

Predictive Study of Monthly Changes in Global Atmospheric Methane Based on Arimr Modeling

Haomiao Wang^{#, *}, Shixian Deng[#], Zixuan He[#]

College of Aviation Meteorology, Civil Aviation Flight University of China, Sichuan 618307, China

* Corresponding author: Haomiao Wang (Email: whm18315251618@163.com)

[#]These authors contributed equally.

Abstract: Atmospheric methane is the second most abundant greenhouse gas in the atmosphere after carbon dioxide, and its concentration changes significantly regulate the intensity of the greenhouse effect, thereby exerting a profound influence on global warming and related climate change. Accurately understanding the changing patterns and future trends of atmospheric methane concentrations is of critical importance for elucidating the mechanisms of climate change, developing scientifically sound and effective emission reduction strategies, and advancing the realization of the global carbon neutrality objective. This study aims to construct a statistical model using the ARIMR model by integrating the characteristics of autoregression (AR), moving average (MA), and regression (R) to conduct predictive analysis of monthly changes in global atmospheric methane concentrations. The research data are sourced from monitoring stations under the World Meteorological Organization (WMO) Global Atmospheric Watch (GAW) program, covering monthly atmospheric methane concentration observations from 1994 to 2021. First, the data were subjected to stationarity tests, with the ADF test used to determine stationarity, and non-stationary data were differenced. Subsequently, the autoregressive correlation function (ACF) plot and partial autoregressive correlation function (PACF) plot were used for model order determination, ultimately selecting the ARIMR(0,1,1) model for predictive analysis. The results indicate that the ARIMR(0,1,1) model with a drift term demonstrates good fitting performance and predictive capability when forecasting changes in atmospheric methane concentrations, with small prediction errors and the ability to effectively continue the original trend of the series. Using the ARIMR model with a drift term to forecast atmospheric methane concentrations for 2025, the results show that atmospheric methane concentrations are expected to continue rising in the future.

Keywords: Atmospheric Methane, Monthly Variations, ARIMR Model, Global Climate Change.

1. Introduction

Climate change has become the core issue of global sustainable development in the 21st century, profoundly affecting the structure and function of natural ecosystems and posing a serious challenge to the economy, health and safety of human societies. As the second largest greenhouse gas after carbon dioxide, the impact of methane on global warming cannot be ignored [1]. In this macro context, the interaction mechanism between methane and climate change is of particular interest. According to the assessment report of the Intergovernmental Panel on Climate Change (IPCC), the Global Warming Potential (GWP) of methane is 28 to 36 times higher than that of carbon dioxide on a 100-year time scale [2-3]. Global atmospheric methane concentrations have shown a significant upward trend since the Industrial Revolution, and this change is closely linked to human activities, mainly agricultural practices (e.g., rice cultivation, ruminant farming), energy production (e.g., coal mining, leakage from natural gas systems), and municipal waste disposal [4-5]. Shi et al.[6] confirmed the significant contribution of these anthropogenic sources to the variation in methane concentration by integrating high spatiotemporal resolution natural and anthropogenic sources. In response to climate change, countries around the globe have proposed carbon neutrality targets, i.e., to achieve a balance between greenhouse gas emissions and sequestration over a certain period of time. Accurate modeling and analysis of monthly changes in atmospheric methane concentrations over the past nearly three decades (1994-2021) is important for better understanding the mechanisms of climate change, predicting

future climate change trends, and formulating effective coping strategies.

In recent years, research on atmospheric methane has been a very popular topic in atmospheric and environmental science research [7], and scholars at home and abroad have done a lot of work on the study of atmospheric methane and climate change, which is mainly categorized into the innovation of observation technology and the improvement of modeling and prediction. At the present stage, the commonly used means to observe methane are mainly the TROPOMI instrument on the Sentinel-5P satellite of the European Space Agency (ESA) and the atmospheric reference stations set up by the National Oceanic and Atmospheric Administration (NOAA) of the United States on a global scale, which provide a large amount of basic data for the study of methane trends and the relationship with the climate. Qian, Xie et al.[8] propose a method to utilize the ground-based high-resolution Fourier transformer (FT) to observe atmospheric methane and climate change. Li et al.[9] proposed a method to collect near-infrared solar absorption spectra using ground-based high-resolution Fourier transform infrared (FTIR) spectroscopy, combined with a nonlinear least-squares (NLS) fitting algorithm to invert and obtain the column-averaged dry-air mixing ratio (DAMR) concentration of atmospheric methane gases in a certain region. . Cheng et al.[10] used the ARMA model to construct the ARMA model based on the dynamics of the global average temperature, and utilized the time-series analysis method to construct the ARMA model based on the global average temperature dynamics data. analysis to construct an ARMA(p,q) model, which can well predict the temperature changes in the short term. Wang et al.[11] used

precipitation and temperature data from Huanghua City to study its evolutionary characteristics and periodic patterns using a variety of methods. They established an ARIMA model for prediction and fitting and obtained the optimal ARIMA model.

On the basis of domestic and international scholars' research, this study analyzes the monthly global atmospheric methane concentration from 1994 to 2021 by constructing an innovative ARIMR model, which overcomes the limitations of the ARIMA model that ignores the external covariates, and the linear regression model that is difficult to capture the characteristics of the time series. This study reveals the intrinsic connection between methane concentration and climate factors, predicts future climate change trends, and provides a scientific basis for coping strategies. The main steps of the study were: first, the ADF test was used to confirm the smoothness of the data, and the autoregressive and moving average orders were determined by combining the ACF and PACF plots to determine the atmospheric methane prediction model. Subsequently, the drift term was quoted and analyzed by comparing the prediction results with and without the drift term, and the prediction was made by the ARIMR model with the drift term. Finally, feasible suggestions are provided based on the future trend of atmospheric methane change.

2. Data processing and model construction

2.1. Data sources and pre-processing

2.1.1. Data sources

In order to ensure the accuracy, reliability and completeness of the data, the atmospheric methane data in this study come from the relevant monitoring stations of the Global Atmosphere Watch (GAW) program of the World Meteorological Organization (WMO), which are selected from the National Meteorological Science Data Center (NMSSDC) as a platform to download the data. These monitoring stations are distributed in different geographical locations around the world, covering a wide range of climatic zones and ecosystem types, and are able to reflect the global atmospheric methane concentration in a more comprehensive manner. In this study, a total of monthly atmospheric methane data observations from January 1994 to December 2021 were obtained, which include the station GAW number, start year, start month, global methane concentration, longitude, latitude, altitude, terrain, and height of sampling ports.

The main website for data queries related to this paper is the National Weather Science Data Center (data source): <http://data.cma.cn/data/cdcdetail/dataCode/G.0029.0003.S001.html>.

2.1.2. Data Preprocessing

Since there are missing values in the officially given data, and the existence of missing values may affect the accuracy and reliability of the model, it is necessary to carry out relevant processing for the missing values. Since the atmospheric methane concentration data present the characteristics of time series, linear interpolation is a suitable choice to deal with the atmospheric methane data, so in this study linear interpolation is used to correlate the missing values.

It is assumed that the values of the time series data at the time points t_1 and t_3 are y_1 and y_3 , respectively, and the value of t_2 is missing at the intermediate time point.

According to the principle of linear interpolation, the value y_2 at the time t_2 can be calculated using the following formula:

$$y_2 = \frac{(t_2 - t_1)y_3 + (t_3 - t_2)y_1}{t_3 - t_1} \quad (1)$$

This method is based on the assumption of continuity of the time series data, and linear fitting is performed between neighboring observations to fill in the missing data points. linear interpolation can effectively maintain the basic trend of the data when dealing with a small number of missing values of the time series data.

2.2. Description of Modeling Methods

2.2.1. ADF test for data smoothness

BP neural network is back propagating, mainly composed of three parts: input layer, middle layer and output layer. The number of nodes in the input and output layers is relatively easy to determine, but the determination of the number of nodes in the hidden layer is a very important and complex problem. The data smoothness test is a prerequisite for constructing the ARM model for time series analysis. Therefore, before carrying out model construction, the data need to be tested for smoothness. The test of time series smoothness generally has the observation method and the unit root test, and the smooth time series data have stable statistical characteristics, such as the mean, variance and self-covariance.

The ADF test is based on the unit root hypothesis, which determines whether a unit root exists in the data by constructing a regression equation that includes a lagged term in the time series. If a unit root exists, the data have non-stationary characteristics such as trend or seasonality. For the atmospheric methane concentration data VALUE, the ADF test statistic is:

$$\tau = \frac{\hat{\rho}}{s(\hat{\rho})} \quad (2)$$

The smoothness of the data is determined by comparing the ADF test statistic with the corresponding critical value. If the test statistic is less than the critical value, the original hypothesis is rejected, i.e., the data is smooth; conversely, if the test statistic is greater than the critical value, the data is non-smooth. If the data is non-smooth, a differencing operation is required. If the data are still not smooth after first-order differencing, it may be necessary to perform multiple differencing until the data reach a smooth state.

2.2.2. Pure randomness test

The Ljung-Box test (Ljung & Box, 1978) is a statistical test based on the autocorrelation of residuals, which is used to verify whether the residuals of a time series model satisfy the white noise assumption. The original hypothesis (H_0) is that "the residuals are not autocorrelated to lag (m) order", and the alternative hypothesis (H_1) is that "there is at least one order of significant autocorrelation". The test statistic is defined as:

$$Q_{LB} = N * (N + 2) \sum_{k=1}^h \frac{\rho_k^2}{N - k} \quad (3)$$

2.2.3. Plot of autocorrelation function and partial autocorrelation function to order

Autocorrelation Function (ACF) plots show the correlation between a time series and its own lagged values. For the atmospheric methane concentration data, the ACF value is (1) when the lag order is (0), and gradually decreases as the lag order increases. When the ACF value rapidly converges to 0 after the first order, it indicates that the correlation between the time series and its past values becomes weaker after that.

Partial Autocorrelation Function (PACF) plots, on the other

hand, show correlation after controlling for the effect of intermediate lagged values. PACF plots can help determine the value of p .

2.2.4. Drift term selection

In the case of differential operations, the drift term refers to the mean value of the sequence after differentiation:

$$\mu = E(1 - B)^d x_t \quad (4)$$

Assume that $\{x_t\}$ is an ARIMA(1, 1, 0) sequence with a drift term.

$$(1 - \phi_1 B)(1 - B)x_t = c(1 - \phi_2) \quad (5)$$

Then the drift term is equal to

$$\mu = E[\nabla x_t] = c(1 - \phi) \quad (6)$$

The arima function only estimates the autoregressive coefficient ϕ , and the constant term c will be ignored, so $\mu = c = 0$. However, the parameter estimation method will make the information of the sequence with drift term missing, which will affect the prediction accuracy of the sequence.

In this study, the Arima function and arima function are used to fit the model through the white noise test plot of the residual series respectively, so as to determine whether the model is significantly valid. The data from 1999 to 2017 are selected as the training set, and the data after 2017 are used as the test set.

3. Results

3.1. Analysis of Model Parameter Results

3.1.1. Analysis of data smoothness results

In order to test the smoothness of the atmospheric methane concentration data, a simple graph test is first performed on the data. The time series graph of atmospheric methane concentration is shown in Figure 1, which can clearly see that the global methane content shows a certain trend of growth, and it can be judged that the atmospheric methane data is a non-stationary series.

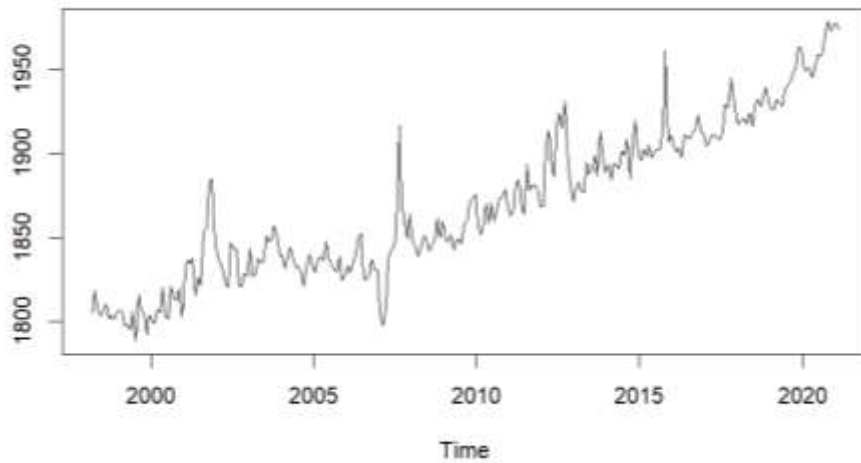


Figure 1 Time series of atmospheric methane concentration from 1994 to 2021

Since the atmospheric methane data is not smooth, it needs to be differenced until the data becomes smooth, so the first-

order difference is applied to the series to extract the linear trend information.

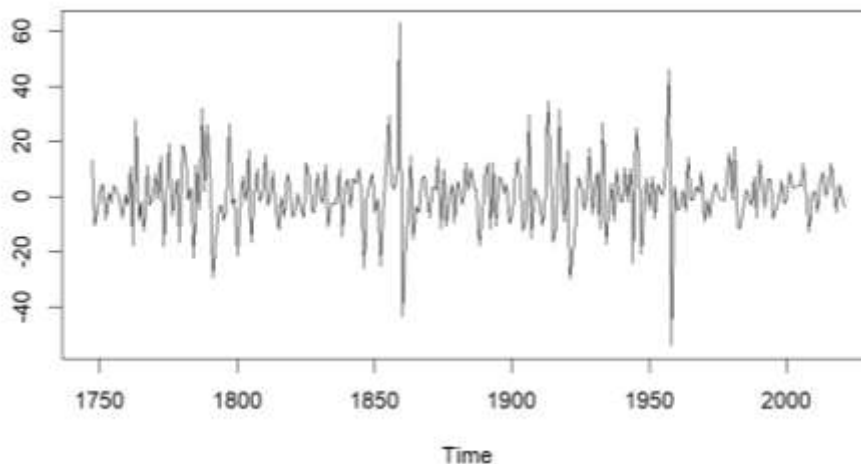


Figure 2 Time series plot after differencing

The time series plot after differencing is shown in Figure 2, which can clearly show that the first-order differencing operation successfully extracts the linear trend from the

original sequence, and the sequence after differencing shows very smooth fluctuations features, the mean and variance are evenly distributed on the time axis, with no obvious

increasing or decreasing trend, and there is no significant periodic fluctuation, which is preliminarily judged to meet the requirement of smoothness, and it can be used for the subsequent construction of the ARIMR model.

3.1.2. Analysis of ADF test results

After the difference using the ADF test of the smoothness

of the sequence, the test results show that the sequence of all the ADF statistics of the P value is less than the level of significance ($\alpha = 0.05$), so it can be assumed that the data after the 1st-order difference to achieve a smooth. The autocorrelation test results are shown in Table 1 :

Table 1 Autocorrelation Test

Lag	Type1: no drift no trend		Type2: with drift no trend		Type3: with drift and trend	
	ADF	p.value	ADF	p.value	ADF	p.value
0	-20.54	0.01	-20.57	0.01	-20.57	0.01
1	-14.94	0.01	-15.00	0.01	-14.99	0.01
2	-12.57	0.01	-12.57	0.01	-12.66	0.01
3	-10.54	0.01	-10.54	0.01	-10.58	0.01
4	-9.36	0.01	-9.51	0.01	-9.52	0.01
5	-9.33	0.01	-9.54	0.01	-9.43	0.01

As shown in Table 1 , when the order of data differentiation reaches the second order, it shows significance ($P < 0.05$), i.e., the original hypothesis is rejected, and the sequence reaches stationarity. Therefore, when constructing the model, second-order differentiated data should be used to construct the ARIMA model.

3.1.3. Analysis of pure randomness results

The results of the pure randomness test are shown in Table 2 , which shows that the P-value of the LB statistic under each order delay is less than the significance level, which indicates that the sequence after differencing is not a white noise sequence, so it can be determined that the sequence after the 1st order differencing is a smooth non-white noise sequence.

Table 2 Pure Randomness Test

Box-Pierce test					
data: cfsj					
X-squared	19.825	df	6	p-value	0.002975
X-squared	27.847	df	12	p-value	0.005825

3.1.4. Analysis of model fixed order results

By looking at the lagged value of the first significant

truncation in the ACF plot, it is possible to determine the order of the moving average term, and the post-differential autocorrelation (ACF) plot of the data is shown Figure 3 :

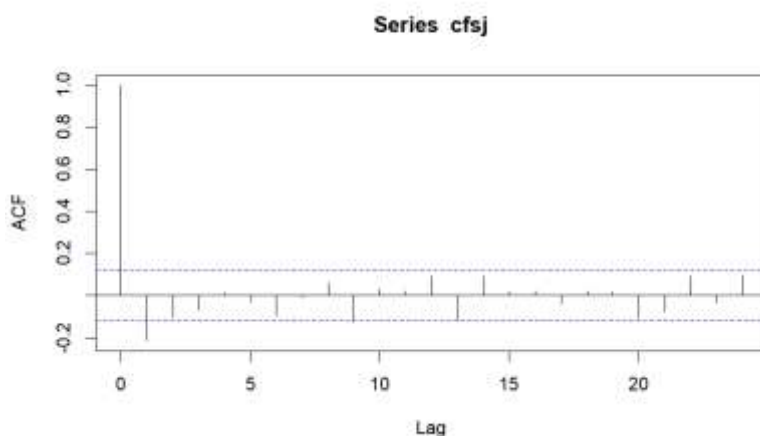


Figure 3 Post-Differential Autocorrelation (ACF) Plot

By looking at the lagged value of the first significant truncation in the PACF plot, one is able to determine the order

of the autoregressive term, and the post-differential partial autocorrelation (PACF) plot for the data is shown Figure 4 :

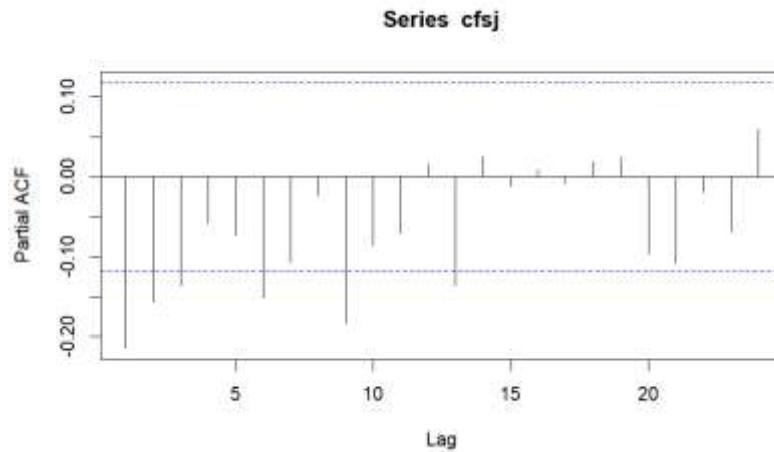


Figure 4 Post-differential Partial Autocorrelation (PACF) Plot

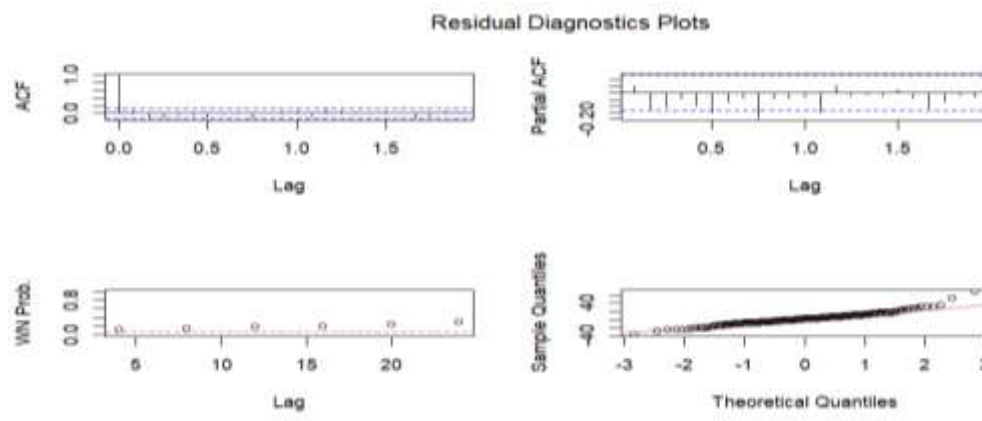
In summary, it can be seen that the model presents a first-order truncated tail characteristic in the autocorrelation plot and a trailing tail characteristic in the partial autocorrelation plot, so the MR(1) model is used to fit the first-order post-differential series, and an ARIMR (0, 1, 1) model is fitted to the atmospheric methane prediction model series, taking into account that a first-order differencing has been carried out earlier.

3.1.5. Analysis of difference results with and without drift term

By Table 3 Model Coefficients for the Un-drifted Model and the Drifted Model and Figure 5 Significance Test for Un-drifted and Drifted Fitted Models it can be seen that the residual test of the models shows that both models hold significantly. The biggest difference between them at this point would be in the prediction where the presence or absence of the drift term would have a significant effect on the predicted trend.

Table 3 Model Coefficients for the Un-drifted Model and the Drifted Model

	Ma1	Ma1	Drift
	-0.3423	-0.3506	0.4825
S.e.	0.0829	0.0842	0.5690
Sigma ²	165.5	166.5	
Log likelihood	-858.31	-857.95	
AIC=1720.62		AIC=1721.91 AICc=1722.02 BIC=1732.04	



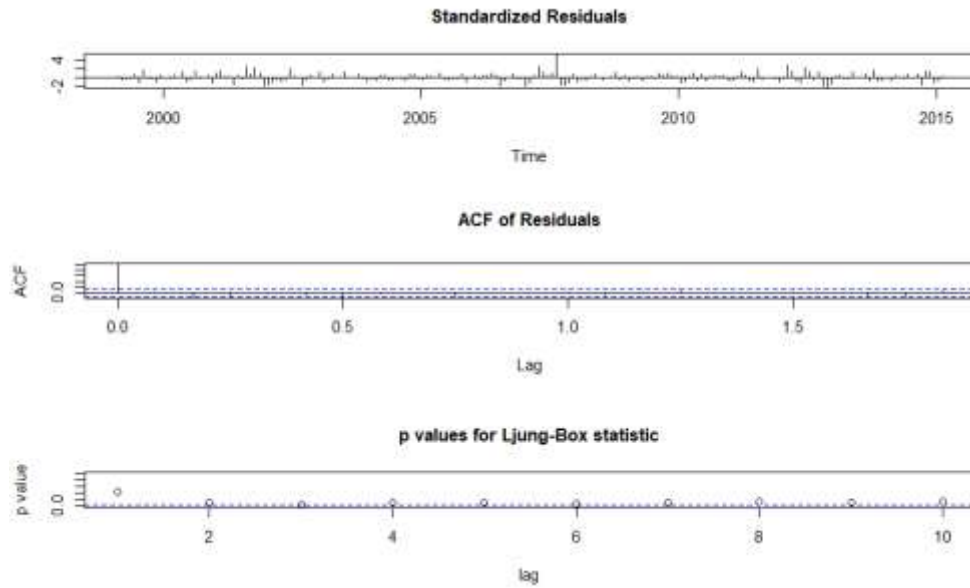


Figure 5 Significance Test for Un-drifted and Drifted Fitted Models

3.2. Analysis of modeling effects

Based on the test set from 2017 onwards, the ARIMA(0,1,1) model with no drift term and the ARIMA(0,1,1) model with drift were fitted, and based on these two fitted models, the series were predicted for a period of 10 periods, and the real values of the last 10 periods of the columns were used as the test set, and compared with the predicted values to get the series of prediction error (error), and compared with the predicted values of the two prediction models, the predicted values of the no drift term model are almost constant with a great prediction error, and the predicted values of the model with drift show a linear increase, but the prediction error becomes large in the short term. The prediction error of the

model without drift term is almost constant and the prediction error is very large, while the prediction value of the model with drift term is linearly increasing and the prediction error is very small in the short term, but the number of prediction periods becomes larger and the prediction difference also becomes larger. In this study, the prediction effect of the model with drift term is significantly better than that of the model without drift term.

Comparison of prediction effect plots with and without drift term model is shown in Figure 6 , the solid line is the observed value of the series, the dashed line is the fitted value of the model, the shaded solid line is the predicted value, the dark shading is the 80% confidence interval of the series, and the light shading is the 95% confidence interval of the series.

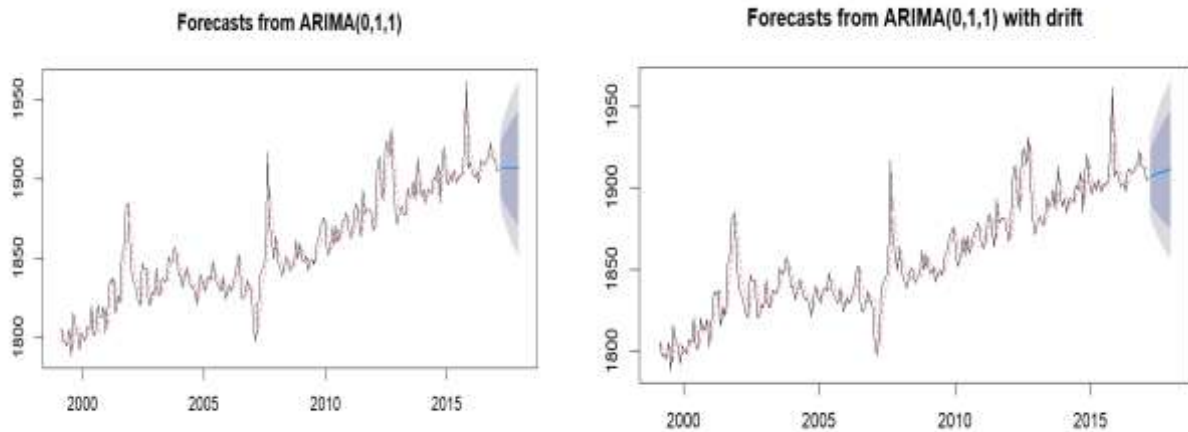


Figure 6 Comparison of Prediction Results Between Models with and Without Drift Terms

As can be seen from Figure 6, both fitting models perform well on the test set data. However, by comparing the errors of the models with and without drift terms, it can be concluded that there is a significant difference in prediction performance.

The prediction errors of the ARIMA(0,1,1) model without drift terms and the ARIMA(0,1,1) model with drift terms are shown in Table 4 :

Table 4 Prediction Errors without Drift Term and with Drift Term Prediction Errors

No.	Prediction error without drift term			Prediction error with drift term		
	Fore1.mean	test.1.10.	Error1	fore2.mean	test.1.10.	error2
1	1906.414	1905.736	4.168535	1907.203	1905.736	3.3801406
2	1906.414	1910.583	4.722535	1907.685	1910.583	3.4516022
3	1906.414	1911.137	3.361535	1908.168	1911.137	1.6080639
4	1906.414	1909.776	1.731535	1908.650	1909.776	-0.5044745
5	1906.414	1908.146	7.007535	1909.133	1908.146	4.2889871
6	1906.414	1913.422	22.544535	1909.616	1913.422	19.3434488
7	1906.414	1928.959	21.173535	1910.098	1928.959	17.4899104
8	1906.414	1927.588	39.142535	1910.581	1927.588	34.9763720
9	1906.414	1945.557	28.150535	1911.063	1945.557	23.5018337
10	1906.414	1934.565	16.309535	1911.546	1934.565	11.1782953

From Table 4, it can be seen that the prediction mean of the model without drift term is almost constant, which basically changes the development trend of the original sequence, and the prediction error is large. The prediction effect of the model with drift term is better, and the predicted value basically

continues the development trend of the original sequence.

3.3. Analysis of prediction results

Predictions of atmospheric methane content in 2025 based on the ARIMA model with drift are shown in Table 5.

Table 5 Prediction of methane content change in the next 10 months based on the drift term

Time	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2025.1	2016.889	1971.538	2062.240	1947.530	2086.247
2025.2	2021.813	1976.021	2067.606	1951.780	2091.847
2025.3	2019.395	1973.166	2065.625	1948.693	2090.097
2025.4	2019.853	1973.190	2066.516	1948.488	2091.217
2025.5	2024.002	1976.909	2071.094	1951.980	2096.023
2025.6	2029.624	1982.106	2077.142	1956.951	2102.296
2025.7	2031.445	1983.505	2079.385	1958.127	2104.763
2025.8	2032.714	1984.356	2081.073	1958.756	2106.673
2025.9	2043.062	1994.288	2091.836	1968.469	2117.656
2025.10	2034.649	1985.463	2083.834	1959.425	2109.872

Based on the above model, the long-term change of atmospheric methane since 1994 can be reasonably portrayed, and important suggestions can be made for the future atmospheric change of methane: the government needs to strengthen the construction of the global methane monitoring network and international cooperation, to promote the integration and sharing of multi-source data, and to establish a unified emission reduction target and policy framework; enterprises should increase the research and development of methane emission reduction technologies and promote them to enhance the efficiency of emission reduction; and the public needs to enhance their awareness and practice emission reduction actions through multiple channels to jointly address the challenge. The public should enhance their knowledge and practice emission reduction actions through multiple channels to meet the challenges together.

4. Conclusions

This study employs an innovative ARIMR model to conduct an in-depth analysis of global atmospheric methane monthly concentrations from 1994 to 2021. It overcomes the limitations of the ARIMA model, which ignores external covariates, and the linear regression model, which struggles to capture time series characteristics. The results indicate that the ARIMR model demonstrates significant advantages in describing the monthly variation characteristics of atmospheric methane. The ADF test was used to confirm data stationarity, and the ACF and PACF plots were employed to determine the autoregressive order and moving average order, leading to the construction of an ARIMR (0,1,1) model for

atmospheric methane prediction. Comparing the prediction results with and without a drift term, the ARIMR(0,1,1) model with a drift term showed better fitting performance, smaller prediction errors, and better continuation of the original sequence trend. The prediction indicates that atmospheric methane concentrations will continue to rise by 2025. Based on the above model, the long-term changes in atmospheric methane since 1994 can be reasonably characterized, and important recommendations can be made for future atmospheric methane changes.

Studying the monthly variation patterns of atmospheric methane holds significant scientific value and practical significance. Although this study analyzed and predicted global monthly atmospheric methane concentrations, it still has limitations such as restricted data sources and models that do not account for external driving factors, which need to be improved and refined in future research.

In the face of the challenges posed by rising global atmospheric methane concentrations, no one can remain unaffected. Only through the collective efforts of governments, businesses, and the public, by taking concrete actions to implement atmospheric methane emissions reductions, can we work together to create a sustainable environment.

References

- [1] Li H L. 2024 Guangdong Province Low Carbon Theme Publicity Event Held in Guangzhou[J]. Environment, 2024, (08): 63-65.

- [2] Umezawa T, Tohjima Y, Terao Y, et al. Long-term and interannual variations of atmospheric methane observed by the NIES and collaborative observation networks[J]. *Progress in Earth & Planetary Science*, 2025, 12(1).
- [3] Chen S, Peng J, Zhu C, et al. Global Response and Mechanism of Methane Cycling in Wetlands Under Elevated Atmospheric CO₂ and Warming[J]. *Global change biology*, 2025, 31(7): e70325.
- [4] Christensen R T, Arora K V, Gauss M, et al. Tracing the climate signal mitigation of anthropogenic methane emissions can outweigh a large Arctic natural emission increase[J]. *Scientific Reports*, 2019, 9(1): 1-8.
- [5] Chen C, Wang E, Yu Q, et al. Quantifying the effects of climate trends in the past 43[J]. 2010, 100(3-4): 559-578.
- [6] Shi Y, Yang Y, Zhou Z, et al. A comprehensive inventory of China's methane emissions: Insights and changes from 2010 to 2020[J]. *Journal of Cleaner Production*, 2025, 519 146089-146089.
- [7] Yao L, Yang D X, Cai Z N H, et al. Analysis of the Current Situation and Development Trends of Atmospheric Methane Observation Satellites in China for Carbon Neutrality and Carbon Pea[J]. *Chinese Journal of Atmospheric Sciences*, 2022, 46(06): 1469-1483.
- [8] Qian Z W, Xie Y, Chen J, et al. Characterization of atmospheric methane using ground-based high-resolution Fourier transform infrared spectroscopy[J]. *Journal of Optics*, 2025, 45(06): 35-43.
- [9] Li L, Xu L X, Wang P X, et al. Drought prediction of summer maize growing season based on conditional vegetation temperature index[J]. *Journal of Agricultural Machinery*, 2020, 51(01): 139-147.
- [10] Cheng Y, Hua Z Q, Huang Y J, et al. Application of ARMA model in predicting global average temperature scenario[J]. *Journal of Inner Mongolia University for Nationalities(Natural Science Edition)*, 2023, 38(02): 103-108.
- [11] Wang C Y, Huang X, Wang Y M, et al. Analytical study of precipitation and temperature in Huanghua city based on wavelet and ARIMA[J]. *Journal of Hebei University of Geology*, 2023, 46(03): 80-85.