Improved Seasonal Prediction of Rainfall over East Africa for Food Security
Forecasting: Statistical Downscaling of CFSv2 and GFDL-FLOR

O. Kipkogei\textsuperscript{1}, A.M. Mwanthi\textsuperscript{1}, J.B. Mwesigwa\textsuperscript{1,2}, Z. K.K. Atheru\textsuperscript{1}, and G. Artan\textsuperscript{1}

\textsuperscript{1} IGAD Climate Prediction and Applications Centre, P. O. Box 10304 – 00100, Nairobi, Kenya;
\textsuperscript{2}Institute for Climate Change and Adaptation, University of Nairobi, P. O. Box 30197-00100, Nairobi, Kenya

**Corresponding Author:**
Oliver Kipkogei
Climate Forecast Downscaling Assistant
IGAD Climate Prediction and Application Centre (ICPAC)

Email: oliver@icpac.net

Mobile: +254726018688
ABSTRACT

The evolution of 2015 El Nino event yielded excessive anxiety about the magnitudes of rainfall expected over the East Africa. In this paper, we evaluate the utility of downscaled forecasts from two GCM models. The Climate Predictability Tool (CPT) was employed to statistically downscale GCM rainfall forecasts over Kenya and Tanzania for the month of October 2015, which usually forms the onset period of the October-December season.

A blend of station and satellite estimates called Climate Hazards Infrared Precipitation with Station (CHIRPS) at 5x5 Km resolution was used as benchmark data to downscale October 2015 model outputs from CFSv2 and GFDL-FLOR over Kenya and Tanzania. Canonical Correlation Analysis (CCA) technique was used to make model correction by generating Empirical Orthogonal Functions (EOFs) on the X (predictor) data then on the Y (observed) data for the training period 1982-2011 for 5 modes. The 5 modes were then used to make cross validated forecasts.

Various skill scores were used to validate the downscaled rainfall forecasts for the month of October 2015. This study revealed that statistical downscaling approach using the CPT was a worthwhile effort since it provided improved forecast skills in terms of Relative Operating Characteristics (ROC), and Pearson Correlation for precipitation forecasts for the month of October over Kenya and Tanzania. CPT is therefore a robust tool that can be applied for skillful downscaling of weather and climate forecasts over East Africa to generate more reliable products applicable for agricultural planning and decision making.
1. Introduction

East African economy is largely dependent on rain-fed agriculture, which is highly vulnerable to the negative impacts of climate variability. Weather and climate shocks frequently impact negatively on the livelihoods of people in the region. For instance, nearly a third of African population dependent on rain-fed food production face chronic food insecurity (Hail, 2005). The main driver of this perennial crisis is the close linkage between weather and climate variability with agricultural production and food security. Improved weather and climate forecasts can provide more reliable climate information that agricultural decision makers can rely on to boost agricultural productivity and food security across the region. A reliable forecasting system is thus of paramount importance for the East African region.

This paper promotes the need for capacity building for the East Africa National Meteorological and Hydrological Services (NMHS) to provide timely and high resolution forecasts that can be used for decision making in vital socio-economic sectors. This will have a strong effect on the sustainability and productivity of the highly susceptible regional economies (Washington and Downing, 1999).

Over the years, there has been an increased need for high resolution climate forecasts for target users in agriculture, hydrology, disaster management and health among others at sufficient lead times. To generate high resolution local climate anomalies, downscaling techniques are applied. These can either be statistical or dynamical. In the dynamical technique, regional nests are used within global circulation models (GCMs) providing boundary conditions to output climate forecasts of higher resolution (Castro et al., 2006). However, this technique is quite costly in terms of computing resources and also requires highly skilled manpower. It’s worthy to note that although dynamical downscaling works
under better representation of land features and atmospheric patterns, the model configurations may introduce errors (Mironov and Raschendorfer, 2001). In statistical downscaling, there is a target set of observations and the statistical model identifies statistical relationships between the course resolution GCM and the observational data (Wilby et al., 1998). This technique is quite effective as it demands minimal computing facilities.

Many studies have been undertaken to improve climate predictability over the Greater Horn of Africa region. A recent contribution (Rowell et al., 2015) noted that the rains in the region exhibited a downward trend which was in contrast with major global climate models that projected an increasing trend for the coming decades, a phenomenon he termed as the East African climate paradox. Kipkogei et al. (2016) combined global output products from four centres using the multimodel superensemble technique and noted that it greatly reduced the forecast errors from day 1 to day 10. Nicholson (2014) examined the predictability of seasonal forecasts over the Greater Horn of Africa using multiple linear regression and cross validation. She showed that surface variables like sea surface temperatures provided lower forecast skills as compared to the atmospheric variables. Rowel et al (2015) further noted that climate change projection has a high degree of uncertainty over Africa with global models exhibiting a broad range of disproportion in the magnitude of their temperature and rainfall change.

Rainfall predictability in the Eastern Africa region is usually derived from several oceanic and atmospheric features related with the seasonal rainfall. Global ocean Sea Surface Temperature (SST) patterns including El Nino Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD) and SST gradients have been utilized extensively for seasonal rainfall prediction (Mutai et al., 1998; Saji et al., 1999; Indeje et al., 2000; Marchant et al., 2007; Gitau et al., 2015). The rainfall seasonality in the region is also modulated by atmospheric circulations such as the pattern of the Inter-tropical Convergence Zone (Okoola, 1999).
systems include tropical cyclones over the western Indian Ocean, the East Africa low level jet and the off-equatorial high pressure cells (Gitau et al., 2015). It is also worthy to note that the ocean-atmosphere coupling is an important interaction that generate seasonal weather patterns.

Over the Greater Horn of Africa (GHA) which is a consortium comprising 11 countries including East Africa, the Inter-Governmental Authority on Development (IGAD) Climate Prediction and Applications Centre (ICPAC) has been spearheading the generation of seasonal forecasts with a lead time of one month to three months over eleven countries. The GHA Climate Outlook Forums (GHACOFs) provides a regional consensus of the expected rainfall anomalies from which national meteorological agencies get a benchmark in generation of downscaled forecasts.

This paper is part of an ongoing work on generation of improved downscaled monthly and seasonal forecasts to promote food security initiatives at both national and farming community levels over East Africa. It is expected that the downscaling technique presented in this paper will be adopted and utilized for generation of high resolution seasonal forecasts and derived agro-meteorological variables over the entire Eastern Africa region.

2.0 Area of Study

Kenya and Tanzania are part of the countries that make up the East African community. Kenya lies within -5° to 5° latitude and 37° to 42° longitude whereas Tanzania lies within -12° to 0° latitude and 28° to 41° longitude.

Rainfall patterns in the region exhibits a bimodal cycle with March, April and May being the ‘long rains’ season while the October, November and December is the ‘short rains’ season. Systems that significantly control rainfall seasonality include the Inter-Tropical
Convergence Zone (Nicholson 2003; Donohoe et al., 2013), Indian Ocean Dipole (Owiti et al., 2008), El Nino Southern Oscillation and associated tele-connections (Terray and Dominiak 2005; Shawn et al., 2007; Kumar et al., 1999), as well as secondary effects resulting from tropical cyclone activities (Parker and Jury 1999), among other factors.

3.0 Data

The data used in this study included rainfall hind casts and forecasts from two global models namely National Centre for Environmental Prediction (NCEP) Coupled Forecast System version 2 (CFSv2), Geophysical Fluid Dynamics Laboratory - Forecast-Oriented Low Ocean Resolution (GFDL- FLOR) and a high resolution (5km by 5km) Climate Hazards Infra Red Precipitation with Station data (CHIRPS). CHIRPS is a blend of actual station data with satellite estimates from Climate Hazards Infra Red Precipitation (CHIRPs) (Funk et al., 2015; Funk and Verdin 2010).

NCEP CFSv2 is a coupled global model with a resolution of 0.937° (T126) with 24 ensemble members (Saha et al., 2006; 2010; and 2014). This model became operational in April 2011 and continues to provide forecasts up to date. The scientists at NCEP used the Climate Forecast System Reanalysis (CFSR) technique (running the model retrospectively) to obtain the best reforecasts (past forecasts) estimates.

GFDL-FLOR (Vecchi et al., 2014) is an improved version of the Climate Model version 2.5 (CM2.5) model (Delworth et al., 2012) and Climate Model version 2.1 (CM2.1) model (Delworth et al., 2006). It has been used extensively to simulate precipitation and temperature over land (Jia et al., 2015) as well as drought patterns (Delworth et al., 2015). It is an important contributor to the North American Multi-Model Ensemble (NMME) for seasonal forecasting (Kirtman et al., 2014). Climate Model version 2.6 (CM2.6) and CM2.5 (Delworth 2012) each has an atmospheric resolution of 0.45° and 0.1° and 0.25° Oceanic resolution respectively whereas the CM2.1 (Delworth 2006) has an Oceanic resolution of 1°.
and an atmospheric resolution of 1.8°. GFDL-FLOR on the other hand has a high land and atmospheric resolution of 0.45° as that in CM2.5 but with a coarser ocean resolution (1°) as that of CM2.5.

The CHIRPS data was used as benchmark analysis in this study. The methodology for blending satellite estimates and actual station data is explained in detail in Funk and Verdin (2010).

4.0 Methodology

In this study, Climate Predictability Tool (CPT) software developed by the International Research Institute was used to statistically downscale the GCM rainfall forecasts. With statistical downscaling there are a target set of observations (CHIRPS) and the statistical model identifies statistical relationships between the course resolution GCM and the observational data.

Canonical Correlation Analysis (CCA) in CPT has the ability to detect and correct bias in amplitude, mean and shape of anomaly pattern when trained on hindcast data and corresponding observations (Mason and Mimmack 2002). For CCA to make a model correction, it starts with the generation of the Principal Component Analysis (PCA) or Empirical Orthogonal Functions (EOF) on the X (predictor) data then on the Y (observed) data for some few modes. CPT uses EOF modes to make cross validated forecasts, then calculates a goodness index summarizing how good the forecasts are (Mason and Chidzambwa 2009). Thereafter, the CCA is done on the amplitude of the PCA modes of the X and Y data to detect the linear rules. The linear rules are then used to make real time forecasts.

EOFs help in identifying the preferred patterns within many variables. EOF clusters points or grids that are strongly related resulting to EOF modes that are used for approximation (Lorenz, 1956). Each value is assigned a weight which can either be positive
or negative with the weights showing a coherent pattern in the spatial domain called EOF loading pattern (eigenvector), (Weare and Nasstrom 1982). The sum of the product of values at each grid point and their corresponding loading weights gives the temporal score (amplitude) for that time. The first EOF mode explains the most variance (eigenvalues) with the second and subsequent ones explaining the remaining variance. Two to six EOF modes have been established to explain the total variability in a data set with the rest just working on the noise (Kim et al., 2003; Kim and Wu 1999).

CPT uses the World Meteorological Organization (WMO) standard forecast verification scores. Among them is the Relative Operating Characteristics (ROC) score which is recognized as an equivalent to the Hanssen-Kuipers Skill score (KSS) when applied to the deterministic forecasts (Hanssen and Kuipers 1965, Swets, 1973; Mason, 1982; Harvey, et al., 1992; Mason and Graham, 1999). ROC is a measure of the quality of the forecasts that relates the hit rates and the corresponding false alarm rates. A ROC score with larger hit rate and smaller false alarm i.e., with the area under the curve of greater than 0.5 shows a skilful forecast whereas one with more false alarms and less hit rate indicates a forecast with worse skill (Mason and Graham 2002; Kharin and Zwiers 2003).

5.0 Results and Discussion

In this study, downscaled forecasts for the month of October 2015 relevant for application in agricultural strategic planning were produced for Kenya and Tanzania. Canonical Correlation Analysis (CCA) technique was used to make model correction by generating Empirical Orthogonal Functions (EOFs) on the X (predictor) data then on the Y (observed) data for the training period 1982-2011 for 5 modes. The 5 modes were then used to make cross validated forecasts.

Figure 1 and 2 show downscaled CFSv2 and GFDL-FLOR model forecasts for October 2015 over Kenya and Tanzania respectively, and their corresponding deviations.
Kenya and the northern part of Tanzania basically lie within the equatorial belt and is associated with a wet signal during the October, November and December season. For the month of October 2015, the rainfall forecast for both the CFSv2 and GFDL-FLOR over Kenya (Figure 1 b, c) indicated that most parts could receive amounts of between 11-100 mm. Generally, the western and central regions were forecasted to be wetter with regions of the northeastern being driest, with about 101-435 mm and less than 10 mm of rainfall respectively. The downscaled forecasts from both models had a close spatial resemblance with the observed rainfall amounts for October 2015 (Figure 1 a, b and c). The computed difference between the observation and forecasts (Figure 1 d, e) shows that both models had tendency to over-forecast in the western parts of Kenya around the Lake Victoria basin. GFDL was better than the CFSv2 model in simulating the precipitation patterns since majority of the regions had deviations in the range of -19 to 20 mm (shaded cyan) as compared to the CFSv2 forecast that had larger coverage of the less than 20mm classes, in the majority of central and eastern regions of Kenya. However, this is with the exception of the coastal region where the GFDL forecasts had a tendency to over-estimate.

Tanzania was forecasted to experience dry conditions in the central and southern parts and wet conditions in the northern regions around the Lake Victoria basin as well as the western and parts of the northen coast (Figure 2 b, c). A similar spatial pattern is depicted by the observation over the same period (Figure 2 a). The computed difference between the observed and downscaled forecasts (Figure 2 d, e) shows that most areas received a fairly good score of between -19 mm and 20 mm. Regions close to water bodies had extreme patterns. Both models overforecasted in areas around the lake Victoria region and underforecasted in the central coastal areas.

Figure 3 shows the forecasts series and cross validated hindcasts (Fig 3a and 3c) for the location 1.02° S latitude and 32° E longitude (a location in Kenya) for the month of
October 2015 for CFSv2 and GFDL-FLOR respectively. The forecasted amounts captured the direction of the observation both in the training and forecast phases though with some few cases of false alarm and underforecasting. Both models gave a forecast within the normal category in the location considered.

Skill scores in terms of ROC and correlation were also computed to gauge on the skill of the downscaled forecasts with a cross-validation window of five years (2011 to 2014). Figure 3b and 3d show the ROC score for the same location for the CFSv2 and GFDL-FLOR respectively. ROC scores for the location 1.02° S latitude and 32° E longitude indicates that CFSv2 had scores of 0.750 and 0.809 whereas the GFDL-FLOR had scores of 0.755 and 0.675 for the forecast categories of above normal and below normal respectively. This implies that the models presented are skilful in capturing categorical forecasts in the given location.

Figure 4 and 5 show the spatial distribution of the ROC scores over Tanzania and Kenya respectively. The figures show that most of the areas over the two countries had ROC values greater than 0.5 with some few patches scoring less than 0.5. One striking area is a small patch over the central Tanzania that scored the lowest ROC score of 0.1 and the western parts of Kenya for both the forecast categories.

Figure 6 shows the Spearman correlation scores for the month of October 2015 for both the CFSv2 (Figure 6a and 6b) and GFDL-FLOR (Figure 6c and 6d) for Kenya and Tanzania. The Spearman correlation gauges the strength of association between two ranked variables. Correlation scores over Kenya for CFSv2 model showed good skills for the south eastern parts with a score of between 0.45 and 0.9 for both CFSv2 and GFDL-FLOR. Over western Kenya, a correlation score of between -0.45 and -0.9 was attained. Most of Tanzania areas had a score of between 0.15 and 0.3. These are possible forecasting gaps that need to be addressed for improved climate forecasts over the East African region.
6.0 Conclusion and Future work

Weather and climate downscaling is an important area for both research and applications. Currently, many techniques are being tested and proposed for use in translating climate forecasts into finer resolutions for improved decision making. Statistical downscaling using the climate predictability tool is one of the techniques.

This study utilized forecast products from two global models namely the National Centre for Environmental Prediction (NCEP) Coupled Forecast System (CFSv2), Geophysical Fluid Dynamics Laboratory -Forecast-Oriented Low Ocean Resolution (GFDL-FLOR) and a high resolution blended CHIRPS rainfall dataset. Error metrics such as Relative Operating Characteristics (ROC) and Spearman Correlation were used for the evaluation of the downscaled forecasts over the two countries. Forecasts for the month of October 2015 were covered in this study.

The main findings of this study shows that downscaled forecasts from both models had a good skill in estimating the October rainfall amounts indicating quality and value addition to the forecast products. There was a good spatial resemblance between the forecasts and the observations. ROC scores for the location 1.02° S latitude and 32° E longitude indicates that CFSv2 had scores of 0.750 and 0.809 whereas the GFDL- FLOR had scores of 0.755 and 0.675 for the forecast categories of above normal and below normal respectively. With an exception of some few patches over Kenya and Tanzania that had low skill, the general results seem to bear the conclusive high skills for the downscaled forecasts.

This paper promotes generation of improved downscaled forecasts to promote food security initiatives at both national and farming community levels. It is expected that the downscaling technique proposed here will be adopted in other climate centers within the Greater Horn of Africa region and beyond and the forecast products utilized for informed policy making and improved decision making. Further research and work is however needed
to improve the existing downscaling techniques and propose new ones. In a future Part II of this work, the results for November, December and the entire OND season will be presented.

Acknowledgements

This is a publication of research findings by IGAD Climate Prediction and Applications Centre (ICPAC). The work was partly supported by CGIAR’s Research Program on Climate Change, Agriculture and Food Security (CCAFS) under a partnership project called “Integrated Food Production and Food Security Forecasting System for East Africa (INAPFS)” in 2015. This paper is part of ICPAC’s contribution to the project under activity 2 [P40A235] which is aimed at developing a robust seasonal climate forecasting system with high spatial and temporal resolution for the East Africa region.
References


Donohoe A, J. Marshall, D. Ferreira, D. McGee 2013: The relationship between ITCZ location and cross equatorial atmospheric heat transport; from the seasonal cycle to
the last glacial maximum. *J. Climate*, **26**, 3597–3618. Doi: http://dx.doi.org/10.1175/JCLI-D-12-00467.1


Kumar K., B. Rajagopalan, M. Cane, 1999: On the Weakening Relationship between the Indian Monsoon and ENSO. *Science 284*(2); 2156-2159; American Association for the Advancement of Science; Washington DC


Mironov D., and M. Raschendorfer, 2001: Evaluation of empirical parameters of the new LM surface layer parameterization scheme: results from numerical experiments including soil moisture analysis. *COSMO technical report 1*, DWD, Offenbach, Germany


Nicholson S.E., 2003: Comments on “The South Indian Convergence Zone and interannual rainfall variability over southern Africa” and the question of ENSO’s influence on southern Africa. *Journal of Climate* **16**: 555 – 562


*J. Climate, 27*, 2186–2208. doi: http://dx.doi.org/10.1176/JCLI-D-12-00823.1.


Figure 1: Downscaled CFSv2 and GFDL-FLOR model forecasts for October 2015 over Kenya (Fig. 1, b and c respectively). Figure 1, d and e shows the difference between the model forecast and October observation. Figure 1a shows the observation for October.
Figure 2: Downscaled CFSv2 and GFDL-FLOR model forecasts for October 2015 over Tanzania (Fig. 2, b and c respectively). Figure 2, d and e shows the difference between the model forecast and October observation. Figure 2a shows the observation for October.
Figure 3: Forecasts and cross validated hindcasts (Fig 3a and 3c) for the location 1.02°S latitude and 32°E longitude (a location in Kenya) for the month of October 2015 for CFSv2 and GFDL-FLOR respectively. Fig 3b and 3d shows the ROC score for the same location for the CFSv2 and GFDL-FLOR respectively.
Figure 4: Relative Operating Characteristics area for the month of October 2015 for the above normal and below normal rainfall forecast categories for Tanzania. Fig 4a and 4b is scores for the CFSv2 model while fig. 4c and 4d shows score for the GFDL-FLOR model.
Figure 5: Relative Operating Characteristics area for the month of October 2015 for the above normal and below normal rainfall forecast categories for Kenya. Fig 5a and 5b is scores from the CFSv2 model while Fig. 5c and 5d shows score for the GFDL-FLOR model.
Figure 6: Spearman correlation scores for the month of October 2015 for CFSv2 (Fig 6a and 6b) and GFDL-FLOR (Fig 6c and 6d) for Kenya and Tanzania