ISSN 2616-5872 Vol.5, Issue 3: 42-48, DOI: 10.25236/AJEE.2023.050307

Predictions and Research about Global Warming Based on ARIMA models

Zitiantao Lin[1], Junxian Kuang[2], Wenhua Li[3]

[1]*Civil Aviation Flight University of China, Guanghan, 618307, China*
[2]*Xinjiang University, Urumchi, 830046, China*
[3]*Zhejiang University of Science and Technology, Hangzhou, 310012, China*

Abstract: *Initially, by processing the data of global average temperature change from 1850 to 2022 and comparing it with the temperature data of March 2022, the data were processed to get the preliminary conclusion of data visualization and an ARIMA model and an SVM model were built by doing the time series in this way.This comparative model can predict the global average temperature at different times in the future and the time to reach 20°C in different regions. Also, the errors MSE, MAE, R^2, and EV of the two models were compared by short-term predictions and the ARIMA model was found to be better. To further improve the model, Pearson correlation coefficients of global temperature with time and geographical location were introduced, and it was found that there was no necessary connection between temperature and time and longitude. Considering the negligible effect of Earth's rotation and revolution on temperature, the weak relationship between latitude and temperature is also negligible. According to the literature[1], the outbreak of natural disasters is directly related to the accumulation of greenhouse gases, so greenhouse gas data are used to train the ARIMA and SVM models, which in turn leads to a Pearson correlation coefficient of 0.6 - 0.7 between greenhouse gases and temperature. Synthesizing the above data and literature, it can be concluded that the main factor of global warming is the irrational use of natural resources by human beings, which leads to the disturbance of the earth's ecological environment and the serious deterioration of the greenhouse effect.*

Keywords: *Temperature Prediction, ARIMA Model, SVM model, Pearson Coefficient, Global Warming*

1. Introduction

With the continuous development of industry, the greenhouse effect of the earth intensified, which in turn led to the imbalance in the capacity of the earth's atmospheric system, so that the energy of the atmospheric system is accumulating, and the annual average temperature of the earth is increasing year by year, global warming is becoming more and more serious. Nowadays, the intensification of global warming has also forced many countries to enter a tight state.

According to research, scientists have shown that the concentration of carbon dioxide (CO_2) is directly related to global warming. Prior to the Industrial Revolution, atmospheric CO_2 concentrations remained around 280 parts per million (ppm) throughout the year. However, with the rise of the industrial age, the concentration of CO_2 in the atmosphere reached 377.7 ppm in 2004, the largest average increase in nearly a decade. Therefore, many relevant scientific research institutions launched a study on atmospheric CO_2 concentration to explore how to mitigate the impact of the greenhouse effect, curb global warming. With the deepening of research, some research institutions have asserted:

(1) In May 2022, the average concentration of CO_2 will reach 421 ppm ------ a peak of the year.

(2) In 2050, the concentration of CO_2 will reach 685 ppm.

In the global warming environment, in order to accurately predict the rise of climate temperature, we build a global temperature change prediction model based on latitude, longitude, and time through NASA and data crawled from the whole network. The model is based on deep learning, and the ARIMA model and SVM model are selected to compare the advantages and disadvantages of prediction, and predict the year when the global average temperature reaches 20°C.

ISSN 2616-5872 Vol.5, Issue 3: 42-48, DOI: 10.25236/AJEE.2023.050307

2. Model Use and Methodology

2.1 Data collection and processing

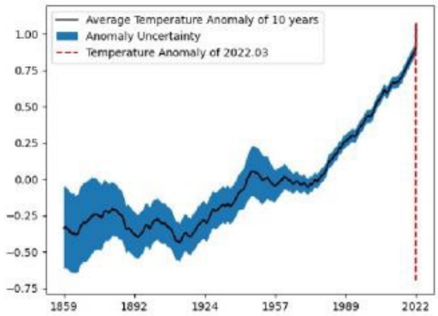
Firstly, export the relevant temperature data from http://berkeleyearth.lbl.gov, divid the data before March 2022 into 10 years as a unit of time, and calculate the global average temperature change and its uncertainty in each unit. From the data, in March 2022, the global sea surface average temperature increased by 1.071°C and its uncertainty is 0.037°C. And the golbal land air temperature increased by 0.888C and its uncertainty is 0.032°C.

The table 1 lists some data (September 2021 to March 2022) of backing material.

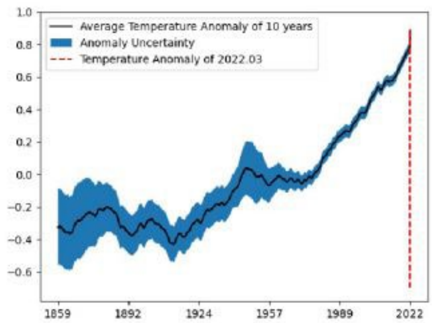
Table 1: The temperature anomaly from September 2021 to March 2022

		sea surface air temperature		land air temperature	
year	month	amplification	uncertainty	amplification	uncertainty
2021	9	0.772417	0.034008	0.881208	0.043325
2021	10	0.775258	0.033992	0.884392	0.043283
2021	11	0.778517	0.034025	0.887417	0.043367
2021	12	0.780742	0.034058	0.889450	0.043383
2022	1	0.784725	0.034042	0.893275	0.043400
2022	2	0.787750	0.034050	0.896192	0.043417
2022	3	0.888000	0.032000	1.071000	0.037000

Obviously, the warming in March 2022 is larger than the previous period. In order to show the temperature change more intuitively, this paper visualizes the data of backing material, as shown in the following Figure 1.



(a) From sea surface temperatures



(b) From land temperatures

Figure 1: Global average temperature anomaly in Sea Ice area

From Figure. 1, fig. 1(a) is the temperature inferred from water temperatures and fig. 1(b) is the temperature inferred from air tempertatures of Sea Ice area. By analyzingthe temperature data of Sea Ice area, the trend of global warming can be obtained. Furthermore, the global temperature is on the rise as a whole. As time goes on, the temperature uncertainty is getting smaller and smaller, and the rise is becoming more and more significant. Until March 2022, the peak of the rise is reached.

2.2 Method Introduction

(1) ARIMA prediction model

The ARIMA model is identified to predict the future values based on the past and present values of the time series, which is simply described by ARIMA (p,d,q). p is the autoregressive term parameter, q is the number of moving average terms, and d represents the number of differences required to make the series a smooth series. The ARIMA model characteristics are described by the three parameters p,d,q.[1]

(2) SVM prediction model

ARIMA is based on a linear approach to fix the order, which can fit well the linear trend of the time series. But on the other hand, prediction can be performed by SVM model. And SVM[2] has obvious advantages in dealing with small samples and nonlinearity, which can circumvent the local minima and overfitting phenomena that are easy to occur in other machine learning algorithms, and has been widely used in the field of temperature prediction.

ISSN 2616-5872 Vol.5, Issue 3: 42-48, DOI: 10.25236/AJEE.2023.050307

3. Model building and solving

3.1 Model building

3.1.1 ARIMA model

Then, we have tried to establish a linear regression model for prediction. However, considering the influence of white noise in the data, therefore, based on the basic multivariate logistic regression, combined with AR regression formula:

$$y_t = \mu + \sum_{i=1}^{p} \gamma_i \cdot y_{t-i} + \varepsilon \tag{1}$$

$\gamma i \cdot yt{-}i + \varepsilon$which can be used to predict data strongly associated with its own historical data. In addition, in order to make full use of historical data and their white noise to predict the future, insert horizontal MA regression formula:

$$y_t = \mu + \sum_{i=1}^{q} \beta_i \cdot \varepsilon_{t-i} + \varepsilon_t \tag{2}$$

into eq. (1). Then establish an ARIMA autoregressive moving average model[3]:

$$y_t = \mu + \sum_{i=1}^{p} \gamma_i \cdot y_{t-i} + \sum_{i=1}^{q} \beta_i \cdot \varepsilon_{t-i} + \varepsilon_t \tag{3}$$

where μ is constant, p and q denote order, γ_i is time parameter, β_i is a two-dimensional vector representing latitude and longitude and ε_t is white noise with time t.

3.1.2 SVM model

SVM[4] is a kind of generalized linear classifier that classifies data in a binary way according to supervised learning. It can transform the problem into a convex quadratic programming problem.

Assume that $\mathrm{f}(x) = \omega x + b$ can predict future temperatures at w. Then the problem of solving high-precision w can be transformed into a quadratic convex programming problem:

$$max\omega$$

$$\text{s.t.} = \begin{cases} y_i - \omega \cdot x_i - b \leq \varepsilon \\ \omega \cdot x_i + b - y_i \leq \varepsilon \end{cases} \tag{4}$$

Then import relaxation variables ξ_i and ξ_i^*, the eq. (4) is converted to

$$\min \frac{1}{2} \left\| \omega \right\|^2 + C \sum_{i=1}^{N} (\xi_i + \xi_i^*)$$

$$\text{s.t.} = \begin{cases} y_i - \omega \cdot x_i - b \leq \varepsilon + \xi_i \\ \omega \cdot x_i + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i , \xi_i^* \geq 0 \end{cases} \tag{5}$$

Where y_i is predicted value, ω is prediction accuracy, x_i is three-dimensional vector with temperature, latitude and longitude, ε is allowable error, C and b is corresponding constant.

3.2 Data training

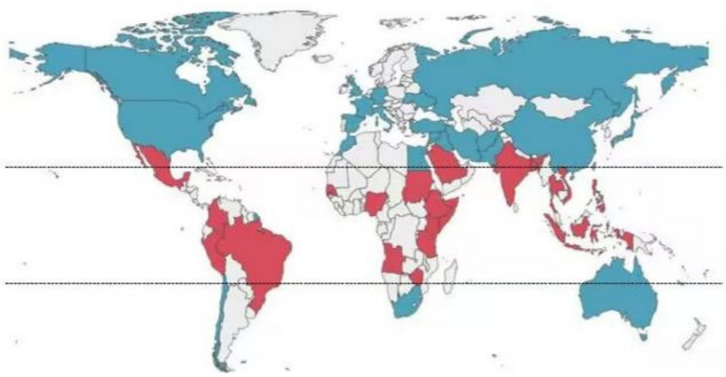


Figure 2: The city covered by the data

By analyzing the longitude and latitude of the dataset after pre-processing the collected data in the

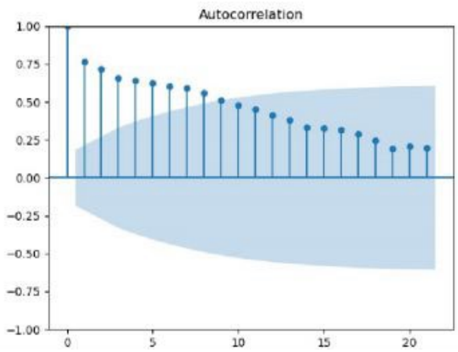
ISSN 2616-5872 Vol.5, Issue 3: 42-48, DOI: 10.25236/AJEE.2023.050307

previous section, we generated a map of the cities covered by the data in Figure 2.

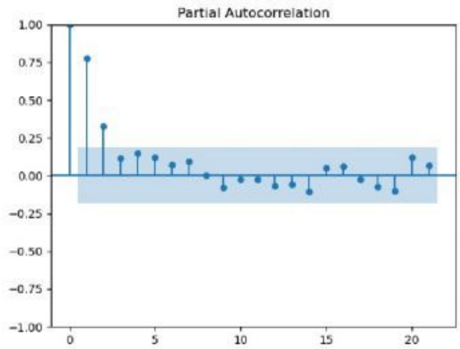
According to the fig. 2, the cities where the data are recorded are geographically unevenly distributed across multiple climate zones. Substituting the data of all cities into the model for training may result in large errors and affect the prediction accuracy. Therefore, we selected nine geographically representative cities (see table 2) in thetropical, subtropical and cold temperate zones according to the distribution of climatic zones, and trained the model with their temperature data.

Table 2: Representative cities of training model

Tropics Zone	Subtropics Zone	Cold Temperate Zone
Singapore	Dhaka	New York
Nairobi	Sao Paulo	Chicago
Mogadishu	Karachi	Toronto
Cali	Kanpur	Montreal
Fortaleza	Santiago	Paris
Abidjan	Cairo	Kiev
Jakarta	Durban	Berlin
Ibadan	Faisalabad	Moscow
Dar Es Salaam	Casablanca	Saint Petersburg



(a) Autocorrelation



(b) Partial autocorrelation

Figure 3: Selection of order of ARIMA

Through model training of the data from 1900 to 2013, The SVM model's training parameter results are C = 10, and the kernel function is radial basis function. According to the autocorrelation and partial autocorrelation detection in figure. 3, the ARIMA model's training parameter results are

p = q = 1 and μ =2. Substituting the above parameters into the model respectively, the corresponding preliminary temperature prediction model is obtained.

3.3 Model analysis and testing

The ARIMA model and SVM model were used to rank and predict the temperature from 2013 to 2022, and the prediction accuracy of the models can be derived when compared with the actual observed values.

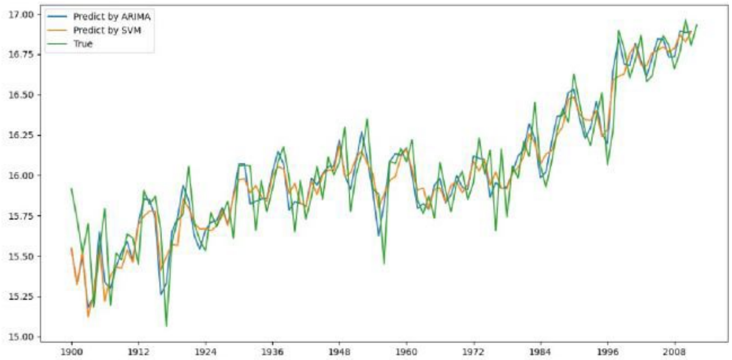


Figure 4: Prediction results of the model

ISSN 2616-5872 Vol.5, Issue 3: 42-48, DOI: 10.25236/AJEE.2023.050307

Figure 4 shows the prediction results of the ARIMA model and SVM model, which visually shows that the prediction results of both models are close to the actual real values, which indicates the effectiveness of both models in temperature prediction.

In addition, the four errors MSE, MAE, R2 and EV were calculated for both models. From Table 3, it can be seen that the ARIMA model has a higher prediction accuracy compared to the SVM model, with an MSE of only 0.000148 and an EV of 0.969808. Therefore, the ARIMA model has a better temperature prediction under that.

Table 3: Prediction error

	ARIMA	SVM
MSE	0.000148	0.000201
MAE	0.011089	0.012089
R2	0.820678	0.756003
EV	0.969808	0.933243

Since temperature is influenced not only by time but also by other factors, for the rigor of the ARIMA model, we introduce greenhouse gases and geographical factors that contribute to the effect of temperature difference into the prediction model. The Pearson correlation coefficient is also used to take these factors into account.

$$\rho_{XY}=\frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y}=\frac{\sum_{i=1}^{n}\frac{(X_i-E(X))}{\sigma_X}\frac{(Y_i-E(Y))}{\sigma_Y}}{n} \tag{6}$$

Combining Cov(X,Y) in terms of geographic location is the population covariance of X and Y, σ is the standard deviation, and E is the expected value. A scatter plot of temperature variation with time and geographic location is plotted using the model.

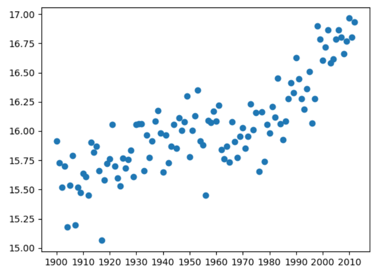
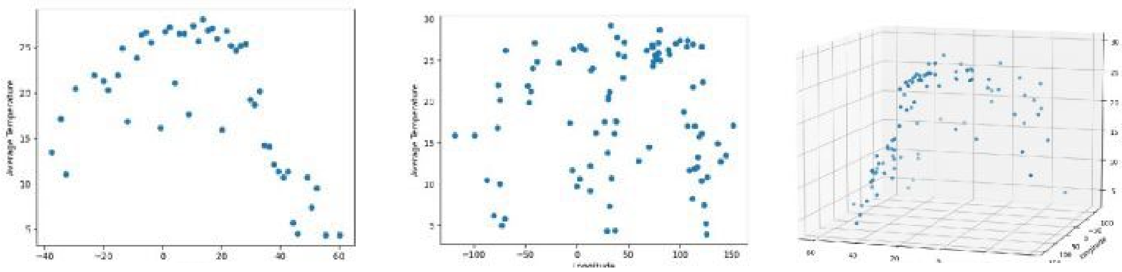


Figure 5: Temperature and time

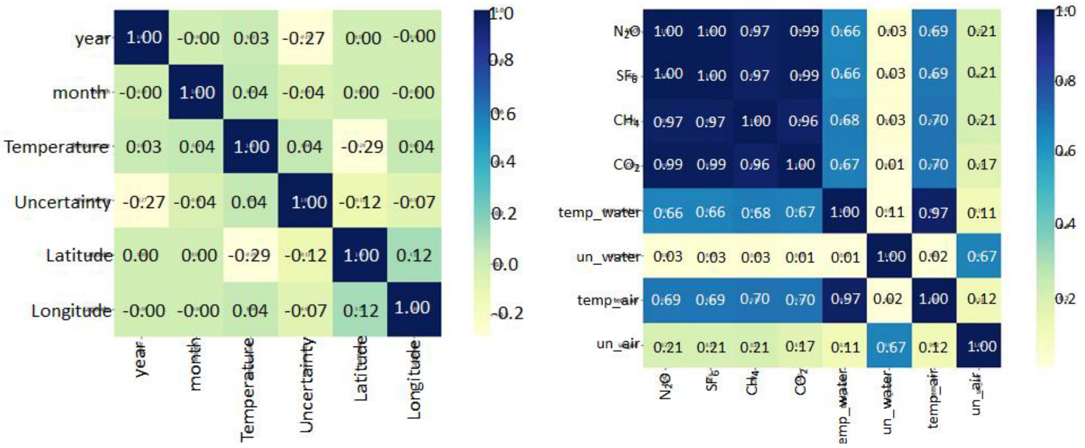


(a) Temperature and latitude (b) Temperature and longitude (c) 3D scatter plot

Figure 6: Temperature and location

The figure. 5 is a scatter plot of temperature changing with time. It can be seen that the correlation between temperature and time is weak. The figure. 6 is a scatter plot of temperature changing with geographical location, where fig. 6(a) represents the relationship between temperature and latitude, and fig. 6(c) represents the relationship between temperature and longitude. It can be seen that the temperature is geographically only correlated with latitude, and the correlation is weak. In addition, the eq. (6) is used to solve the Pearson correlation coefficient between these variables (see fig. 7(a)).

ISSN 2616-5872 Vol.5, Issue 3: 42-48, DOI: 10.25236/AJEE.2023.050307



(a) Temperature and time and geography *(b) Temperature and greenhouse gases*

Figure 7: Pearson correlation coefficient

It can be seen from figure. 7(a) that there is a weak correlation $\rho = -0.29$ between temperature and latitude, and almost no correlation with time and accuracy, whose correlation coefficient are only $\rho = 0.035$ and $\rho = 0.04$.

In the figure, we selected the four most representative greenhouse gases N_2O, SF_6, CH_4 and CO_2 for the analysis. By analyzing the data of these four gases and visualizing the data and bringing them into the ARIMA model to solve the Pearson correlation coefficient matrix, the output from the data in Fig. 7(b), compared with Fig. 7(a), shows that (a), the influence factors in our model are very low in relation to latitude and longitude, but the greenhouse gases have a large weight on the temperature model, with an average influence parameter of 0.6-0.7 between. The correlation between temperature and N_2O is $\rho=0.66\pm0.03$. The correlation between temperature and CH_4 is $\rho=0.68\pm0.03$. While the correlation between temperature and CO_2 is $\rho=0.67\pm0.01$.

With the collected GHG data, the model is further trained using the ARIMA model. We found that the effect of greenhouse gases on temperature increase (greenhouse effect) is much higher than the normal case. Its change as an influence factor relative to geographic location or time, the change in perturbation temperature occupies the main part of the overall influence coefficient.

Also corroborating the literature, before the industrial revolution, the mass fraction of greenhouse gases represented by carbon dioxide The mass fraction of greenhouse gases represented by carbon dioxide was only 275 ppm, but after the industrial revolution, the concentration of carbon dioxide has reached a staggering 414.35 ppm, and the content of greenhouse gases in the air has increased by an average of 1.0 - 1.2 ppm per year, and the most CO2, the main constituent gas of greenhouse gases, also shows an average increase of 0.3 ppm per year. At the same time, the global temperature has increased by 1-1.3 °C in just 50 years after the Second World War, with an observed warming of (1.1 ± 0.1) °C in 2019 [5].

4. Conclusions

In this study, firstly, by using the global temperature situation data from 1850 to 2020, the ARIMA model is applied to construct a prediction model about the future temperature under the time series, while the output of this model is compared with the prediction result of SVM model, and the advancedness of ARIMA model is further determined by MSE, MAE, R2 and EV.

In fact, temperature changes are not only influenced by time. In order to further establish the influence of the influence factors in the ARIMA prediction model on the predicted temperature situation and improve the accuracy and fitness of the model, factors such as geographic location and air composition content are introduced to find out the influence situation to which the model is subjected by Pearson correlation coefficient.

After the above study, this model can find the main factors affecting the global warming from the mathematical point of view, and at the same time, it can be helpful in predicting the temperature and climate development in various regions of the world, and help governments and environmental protection agencies to play a certain role in detecting the temperature anomalies caused by temperature

ISSN 2616-5872 Vol.5, Issue 3: 42-48, DOI: 10.25236/AJEE.2023.050307

increase and human activities.

References

[1] Tan B, Zhang Z, Zhang Y. Forecasting method of e-commerce cargo sales based on ARIMA-BP model[C]// 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA). IEEE, 2020.*
[2] Zhang Ge, Bai Jiao, Zhou Zhipeng, Cheng Qian. Prediction of TVOC concentration in museum scripture library based on ARIMA-SVM model [J]. HVAC, 2022, 52(11): 100-103. DOI: 10.19991 /j.hvac1971. 2022.11.15.
[3] Y. H. Wang. Global temperature prediction based on arima model and lstm neural network. Scientific and Technological Innovation, (35):166-170, 2021.
[4] J. Y. Lv, J. N. Du, M. Cao, and X. F. Fan. Carbon emissions trading price prediction using arima-svm model. Xi'an University of Science and Technology, 40(03):542-548, 2020.
[5] Z. C. Zhao, Y. Luo, and Huang J. B. Evidences of human influence on global warming. Department of Earth System Science, 19:18-23, 2021.