Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/40/20230665

The prediction and analysis of global climate change based on SARIMA

Dongyao Liu

Jinan Foreign Language School, Jinan, 250108, China

1811000211@mail.sit.edu.cn

Abstract. Global climate change is a significant challenge that the world is currently facing. Accurate prediction of global climate change is essential for environmental protection, agricultural production, and social development. This study explores the utilization of the Seasonal Autoregressive Integrated Moving Average (SARIMA) model for forecasting global climate change. The SARIMA model is a machine learning algorithm that can effectively capture seasonal patterns and non-linear characteristics of climate data. The study initiates by performing data preprocessing tasks, which encompass data cleaning, managing missing values, and converting the data into a suitable format for analysis. The SARIMA model is then constructed, considering the seasonality and autocorrelation of the climate data. Historical climate data is used to train the SARIMA models, which are then utilized to forecast future global climate changes. The predictive performance of the models is evaluated to validate the effectiveness and accuracy of the SARIMA model in global climate change prediction. Experimental results indicate that the SARIMA model effectively captures the underlying patterns and dynamics of the climate data. The accurate predictions of the SARIMA model have practical implications for understanding and forecasting global climate change. These forecasts provide insightful information for policy formulation and decision-making, aiding in the development of innovative strategies to mitigate and adapt to climate change.

Keywords: Global Climate Change, SARIMA Model, Prediction, Data Preprocessing, Seasonality.

1. Introduction

Global climate change is a significant contemporary global challenge [1]. With the increase in global temperatures, the rise in extreme weather events, and the instability of ecosystems, accurate prediction of global climate change is crucial for environmental protection, agricultural production, and social development. [2,3] In response to this challenge, many researchers and organizations have been devoted to developing reliable climate prediction models [4].

In previous research, various methods have been used to address the issue of climate change prediction. Some researchers have employed traditional statistical methods, such as linear regression models, to establish climate prediction models. However, due to the complexity and non-linear characteristics of the global climate system, traditional statistical methods have limitations in predicting long-term trends and seasonal variations. In recent years, time series analysis methods have been widely applied in the field of climate change prediction [5-7]. The research aims to utilize the SARIMA Model, a machine learning algorithm, to comprehensively comprehend and effectively leverage existing datasets and simulations. By integrating the SARIMA Model into the analysis, a deeper understanding of the data can be achieved, enabling accurate predictions and extraction of valuable insights. The SARIMA Model possesses an inherent capability to capture seasonal patterns and seamlessly integrates with machine learning techniques, facilitating precise predictions and extraction of valuable insights from the data. This study focuses on harnessing the potential of the SARIMA Model to analyze and leverage existing datasets and simulations, contributing to the advancement of knowledge in this field. Moreover, the SARIMA model has gained recognition for its exceptional ability to capture the seasonality and autocorrelation present in time series data. This distinctive attribute of the SARIMA Model enables more precise predictions of climate change trends, offering valuable insights into the dynamics of such phenomena [6,8,9].

The primary goal of this study is to predict global climate change using the SARIMA model. Specifically, the global climate data will be collected and preprocessed to ensure its suitability for analysis. The preprocessing steps encompass data cleaning to remove errors or inconsistencies, handling missing values, and converting the data into an appropriate format for analysis. The purpose of the preprocessing phase is to ensure the quality and integrity of the data, enabling accurate and reliable predictions using the SARIMA model. then construct SARIMA models by considering the seasonality and autocorrelation of the climate data. The models will be trained using historical climate data and utilized to forecast future global climate changes. Additionally, the predictive performance of the model is evaluated to validate the effectiveness and accuracy of the SARIMA model in global climate change prediction. The experimental results indicate that the SARIMA parameters were well-fitted, as the predicted values align closely with the real values and exhibit the expected seasonal pattern. This suggests that the SARIMA model effectively captures the underlying patterns and dynamics of the climate data. The accurate predictions of the SARIMA model have practical implications for understanding and forecasting global climate change. These insights facilitate the design of policies and decision-making, enabling the creation of potent countermeasures and climate change adaptation plans. Policymakers can use these insights to implement measures such as reducing greenhouse gas emissions and protecting vulnerable ecosystems. The accurate predictions also help in anticipating and preparing for potential impacts, guiding the development of long-term strategies to address climate change and ensure a sustainable future.

2. Methodology

2.1. Dataset description and preprocessing

The information is derived from a recent collection conducted by Berkeley Earth, a company associated with Lawrence Berkeley National Laboratory [10]. The Berkeley Earth Surface Temperature Study consolidates temperature reports from 16 pre-existing archives, totaling 1.6 billion data points. The data set includes time, average temperature, average temperature uncertainty, country, latitude and longitude, and other parameters. It is necessary to do some preprocessing on the dataset. In the case of Rio de Janeiro, first, create a pivot table to plot the monthly temperatures through the years. The series has some seasonality, the higher temperatures are around November and February and the lower are between July and September. This means that a seasonal ARIMA model can be used. Plot the average temperature trend over a hundred years, it can be confirmed that there is a constant increasing trend and that the average temperature increased from 23.5º to 24.5º, that's 4.25% in over 100 years.

2.2. Proposed approach

The purpose of this thesis is to study the prediction of global climate change using the SARIMA model. The SARIMA model has significant significance and characteristics in global climate forecasting experiments. Global climate data often exhibit pronounced seasonal variations and non-stationarity, which can be effectively captured and predicted by the SARIMA model. The overall experiment is divided into four parts. Firstly, data preprocessing was performed, including data cleaning, missing values processing, and feature selection. Next, the SARIMA model is used to construct the analytical model, which can make predictions of future global climate change. The third part is the performance analysis, which verifies the effectiveness and accuracy of the model in global climate change prediction. Figure 1. depicts the procedure of the experiment.

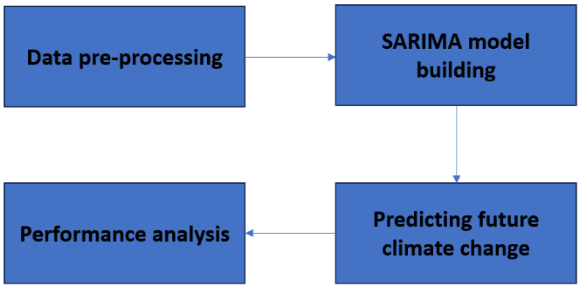

Figure 1. The process of the study.

2.2.1. SARIMA. By considering both seasonal and non-seasonal components, the SARIMA model can handle long-term and short-term dependencies and provide predictions at different time scales, ranging from hourly to yearly. Its structure includes components such as autoregression, differencing, and moving averages for non-seasonal patterns, as well as seasonal autoregression, seasonal differencing, and seasonal moving averages for capturing seasonal patterns. In global climate forecasting experiments, the SARIMA model serves as an important analytical tool for scientists, meteorologists, and policymakers, enabling them to understand trends, seasonal patterns, and extreme weather events in global climate change, and providing scientific basis and predictive capabilities for decision-making and planning. The parameters of the proposed model are divided into seven blocks.

1. p: Number of lag observations in the autoregressive model, capturing correlation with previous observations.
2. d: Number of differencing operations to transform a non-stationary time series into a stationary one.
3. q: Number of lagged forecast errors in the moving average model, capturing correlation with previous forecast errors.
4. P: Number of lag observations in the seasonal autoregressive model, capturing seasonal correlation.
5. D: Number of seasonal differencing operations to eliminate seasonal patterns.
6. Q: Number of lagged forecast errors in the seasonal moving average model, capturing seasonal average fluctuations.
7. S: Seasonal period, representing the length of the repeating seasonal pattern.

The examination of the autocorrelation function (ACF) and partial autocorrelation function (PACF) helps determine the optimal parameter values. Additionally, information criteria such as AIC or BIC can be used to evaluate the performance of different parameter combinations and select the one with the lowest information criterion value.

3. Result and discussion

Upon completion of the model training process, a visual representation was generated to evaluate the residuals.

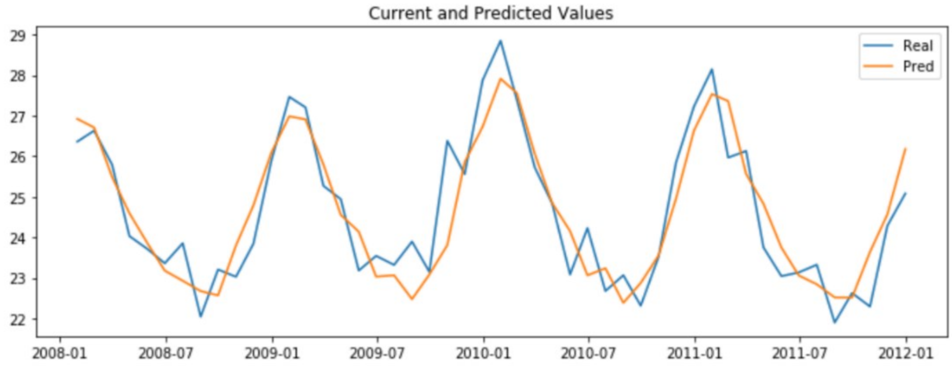

Figure 2. Current and Predicted value chart.

Figure 2. shows a line chart depicting the temporal progression of both current and predicted values for the city of Rio de Janeiro from 2008 to 2012 was generated. Based on the analysis of the plots presented above, it is evident that the predicted values align remarkably well with the current data.

Figure 3. Residuals and Predicted value chart.

Figure 3 shows the scatter plot analysis indicates that there is no apparent correlation between the magnitude of the errors and the corresponding temperature values. The errors, ranging from -1.5 to +1.5, do not exhibit any discernible trend or pattern as the temperature increases.

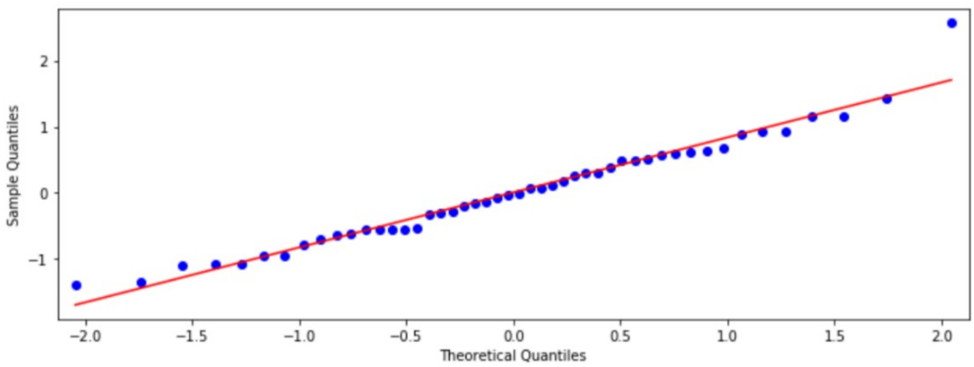

Figure 4. Quantile-quantile chart.

The QQ plot analysis reveals a normal distribution pattern with a few minor outliers in Figure 4. The majority of the data points conform to the expected normal distribution, indicating a satisfactory fit between the observed data and the theoretical normal distribution. However, a few scattered outliers deviate from the expected pattern, suggesting the presence of some exceptional observations. Despite these outliers, the overall shape of the QQ plot indicates that the data aligns reasonably well with the normal distribution. This finding reinforces the assumption of normality and supports the validity of applying statistical techniques that rely on the normal distribution assumption to further analyze the data.

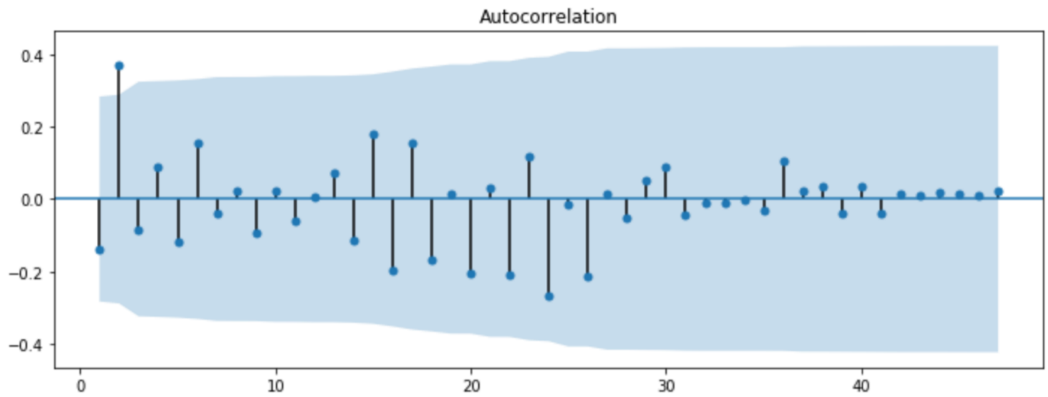

Figure 5. Quantile-quantile chart.

The autocorrelation plot in Figure 5 exhibits a notable positive spike that surpasses the confidence interval in proximity to the second lag. However, based on the analysis conducted, it is deemed unnecessary to implement further modifications. This finding suggests that the observed data exhibits a certain degree of temporal dependence, particularly in the vicinity of the second lag. Despite this positive spike, it is determined that the existing model or methodology adequately captures the underlying patterns and relationships within the dataset. Therefore, it is concluded that additional adjustments or refinements are not warranted at this stage of analysis.

In the final stage of analysis, as Figure 6, the focus shifts towards extrapolating predictions for the test set, specifically targeting the last 12 months. This extrapolation aims to extend the model's forecasting capabilities beyond the observed data and into the future period. By applying the established model to the test set, valuable insights can be obtained regarding the anticipated trends and patterns that may unfold during the specified 12-month period.

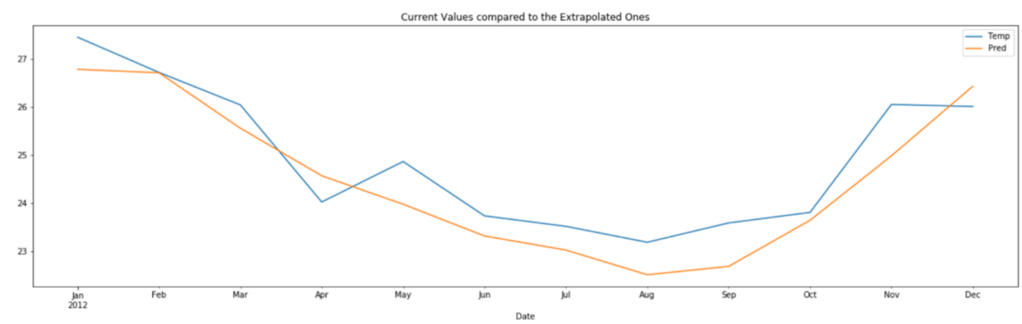

Figure 6. Quantile-quantile chart.

The SARIMA parameters were effectively calibrated, resulting in a satisfactory fit between the predicted values and the actual data. The model successfully captures not only the overall trends and patterns in the dataset but also the inherent seasonal variations. The observed alignment between the predicted and real values, as well as the replication of the seasonal pattern, indicates the model's robustness and accuracy in capturing the underlying dynamics of the data. This finding provides compelling evidence of the model's reliability and its ability to generate accurate predictions, thereby supporting its suitability for forecasting and decision-making purposes.

4. Conclusion

This research study focuses on the prediction of global climate change using the SARIMA model, capturing seasonal patterns and non-linear characteristics to forecast future global climate changes. The research initiates with data preprocessing, involving tasks such as cleaning, handling missing values, and preparing the data for analysis. The SARIMA model is then constructed, taking into account the seasonality and autocorrelation of the climate data. To assess the efficacy and precision of the SARIMA model in projecting global climate change, several experiments are carried out. Experimental results demonstrate that the SARIMA model effectively captures the underlying patterns and dynamics of the climate data, providing accurate predictions. These predictions provide valuable insights for decision-making and policy formulation, facilitating effective strategies to address and adapt to climate change.

In future work, the research will consider the impact of greenhouse gas emissions as the research objective for the next stage. The focus will be on analyzing the contribution of greenhouse gas emissions to global climate change, further enhancing our understanding of the factors driving climate change. This will help develop the field's understanding and allow for more thorough and accurate projections of climate change patterns.

References

[1] Abbass K Qasim M Z Song H et al 2022 A review of the global climate change impact adaptation and sustainable mitigation measures Environmental Science and Pollution Research 29(28): pp 42539-42559
[2] Kang Y Khan S Ma X 2009 Climate change impacts on crop yield crop water productivity and food security–A review Progress in Natural Science 19(12): pp 1665-1674
[3] Shad M Sharma YD & Singh A 2022 Forecasting of monthly relative humidity in Delhi India using SARIMA and ANN models Model Earth Syst Environ 8: pp 4843–4851
[4] Yerlikaya B A Ömezli S Aydoğan N 2020 Climate change forecasting and modeling for the year 2050 Environment climate plant and vegetation growth pp 109-122
[5] Parasyris A Alexandrakis G Kozyrakis G V et al 2022 Predicting meteorological variables on local level with SARIMA LSTM and hybrid techniques Atmosphere 13(6): pp 878
[6] Valipour M 2015 Long - term runoff study using SARIMA and ARIMA models in the United States Meteorological Applications 22(3): pp 592-598
[7] Manigandan P Alam M D S Alharthi M et al 2021 Forecasting natural gas production and consumption in United States-Evidence from SARIMA and SARIMAX models Energies 14(19): p 6021
[8] XHABAFTI M SINAJ V 2022 Weather forecasting based on the application of SARIMA models CIRCULAR ECONOMY p 549
[9] Zia S 2021 Climate Change Forecasting Using Machine Learning SARIMA Model iRASD Journal of Computer Science and Information Technology 2(1): pp 01-12
[10] Dataset https://www.kaggle.com/datasets/berkeleyearth/climate-change-earth-surface-temperature-data