

Forecasting Drought in Rwanda Using Time Series Approach Case Study: Bugesera District

Gashumba Kaminuza Pascal¹, Dr. Joseph K. Mung'atu²

¹Jomo Kenyatta University of Agriculture and Technology, Kenya

²Jomo Kenyatta University of Agriculture and Technology, Kenya

Abstract: *Drought is a global phenomenon that occurs almost in all landscapes causing significant damage both in natural environment and in human lives. Early detection of droughts helps to implement drought mitigation strategies and measures, before they occur. This study finds it necessary to use mathematical and statistical techniques to build models that will be used to forecast occurrences of drought in Bugesera district which is the area in Rwanda hardest hit by droughts. The developed ARIMA model was applied to forecast droughts using monthly rainfall data of Bugesera district, Eastern province of Rwanda from January 1970 to December 2015. The ARIMA generator has proven to be relatively portable across different systems, provide a good source of practically strong random data on most systems. Using the computational analysis of ARIMA approaches, better and accurate results were obtained from non-stationary monthly rainfall data. The forecasted results were compared with the observed results and the results show reasonably good approximation. So the obtained ARIMA model can be used to forecast droughts up to 5 months of lead-time with reasonably accuracy as it produce the lower error value. Furthermore, the study revealed that the model can improve the current forecast future conditions, which will allow the communities to plan ahead the water management activities during droughts.*

Keywords: Forecasting drought in Rwanda, Time series approach

1. Introduction

Drought has long been recognized as one of the most insidious causes of human misery and being the natural disaster that annually claims the most victims. Its ability to cause widespread misery is actually increasing (Perez, 1995). Drought is a temporary, recurring natural disaster, which originates from the lack of precipitation and brings significant economic losses. Water scarcity can be said as the cause and effect of drought. Drought occurs whenever and whenever the links in hydrological cycle is broken or is destabilized. Drought is a slow poison, no one knows when it creeps in, and it can last any number of days and its severity cannot be predicted. It is not possible to avoid droughts. But drought preparedness can be developed and droughts impacts can be managed. The success of both depends, amongst the others, on how well droughts are defined and drought characteristics quantified (Smakhtin., 2004). As the water is vital to all the perception of drought varies with the people concerned: via-meteorology, deficit in rainfall amount; water resources, low river-flow level or reservoir storage level or decreased ground water level, agriculture-deficit or no water during critical crop watering time leading to growth deficiency and poor crop yield, economy profitability descent, and commerce-food items are in short supply (R.Nagarajan, 2004) drought definitions vary from region and may depend upon the dominating perception, and the task for which it is defined. Drought always starts with the lack of precipitation, but may (or may not, depending on how long and severe it is) affect soil moisture, streams, groundwater, ecosystems and human beings. This leads to the identification of different types of drought (meteorological, agricultural, and hydrological) which reflect the perspectives of different sectors on water shortages. A drought is prolonged, abnormally dry period, when there is not enough water for users normal needs. Conceptually it can be defined as a protracted period of deficient precipitation resulting in extensive damage to

crops, resulting in loss of yield (center, 2005). Drought is an interval of time, generally of the order of months or years in duration, during which actual moisture supply at a given place rather consistently falls short of the climatically expected or climatically appropriate moisture supply (Palmer, 1965).

2. Statement of the Problem

Drought is a dangerous hazard of the nature that causes significant damage both in natural environment and in human lives. It originates from a deficiency of precipitation over an extended period of time, usually a season or more. Although mathematical and statistical techniques are widely used to forecast occurrences of drought, however such methods are rare used for Rwanda situations. This study finds it urgent to use mathematical and statistical techniques to build models that will be used to forecast occurrences of drought in Bugesera district.

3. Justification

Despite the community's exposure and experience with drought that has become common phenomenon in Rwanda, by and large many households have remained vulnerable to subsequent Droughts. A lot work on a forestation, rainwater harvesting, and water storage (Dams) assessment has been extensively covered by Government of Rwanda through SPAT (Strategic Plan for Agriculture Transformation). These have been partly government and produced controversial results. Therefore, the scope of this work is different from these studies as it attempts to focus in detail on Drought Forecasting in Rwanda using Time-series approach. In addition to this, the Drought forecasting is necessary to know what is going. To happen and put in place a Drought early warning system that will involve indecision making about the future Drought.

4. Objectives of the Study

4.1 General objective

The main objective of this research was to develop the statistical model that forecast the occurrence of droughts in Bugesera district, Rwanda and use it to forecast future Droughts in Bugesera district, Rwanda.

4.2 Specific objectives

The objectives of this research were:

- i. To fit the autoregressive integrated moving average of the time series forecasting droughts in Rwanda.
- ii. To determine the statistical time series model that suggests forecasting droughts in Rwanda.
- iii. To select the autoregressive integrated moving average that corresponds to forecasting droughts in Rwanda.
- iv. To fit the observed and forecasted values in Rwanda using autoregressive moving average model.

5. Research hypotheses

$H_0 : \phi = 1$ (*series contains a unit root*)

$H_1 : \phi < 1$ (*series is stationary*)

6. Methodology

6.1 Data analysis

6.1.1 Introduction

The data that used in this study were secondary data of the monthly rainfall of three stations (Karama, Ruhuha and Nyamata) of Bugesera district from January 1970 to February 2016. Referring to what have seen in literature review, this stage deals with the formulation of ARIMA model that fit the existing data to forecast the occurrences of drought in Bugesera. Building an ARIMA (p, d, q) model requires us to first determine the differencing (d orders), the AR terms needed (AR orders, p) and the MA terms needed (MA orders, q). This process is called model identification. We tried to use the fewest number of these terms as possible. Once they have been identified, we estimated the AR and MA parameters and check that the model fits adequately. When a reasonably fitting model has been derived, it was used to generate forecasts, test for interventions, and predict the values and explore the relationship with other series.

6.1.2 Series Plots and Stationarity

Examination of a time series plot allowed us to evaluate some necessary conditions for fitting an ARIMA model. The most important of these is stationarity. If we discover that our series is non-stationary, we may be able to render it stationary by differencing. Another aspect of stationarity is that the variance of the series should remain constant. In some series the variance changes as a function of the level of the series, either because of the nature of the

measurement or because of floor or ceiling effects. If the variance changes, we should consider transforming the series before analyzing it (Box, 1976). However, care must be taken when using transformed series for forecasting (Pankratz, 1983) or testing interventions (Hay, 1980).

6.1.3 Autocorrelations and Partial Autocorrelations

We used the autocorrelation and partial autocorrelation function to determine p orders and q orders and confirm the differencing that was suggested by the plot. The autocorrelation at lag k , ACF (k), is the (linear) Pearson correlation between observations k time periods (lags) apart. If the ACF (k) differs significantly from zero, the serial dependence among the observations must be included in the ARIMA model (Durdu, 2010). Like the ACF (k), the partial autocorrelation at lag k , or PACF (k), measures the correlation among observations k lags apart.

However, the PACF (k) removes, or "partials out," all intervening lags. In general, for autoregressive processes ARIMA ($p, 0, 0$): the ACF declines exponentially and the PACF spikes on the first p lag. By contrast, for moving average processes ARIMA ($0, 0, q$): the ACF spikes on the first q lags and the PACF declines exponentially. Mixed processes ARIMA (p, d, q) decline on both ACF and the PACF. Finally, if the ACF or PACF declines slowly, the process is probably not stationary and therefore should be differenced. Usually, a large number of significant autocorrelations or partial autocorrelations indicates non-stationarity and a need for further differencing.

7. Empirical Study

7.1 Model Development

Time series model development consists of three stages identification, estimation, and diagnostic checking (Box, 1976). The identification stage involves transforming the data (if necessary) to improve the normality and the stationarity of the time series and determining the general form of the model to be estimated. Graphical procedures are used (plotting the data over time and plotting the acf and pacf) to determine the most appropriate specification. During the estimation stage the model parameters are calculated using the method of moments, least square methods, or maximum likelihood methods. This involves model checking i.e. determining whether the model specified and estimated is adequate. Box and Jenkins suggest two methods: Overfitting and residual diagnostics. Overfitting involves deliberately fitting a larger model than that required to capture the dynamics of the data as identified in identification stage. If the model specified at identification stage is adequate, any extra terms added to the ARMA model would be insignificant. Residual diagnostics simply checking the residuals for evidence of linear dependence which, if present, would suggest that the model originally specified was inadequate to capture the features of the data. The acf, pacf or Ljung-Box tests could be used. Finally, diagnostic checks of the model are performed to reveal possible model inadequacies and to assist in selecting the best model.

7.2 Identification Stage

The data series are plotted against time, except in 1994 and 1995 because of Genocide in Rwanda.

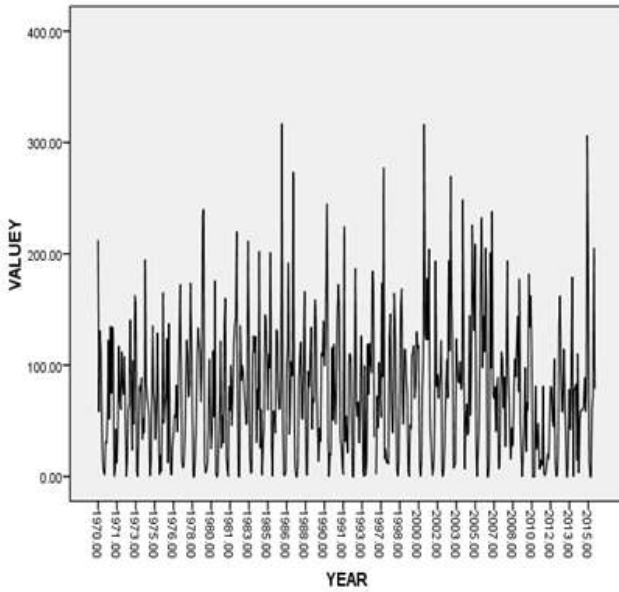


Figure 1: Plot of Monthly Rainfall against Time

7.3 Test for stationarity

This test will help us to know if our data are stationary or not using Augmented Dickey-Fuller test and including intercept, Trend and Intercept or none of these (no intercept, no Trend). If the absolute value of calculated ADF is greater than the absolute value of 5% critical value, this means that there is stationarity.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
YT(-1)	-0.216723	0.030381	-7.133470	0.0000
D(YT(-1))	-0.156402	0.042419	-3.687120	0.0002

Figure 2: ADF test on YT: Rainfall

7.4 Test of Stationarity for First Difference

This figure represents plot for first difference, this plot seems to be stationary. This will be confirmed with unit root test of the first difference.

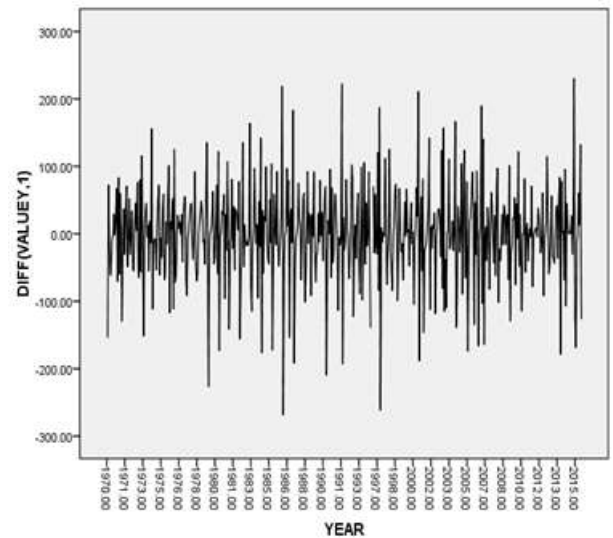


Figure 3: Plot for first difference

7.5 Autocorrelation and Partial Autocorrelation Plots

As we found that the first difference of our series is stationary, we have to proceed with the identification of tentative models, by using the correlogram of stationary data. The first difference of monthly rainfall turns out to be a stationary process. The next step is to choose the ARMA model that best fits this data using a correlogram.

Correlogram of D(RAINFALL)

Date: 11/17/16 Time: 16:15
 Sample: 1970M01 2015M12
 Included observations: 549

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
1		1	-0.262	-0.262	37.999	0.000
2		0.080	-0.160	-0.160	41.549	0.000
3		-0.168	-0.259	-0.259	57.238	0.000
4		-0.081	-0.267	-0.267	60.866	0.000
5		0.018	-0.209	-0.209	61.053	0.000
6		0.137	-0.050	-0.050	71.545	0.000
7		0.071	0.019	0.019	74.322	0.000
8		-0.127	-0.130	-0.130	83.388	0.000
9		-0.154	-0.261	-0.261	96.695	0.000
10		-0.054	-0.302	-0.302	98.336	0.000
11		0.054	-0.328	-0.328	99.967	0.000
12		0.276	-0.080	-0.080	142.97	0.000

Figure 4: ACF and PACF of first difference

Table 1: ACF and PACF Behaviours

Model	ACF Behaviour	PACF Behaviour
AR(p)	Decays Gradually	Spike in Lag p
MA(q)	Spike in Lag q	Decays Gradually
ARMA(p, q)	Decays Gradually	Decays Gradually

The figure 4 above is called a correlogram of the first difference in monthly rain fall and the figure 7 is used to identify which kind of model we should be using AR, MA, or mixture model ARIMA. Lags are individually statistically significantly different from zero. ACF decays gradually and PACF has a large spike at first lag; this can be AR (p), MA (q) or an ARIMA (p, d, q). Based on the output of correlogram of first difference, we may tentatively choose these models: AR (1), ARIMA (1, 1, 0), AR (2) and MA (2). We will have to apply diagnostic tests to find out if the chosen ARIMA model is reasonably accurate.

7.6 Estimation Stage

As we have many tentative models, we choose one based on characteristics of a good model. Through these we choose ARIMA (1, 1, 0).

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.039250	2.276548	0.017241	0.9863
AR(1)	-0.265889	0.041350	-6.430209	0.0000

Figure 5: Estimation of Equation

From the Eviews output, estimating an ARIMA (1, 1, 0) of the difference of quantity of monthly rainfall.

$$AR(1): Y_t = \phi_1 Y_{t-1} + \varepsilon_t \quad (4.3)$$

$$\text{In backshift notation: } (1 - B\phi_1) \tilde{y}_t = \varepsilon_t \quad (4.4)$$

Where $\tilde{y}_t = y_t - \mu$ and \tilde{y}_t are the deviations. μ is the mean or

$$(y_t - \mu) - \phi_1 (y_{t-1} - \mu) = \varepsilon_t \quad (4.5)$$

This gives $y_t - \mu = B\phi_1 y_t - \phi_1 \mu + \varepsilon_t$ and $B y_t = y_{t-1}$,

$y_t = \mu - \phi_1 \mu + \phi_1 y_{t-1} + \varepsilon_t$ by transformation we obtain

$$y_t = c + \phi_1 y_{t-1} + \varepsilon_t \text{ where } c = \mu(1 - \phi_1) \text{ is the constant.} \quad (4.6)$$

Output estimation:

$$DY_t = 0.039250 - 0.265889DY_{t-1} \quad (4.7)$$

$|-0.265889| < 1$: Stationarity condition is confirmed.

7.7 Diagnostic Checking

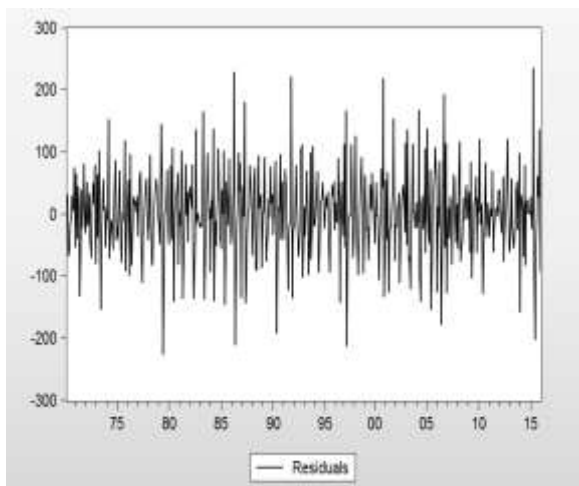


Figure 6: Plot of Residuals of our Estimated Model

From the figure 6 we do not know if the residuals are stationary or not. To know it, let us use ADF test

Correlogram of RESID

Date: 11/06/16 Time: 13:59
 Sample: 1970M01 2015M12
 Included observations: 547

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1		-0.041	-0.041	0.9087	0.340
2		-0.216	-0.218	26.643	0.000
3		-0.241	-0.274	58.680	0.000
4		-0.139	-0.258	69.360	0.000
5		0.031	-0.171	69.876	0.000
6		0.186	-0.011	89.084	0.000
7		0.081	-0.030	92.730	0.000
8		-0.167	-0.206	108.35	0.000
9		-0.222	-0.300	135.82	0.000
10		-0.091	-0.341	140.44	0.000
11		0.122	-0.327	148.75	0.000
12		0.377	-0.043	228.63	0.000

Figure 7: Correlogram of Residuals of our Estimated Model

Looking at this output of the residuals of first difference of our data series up to lag 12, none of the autocorrelation and partial autocorrelation is individually significant.

Augmented Dickey-Fuller unit root test:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
T(-1)	-1.156592	0.062032	-18.64504	0.0000
D(T(-1))	0.110775	0.043180	2.565404	0.0106
C	0.112065	2.853441	0.039273	0.9687

Figure 8: Augmented Dickey-Fuller Unit Root Test of Residuals

In the figure above:

$$\Delta T_t = 0.112065 - 1.150592T_{t-1} + 0.110775T_{t-1} \quad (4.8)$$

Where ΔT_t the change of residuals. As our calculated ADF (-18.64504) this absolute value is greater than the 5% critical Value (-2.86663) this means that our series is stationary in the residuals (Presence of homoskedasticity in the series). The errors of the model appear to be a white noise process, mean 0 and constant variance. The Augmented Dickey Fuller test rejects the hypothesis that the model's error term is non-stationary. The diagnostic check of the models performance validates unbiased estimates.

7.8 Forecasting

One of the most useful models for forecasting is the ARIMA model. To produce dynamic forecasts the model needs to include lags of either the variables or error terms. Breaking down the lags and first differences yields the final forecasting model that will be used to forecast the quantity of monthly rainfall. As we found that our model is $DY_t = 0.039250 - 0.265889DY_{t-1}$, we are going to forecast based on it. Only five first months will be forecasted in 2016

$$\begin{aligned}
 DY_{(2018)_t} &= 0.039250 - 0.265889DY_{(2017)_t} = 0.039250 - 0.265889(-59.65) = 15.9 \\
 DY_{(2018)_t} &= Y_{2018_t} - Y_{2017_t} = 15.9 \Rightarrow Y_{2018_t} = 15.9 + Y_{2017_t} = 15.9 + 89 = 104.9 \\
 DY_{(2018)_t} &= 0.039250 - 0.265889DY_{(2015)_t} = 0.039250 - 0.265889(-6.25) = 1.7 \\
 DY_{(2018)_t} &= Y_{2018_t} - Y_{2015_t} = 1.7 \Rightarrow Y_{2018_t} = 1.7 + Y_{2015_t} = 1.7 + 58 = 59.7 \\
 DY_{(2018)_t} &= 0.039250 - 0.265889DY_{(2013)_t} = 0.039250 - 0.265889(75.9) = -20.14 \\
 DY_{(2018)_t} &= Y_{2018_t} - Y_{2013_t} = -20.14 \Rightarrow Y_{2018_t} = -20.14 + Y_{2013_t} = -20.14 + 76.1 = 55.96 \\
 DY_{(2018)_t} &= 0.039250 - 0.265889DY_{(2011)_t} = 0.039250 - 0.265889(107.5) = -285.8 \\
 DY_{(2018)_t} &= Y_{2018_t} - Y_{2011_t} = -285.8 \Rightarrow Y_{2018_t} = -285.8 + Y_{2011_t} = -285.8 + 107 = -21.2 \\
 DY_{(2018)_t} &= 0.039250 - 0.265889DY_{(2011)_t} = 0.039250 - 0.265889(598.22) = -159.02 \\
 DY_{(2018)_t} &= Y_{2018_t} - Y_{2011_t} = -159.02 \Rightarrow Y_{2018_t} = -159.02 + Y_{2011_t} = -159.02 + 183 = 23.98
 \end{aligned}$$

Figure 9: Results of Forecasting

In the following table we summarize the results

Table 2: Summary of the Results

Months	Forecasted values	Observed values
1	104.9	105.5
2	59.7	62.4
3	55.96	57.6
4	21.2	20.7
5	23.98	18.9

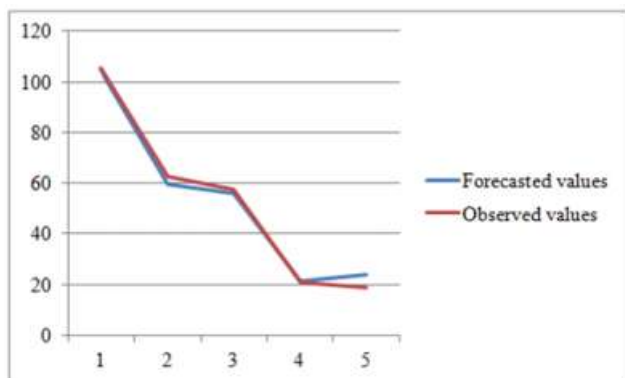


Figure 10: Observed and Forecasted Time Series Data

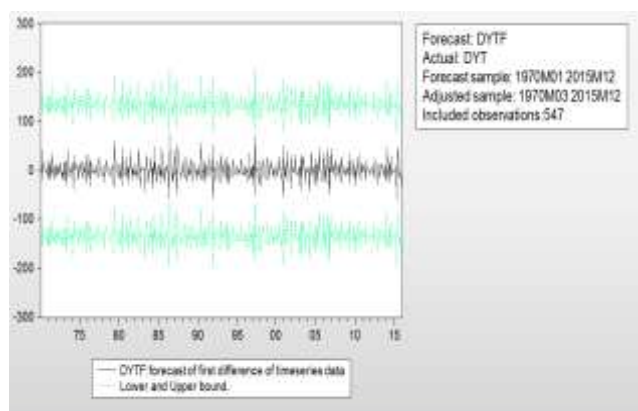


Figure 11: Plot of Model Fitting

7.9: Discussion of the Findings

The forecast was done for 5 months using the best model from historical data. It is observed that the forecasted data follows the observed data very closely. Therefore the selected best model from ARIMA building approach using a time series data of monthly rainfall from 1970 to 2015 can be used for forecasting droughts. We can conclude that the model is valid for forecasting of monthly rainfall. Adding explanatory variables to the ARIMA model could

enhance the accuracy of the model for forecasting droughts. The proposed explanatory variable, monthly rainfall is not difficult to obtain and is feasible to be incorporated in a predictive model.

8. Conclusions and Recommendations of the study

8.1 Conclusion

ARIMA model offers a good technique for predicting the magnitude of any variable. Its strength lies in the fact that the method is suitable for any time series with any pattern of change and it does not require the forecaster to choose a priori the value of any parameter. Its limitations include its requirement of a long time series. Like any other method, this technique also does not guarantee perfect forecasts. Nevertheless, it can be successfully used for forecasting long time series data. In our study the developed model for monthly rainfall was found to be ARIMA (1, 1, 0). The model can be used by researchers for forecasting drought in Bugesera. However, it should be updated from time to time with incorporation of current data. The main objective of this research is to develop the statistical model that forecast the occurrence of droughts in Bugesera District, Rwanda and use it to forecast future droughts in Bugesera District, Rwanda. Region mostly exposed to rainfall deficits which have been determined from information on dates of beginning and ends of rainy seasons and from the analysis of short droughts (dry spells) data. The ARIMA generator has proven to be relatively portable across different systems, provide a good source of practically strong random data on most systems. Using the computational analysis of this ARIMA approaches, better and accurate results were obtained from the no-stationary monthly rainfall. From the finding of this study, it showed that ARIMA (1, 1, 0) was found to be the best model as it produced the lowest error value. Therefore, the overall results of this study suggested that the model can give a better statistical technique. This study finished by finding the good model that can be used by researchers for forecasting occurrence of drought in Bugesera, Rwanda. The obtained results show that the observed and forecasted monthly rainfall has the strong relation without significant error. Means that the formulated model to forecast the monthly rainfall is fitting the used data adequately. In general, it can be concluded that whether Bugesera is in for a drought or much less rains than normal.

8.2 Recommendations

This study was focused on forecasting drought using monthly rainfall as a drought indicator. The time series models have proved to be a powerful tool in forecasting. MIDMAR and REMA today could use the statistical models for different interest without losing much money for going to the field every time for investigations or survey. There are several actions, in addition to what is already being done. While some research regarding climate change in Bugesera has been conducted, more such research is needed in order to better understand the interplay between increasing air temperatures (Climate

change) and precipitation (the ultimate source of all the state's water supplies).

References

- [1] A.K, M. (2005). *Steam flow Drought time series forecasting*.
- [2] Anderson, O. (1976). *Time sries analysis and Forecasting*. Butterworth 182.p.
- [3] Bhayan. (2004). *Objective quantification of Drought Severity and Duration*.
- [4] Box, G. J. (1976). *Timeseries analysis:forecasting and control,2edn.,Holden-Day*. San Francisco.
- [5] center, N. D. (2005). *What is Drought*. Retrieved from <http://www.drought.unl.edu/whatis/concept.htm>
- [6] Chatfield. (2003). *The analysis of time series*.
- [7] Chemonics International Inc. (2003). *Rwanda Environmental Threats And Opportunities Assessment*. No. 818 under the Biodiversity & Sustainable Forestry (BIOFOR) IQC USAID Contract No. LAG-I-00-99-00014-00.
- [8] Cristina F. (2008). *Drought forecasting using stochastic model*.
- [9] David, P. (1990). *Timeseries Forecasting*. New York,NY,USA: Springer.
- [10] DFID. (2009). *A citation of government Report*. Kigali,Rwanda: DFID.
- [11]Dickey, D. (1979). Distribution of Estimators for Time Series Regressions with Unit Root. *Journal of American Statistical Association* , 74,427-31.
- [12] Durdu, O. (2010). *Application of linear Stochastic models for Drought forecasting in Buyuk Menderes river basin*.. Western Turkey.
- [13]F, C. (2008). *Drought forecasting using stochasting model*.
- [14] Gardner. (1985). *Agriculture in the Yarra Catchment,Discussion Paper No.5,Yarracare Waterway Series*. Melbourne,Australia.
- [15] Hay, C. M. (1980). *The relationship of Drought frequency and duration to time scales*.
- [16]Inc, C. I. (2003). *Rwanda Environmental Threats And Opportunities Assessment*. Task order No. 818 under the Biodiversity & Sustainable Forestry (BIOFOR) IQC USAID Contract No. LAG-I-00-99-00014-00.
- [17]K., M. a. (June 21,2005). *Steam flow Drought timeseries forecasting*.
- [18]Mishra, a. A. (2005). *Drought forecasting using stochastic models*.
- [19]Obasi, G. (1994). "WMO'S Role in the international Decade for Natural Disaster Reduction". Bulletin of the American Meteorological society 75(1).
- [20]Obasi, G.O.P. (1994). *Role in the international Decade for Natural Disaster Reduction*. New York: American Society 75(1).
- [21]Palmer. (1965). *Metreological Drought*. Washington D.C: Department of Commerce Weather Bureau.
- [22]Pankratz. (1983). *Esimating the dimension of a model*.
- [23]Perez, E. P. (1995). *Natural Hazards:Drought*.
- [24]Perez,E.and P.Thompson. (1995). *Natural Hazards: Drought*.
- [25]R.Nagarajan. (2004). *Drought vulnerability Assessment Using Geoinformatics*. Bombay: CSRE.India Institute of Technology.
- [26]Salas et al, .. (2003). *Time series forecasting*.
- [27]Salas et al, C. (Time series forecasting). 1980;2003.
- [28]Salas et al., C. (2003). *Time series forecasting*.
- [29]Scoones, I. w. (1996). *Hazards and Opportunities. Farming Livelihoods in Dryland Africa:lessons from Zimbabwe*. London: Zed Books, 267p.
- [30]Smakhtin., V. D. (2004). *Review Automated Estimation and Aalyses of Drought Indices in South asia.W.P.83*. Colombo,SriLanka: International Water Management Institute.
- [31]Steams, H. T. (1967). *Geology of the Hawaiian Islands*.
- [32]Thompson, P. P. (1995). *Natural Hzards Drought*.
- [33]Thyer, M. (2001). *Modeling long-term persistence in hydrological time series,Ph.D Thesis*.. Australia: University of Newcastle.
- [34]WFP/FEWS-NET. (2003). *Rwanda Vulnerability Baseline Report*. Kigali: World Food Programme of the United(WFP).
- [35]Wilhite, D. A. (1985). *Understanding the Drought Phenomenon: the role of*. London: Water International, No.10, pp.111-120.