

A Statistical Model for Seasonal Rainfall Forecasting over the Highlands of Eritrea

Mehari Tesfazgi Mebrhatu¹, M. Tsubo² and Sue Walker¹

¹University of the Free State, Department of Soil, Crop and Climate Sciences, P. O. Box 339, Bloemfontein 9300, South Africa.

Email: MebrhaMT@Sci.Uovs.ac.za; walkers.sci@mail.uovs.ac.za

²The University of Queensland, School of land and food Sciences, Brisbane, Qld 4072, Australia. E-mail: m.tsubo@uq.edu.au

Abstract

A deterministic model was developed to investigate how global rainfall predictors relate to the two main rainy months in the highlands of Eritrea. The main aim of looking at these relationships was to develop a simple statistical model for forecasting rainfall amount. Farmers could make better management decisions if they had a better assessment of the forthcoming season. In a preliminary step, in order to identify the most influential rainfall predictor, a correlation matrix and stepwise regression of 10 predictors with different lags were analysed. The influence of the southern Indian Ocean Sea Surface Temperature was identified as the most influential predictor for the highlands of Eritrea. A model was developed and validated giving a promising result.

Media summary

The SST of the southern Indian Ocean was identified as a major influence on summer rainfall in the highlands of Eritrea.

Key words

Highlands of Eritrea, Indian Ocean SST, Jack-knife cross-validation

Introduction

Given an improved understanding of Eritrean climate, numerous benefits could be expected in many related activities: better management of agriculture and water resources stemming from more reliable seasonal predictions. Seasonal forecasting has good prospects for early warning of low rainfall totals to help prepare for, and mitigate the effect of, famine, which so often results in Eritrea. The need for providing accurate forecasts for a coming rainfall season is becoming more and more necessary.

Most parts of Eritrea (latitude 12°40' - 18°02' N and longitudes 36°30' - 43°20' E) receive rainfall from the South-western Monsoon winds during the spring and summer months (April to October) (FAO, 1994). The rainfall is mainly convective and thus it is localised and its spatial and temporal distribution as well as the intensity of rainfall is highly variable. "Short rains" fall in April/May and the "main rains" in July and August. Interannual variability of rainfall in East Africa results from complex interactions of forced and free atmospheric variations (Ogallo, 1988). There have been several recent studies examining connections between observed rainfall and a number of large-scale climate signals (Clark *et al.*, 2003). Studies have also looked for prediction links based on correlation with raw station data (Nicholls, 1981). Promising seasonal forecast skill for the Oct-Nov-Dec "short" rains using multiple regression techniques have been found for East Africa and predictors based on eigenvectors of global sea surface temperatures (Mutai, *et al.*, 1998).

The objective of this research is to develop a simple statistical model forecast for the peak rainy months (July-August) during which 65% of the total annual rainfall in the highlands of Eritrea occur.

Data

The data used in this assessment are the long-term monthly rainfall amounts for two representative stations (Asmara and Mendefera) of the highlands of Eritrea and 10 predictors from the Pacific, Atlantic and Indian Oceans from year 1950 to 2000. Thus: -Asmara station (the central zone): Latitude (N)-15°21'00"; Longitude (E)-33°55'30"; Altitude-2150m. And Mendefera station (the southern zone): Latitude (N)-14°53'30"; Longitude (E)-38°49'10"; Altitude-2060m. The rainfall data were quality checked by homogeneity testing using cumulative partial sum technique (Mebrhatu, 2003). Emphasis was placed on areas with significant water resources and rain-fed agricultural production.

The ten predictors selected and listed below have been identified by many authors to be related to large-scale eastern African climate. These are: Niño1+2 (0-10°S, 90-80°W); Niño3 (5°N-5°S, 150-90°W); Niño4 (5°N-5°S, 160°E-150°W); Niño3.4 (5°N- 5°S, 170-120°W); North Atlantic SST (5-20°N, 60-30°W); South Atlantic SST (0-20°S, 30°W-10°E); Global Tropics SST (10°S-10°N, 0-360); South Indian SST (0-15°S, 45-60°E); Southern Oscillation Index (SOI) and Pacific Decadal Oscillation (PDO). These parameters were tested to determine which have a significant influence on Eritrea rainfall and can be used in a prediction model. Source of data: - Cpc-NOAA and South African Weather Service.

Model development

Correlation matrix

A correlation matrix is useful in showing how strong each independent variable is related to the dependent variable at different lag times. Thus, a matrix was set up for 11 candidates including rainfall itself as one of the candidates to show all possible correlation coefficients between all variables. For each candidate monthly lag (1-12) was calculated, such that lag 1 is a 1-month difference between the predicted rainfall and the candidates. From the resultant correlation matrix it was found that lag11, lag12 between the previous year's rainfall and the coming rainy season are significant, which can be explained by the autocorrelation property of the long-term rainfall pattern.

Stepwise and Multiple regression

In order to identify the most influential predictor for the peak rainy months (Jul-Aug, JA) which is the main focus, all the predictor data were categorised into 6 sets consisting of 2 month sums of Jan-Feb (JF), Mar-Apr (MA), May-Jun (MJ), Jul-Aug (JA), Sep-Oct (SO) and Nov-Dec (ND). A forward stepwise regression procedure was employed for further analysis to select the most significant candidate from the 60 variables. These 60 variables represent from lag 1 to 6 for each of the 10 predictors for the 6 categories of the 2-month sum. Stepwise regression is a technique for choosing the variables to be included in a multiple regression model. It is important to note that when one deals with multiple regression one has to be aware of multi-collinearity. Multi-collinearity was checked for each step (Table 1).

From multiple regression, it was found that (lag 1 to 6) of the South Indian Ocean SST, (lag 6) of South Atlantic Ocean ($R^2 = 0.63$) and (lag 3) of North Atlantic Ocean ($R^2 = 0.61$) have an influence on the amount of rain received during the peak rainfall months (JA). R^2 value for South Indian Ocean SST lags 1 to 6 is indicated in Table 1. For the sake of simplicity it was decided to focus on the highly influential region that was identified as the South Indian Ocean SST, but ignoring lag 1 from an operational point of view, being hardly useful as it is only available at the beginning of July. Moreover, as was mentioned before the rainfall amount of the previous year has an influence on the current rainy season. Thus the model has been developed using the lags of the South Indian Ocean SSTs and the lags of rainfall amount data. The difference in rainfall between JA (July-August) and ND (November-December) (lag 4) can be a function of the South Indian Ocean SSTs as follows:

$$R_{i,j} - R_{i,j-4} = f(\text{SST}_{\text{S. Indian}}) + \text{Error}; \text{ If } R_{i,j} < 0, R_{i,j} = 0$$

where $R_{i,j-4}$ is total rainfall amount of the previous year for ND, and $R_{i,j}$ is current rainfall amount for JA (mm).

From multiple regression estimates, the standardised formula for the peak rainy months (JA) is given as:

- Asmara station (the central zone)
 $R_{i,j} = R_{i,j-4} + (.03(SIn_{lag2})+.25(SIn_{lag3})-.08(SIn_{lag4})+.64(SIn_{lag5})-.86(SIn_{lag6})) R^2 = 0.80$
- Mendefera station (the southern zone)
 $R_{i,j} = R_{i,j-4} + (.01(SIn_{lag2})+.32(SIn_{lag3})-.05(SIn_{lag4})+.59(SIn_{lag5})-.85(SIn_{lag6})) R^2 = 0.84$

Where SIn = South Indian Ocean SST (?C).

Table 1. Multiple regression statistics for rainfall and different lags of South Indian SST.

months	variables	Regression coefficient	S.E. (%)	t-value	Probability level	Power (5%)	R ²	Multi collinearity	
								Variance inflation	tolerance
July- August (Asmara)	Intercept	129.232	540.26	0.239	0.811	0.056	0.79		
	Indian lag2	4.049	4.791	0.845	0.398	0.134		1.904	0.525
	Indian lag3	33.461	4.996	6.697	0.000	0.999		2.081	0.481
	Indian lag4	-10.50	5.947	-1.766	0.078	0.421		2.923	0.342
	Indian lag5	85.348	5.002	17.064	0.000	1.000		2.063	0.485
	Indian lag6	-114.67	4.790	-23.939	0.000	1.000		1.894	0.528
July- August (Mendefera)	Intercept	-141.41	540.26	-0.285	0.776	0.059	0.84		
	Indian lag2	1.421	4.791	0.323	0.747	0.062		1.904	0.524
	Indian lag3	43.849	4.996	9.555	0.000	1.000		2.081	0.481
	Indian lag4	-6.95	5.947	-1.273	0.204	0.245		2.924	0.344

Indian lag5	80.546	5.002	17.533	0.000	1.000	2.063	0.485
Indian lag6	-116.31	4.790	-26.436	0.000	1.000	1.884	0.537

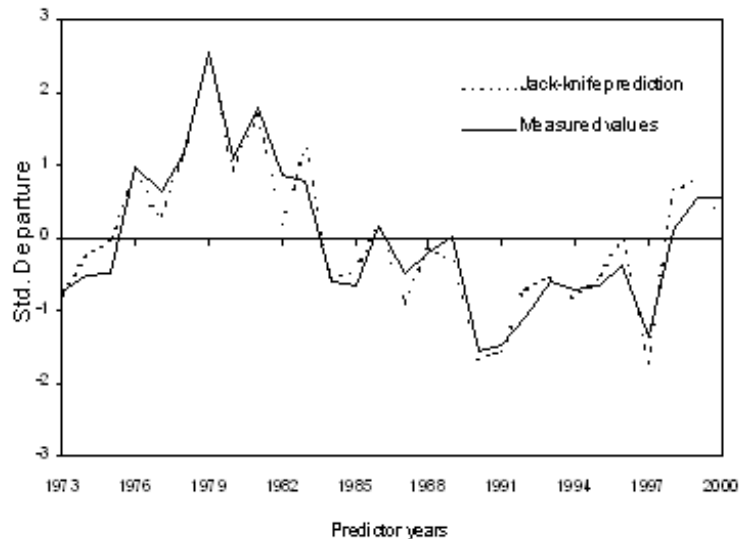
Model Validation

Model validation is performed with the objective of assessing the performance of the model and to uncover any possible lack of fit. Two methods were used in order to determine the model accuracy. These are the Jack-knife method of cross-validation (Jury, *et al.*, 1997) and hit rate. Trial forecasts (hindcasts) were made using the Jack-knife method. Jack-knife forecasts were made for every year in the data set period using equations calculated using the majority of the remaining years in the data period. The forecast year is always excluded from the regression equation. The process is repeated by removing the next year and so on. Thus the coefficients of the predictors in the equations change from year to year.

Jack-knife multiple regression forecasts are plotted against measured rainfall amount in Figure 1. The D-index (Willmott, 1981) of 0.91 and 0.79 for the Asmara station and Mendefera station is very high for a climate prediction. The anomaly departures calculated by subtracting the mean and dividing by the standard deviation.

Hit rate also made for validation. The data were separated into two continuous discrete segments of time. Half of the data (1950-75) was used for training and another half (1976-2000) of the data for validation of the model. For the hit rate a probability <33% is taken as below normal; 33% - 66% probability as near normal and the probability > 66% used as above normal. From the cumulative probability distribution the 33% and 66% were calculated. That is 261.4mm & 381.7mm for the central zone and 297mm & 347.5mm for southern zone for 33% and 66% respectively. The hit rate for the model is very good for both stations that is 80% for central and 70% for southern zone.

(a)



(b)

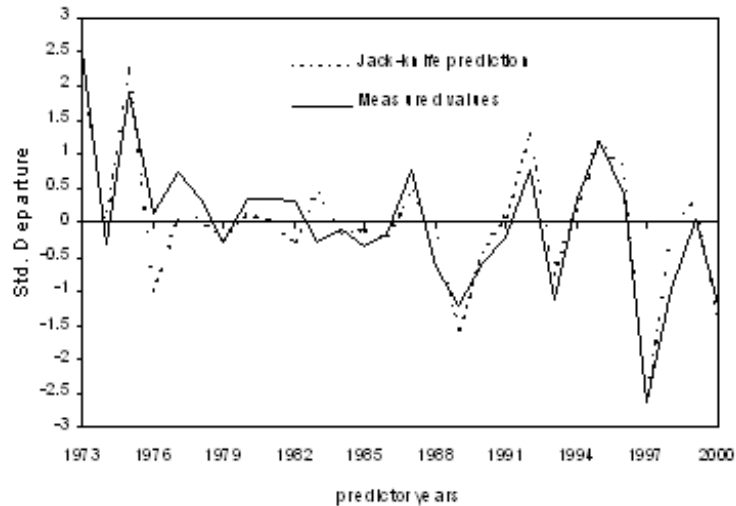


Figure 1 Jack-knife skill tests for South Indian SSTs (lag_{2-6}), showing rainfall anomalies versus Jack-Knife predictions for (a) the Central zone [$R^2 = 0.89$; $D = 0.91$] and (b) the southern zone [$R^2 = 0.81$; $D = 0.79$] for peak rainy months of July -August

Conclusion and recommendations

The Indian Ocean SST was identified to have a large influence on the rainfall of the highlands of Eritrea. A deterministic model was developed to predict rainfall amount for the peak rainfall months of July-August. The multiple regression model uses rainfall amount from the previous November-December and SST from the South Indian Ocean during the previous year (five values of South Indian Ocean SST for MJ, MA, JF, ND and SO). The model could provide a measure that can be used in developing a seasonal forecast system, which would allow farmers to make better management decisions. An inspection of the results of model validation using different statistics clearly indicated that the model is reproducing and describing the pattern of the rainfall for the site of interest.

A further study should be done for a more specific seasonal forecast that relates to crop yields, rather than just rainfall prediction, which is very critical for drought prone countries like Eritrea. Improved prediction of expected rainfall behaviour in the approaching crop season enables improved decision making at the field level and outlooks need to be developed for crop specific information.

References

- Clark CO, Webster PJ and Cole JE (2003). Interdecadal variability of the relationship between the Indian Ocean zonal model and east African coastal rainfall anomalies. *Journal of Climate* 16: 548-553.
- FAO (1994). *Agricultural Sector Review and Project Identification*, Food and Agricultural Organization of the United Nations, Rome.
- Jury, M. R., Mulenga, H. M., Mason, S. J. & Brandao, A. (1997). *Development of an Objective Statistical System to Forecast Summer Rainfall over Southern Africa*. WRC Report No 672/1/97. Water Research Commission, Pretoria, 45 pp.
- Mutai CC, Ward MN and Colman AW (1998). Towards the prediction to the East Africa short rains based on the Sea Surface Temperature - Atmospheric Coupling. *International Journal of Climatology* 18: 975-997.

Nicholls N (1981). Air-sea interaction and the possibility of long-range weather prediction in the Indonesian archipelago. *Monthly Weather Review* 109: 2435-2443.

Ogallo LA (1988). Relationship between seasonal rainfall in East Africa and Southern Oscillation. *Journal of Climate* 8: 34-43

Willmott CJ (1981). On the validation of models. *Physical Geography* 2: 184-194.