layer unit output vector
$w_{i,j}, i = 1,2,\cdots,p, j = 1,2,\cdots,p$	Link weights from the output layer to the intermediate layer

$v_{jt}, j=1,2,\cdots,p, t=1,2,\cdots,q$	Link weights from the middle layer to the output layer

2.2 Problem Hypothesis

(1). Assuming that the model shows no change in the factors affected by CO2 concentration and temperature during the prediction process, all environmental factors before and after the prediction are the same.

(2). Assume that there will be no great breakthrough in human technology, and that the existing energy and transportation will remain the mainstay.

(3). It is assumed that the data and information collected in this paper and the annexed data and information are true and reliable, and can accurately reflect the basic laws of global climate change.

(4). assumes that environmental factors will not be influenced by significant objective factors in the model's predictions, such as lack of resources, etc.

(5). Assume that the data given in the annex is true and valid and in accordance with objective laws.

2.3 Data Analysis and Data Pre-Processing

By analyzing the data we can conclude that there are more vacant values in the data, so for the data we need to perform data pre-processing. Data pre-processing is generally performed by data analysis, data cleaning, and data correction.

The data were first analyzed and it was concluded that the date data of the dataset done were rather mixed, and in order to count the monthly average data, it was necessary to unify the dates, the results of which are shown in Figure 1 below.

dt	AverageTemperature	AverageTemperatureUncertainty	City	Countr	Latitud	Longitu
2013/1/1	2.347	0.98	Kabul	Afghanist	34.56N	70.05E
2013/2/1	4.802	1.198	Kabul	Afghanist	34.56N	70.05E
2013/3/1	8.681	2.216	Kabul	Afghanist	34.56N	70.05E
2013/4/1	14.936	0.399	Kabul	Afghanist	34.56N	70.05E
2013/5/1	21.245	0.65	Kabul	Afghanist	34.56N	70.05E
2013/6/1	25.859	0.575	Kabul	Afghanist	34.56N	70.05E
2013/7/1	26.805	0.467	Kabul	Afghanist	34.56N	70.05E
2013/8/1	24.974	0.459	Kabul	Afghanist	34.56N	70.05E

Figure 1. Date Unification Process

Through our observation, there are a large number of missing values in the attached data. Since the amount of data is particularly large and also this question is analyzed by monthly average temperature or annual average temperature level, so the vacant data has a small impact on the results and can be ignored, so the data with vacant values are eliminated from the process.

With Excel's locate function, all the locations where data has missing values are located so that they can be cleared with one click.

According to the above data steps for processing, a relatively accurate and valid data set can be derived, then we apply Excel's pivot function to analyze the data and calculate the average temperature of all dates for each region, and the temperature of the eastern longitude region is shown in Figure 2 below.

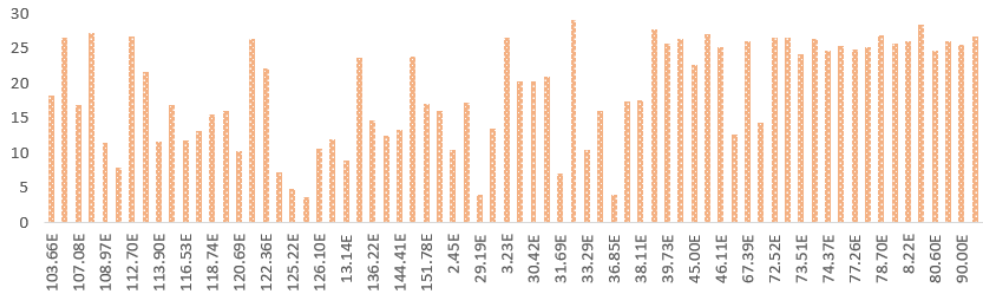

Figure 2. Average temperature data for the eastern longitude region

From Figure 2, the average data for each region can be derived, and by calculation, the global annual average data can be derived as shown in the figure below.

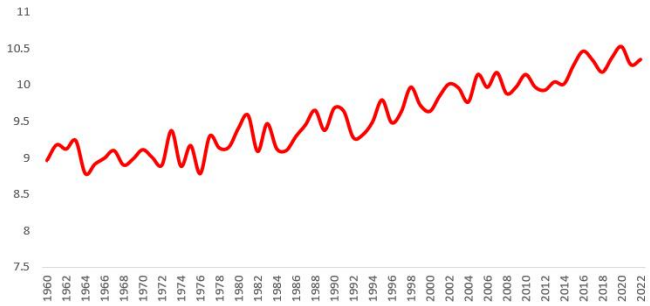

Figure 3. Average annual temperature

Based on the data set screened and calculated from the above steps, subsequent studies of temperature and other factors were conducted.

3. Analytical Methods and Models

In this paper, a first-order difference temperature model and Excel pivot function are used to compare the temperature increase in March 2022 with the temperature increase in January 2013 to March 2002 over a 20-year period. The gray-scale temperature prediction model (GAGM) optimized based on adaptive genetic algorithm and data-based BP neural network prediction model (BP) were constructed to predict the temperature data in 2050 and 2100. In addition, according to the location and height data of the country location and degree, the temperature factor correlation evaluation model (FAEM) is constructed, and the relationship between temperature and time and location height is calculated by the iterative model comprehensive training. Environmental factors such as volcanic eruption, forest fire and COVID-19 were introduced into the FAEM evaluation model as key influencing factors, and the degree of their influence on temperature was analyzed, and the influence of these factors on temperature was correlated by the score coefficient of their characteristic relationship.

3.1 Temperature Trend Evaluation

The first visualization is based on the global monthly average data obtained above from January 2013 to March 2022, as shown in Figure 4 below.

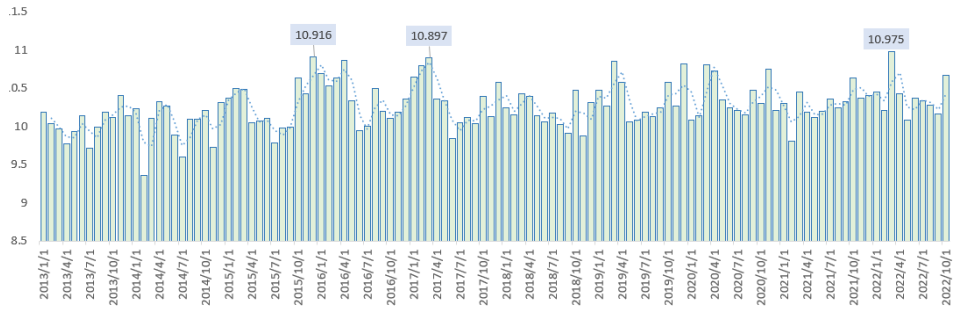

Figure 4. Global monthly average data from January 2013 to March 2022

A preliminary comparison of temperatures can be made according to Figure 4, due to the data of 10.975 for March 2022, which is greater than 10.916 for October 2016, but the difference is not significant enough to require a more accurate method for evaluation. First-order differencing is an effective way to distinguish data dominance. Therefore, this question constructs a first-order differential temperature model for analysis. The rules of the constructed first-order difference model are shown below.

$$\mu_i = \xi_i - \xi_{i-1} \tag{1}$$

μ_i represents the temperature growth value in month i. Then applying equation (1) to calculate the global monthly average data differential from January 2013 to March 2022, the result can be obtained as shown in the figure below.

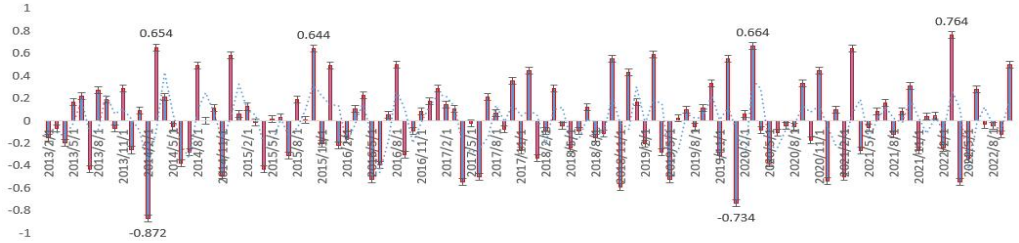

Figure 5. Monthly graph of first-order differential temperature data

According to Figure 5, it can be concluded that the first-order differential temperature value of 0.764 in March 2022 is among the maximum values from January 2013 to March 2022, which leads to the conclusion that there is a greater increase in temperature in March 2022 than in the previous decade of months.

3.2 Construction of Prediction Models

3.2.1 BP Neural Network Temperature Prediction Model

The BP neural network model is a basic system model for solving problems by analyzing the "weights" of the input information on the output results. The BP algorithm consists of two processes: forward propagation of the signal and backward propagation of the error. That is, the error output is calculated in the direction from the input to the output, while the adjustment of the weights and thresholds is performed from the output to the input. In forward propagation, the input signal acts on the output node through the implied layer and undergoes a nonlinear transformation to produce the output signal, and if the actual output does not match the desired output, it is transferred to the backward propagation process of the error. The error back-propagation is to back-propagate the output error through the implied layer to the input layer by layer and apportion the error to all units in each layer, using the error signal obtained from each layer as the basis for adjusting the weight of each unit. By adjusting the connection strength of the input nodes to the hidden layer nodes and the connection strength of the hidden layer nodes to the output nodes and the threshold value, the error decreases along the gradient direction, and after repeated learning training, the network parameters (weights and thresholds) corresponding to the minimum error are determined and the training is stopped. In this case, the trained neural network is able to process the input information of similar samples and output the non-linear transformed information with minimum error on its own.

(1) Model building process

Step 1: Perform initialization. The weights $w_{i,j}$ and v_{jt}, and thresholds θ_j and γ_t of each neuron are assigned random numbers within the interval (-1, 1).

Step 2: A set of learning samples $X_n = (\mathrm{x}_1, \mathrm{x}_2, \cdots, \mathrm{x}_n)$ and target samples $\mathrm{T}_n = (\mathrm{y}_1, \mathrm{y}_2, \cdots, \mathrm{y}_n)$ are selected in the Appendix.

Step 3: Compute the neuronal input s_j of the hidden layer using Equation (2) for the learning samples $X_n = (\mathrm{x}_1, \mathrm{x}_2, \cdots, \mathrm{x}_n)$, weights $w_{i,j}$ and threshold θ_j. Then, compute the output b_j of the intermediate layer units using the input values.

$$s_j = \sum_{j=1}^{n} w_{ij}\alpha_j - \theta_j, j = 1,2,\cdots,\mathrm{p} \quad (2)$$

$$\mathrm{b}_j = f\left(s_j\right), j = 1,2,\cdots,\mathrm{p} \quad (3)$$

Step 4: Compute the output layer neurons L_t using the implicit layer output b_j, the output layer weights $w_{i,j}$ and threshold γ_t, and the output layer result Z_t using the transfer function

$$\mathrm{L}_t = \sum_{j=1}^{n} v_{ij} b_i - \gamma_t, t = 1,2,\cdots,q \quad (4)$$

Step 5: The difference between the target vector $\mathrm{T}_n = (\mathrm{y}_1, \mathrm{y}_2, \cdots, \mathrm{y}_n)$ and the actual vector $Z_n = (\mathrm{z}_1, \mathrm{z}_2, \cdots, \mathrm{z}_n)$ of the network is calculated to obtain d_t.

$$d_t = (\mathrm{y}_t - c_t)\mathrm{c}_t(1 - \mathrm{c}_t), \mathrm{t} = 1,2,\cdots,\mathrm{q} \quad (5)$$

Step 6: calculate the intermediate layer general error e_j using the weights v_{jt}, the output

layer error d_t and the intermediate layer output value b_j.

$$e_j = \left[\sum_{t=1}^{q} d_t v_{jt}\right] b_j\left(1-b_j\right) \tag{6}$$

Step 7: The general error d_t of each unit in the output layer is used with the output value b_j of the intermediate layer unit to correct the connection weights v_{jt} and threshold γ_t.

$$v_{jt}\left(N+1\right) = v_{jt}\left(N\right) + \alpha d_t b_j \tag{7}$$

$$\gamma_t\left(N+1\right) = v_{jt}\left(N\right) + \alpha d_t \tag{8}$$

Where $t = 1,2,\cdots,q; j = 1,2,\cdots,p; 0 < \alpha < 1$.

Step 8: The middle layer neuron error e_j is used to adjust the connection weights $w_{i,j}$, and the input X_n of the input layer neuron is used to adjust the threshold θ_j.

$$w_{jt}\left(N+1\right) = v_{jt}\left(N\right) + \beta e_t a_i \tag{9}$$

$$\theta_j\left(N+1\right) = \theta_j\left(N\right) + \beta e_t \tag{10}$$

Where

$i = 1,2,\cdots,q; j = 1,2,\cdots,p;. < \beta < 1$

.

Step 9: The next sample is randomly selected and substituted into the neural network for training, and Step 3 is executed until the training global error of the network reaches the network convergence value and learning is completed (Chou, C. H., 2016).

(2) Model Parameter Setting

The coefficients for the goodness-of-fit determination of the model are shown below.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y_i})^2} \tag{11}$$

The goodness-of-fit coefficient takes values in the range of (0,1), and the closer the value is to 1, the better the performance of the model.

3.2.1.1 BP Neural Network Prediction Model Forecasting

MATLAB was applied to program the model separately, and the data from 1900 to 2012 were used as the training data set and the data from 2012 to 2022 were used as the test data set, which were imported into the BP neural network prediction model, and the training results and test results of the model are shown below.

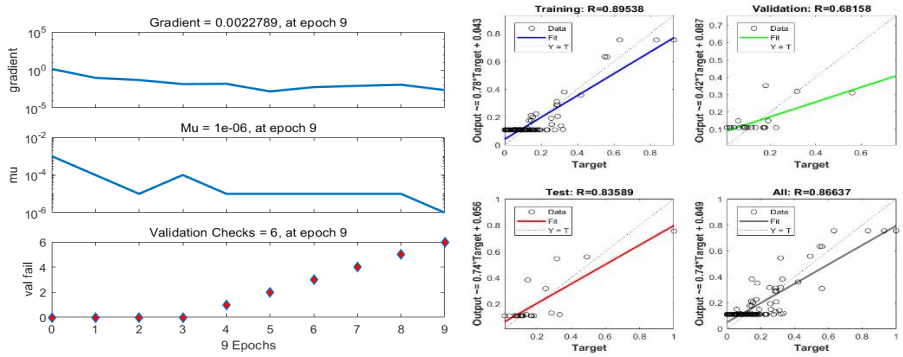

Figure 6. Training prediction graph of BP neural network prediction model

According to the above figure it can be concluded that the training gradient of the BP neural network prediction model reaches the level of 10-6 and the training effect of the model is good, and according to Figure 6 it can be concluded that the overall goodness of fit of the model is 0.87 and the training effect, testing effect, fitting effect, and prediction effect of the model is good.

The data from 1900 to 2022 were used as the training data set and imported into the BP neural network temperature prediction model, by taking 2021 to 2101 as the predicted object, the predicted results are shown in the following figure.

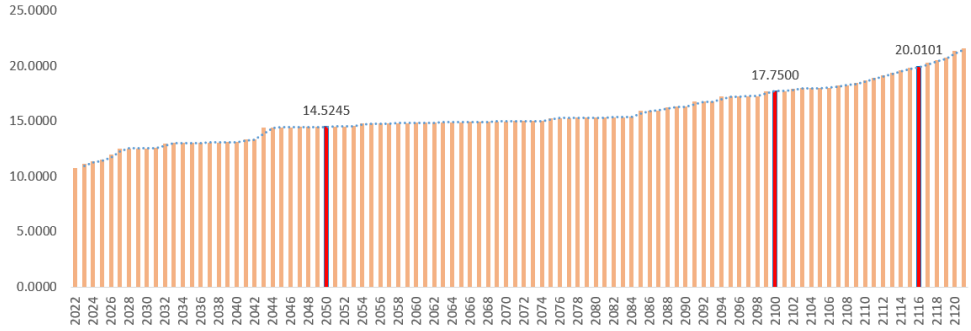

Figure 7. BP neural network temperature prediction model results

Analyzing the projected data in Figure 7, we can conclude that the temperatures in 2050 and 2100 are 14.52 and 17.75, respectively, neither of which is equal to 20°C. Therefore, we do not agree with the prediction that the average temperature of the global observation points will reach 20.00°C in 2050 or 2100. By our continued prediction, the data of 20.0101°C in 2116 was obtained, i.e., the year in which the average temperature of 20°C was reached was in 2116.

3.2.2 Grey Temperature Prediction Model Based on Adaptive Genetic Algorithm

(I) The Process of Gray Temperature Prediction Model

Gray forecasting refers to the use of the GM (1, 1) model to estimate and predict the development of the system's behavioral characteristics, as well as to estimate and calculate the moment of occurrence of anomalies in behavioral characteristics, and to study the future temporal distribution of events occurring within a specific time zone, etc. In essence, these works treat "stochastic processes" as "gray processes" and "random variables" as "gray variables" and are mainly treated by the GM (1, 1) model in gray system theory. The temperature growth process in this question is a black box process, because the factors affecting the temperature growth are unknown, so according to the characteristics of the data in this question there is a gray process, the construction of gray temperature prediction model for the relevant predictions.

(1) Gray temperature prediction model based on the level ratio test

First of all, in order to ensure the feasibility of the modeling approach, it is necessary to do the necessary test processing on the known data columns. Using the data of the sample (130 data in total) as the original reference, the data set the reference data as $x^{(0)}=\left(x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(132)\right)$ to calculate the series of series (Yang H-L, Liu J-X & Zheng B., 2011)

$$\lambda(k)=\frac{x^{(0)}(k-1)}{x^{(0)}(k)},k=2,3,\cdots,132\triangleright \quad (12)$$

If all the level ratios $\lambda(k)$ fall within the admissible coverage

$$\Theta=\left(e^{-\frac{2}{133}},e^{\frac{1}{133}}\right)=(-0.0095,0.0091),$$

the series $x^{(0)}$ can be used as data for model $GM(1,1)$ for gray prediction. Otherwise, the necessary transformations need to be done on the series $x^{(0)}$ to make it fall within the admissible cover. That is, take the appropriate constant c and make a translation transformation

$$y^{(0)}(k)=x^{(0)}(k)+c,k=1,2,\cdots,132 \quad (13)$$

Then the series $y^{(0)}=\left(y^{(0)}(1),y^{(0)}(2),\cdots,y^{(0)}(n)\right)$ such that the series ratio

$$\lambda_y(k)=\frac{y^{(0)}(k-1)}{y^{(0)}(k)}\in\Theta,k=2,3,\cdots,\mathrm{n} \quad (14)$$

Importing the attached temperature data and using MATALB to calculate $\lambda_y(k)$ = 0.007125 while $\lambda_y(k)\in\Theta$, a gray prediction model can be built.

(2) Establishment of temperature prediction model based on gray process

By building model $GM(1,1)$ equation (16), the predicted values can be obtained

$$\hat{x}^{(1)}(k+1)=\left(x^{(0)}(1)-\frac{b}{a}\right)e^{-ak}+\frac{b}{a},k=0,1,\cdots,n-1,\cdots \tag{15}$$

And

$$\hat{x}^{(0)}(k+1)=\hat{x}^{(1)}(k+1)-\hat{x}^{(1)}(k),k=1,2,\cdots,n-1,\cdots \tag{16}$$

(II) Adaptive Genetic Algorithm Optimized Gray Prediction Model

The temperature growth process in this question is a black box process, because the factors affecting the temperature growth are unknown, so according to the data in this question there are characteristics of the gray process, because the need to predict longer data, while there is variability in the data, the application of the prediction method of the gray process, the prediction has limitations, while the lack of data correction, will lead to a sharp decline in prediction accuracy. In order to increase the accuracy of prediction and achieve accurate prediction, and to achieve data correction and training, we apply the self-searching and self-learning functions of the adaptive genetic algorithm to optimize the prediction algorithm in the gray range, i.e., to construct a gray temperature prediction model based on the optimization of the adaptive genetic algorithm.

(1) Optimization process

Genetic algorithms have inherent implicit parallelism and better global search capability; they use a probabilistic search method that automatically acquires and guides the optimized search space, adaptively adjusts the search direction, and does not require defined rules. Using the exact algorithm will not be able to compute the global optimal solution in an acceptable amount of time. The adaptive genetic algorithm has good convergence and is a global optimization algorithm with less computational time for the same computational accuracy. Also combined with the mean iteration, it will get better results. The optimization process is mainly to iterate the gray space after the variation operation, and derive the optimal value in the gray space of the current process, and the flow chart of the optimization is shown below.

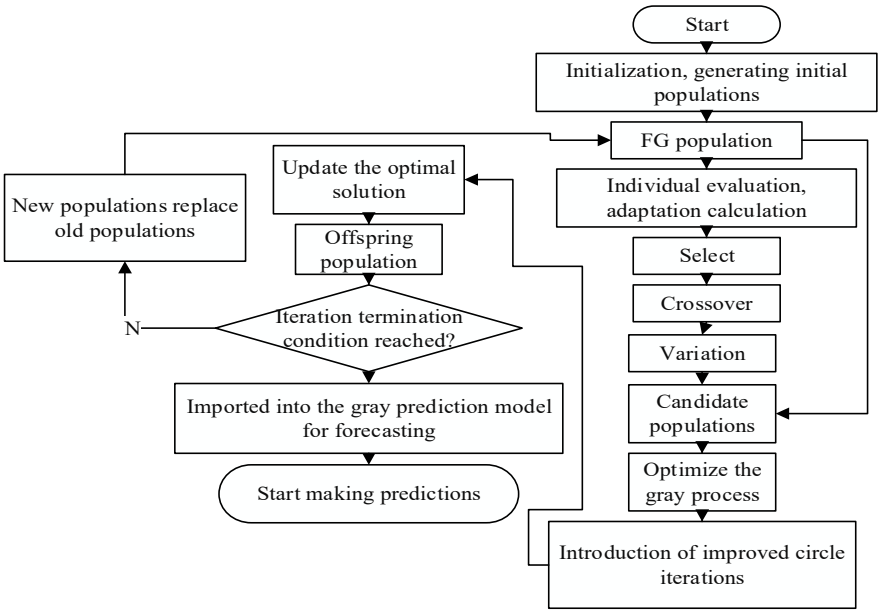

Figure 8. Optimized gray temperature prediction process based on adaptive genetic algorithm

(2) Creation of genetic algorithm optimization model

According to the flow in Figure 8, the process for performing the optimization of the specific design algorithm is shown below.

Step 1: Individual coding design

The pre-processed data were imported, the retail transaction price of used cars was used as the dependent variable and the rest of the factors were used as dependent variables, and then the

data were categorized for individual coding, which was done by MATLAB.

Step 2: Initial population generation

In temperature generation, all constraints in the model need to be satisfied. If an individual solution is generated randomly, the possibility of that individual becoming a feasible solution is small, so it is controlled by constraints during the generation of the initial population to ensure that it becomes a feasible solution.

Step 3: Roulette Selection

The basic idea of roulette selection is that the better the survivability the greater the probability that an individual will be selected, a random sampling method with put-back. Roulette selection calculates the probability of occurrence of an individual in the population based on the fitness of the individual, and randomly selects individuals according to the probability of occurrence to form the self-band. The steps are as follows.

I. Calculate the fitness of each individual $fitness(i)$, $i = 1, 2, \ldots, popsize$

II. Calculate the probability of an individual being inherited to the next generation:

$$p_i = \frac{fitness(i)}{\sum_{j=1}^{poprize} fitness(j)} \tag{17}$$

III. Calculate the cumulative probability of an individual.

$$p_i = \frac{\sum_{k=1}^{i} fitness(k)}{\sum_{j=1}^{poprize} fitness(j)} \tag{18}$$

IV. A random number r in the interval [0,1] is randomly generated and if $P_{i-1} \le r \le P_i$, individual i is selected. The cumulative probability of occurrence in each generation of the population is used to select individuals into the next generation population by generating a random number. This makes the design of the fitness function important and directly uses the objective function values.

$$z = \left\| \overrightarrow{c_1(t)} - \overrightarrow{c_2(t)} \right\|_2 \tag{19}$$

As a fitness function, the best offspring are screened into the genetic operation based on the fitness function value.

Step 4: Single point cross

Crossover operation means that two individuals are randomly selected from the population and the superior characteristics of the parent string are passed on to the offspring string by exchanging the combination of the two chromosomes, thus producing new superior individuals. Since individuals are coded in real numbers, the crossover operation is performed using the real number crossover method by first searching and determining whether there are two chromosomes in the Sizepop range with the same coding feature for the gene fragment, and if so, performing the crossover operation with the kth chromosome vk and the lth chromosome vl at position j as:

$$v_{kj} = v_{ij}(1-b) + v_{lj}b \tag{20}$$

$$v_{lj} = v_{lj}(1-b) + v_{kj}b \tag{21}$$

where b is a random number in the interval [0,1].

Step 5: Variation processing

The main purpose of the mutation operation is to maintain the diversity of the population. The mutation operation randomly selects an individual from the population, searches for and determines whether there is a gene segment in the Sizepop range with the same coding characteristics of the chromosome, and if so, mutates the randomly generated coding position in these individuals to produce a better individual. The operation of the jth gene vij of the ith individual for mutation is.

$$v_{ij} = \begin{cases} v_{ij} + (v_{ij} - v_{\max}) f(g), r \ge 0.5 \\ v_{ij} + (v_{\min} - v_{ij}) f(g), r \ge 0.5 \end{cases} \tag{22}$$

Where, v_{max} is the upper limit of gene v_{ij}; v_{min} the lower limit of gene v_{ij}; $f(g) = r_2(1 - g / G_{\max})^2$, r_2 is a random number, g is the current iteration number, Gmax is the maximum evolution number, and r is a random number in the interval [0,1].

Step 6: Elite retention

In the resulting offspring population, the individual with the highest fitness is selected for comparison with the current optimal individual (elite), and the higher fitness becomes the new

optimal individual (elite).

Step 7: Termination Conditions

If the maximum number of generations (maxgen) is satisfied, the algorithm stops, otherwise it returns to Step 2.

Step 8: Global loop

At each moment, the algorithm flow of Step 2-Step 7 is called in turn, and the dynamic planning process completed to output the optimal value.

Step 9: Introduce the grey prediction model

Programming is performed simultaneously with the genetic algorithm output direct gray prediction model.

3.2.2.1 GAGM Model Prediction

MATLAB was applied to program the model, and the data from 1900 to 2012 were used as the training data set, and the data from 2012 to 2022 were used as the test data set and imported into the GAGM temperature prediction model, and the predicted results are shown below.

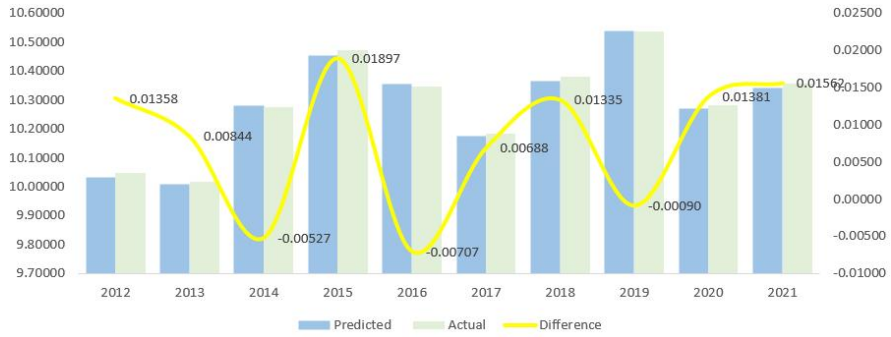

Figure 9. GAGM temperature prediction model test results

According to Figure 9, the fit of the GAGM temperature prediction model is 0.955, and the maximum value of its prediction error is 0.018°C.

The data from 1900 to 2022 were used as the training data set and imported into the GAGM temperature prediction model. Since the GAGM model has self-learning and self-searching functions, it can correct the predicted data and incorporate them into the training database of the prediction, and the prediction results from 2021 to 2101 were used as the prediction object as shown in Figure 10.

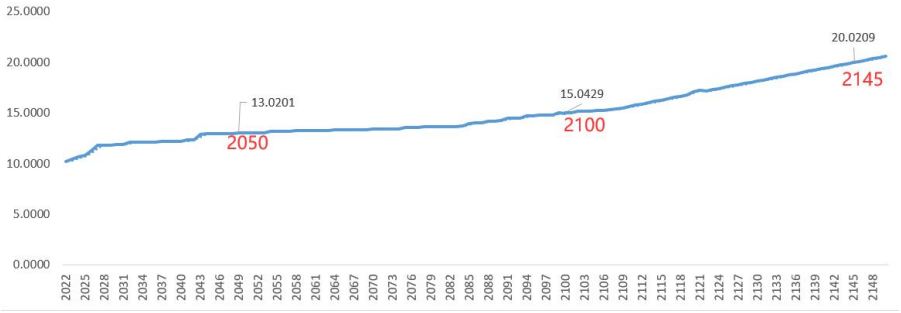

Figure 10. GAGM prediction results

Analyzing the predicted data in Figure 3.3.2.3, we can conclude that the temperatures in 2050 and 2100 are 13.02 and 15.4, respectively, neither of which is equal to 20°C. Therefore, we do not agree with the prediction that the average temperature of the global observation points will reach 20.00°C in 2050 or 2100. By continuing our prediction, we arrive at 20.0209°C in 2146, which means that the year in which the average temperature of 20°C was reached was in 2146.

3.2.3 Construction of the FAEM Evaluation Model

We developed an impact temperature scoring model based on FAEM. After data preprocessing,

the above time and location data are taken as input data, and the scores affecting temperature are taken as output data. By iteratively and continuously fitting the residuals of the current model, the weak learners are merged into strong learners. FAEM has several steps, such as second-order Taylor expansion of the objective function, calculation of leaf node values and tree complexity expression. The following (Zhu Yangbao, 2020) will analyze its creation process in detail.

Assuming that there are n samples, then the data set with m-dimensional features for each sample is $D=\{(x_i,y_i)\}\left(|D|=n,x_i\in\Re^m,y\in\Re\right)$. The prediction function using a combination of K basis functions is.

$$\hat{y}_i=\phi(X_i)=\sum_{k-1}^{K}f_k(X_i),f_k\in F \tag{23}$$

Where $F=\left\{f(X)=w_{q(x)}\right\}(q:\Re^m\to T,w\in\Re^T)$ is the function space containing all regression trees, q represents the mapping from the sample space to the leaf nodes, i.e., the structural parameters of the regression tree; w_i represents the value of the ith leaf node, T is the total number of leaf nodes of the regression tree, and each weak learner f_k has its own independent tree structural parameters q and complete w of leaf nodes. Given a sample point x_i, the order is to divide the sample points into different leaf nodes using the K regression tree rule (determined by q), and the w values obtained in each regression tree are summed to obtain the predicted value $\hat{y}_i$.

Input: training data set $D=\{(x_i,y_i)\}\left(|D|=n,x_i\in\Re^m,y\in\Re\right)$ with iteration number L.

Output: the trained FAEM algorithm model.

Step 1: Initialize the model

Initialize the model to a constant.

Step 2: Tree structure iteration function

Use equation (24) as the objective function $(t\geq 1)$ for the optimization of step t, i.e., the scoring function for the tree structure.

$$Q_{bj}^{(t)}=\frac{1}{2}\sum_{j=1}^{T_r}\frac{G_j^2}{H_j+\lambda}+\gamma T_r \tag{24}$$

Then the forward distribution algorithm is applied to calculate and improve it, and after the tth iteration, the objective function is expressed as

$$Q_{bj}^{(t)}\approx\sum_{i=1}^{N}l(y_i,\hat{y}_i^{t})+\Omega(f_t)=\sum_{i=1}^{N}l(y_i,\hat{y}_i^{t-1}+f_t(a_i))+\Omega(f_t) \tag{25}$$

Where the former is the loss function, the latter is the complexity, and $f_t(a_i)$ denotes the tth iteration is the new tree generated.

Step 3: Taylor expansion calculation

A Taylor expansion of equation (25) yields the following equation.

$$\begin{cases}Q_{bj}^{(t)}\approx\sum_{i=1}^{N}\left[l(y_i,\hat{y}_i^{t-1})+g_if_t(x_i)+\frac{1}{2}h_if_t^2(x_i)\right]+\Omega(f_t)\\ g_i=\partial_{\hat{y}_i^{(t-1)}}l\left(y_i,\hat{y}_i^{(t-1)}\right)\\ h_i=\partial^2_{\hat{y}_i^{(t-1)}}l\left(y_i,\hat{y}_i^{(t-1)}\right)\end{cases} \tag{26}$$

Where g_i and h_i are the first- and second-order partial derivatives of the error terms, respectively. Continuing the simplified calculation of equation (26) leads to.

$$\begin{cases} f_t(a_i) = \sum_{j=1}^{T}\left[\left(\sum_{i\in I_j} g_i\right)w_j + \frac{1}{2}\left(\sum_{i\in I_j} h_i + \lambda\right)w_j^2\right] + \gamma T \\ H_j = \sum_{i\in I_j} h_i \\ G_j = \sum_{i\in I_j} g_i \end{cases} \tag{27}$$

According to equation (27) the extreme value point can be calculated.

$$\lambda_j^* = -\frac{G_j}{H_j + \lambda} \tag{28}$$

The target value is then shown in equation (28).

Step 4: Enumeration calculation

Enumerate different tree structures such that the tree with the smallest objective function, i.e., the optimal structure, is added to the model, where the leaf node of each tree splits only at the best splitting point of the best splitting feature, and both the best splitting point and the best splitting feature are determined by the splitting gain, which is calculated as shown in equation:

$$Gain = \frac{1}{2}\left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} + \frac{G_L + G_R}{H_L + H_R + \lambda} - \gamma\right] \tag{29}$$

Where G_L , H_L and G_R , H_R are the first-order derivatives and second-order derivatives of the left and right nodes, respectively. The fraction of the left and right nodes after the split minus the fraction before the split is the split gain. That is, the performance of the learner is evaluated based on the splitting gain, and if the gain is positive, the splitting will improve the model performance; if the gain is negative, the splitting will be stopped.

The FAEM evaluation model is constructed according to the above steps, and it focuses on the evaluation of indicator relationships by means of scoring. The rating interval is (-1, 1), where the closer the rating is to -1, the more it shows a negative correlation, the closer it is to 1, the more it shows a positive correlation, and the closer it is to 0, the more it shows no correlation. Then the model can be derived as shown in equation (30) below.

$$\hat{\zeta}_i = \sum_{t=1}^{l} f_t(a_i) \tag{30}$$

3.2.3.1 Analysis of the Relationship Between Time, Position and Temperature

Applying MATLAB to program the model and importing the relevant data for calculation, the values of the model can be derived as shown in the table below.

Table 2. FAEM influence temperature scoring model results

Indicators	East longitude	Western Longitude	Latitude North	South latitude	Time
ζ_i	0.881	-0.825	0.874	-0.922	0.941

According to Table 2, it can be concluded that the influence temperature scores of east longitude, west longitude, north latitude, south latitude, and time are 0.881, -0.825, 0.874, -0.922, and 0.941 respectively, indicating that they all have an extremely strong correlation with temperature, and the degree of bit height is negatively correlated with temperature, and when the degree of bit height is greater, the temperature is lower; the location relationship presents a geographical gap, when east longitude (north latitude), temperature is positively correlated when the east longitude (north latitude), and the west longitude (south latitude) temperature shows a negative correlation, and with the rotation of time (season) its positive and negative relationship will also rotate back, that is, it can be concluded that when the location of different performance of the temperature is different, when the time rotation, the temperature of its location will also change with who.

In summary, we can conclude that there is a correlation between location, time and temperature, with the same time and different temperatures at different locations, and the same location and different temperatures at different times.

3.2.3.2 Influence of Natural Factors

Firstly, the effects of natural factors such as volcanic eruptions, forest fires and COVID-19 need to be analyzed, the results of which are shown in Table 3 below.

Table 3. Analysis table of the results due to natural factors

Natural Factors	Impact
Volcanic eruptions	A large amount of greenhouse gases, mainly CO2, will be produced.
Forest fires	Reduces vegetation cover and thus carbon sequestration, while fires burn a lot of vegetation and produce a lot of CO2
COVID-19	Reduced traffic flow, resulting in reduced CO2

The analysis in Table 4 leads to the conclusion that the main factor of the influence of the three natural factors is CO2, and they will cause the CO2 concentration to become larger.

In summary, it can be concluded that the most important factor of the above-mentioned natural factors affecting temperature is CO2 concentration, so we need to find out the CO2 concentration data and temperature data from 1960 to 2020 for analysis. The obtained data set was substituted into the FAEM model for calculation, and the evaluation score of CO2 concentration was calculated as 0.967. Then it can be concluded that the CO2 concentration is positively correlated with the temperature increase, and the correlation factor is 0.967. Then it can be calculated that when the CO2 concentration increases by 4.5 ppm, the temperature increases by 0.1°C.

In summary, it can be concluded that natural factors such as volcanic eruptions, forest fires, and COVID-19 mainly affect CO2 concentration and thus temperature, and when CO2 concentration increases by 4.5 ppm, the temperature increases by 0.1°C.

3.2.3.3 Analysis of Factors Affecting Global Temperature

Factor Acquisition

A total of nine factors affecting the temperature variation were obtained from the references, as shown in the following table.

Table 4. Factors affecting temperature change

Serial number	Indicators	Abbreviations
1	CO2 concentration	COND
2	Energy Extraction	NYKC
3	Energy consumption	NYXH
4	GDP	GDP
5	Average increase in population density	PJRK
6	Unit greenery area	LZMJ
7	Precipitation	JSL
8	Technology level	KJSP
9	National education level	GJJY

The 1960-2021 data of nine global indicators, including COND, NYKC, NYXH, GDP, PJRK, LZMJ, JSL, KJSP, and GJJY, were obtained from open-source data sites and references, and the nine indicators were used as independent variables and the global average annual temperature was used as the dependent variable to construct the FAEM model imported from the above construction.

Solution to the factor relationship

First of all, the acquired data will be visualized and analyzed. Data standardization (normalization) processing is a basic work of data mining, and different evaluation indicators often have different levels and units of measurement, and such a situation will affect the results of data analysis. The indicators in the annexes have different data outlines. In order to eliminate the influence of the outline between indicators, data standardization is required to address the comparability between data indicators. After the raw data are processed by data standardization, the indicators are in the same order of magnitude and suitable for comprehensive comparative evaluation. The following are two commonly used normalization methods.

(1) Min-max normalization (Min-Max Normalization)

Also known as deviation normalization, it is a linear transformation of the original data so that the resultant values are mapped between [0-1]. The transformation function is as follows.

$$x_{ij} = \frac{\max\left\{x_{1j},...,x_{nj}\right\} - x_{ij}}{\max\left\{x_{1j},...,x_{nj}\right\} - \min\left\{x_{1j},...,x_{nj}\right\}} \tag{31}$$

Where max is the maximum value of the sample data and min is the minimum value of the sample data. One drawback of this method is that when new data are added, it may lead to changes in max and min, which need to be redefined.

(2) Z-score normalization method

This method normalizes the data by giving the mean and standard deviation of the original data. The processed data conform to the standard normal distribution, i.e., the mean value is 0 and the standard deviation is 1. The transformation function is

$$x^* = \frac{\left(x - \mu\right)}{\sigma} \tag{32}$$

Where μ is the mean of all sample data and σ is the standard deviation of all sample data.

In this paper, the Z-score standardization method is selected to apply equation (32) to standardize the 1960-2021 data sets of nine global indicators, such as COND, NYKC, NYXH, GDP, PJRK, LZMJ, JSL, KJSP, and GJJY, respectively, and the standardization results are calculated by applying the function of Excel pivot to them. The standardized results are shown in Figure 11 below.

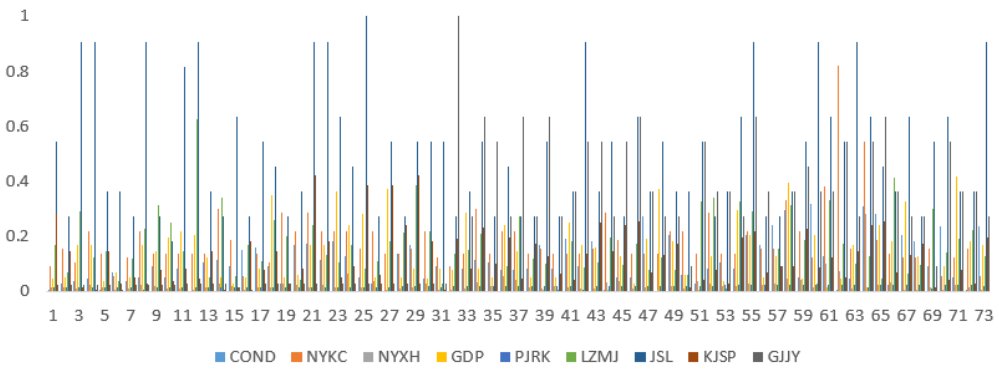

Figure 11. Standardization results

MATLAB was applied to construct the model, and the other factor data and temperature data from 1960 to 2021 were substituted into the model, and the output of the FARM model coefficients are shown below.

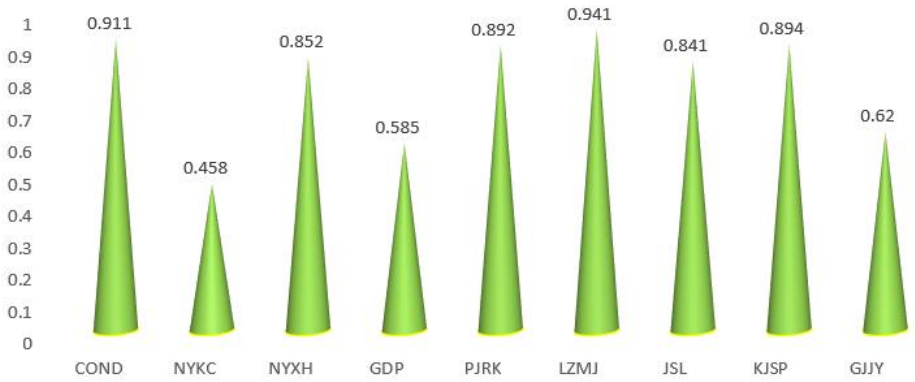

Figure 12. FARM indicator coefficients

From the above figure, it can be seen that COND, NYKC, NYXH, GDP, PJRK, LZMJ, JSL, KJSP, and GJJY yielded their FARM correlation coefficients of 0.911, 0.458, 0.852, 0.585, 0.892, 0.941, 0.841, 0.894, and 0.620, respectively, indicating that COND, NYXH, PJRK, LZMJ, JSL, and KJSP have coefficients greater than 0.8 and have a large correlation, thus indicating a significant positive correlation between COND, NYXH, PJRK, LZMJ, JSL, KJSP, and temperature. In summary, it is concluded that the indicators of COND, NYXH, PJRK, LZMJ, JSL, and KJSP are the main causes of the global temperature change.

3.2.3.4 Measures to Curb Global Warming

Based on the above study, it can be concluded that the main causes of global temperature change are indicators such as COND, NYXH, PJRK, LZMJ, JSL, KJSP, etc. Then for the influence of these factors, we propose effective measures to slow down the temperature increase in Figure 13 below.

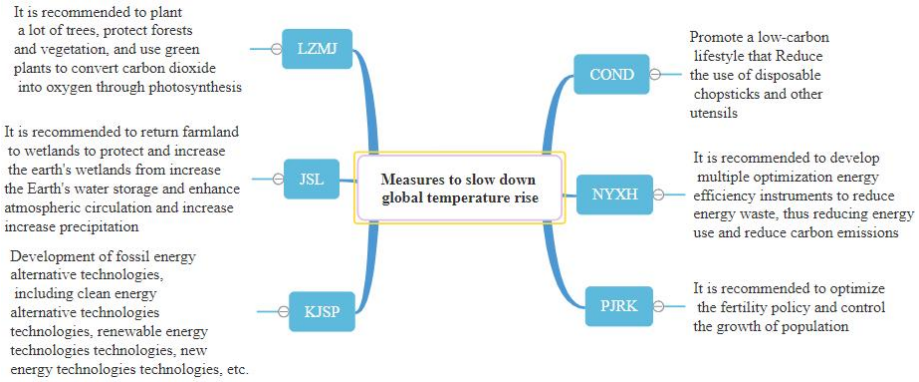

Figure 13. Effective measures to mitigate global warming

From Figure 13, we derive the feasible and effective measures that we use to curb or slow down global temperature warming based on the study above.

3.2.4 Contrast Analysis

3.2.4.1 Model Error Testing

●Residual test:

Let the residual be $\varepsilon(k)$. Calculate

$$\varepsilon(k)=\frac{x^{(0)}(k)-\hat{x}^{(0)}(k)}{x^{(0)}(k)},k=1,2,\cdots,n \quad (33)$$

Here $\hat{x}^{(0)}(k)=x^{(0)}(1)$, if $\varepsilon(k)<0.2$, is considered to meet the general requirements; if $\varepsilon(k)<0.2$, is considered to meet the higher requirements (Zeng Linghui & Zhou Ailian, 2016).

●Grade ratio deviation value test:

Firstly, the step ratio $\lambda(k)$ is calculated from the reference data $x^{(0)}(k-1),x^{(0)}(k)$, and then the corresponding step deviation is found by the development factor a.

$$\rho(k)=1-\left(\frac{1-0.5a}{1+0.5a}\right)\lambda(k) \quad (34)$$

If $\rho(k)<0.2$, the general requirements are considered to be met; if $\rho(k)<0.1$, the higher requirements are considered to be met.

●**Mean Square Error Test:**

$$MSE=\frac{1}{n}\sum_{i-1}^{n}(f(x_i)-y_i)^2 \quad (35)$$

Where n is the number of data samples, $f(x_i)$ is the model prediction, y_i is the ideal optimum (i.e., the reference value), and μ_i is the average of the model predictions. The smaller the value of MSE, the smaller the error between the predicted value of the model and the ideal optimal value, and the higher the prediction accuracy.

The values of the two models were 0.1580, 0.192, 0.113; 0.087, 0.074, 0.095, respectively, by MATLAB's residual and cascading deviation tests and mean square error test. It shows that the BP neural network temperature prediction model achieves a general level of prediction accuracy and its fit is high, while the GAGM temperature prediction model achieves a high level of prediction accuracy and its fit is relatively high.

3.2.4.2 Comparative Analysis of Models

Combining the above prediction data as well as the error test, the comparison of the two models is shown in the following table.

Table 5. Table of model comparison results

Models	$\rho(k)$	$\varepsilon(k)$	MSE	R^2	Effect
BP	0.158	0.192	0.113	0.87	Butter
GAGM	0.087	0.074	0.095	0.96	Generally

According to Table 5, it can be concluded that the grade deviation, residual, and mean square error of the GAGM model are 0.087, 0.074, and 0.095, respectively, which are smaller than BP, and the GAGM model achieves a high level of goodness of fit, which can fit the past data perfectly and predict the future data accurately. In summary, it can be concluded that the GAGM temperature prediction model outputs the most accurate prediction values.

4. Conclusion

We obtained by constructing a first-order differential temperature model the analysis yielded a larger temperature rise in March 2022 than that observed in the last 10 years. Secondly, according to the data, we build the prediction model of BP (BP) based on BP neural network and the gray temperature prediction model (GAGM) optimized based on adaptive genetic algorithm. By calculating the imported data, we conclude that the average temperature of the global observation points reached 20℃ in 2050 and 2100, and the continuous prediction year reached 20℃ in 2116 and 2145 respectively. Further, the model conducted error testing and found that the rank deviation, residual and mean square error were 0.087, 0.074 and 0.095, less than BP, and the GAGM model had a goodness of fit of 0.96, which could fully fit past data and accurately predict future data. Afterwards, we conclude that the GAGM temperature prediction model outputs the most accurate prediction of the (Zhu Yangbao, 2020). According to the national location data, and then to the position of each country time as the independent variable, temperature as the dependent variable, temperature factor correlation evaluation model (FAEM) construction, we conclude that the correlation between location, time and temperature, the same time, different temperature is different, the same position, different time temperature is different. Next, we analyzed key environmental factors such as volcanic eruptions, forest fires, and COVID-19. The carbon dioxide concentration and temperature data were imported into the FAEM evaluation model, and the correlation coefficient was 0.967. When the carbon dioxide concentration increased by 4.5 ppm, the temperature increased by 0.1 C, that is, natural disasters will cause the temperature increase. Furthermore, nine factors affecting environmental factors, including COND, NYKC, NYXH, GDP, PJRK, LJMJ, JSL, KJSP, GJJY, and these factors as independent variables and

temperature as dependent variables were imported into the FAEM model. By comparing the magnitude of the correlation coefficient, it is concluded that COND, NYXH, PJRK, LZMJ, JSL, and KJSP are the main causes of global temperature change. Finally, the model results are summarized, and suggestions for optimizing the energy structure and planting trees to effectively alleviate the global temperature rise are proposed.

5. Model Evaluation

5.1 Advantages of the Model

1). Adaptive genetic algorithm has good convergence and is a global optimization algorithm with less computational time under the same computational accuracy. The combination of mean iteration at the same time will give better results. So we apply adaptive genetic algorithm to optimize the exponential smoothing prediction model and the gray prediction model, first apply the genetic algorithm model to iteratively analyze the data to derive the best input weights for each of its factors and output the global optimal value, square then facilitate the gray prediction to build the network so that the training results and the prediction accuracy can reach a high level.

2). The constructed FAEM model divides the sample data into N boxes according to their ranges, accumulates the gradient gain values in the boxes, calculates the gradient gain for each candidate segmentation point in the boxes, and saves the accumulated gradient and the number of samples in the boxes for each box separately. By means of histogram, although the accuracy of the segmentation becomes worse, it has little effect on the final result. On the one hand, it can improve the computational efficiency, and on the other hand, this coarser segmentation point can play a kind of regularization effect. The final better global iteration was performed to derive the factor influence temperature scores.

Fund Project

This research was financially supported by Higher Education Research Project of Chongqing Higher Education Association (CQGJ21B058), Chongqing Higher Education Teaching Reform Research Project (223280), Chongqing University of Technology Undergraduate Education Teaching Reform Key Research Project (2121ZD02).

References

C. Maji, B. Sadhukhan, S. Mukherjee, S. Khutia and H. Chaudhuri. (2021). Impact of climate change on seismicity: a statistical approach, *Arab J Geosci, 14*(24), 2725.

Chou, C. H. (2016). *Machine learning*. Beijing: Tsinghua University Press.

Deng Julong. (1986). *Gray prediction and decision making*. Huazhong University of Technology Press.

F. Becker and Z. Li. (1995). Surface temperature and emissivity at various scales: Definition measurement and related problems, *Remote Sens. Rev., 12*(3), 225-253.

F. Jacob, A. Olioso, X. F. Gu, Z. Su and B. Seguin. (2002). Mapping surface fluxes using airborne visible near infrared thermal infrared remote sensing data and a spatialized surface energy balance model, *Agronomie, 22*(6), 669-680.

J. Huang, J. Zhang, Z. Zhang, C. Y. Xu, B. Wang and J. Yao. (2011). Estimation of future precipitation change in the Yangtze River basin by using statistical downscaling method, *Stoch. Environ. Res. Risk Assess, 25*(6), 781-792.

J. T. Chu, J. Xia, C. Y. Xu and V. P. Singh. (2010). Statistical downscaling of daily mean temperature pan evaporation and precipitation for climate change scenarios in Haihe River China, *Theor. Appl. Climatol, 99*(1-2), 149-161.

Liu Xingdong and Yang Xiwu. (2007). Analysis of temperature impact to track on asphalt pavement, *Technology of Highway and Transport*, (3), 66-69.

Ma Jian, Liu Zhu, Shi Yurong and Luo Bingxian. (2021). Zheng Feng Multimodal solar flare prediction model based on sunspot group data, *Spacecraft environmental engineering, 38*(03), 256-262.

Ma Jian, Liu Zhu, Shi Yurong and Luo Bingxian. (2021). Zheng Feng Multimodal solar flare prediction model based on sunspot group data, *Spacecraft environmental engineering, 38*(03), 256-262.

R. M. Hirsch and J. R. Slack. (1984). A Nonparametric Trend Test for Seasonal Data with Serial Dependence, *Water Resources Research, 20*(6), 727-732.

Takashi Asaeda, Vu Thanh Ca and Akio Wake. (1996). Heat storage of pavement and its effect on the lower atmosphere, *Atmospheric Enuiromnent*, *30*(3), 413-427.

Wang Qing. (2011). A study of some key issues in integrated learning. Shanghai: Fudan University.

Wang Yunjie and Ma Jun. (2015). Analysis and prediction of sunspot time series based on Wavelet and BP neural network, *Journal of Qinghai Normal University (NATURAL SCIENCE EDITION)*, *31*(02), 10-13.

Xia Huiwei. (2020). Research on network traffic analysis and prediction based on decision tree integration and width forest. Nanjing University of Posts and Telecommunications.

Xiao Yang. (2016). Coal price forecasting model based on co-integration theory and artificial neural network. Xi'an University of Science and Technology.

Yang H-L, Liu J-X, Zheng B. (2011). Improvement and application of the gray prediction GM(1,1) model. *Mathematical Practice and Understanding*, *41*, 39-46.

Zeng Linghui. Zhou Ailian. (2016). Genetic algorithm-based optimization model of urban underground logistics network. *Transportation Science and Engineering*, *3*(22), 84-93.

Zhang Xuan. (2016). Fault diagnosis of motor production line based on genetic algorithm optimized BP neural network. Hefei University of Technology.

Zhu Yangbao. (2020). Multi-factor stock selection scheme design based on XGBoost and LightGBM algorithms. Nanjing University.