iJET EXPLORE International Journal of Engineering & Technology
CENTRE OF PROFESSIONAL RESEARCH PUBLICATIONS
ISSN 2049-3444

International Journal of Engineering and Technology Volume 3 No. 2, February, 2013

Evaluation of Best Fit Probability Distribution Models for the Prediction of Rainfall and Runoff Volume (Case Study Tagwai Dam, Minna-Nigeria)

Busari Afis Olumide[1,3], Mohammed Saidu[2,3], Ajibola Oluwasesan[4]

[1]Coastal and Hydraulic Engineering, Department of Civil & Environmental Engineering,
Hong Kong Polytechnic University, Kowloon, Hong Kong
Department of Environmental Engineering, Faculty of Civil engineering
Universiti Teknologi, Malaysia (UTM)
[3]Water Resources & Environmental Engineering, Department of Civil Engineering,
Federal University of Technology, Minna, Niger State, Nigeria
[4]Department of Civil Engineering
University College, London, UK

ABSTRACT

For the significance of monitoring hydrometric data and subsequent prediction, various probability distribution models were fitted to various rainfall and runoff for the Tagwai dam site in Minna, Niger State, Nigeria. This is to evaluate the model that is best suitable for the prediction of their values and subsequently using the best model to predict for both the expected yearly maximum daily-rainfall and yearly maximum daily-runoff at some specific return periods. Many models were evaluated for the rainfall runoff and the best model was selected based on the statistical goodness of the fit test. The Normal distribution model was found most appropriate for the prediction of yearly maximum daily-rainfall of 131.21mm and the Log-Gumbel distribution model was the most appropriate for the prediction of yearly maximum daily-runoff of 1124.73 m^3/s. Based on this, the Normal distribution and the Log-Gumbel distribution are good for the estimation of rainfall and as well as runoff volume respectively for the Tagwai dam site. Hence, the establishment of these models has made the prediction of both the rainfall and runoff volume for Tagwai dam possible.

Keywords: *Probability distribution models, predicted rainfall, runoff, Tagwai dam*

1. INTRODUCTION

This paper is aimed at developing probability distribution models for both rainfall and runoff volumes for Tagwai Dam in Minna, Niger State of Nigeria using four probability distribution models namely Normal, log-Normal, Gumbel and log-Gumbel. These models were tested using Chi-Square (X^2) test, probability plot correlation coefficient (r) and coefficient of determination (R^2) so as to determine the probability distribution model that best fit the dam and which could be recommended for the prediction of future occurrence. With the statistical test a model that is most suitable for the prediction of future occurrence of the dam site could be selected. The prediction accuracy of rainfall-runoff models depends primarily on the quality of rainfall input. It is normally assumed that point rainfall measurements obtained at rain gauge locations are representative of the true rainfall field, and thus they are widely used in modelling applications.

The study area is the Tagwai basin with catchment Area of approximately 110km^2, located within Chanchaga local government area of Niger State (see Figure 1). The Tagwai dam is located along Chanchaga road with the road named after the dam (Tagwai dam road). The dam has a design capacity of 28.7MCM with a height of 24m, length of embankment is 1.80 km, it is a zone earth filled dam. The dam was commissioned in November 1978. The dam mainly serves Minna, Paiko, Badegi water works. Data were gotten from the Niger State Water Board Area office in Minna, Upper Niger River Basin development Authority, Minna and the Nigerian Meteorological Agency (NIMET) office, Minna International Airport.

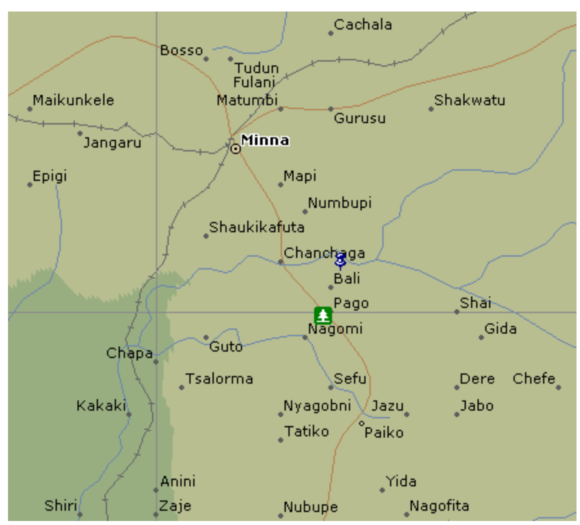

Figure 1: Chanchaga basin location map

2. MODEL SIGNIFICANCE TO RIVER BASIN DEVELOPMENT

Estimating the rate of rainfall as well as quantity of runoff volume at the natural catchment area at a certain/particular projected time is very necessary. The significance of rainfall and runoff for proper design and planning of hydraulic structure and water or hydrological based schemes necessitates a need for developing models which can be used for estimating future rainfall and runoff volume occurrence and that the best distribution model be adopted for the analysis to predict the most accurate future occurrence data (Kroll and Vogel, 2002 and Castellarin, 2007). In order to attain the maximum information from the observed records and to evaluate the most possible nature about the corresponding data, hence the search for adequate probability distribution function has been found to be the subject of several studies, which are aimed at obtaining the model that best fit the observed rainfall and runoff volume. Nevertheless, it is known that rainfall can vary significantly between the rain gauge locations. (Cicioni et al, 1993, and Connell and Pearson, 2001) showed that runoff predictions can vary by up to 100% between rain gauge models and spatially correct models, depending on the storm track. For this reason, the above assumption can lead to inaccuracies in model prediction. This will ultimately result in problems when the model is used in design.

3. METHODS: COLLECTION OF DATA, ANALYSIS AND STATISTICAL TEST

3.1 Data Collection

The hydrological data of the study area were obtained from the Niger State Water Board Area Office in Minna, Upper Niger River Basin Development Authority, Minna and the Nigerian Meteorological Agency (NIMET) at Minna International Airport.

3.2 Data Analysis

For the purpose of this paper, the rational method and the Flood-frequency technique has been used. The rational method was used to calculate the runoff using equation (1) and the Flood-frequency studies have been used for the frequency analysis of the values of the daily depth-duration data of the rainfall (hydrologic data series) for Tagwai dam catchment area.

Rational method was adopted for the runoff estimation using a rainfall of uniform intensity and very long duration occurring over the Tagwai dam basin. As t_c is time of concentration, it is obvious that if the rainfall continues beyond t_c, the runoff will be constant and at the peak value. The peak value of the runoff using rational method is given by (Kjeldsen et al 2002 and Lee et al 2007):

$$Q_p = C\,A\,i\,;\ for\ t \geq t_c \quad (1)$$

where C is the coefficient of runoff,
A is the catchment area of the dam and
i is the rainfall intensity.

Applying the rational method, the inflow design assumes a return period of 100 yrs and rainfall duration of 4 hours, and a runoff coefficient of 0.5 because of the flat large area of infiltration and percolation that will occur.

The depth-duration rainfall data from 1971 to 2009 of Minna, Niger State was obtained from the NIMET. And the average depth and duration was obtained for each month. The values maximum value of the average depth and duration was used to obtain a depth-duration curve.

Another data of the yearly maximum daily-rainfall was obtained from the Upper Niger River Basin Development Authority, Minna. The values of the yearly maximum daily-rainfall were traced out on the depth-duration curve to obtain the equivalent duration of the rainfall.

The analytical tools that was adopted in the flood-frequency studies for this research work is the probability distribution function which allows for the understanding of the behavioural pattern of both rainfall and runoff volume and to provide a basis for future forecasting. The methods employed include Gumbel, log Gumbel, Normal and log normal distribution. The procedure involves the ranking of the events by the strutting from the highest value and so on in descending order. The reoccurrence interval otherwise known as the return period represented by T_r is stated equation (5).

The extreme value distribution introduced by Gumbel is commonly known as Gumbel's distribution was used. According to this theory of extreme events, the probability of occurrence of an event equal to or larger than a value X_0 is

$$P\,(X \geq X_0) = 1 - e^{-e^{-y}} \quad (2)$$

e here is the natural logarithm and y is a dimensionless variable

$$Yt = -ln\left(-ln\left(1 - \frac{1}{T}\right)\right) \quad (3)$$

$$Yt = -ln\big(-ln(1 - P)\big) \quad (4)$$

T is the return period

$$Tr = \frac{n+1}{m} \quad (5)$$

The event X of return period is now defined as

$$Q_t = Q_{av} + Q\ (0.78Y_t - 0.45) \quad (6)$$

where Q_{av} is the average of all values Q in the standard deviation of the series and Y is the reduced variant (Madson et al, 1997).

Log Gumbel distribution also follows the same step as Gumbel distribution but the runoff volume and rainfall have to be e converted to logarithm value and an event Q having return period T_r yrs is described by log Gumbel model with a general equation of the form;

$$LogQ_1\ LogQ_{av} + \sigma\ LogQ\ (0.78Y_t - 0.45) \quad (7)$$

Where $Log\ Q_{av}$ is the average of all values of Q,

σ LogQ is the standard deviation of the series and y is the reduced variant.

The normal probability distribution is another known statistical model and often the most frequently used. The normal distribution model is for random variable with Parameter mean (u) and Standard deviation (S). The general formula for it is given as

$$Q_t = Q_{av} + K\ \sigma \quad (8)$$

Also, the Log normal probability distribution model was also adopted for random variable with parameter mean (u) and standard deviation (S). It has a general formula of the form (Castellarin et al, 2005).

$$Log\ Q_t = Log\ Q + K\ \sigma Log\ Q \quad (9)$$

where Q and σlogQ are the respective mean and standard deviation of the logarithms of the variables. The Parameter K in the equation (8) and (9) can be selected for a particular probability or return period from a table in U.S Water resources Council and other literatures.

3.3 Statistical (Goodness of fit) Test

The acceptability and reliability of the fitting (probability distribution) models were tested using the three statistical tests often known as goodness of fit test. In order to ascertain whether the data is significantly correct from the theoretical distribution, fit test are required (Zhang and Hall, 2004). The statistical tools include: Chi-square (X^2), probability plot correlation coefficient (r) and coefficient of determination (R^2).The Goodness of fit test was carried out in accordance to standard procedure (Onos and Bayazit, 1995, and Vogel et al, 2007). The observed variables and estimated variables were both used to determine how well both the correlation coefficient and coefficient of determination best fit the models, the values of both lies between -1 and 1. The closer the results are to 1, the higher the chances of being accepted. This Chi-square test was also used for the best-fit models. If the value of the ratio of the calculated Chi-square to the value of the Chi-square in Table i.e. $\left(\frac{X2\ cal}{X2\ tab}\right)$ gives a value close to 1, then we have the best probability distribution model for Tagwai dam.

4. RESULTS AND DISCUSSION OF RESULTS

4.1 Results

The year maximum daily rainfall was determined and the maximum runoff from the depth -duration curve using the rational method. The results so obtained from the rainfall and runoff for each probability distribution models are presented in Table 1 and Table 2 respectively.

Table 1: Model parameters for maximum daily rainfall

Parameter	**GUMBEL**	**LOG-GUMBEL**
Average	88.80	1.94
Std. Deviation	18.20	0.10
Skew	-0.52	-0.87
Cor. Coeff. (r)	0.89	1.06
Coef. Det. (R^2)	0.88	0.77
Chi-Square (X^2)	13.96	21.51
Standard Error (Se)	5.71	0.0395
Parameter	**NORMAL**	**LOG-NORMAL**
Average	88.80	1.94
Std. Deviation	18.20	0.10
Skew	-0.52	-0.87
Cor. Coeff. (r)	0.93	1.05
Coef. Det. (R^2)	0.96	0.90
Chi-Square (X^2)	4.97	9.25
Standard Error (Se)	3.43	0.0269

Table 2: Model parameters for runoff

	GUMBEL	**LOG-GUMBEL**
Average	650.01	2.80
Std. Deviation	138.46	0.08
Skew	1.80	1.49
Cor. coef. (r)	0.87	0.96
Coef. Det. (R^2)	0.88	0.99
Chi-Square (X^2)	89.00	74.19
Standard Error (Se)	42.04	0.0197
	NORMAL	**LOG-NORMAL**
Average	650.01	2.80
Std. Deviation	138.46	0.08
Skew	1.80	1.49
Cor. coef. (r)	0.91	0.78
Coef. Det. (R^2)	0.75	0.81
Chi-Square (X^2)	201.45	136.08
Standard Error (Se)	63.88	0.0321

The adopted equations for the probability distribution models are presented in Table 3.

Table 3: Equations for probability distribution models

Model	Yearly maximum daily-rainfall Yearly maximum daily-runoff
Gumbel	$R_p = 80.61 + 14.20Y_t$ Q_p = 587.703 +108Yt
Log Gumbel	R_p = Antilog (1.895 + 0.078Y_t) Q_p = Antilog (2.764 + 0.0624Y_t)
Normal	R_p = 88.80 + 18.20K Q_p = 650.01 + 138.46K
Log Normal	R_p = Antilog(1.94 + 0.10K) Q_p = Antilog (2.80 + 0.08K)

The equation that relates the estimated values with observed values in the statistical test is given. The goodness of fit test was performed for the entire probability distribution model for the dam site. The result for the best fit model for the Tagwai dam site is presented Table 4.

Table 4: Output of best fit model

PARAMETERS	BEST FIT MODEL	X^2	R^2	r
Yearly Maximum daily-rainfall	Normal distribution	4.97	0.96	0.93
Yearly Maximum daily-runoff	Log Gumbel distribution	74.19	0.99	0.96

From Table 3, the best fit model for Tagwai dam yearly maximum daily-rainfall is the Normal Distribution model with equation given as $R_p = 88.80 + 18.20K$ while that of the yearly maximum daily-runoff is the Log Gumbel Distribution Model with equation given as Q_p = Antilog $(2.764 + 0.0624Y_t)$.

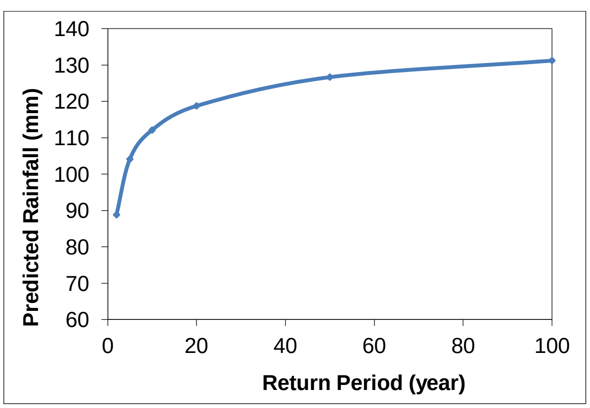

Figure 2: Predicted frequency curve for maximum rainfall

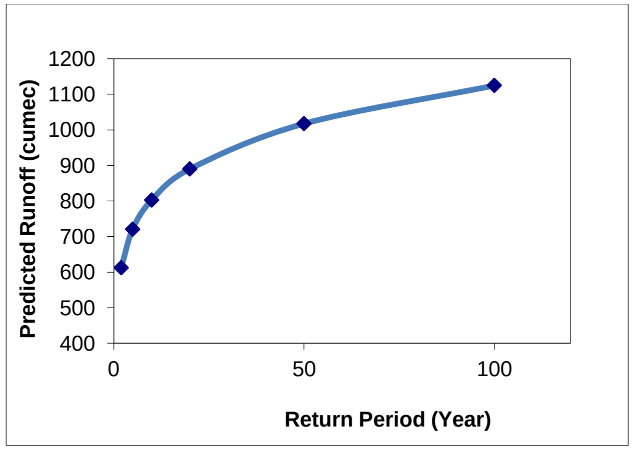

Figure 3: Predicted frequency curve for maximum runoff

4.2 Discussion of Results

The yearly maximum daily-rainfall for the dam site has values of (X^2 cal), R^2, r, and Se equals 4.97, 0.96, 0.93 and 3.43 respectively for the Normal distribution. It can be seen clearly from the result that the model gives a value close to 1 for the correlation coefficient (r), which shows that there is a close linearity between the observed and predicted rainfall. Also, the coefficient of determination (R^2) gives a higher value close to 1 when compared with other models and which implies that the model is the strongest.

The yearly maximum daily-runoff of the dam site has values of (X^2 cal), R^2, r, and Se as 74.19, 0.99, 0.96 and 0.0197 respectively for the Log Gumbel distribution. It can be seen clearly from the result that the model gives a value close to 1 for the correlation coefficient (r), which shows that there is also a close linearity between the observed and predicted runoff. Similarly R^2 gives a higher value close to 1 when compared to other R^2 of other models and which all implies that the model is the strongest. Also, checking the graph and comparing it with the table for yearly maximum daily-rainfall (Normal distribution) in Table 3, the predicted rainfall was 131.206mm while the maximum observed rainfall was 118.60mm, therefore the model is absolutely satisfactory. The predicted maximum runoff has a value of 1124.733m^3/s while the maximum observed runoff has a value of 1081.80m^3/s which still conforms to standard.

Since the objective of this research work is to develop and determine theoretically best probability distribution models using statistical test and also the prediction of rainfall and runoff at various return periods based on the selected best fit model for the dam site of this project work. The result of this statistical test i.e. the Chi-square test (X^2), correlation coefficient (r), coefficient of determination (R^2) and standard error (Se) was adopted in the selection of the best fit probability distribution model used for the prediction of both the yearly maximum daily-rainfall and runoff for the dam site and is presented in table 4. Also, the equation of the best fit probability model which form the basis of the frequency curves for the respective rainfall and runoff for the dam site are also presented in Figures 2 and 3.

5. CONCLUSION

Various probability distribution models were fitted to various rainfall and runoff data of Tagwai dam site in Minna, Niger State to evaluate the model that is best suitable for the prediction of values and subsequently using the best model to predict for both the expected yearly maximum daily-rainfall and yearly maximum daily-runoff at some specified return period.

Many models were developed for the rainfall and runoff and the best model was selected based on the statistical goodness of fit test.

The Normal distribution model was found to be most appropriate for the prediction of the yearly maximum daily-rainfall while the Log-Gumbel distribution model was also found to be the most appropriate for the prediction of yearly maximum daily-runoff. The establishment of the best fit probability distribution model would be a useful guide in the prediction of various rainfall and runoff volume for dam site.

ACKNOWLEDGEMENTS

The research work reported was carried out at Federal University of Technology, Minna, Nigeria. The authors are grateful to the Niger State Water Board (NSWB) Minna and Upper Niger River Basin Development Authority (UNRBDA), Minna for releasing data for this research.

REFERENCES

[1] Castellarin, A. (2007). Probabilistic envelope curves for design flood estimation at ungauged sites. Water Resources Research, 43.

[2] Castellarin, A., Vogel, R.M., and Matalas, N.C. (2005).Probabilistic behavior of a regional envelope curve. Water Resources Research, 41.

[3] Cicioni .G, Guiliano .G, and Spaziani F.M (1993). Best fitting of probability functions to a set of data for flood studies. Floods and droughts proc 2nd international symposium in Hydrology 11-13 September, 1993.

[4] Connell, R. J. and Pearson, C. P. (2001). Two-component extreme value distribution applied to Canterbury annual maximum flood peaks. Journal of Hydrology (NZ), 40(2):105–127.

[5] Kjeldsen, T. R., Smithers, J. C., and Schulze, R. E. (2002). Regional flood frequency analysis in the Kagouls – Natal province, South Africa, using the index flood method. J. Hydroa., 255:194–211.

[6] Kroll, C. N. and Vogel, R. M. (2002). Probability distribution of low streamflow series in the United States. Journal of Hydrologic Engineering, 7:137–146.

[7] Lee, H., Zehe, E. and Sivapalan, M. (2007). Predictions of rainfall-runoff response and soil moisture dynamics in a microscale catchment using the CREW model. Journal of hydrology and earth system science, Volume II, Page 819-849.

[8] Madsen, H., Pearson, C. P., and Rosbjerg, D. (1997). Comparison of Annual Maximum Series and Partial Duration Series Methods for Modeling Extreme Hydrologic Events. 2 Regional Modeling. Water Resour. Res., 33(4):759–769.

[9] Onos .A and Bayazit .B (1995). Best distributions of largest Available flood samples. Journal of hydrology (pp. 195-208).

[10] Vogel, R.M., Matalas, N.C., England Jr., J.F., and Castellarin, A. (2007). An assessment of exceedance probabilities of envelope curves, Water Resources Research, 43.

[11] Zhang Jingyia and Hall, M.J.(2004). Regional flood frequency analysis for the Gan-Ming River basin in China. Journal of hydrology. Volume 296, Page 98-117.