Time Series Analysis of Greenhouse Gas Emission Based on ARIMA and LSTM

Jiajun Zhang*

School of Information Science and Engineering, Lanzhou University, Lanzhou, China

*Corresponding author: zhangjiajun2021@lzu.edu.cn

Abstract. Greenhouse gas emissions have become a topic of great concern, and research on the prediction of greenhouse gas emissions is urgently needed. In this paper, based on the GHG emission data from 1990-2018, it applied ARIMA model and LSTM model to predict future GHG emissions, and evaluated their prediction performance using MAE. According to the analysis, the results show that the ARIMA model can more accurately capture the trend and seasonal characteristics in the greenhouse gas emission data, and generate prediction results that match the actual observations. Moreover, this study confirmed the effectiveness and feasibility of ARIMA and LSTM models in greenhouse gas emission prediction. At the same time, one must also be aware that greenhouse gas emission forecasts still face limitations such as data reliability, model assumptions, and policy uncertainties. Future research can further improve model performance and explore more comprehensive predictive models to improve accuracy. Overall, these results shed light on guiding further exploration of gas emission analysis.

Keywords: Greenhouse gas emission; time series; ARIMA; LSTM.

1. Introduction

In recent years, climate change and environmental issues have attracted much attention, among which greenhouse gas emissions are a key topic. Greenhouse gases are substances like carbon dioxide (CO2), methane (CH4), nitrous oxide (N2O), and others that absorb and re-radiate solar radiation in the atmosphere. Human activities including the burning of fossil fuels, deforestation, and agricultural practices all contribute to the emissions of these gases. With the rapid advancement of global industrial and economic development, greenhouse gas emissions are showing an increasing trend, which also has a huge impact on environmental protection and atmospheric safety.Global emissions are still rising despite international efforts to tackle climate change, most recently with the adoption of the Paris Agreement in 2015. For example, global CO2 emissions increase by 6% from 2020 to 2021 [1]. In addition, with the development of technology, natural gas vehicles began to rise.These vehicles are particularly appealing in developing nations due to the inexpensive cost of natural gas. China has 5 million NGVs. Even though they are less noticeable, CNG-powered vehicles are also present in industrialized economies. For instance, the US has 1.5 million vehicles that run on natural gas [2]. In order to better understand the changing trend and possible future direction of greenhouse gas emissions, it is an important method to use forecasting tools to predict its development trend. Time series forecasting is a technique for building models of future trends, based on available historical data. Researchers collect and analyze large amounts of emissions data, taking into account a variety of factors such as economics, policy, and technology. By building the right models and algorithms, they can predict future greenhouse gas emissions. As training samples for time series prediction, create a set of input-target pairings.

New observations are added to the collection and the earliest observations are discarded each time a new time series is constructed [3].The main purpose of time series forecasting is to predict future trends and changes so that relevant departments can formulate effective environmental policies and climate improvement plans to reduce emissions, increase the use of renewable energy, and promote the development of clean technologies. When doing time series forecasting, this study uses traditional statistical methods and machine learning techniques. And to increase the forecast's accuracy and dependability, take into account some variables (historical data, seasonality, periodicity, etc.).

Through time series forecasting, one can more clearly understand the development trend of greenhouse gas emissions, which is very important for formulating global climate change policies and environmental protection goals. Prediction results can help governments and relevant departments take appropriate actions to promote development and mitigate carbon emissions.

GHG emissions are one of the main causes of climate change. By studying and analyzing the emission of greenhouse gases, one can better understand the law of climate change. It is helpful in determining the causes of the greenhouse effect and global warming and offers a scientific basis for developing policies and actions to combat climate change. On the other hand, greenhouse gas emissions are a major issue worldwide. Research on greenhouse gas emissions provides important guidance for the formulation of environmental policies and international cooperation. Based on time series forecasting and model building, one can predict future emission trends and assess the impact of different policies and actions on emissions. This helps policymakers formulate and implement emission reduction strategies, promote sustainable development, and promote international cooperation to mitigate climate change. In addition, research on greenhouse gas emissions can help set national and global emission reduction targets. Through forecasts of future emissions, governments and relevant departments can determine the measures that need to be taken to achieve emission reduction targets. Research on greenhouse gas emissions can also promote technological innovation and the development of clean energy. Through prediction, one can develop new technologies and support the clean energy transition. This will help promote the application and promotion of renewable energy, energy efficiency and low-carbon technologies, and promote the development and use of sustainable energy. In addition, research can assess the effectiveness of existing policies and measures and suggest improvements. The main issue to study the emission of greenhouse gases is to protect the environment and achieve sustainable development. Through the prediction and analysis of emission data, one can better understand the changes in global temperature, and can better assess the impact of climate change on ecosystems, biodiversity and human society.

In 1896, Swedish scientist Svent Ahrens published a paper in which he proposed for the first time that carbon dioxide is a key factor in the change of Earth's climate. The United Nations World Meteorological Organization (WMO) and the United Nations Environment Program (UNEP) founded the Intergovernmental Panel on Climate Change (IPCC) in 1979. The IPCC provides policymakers with the latest scientific information and recommendations on greenhouse gas emissions and climate change by evaluating the results of international scientific research. And in the early 20th century, scientists began to observe the concentration of carbon dioxide in the atmosphere and established weather stations and laboratories to monitor the emission of greenhouse gases. The Rio de Janeiro city hosted the signing of the United Nations Framework Convention on Climate Change (UNFCCC) in 1992. The Convention aims to stabilize greenhouse gas concentrations such that they do not affect living things or the environment. The adoption of the Paris Agreement in recent years has encouraged international efforts to combat climate change [1]. The accord requires all nations to adopt steps to keep global warming to 1.5°C.

Research on the relationship between greenhouse gas emissions and climate change can infer the extent of negative impacts of greenhouse gas emissions on different countries [4]. With the emission of greenhouse gases, the temperature of the atmosphere continues to rise, causing the temperature to rise above the prescribed temperature limit, causing anomalous changes in temperature in most regions. Through certification in the global carbon market and using the AMS-III.S. method to calculate greenhouse gas emissions over a year, researchers evaluated the switch from diesel to compressed natural gas as a cost-effective solution for reducing greenhouse gas emissions in the private road transport sector. And concluded that the total amount decreased by 7% [5]. Larissa Fait and his team's research on the relationship between inefficiencies in current production technologies and the untapped potential for emission reductions could provide useful information for policymakers and international climate policy. Finally, using inefficiency measures, they conclude that reducing inefficiencies can make a considerable contribution [1].

Greenhouse gas emissions have serious impacts on global climate change and environmental health. According to scientific research, the emission of greenhouse gases such as carbon dioxide, methane and nitrous oxide has led to an increase in global average temperature and an increase in extreme weather events and other adverse effects. Facing this series of problems, the research on the prediction of greenhouse gas emissions has become particularly important.

In recent years, many research institutes have analyzed and predicted the trends and changes of greenhouse gas emissions. These investigations have employed a range of techniques, including time series analysis, machine learning algorithms, and statistical models, among others. However, despite some progress, existing studies still have some shortcomings in terms of accuracy, data reliability, and timeliness. For example, reliable emissions data are often lacking or difficult to obtain in some regions and industries, which hinders accurate emissions projections. In addition, the complexity and diversity of greenhouse gas emissions also increase the difficulty of data collection and analysis. Recognizing the limitations and deficiencies of existing research, this study used more timely and complete data in this study to supplement and improve the accuracy of greenhouse gas emission predictions. By combining new datasets and improved model approaches, it is hoped to provide more reliable, accurate and comprehensive projections of greenhouse gas emissions.

2. Data and Methods

The data set comes from the Kaggle website. The data were collected from all countries in the world and covered almost 30 years (1990-2018). Data was collected from the Climate Watch Data Portal using the CAIT data source. Through the analysis and understanding of the data, one can deeply understand the role of methane, fluorine and other greenhouse gases and the general trend of global greenhouse gas emissions. Greenhouse gas emissions are time-series, so time-series models can be used to predict them. This study will use ARIMA (Auto regressive Integral Moving Average Model) and LSTM (Long Short-Term Memory Network) are two commonly used models for time series forecasting. They can help analyze trends and periodicity in data and make future predictions based on historical observations.

Auto regressive integral moving average (ARIMA) is the most commonly used time series prediction model [6]. It models based on the auto correlation and moving average properties of time series. ARIMA models can capture features such as long-term trends, seasonality, and periodicity in the data. This model is divided into three types: auto regressive (AR) method, moving average (MA) method and ARIMA method [7]. The model mainly has three parameters: p (auto regressive order), d (difference order) and q (moving average order). Additionally, very critically ARIMA has been shown to be ineffective in determining the nonlinear relationships involved [8]. The steps of the ARIMA model construction method are shown in the flowchart in Fig. 1.

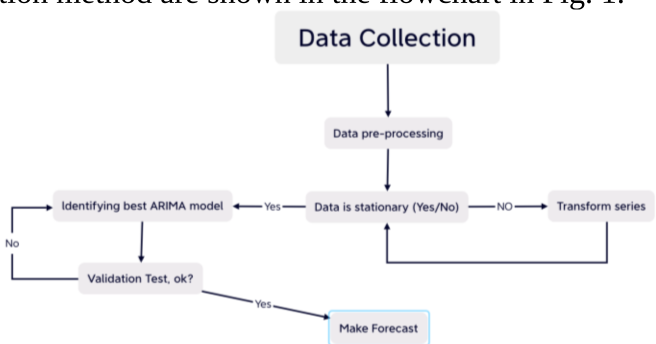

Fig. 1 Steps to build an ARIMA model.

LSTM is a variant of recurrent neural networks (RNN) that can identify long-term dependencies (seen from Fig. 2) [9]. Relying on the ability of LSTM to memorize patterns in both short-term and long-term, it is considered to be a good method to overcome the problem of time series forecasting [10]. LSTM models efficiently process time series data by using memory cells and gate mechanisms to remember and forget information. Compared to traditional feed-forward neural networks, LSTMs can handle temporal dependencies better and do not suffer from gradient vanishing or exploding gradients. When predicting greenhouse gas emissions, LSTM can learn nonlinear relationships in greenhouse gas emission data and make accurate predictions.

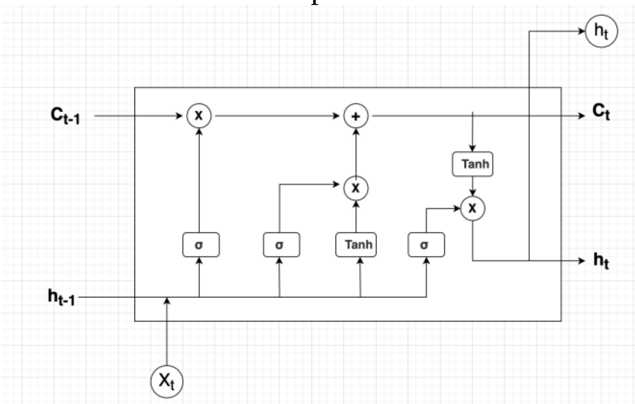

Fig. 2 The internal structure of the LSTM

This study will use Mean Absolute Error (MAE) for evaluation. MAE is used to measure the mean absolute error between the forecast result and the actual observation value, and it is one of the commonly used indicators to measure the accuracy of the forecast model [11]. To calculate the MAE, Take the absolute difference between each anticipated value and its matching actual observed value to determine the MAE. Then, add up and average these differences to determine the MAE.

3. Results and Discussion

For the analysis, this tudy utilized a dataset that compiled information on global greenhouse gas emissions from various nations and totaled those emissions for each year. First, all the data of "world" is extracted (seen from Fig. 3), and then the initial visualization of the data is performed. The purpose of doing this is to facilitate us to have a complete understanding of the data as a whole.

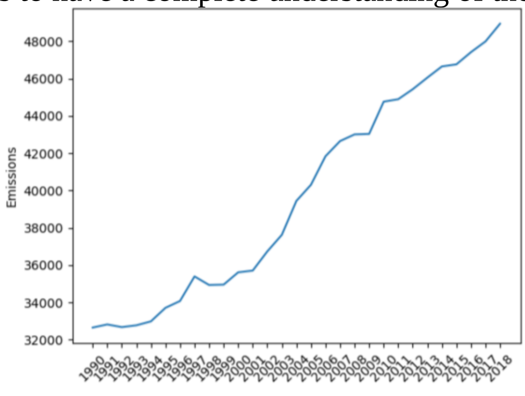

Fig. 3 Historical data visualization.

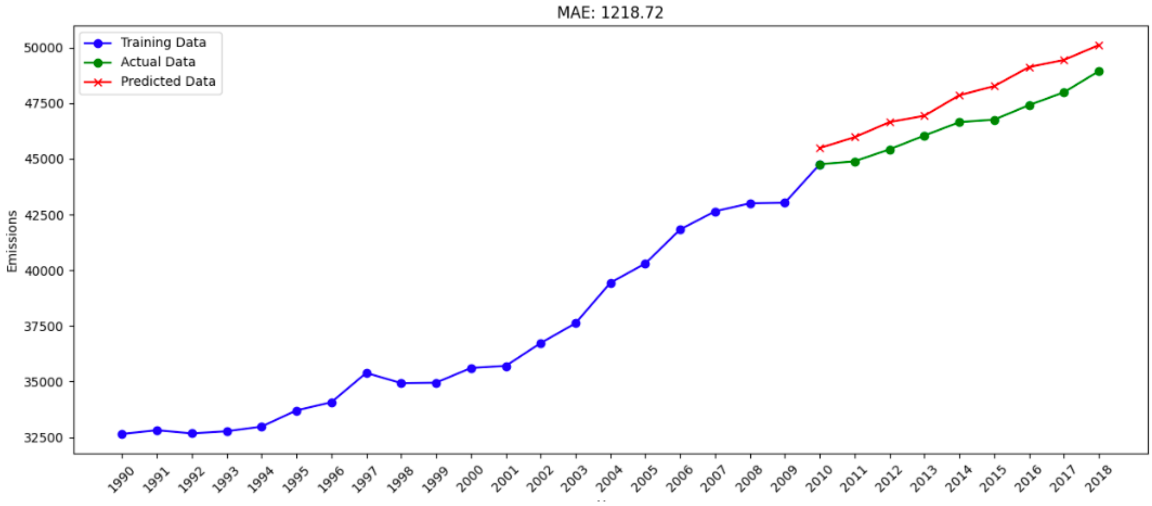

Fig. 4 Predicted GHG Emissions for World in ARIMA

The original data includes data between 1990-2018. This selected the data from 1990-2010 as the training set and the data from 2010-2018 as the test set. Specify the three parameters p, d, q required by the ARIMA model as (5, 1, 5) [12]. Finally, ARIMA model fitting is carried out. The emissions from 2010 to 2018 were forecasted using the training set's data, and the MAE value was then calculated to assess the model's level of accuracy. The prediction results are shown in Fig. 4.

When using the LSTM model for forecasting, select the data from 1990-2010 as the training set, then predict the data from 2010-2018, and finally calculate the value of MAE based on the forecast data and original data from 2010-2018. Before selecting the training set, serialize the emission data to make it more suitable for the LSTM model. The results are shown in Fig. 5.

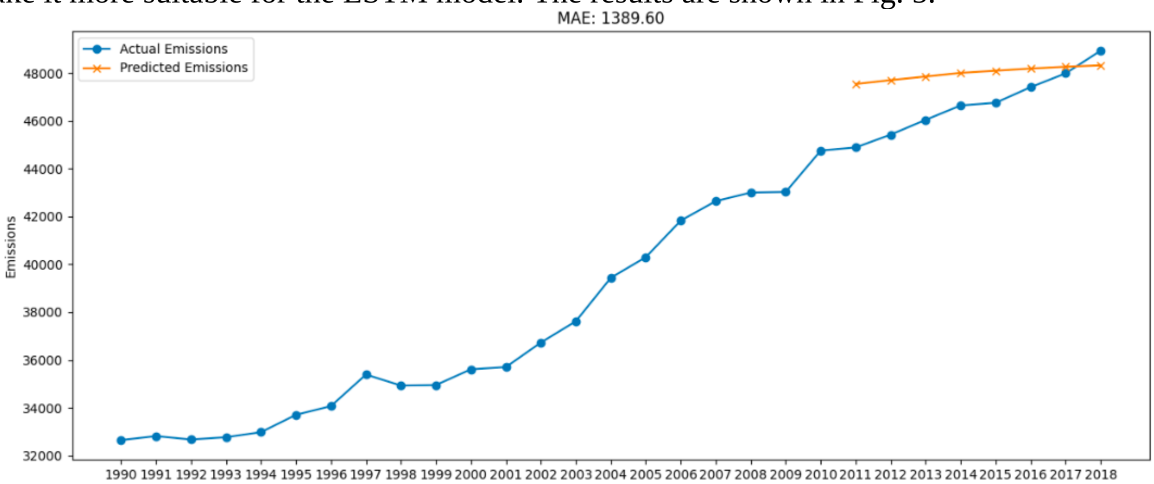

Fig. 5 Predicted GHG Emissions for World in LSTM

By observing the two pictures provided before, one can clearly see that the MAE value using the ARIMA model is 1217.82, and the MAE value using the LSTM model is 1389.60. Here, one needs to explain that the calculated MAE is relatively large because the data used is relatively large, because MAE represents the sum of the absolute value of the difference between the predicted value and the actual value, the MAE result becomes larger. Through the comparison of the two MAEs, the prediction results of greenhouse gas emissions from the ARIMA model are closer to the true values than those from the LSTM model.

4. Limitatons and Prospects

Despite the efforts in completing this research, there are still some limitations to consider. One of the important limitations is the reliability of the data. Existing data on greenhouse gas emissions rely on monitoring and reporting systems that can be inaccurate, inconsistent, and not up to date. This may affect the accuracy of the forecast. Therefore, timely data updates will help improve the accuracy of model predictions. Another limitation is the dependence on the model. Models used to predict greenhouse gas emissions often include assumptions about changes in the economy, technology, etc. And these assumptions are affected by uncertain factors, which may lead to loss of prediction accuracy. In addition, the rate at which the model is updated differs from the actual rate of progress, which may also lead to deviations in the predicted emission levels. Accelerated model updates, improved model accuracy, reliability and predictive power are critical to achieving more accurate GHG emission projections.

5. Conclusion

To sum up, this study explore and compare the effects of using ARIMA and LSTM models to predict greenhouse gas emissions, and use the MAE evaluation metric to compare their prediction performance and accuracy. One can infer the following information and conclusions from this investigation. First, one finds that the ARIMA model exhibits good performance in greenhouse gas emission predictions. The ARIMA model is able to capture trends, cyclical and seasonal characteristics in time series data, so it is effective for forecasting greenhouse gas emission data with obvious time correlation. Furthermore, one compares the predictive performance of ARIMA and LSTM models. Based on the comparison of MAE values, the ARIMA model showed better prediction results than the LSTM model. Finally, although ARIMA and LSTM models show good performance in greenhouse gas emission predictions, there are still limitations such as data reliability, model assumptions, and policy uncertainties. Future research can consider how to improve data quality, speed up model updates, and add more influencing factors to predictive models. In summary, these models are effective and have certain advantages in GHG emission prediction. These findings have important guiding significance for the formulation and optimization of greenhouse gas emission reduction policies and sustainable development planning.

References

[1] Larissa F, Krüger J J, Tarach M, et al. Trend Projections of Greenhouse Gas Emission Reduction Potentials: A Bootstrap-Based Nonparametric Efficiency Analysis. Available at SSRN 4107819, 2022.

[2] Pessoa J P, Santos R A, Chimeli A. Natural Gas Vehicles: Consequences to Fuel Markets and the Environment. NV report, 2023.

[3] Bloch D A. Machine learning: Models and algorithms. Available at SSRN 4493977, 2023.

[4] Sharma S, Saxena A K, Bansal M. Forecasting of GHG (greenhouse gas) Emission using (ARIMA) Data Driven Intelligent Time Series Predicting Approach. 2022 7th International Conference on Communication and Electronics Systems (ICCES). IEEE, 2022: 315-322.

[5] Serrano-Guevara O S, Huertas J I. Greenhouse gas emissions reduction from fuel switching in Heavy Duty Vehicles and its application in the international carbon market. 2022 International Symposium on Electromobility (ISEM). IEEE, 2022: 1-8.

[6] Luo S, Xu S, Luan L, et al. Short-term Prediction of Electricity Consumption in Distribution Network during Peak Summer Period Based on ARIMA Model. 2021 IEEE 10th Data Driven Control and Learning Systems Conference (DDCLS). IEEE, 2021: 328-332.

[7] Putra J A, Basbeth F, Bukhori S. Sugar production forecasting system in PTPN XI semboro jember using autoregressive integrated moving average (ARIMA) method. 2019 6th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI). IEEE, 2019: 448-453.

[8] Al Amin M A, Hoque M A. Comparison of ARIMA and SVM for short-term load forecasting. 2019 9th annual information technology, electromechanical engineering and microelectronics conference (IEMECON). IEEE, 2019: 1-6.

[9] Prakash S, Jalal A S, Pathak P. Forecasting COVID-19 Pandemic using Prophet, LSTM, hybrid GRU-LSTM, CNN-LSTM, Bi-LSTM and Stacked-LSTM for India. 2023 6th International Conference on Information Systems and Computer Networks (ISCON). IEEE, 2023: 1-6.

[10] Lu F, Lv J, Zhang Y, et al. Ultra-Short-Term Prediction of EV Aggregator's Demond Response Flexibility Using ARIMA, Gaussian-ARIMA, LSTM and Gaussian-LSTM. 2021 3rd International Academic Exchange Conference on Science and Technology Innovation (IAECST). IEEE, 2021: 1775-1781.

[11] Jonathan B, Rahim Z, Barzani A, et al. Evaluation of mean absolute error in collaborative filtering for sparsity users and items on female daily network. 2019 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS). IEEE, 2019: 41-44.

[12] Hosseini S M, Saifoddin A, Shirmohammadi R, et al. Forecasting of CO2 emissions in Iran based on time series and regression analysis. Energy Rep, 2019, 5: 619–631.