OPEN ACCESS Freely available online

Journal of Climatology & Weather Forecasting

Research Article

Forecasting Rainfall in Tanzania Using Time Series Approach Case Study: Dar es Salaam

PAPanga[1*], S.M.Nyimvua[1], I.S.Mbalawata[1]

Department of Mathematics, National Institute of Transport (NIT), Tanzania, East Africa

ABSTRACT

The prediction of rainfall on monthly time scale has been attempted by a number of researchers by using different time series techniques at different time periods around the world. It is challenging to forecast rainfall at monthly time scale because of spatial and temporal random variation caused by a numbers of dynamic and environmental factors. In this paper, an attempt has been made to develop a Seasonal Autoregressive Integrated Moving Average (SARIMA) Model to analyze long term monthly rainfall data of Dar es Salaam region in Tanzania for the period of fifty three years (1961 to 2014). Rainfall observations were discovered to have Seasonality and also non-stationarity and hence differencing and Seasonal differencing was used to attain stationarity. Rainfall data were found to have two seasons namely October to December (OND) and March to May (MAM) .The analysis exhibited that the Seasonal ARIMA model which is satisfactory in describing the monthly rainfall data in Dar es Salaam Tanzania is SARIMA (2, 1, 1)(1, 1, 1)12. The model was then used for predictions of monthly rainfall values from January 2015 to December 2024. The forecasting results showed that monthly rainfall values have a decreasing trends, hence that may be a threat to agriculturists and water managers in the region. The study will be useful to decision makers for the region of Dar es Salaam to establish priorities and strategies based on the impacts posed by the variation of rainfall.

Keywords: Seasonal arima models; Forecasting; Time series; MAPE; RMSE

INTRODUCTION

For a decades now, the trends of rainfall instability and its impacts has been a crucial climatic problems facing different nations. The scenario has been linked directly or indirectly with global warming, which pose its impacts to a number of sectors particularly agriculture and tourism whose contribution is vital to any countries economy. Vital sectors of the Tanzania economy such as agriculture, fishing, to mention few, depend on rainfall. Considering this case for the survival and growth of the plants, rainfall is required, though too much or too little is still a problem. On the other hand, heavy precipitation is associated with floods, loss of people's lives and outbreak of diseases. So it is unavoidable to have untimely effects of dynamics of rainfall patterns. The United Republic of Tanzania alludes that the impacts of rainfall instability will persist to torture agriculture, biodiversity, livelihood, health and other sectors. Hence, early indication may help to solve a number of problems associated with dynamic of rainfall trends.

Variation of climate has been a topic in many parts of the world due to its immediate effects on people's lives. Over the past decades Tanzania has witnessed the increase of climatic events such as floods, which are linked with grievous ecological and socio-economic intimidation like loss of lives and destruction of structural design. Serious floods that have recently tortured many parts of the Tanzania include that of 2006, 2009, 2010, 2011, 2012, 2014, 2016 and 2017. It is evident that the coastal areas of Tanzania including Dar es Salaam region will encounter a number of damages due to increasing trend of temperature plus instability in precipitation. The trends of precipitation in the region is dynamic and there is variability in rainfall caused by a number of different time scales from daily to decadal. It has been noticed that the trend and variability of rainfall will continue at a longer timescale. Therefore, there is a need for a suitable prediction method to be applied in forecasting precipitation patterns.

***Correspondence to:** Panga PA, Department of Mathematics, National Institute of Transport (NIT), Tanzania, East Africa, E-mail: pangapaul16@gmail.com

Received: February 15, 2021; **Accepted:** March 01, 2021; **Published:** March 08, 2021

Citation: Panga PA (2021) Forecasting Rainfall in Tanzania Using Time Series Approach Case Study: Dar es Salaam. J Climatol Weather Forecast 9: 272.

Recently, Time series analysis and forecasting was observed to be an important tool when applied in studying the variations and trends of different hydo-meteorological variables such as precipitation, humidity, temperature, streamflow and many other environmental parameters. Various published papers have analyzed precipitation by using Time series Box and Jenkins ARIMA and SARIMA approaches, which gives the usefulness of modelling rainfall from different parts of the world. Most of the observations and time series modelling results of the mentioned studies have declared projected instability in rainfall patterns. However, there are limited or no published papers that have attempted to understand, analyze, model and predict rainfall by using Box and Jenkins ARIMA approach in Tanzania particularly Dar es salaam. Therefore, this paper would seem to be the first application of the Box and Jenkins ARIMA approach for rainfall in Dar es Salaam, Tanzania.

In this study, we first check if the condition of stationarity in the time series data is attained, then followed by finding the appropriate time series model for monthly rainfall data by using previous available data from January 1961 to December 2014 of Dar es Salaam region in Tanzania. Second, we will check if the parameters are sensitive to the time series models and finally, we will predict the future trends of rainfall values by using the time series model developed. Box and Jenkins methodology will be used in developing the time series model. The approach flows through identification of the model, estimation of the model parameters, diagnostic checking of the selected model and lastly the use of the model in forecasting purposes.

Different researchers allude that socio economic development of the developing countries like Tanzania are hindered by the trends and patterns of climatic extremes. Efforts like achieving Millennium Development Goals (MDG), Sustainable Development Goals (SDG) and National Developmental Vision (Visions 2025) which are associated with reducing poverty, hunger and promoting food security are hampered by floods and natural disasters like drought, thus if not managed properly the prolonged impacts will continue in the future. Hence this study is of practical importance for providing information to decision makers, planners, climatologist, meteorologist and others in predicting the future rainfall. The paper is organized as follows. Section 2, describes study area, data and methodology for fitting time series models. Finally, the results of the appropriate time series model and their prediction are discussed in Section 3 and finally we give conclusion.

MATERIALS AND METHODS

Data collection

The Tanzania Meteorological Agency (TMA), Tanzania, is the responsible organization for the collection and publication of meteorological data. It is a government agency responsible for meteorology issues in the country. The data used in this paper are completely secondary in nature and they are collected from TMA. The daily rainfall data from the period of January 1961-December 2014 of Dar es Salaam region of Tanzania are used. The given daily rainfall data was converted into monthly data by using the totaling approach. Since rainfall data are time dependent data then monthly data was converted into time series data and then smoothing moving average of five data points was applied.

Study area

Dar es Salaam is one among the thirty regions of Tanzania, lying at the latitudes of 6˚52 South and longitudes of 39˚12 East. It is among highly populated coastal regions with population of 6,368,272 covering the area of 1,393km. The region constitutes of five districts which are Kinondoni, Ubungo, Kigamboni, Ilala and Temeke.

Dar es Salaam region is characterized by tropical type of climate with higher degree of hotness, high humidity and average annual precipitation of over 1000 mm. The region is characterized by bimodal rainy seasons. The longer rain falls from March to May (MAM) and shorter rains fall from October to December (OND). The map of Tanzania and the extract of Dar es Salaam region from the map are exhibited in Figure 1.

Figure 1: A map of Tanzania and the extracted map of Dar es Salaam.

Moving average smoother

A smoother is one of the most useful tools for expressing the trends of the dependent variable as the function of one or more regressors. Smoother is used when it happens the amount of horizontal scatter in data places difficulties in seeing the trends. The methods is commonly used in univariate time series analysis.

According to Time series is a sequence of data arranged in chronological order. Fundamental principal of the time series is to understand the historical trends of the data at a particular time. Normally if the previous values are well described then they can be used to forecast the future values of the series. Time series data can either be discrete if it is recorded at a discrete time point or continuous if it is recorded at every instance of time. Time series observations for rainfall and temperature which are recorded continuously can be converted to discrete time points. Mathematical technique for modelling precipitation values is a stochastic process. A number of probability models have been developed to understand weekly, monthly and annual precipitation, however nowadays monthly rainfall are analyzed by using time series models. Time series models have been thoroughly studied by Box and Jenkins (1976). Till now, a number of researchers still use this method due to its effectiveness in forecasting purposes.

Box and jenkins algorithms: Back in 1976, Box and Jenkins give a methodology (Presented in the Table below) in time series analysis to find the best fit model using the previous values to give the predicted values. One of the advantages of Box and Jenkins ARIMA time series model is that, it has an ability to generate the sequence of historical data and produce mathematical formula which will then be used to generate forecasted values. Also some articles have approved Box and Jenkins methodology as a very strong tools for giving solution of the prediction problems due to its ability to provide very tremendous correct prediction of the time series and also it yields a framework to develop the model and do analysis. The aims of using Box and Jenkins Prediction approach are to look for suitable formula that will force the error term to show no change in pattern and must be as small as possible. In this study the approach is used to develop the model and do prediction of rainfall values. The Conceptual framework of Box and Jenkins modelling approach is given in the Table below.

1	Plot the Series
2	Is variance stable?
3(a)	No, apply transformation, go to 1
3(b)	Yes, continue
4	Obtain ACFs and PACFs
5	Is mean stationary?
6(a)	No, apply regular and seasonal differencing
	Yes, continue
6(b)	Model selection
7	Estimate parameter values
8	Are the residual uncorrelated?
9	No, modify the model, go to step 5
10(a)	Yes, continue
10(b)	Forecast
11	

Table1: Box and Jenkins Algorithms.

Stationarity check: Stationary time series is attained when the probability distribution properties such as mean and variance remain fixed at a time, and they must not be uncorrelated in the series. It should be noted that, before starting any time series modelling, you must check whether or not the series is stationary and uncorrelated. A number of methods can be used to convert non-stationary time series into stationary time series. These methods are Augmented Dickey Fuller Tests (ADF), Phillips-Peron (PP) and Kwiatkowski-Phillips-Schmidt- Shin (KPSS) test, ACF and PACF plots.

Parameter estimation: After completing identifying the tentative model, the next step is to estimates the parameters of the model. The model parameters are estimated by using the maximum likelihood estimates (MLE). Normally MLE is used in finding the parameters that maximize the probability of observations.

Model selection: The selection of the model to be used is based on the criterion test. The Autocorrelation and Partial autocorrelation function have shown the usefulness in identifying the orders of the model. They also help in proposing where the model comes from although the required model is chosen based on the test statistic such as Bayesian Information Criterion (BIC) and Akaike Information criterion (AIC).

Diagnostic checking: The selected models must have as smaller errors as possible. Normally the model diagnostic checking is accomplished by carefully analysis of the residual series, histogram of the residual, normal QQ plots and diagnostic test. So in order to check whether the model residuals follows a white noise property (that is the process is stationary and independent) then the following test is used.

Forecasting: In order to make the correct decision about the time series prediction, suitable forecasting tools are required. The suitable selected model is not the criteria that the model is the best for prediction purposes. So, in order to get the appropriate forecasting model, measures of errors such as Mean Absolute Error (MAE), Mean Square Error (MSE) and Root Mean Square Error (RMSE) must be performed well to be assured that the obtained model is exactly the required model for forecasting the time series observations.

RESULTS AND DISCUSSIONS

Time plot for rainfall data

Variation in time of the rainfall values are presented in the Time Plot. The Plot exhibited a set of values taken at different time points and graphed in a time series Figure 2. The Smoothed plot for rainfall data in Dar es Salaam region was shown in Figure 1. From the plot, we can observe that there is seasonal cycle in the series at the same time the variance were observed to be more varying from the mean, hence this indicates that the rainfall time series is not stationary. However, for the case of trends, it is not easy to depict it. Clearly the rainfall plot seems to have the strong yearly circle.

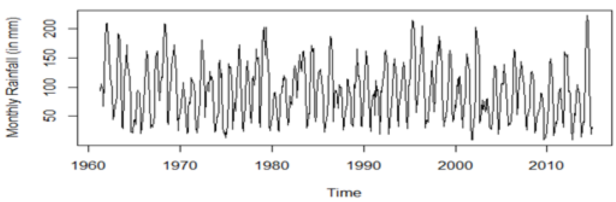

Figure 2: Smoothed time series plot of monthly rainfall from January 1961 to December 2014.

Seasonal ARIMA modelling

The decomposition of monthly rainfall time series data was plotted in order to see whether the time series has trends, seasonal, cyclic and random components. We plotted Year on X-axis and the observed monthly rainfall on Y-axis Figure 2. However it has been observed that it is difficult to interpret the trends based on visual inspection technique. So in order to scrutinize the trends, we decompose the average monthly rainfall data by additive decomposition approach by the statistical software R as observed in Figure 3. From the decomposition upward trends are depicted in some years, for instance from 1961 to 1970 and 1976 to 1980. In Figure 3, it seems that there is sturdy seasonal cycle in the monthly rainfall data set.

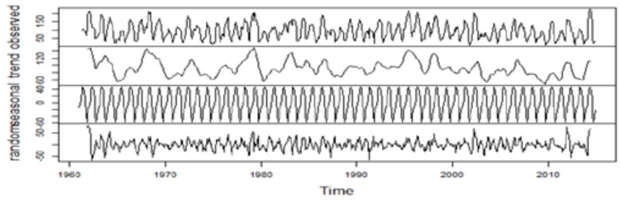

Figure 3: Decomposition of Smoothed Monthly Rainfall Series.

Ordinarily, the Box and Jenkins methodology works under assumptions that the time series are stationary and serially correlated. In this study, graphical inspection and unit root tests were the most selected techniques for stationarity test of monthly rainfall data. In Figure 4, the ACF and PACF plots are drawn using 60 lags on X-axis and values for autocorrelation on the Y-axis. The seasonal autocorrelation relationships dominate the two plots. Hence due to the appearance of strapping seasonality and upward trends, we conclude that the average monthly rainfall series is not stationary. This result was supported by ADF test statistics.

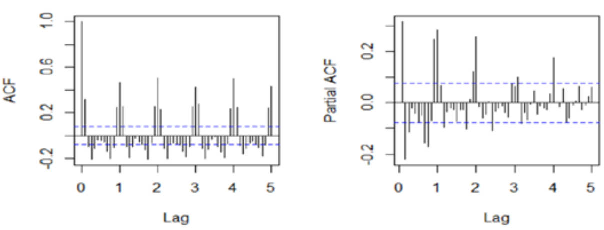

Figure 4: Autocorrelation and Partial autocorrelation function of Monthly Rainfall.

Based on Box and Jenkins approach, before fitting the model either ARIMA or SARIMA, we must make sure that the stationarity condition is attained. This is accomplished by performing seasonal differencing of the average monthly rainfall data, so as to eliminate seasonal characteristics. After conducting seasonal differencing series we use Graphical approach (ACF and PACF) and Unit root test (ADF and KPSS tests), in order to check if the stationarity condition is achieved. For the ACF and PACF plots shows that as the number of lags increases there is slow decays of ACF and PACF plots also most of the lags are outside the 95 percent confidence limits, which is the confirmation that the series is not stationary. Also the unit roots test (ADF and KPSS) gives p-values of 0.07 and 0.01. The interpretation for these two test is that, for ADF the p-value is greater than 0.05 which means the null hypothesis should not be rejected, also for the KPSS test the p-values is less than 0.05 which means that the null hypothesis should be rejected. Hence it shows that the series are not stationary. Thus after performing seasonal differencing the seasonal characteristics is not observed any more Figure 5. Most of the spikes lies within the confidence limits except very few individual correlations appear larger compared with the confidence limits. At the non-seasonal level, results show that ACF spikes at lag 1, and goes off after lag 1 while PACF exhibit a significant spike at lag 1 and cut off at lag 1. At the seasonal level, ACF are observed to have spikes at lag 12 and also cut lag 12 while PACF goes off after lag 12. However little number of lags are faintly observed outside the confidence limits. Also the statistical tests for stationarity results show that P-value for ADF test is 0.01 and P-value for KPSS is 0.1 which is greater than 0.05 (level of significance), so we cannot reject the null hypothesis of trends-stationarity. The results from these tests concludes that the monthly rainfall differenced series is stationary.

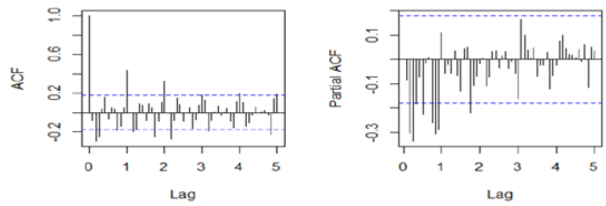

Figure 5: ACF and PACF plot for seasonal differenced Monthly Rainfall series.

Model validation

In order to check accuracy and forecasting capability of the picked model, the actual values and the fitted ones were plotted together and presented in the Figure 6. The rainfall data from January 2005 to December 31, 2014 were designed as the test sets and were used to assess the ability of the models to fit the original data. The red and blue lines are the fitted and actual values respectively. The plot exhibited that, the fitted values are very close to the original data. This indicates that the selected model for monthly rainfall is the better one for the set of data.

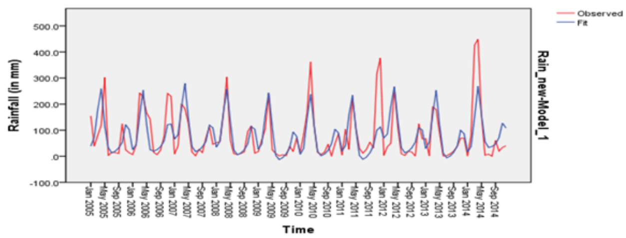

Figure 6: Observed and fitted values of Monthly Rainfall series.

CONCLUSION

The study of Rainfall variability is very important since rainfall is one of the crucial elements of climate parameters. In this study we use Time series analysis, where the monthly rainfall data were observed to have stochastic trends, seasonal variation and random movements. High fluctuation from central mean were observed from descriptive statistics. The monthly data followed model which gives lowest AIC, AICc and BIC. Based on diagnostic checking, the model were considered to be the best model with stationary, white noise and few outliers of residuals. Then the model was used to forecasts monthly rainfall values for ten years that is from 2015 to 2024. The research results has divulged a decreasing trends that may be a harm to sectors like agriculture, tourism, to mention few. Hence this study is of useful for policy implications of Dar es Salaam region in Tanzania and the rest of the world which share the same equatorial zone.

REFERENCES

1. Mohanty SK, Sai Leela K. Textbook of Immunology, Jaypee Brothers Medical Publishers Ltd, Daryaganj, New Delhi, 2014.
2. Yuen GJ, Demissie E, Pillai S. B lymphocytes and cancer: a love-hate relationship. Trends Cancer 2016;2(12): 747-757.
3. Maccalli C, Volonte` A, Cimminiello C, Parmiani G. Immunology of cancer stem cells in solid tumours: A review. Euro JCancer 2014;50(3): 649-655.
4. Sinha PK. The common genesis of all cancers. JCarcinog Mutagen 2019;10(3):342.
5. Corces-Zimmerman MR, Majeti R. Pre-leukemic evolution of hematopoietic stem cells: the importance of early mutations in leukemogenesis. Leukemia 2014;28(7): 2276-2282.
6. Watson NFS, Ramage JM, Madjd Z, Spendlove I, Ellis IO, Scholefield JH, et al. Immunosurveillance is active in colorectal cancer as downregulation but not complete loss of MHC class I expression correlates with a poor prognosis. Int J Cancer 2006;118(1) :6-10.
7. Yoo SH, Keam B, Ock CY, Kim S, Han B, Kim JW, et al. Prognostic value of the association between MHC class I downregulation and PD-L1upregulation in head and neck squamous cell carcinoma patients. Scientific Rep 2019; 9(5):7680.
8. Goeppert B, Frauenschuh L, Zucknick M, Roessler S, Mehrabi A, Hafezi M, et al. Major histocompatibility complex class I expression impacts on patient survival and type and density of immune cells in biliary tract cancer. British J Cancer 2015;113(3):1343-1349 (2015).
9. Lee JH, Shklovskaya E, Lim SY, Carlino MS, Menzies AM, Stewart A, et al. Transcriptional downregulation of MHC class I and melanoma de- differentiation in resistance to PD-1inhibition. Nature Comm 2020;11(4):1897.
10. Byrne EH, Fisher DE. Immune and molecular correlates in melanoma treated with immune checkpoint blockade. Cancer 2017;123(11): 2143-2153.
11. Huang X, Zhang X, Bail X, Liang T. Eating self for not be eaten: pancreatic cancer suppresses self-immunogenicity by autophagy-mediated MHC-I degradation. Signal Transd Targ Ther 2020;5(6):94.
12. Inoue M, Mimura K, Izawa S, Shiraishi K, Inoue A, Shiba S, et al. Expression of MHC Class I on breast cancer cells correlates inversely with HER2 expression. Onco Immun 2012;1(7):1104-1110.
13. Zagzag D, Salnikow K, Chiriboga L, Yee H, Lan L, Ali MA, et al. Newcomb, downregulation of major histocompatibility complex antigens in invading glioma cells: stealth invasion of the brain. Lab Invest 2005;85(3):328-341.
14. Morrison BJ, Steel JC, Morris JC. Reduction of MHC-I expression limits T-lymphocyte-mediated killing of Cancer-initiating cells. BMC Cancer 2018;18(4):469.
15. Cornel AM, Mimpen IL, Nierkens S. MHC class I downregulation in cancer: underlying mechanisms and potential targets for cancer immunotherapy. Cancers 2020;12(7):1760.
16. Sheikhpour E, Noorbakhsh P, Foroughi E, Farahnak S, Nasiri R, Neamatzadeh H. A survey on the role of interleukin-10 in breast cancer: a narrative. Repo Biochem Mole Biol 2018;7(1):30-37.
17. Gonzalez Garza MT, Cruz Vega DE, Bernal CM. IL10 as Cancer Biomarker. Translational Res Cancer 2019;10(2)25.
18. Seyfried TN, Huysentruy LC. On the origin of cancer metastasis. Crit Rev Oncog 2013;18(2):43-73.