**Time Series Analysis on Climate Data Using SARIMA Model**

P L M W T P Jayasinghe

Faculty of Science, University of Peradeniya

Faculty of Information Technology, University of Moratuwa

[thisarajayasinghe07@gmail.com](mailto:thisarajayasinghe07@gmail.com)

Abstract*—* **The issue of climate change is a sensitive; it impacts the environment and the population. This study therefore employs data on temperature change around the world to examine change with time, with data collected between the years 1983 and 2008. In our analysis we use time series analysis with the SARIMA model, which allows to assess seasonal and long-term fluctuations in temperature. The temperature data obtained in this work is obtained from the Climatic Research Unit (CRU) of the University of East Anglia, and other atmospheric data including CO2, N2O, and methane concentrations are sourced from the ESRL/NOAA Global Monitoring Division. We also look at other parameters such as solar intensity, aerosols and ENSO, which affect global temperatures. This research aims at finding trends in temperature fluctuations in the past and forecast on changes in temperatures in the future. From this analysis, it will be easy for individuals to understand the effects of climate change and thus help cope with global warming.**

Keywords— Climate Change, Time Series Analysis,SARIMA Model

I. INTRODUCTION

Global climate change is one of the most significant issues in the world at the moment. In the past century there have been evidences of global warming, and changes that have affected climate, water levels and other conditions in the world. It is important to understand all of them if one has to know ways on how to respond to such climate issues whether positively or negatively. Global warming and cooling are of particular interest to this research, with regards to the period between 1983 and 2008 being of special interest. Analyzing that period, it is possible to determine the patterns that exist between the years examined and the speed with which global warming has proceeded.

So, as a person that hold a honors degree in Statistics and has knowledge of Time Series Analysis, I decided to do a Time Series Analysis and prediction on climate data.

To do this, we will use a statistical method that is known as time series analysis method. It jointly provides an understanding of the behavior of the data collected on different facilities over the years, such as temperature, which increases year after year. Our goal in drawing similar lines for climate data is to learn something about how global temperatures have in fact behaved and how they may be expected to behave in the future.

One of the practical implications of using the results of this research is that it can speak to policymakers, scientists, and environmental organizations. Knowledge of alterations in the previous temperatures will aid in the formulation of better estimates on subsequent climate circumstances and the steps that individuals ought to undertake to address climate change.

II. LITERATURE REVIEW

[1] Using time series model is quite common in meteorology and hydrology to establish the numerical relationship for several climatic parameters like rainfall, temperature and many others. Out of all the methods available SARIMA Seasonal Autoregressive Integrated Moving Average is proved to be one of the most effective in regards to forecasting time series data that has got a seasonality element in it. SARIMA models have been used In several studies to analyze rainfall patterns and pointed out that the data must be transformed to ensure it is stationary before application of the model.

The paper titled “Modelling and Forecasting of Rainfall Time Series Using SARIMA” uses SARIMA models on monthly, weekly and daily monsoon rainfall data. The procedures highlighted in the study include the Box-Cox transformation and differencing in order to arrive at stationarity. Box-Cox transformation is widely used in the time series regression analysis when there is evidence of non-constant variance in the data or the data is skewed. This serves to help reduce the variance as well in a bid to prepare the data for further analysis. Altogether, the authors used ACF and PACF, which form the basis for modeling selection, and AIC and SBC to select a proper SARIMA model with an appropriate order.

That is why in the literature much attention is paid on checking residuals for randomness. In the analysis part, Q statistics of Ljung-Box were utilized in order to verify that the residuals of the established SARIMA models did not contain a necessary additional modeling. This is actually common in time series models where residuals should not have some form of pattern for the model to hold. In addition, the mean, the standard deviation, and the Nash-Sutcliffe targets which were used to establish the efficiency and accuracy of models in modeling rainfall data proved that SARIMA models were proficiently promising. This is in line with earlier research works that emphasized the importance of statistical test for the times series data in the process of making future forecasts, and that the present study considers the adoption of the F-test to further confirm the ability of the model to make reliable predictions.

Another method used in the study was long-term based where the SARIMA model was estimated on rainfall for up to 14 year from 2014 to 2027. This is the same as other climate prediction analyses where long term projections are essential in understanding future climate patterns and making decisions in sectors such as agricultural, water and land use and other sectors.

However, in this study, we also respond to the equation threshold of heteroscedasticity (ARCH effect) in time series modeling. In Engle’s (1982) article about conditional heteroscedasticity it was established that while testing various time series more often than not large outcomes and small outcomes occur in clusters; therefore, violating homoscedasticity assumption. This is especially true for environmental time series for which the residuals are known to be heteroscedastic. The Box-Cox transformation for variance stabilization was determined to be useful in addressing this concern. This technique has been well applied in many applications such as meteorology, hydrology etc. where accuracy of time series models is essential.

In conclusion, the analysis of the rainfall and other environmental variables using the SARIMA model correlated with the Box-Cox transformation as well as the statistical evaluation techniques presents a good methodological model in the field of the environmental forecasting. The conclusions derived from this study also support earlier research suggesting that SARIMA is a commendable technique for times series forecasting, particularly where there are issues of seasonality. The approach followed in this research offers a good platform for extension of same method on other climatic data for the analysis and modelling of long-term climate changes.

[2] The forecasting of the time series is very useful for the identification of future values depending on the historical data and it is used basically in meteorology finance and others. Forecasting time series with seasonal cycles is quite efficient with use of ARIMA – Autoregressive Integrated Moving Average and its variant, SARIMA – Seasonal ARIMA. The SARIMA model builds on the concept of ARIMA by including seasonal differencing and seasonal autoregressive, and moving average terms and is ideal for climate and weather data.

In the study "Time Series Forecasting of Temperatures using SARIMA: In the paper “The difference and changing point of the structure and strength of loin musculature between endurance runners and sprinters – An Example from Nanjing”, the authors look at the monthly mean temperature for Nanjing from 1951 to 2017 and apply SARIMA. The data for training and testing are divided into two periods: For training, the data is from 1951–2014 and for testing 2015 to 2017. The authors, through the use of the SARIMA model, forecast future temperature changes and assess the accuracy through the calculation of Mean Squared Error (MSE). The results shown here do prove the model and hence the model is justifiable enough as the MSE is still quite low ranging from 0.84 to 0.94 which is an indication of improved fitness of the SARIMA model to the temperature data. These forecasted values are highly accurate to the observed values as shown above further validating the above forecasted model.

The study concentrates on the choice of models for decomposition and forecasting of time series data and reveals that the SARIMA model of temperature data can capture seasonality. Implications of this finding also indicate that SARIMA could be effectively employed for long-term temperature prediction for meteorological and climate study application.

[3] Selected areas include the Indian Sunderbans and the site is the world’s largest tropical mangrove forest, which is vulnerable to climate change and climate change impacts such as SLR and anthropogenic pressures. Sunderbans are potential in terms of ecosystem services and maintains the manifold species of plant and animal. These forests have some special root systems that help in trapping sediments, are vital in coast shield. However, the Sunderbans are facing accelerated rate of land change because of climate change via SLR which has lowered the accretion/erosion undertaking at various factions of the delta.

Some researchers have tried to analyze the effects of SLR on the Sunderbans and most of them have pointed out that while there is erosion on a particular stretch, there is accretion on another because of the difference in the hydro graphical and morphological situations of the area. For instance, where the western sector of the Indian Sunderbans is eroding, the central sector has aggregated, probably through the deposition of fluvial sediments. Moreover, proliferation of barriers upstream decreased the salinity of the river which has led to decrease in the species richness, for example, there is low density of phytoplankton in some sectors of the river and thus the productivity of these sectors will be low.

Another aspect that the study elaborates on is the changes in the values of density and species diversity according to the salinity feature; it is for example shown that the phytoplankton species diversity significantly distinguishes the western and central sectors of the investigated area. Stenohaline species have intruded into the more saline central sector thus changing the ecological environment.

The results presented here demonstrate the influence of climate change and anthropogenic pressures on the Sunderbans’ biota and highlight the freshwater management and mangrove conservation strategies to safeguard this prone ecosystem.

[4] The analysis of global temperatures in the recent work has shown considerable warming in the course of the past century, and the decade from 2011 to 2020 has been considered as being 1,10 ± 0,12C above in average for the period of 1850-1900, thus correlating quite well with the key climate reports. This stability helps to support the quality of temperature data analyzed and used in multiple experiments.

Climate change is described to have occurred in the decade of 2015-2020, and 2016 is estimated to be the warmest year. This occurred during the El Niño phase of 2015/2016, which WHO says has a massive impact on world temperatures. These climatic events therefore have a very important role in short term changes in global temperature and therefore are categorised as factors necessary for calculating yearly alterations.

Globally and regionally, temperatures were above the 1981-2010 temperatures most of the time across the globe. The first decade of the millennium remained the warmest in the history of such indicators for a number of regions, including Africa, Asia, and Europe suggesting that climate change is far from being a universal phenomenon. Awareness of these variations cannot be overemphasized when it comes to the measurement of the impacts of climate change on ecosystems, weather and people.

There were mild-hi cold variations in which 2011 was the coldest year in the decade due to a previous La Nina occurrence. This shows that warm as well as cold climate events should both be taken into account while evaluating temperature changes around the globe. These peculiarities may affect the year average temperatures in a significant way and can belong to the reasons for climate change on the Earth.

A specific sub-trend noted in the Arctic area revealed the highest positive temperature deviations, more than 2 degrees Celsius in the modern period relative to the baseline year of 1981-2010. This disproportionate warming in polar areas has implications for changes in climatic patterns of global warming specially focused on sea level and atmospheric circulation.

[5] Statsmodels is a much wider Python library that greatly enhances the opportunity of time series analysis with an array of statistical tools including AUTORE SSP, AUTORE MAV, and VECTOR AUTORE SSP. Originally designed for linear regression and other types of regression methods named econometric models, Statsmodels new features incorporate the field of time features data used broadly in economy, finance, sales and etc. These models presuppose that the time series data is continuous and equidistant, and contains no gaps and that these are ideal for forecasting and examining a temporal structure. Though, in the Statsmodels project the improvements in this aspect look rather considerable now, the perspectives of its further development are to focus on the better integration and, especially, usability and efficiency enhancements, with the ultimate goal to make Python into one of the top environments for the applied statistics and econometrics, where the models in question are of paramount importance.

III. METHODOLOGY

In this research, we employ the SARIMA model to estimate global temperature data over the time period of 1983 to 2008. These models are commonly known as time series models of which SARIMA is an example in use to analyze and forecast data. This is an acronym of Seasonal Autoregressive Integrated Moving Average, in which the figures refer to the fact that the model is sensitive to periodic variations and trends.

First, we import the temperature data into our computer in the programming language we’re using, which is Python. Yearly temperature data is found in the data set ranging from the year 1983 to the year 2008. In order to model our data, before applying the SARIMA model it is best to check the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). These functions allow us to recognize what kind of relation is between the data points and define which parameters of SARIMA are to be used.

ACF demonstrates how each point in the data is about related to the previous points in the data. This assists us in choosing the ‘moving average’ component of the model.

PACF clarifies the relationship of the data points after factoring out the previous data point influences. This allows us to select or, more accurately, discard all of the “autoregressive” components of the model.

After determining the best parameters to use, we fit the SARIMA model into the data. Seasonal variations are incorporated also in the model to enable the understanding of the changes in temperature during a particular season. Having fitted the model, we make forecasts about the subsequent temperatures.

This approach helps one to understand how temperatures around the globe have varies in the past and can be used to estimate how they will vary in the future. SARIMA can help us control for long-term structural shifts as well as cycles of temperature fluctuations.

III. DATA

For this research, we use a data set that consist of several variables that are in form of climate and atmospheric condition of the world. The key variable we pay attention is Temperature, which can be described as the deviation in degrees Celsius of global mean temperature of a particular epoch from a standard value. The temperature data presented here is time-series data, ranging from January, 1983 to December, 2008 and sourced from CRU, University of East Anglia.

Besides many other variables which describe certain atmospheric conditions, the dataset also contains several other essential items. These variables are important because they affect temperature at which the world we live is warmed.

They are as follows

Year - The year of the observation.

Month - The month of the observation.

Temp - The temperature difference (in degrees Celsius) between the average global temperature and a reference value.

During the analysis, I merged Year and Month columns to a single YearMonth column.

CO2 (Carbon Dioxide) - The concentration of carbon dioxide in the atmosphere, measured in parts per million by volume (ppmv).

N2O (Nitrous Oxide) - The concentration of nitrous oxide in the atmosphere, measured in parts per million by volume (ppmv).

CH4 (Methane) - The concentration of methane in the atmosphere, measured in parts per million by volume (ppmv).

CFC.11 (Trichlorofluoromethane) - The concentration of CFC-11 in the atmosphere, measured in parts per billion by volume (ppbv).

CFC.12 (Dichlorodifluoromethane) - The concentration of CFC-12 in the atmosphere, measured in parts per billion by volume (ppbv).

These atmospheric concentration data points are obtained from the ESRL/NOAA Global Monitoring Division.

Additionally, the dataset contains other variables that help in understanding factors affecting climate

Aerosols - The mean stratospheric aerosol optical depth at 550 nm. This is linked to volcanic activity, as volcanic eruptions add particles to the atmosphere that reflect the sun’s energy.

Data source - Godard Institute for Space Studies at NASA.

TSI (Total Solar Irradiance) - The total amount of energy received from the sun, measured in watts per square meter (W/m²). Solar activity, including sunspots, affects this variable.

Data source - SOLARIS-HEPPA project website.

MEI (Multivariate El Nino Southern Oscillation Index) - A measure of the strength of the El Nino/La Nina phenomenon, which affects global weather patterns.

Data source - ESRL/NOAA Physical Sciences Division.

More specifically, the Temperature column is the feature of interest for this study because it contains the information on the shifts in the Earth’s temperature. The other variables are relevant in giving background information which can be used in further analysis concerning the change in global temperature.

I. References

[1] P. P. Dabral and M. Z. Murry, "Modelling and forecasting of rainfall time series using SARIMA," *Environmental Processes*, vol. 4, pp. 399–419, 2017, doi: 10.1007/s40710-017-0226-y.

[2] T. Dimri, S. Ahmad, and M. Sharif, "Time series analysis of climate variables using seasonal ARIMA approach," *Journal of Earth System Science*, vol. 129, no. 1, 2020, doi: 10.1007/s12040-020-01408-x.

[3] A. Raha, S. Das, K. Banerjee, and A. Mitra, "Climate change impacts on Indian Sunderbans: a time series analysis (1924–2008)," *Biodiversity and Conservation*, vol. 21, no. 5, pp. 1289–1307, 2012, doi: 10.1007/s10531-012-0260-z.

[4] J. Hansen, et al., "Global temperature change," *Proc. Natl. Acad. Sci.*, vol. 103, no. 39, pp. 14288-14293, 2006.

[5] W. McKinney, J. Perktold, and S. Seabold, "Time Series Analysis in Python with statsmodels," *SciPy*, 2011.