Climate Change Prediction Using Time Series Analysis

T. Auna Devi[1], M.Esakkiraj[2]
[1, 2] Dept of Master of Computer Application
[1, 2] Francis Xavier Engineering college, Vannarpettai, Tirunelveli.

Abstract- *Predicting climate change is a crucial task that calls for precise and trustworthy data analysis techniques. Time series analysis is one such method that can be used to predict future climate patterns by examining past data. The purpose of this project is to use historical temperature and precipitation data to anticipate climate change patterns using time series analytic techniques. Different time series models, including ARIMA, SARIMA, and exponential smoothing, are used to analyse the data. Different performance metrics, including mean squared error and mean absolute error, are used to evaluate the models. The findings demonstrate that time series models may reliably forecast future climatic patterns, with the SARIMA model yielding the best outcomes for the studied datasets.The findings of this study can be utilized to inform policy decisions and guide future climate change research. Significant trends were found in 15 of the temperature series, and there were trends in precipitation in only five of them.The findings for the trends are addressed in relation to those of other, more in-depth research conducted in the various regions, confirming whether the tendency held steady over time. Twelve-month predictions were created using the ARIMA models by evaluating the errors against the observed data. More than fifty percent of both series were modeled. These models' predictions could be helpful for several seasonal job planning issues, including wildfires, pests and diseases, and agricultural crops.*

Keywords- Time series; Climate change; ARIMA; peninsular Spain

I. INTRODUCTION

In a comparative study of statistical and neuro-fuzzy network models for forecasting the weather of Goztepe, Istanbul using Adaptive Neuro-Fuzzy Inference System (ANFIS) and Autoregressive Integrated Moving Average (ARIMA) models, ANFIS performed slightly better than ARIMA evaluating the RMSE and R2.[1] Ahafo Region of Ghana; Int. J. Afrifa-Yamoah E 2015 Application of ARIMA models in forecasting monthly average surface temperature of Brong Stat. Appl.5(5) 237–246.

Water availability is highly influenced by variability of weather parameters. Minimum temperature and relative humidity are important parameters that have been sidelined in many water resources management projects. In this study, Autoregressive Integrated Moving Average (ARIMA) models were identified and diagnosed in order to forecast minimum temperature and relative humidity of the study area. The findings of the study show that minimum temperature was high during the dry season, when relative humidity was low. [2] Aiyeloyun O and Olodo A 2017 Forecasting one decade ahead minimum temperature and relative humidity for water resources management in lower Niger; J. Water Security.

This paper describes the application of a K-nearest neighbour weather-generating model that allows resampling with perturbation of the observed data to simulate (1) duration of extreme wet spells, and (2) duration of extreme dry spells in the Brahmaputra River Basin. The results of the simulation of extreme events carried out by the K-NN model clearly indicated that the model generated unprecedented extreme events that were not seen in the observed record. [3]Rajagopalan B and Lall U 1999 A k- nearest neighbour simulator for daily precipitation and other variables; Water Resource Res. 35(10) 3089– 3101.Temperature and precipitation fluctuations are the main factors that influence climatic changes. Research assessing climate change is frequently based on the examination of temperature and precipitation time series (Babazadeh and Shamsnia 2014; Balibey and Serpil 2015). In strategic planning for the management of natural disasters like floods and droughts, both of these variables are crucial. Evaluating trends and seasonality in the data is a necessary step in any time series analysis. While seasonality refers to variations in the data at regular short intervals like weekly, monthly, biyearly, quarterly, etc., trends are the long-term growth or decrease in the time series (Pazvakavambwa and Ogunmokun 2013; Wang et al.2013). When arriving at water, temperature and precipitation forecasts are frequently needed. Making water management decisions at the basin scale frequently requires the use of temperature and

precipitation forecasts. For prediction of weather at regional and national levels, various precipitation forecasting methods are available. The best methods for forecasting weather include regression analysis, Auto Regression Integrated Moving Average (ARIMA), genetic algorithms, Adaptive Splines Threshold Autoregressive (ASTAR), Support Vector Machines (SVMs), and K-nearest Neighbour (K-NN). By fitting a regression model, regression analysis determines the degree of relationship between a dependant variable and a number of independent predictor variables. If there are more than two predictor variables, the regression analysis is referred to as a multiple regression analysis. The ARIMA model is widely used in the field of water resources by numerous scientists in recent years, notably in the last few decades, in light of the aforementioned weather forecasting models. The finest Bt model for comprehending the forecast of climatic change, according to them, is Auto Regression Integrated Moving Average (ARIMA) and Seasonal Auto Regression Integrated Moving Average (SARIMA).

II. RELATED WORK

Manfred Mudelsee et al.,[1] The climate during the Cenozoic era changed in several steps from ice-free poles and warm conditions to ice-covered poles and cold conditions. Since the 1950s, a body of information on ice volume and temperature changes has been built up predominantly on the basis of measurements of the oxygen isotopic composition of shells of benthic foraminifera collected from marine sediment cores.

Andreas Hamannet et al.,[2] The quality of ClimateEU estimates was evaluated against weather station data for a representative subset of climate variables. Dynamic environmental lapse rate algorithms employed by the software to generate scale-free climate variables for specific locations lead to improvements of 10 to 50% in accuracy compared to gridded data. We Conclude with a discussion of applications and limitations of this database.

Thomas J.Crowley et al.,[3] Recent reconstructions of Northern Hemisphere temperatures and climate forcing over the past 1000 years allow the warming of the 20th century to be placed within a historical context and various mechanisms of climate change to be tested. Comparisons of observations with simulations from an energy balance climate model indicate that as much as 41 to 64% of pre anthropogenic (pre-1850) decadal-scale temperature variations was due to changes in solar irradiance and volcanism.

Daniel S.Wilks et a.,l[4] While large-scale climate models (GCMs) are in principle the most appropriate tools for predicting climate changes, at present little confidence can be placed in the details of models for investigation of potential impacts of climatic change requires daily data pertaining to small spatial scales, not the monthly-averaged and large-scale information typically available from the GCMs.

Shamshad Ahmad et a.,l[5] The model prediction results show that the forecast data Bts well with the trend in the data. However, over-predictions are found in extreme rainfall events and temperature results. The information of patterns and trends can assist as a prediction tool for development of better water management practices in the area.

N Z A Hamid et al[6] Analysis and prediction of temperature time series is important because temperature changes can affect human's health. The objectives of this study are to analysis and predict the temperature series in Jerantut, Pahang, Malaysia using chaotic approachModelling through chaotic approach divided into two stages; reconstruction of phase space and prediction processes. Through the reconstruction of phase space, a single scalar time series is rebuilt into a multidimensional phase space.

M Uma Maheswar Rao et al[7] The main focus in this study is the area of monsoon track, the coastal region of the eastern part of India. The analysis is done from the climatic data of 20 randomly selected grid points from the study area. The climatic parameters used in this study are Maximum temperature, precipitation, wind speed, relative humidity, and solar radiation. ARIMA short term forecasting model was used to predict all these climatic parameters for the next decade which is hypothesis tested.

Deepa Tyagi et al[2] Rainfall is the most essential stochastic phenomenon which plays an important role in the Indian agriculture sector and is necessary for economic growth of the country. These days, the prediction of rainfall has become a most challenging task as it is drastically affected by climate changes due to the worsen effects of global warming.For the accurate and timely rainfall predictions, in this article seasonal Naive, triple exponential Smoothing and seasonal ARIMA time series models have been applied and the comparison of accuracy of forecasts of these time series models has been checked through various scale dependent error forecast methods and the residual analysis.

Abhijit Kocharekar et al[9] In many areas, accurate projections of future occurrences are crucial, one of which is the tourism industry. Usually counter- trials and towns spend an enormous quantity of cash in planning and preparation to accommodate and benefit visitors. Precisely predicting the amount of visits in the days or months that follow could assist both the economy and tourists.

Israel Goytom et al[10] Climate change is already altering the probabilities of weather hazards. Accurate prediction of climate extremes can inform effective preparation against weather-induced stresses. Accurately forecasting extreme weather events is a task that has attracted interest for many years. Classical and to a lesser extent, machine learning-based approaches have handled this issue; however, such systems are hard to tune or scale.

III. THEORY

The existing work is;the world is grappling with a major problem: climate change. Rising sea levels, melting ice caps and glaciers, more intense storms and hurricanes, more droughts and wildfires, and increased precipitation in some parts of the world while experiencing less precipitation in other parts of the world are all effects of the climate changes already affecting our planet. Flooding from excessive precipitation can result in both property damage and fatalities. Machine learning methods and climate models can be used to forecast extreme precipitation. The ability of climate models to forecast weather and climatic patterns is improving. They still struggle to correctly predict extreme precipitation, though.

The Proposed it that variables that define the weather, such as the minimum and maximum temperatures, the average rainfall, etc., change constantly over time, establishing a time series of each variable. A forecasting model can be created using this time series data.

A 1. Research Methodology

In the beginning, we gathered global temperature data. We also offer a number of essential charting programmes, such as Maplotlib, Seaborn, and Bokeh, as well as the Numpy and Pandas data processing libraries. Last but not least, our models were created using the machine learning programmes XGBoost and scikit-learn. Our data is then denoised and given a clear visual representation. The first issue we identified was the occurrence of NaN values. It was probably much harder to infer the average land temperature over the world in the 18th century. We learn that the datasets is only useful after the year 1800 after utilizing the dropna() method of data cleaning. The fact that collated 19th century temperature data are frequently cited by climate specialists is consistent with this. We are just concerned in the Land Average Temperature and its 95% Uncertainty. We understand that it will be challenging to provide the yearly temperatures if the data is preserved in this format. Therefore, we downsized the data and produced a smaller data frame.

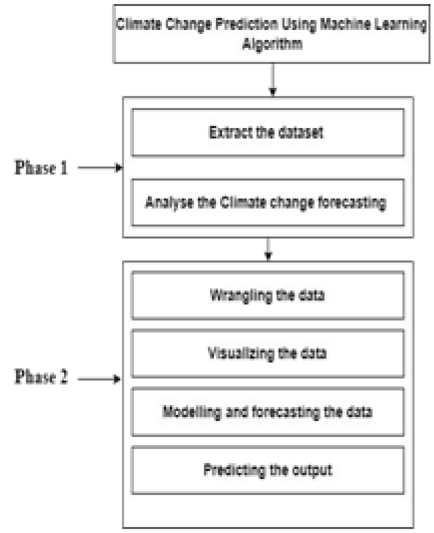


Figure 1: Research methodology

A 2. Algorithm Implementation

TIME SERIRES DATA

cross industries, organizations commonly use time series data, which means any information collected over a regular interval of time, in their operations. Examples include daily stock prices, energy consumption rates, social media engagement metrics and retail demand, among others.

TIME SERIES ANALYSIS

Analyzing time series data yields insights like trends, seasonal patterns and forecasts into future events that can help generate profits. For example, by understanding the seasonal trends in demand for retail products, companies can plan promotions to maximize sales throughout the year. When analyzing time series data, you should undertake a number of steps. First, you need to check for stationarity and autocorrelation. Stationarity is a way to measure if the data has structural patterns like seasonal trends. Autocorrelation occurs when future values in a time series linearly depend on past values. You need to check for both of these in time series data because they're assumptions that are made by many widely used methods in time series analysis. For example, the autoregressive integrated moving average (ARIMA) method for forecasting time series assumes stationarity. Further, linear regression for time series forecasting assumes that the data has no autocorrelation. Before conducting these processes, then, you need to know if the data is viable for the analysis.

During a time series analysis in Python, you also need to perform trend decomposition and forecast future values. Decomposition allows you to visualize trends in your data, which is a great way to clearly explain their behavior. Finally, forecasting allows you to anticipate future events that can aid in decision making. You can use many different techniques for time series forecasting, but here, we will discuss the autoregressive integrated moving average (ARIMA).We will be working with publicly available airline passenger time series data, which can be found here.

TIME SERIES ANALYSIS IN PYTHON

Time series forecasting is a method in the statistics field to analyze historical data with a time component and create a prediction based on it. Some classic examples of time series forecasting methods are Moving Average, ARIMA, and Exponential Smoothing. Time series forecasting is the use of a model to predict future values based on previously observed values.

IV. EXPERIMENTS AND RESULTS

A 1. STIMULATION ENVIRONMENT

The Jupyter Notebook is an open-source web application; that enables data scientists to create and share documents that include live code, equations, computational output, visualizations, multimedia resources, and explanatory text. Using only 'shift + enter' keys, your code in Jupyter Notebook gets executed, which helps to identify whether the code works or not. Because of this, it has become one of the day-to-day tools used by data scientists as it is easy to explore and plot data.Using only 'shift + enter' keys, your code in Jupyter Notebook gets executed, which helps to identify whether the code works or not. Because of this, it has become one of the day-to-day tools used by data scientists as it is easy to explore and plot data.

Colab is a free Jupyter notebook environment that runs entirely in the cloud. Most importantly, it does not require a setup and the notebooks that you create can be simultaneously edited by your team members - just the way you edit documents in Google Docs. Colab supports many popular machine learning libraries which can be easily loaded in your notebook.

A 2.ARCHITECTURE DIAGRAM

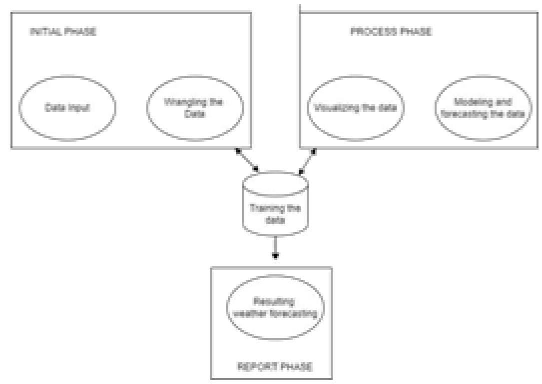


Figure 2: Architecture Diagram

The steps invovled in this processes are,

a) Data Collection:

In this module the raw is collected data from different data set. Then the data set is changed as per need. This raw data cannot be predicted directly. So, it is needed to clean and pre-process.

b) Data Pre-processing:

In this module the data is cleaned. After cleaning of the data, the data is grouped as per requirement. This grouping of data is known as data clustering. Then check if there is any missing value in the data set or not. It there is some missing value then changes it by any default value. Total process before the prediction is known is data pre-processing. After that the data is used for the prediction and forecasting step.

c) Data Prediction and forecasting:

In this step, the pre-processed data is taken for the prediction. This prediction can be done in any process which are mentioned above. But the Linear Regression algorithm scores more prediction accuracy than the other algorithm. So, in this project the linear regression method is used for the prediction. For that, the pre-processed data is splitted for the train and test purpose. Then a predictive object is created to predict the test value which is trained by the trained value. Then the object is used to forecast data for next few years.

d) Visualization:

In this step, the predicted and forecasted data is used to provide a graphical interface separately. At first the predicted data is plotted in a graph separately with the help of matplot library.

A3. OUTPUT SCREEN

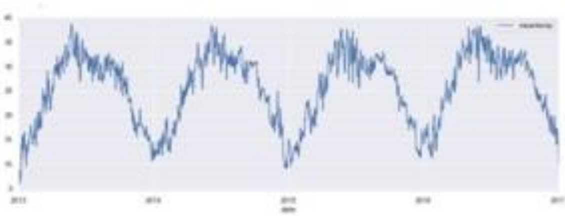


Figure 3 : It is shows kinda of a constant mean and variance.

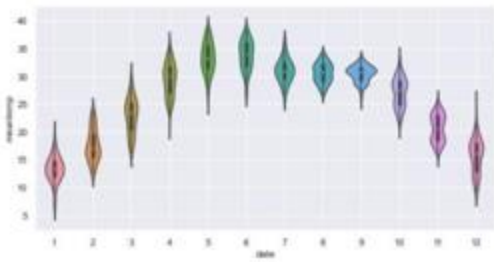


Figure 4: Viol in plot

Figure 5: Trend

Figure 6: Seasonal

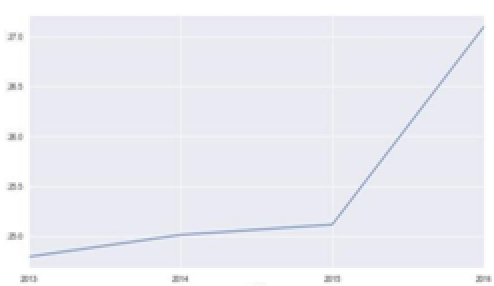


Figure 7::increase rate between 2015 and 2016 with 3 degrees.

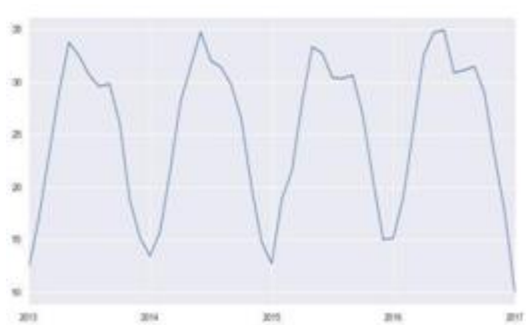


Figure 8: Training the model to predict the data

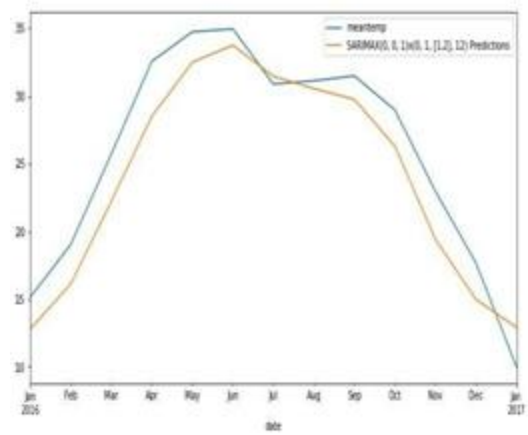
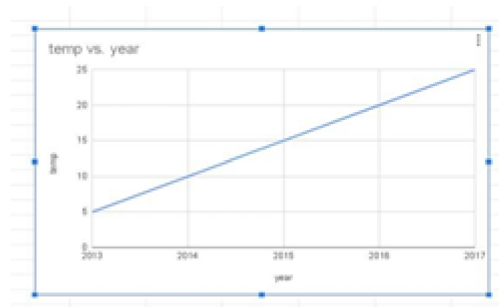


Figure 9 : The average temp in each month throught the year.

A1.Performance Metrics



V. DISCUSSION AND CONCLUSION

climate change caused by the climate's dynamic structure. In the current study, temperature and precipitation data time series were examined, and the best Batted ARIMA model was discovered after seasonality was eliminated. The same model was then used for forecasting. Extreme rainfall occurrences are found to be over-predicted by the

precipitation forecast findings, whereas other rainfall events are found to be predicted correctly.In cases of extreme events, such as precipitation that is either below or above normal, the model outperforms prediction. While it is decreasing over time, the average maximum and minimum temperatures are rising over time. According to this, the colder area at higher elevations is becoming colder while the warmer area at lower elevations is becoming warmer. The study's findings support the conclusion that the results of SARIMA modelling for precipitation and temperature forecasting will help researchers and policymakers create strategies for effective scheduling of flood prediction.

VI. FUTURE SCOPE

Predicting climate change is a complex and difficult problem that calls for extensive data analysis over a lengthy period of time. Time series analysis is a potent method that can be applied to such data to extract meaningful information and forecast future trends. Here are a few probable future developments for time series analysis- based climate change prediction:

Use machine learning methods: The accuracy of time series analysis can be increased by using machine learning methods like support vector machines, random forests, and artificial neural networks. These methods can recognise intricate data patterns and produce more precise predictions.

Include more data sources: Climate change is a global phenomenon that is affected by many different factors. By including data from sources such as satellite observations, ocean measurements, and atmospheric sensors, time series analysis can provide a more comprehensive view of climate change and make more accurate predictions.

Improve data quality: The accuracy and quality of data used in time series analysis can greatly affect the results. Improving data quality by reducing noise, correcting errors, and filling missing data can enhance the accuracy of climate change predictions.Employ ensemble models to increase forecast accuracy. Ensemble models aggregate the outcomes of various models. The risk of over fitting can be decreased and the accuracy of climate change predictions can be increased by using ensemble models.

Including uncertainty analysis: Because of the complexity of the system and the sheer number of variables at play, predicting climate change is inherently uncertain. Time series analysis can produce more accurate and trustworthy projections of future climate change by using uncertainty analysis techniques like Monte Carlo simulations.

REFERENCES

[1] Abdul-Aziz A R, Anokye M, Kwame A L, Munyakazi and Nsowah- Nuamah N N N 2013 Modelling and Forecasting rainfall pattern in Ghana as a seasonal arima process: The case of Ashanti region; IJHSS 3(3) 224–233.

[2] Afrifa-Yamoah E 2015 Application of ARIMA models in forecasting monthly average surface temperature of Brong Ahafo Region of Ghana; Int. J. Stat. Appl. 5(5) 237–246,

[3] Aiyelokun O and Olodo A 2017 Forecasting one decade ahead minimum temperature and relative humidity for water resources management in lower Niger; J. Water Security 3 jws2017006,

[4] Akinbobola A, Okogbue E C and Ayansola A K 2018 Statistical modelling of monthly rainfall in selected stations in forest and savannah eco-climatic regions of Nigeria; J.Climatol.

[5] Babazadeh H and Shamsnia S A 2014 Modeling climate variables using time series analysis in arid and semi-arid regions; African J. Agr. Res. 9(26) 2018–2027.

[6] Bahadir M 2012 The analyses of precipitation and tempera- ture in Afyonkarahisar (Turkey) in respect of Box Jenkinstechnique; J. Acad. Soc. Sci. Studies 5(8) 195–212.

[7] Balibey M and Serpil T 2015 A Time series approach for precipitation in Turkey;GU J. Sci. 28(4) 549–559.

[8] Bari S H, Rahman M T, Hussain M M and Ray S 2015 Forecasting monthly precipitation in Sylhet City using ARIMA Model; Civil Environ. Res. 7(1) 69–78.

[9] Box G E P and Jenkins G M 1970 Time Series Analysis,Forecasting and Control; Holden-Day San Francisco, CA. Buish T A and Brandsma T 2001 Multisite simulation of daily precipitation and temperature in the Rhine Basin by nearest-neighbor resampling; 1 QWater Resour. Res. 37(11) 2761–2776.

[10] Dhawal H and Mishra N 2016 A Survey on rainfall prediction techniques; Int. J. Comput. Appl. 6(2) 1797–2250.

[11] Dwivedi D K, Sharma G R and Wandre S S 2017 Forecasting mean temperature using SARIMA Model for Junagadh City of Gujarat; IJASR 7(4) 183–194.

[12] Hazarika J, Pathak B and Patowary A N 2017 Studying monthly rainfall over Dibrugarh, Assam: Use of SARIMA approach; Mausam 68(2) 349–356.

[13] Huntra P and Keener T C 2017 Evaluating the impact of meteorological factors on water demand in the Las Vegas Valley using time-series analysis: 1990 2014; Int. J. Geo- Inf. 6 249.

[14] Kakade S B and Kulkarni A 2016 Prediction of sumer monsoon rainfall over India and its homogeneous regions; Meteorol. Appl. 23 1–13,

[15] Kaushik I and Singh S M 2008 Seasonal ARIMA model for forecasting of monthly rainfall and temperature; J. Environ. Res. Dev. 3(2) 506–514.

[16] Machekposhti H K, Sedghi H, Telvari and Babazadeh H 2018 Modelling Climate Variables of Rivers Basin Using Time Series Analysis (Case Study: Karkheh River Basin at Iran); Civil Eng. J. 4(1) 78–92.`

[17] Mahmud I, Bari S H and Rahman M T U 2017 Monthly rainfall forecast of Bangladesh using autoregressive integrated moving average method; Environ. Eng. Res. 22(2) 162–168.
[18] Meher J and Jha R 2013 Time-series analysis of monthly rainfall data for the Mahanadi River Basin, India; Sciences in Cold and Arid Regions 5(1) 73.
[19] Murat M, Malinowska I, Gos M and Krzyszczak J 2018 Forecasting daily meteorological time series using ARIMA and Regression models; Int. Agrophys. 32(2) 253–264,
[20] Naz S 2015 Forecasting daily maximum temperature of Umea; Department of Mathematics and Mathematical Statistics, Umea University Master Thesis.
[21] Partheepan K, Jeyakumar P and Manobavan M 2005 Development ofa time-series model to forecast climatic data in the Batticaloa District Sri Lanka; Water Professionals' Day Symposium.
[22] Pazvakawambwa G T and Ogunmokun A A 2013 A Time-series forecasting model for Windhoek rainfall Namibia; UNAM, pp. 1–11.
[23] Rajagopalan B and Lall U 1999 A k nearest neighbour simulator for daily precipitation and other variables; Water Resour. Res. 35(10) 3089– 3101.
[24] Rahman A and Hasan M M 2017 Modelling and forecasting of carbon dioxide emissions in Bangladesh using autoregressive integrated moving average (ARIMA) models; Open J. Statistics 7 560–566