Available online at www.worldscientificnews.com

World Scientific News

WSN 24 (2015) 8-17 EISSN 2392-2192

Statistical Analysis of Some Meteorological Variables Data for Sokoto and its Vicinity

Z. Hassan[1] **and T. B. Ajibola**[2]

[1]Department of Physics, Adamu Augie College of Education Argungu, Kebbi State, Nigeria

[2]Department of Physics, University of Ilorin, Ilorin, Kwara State, Nigeria

ABSTRACT

This study aims at studying the characteristics of some meteorological variables in Sokoto and its vicinity using probability distribution models. The thirty one (31) years data (i.e 1980-2010) were collected at Nigerian Meteorological Agency (Nimet) Oshodi and the data were subjected to various probability distribution analyses in order to resolute the best fit probability functions for each meteorological variables. The variables measured consist of Relative humidity, Rainfall, Temperature, Sunshine hours, Solar radiation, wind speed and Evaporation pitche. Whereas the probability distribution models adopt were Normal, Gumbel, Pearson type III and Log-Pearson type III distribution functions. Numerical equation were recognized and used to forecast the variables. Goodness of fit tests such as chi-square, correlation coefficient and coefficient of determination were carried out to obtain the reliability of the forecasted values. The model that satisfies the statistical tests conditions mostly was selected as the best fit model. The study revealed that Rainfall, wind speed, evaporation pitche are best fitted by Log-pearson type III probability distribution model, whereas the Relative humidity, solar radiation, sunshine hours and temperature the best model is Gumbel probability distribution.

***Keywords*:** Meteorological variables, Gumbel, Normal, Pearson and Log-pearson type III, some meteorological variables, relative humidity, solar radiation, Statistical Analysis

1. INTRODUCTION

In frequency analysis, an assumed probability distribution was fitted to the accessible data to approximate the extent matching to return periods and the appropriate distribution models that represent the data are chosen. The choice of the probability distribution model was almost random as no physical basis was accessible to reduce the use of any particular function and the search for the proper distribution function has been the subject of several studies (Warren et al, 1972, Viessman et al, 1989, Wilson, 1990 and Salami, I.T, 2009). Gary and Robert (1971) studied the normal, log-normal, square-root-normal and cube-root-normal frequency distributions of meteorological data for Texas. The results of this research shows that the precipitation data conform to the square-root-normal distribution, while evaporation and temperature data conform to all of the frequency distributions tested. The evaporation, temperature and precipitation data were further fitted to the Gumbel extreme-value and to the log-Pearson type III distributions. Precipitation data fit the log-Pearson type III distribution more adequately than the Gumbel distribution, while both the evaporation and temperature data conform very well to Gumbel distribution. Ogunlela (2001) in a stochastic analysis of rainfall event in Ilorin used probability distribution function. He concluded that the log-Pearson type III distribution best described the daily rainfall data while the normal distribution best characterizes the monthly rainfall for Ilorin.

The effect of Asa dam on meteorological variables in Asa river basin using time series analysis was studied by (Salami, Bankole, 2003). The researchers divided the period into three different stages which includes; the pre- dam period, dam construction period and post dam period. In their study they concluded that rainfall, evaporation, and relative humidity increased considerably during the post- dam period compare to the pre-dam and dam construction period, while variables such as temperature and wind speed decreased during the post dam compared to pre-dam and construction period. Salami (2007) fitted various probability distribution models to the annual mean values of the meteorological variables (rainfall, temperature, sunshine, humidity, evaporation and wind speed) at Jebba hydropower station to evaluate the model that was most appropriate for the prediction of these variables. Gumbel probability distribution models best fit humidity, temperature, evaporation and sunshine, while log-Pearson distribution model best fit rainfall and wind speed. Salami and Yusuf (2009) studied the application of probability distribution models on studying the characteristics of meteorological variables in Ibadan and environs; they concluded that the best probability distribution obtained for the rainfall, wind speed and sunshine hours can best be fitted by log-Pearson type III distribution, which occupies 50% of the variables, while log-Gumbel, log-Normal and Pearson type III respectively best for relative humidity, evaporation and temperature. They suggested that similar studies may be conducted for other areas of Nigeria to obtain more conductive conclusions.

2. DATA AND ANALYSIS

The meteorological variables were obtained at the Nigerian Meteorological Agency (Nimet), Oshodi. The data collected include rainfall, relative humidity, evaporation pitche, sunshine hour, solar radiation,temperature, and wind speed. A total of 31 years data (1980 – 2010) were collected.

The summaries of statistics for the meteorological variables are presented in Table 1. The data are ranked according to welbull's plotting position and the corresponding return periods were estimated. The ranked data were evaluated with four methods of probability distribution functions to determine the best – fit functions. The methods include; Gumbel (EVI type1), Normal (N), Log-pearson type III (LP_3) and Pearson type III (P) probability distribution models. Four statistical goodness of fit test were used for the selection of the best fit models.

Table 1. Summary of statistics for meteorological variables (1980 – 2010).

Climate Variables	Parameters					
	Mean Value, x	Standard Deviation, σ	Skewness Coefficient	Coefficient of Variation	Maximum Values	Minimum Values
Evaporation Pitche **(m**m)	12.76	2.96	–0.64	0.23	17.59	3.93
Rainfall (mm)	57.31	12.19	1.00	0.21	95.56	31.10
Relative Humidity (%)	42.86	2.52	–0.60	0.06	46.42	37.33
Sunshine Hour (hr)	8.03	0.75	–2.40	0.09	8.80	4.98
Solar Radiation	16.40	0.80	–1.30	0.05	17.59	13.85
Wind-Speed (m/s)	7.80	0.92	–0.26	0.12	9.61	5.69
Temp. (°C)	35.29	0.60	–0.51	0.02	36.38	33.94

Probability distribution analysis was carried out in accordance with standard procedure (warren et al, 1972); viessman et al, (1989); Mustapha and Yusuf (1999) and Topaloglu (2002)]. The mathematical expression obtained for various functions are presented in Table 2. The mathematical expressions obtained for each function were used to predict the variables based on the estimated returned periods and were also used in performing the statistical tests (goodness of fit tests) for the selection of the best fit models.

Table 2. Mathematical expression for probability distribution models.

Climate variables	Distributions	
	Normal	Gumbel
Evaporation Pitche	$Ep = 12.76 + 2.96*K$	$Ep = 12.76 + 2.96*(0.78Y_T - 0.45)$
Rainfall (mm)	$Rp = 57.31 + 12.19*K$	$R_P = 57.31 + 12.19*(0.78Y_T - 0.45)$
Relative Humidity(%)	$R_{HP} = 42.86 + 2.52*K'$	$R_{HP} = 42.86 + 2.52*(0.78Y_T - 0.45)$
Sunshine Hour (hr)	$SH_P = 8.03 + 0.75*K$	$SH_P = 8.03 + 0.75*(0.78Y_T - 0.45)$
Solar Radiation (ml)	$SR_P = 16.40 + 0.80*K$	$SR_P = 16.40 + 0.80*(0.78Y_T - 0.45)$
Wind-Speed (m/s)	$WS_p = 7.80 + 0.92*K$	$WS_p = 7.80 + 0.92*(0.78Y_T - 0.45)$
Temp. (°C)	$Tp = 35.29 + 0.60*K$	$Tp = 35.29 + 0.60*(0.78Y_T - 0.45)$

Table 2. Continue

Pearson type III	**Log Pearson type III**
Ep = 12.76 + 2.96*K'	Log Ep = 1.09+ 0.12* K'
Rp = 57.31+ 12.19*K'	Log Rp = 1.75 + 0.09* K'
R_{HP} = 42.86 + 2.52*K'	Log R_{HP} = 1.63+ 0.03*K'
SH_P = 8.03 + 0.75*K'	Log SH_P = 0.90+ 0.05*K'
SR_P = 16.40 + 0.80*K'	Log SR_P = 1.21 + 0.02*K'
WS_p = 7.80 + 0.92*K'	Log WS_p = 0.89 + 0.05*K'
Tp = 35.29 + 0.60*K'	Log Tp = 1.55 + 0.01*K'

Table 3. Summary of the statistical test score results for each meteorological variables.

Climate Variables	Best Fit Distribution Model	Chi-Square Test X^2	PPCC (r)	R^2	Total
Ep (Mm)	Normal	1	1	1	03
	Gumbel (EVI)	3	4	4	11
	Pearson	2	2	2	06
	Log Pearson	4	3	3	10
Rainfall (mm)	Normal	2	2	2	06
	Gumbel	3	4	4	11
	Pearson	1	1	1	03
	Log Pearson	4	3	3	10
RH (%)	Normal	2	3	3	08
	Gumbel	3	4	4	11
	Pearson	1	1	1	03
	Log Pearson	4	2	2	08
SH (hr)	Normal	1	1	2	04
	Gumbel	3	4	4	11
	Pearson	2	3	3	08
	Log Pearson	4	2	1	07
Solar Rad	Normal	1	1	2	04
	Gumbel	3	4	4	11
	Pearson	2	3	3	08
	Log Pearson	4	2	1	07
W-Speed (m/s)	Normal	2	2	3	07
	Gumbel	3	4	4	11
	Pearson	1	1	1	03
	Log Pearson	4	3	2	09
Temp.(^{0}C)	Normal	2	2	3	07
	Gumbel	3	4	4	11
	Pearson	1	3	2	06
	Log Pearson	4	1	1	06

In order to determine the best-fit model(s) at each station. Probability distribution models were subjected to four (4) statistical tests (goodness of fit tests). The statistical tests include chi-square (X^2), fisher's (F) test, probability plot coefficient of correlation (r) and coefficient of determination (R^2).

The statistical tests were carried out in accordance with standard procedure (Chowdhury and Stedinger (1991); Adegboye and Ipinyomi (1995); Dibike and Solomatine (1999); Murray and Larry (2000)). The assessment of the probability distribution models was based on the total test score obtained from all the tests. Test scores ranging from zero to six (0-6) is awarded to each distribution model based on the criteria that the distribution (s) with the highest total score is or are chosen as the best distribution model (s) for the data of a particular variable. In general, the distribution best supported by a test is awarded a score of six (6), the next is awarded five (5), and so on in descending order.

A distribution is awarded a zero (0) score for a test if the test indicates that there is a significant difference between the variable estimated by the distribution model and the observed variables. The overall ranking results are presented in Table 3, while the best fit probability distribution models are presented in Table 4.

Table 4. Goodness of fit tests and the selected models for each variable.

Climate Variable	Best Fit Model	Second Best Fit Model	Best Fit Model Equation
Evaporation (mm)	Log Pearson III	Gumbel (EVI)	Ep = 1.09 + 0.12*K'
Rainfall (mm)	Log Pearson III	Gumbel (EVI)	Rp = 1.75 + 0.09*K'
R/Humidity (%)	Gumbel (EVI)	Log Prearson III	$RH_P = 42.86 + 2.52*(0.78Y_T – 045)$
S/Hours (hr)	Gumbel (EVI)	Log Pearson III	$SH_P = 8.03+ 0.75*(0.78Y_T – 045)$
S/Radiation	Gumbel (EVI)	Log Pearson III	$SR_P = 16.40 + 0.80*(0.78Y_T – 045)$
Wind Speed (m/s)	Gumbel (EVI)	Log Pearson III	$WS_P = 7.80 + 0.90*(0.78Y_T – 045)$
Temp. (°C) max	Gumbel (EVI)	Log Pearson III	$Tp = 35.29 +0.60*(0.78Y_T – 045)$

2. 1. Graphical comparison of the observed and predicted value using best fit models.

In order to compare the observed data with the predicted data, both the observed and the predicted data based on the best probability distribution models for each variable are plotted and presented in Figures 1 – 7. The graphical comparison show very close relationship between the observed and the predicted values, hence graphs confirmed the appropriateness of the selected models.

Figure 1. Relationship between observed and Log-Pearson distribution model predicted evaporation pitche data

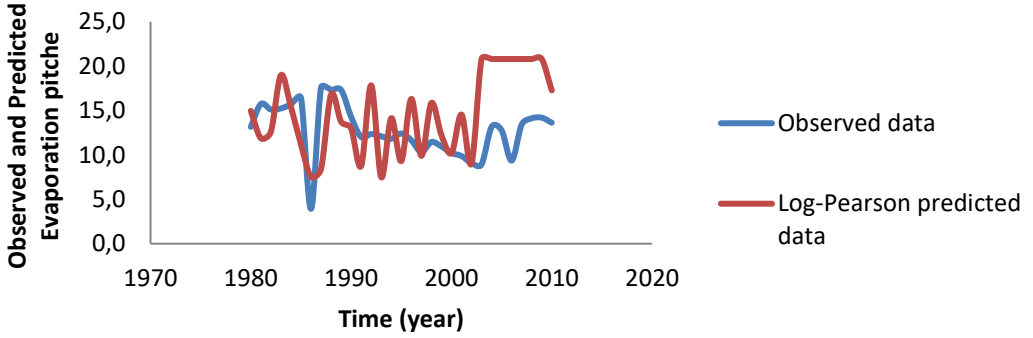


Figure 2: Relationship between observed and Log-Pearson distribution model predicted rainfall data

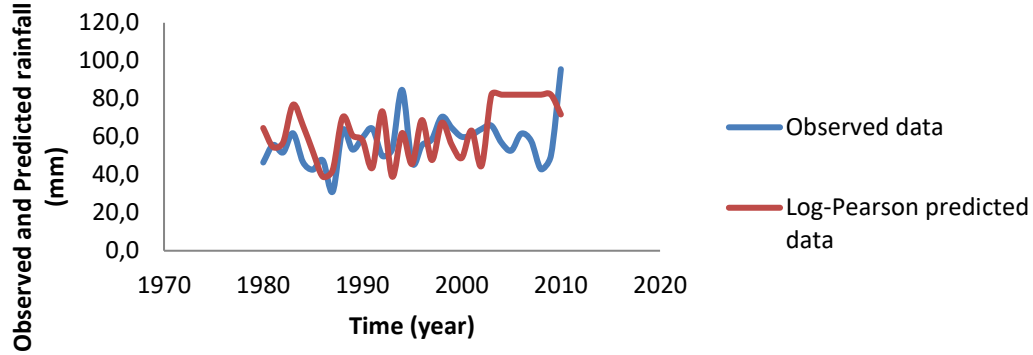


Figure 3: Relationship between observed and Gumbel distribution model predicted Relative humidity data

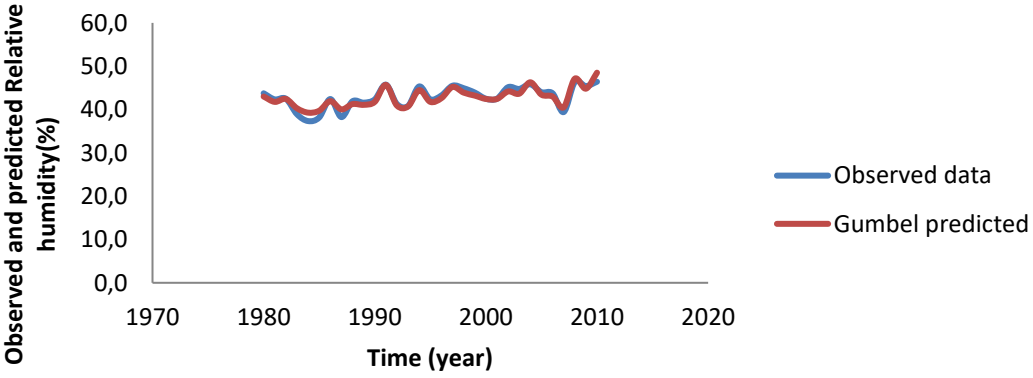


Figure 4: Relationship between observed and Gumbel distribution model predicted Sunshine hours data

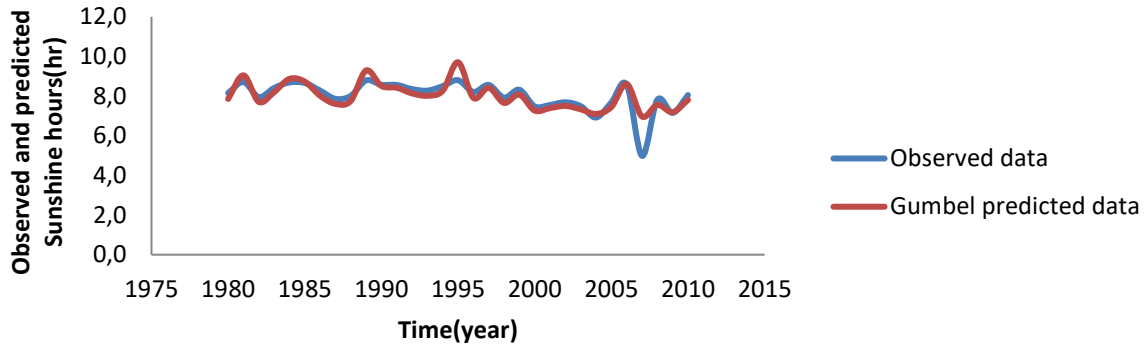


Figure 5: Relationship between observed and Gumbel distribution model predicted Solar radiation data

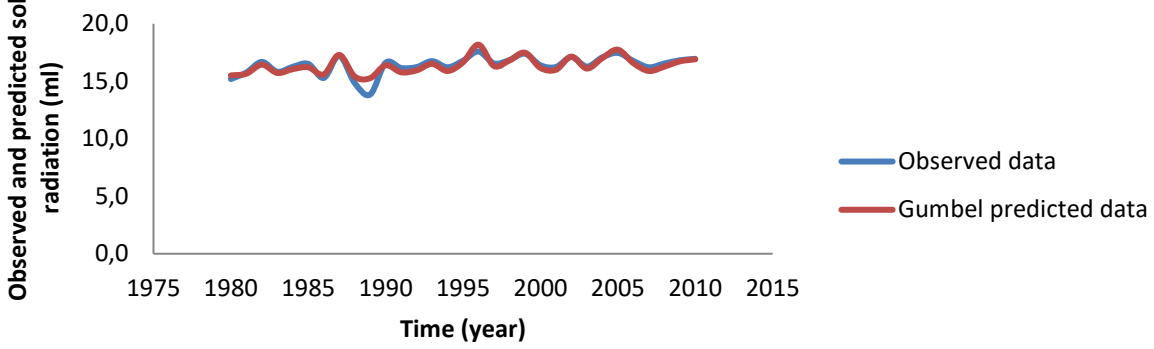
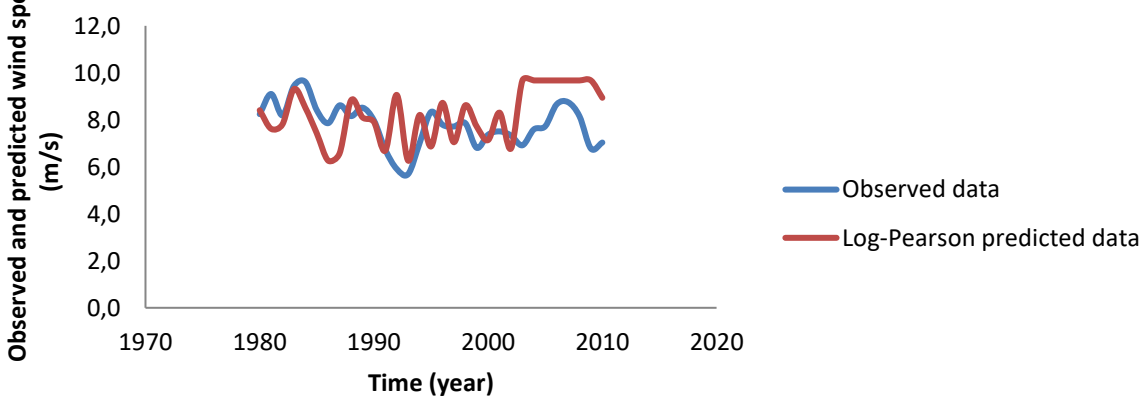
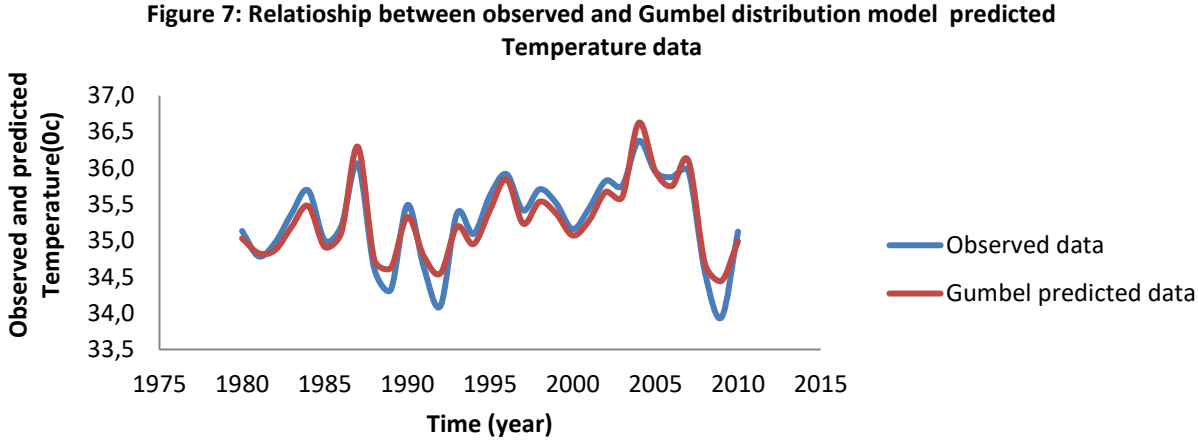


Figure 6: Relationship between observed and Log-Pearson distribution model predicted wind speed data





3. RESULTS AND DISCUSSION

The meteorological variables were obtained for a period of thirty one years (1980 – 2010) for Sokoto City in Nigeria and the average values were predictable. The data were evaluated with various probability distribution functions to determine the best fitting model; the summary statistic was presented in Table 1, while the mathematical representations of the evaluated probability functions were presented in Table 2. For the purpose of theoretical determination of best fit probability function, statistical tools (goodness of the fit tests) were adopted. The results of the statistical test score (the goodness of fit tests) and the best fit models were presented in Table 3 and 4 respectively. The relationship between the observed and predicted value based on the selected best probability distribution for each variable are presented in Figures 1 – 7. The charts help to confirm the appropriateness of the model selected as the best fit.

The statistical test score results at each station presented in Table 3 was used to decide which of the probability model (s) best fit the variables. Examination of the goodness-of-fit test results reveals that in many cases there was very little difference between the various distributions for each variable. Furthermore, the goodness of fit assessment for all the seven variables also indicated that no one distribution ranked consistently best of all the variables. However, the overall ranks for four meteorological variables, (relative humidity, temperature, solar radiation, and sunshine hour) combined show that Gumbel was best to describe them. While, the overall ranks for rainfall, wind speed and evaporation pitche show that Log-Pearson type III was the best for their prediction.

4. CONCLUSION

From the results of four frequency distributions applied in this study, it suggest that the best probability distribution obtained for the relative humidity, temperature, solar radiation, and sunshine hour can best be fitted by Gumbel distribution, which occupies 57% of the variables, while Log-Pearson type III best for for rainfall, wind speed and evaporation pitche. The outcome was relied on the results of three goodness-of-fit tests, Chi-Square, Correlation

Coefficient, Coefficient of Determination. The proposed assessment procedure has been successfully used to identify the best probability distributions that could provide accurate values for the considered meteorological variables for Sokoto, Nigeria. Since this study is focused on Sokoto, the results are just applicable to this study site. As for the factors affecting the probability distribution of meteorological variables for the entire Nigeria, it is suggested that similar studies may be conducted for other areas of Nigeria to obtain more conductive conclusions.

References

[1] Adeboye O.S. and Ipinyomi R.A. (1995). Statistical tables for class work and examination. Tertiary publications Nigeria Limited, Ilorin, Nigeria, pp. 5-11

[2] Chowdhury J.U. and Stedinger J.R. (1991). Goodness of fit tests for regional generalized extreme value flood distributions. *Water resources. Res.* 27(7): 1765-1776.

[3] Dibike B.Y. and Solomatine D.P. (1999). "River flow forecasting using Artificial Neural Networks". Paper presented at European Geophysical Society (EGS) XXIV General Assembly. The Hague. The Netherlands, 1-11.

[4] Gary E.O. and Robert A.C. (1971). "Applications of Climatology and Meteorology to Hydrologic Simulation". Technical Report No. 38, Texas Water Resources Institute, Texas

[5] Mustpha S. and Yusuf M.I. (1999). "A Textbook of Hydrology and Water Resources", Jenas Prints & Publishing Co., Abuja, Nigeria, pp. 164-184, First edition.

[6] Murray R.S. and Larry J.S. (2000). "Theory and problems of statistics" Tata Mc Graw – Hill Publishing Company Limited, New Delhi, pp. 314-316, Third edition.

[7] Ogunlela A.O (2001) "Storchastic Analysis of Rainfall Events in Ilorin, Nigeria". *Journal of Agricultural Research and Development*, 1(2): 39-50.

[8] Salami A.W and Bankole O.O. (2003). Impact of Asa dam on meteorological variables in Asa River Basin Ilorin, Nigeria. Technical paper presented to the department of Civil Engineering, University of Ilorin, Ilorin. Nigria.

[9] Salami A.W. (2007). Probability distribution analysis for hydrometeorological variables at Jebba hydropower station. Accept for publication in Nigeria Journal of Science and Technical Research.

[10] Salami A.W. and Yusuf I.T. (2009). Application of Probability Distribution Models on Studying the Characteristic of Meteorological Variables in Ibadan and its Environs. *Journal of engineering and Technology*, 4; 46-55.

[11] Topaloglu F. (2002). "Determining Suitable Probability Distribution models for flow and precipitation Series of the Seyhan River Basin". Turk Journal of Agric. 23: 189-194.

[12] Viessman W., Krapp J.W. and Harbough, T.E. (1989). "Introduction to hydrology". Harper and Row Publishers Inc., New York, pp. 675-695, Third edition.

[13] Warren V., Terence E.H. and John W.K. (1972). "Introduction to hydrology" Educational Publishers, New York, pp. 106-141, Second edition.

[14] Wilson E.M. (1990). Engineering Hydrology. Macmillan Press Ltd. 3rd edition. Houndmills, Basing Stockes Hampshire and London.

(Received 08 October 2015; accepted 22 October 2015)