Predictions of Future Aspects of the Rainy Season Using Simple and Multiple Linear Regression Analysis- A Case Study of Chingóme Mission Daily Rainfall Data in Zambia

Urban Nchimunya Haankuku
The University Of Zambia
School Of Natural Sciences
Department of Mathematics & Statistics
P.O. Box 32379, Lusaka, Zambia

Abstract

This article demonstrates the point and interval predictions of the dependent variables Y using both simple and multiple linear regression analyses for the given independent X variables. The primary methodology was analysis of quantitative data collected at Chin'gombe mission, northern part of Zambia, weather station for a period of 25 years. The article begins by justifying why a particular approach was used for analysis by testing the available data for randomness. Since, time trends where not evident, a classical approach was adopted which involved the construction of models that reflect the available data as closely as possible. A distribution with two parameters was preferable for greater flexibility, hence, the truncated exponential distribution with two unknown parameters was investigated instead of other distributions such as; lognormal, gamma, or weibull. But the predictions obtained were not particularly informative for agricultural planning, water management, and designing purposes. The article also shows methods of analysis of daily rainfall data using both simple and multiple linear regression analysis and demonstrates how the models derived can be of direct use in agricultural planning, water management and designing. The correlations between the onset and end of rainy season dates were also investigated. The results obtained showed that there is no correlation between the onset and end of rainy season.

Introduction

Throughout the world considerable efforts are devoted to the collection of rainfall data. Records for many years exist and most countries like Zambia, now has a reasonable dense network systems of rainfall stations. Factors controlling the distribution of rainfall over the earth's surface are the belts of converging ascending airflow, air temperature, moisture bearing winds, ocean currents and distance in land from the coast and mountain ranges. Rainfall distribution is one of the primary elements in the climate and a factor of tremendous importance in the distribution of plant and animal life. In addition to daily, monthly and annual totals, the depth of individual rainfalls and their intensity (amount of rain falling during a specific period of hours or minutes) and other pertinent facts are recorded. Floods in Zambia have caused significant damage to properties, crops and loss of life in the past. In order to mitigate against such negative impacts, accurate predictions of onset/end of the rainy season, distribution of the rainfall within the season, rainfall intensity and total rainfall are potentially of great benefit for agricultural planning, water management and designing.

This article demonstrates methods of predicting future aspects of rainfall for a given rainy season and shows how the models derived can be used to give results that are of direct use in agricultural planning, water management and designing (such as ; planning on what variety of crop to use, construction of reservoirs, airports, road constructions, dams, and flood controls etc). Although various climatic variables interact with the crops in complex ways, rainfall is the crucial factor in most parts of the Tropics.

For effective agricultural planning on a rational and scientific basis, appropriate analysis of rainfall procedures need to be set up. Fairly accurate predictions within a given season or future seasons are of great importance to Agriculturalists and Designers. Agriculturalists and Designers need to know with confidence when the rainy season is expected to begin/end, amount of rainfall expected, distributions of the rainfall, rain intensity, length of dry and wet spells.

Definition of variables

There has never been any rainfall recorded in Zambia before 23rd September, and this date was used as an origin in this study.

x_1 = The beginning of the rainy season (meteorological definition) measured in days from 23rd September

x_2 = The first day of at least two consecutive days with rainfall and the accumulated rainfall is $\geq$ 20mm measured in days from 23rd September.

x_3 = Total rainfall from 23rd September to 30th November

x_4 = Total rainfall from 23rd September to 31st December.

Y_1 = The end of the rainy season (meteorological definition) measured in days from 23rd September.

Y_2 = Total rainfall for the rainy season

Meteorological definitions:

Rain: Liquid water droplets that falls from the atmosphere, having diameters greater than drizzle (0.5mm).

Beginning of rainy season: A block of seven days with the first three days with rainfall and the total rainfall $\geq$ 20mm, the first day is called the beginning of the rainy season

End of rainy season: A block of seven days with the last three days with rainfall and the total rainfall $\geq$ 20mm, the last day marks the end of the rainy season.

Literature Review

Southern African Development Community (SADC) Drought Monitoring Centre (DMC) ; obtain forecast based on probabilities derived from 30 years of historical rainfall data. The centre considers the state of the global oceanic – atmospheric system and its implications for the region. This forecast is for aggregate rainfall over three month periods and for relatively large areas. This means local and month-to-month variations are not considered. This forecast is influenced by the current E1 Niño (El Niño events are large climate disturbances which are rooted in the tropical Pacific Ocean, and occur every 3 to 7 years. They have a strong impact on the continents around the tropical Pacific, and some climatic influence on half of the planet). International research Institute for climate Prediction: Based on historical data, the United States Geological Survey (USGS) analysis shows that when October rainfall is above average, January rainfall that follows is always drier than average. On the basis of this study, (USGS) conclude that a dry spell is likely to occur during the second half of the season. This analysis relates to regional trends, and the local trends might not follow what has been observed as a region-wide phenomenon. The USGS through Famine Early Warning System Network (FEWSNET), in collaboration with the SADC Regional Remote Sensing Unit (RRSU) has released a forecast on likely end season maize crop performance based on likely rainfall performance. The model uses the Water Requirement Satisfaction Index (WRSI) which estimates the percentage of a crop's total seasonal water requirement that is met from available rainfall and soil moisture. Using this index they are able to forecast whether the yield will be normal, below normal or above normal (Dlamini, 2002). According to Adams, et al (2011) one of the problems for climatologists was the difficulty of predicting rainfall. Rainfall involves a lot of different processes some of which cannot be captured in long term projections. Very little is known about how much climate change would affect one of Australia's primary rain delivery systems E1 Niño.

Odekunie, Balogun, and Ogunkoya (2005), generated models for predictions of rainfall onset and retreat, using composite rainfall promoting factors namely, sea surface temperature of the tropical Atlantic ocean, land/sea thermal contrast, the tropical Atlantic ocean, surface location of the inter-tropical discontinuity, the land surface temperature, rainfall, and temperature data were collected from selected locations in Nigeria. The results obtained showed that the hypothesized rainfall promoting factors are efficient in predicting rainfall onset and retreat in Nigeria. Sunyoung, Sungzoon, and Wong (1998), predicted daily rainfall at 367 locations based on daily rainfall at nearly 100 locations in Switzerland. The results obtained revealed that Radial Basis Function Network (RBF) Networks produced good predictions while the linear models produced poor predictions. Campling, Globin, and Feyen(2001), investigated whether a climatologically variogram model was appropriate for mapping rainfall taking into consideration the changing rainfall characteristics through the wet season.

The research showed that although distinct wet season phases could be established based on the temporal analysis of daily rainfall characteristics, the interpolation of daily rainfall across medium sized catchments based on spatial analysis was better served by using the global rather than the wet season phase climatologically variogram model.

Possible causes of rainfall variation are E1 Niño, quasi-Bienial Oscillation and surface air temperature which were found to have the same correlation with the large scale features of the summer monsoon zonal as well as all- India series, with 20 selected regions (Singh, Sontakke, 1999). The International Research Institute for climate prediction believes that a scientific breakthrough in the prediction of climate variations such as E1 Niño could provide for improved management in sectors of agriculture, food security, public health and water resources, and materially improve the lives of people in many parts of the world (Zebiak, 2005). The short term rainfall prediction system was developed with the longest prediction lead time of six hours, the grid interval of 2.5km and the update interval of prediction of every 10 minutes. The results of the extrapolation model and meso-scale atmospheric model were combined to provide accurate rainfall prediction data for up to 6 hours ahead (Masashi, Koki, Shuichi, 2004). There is an overwhelming scientific consensus that carbon dioxide in the atmosphere released by burning coal, oil, and other fossil fuels has begun to warm the planet. The surface temperature of the eastern Atlantic Ocean has increased by nearly one degree Celsius in the last century, and the temperature will likely increase in the future due to warmer sea-surface temperatures and annual weather patterns associated with global warming (Batten; et al, 2007).

Statement of the problem

Records of many years of daily rainfall data exist and most countries like Zambia, now have a reasonable dense network systems of rainfall stations. With this available data Agriculturalists and other water users would like to know with confidence some aspects of the rainy season before the rainy season for planning purposes. There has been little analysis of daily rainfall data in Zambia which would guide agriculturalists and water users.

Objectives

The main objective of this paper is to demonstrate methods of predicting certain aspects of rainfall in any given season, given variables of the rainfall in the first part of each rainy season.

Methodology

With the available data, it was necessary to begin by testing whether the given data display a time trend. If a significant time trend was observed, then a time series approach was going to be used for the analysis of data. The run tests for randomness and serial correlation used for the tests of randomness indicated that the data available for analysis may be assumed to be purely random, since a time trend was not evident.

Therefore, a classical approach was used which involves the construction of models that reflects the available data as closely as possible. There was no reason why any one particular distribution can be used since there has been no study made of the various possibilities. But a distribution with two parameters was preferable for greater flexibility. Possible distributions are: lognormal, gamma, weibull and truncated exponential. In this paper a truncated exponential distribution was investigated. The advantage of the truncated exponential is that Lawless (1979) has derived expressions for upper 100p percent bounds for a predicted value. The predicted values of the dependent variables using this model were used.

In this article simple and multiple linear Regression models were also used for the Analysis. The choice of these models was to determine variables which would produce accurate predictions of the aspects of the rainy season.

Applications of the project

In Zambia, where annual rainfall is moderate to meager over large parts of the country, and year to year variations are large, unpredictable weather can prove to be a serious obstacle to stable and systematic agricultural production. For effective agricultural planning on a rational and scientific basis appropriate analysis of rainfall procedures within a given season or for future seasons are of great importance to agriculturalists and designers.

Agriculturalists and designers need to know with confidence when the rainy season is expected to begin/end, and how much rain is likely to fall in a given season. They may also want to predict reasonable expectations of dry spells, the distributions of the rainfall within the season.

Fairly accurate predictions of the beginning/end of rainy season, total amount of rainfall, distributions of rainfall within the season in any given rainy season would assist in planting decisions, water management and transportation of agricultural produce.

The period of maturity for various crops differs from one crop to another. The table below shows various crops grown in Zambia and their maturity periods (Appendix 1).

Maize is a major cereal staple food in Zambia. It is a sub tropical plant which prefers hot sunny conditions with reliable and evenly distributed rainfall. The crop is grown extensively in southern, central and eastern provinces where soils and climate are conducive for high yields. In the northern higher rainfall areas, and in the drier parts of western province, Gwembe, and Luangwa valleys, yields are lower and production is less. For a maize grower to plan properly on when to plant, the type of variety to be used, information on the beginning/end of the rainy season and the distribution of the total rainfall in that rainy season is needed. Maize is one crop which needs enough moisture in the soil at the time of germination and pollination otherwise the crops suffer from moisture stress and that would result in poor yields. At present the planting decisions are often associated with beliefs about annual rainfall patterns which haven't been objectively studied. People have beliefs such as: Early beginning of rainy season means early ending of the season, late beginning means late ending of the season. On the bases of these beliefs, people normally make their planting decisions.

In some years rainfall may have been adequate if planting had been done at the right time, which means inadequacy of the rainfall may be due to poor planning based on beliefs. Rainfall analyses also help in transportation planning. The produce needs to be taken for sale, therefore, an early estimate of the size of the yields would help to prepare storage, organize transport and arrangements of shifting the agriculture produce from one place to another. Poor transport system would result in wastage of agricultural produce. This type of planning can be done effectively if rainfall analysis is well established.

The other application of rainfall analysis is on water management. Long term average rainfall is very important for water conservation, although the variability of rainfall tend to be from year to year. Distribution of total rainfall in any given season is one of the most important aspects of rainfall analysis for agriculturalists and other water users.

The run test for randomness

The run test is distribution free. It serves to test the independence (the random order) of sample values. A run is a sequence of identical symbols preceded or followed by other symbols. Runs are obtained not only for dichotomous data but for measured data that are divided into two groups by the median. The null hypothesis H_o that the sequence is random is a two sided problem opposed by the alternative hypothesis H_A that the given sequence is not random. In the one sided question the H_o is opposed either by H_{A1} (cluster effect) or H_{A2} (regular effect). The critical bounds $r_{lower} = r_l$ and $r_{upper} = r_u$, for $n_1, n_2 \leq 20$ are given below (where n_1 and n_2 are the number of times the two symbols appears). Two sided test: $r_l < \hat{r} < r_u$; H_o is retained, otherwise, H_o is rejected

One sided test: H_o is rejected against H_{A1} (resp. H_{A2}), if $\hat{r} \leq r_l$ or $\hat{r} \geq r_u$

The run test can also serve to test the null hypothesis that two samples of about the same size originate from the same sample.

In this study a classical approach was used since a time trend was not evident from the run test of randomness, for all the given X and Y variables under consideration.

Distribution of the rainfall

In order to determine the distribution and intensity of rainfall for a period of time, box plots gives sufficient details of short term variability's in heavier rainfall periods and facilitate year to year comparisons of rainfall patterns. Given the daily rainfall data at a particular weather station, for each rainy season divide the total amount of rainfall into 25%. This will give four intervals with equal amount of rainfall in each interval. The length of each interval depends on the intensity of the total rainfall received in that period. Then variability's and year to year comparisons rainfall patterns can be established.

Truncated Exponential Distribution

Suppose that X has a truncated exponential distribution of density

$$f(x) = \frac{1}{\alpha} e^{-(x-\beta)/\alpha} \ , \ x > \beta$$

Where α and β are unknown parameters. Using results derived by Lawless J.F (1977). it may be shown that given the sample values $x_1, x_2, \ldots, x_n$ the predictive distribution of X is such that:

$$P(x \le c) = 1 - \frac{n}{n-1}\left(1 + \frac{c - x'}{\acute{u}}\right) - (n-1) \ , \ for \ c \le x$$

Where $x_1^{'}$ is the smallest sample value and $u = \sum_{i=1}^{n} x_i - n x_1^{'}$

Thus for any given p, $P(x \le c) \ge \frac{1}{n+1}$, we find that $c = x_1^{'} + u\left\{\left[\frac{n}{(n+1)(1-p)}\right]^{\frac{1}{n+1}}\right\} - 1$

Which is an upper 100p percent bound for the predicted value of x, for selected values of p for 2011/12 rainy season?

Applying these results to the various X's and Y's under consideration, appendix 2 below give the 100p percent upper bounds for the predicted X and Y for the selected values of P. When appropriate, the value of c (in days) is converted into a date.

In the appendix 2, the c is an upper 100p percent bound for the values of X and Y in the next season given by the truncated exponential distribution.

The results displayed are not particularly informative for Agriculturalists. It is possible that an alternative distribution model such as the lognormal or Weibull would have been a better choice. But the advantage of the truncated exponential distribution was that an explicit expression can be obtained for the prediction bounds.

Linear Models

This model illustrate a variable Y which is thought to be a linear combination of one or more variables, usually measured on the same observational unit, and from a series of such measurements estimate the coefficients in the linear combinations.

Suppose that for given values of x_1 and x_2 , we assume the average value of y in the population of units (of which we are assuming , we have just a random sample) which is of the form:

$$E(y) = a_o + b_1 x_1 + b_2 x_2 \quad \ldots\ldots.. (1)$$

In this equation (1), E(y) represents the average or expected value of y corresponding to values of x_1 and $x_{2;}$ and a, b_1 ,and b_2 are some unknown constants. Regression Analysis simply uses the observations on y and the x's to obtain estimates of the b's in equation (1).

Let the linear model be

$y_i = a + \sum_{i,j} b_i x_{i,j} + e_i$ …….(1a) , and $e_i = y_i - E(y_i)$ represents the extent to which an observed y_i differs from its expected value $E(y_i)$.

Thus the e_i 's include all manners of discrepancies between observed y's and their expected values and as such, are considered as random variables, usually called random errors or random residuals. A change in notation is now made: in place of a write b_o, and then for b_o write $b_o x_{io}$ with all values of x_{io} being unity, this give

$$y_i = b_o x_{io} + b_1 x_{i1} + b_2 x_{i2} + \ldots + b_n x_{in} + e_i \quad \ldots..(2)$$

Equation (2) occurs for every set of observations y_i , $x_{io} = 1$, x_{ij} $(j = 1,2,\ldots,n)$ that constitute the data: Hence, equation (2) exists for $i = 1,2,\ldots,n$.

Define the following matrix X and vectors Y, b, and e as follows:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix}; \quad X = \begin{bmatrix} x_{1o} & x_{11} & x_{12} & . & . & . & x_{1n} \\ x_{2o} & x_{21} & x_{22} & . & . & . & x_{2n} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ x_{no} & x_{n1} & x_{n2} & . & . & . & x_{nn} \end{bmatrix}; \; b = \begin{bmatrix} b_o \\ b_1 \\ . \\ . \\ . \\ b_n \end{bmatrix}; \; e = \begin{bmatrix} e_1 \\ e_2 \\ . \\ . \\ . \\ e_n \end{bmatrix} \quad .....(3)$$

Then the complete set of equations represented by equation (2) is:

$$Y = Xb + e \quad ......(4)$$

Equation (4) is called the equation of the model, or model equation. Equation (4) is the basis for estimating b from the observations arrayed in X and Y.

Numerous other models could be postulated, many of which are non linear in their parameters. For example

$$y_i = \alpha x_1^{\beta} x_2^{\gamma} \quad ....(5)$$

This non linear model because the parameters α, β *and* γ occur nonlinearly; but

$E(y) = a + b_1 x_1 + b_2 x_1^2$ is a linear model because it is linear in a and the b's.

However, many non linear models can often be rearranged to be linear. This means nonlinear relationships between variables can often be reduced to linear ones and can be studied analytically using linear regression. Which means that linear model is a technique that has wide applications.

Parameter estimation

The usual estimation procedure is to apply the method of "least square "to (4). It results in choosing the estimator of b those values of b_o , b_i (i = 1,2,…, n) which minimize the expression

$$\sum_1^n [y_i - E(y_i)]^2 = \sum_{i=1}^n [y_i - (\sum b_i x_{ij})]^2 = \sum_1^n e_i^2$$

$$= (Y - Xb)^{'} (Y - Xb) = e^{'} e$$

If $e = \{e_i\}$ for i = 1,2,…,n, the sum of squares is;

$$S = \sum_{i=1}^n e_i^2 = e^{'} e = (Y - Xb)^{'} (Y - Xb)$$

$$= Y^{'} Y - b^{'} X^{'} Y - Y^{'} Xb + b^{'} X^{'} Xb$$

Since $Y^{'} Xb$ is a scalar it is equal to transpose $b^{'} XY$ and hence

$$S = Y^{'} Y - 2b^{'} X^{'} Y - b^{'} X^{'} Xb$$

Minimizing this with respect to b means $\frac{\partial S}{\partial b} = 0$ where a vector of differential Operator is $\frac{\partial}{\partial b}$.

Therefore, $\frac{\partial S}{\partial b} = -2X^{'} Y + 2X^{'} Xb$ and equating this to a null vector leads to the consistent equations

$$X^{'} Xb = X^{'} Y \quad ......(6)$$

With solution

$$b = (X^{'} X)^{-1} X^{'} Y \quad ..... \quad (7)$$

Since every x_{io} in (1_a) is unity, and every x_{ij} are observations, X is a real matrix. Therefore, $X^{'} X$ of (6) is nonnegative definite (n.n.d), and practically all instances of regression analysis X has full column rank, making $X^{'} X$ probability density (p.d.). Hence, $(X^{'} X)^{-1}$ of equation (7) exists.

Notation: b = vector of parameter and $\hat{b}$ = vector of estimator

$\hat{b} = (X'X)^{-1} XY$, and for any specified $X_o = [1, x_1, ..., x_n]$, the least squares estimates of the observed values Y_o of Y is $Y_0 = X_o^{'} b'$ where sampling variance is given by; $\text{var}(\hat{Y}_o) = [1 + X_o^{'}(X^{'}X)^{-1} X_o]^2$.

Furthermore, the $100(1-\alpha)$ per cent predicted interval for Y_o has the limits;

$Y_0 \pm t_{\alpha/2, n-k} \sqrt{MSE(1 + (X_o^{'} X)^{-1} X_o))}$, where MSE is the mean sum of errors, with $k-n-1$ degrees of freedom, (where k is the number of regressor variables) and is calculated as; $MSE = \frac{(Y - X^{'}\hat{b})(Y - X^{'}\hat{b})}{n-k-1}$

Hypothesis testing on the slope

We are often interested in testing hypotheses and constructing prediction intervals about the model parameters. These procedures require that we make an assumption that the errors are normally distributed and independently distributed with mean 0 and variance σ^2.

A very important special case is H_o; that the gradient of the regression line is zero. Failing to reject H_o implies that there is no linear relationship between X and Y. Note that this may imply either that X is of little value in explaining the variation in Y and that the best estimate of Y for any X is $Y = \bar{Y}$ or that the true relationship between X and Y is not linear. Rejecting H_o could mean either that the straight line model is adequate or that even though there is a linear effect of X, better results could be obtained with the additional of higher order polynomial terms in X.

To test the hypothesis H_o, an analysis of variance procedure is used.

The test statistics is;

$$F_o = \frac{MSR}{MSE}$$

Where MSR is the regression mean square and MSE is the residual mean square or mean square error.

The test for significance of regression may also be performed using the t-test equation with H_o , say

$$t_o = \frac{\hat{b}}{MSE/S_{xx}} = \frac{MSR}{MSE}$$

Although the t-test for H_o, is equivalent to the F-test, the t-test is somewhat more adaptable as it could be used for one-sided alternative hypothesis (either $H_1 : b < 0$ *or* $H_1 : b > 0$) while the F-test considers only the two-sided alternative. The inability to show that the slope is not statistically different from zero may not necessarily mean that Y and X are unrelated. It may mean that our ability to detect this relationship has been obscured by the variance of the measurement process or that the range of values of X is inappropriate. A great deal of non statistical evidence and knowledge of the subject matter in the field is required to conclude that H_o is true.

Adequacy of models

It is important to have some checks that the assumptions made in a regression analysis are valid for the problem under consideration. Consider the residuals (errors)

$e_i = Y_i - \hat{Y}_i$, $i = 1,2,...,n$, Where $Y_i = X_i^{'}\hat{b}$, with $X_1^{'} = [1, x_{i1}, ..., x_{in}]$.

It is necessary to consider the validity of the assumption used on the linear model before it is used. Gross violation of the assumptions may yield an unstable model in the sense that a different sample could lead to a total different model with different conclusions.

The t or F statistics or R^2 cannot detect departures from the underlying assumptions. We shall assume that given $x_i = (x_{1i}, x_{2i}, ..., x_{ni})$, the observed values of Y_i are such that $Y_i = b + b_1 x_i + e_i$, $i = 1,2,...,n$, where e_i are independent and normally distributed with variance σ^2.

For any X_i the value of Y on the least square line is: $\hat{Y} = \hat{b}_o + \hat{b}_i X_i$ and the residuals are;

$e_i = Y_i - \hat{Y}_i, \quad i = 1,2,...,n$ This is the deviation between the data and the fitted values. There are many ways of assessing the adequacy of the model. The usual checking methods are as follows: normal probability plot, the plot of the standardized residuals, plot of standardized residuals against standardized predicted Y_i and the coefficient of determination. In this article the assumptions were checked and they satisfied the models adequacy.

Predicting Y_1 (end of rainy season)

In this section prediction of the end of the rainy season was considered having observed $x_1, x_2, x_3, and\ x_4$ for the 20011/12 rainy season. Both simple and multiple linear regression analysis were not significant for all x_i. It is not easy to predict the end of the rainy season, if parts of the aspects of the rainy seasons are known.

Predicting Y_2 (total rainfall for the season)

In this section prediction of the total rainfall for 2011/12 was considered having observed the values of $x_1, x_2, x_3, and\ x_4$ for the season. Appendix 3 below shows the least squares estimates and 95 percent confidence interval for all possible regressions.

Conclusion

The least squares estimates given below (Appendix 3) give the predicted total rainfall for the 2011/12 season for various independent $x_i, (i = 1,2,3,4)$ and the predicted 95 percent confidence interval. The actual total rainfall for this season was 836.3mm, and this value was closely estimated by x_2 with $\hat{Y} = 772.3mm$, with confidence interval of 424.2mm to 1120mm.

The correlation coefficient calculated between the onset and withdraw dates indicated no relationship between the onset and withdraw dates of the rainy season in Zambia.

Zambia is an inland country and receives its rainfall through the converging ascending airflow, and air temperature. The moisture bearing winds from the Indian Ocean through Mozambique, Malawi, and Tanzania dries up by the time they reach Zambia due to distance from the ocean. Hence, predictions of future aspects of rainy season may be more accurate if variables such as; converging ascending airflows, air temperature; moisture bearing winds, daily rainfall, sea-surface temperatures, and altered wind patterns associated with global warming were considered for the analysis.

References

Vivelianandan N: (2007): - Analysis of trend in rainfall using non parametric statistical methods (AIP conference proceedings, vol.923, pp 101-113.AIP homepage)

Campling P; Gobin A; Feyen J: (2001): Temporal and spatial rainfall analysis across a humid tropical catchment. (Hydrological processes, vol 15, issue 3, pp 359-375)

Odekunie,T.O., Balogun, E.E., and Ogunkoya,O.O.: (2005): On the prediction of rainfall onset and retreat dates in Nigeria. (Theoretical and Applied climatology, vol, 81, number 1-2).

Sunyoung Lee; Sungzoon Cho; Patrick M Wong: (1998): Rainfall prediction using artificial Neural Networks. (Journal of geographical information and decision Analysis, vol,2 No. 2, pp 233- 242).

Yang, M,J; Benj D; Wang,S,C: (2004): Ensemble prediction of rainfall during the 2000-2001 Mei-Yu seasons: evaluation(journal of geographical research, vol,109, issue Di8 cite ID 18203(JGRD Home page)).

Singh ,Nityanand; Sontakke,N,A: (1999): On the variability and prediction of rainfall in the post-monsoon season over India.(International journal of climatology, vol.19, issue 3, pp. 309-339).

Adams, D. K., Rui M. S., Jair M. F. M, (2011): GNSS Precipitable Water Vapor from an Amazonian Rain Forest Flux Tower. *J. Atmos. Oceanic Technol.*, **28**, 1192–1198. Vol 28. No.10

Stephen E. Zebiak (2005): International Research Institute for Climate prediction (2005-2010)

Dlamini, E.M.: (2002) : Season rainfall outlook for the period October 2001 to march 2002(SADC Drought Monitoring Centre. MANGOCHI, MALAWI; 17 to 21 September 2001)

Masachi, Terada; Koki Kataoka; Shuichi Ikebuchi: (2004) ; The development of short Term rainfall prediction system in mountainous Region by the combination of extrapolation model and Meso- scale atmospheric model. (Sixth international symposium on hydrological applications of weather radar).

Lawles, J.F, (1977) : Prediction intervals for the two – parameter exponential distribution Technique.

Battern Kit; Bullock Jane; Goldstein B; Haddow G; Hendricks B; Manlove K; Weiss D.J ; (2007) : Forecast: Storm Warnings; (Centre for American Progress)(www.americanprogress.org).

Appendix 1

MAIZE	**MATURITY PERIOD**
Variety	
MM501	130-135 Days
MM502	140-145 Days
MM504	135-140 Days
MM601	140-145 Days
MM603	145-150 Days
MM604	145-150 Days
MM752	155-160 Days
MMV600	130-135 Days
MMV400	100-110 Days

Potato	**Maturity**
Variety	
Arka	115-125 Days
Baraka	125-130 Days
Pent land dell	115-125 Days
Pimpernel	130-140 Days
Up to date	115-125 Days

Soya Beans	**Maturity period**
variety	
Jupita	140 Days
Santa Rosa	110 Days
Kalenga	115 Days
Tania	120 Days
Hermon 147	125 Days
Magoye	130 Days

Sorgham	**Maturity period**
Variety	
ZSR – 1	115 – 125 Days
Framida	100 – 110 Days

Sunflower	**Maturity period**
Variety	
CCA – 75	130 – 135 Days
CCA 81	110 – 115 Days
CH 258	105 – 110 Days

Appendix 2

100p percent upper bounds for the predicted X for selected values of p for 2011/12

variable	$p = p(X \leq c)$	0.5	0.6	0.7	0.8	0.9	0.95
X_1	c date	47 9/11/11	56 18/11/11	67 29/11/11	84 16/12/11	133 8/2/12	143 18/2/12
X_2	c date	35 28/10/11	39 1/11/11	46 8/11/11	55 17/11/11	70 2/12/11	122 24/1/12
X_3	c	28mm	101mm	126mm	161mm	224mm	288mm
X_4	c	154mm	316mm	372mm	452mm	592mm	736mm
Y_1	c date	173 14/3/12	185 26/3/12	198 8/4/12	219 29/4/12	254 3/6/12	309 28/7/12
Y_2	c	537mm	885mm	1006mm	1178mm	1479mm	1789mm

Appendix 3

Least square estimates for all possible regressions

Dependent variable Y_2								
Independent variable	β_o	β_1	β_2	β_3	β_4	F	$\hat{Y}$	95% Conf. interval.
x_1	38.11537	-0.010839				1.36306		
x_2	19.48357		0.29517			4.85680*	772.3mm	424.18mm-1120mm
x_3	25.28937			1.63845		6.00553*	763.0mm	462.53mm-1063.5mm
x_4	22.22603				0.73620	5.15837*	708.5mm	352.81mm-1064.01mm
x_1 , x_2	24.63462	-0.06881	0.27128			2.67742		
x_1 , x_3	26.40394	-0.01495		1.58400		2.88684		
x_1 , x_4	25.55278	-0.04267			0.67792	2.59226		
x_2, x_3	14.70821		0.26993	1.52233		5.97045*	739.4mm	422.4mm-1056mm
x_2, x_4	11.06527		0.28254		0.70629	5.73357*	684.1mm	372.1mm-995.9mm
x_3, x_4	22.37749			1.10177	0.39202	3.41257		
x_1 , x_2 , x_3	12.32716	0.02772	0.27794	1.61985		3.84532*	688.9mm	364.2mm-1013.5mm
x_1, x_3, x_4	22.91445	-0.00692		1.08032	0.38926	2.17377		
x_2, x_3, x_4	11.61853		0.27190	0.96626	0.40557	4.35676*	700.9mm	391.2mm-1010.9mm
x_1, x_2, x_3, x_4	8.31311	0.03712	0.28270	1.07586	0.42089	3.17669*	676.8mm	360.2mm-993.7mm

* means significant