1.0 INTRODUCTION

Tanzania’s economy depends on agriculture as about 90% of its population who live in rural areas is engaged in agricultural activities. These agricultural activities are rain dependent. The seasonal rainfall over Tanzania is highly variable both in time and space. Hence there is a need for forecasts of seasonal rainfall in order to optimize agricultural yields and to provide early warning for impending adverse weather/climate.

Accurate long range forecasts of mean circulation and precipitation patterns are important and help to anticipate seasonal climate anomalies such as drought and potential flooding situations. The skill of the forecast is heavily dependent on the performance of operational forecast models.

While the short and medium range of weather forecasting using various techniques are currently operational in many forecasting centers for a longer period, the operational of long (extended) range weather forecasting has lagged behind. However, the techniques of long range weather forecasting have substantially improved over the recent past (1990’s and early 2000’s). The methods that are used to give long range weather forecast can be classified into three major categories, namely statistical, dynamical and dynamical–statistical method.

The statistical methods are based on empirical relationship derived from historical records. Studies on long range weather forecasting using statistical methods include Farmer (1987), Folland et al (1991), among others. The major set back of these techniques is that they are non-deterministic.

The dynamical method involves detailed integration on computer of the time-dependent equations of motions and state that govern the atmosphere. In dynamic methods for long range weather forecasting, the General Circulation models (GCMs) are used, physically based dynamical model should outscore statistical models, if the ocean–atmosphere system contains sufficient inherent predictability Barnston et al (1994). Forecasting skill using dynamical model of the atmosphere can be achieved by first predicting Sea Surface Temperature (SSTs) for the regions known to be partly responsible for the rainfall variability over land. GCM forecasts can never be perfect because of errors unavoidable present in initial conditions and because of model deficiencies (Holton, 1979).

Shukla, (1981) and Mechoso et al, (1988) carried out numerical simulation studies and observed that GCM can be used to provide skillful seasonal rainfall forecasts for many parts of the globe, especially in the tropics. However the GCMs are usually coarse and do not resolve the local features. In order to resolve the local feature, down scaling is required.

Glahn and Lowry (1972) observed that Model Output Statistics (MOS) approach give better results when incorporating GCM forecasts information into statistical weather forecasts. The preference for MOS approach is derived from its capacity to include directly in the regression equations the influences of specific characteristics of different NWP models at different projections into the future. MOS use quantities from NWP, output as predictor variables in both development and implementation of the statistical equation. The regression equation serves two purposes; that is, it eliminates biases in the original model data and it optimizes the conversion of the model data into forecast information.

It has been found that better forecasts can be derived from a combination of statistical and dynamical methods. In this approach dynamical forecast output is improved over a given area of interest by using statistical methods. Two of the commonly used dynamical–statistical methods are perfect prognosis (PP) and Model Output Statistical (MOS). In this study the MOS approach was used to improve the forecast of the numerical model over Tanzania.

1.1 The study Area and its Climatology

Tanzania is bounded by latitudes $11^\circ$ S to $1^\circ$ S and longitude $29^\circ$ E to $40^\circ$ E (figure 1). The main rainfall generating system over the study area is the convergence zone between the northeast monsoon and the southeast monsoon referred to as inter-tropical convergence zone (ITCZ) (Asnani, 2005). The ITCZ moves northward passing over East Africa during March to May (MAM), while the southward movement takes place during October to December (OND) (Okoola, 1999). Therefore the variation in its position and intensity greatly affect the rainfall amount over Tanzania. Its intensity and position depends on the location, orientation and position of the subtropical anticyclones (Ininda, 1987). The activeness of the ITCZ depends on the divergent or convergent nature of the two monsoons, the moisture depth and vertical structure as well as the stability in the upper levels (Griffiths, 1972).
Table 1. List of Stations used in the study

<table>
<thead>
<tr>
<th>Number</th>
<th>Station Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bukoba</td>
</tr>
<tr>
<td>2</td>
<td>Musoma</td>
</tr>
<tr>
<td>3</td>
<td>Mwanza</td>
</tr>
<tr>
<td>4</td>
<td>Collondo</td>
</tr>
<tr>
<td>5</td>
<td>Engara</td>
</tr>
<tr>
<td>6</td>
<td>Sema</td>
</tr>
<tr>
<td>7</td>
<td>Lushoto</td>
</tr>
<tr>
<td>8</td>
<td>Tanga</td>
</tr>
<tr>
<td>9</td>
<td>Singida</td>
</tr>
<tr>
<td>10</td>
<td>Tabora</td>
</tr>
<tr>
<td>11</td>
<td>Kigoma</td>
</tr>
<tr>
<td>12</td>
<td>Mpanda</td>
</tr>
<tr>
<td>13</td>
<td>Dodoma</td>
</tr>
<tr>
<td>14</td>
<td>Ilongo</td>
</tr>
<tr>
<td>15</td>
<td>Dar es salam</td>
</tr>
<tr>
<td>16</td>
<td>Iringa</td>
</tr>
<tr>
<td>17</td>
<td>Mahanga</td>
</tr>
<tr>
<td>18</td>
<td>Mtwara</td>
</tr>
<tr>
<td>19</td>
<td>Tunduru</td>
</tr>
<tr>
<td>20</td>
<td>Songea</td>
</tr>
<tr>
<td>21</td>
<td>Mbeya</td>
</tr>
<tr>
<td>22</td>
<td>Sumba</td>
</tr>
</tbody>
</table>

The northern sector of the country covering the northern coast and Zanzibar, north Eastern highlands and Lake Victoria basin experiences two rain seasons (bimodal) thus March to May (MAM) with peak in April and October to December (OND),with peak in November.

The southern, western and central part of the country has one rainfall season (unimodal) which occurs between November to April.

Local features such as topography and large water bodies generate mesoscale systems which also play a key role in modulating the weather and climate over the country leading to spatial variation of rainfall.

1.2 General Circulation Models

General Circulation Models (GCMs) are highly sophisticated atmospheric and ocean mathematical models, which use the physical principles of hydrodynamics. These models also include representations of land surface process, sea ice related processes and many other complex processes of the climatic system (Mutemi, 2003). Unlike empirical models, the GCMs involve mathematical representation of the physical processes in the atmospheric and oceanic and ocean-atmosphere interaction. These equations are solved numerically with computers using a three–dimensional grid over the globe. For climate, typical resolutions are about 200 –300 km in the horizontal and 1 km in the vertical in atmospheric GCMs, often with a higher vertical resolution near the tropopause and stratosphere (IPCC, 1995). Many physical processes such as those related to clouds take place in much smaller spatial scales and therefore cannot be properly resolved and modeled explicitly, but their effects must be included in a simple way by taking advantage of physical based relationships based on association with large scale variables. This technique is knows as parameterization.

GCMs have two uses. They are used to predict the future climate from the given initial conditions and secondly they are used to determine the response of model atmosphere in variations in external forcing such as change in vegetation or internal forcing such as changes in the greenhouse gases. This is referred to as sensitivity studies. Sensitivity studies are an important tool of assessing the performance of the models. These studies are also useful in the interpretation of the model products. The global and regional performances
2.2 Analysis of temporal rainfall pattern and Assessment of Model Skill

The temporal pattern of rainfall was studied using standardized rainfall anomalies. Standardized rainfall index enable comparison of rainfall from different stations. The standardized rainfall index was derived by subtracting the mean from the observed value and dividing by the standard deviation.

The simulation skill was using the Pearson correlation coefficient and the percentage variance explained.

2.3 The description of the Numerical Model used in the Study

The General Circulation Model used in this study is the ECHAM model (version 5.2) which evolved from the numerical weather prediction model developed at the European Centre for Medium Range Weather Forecasting, ECMWF (therefore the first part of its name: EC) and a comprehensive parameterization package developed at Hamburg (therefore the abbreviation HAM). This model has been described in detail by Roeckner et al. (2003). In its standard configuration, the climate model uses a pressure based terrain-following vertical coordinate. The horizontal representation is configured using spectral transform techniques at triangular 42 spectral truncation (T42), which is equivalent to a spatial resolution of about 2.8$^\circ$ longitude by 2.8$^\circ$ latitude grids (64 points in the vertical respectively). A semi-implicit time scheme (a scheme in which the unknown or future time level is only applied to terms in the primitive equations that generate fast gravity waves, and the remaining terms handled explicitly) is used for equations of divergence, temperature and surface pressure. The time step for the T42 resolution taken as 24 minutes (Asselin, 1972). The model is forced using the Atmospheric Model Inter-comparison Project (AMIP2) sea surface temperature and sea-ice blended dataset after Reynolds (1988) and Alexander and Mobley (1976). The governing equations of the climate model are presented below.

2.3.1 The Governing Equations

The governing equations and their spectral representation for a general pressure-based terrain-following vertical coordinate $\eta(p$, of different GCMs vary considerably due to many factors. These include resolution, difference in representation of sub grid scale processes such as convective parameterization, and such subjective factors as the tuning of various parameters in the model with their range of uncertainty to optimize model performance in the region of interest to the modelers. In this study, the ECHAM general circulation was used to stimulate seasonal rainfall.

2.0 DATA AND METHODOLOGY

The data sets which were used in this study included, the observed monthly station rainfall over Tanzania for the period 1971-2004, the grided global observed rainfall data, global observed winds from NCEP reanalysis data and the ECHAM model simulated data over Tanzania.

The monthly rainfall data were obtained from Tanzania meteorological Agency (TMA). The ECHAM model forecasts data over Tanzania, global observed rainfall data and global observed winds were obtained from climate modeling section of IGAD climate prediction and the application centre (ICPAC).

2.1 Data Quality Control

There were gaps in the observed rainfall data. Since the study required continuous data, the missing data was estimated using the Arithmetic mean method. However, this was done only for the stations with missing data set less than ten per cent. Those with more data set missing were not included in the study.

In order for the results from data analysis to have proper interpretation, it must first subjected to quality control. The purpose of data quality control is to detect and remove errors in the data sets. One of the causes of the errors in the data sets is discontinuity. Discontinuities in the data that may occur as a result of non-natural influences like changes in observation schedules and methods, instrumental changes, shifting of station sites, urbanization, and other human processes (WMO, 1966). Heterogeneity makes records not comparable over long time periods and between different stations.

The methods that are commonly used to detect and correct heterogeneity in the data set are single and double mass curves. The single mass curve was used in this study to tests for the homogeneity of the data. The mass curves for most of the stations were nearly straight lines showing that the data from most stations are homogeneous.
(1) The parameterized spectral tendencies $P_u, P_v, P_T$, and $P_{q_i}$ in equations 7 – 9 are expressed as given in equations 10 – 13 below.

$$P_u = -g \cos \theta \left( \frac{\partial p}{\partial \eta} \right)^{-1} \frac{\partial J_u}{\partial \eta}$$

$$P_v = -g \cos \theta \left( \frac{\partial p}{\partial \eta} \right)^{-1} \frac{\partial J_v}{\partial \eta}$$

$$P_T = \frac{1}{c_p} \left[ Q_R + Q_L + Q_D - g \left( \frac{\partial p}{\partial \eta} \right)^{-1} \left( \frac{\partial J_s}{\partial \eta} - C_p T (\delta - 1) \frac{\partial J_q}{\partial \eta} \right) \right]$$

$$P_{q_i} = S_{q_i} - g \left( \frac{\partial p}{\partial \eta} \right)^{-1} \frac{\partial J_{q_i}}{\partial \eta}$$

where $g$ is acceleration due to gravity, $C_p$ is the specific heat capacity at constant pressure given by:

$$C_p = C_{pd} (1 + (\delta - 1)q_v)$$

In equations 10 – 13, $J_u$ and $J_v$ are parameterized vertical
heating due to radiation, phase changes and internal dissipation of kinetic energy associated with the $P_u$ and $P_v$ terms, respectively. $S_{qi}$ denotes the rates of change of $q$ due to phase changes and precipitation formation. The K terms in equations 2 – 4 represent the influence of unresolved horizontal scales. Their treatment differs from that of the P terms in that it does not involve a physical model of sub grid scale processes, but rather a numerically convenient form of scale selective diffusion of a magnitude determined empirically to ensure a realistic behaviour of resolved scales.

2.4 Down Scaling of Model Output.

The linear regression model for generating the station rainfall using the ECHAM output rainfall predictors

$$y_i = a + b_i x_i$$  \hspace{1cm} (15)

Where $y_i$ is the station rainfall, $a$ is constant, $b_i$ is a regression constant and $x_i$ is the ECHAM output for the grid box in which the station is lying.

3.0 RESULTS AND DISCUSSION

The results obtained using the various analysis methods described in section 2 are presented and discussed below.

3.1 Annual Cycle of Rainfall and Atmospheric Circulation

The process of model validation involves comparison of simulated and observed climate. This is achieved by graphical comparison of simulated and observed fields, as well as statistical measures designed to assess the model’s ability to reproduce the observed spatial patterns.

Figure 2 shows the simulated and observed annual cycle of rainfall at various locations. Annual cycles were obtained by averaging of observed and model simulated rainfall for the periods of 1950-2000. The results show that the model capability of reproducing the annual cycle of rainfall at various regions over Tanzania. It can be seen the model simulated the Unimodal type of rainfall is observed over the southern, western and central part of the country and bimodal type of rainfall in the northern part of the country.

Comparison was made between the simulated and observed large scale mean flow pattern over Tanzania. Figure 3 show the flow
which are the representative months for the four standard seasons. From the streamlines it can be seen that the model captures the main flow pattern during each season. However minor localized discrepancy between the model simulation and observed which may be attributed to inability of the model to capture the influence of the mesoscale features.

Fig 2(d) Annual cycle of model and observed over Mtwar

Fig 3a January surface winds for simulated and observed in vector form

Fig 3b April surface winds for simulated and observed in vector form

Fig 3c July surface winds for simulated and observed in vector form

Fig 3d October surface winds for simulated and observed in streamlines
3.2 The Simulated Rainfall Using the ECHAM and the Downscaled Results

Figures 4 to 5 show time series of both the simulated and observed rainfall for the two main rainfall seasons, OND and MAM, for the period 1950 to 2000. The figures compare the observed against both the model output and the downscaled rainfall. The results show that the ECHAM was able to capture the major dry and wet events during both seasons. For example, during the MAM season the dry periods were 1953-54, 1959-60, 1973-74, 1996-97 and wet periods were 1955, 1962,1974, 1979-80, 1992. However the model over estimated the MAM rainfall in 1998, 1955, 1960, 1976, 1993 and under estimated the MAM rainfall during 1959, 1970, 1985 and 1994. There was a good march for the observed and simulated rainfall during the OND season, although the model underestimated the wet periods of 1962, 1967.

Table 2. Regression equation for simulating the rainfall from the model output (Only the stations where the MOS increased the skill by at least 20% are shown)

<table>
<thead>
<tr>
<th>STATION</th>
<th>OND</th>
<th>MAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bukoba</td>
<td>-609.2+68.8*ECHAM</td>
<td></td>
</tr>
<tr>
<td>Loiondo</td>
<td>-157.4+47.4*ECHAM</td>
<td></td>
</tr>
<tr>
<td>Same</td>
<td>-157.4+47.4*ECHAM</td>
<td></td>
</tr>
<tr>
<td>Dodoma</td>
<td>-40.2+26.6*ECHAM</td>
<td></td>
</tr>
<tr>
<td>Mbeya</td>
<td>44.7+25.4*ECHAM</td>
<td></td>
</tr>
<tr>
<td>Mtwara</td>
<td>-103.6+46.8*ECHAM</td>
<td></td>
</tr>
</tbody>
</table>

In general the ECHAM model was more skillful in simulating the October-December rainfall season compared to MAM. This may be attributed to the fact that the model skill is driven by observed SSTs. Past studies have indicated strong correlations between global SSTs and seasonal rainfall over the region during October to December months compared to March to May periods. Also October to December was normal occurrence of peak ENSO events (Mutemi, 2003, Ogallo et al., 1988, Ininda, 1995)

From figures 4 and 5 it can be seen that there was improvement in the skill of simulation for both the OND and MAM seasonal rainfall when the model output statistics (MOS) was used. The MOS used in this study involved developing regression equations from station data as predictand and ECHAM outputs as predictors. Table 2 shows the equations of the stations for the MOS increased the explained variance by more than 20%. The OND season had more station where the skill improved significantly when MOS was applied. For example for Same the explained variance increase from 37% for ECHAM forecasts to 69% when MOS was applied.
4.0 CONCLUSIONS.

The study analyzed the rainfall and the circulation over Tanzania simulated by the ECHAM model forced with observed Sea-Surface Temperatures over the 1950-2000 period. The model captured well the main features observed over Tanzania. The annual cycle and interannual rainfall variability were reproduced by the model especially during the OND season. This suggests that the model can be used for predicting seasonal rainfall over Tanzania.

MOS as a technique of down-scaling significantly improved the prediction skill especially during the OND season. It was therefore concluded that the MOS technique may be a good method to be incorporated with NWP to improve the forecast skill.
REFERENCES


Griffiths, J.M, 1972: Climate of Africa World survey of climatology vol. 10


WMO, 1966: Some methods of climatological analysis, WMO Tech. Note no. 81, WMO, Geneva