Prediction of Malawi Rainfall from Global Sea Surface Temperature Using a Simple Multiple Regression Model

Chisomo Patrick Kumbuyo¹, Katsuyuki Shimizu², Hiroshi Yasuda³ and Yoshinobu Kitamura⁴

Abstract: This study deals with a way of predicting Malawi rainfall from global sea surface temperature (SST) using a simple multiple regression model. Links between Malawi rainfall and SST based on statistical correlations were evaluated and selected as predictors for the model. Monthly rainfall data from nine stations in Malawi grouped into two zones on the basis of inter-station rainfall correlations were used in the study. The predictors for Zone 1 model were identified from the Atlantic, Indian and Pacific oceans while those for Zone 2 were identified from the Indian and Pacific Oceans. The correlation between the fit of predicted and observed rainfall values of the models were satisfactory with r = 0.81 and 0.54 for Zone 1 and 2 respectively (significant at less than 99.99%). The results of the models are in agreement with other findings that suggest that SST anomalies in the Atlantic, Indian and Pacific oceans have an influence on the rainfall patterns of Southern Africa. We conclude that SST in the Atlantic, Indian and Pacific Oceans is correlated with Malawi rainfall and can be used to predict rainfall values.

Keywords: Malawi rainfall; forecast model; predictors; model validation; SST

1 Introduction

The economy of Malawi is heavily dependent on agriculture which is mostly rain-fed and occupies 86% of its workforce, and making up 38% of its GDP and 90% of its export earnings (World Bank, 1990). The total rainfall received determines the type of farming and which crops can be cultivated. The failure of rains for more than one month in the rainy season impacts heavily on agriculture and the economy of Malawi.

Rainfall in Africa varies both spatially and temporally. Previous studies have reported that variability of African rainfall is strongly related to sea surface temperature (SST) over the Atlantic, Indian and Pacific Oceans. Hirst and Hastenrath (1983), Nicholson and Entekhabi (1987) and Nyenzi (1988) have conducted studies that have looked at the relationship between SST in the global oceans and African rainfall. Mason (1992) performed correlation analysis on rainfall and SST with easterly (westerly) quasi biannual oscillation (QBO) and found that the Ocean region associated with greatest rainfall response was the Indian Ocean south east of South Africa (QBO easterly), followed by south Atlantic Subtropical Convergence region with westerly QBO and lastly Benguela system (South East Tropical Atlantic) with westerly QBO. According to Jury and Pathack (1991) the correlation between SSTs and convection (Outgoing Longwave Radiation) shows that a warming of waters within the cyclogenesis region northeast of Madagascar triggers decreased rainfall in Southeast Africa. According to Jury et al. (1991), the increased low level westerly anomalies off the northern tip of Madagascar reduces the inflow of moisture and subsequently limits rainfall over southern Africa. Likewise low-level easterly anomalies south of Madagascar in the region 25–35°S and 40–55°E, result in cyclonic circulation centered over Madagascar in dry summers and oppose convective outflows. Strong convective activity across equatorial eastern Africa, northeast Madagascar and the Southwest Indian Ocean matches with below normal convective activity across southern Africa.

The recurrence of droughts over southern Africa in recent years has contributed to an increased interest in research scientists to focus more attention on the study of climate variability and its predictability (Mwalulirwa, 1999). The farming community in Malawi requires an advance warning of drought/floods; however, the oceanic thermodynamic patterns and corresponding atmospheric circulation associated with extreme weather conditions are not well known. Statistical models for the prediction of seasonal rains are unverified (Jury and Mwalulirwa, 2002).

Some of the approaches used for long range forecasting of rainfall are; (a) statistical method, which uses the historical relationships between rainfall and global atmosphere parameters (Sahai et al., 2003); (b) empirical method, which uses time series of past rainfall data (Iyengar and Raghukanth, 2004); And (c) dynamical method, which uses Atmospheric General Circulation Models (AGCM) and oceans to simulate rainfall (Wang et al., 2005).

Malawi relies on the Southern Africa Regional Climate Outlook Forum (SARCOF), a regional grouping of climate scientists from Southern Africa Development Community (SADC) member states to develop seasonal forecasts. SARCOF meets annually before the start of the rainy season to develop seasonal forecasts for the region. The forecast covers the rainfall season from October to March and is relevant to seasonal time scales and relatively large
areas. The forecast is based on dynamic and statistical models that use scientifically established relationships between rainfall over southern Africa and SST.

Literature on studies about rainfall forecasting in Malawi is scarce. Jury and Mwafulirwa (2002) is one notable study that was carried out to predict Malawi rainfall. The study looked at the climate variability in Malawi, statistical associations with global circulation features and its predictability. The study reported that the three area SST index, formulated to capture ENSO modulated Rossby wave pattern, had the most influence in predicting Malawi rainfall, followed by air pressure over the East Indian Ocean and the stratospheric zonal wind anomaly (QBO). The study found skilful results with a 55% hindcast fit and two thirds of tercile categories correctly forecast in independent test.

This study therefore aims at developing simple regression models that can be used to predict Malawi rainfall. The

Figure 1: Map of Southern Africa showing location of Malawi and map of Malawi showing location of rain gauge stations

Figure 2: Mean monthly rainfall for Zone 1 (top) and Zone 2 (bottom)
study will exploit the relationship between Malawi rainfall and SST in the global oceans and use these as predictors in the models.

2 Study area
Malawi is located within tropical southeastern Africa (lat. 9–15°S, long. 32–36°E). Malawi has an area of 118,000 km² and is dominated on its eastern side by Lake Malawi, which is part of the Great African Rift Valley (Kumbuyo et al., 2014).

Malawi experiences a subtropical climate that is relatively dry and strongly seasonal, with an austral summer rainy season. A cool, dry season is experienced from May to August; hot, dry season from September to October; and rainy season from November to April, with annual rainfall averaging 725–2500 mm (www.metmalawi.com/climate). The climate of Malawi depends on the position of the Intertropical Convergence Zone (ITCZ), which marks the convergence of the northeasterly monsoon and southeasterly trade winds, and during the rainy season it oscillates over the country from north to south, often connecting with troughs in the Mozambique Channel (Torrance, 1972). Other meteorological systems that bring rain to Malawi include the northwest monsoon and tropical cyclones from the West Indian Ocean. These systems bring rain that falls throughout the country.

3 Materials and methods
Rainfall data were obtained from the Department of Climate Change and Meteorological Services for the period 1981–2011 from nine rainfall stations in Malawi (Figure 1) and the rainfall data were quality checked with respect to the World Meteorological Organization standards to ensure that they were stationary, consistent, and homogeneous. The nine rainfall stations were zoned into two zones on the basis of inter-station rainfall correlations and mean monthly rainfall for Zone 1 and Zone 2 were calculated (Figure 2) as described in Kumbuyo et al. (2014). Mean monthly rainfall patterns in Figure 2 show apparent difference between the two zones with zone 1 indicating bimodal pattern and zone 1 having a single peak.

SST data for the period 1975–2012 were obtained from the British Atmospheric Data Centre (BADC). We used the BADC HadISST 1.1 dataset found at www.badc.nerc.ac.uk/view/badc.nerc.ac.uk_ATOM_dataent_hadi. SST data for 1870 to the present. HadISST 1.1 contains gridded monthly mean SST, which are updated monthly for Zone 1 and Zone 2 were calculated (Figure 2) as described in Kumbuyo et al. (2014). Mean monthly rainfall patterns in Figure 2 show apparent difference between the two zones with zone 1 indicating bimodal pattern and zone 1 having a single peak.

We used the BADC HadISST 1.1 dataset found at www.badc.nerc.ac.uk/view/badc.nerc.ac.uk_ATOM_dataent_hadi to obtain SST data from 1870 to the present. SST data for Zone 1 and Zone 2 were calculated (Figure 2) as described in Kumbuyo et al. (2014). Mean monthly rainfall patterns in Figure 2 show apparent difference between the two zones with zone 1 indicating bimodal pattern and zone 1 having a single peak.

SST data for the period 1975–2012 were obtained from the British Atmospheric Data Centre (BADC). We used the BADC HadISST 1.1 dataset found at www.badc.nerc.ac.uk/view/badc.nerc.ac.uk_ATOM_dataent_hadi to obtain SST data from 1870 to the present. SST data for Zone 1 and Zone 2 were calculated (Figure 2) as described in Kumbuyo et al. (2014). Mean monthly rainfall patterns in Figure 2 show apparent difference between the two zones with zone 1 indicating bimodal pattern and zone 1 having a single peak.

Among other areas. In the southern Atlantic, significant positive correlations were observed east of Brazil (AO1, 20–25°S, 40–33°W). Other areas of significant correlation were; east and north east of Australia (PO4, 10°N–20°S, 145–155°E); along Philippines (PO5, 20–5°N, 145–155°E) and south east of Madagascar (IO3, 30–37°S, 55–67°E) as shown in Figure 3. For Zone 2, areas of significant correlation were observed around Australia (PO5, 5–10°S, 155–165°E, PO8, 35–40°S, 155–160°W, PO1, 10–35°S, 90–110°E); off the coast of Chile in South Pacific Ocean (PO2, 15–30°N, 80–120°W, PO3, 35–45°S, 75–85°W) and near Hawaii Islands in North Pacific Ocean (PO6, 26–28°N, 150–165°W) as shown in Figure 4. The correlation coefficients ranged from 0.36 to 0.47; these r values corresponded to type I error rates (p) of 5% and 1%, respectively.

Regions with significant correlation (r>0.36 significant at 5 % error) were extracted from the maps using National Oceanic and Atmospheric Administration extracting tool available at www.esrl.noaa.gov/psd/data/timeseries/. The regions are shown in Figure 2 and Figure 3. These were taken as the predictors. Predictors had to meet the following requirements;

a) A good relationship with Malawi rainfall (correlation of r>0.36 significant at 5 % error)

b) A reasonable lead time (months)

Table 1 presents a summary of the final predictors used in development of Zones 1 and 2 models and their spatial location.

3.1 Model description
Multiple linear regression models are generally used in developing long range forecasting models. Correlations between SST and rainfall were performed to identify predictors to include in forecast models. The step-wise multiple regression method was used within the R software package (available from http://cran.r-project.org). Malawi rainfall was the target objective and the multiple linear regression model was denoted by:

\[
y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon \quad (1)
\]

where

- y is the response variable
- \(\beta_0\) is the intercept
- \(\beta_i\) is slope coefficient for the first explanatory variable \(x_1\)
- \(\beta_2\) is slope coefficient for the second explanatory variable \(x_2\)
- \(\beta_k\) is slope coefficient for \(k^{th}\) explanatory variable \(x_k\)
- \(\varepsilon\) is the error

The selection procedure used in step-wise multiple regression of the model was the forward and backwards procedure. It is used to control the selection of variables into the model. Variables were examined to be entered or removed with the aim of obtaining a model with a high degree of fit.
Figure 3: Correlation map for Zone 1 showing regions where predictors were extracted

Figure 4: Correlation map for Zone 2 showing regions where predictors were extracted

Table 1: Details of final predictors used for Zone 1 and Zone 2 summer rainfall prediction

| Area  | Predictors | Spatial domain       | Descriptive location               |
|-------|------------|----------------------|------------------------------------|
| Zone 1| AO1        | 20–25°S, 40–33°W     | East of Brazil                     |
|       | AO2        | 17–10°N, 40–25°W     | Off Cape Verde Island              |
|       | AO3        | 15–10°N, 83–65°W     | Along Venezuela and Colombia       |
|       | IO2        | 30–37°S, 55–67°E     | South east of Madagascar           |
|       | PO2        | 3°N–3°S, 45–125°W    | Along the equator in Pacific Ocean |
|       | PO3        | 25–35°S, 160–175°E   | East of Australia                  |
|       | PO4        | 10°N–20°S, 145–155°E | North East of Australia            |
|       | PO5        | 20–5°N, 145–155°E    | Off Philippines                    |
|       | PO6        | 35–40°S, 155–160°W   | South east Australia               |

| Zone 2| PO2        | 15–30°S, 120–80°W    | Off Chile coast                    |
|       | PO4        | 45–37°N, 160–170°E   | Off Japan coast                    |
|       | PO5        | 5–10°S, 155–165°E    | North east Australia               |
|       | PO6        | 35–40°S, 155–160°W   | South east Australia               |
### 3.2 Model validation

The model was validated using the correlation coefficient \((r)\), bias \((BIAS)\) and \(RMSE\) between observed and predicted summer rainfall. They are calculated as follows:

\[
r = \frac{\sum (Y - \bar{Y})(Y' - \bar{Y})}{\sqrt{\sum (Y - \bar{Y})^2 \sum (Y' - \bar{Y})^2}}
\]

\[
BIAS = \frac{\sum (Y - Y')}{n}
\]

\[
RMSE = \sqrt{\frac{\sum (Y' - Y)^2}{n}}
\]

where; \(\bar{Y}\) and \(\bar{Y}'\) are the sample averages of the \(Y\) and \(Y'\), respectively, and \(n\) is the number of data.

### 4 Results and discussion

Table 2 shows the model validation statistics for models that were developed for Zone 1 and Zone 2.

| No. | Model value \(r\) | Model value \(RMSE\) | Model value \(BIAS\) |
|-----|------------------|----------------------|----------------------|
| Zone 1 | 0.81* | 0.01 | -0.03 |
| Zone 2 | 0.54* | 0.003 | 0.014 |

*significant at 99.99% level of significance

For Zone 1, significant correlations between summer rainfall and SST were obtained at 7 months lag time. The regions of significant correlation were in Atlantic, Indian and Pacific Oceans. The areas of significant correlation were used as the predictors of the model. The predictors were modeled in \(R\) software using the regression function and selection of predictors was based on their significant correlation with summer rainfall, \(t\) value and residual standard error.

The regression equation for Malawi rainfall, Zone 1 model is given below:

\[
MSR_{Z1} = -1684.7 - 21.2AO_1 - 30.1AO_2 - 59.8AO_3 + 21.6PO_2 + 68.9PO_3 - 28.3PO_4 + 102.3PO_5 + 41.3IO_2
\]

where; MSR\(_{Z1}\) is Malawi rainfall for Zone 1 and AO, IO and PO are predictors as described in Figure 2.

Figure 5 shows the results of predicted rainfall against the observed rainfall for Zone 1. There is a general agreement between the observed and predicted values except for the period between 1982-83; 1992 and 2001 that were as a result of El Nino. The model overestimated values in 1981, 1991, 1994, 2003 and 2004. The model correlation coefficient \((r)\) was 0.54, significant at 99.99% level of significance. The model \(RMSE\) and \(BIAS\) were 0.003 and 0.014 respectively.

The observed and predicted rainfall time series for Zone 2 are shown in Figure 6.

For Zone 2, significant correlations between summer rainfall and SST were obtained at 11 months lag time. The regions of significant correlation were in Indian and Pacific Oceans.

The areas of significant correlation were extracted as indicated in Figure 4 and these were the predictors of the model. The final regression equation of the model for Zone 2 is given below:

\[
MR_{Z2} = 38.9PO_5 - 8.6PO_2 - 14.2PO_4 - 8.7PO_6 - 501.9
\]

where; MR\(_{Z2}\) is Malawi Rainfall for Zone 2 and IO and PO are predictors as described in Figure 3.

The observed and predicted rainfall time series for Zone 2 is shown in Figure 6.

It can be seen that the model has successful predictions and some series failures. The model successfully predicted dry spells that were experienced in Malawi during 1982-83; 1992 and 2001 that were as a result of El Nino. The model overestimated values in 1981, 1991, 1994, 2003 and 2004. The model correlation coefficient \((r)\) was 0.54, significant at 99.99% level of significance. The model \(RMSE\) and \(BIAS\) were 0.003 and 0.014 respectively.

The results of the Zone 1 model are in agreement with what different researchers have found about the relationship between Southern Africa rainfall and SST. For example Nicholson (1987) reported that the ENSO signal in African rainfall variability is a manifestation of the influence of the...
ENSO on SST in the Atlantic and Indian Oceans which in turn, influences rainfall. Reason (2001) noted that when the SST is warm(cold) to the south of Madagascar and cool (warm) off Western Australia, increased (decreased) summer rainfall occur over large areas of Southern Africa.

Williams et al. (2008) found that both decreasing SST in the central South Atlantic and increasing SST off the coast of southwestern Africa lead to a demonstrable increase in daily rainfall and rainfall extremes over Southern Africa.

It is evident that SST in Indian and Pacific Oceans are playing an important role in summer rainfall for Zone 2. Xie and Arkin (1996) reported that when warm SST anomalies occur to the south of Madagascar and cool anomalies occur off Western Australia, summer rains over southeastern Africa are enhanced. Furthermore, these authors suggest that the impact on rainfall is related to a weakening of the maritime ITCZ over the Indian Ocean and enhanced moisture transport towards southeastern Africa by stronger Southwesterlies. Ratnam et al. (2014) in their study of the remote effects of El Nino and Modoki events on austral summer precipitation of SA have reported that the differences in the spatial distribution of precipitation over southern Africa are seen to be related to the SST anomalies of the equatorial Pacific through atmospheric teleconnections.

5 Conclusions
Prediction of Malawi rainfall was done by first analyzing the linkage between Malawi rainfall and SST in order to determine predictors for the models. Significant correlations between rainfall and SST in the Atlantic, Indian and Pacific oceans were found. The correlations are significant in Indian Ocean; Pacific Ocean and between the equator and 20°N latitude in the Atlantic Ocean. These results are in agreement with reported previous studies. To predict Malawi rainfall, these SST time series were applied to a simple multiple regression model. The fit of the predicted and observed rainfall for both Zones 1 and 2 were satisfactory.

There is need increase the sample size of the rain gauge stations to in order to get improved results. The study also focused only on relationship between SST and Malawi rainfall. Other factors like sea level pressure and outgoing longwave radiation if taken into consideration can greatly improve the results of the model.

Acknowledgements
The authors are grateful to Tottori University for financial assistance to carry out this research. The authors are also grateful to the National Oceanic and Atmospheric Administration for providing SST data. The British Atmospheric Data Centre is also acknowledged for providing some of the data used in the study.

References
[1] Hirst, A.C., and Hastenrath, S. (1983): Atmosphere-Ocean mechanisms of climate anomalies in the Angola–tropical Atlantic sector, J. Phys. Oceanogr., 13, pp.1146–1157.
[2] Iyengar, R.N., and Raghukanth, S.T.G. (2004): Intrinsic mode functions and a strategy for forecasting Indian monsoon rainfall. Met. and Atmos. Physics. 90, pp.17-36
[3] Jury, M.R., and Mwafurirwa, N.D., (2002): Climate variability in Malawi, Part I: Dry summers, statistical associations and predictability, Int. J. Climatol., 22, pp.1289–1302.
[4] Jury, M.R., and Pathack, B. (1991): A study of climate and weather variability over the tropical southwest Indian Ocean, Meteorology and Atmospheric Physics. 47, pp.37-48.
[5] Jury, M.R., Pathack, B., and Legley, D.M. (1991): Structure and variability of surface atmospheric circulation anomalies over the tropical South West Indian Ocean in the austral summer. South Africa Journal of Marine Science. 11, pp.1-14.
[6] Kumbuyo, C.P., Yasuda, H., Kitamura, Y., and Shimizu, K. (2014): Fluctuation of rainfall time series in Malawi: An analysis of selected areas, Geofizika Journal, 31, pp.13–28, DOI: 10.13233. gzf. 2014.31.3.
[7] Mason, S.J. (1992): Sea surface temperature and South African rainfall variability. Ph.D. Thesis. University of the Witwatersrand. South Africa.
[8] Mwafurirwa, N.D. (1999): Climate variability and predictability in tropical southern Africa with focus on dry spells over Malawi. MSc. Thesis. University of Zululand. South Africa.
[9] Nicholson, S.E. (1987): Rainfall variability in southern and equatorial Africa, its relation to Atlantic sea surface temperatures and the southern oscillation, Amer. Meteor. Soc., pp.657–676.
[10] Nicholson, S.E., and Entekhabi, D. (1987): The quasi-periodic behaviour of rainfall variability in Africa and its relationship to the Southern Oscillation, Archiv fur meteorologie. geophysik und bioklimatologie. Ser. A., 34, pp.311–348.
[11] Nicholson, S.E. (1997): An analysis of the ENSO signal in the tropical Atlantic and western Indian oceans, Int. J. Climatol., 17, pp.345–375.
[12] Nyenzi, B.S. (1988): Mechanisms of East African rainfall variability. PhD Thesis, Florida State University, USA.
[13] Ratnam, J.V., Behera, S.K., Masumoto, Y., and Yamagata, T. (2014): Remote effects of El Nino and Modoki events on the austral summer precipitation of southern Africa. J. Climate, e-View, doi://http://dx.doi.org/10.1175/JCLI-D-13-00431.1
[14] Reason, C.J.C. (2001): Subtropical Indian Ocean SST dipole events and southern African rainfall, Geophys. Res. Lett., 28 (11), pp.2225–2227.
[15] Sahai, A.K., Grimm, A.M., Satyan, V., and Pant, G.B. (2003): Long-lead prediction of Indian summer monsoon rainfall from global SST evolution, Climate Dynamics 20, pp.855-863.
[16] Torrance, J.D. (1972): Malawi, Rhodesia and Zambia, In Climates of Africa. Ed. Griffiths, J.F., World Survey of Climatology, Vol. 10, Elsevier, Amsterdam, pp409-460
[17] Wang, B., Ding, Q., Fu, X., Kang, I., Jin, K., Shukla, J., and Doblas-Reyes, F. (2005): Fundamental challenge in simulation and prediction of summer monsoon rainfall, Geophys. Res. Lett., 32 (15), L15711, doi. 10.1029/2005GL 022734.
[18] Williams, C.J.R., Kniveton, D.R., and Layberry, R. (2008): Influence of South Atlantic Sea Surface Temperatures on Rainfall Variability and Extremes over Southern Africa, J. Climate, 21, pp.6498–6520.
[19] World Bank (1990): ‘Malawi food security report.’ Report No. 8151 MAI. Washington, D.C., USA.
[20] Xie, P., and Arkin, P.A. (1996): Analysis of global monthly precipitation using gauge observations, satellite estimates and number model predictions, J. Climate, 9, pp.840–858

Discussions open until December 31, 2015