Proceedings of the Fourth International Conference on Engineering
and Technology Research February 23 - 25, 2016 ISBN: 978-2902-58-6 Volume 4

EVALUATION OF BEST FIT PROBABILITY DISTRIBUTION MODELS FOR THE PREDICTION OF INFLOWS INTO KAINJI RESERVOIR, NIGER STATE, NIGERIA

Mohammed J.Mamman.[1], Otache, Y, Mathins[2], Jibril Ibrahim[3], Mohammed I. S[4]
[1]Department of Agricultural Technology,
College of Agriculture Mokwa, Niger State Nigeria.
[2]Department of Agricultural and Bioresources Engineering,
Federal University of Technology, Minna, Niger State, Nigeria.
3Department of Agricultural Technology,
College of Agriculture Mokwa, Niger State Nigeria.
[4]Department of Agricultural and Bioresources Engineering,
Federal University of Technology, Minna, Niger State, Nigeria.
Email:ikwan1565@gmail.com

Abstract

The analysis of time series is essential for building mathematical models to generate synthetic hydrological records, to forecast hydrological events, to detect intrinsic stochastic characteristics of hydrologic variables as well as to fill predict of forecast record. To this end various probability distribution models were fitted to river inflows of Kainji Reservoir in New Bussa, Niger State, Nigeria. This is to evaluate the probability function that is best suitable for the prediction of their values and subsequently using the best model to predict for both the expected maximum and minimum monthly-inflows at some specific return periods. Three models i.e., Gumbel (EVI), log- Normal and Normal were evaluated for the inflows and the best model was selected based on the statistical goodness of fit test. The values of goodness of fit test for Kainji hydropower dam are $r = 0.96$, $R^2 = 0.99$, $Se = 0.0087$ $X^2 = 0.0054$, for Gumbel (EVI), $r = 0.79$, $R^2 = 0.85$, $Se = 0.02$, $X^2 = 0.31$ for Log- Normal, and $r = 0.0.75$, $R^2 = 0.0.68$, $Se = 0.056$, $X^2 = 1376.39$ for Normal. For the Kanji hydropower dams the Gumbel (EVI) model gave the best fit. These probability distribution models can be used to predict future reservoir inflow at the Kainji hydropower dam.

Keywords: *Probability distribution models, predicted inflow, Kainji dam, fit tests*

1. Introduction

River is a natural stream of water flowing in a channel to the sea or to the lake or joining another river, the area drained by a river and its tributaries is called river basin. There are two major rivers in Nigeria; the river Niger and river Benue. River Niger and its tributaries has great potential for the socio-economic transformation of the West African sub-region including Nigeria. Kainji hydropower dam is situated along the Niger River, a dam is a barrier built across a water body to hold back the flow of water, thereby creating a large body or pool of water called reservoir. Dams can be classified according to the purpose for which they are meant and also according to the materials used in constructing them, such as concrete, earth fill etc. The functions of dam include: regulation of the river flow, controlling of the water release in accordance to the electricity production or irrigation requirement. Despite the usefulness of dams, they usually have significant impact on economy, geology, environment, hydrology and meteorological variables.

The probability distribution is a hydrological tool most widely used in flood estimation and prediction. The importance of reservoir inflow analysis at any hydropower dam to our daily life makes it imperative that the appropriate probability distribution model be established to determine the discharge into the reservoir. Olukanni and Salami (2008), Larry and Murray (2000) stated that the choice of the probability distribution model is almost arbitrary as no physical basis is available to rationalise the use of any particular function. In general, the search for a proper distribution function has been the subject of several studies. Salami (2004) studied the flow along the Asa River and established probability distribution models for the prediction of the annual flow regime. For minimum and maximum flows, log-Pearson type III (LP3) and Gumbel extreme value type I (EVI) respectively were recommended. Salami (2002) considered flood levels at four gauging stations along the River Niger, below the Jebba hydropower dam. The maximum and minimum flood level data were fitted with four probability models and compared graphically with the observed data. The EVI type 1 distribution fit the data best and it was used to predict flood levels with return periods of 10, 50 and 100 years. Bayazit and Önöz (2000) dealt with the probability distribution of largest available flood sample with the aim of determining the distribution that best fit the observed flood. Olokanni and Salami (2008) reported that the Water Resources Council of the USA conducted a study with the objective of developing a uniform technique of determining flood frequency. The work applied the available methods to flood records at 10 stations in various parts of the USA. Record length varied and five methods were used, namely Gamma, EVI, log-Gumbel, log-normal (LN) and LP3 distributions. However, no statistical test was applied to determine the goodness of fit, instead flood discharge for various return periods (2 – 50 years) were obtained from the probability plot and compared with the corresponding values from the five hypothesized distributions. Among these methods, the LP3 distribution was preferred for common use, and for being capable of fitting skewed data. Salaudeen and Yusuff (2008) considered LN, LP3, EVI distributions for the flood data from 108 stations in Italy. Statistical tests such as chi-square (X^2), Kolmogrov-Smirnov (KS) and probability plot correlation coefficient (ppcc) were applied ~~and the~~ best fitting distribution was found to be LN by the chi-square test while EVI and LP3 were found to be the best by the other test. Olokani and Salami (2008) reported that 1000 years floods at 300 stations in the USA with four different models (LN, Gamma, log- Gumbel and LP3). LN and LP3 came close to reproducing the expected exceedences and were concluded to be the best. Vogel *et al.*, (1993) explored the suitability of various models applied to the flood flow data at 38 sites in the southwest USA. The probability distribution models adopted include N, LN, EVI type 1, and LP3, which were compared graphically with the observed data. Ajayi *et al.*, (2007) estimated the occurrence of flood events and its frequency at the lower Niger basin, Nigeria, using hydrological data including river discharges, runoff records and meteorological data from different gauging stations within the basin. The data collected were subjected to various statistical analyses and plotting position and probability distributions were determined. The results showed that various plotting positions and probability distributions could be used to fit the available discharge records of the River Niger. The EVI distribution was the best of the applied models for peak average reservoir inflow and peak discharge at the River Kaduna (Wuya Gauging Station). The LN distribution best predicted the peak runoff discharge of River Niger (Lokoja Gauging station) and peak discharge at Baro Gauging Station. The predicted models that compared favorably with the observed values are considered the best distribution models. Busari *et al.*, (2013) evaluated best fit probability distribution models for the prediction of rainfall and runoff volume in Tagwai dam, Minna-Nigeria, 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. According to Chowdhury and Stedinger (1991), various probability distribution models were fitted to the peak reservoir inflows and the suitable model was selected based on the goodness of fit tests. The Gumbel (EVI type 1) probability distribution model was found to be appropriate for Kainji. This study focuses on the evaluation of three methods of probability distribution analysis for the prediction of mean reservoir inflow at Kainji hydropower dam in Nigeria. This study could serve as a guide to institutions and dam managers in determining available flow that will generate maximum discharge for hydropower dams and prevent flood waters overtopping the dam, thereby causing subsequent release of flood wave and averting loss of life and properties (Binnie, 1981). The information can also be a valuable tool for flood forecasting.

Material and Methods

The Study Area

Geographically, Kainji hydroelectric dam is located in New Bussa town now headquarter of Borgu local government area of Niger State, Nigeria. The lake created behind the dam span between latitude 9° 8′ to 10° 7′ and between longitude 4° 5′ to 4° 7′ E with reference point 9.54 N and 4.38 E northwest of the Federal Capital Territory (FCT, Abuja) (Dukiya,2013).

Hydrology of the Niger River system

The average rainfall at the headwaters of Niandan and Milo rivers which is the source of the Niger at the Fouta Djallon Mountains in Guinea and its exit to the sea in Nigeria is 2200mm. The river flow regime is characterized by two distinct flood periods occurring annually namely the White and Black floods. The black flood derives its flow from the tributaries of the Niger outside Nigeria (flow lag October to May) and arrives at Kainji reservoir (Nigeria) in November and lasts until March at Jebba after attaining a peak rate of about 2,000m^3/sec in February (Oyebande *et al*, 1992). The White flood is a consequent of flows from local tributaries especially the Sokoto-Rima and Malendo river systems. The White flood is heavily laden with silts and other suspended particles (flow lag June to September) and arrives Kainji in August in the pre-Kainji Dam River Niger having attained a peak rate of 4,000 to 6,000m^3/sec in September-October in Jebba. The critical low flow period into the Kainji reservoir is March and July each year.

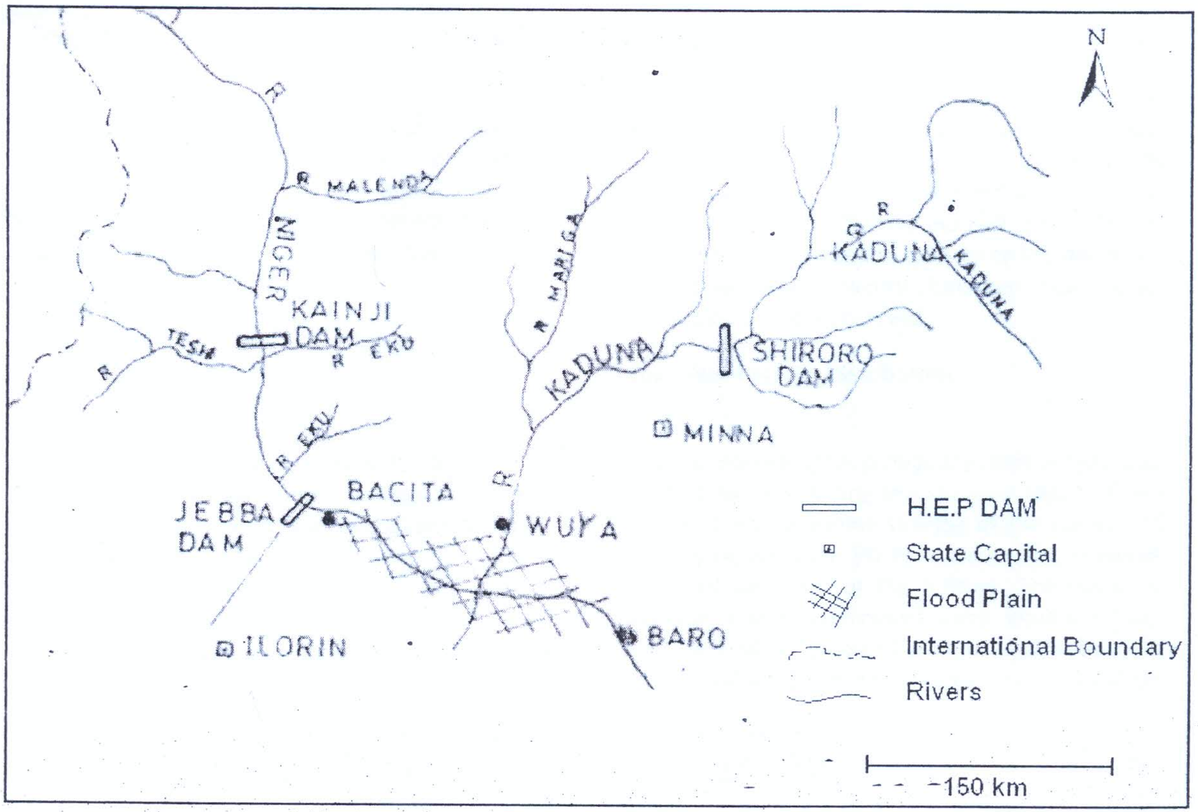

Fig.1: Location of Kainji Hydroelectric Dam.
Source: Salami: (2013)

Data Collection

The reservoir inflow data for a total of 25 years (1990 - 2014) were collected from the hydrological unit of Kainji hydropower station in Nigeria.

Data Analysis and Evaluation of Probability Distribution Models

The mean inflows data were evaluated with three methods of probability distribution to determine the best probability distribution function for the inflows. The methods adopted were; Gumbel Type 1 (EV1), log – normal (LN) and Normal distribution (N) respectively.

Gumbel (EVI Type 1)

According to Wilson (1990) and Salami (2004), the Gumbel (EV1) distribution model is based on the probability that any of the events would equal or exceed a particular value having return period (Tr) as given in equation below: (6).

The equation for determining reduced variate is given as in equation (1)

Y_T = - ln (- ln(1- p) (1)

where p is the probability of occurrence of event

The general form of the Gumbel distribution is given as in equation (2)

$Q_{T_r} = Q_{av} + \sigma(0.78Y_T - 0.45)$, (Yusuf and Salami, 2009). (2)

Where Q_{av} the average of all values of inflows, σ is the standard deviation of the series and Y_T is the reduced variate. The means and standard deviation of the inflows were determined by employing equation (3) and (4) respectively.

employed as stated below:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{N} x_i \qquad (3)$$

where $\bar{x}$ =The measure of central tendency n = the size of the sample, x is the observed parameter, Σ is the summation symbol and N is the number of observations.

$$\sigma_p = \frac{\sqrt{\Sigma(x-\mu)^2}}{n-1} \qquad (4)$$

σ_p= standard deviation, x = is the observed parameter, μ= mean, n = the size of the sample

The average monthly inflows were ranked based on the magnitude of the inflow and the return period computed employing equation (6) as in the works of Yusuf and Salami, (2009).

$$\text{Tr}=\frac{N+1}{m} \qquad (5)$$

where Tr is the return period, m is the series of events ranking and N is the number of observations in the series. The probability of occurrence of the inflow was computed using equation (6).

$$\text{P}=\frac{1}{Tr} \qquad (6)$$

The reduced variate (Y_T) was computed as in equation (7), this in accordance with works of Yusuf and Salami, (2009).

Y_T = -ln (- ln (1-p) (7)

Where Y_T is the reduced variate and P is the probability of occurrence of an event.

The values of the mean, standard deviation of the inflows obtained were substituted into the general Gumbel equation (2) to obtain a new Gumbel distribution (EV1) model equation (8) for the reservoir inflows. This equation is employed to simulate the inflows of the reservoir and used to determine its best fit.

$$Q_T = 817.14 + 675.28y_T \qquad (8)$$

The computed reduced variates Y_T obtained were then substituted into equation (8) to obtain the simulated values of inflow Q_T. This is to ascertain the fitness of the model and whether the data is significantly correct (Zhang and Hall, 2004). The R^2 and r values were used to determine the degree of correlation and linearity between the model prediction and the observed.

Log – Normal (LN) Distribution

The log-normal (LN) probability distribution was applied by first finding the seasonal mean of the inflow, then determine the log of the mean and employing equation (9) to evaluate the seasonal standard deviation of the inflows. The standard variable z was determined using equation (10), and the z values obtained were used to determine the K values from the probability distribution table.

$$\sigma = \sqrt{\Sigma\left(\frac{x-\bar{x}}{n-1}\right)^2} \qquad (9)$$

$$z = \frac{\log x - \log \mu}{\sigma} \qquad (10)$$

where logx is the logarithm of the inflows and logμ is the seasonal mean of the inflow (Zhou, 2000).The general relationship for the log- normal probability distribution (equation 11) was then employed as in the works of (Yusuf and Salami, 2009).

$$\text{Log}Q_T = \overline{logQ} + K\sigma_{logQ} \qquad (11)$$

Where $\overline{logQ}$ depicts the mean of the log of inflow.

The values of average mean of the log of the inflows and the standard deviation obtained were substituted into equation (11) and a new log-normal probability distribution model for the reservoir inflows obtained, equation (12)

$$\text{Log}\ Q_T = 2.8764 + 0.5384K_T \qquad (12)$$

The obtained values of K_T were then substituted into equation (12) to have the simulated values of the $\text{log}Q_T$. This was done to test the fitness of the log-normal distribution model.

Normal or Gaussian Distribution

The general relationship for the normal probability distribution is given as in (equation 13) (Busari, *et. al.*, 2013). The normal probability distribution was determined by first finding the averages of the reservoir inflows and then determining the standard variable (z) employing equation (14).

$$Q_t = Q_{AV} + K\sigma \qquad (13)$$

$$Z = \frac{x-\mu}{\sigma} \qquad (14)$$

The z values computed were used to obtain k values from normal probability distribution table. The values of the average inflow and standard deviation were substituted into equation (13) to obtain a new normal probability distribution model (equation 15) for the reservoir inflows

$$Q_t = 1206.72 + 865.7371K \qquad (15)$$

The obtained k values were then substituted in equation (15) to obtain the simulated values of the inflow after which the observed and the simulated values of the inflows were plotted to test the fitness of the developed normal probability distribution model.

Testing of the probability distribution models

The acceptability and reliability of the developed probability distribution models were tested by using statistical test (goodness of fit test); these are; chi – square (X^2); probability plot coefficient of correlation (r), coefficient of determination (R^2) and standard error of estimate (Se). The equations respectively are presented below:

Chi- square (X^2) test

The expression for the analysis of chi-square is as in equation (16).

$$X^2 = \sum_{j=1}^{N} \frac{(O_j - e_j)^2}{e_j} \qquad (16)$$

where O is observed flow; e is predicted flow, N is total frequency and level of confidence is 95%

Probability Plot Coefficient of Correlation (r)

Equation (17) was adopted for the estimation of the probability plot coefficient of correlation (r).

$$r = \pm\sqrt{\frac{\sum(Q_{est} - Q_{mean})^2}{\sum(Q_{obs} - Q_{mean})^2}} \qquad (17)$$

where Q_{est} is the value of inflow estimated with the probability function, Q_{mean} is the mean value of the observed inflow and Q_{obs} is the value of the observed inflow.

Coefficient of Determination (R^2)

According to Dibike and Solomatine (1999), the coefficient of determination (R^2) is given as in equation (18). This is to determine the strength between the observed inflow and the predicted inflows.

$$R^2 = \frac{E_O - E}{E_O} \qquad (18)$$

Where $E_o = \sum_{i=1}^{N}(Q_{i(obs)} - Q_{i(mean)})^2$,$E = \sum_{i}^{N}(Q_{i(obs)} - Q_{i(est)})^2$

$Qi_{(est)}$ is the model output in the ith time period, $Q_{i(obs)}$ is the observed data in the same period and $Q_{i(mean)}$ is the mean over the observed periods.

Standard Error of Estimate (Se).

The relationship in equation (20) was adopted in the estimation of standard error (Se)

$$Se = \sqrt{\frac{1-r^2}{N-2}} \qquad (19)$$

The reservoir inflow data were evaluated using various probability distribution functions to determine the best fitting model; the mathematical representations of the evaluated probability functions are presented in table 1. Also for the purpose of theoretical determination of best fit probability function, statistical tools (goodness of fit test) were adopted. The result of goodness of fit tests and best fit models are presented in table 2.

Figure 2, 3, and 4 compares the average monthly inflow of Kainji reservoir to the model predicted inflow of the reservoir. The distribution models are the extreme value type I (EVI), the log – normal, and Normal distribution. The inflow at Kainji reservoir station has values of X^2, R^2, r, Se as 0.0054, 0.99998, 0.95518 and 0.00876 respectively for Gumbel distribution. From this result, the value of the ratio of calculated chi- square to the table chi-square is less than one and model gives the value of 0.955518 for correlation coefficient (r), R^2 square value of 0.99998 and Se value of 0.00876 which shows that the model compare favourably and with stronger linearity between the observed and the predicted reservoir inflow. For the log-normal distribution the goodness of fit values obtained (table 2) compare less favourably between the observed and simulated values of the inflow but better than the Normal distribution. Also based on the graphical comparison (figure 2, 3.4) the Extreme value type I (EVI) distribution model is a better fit than the other probability distribution models. Hence, Extreme value type I (EVI), is the most appropriate model for the reservoir inflow at Kainji reservoir.

Table 1 Model Equations for the Probability Distributions

S/no	Hydropower Dam	Probability distribution	Developed equation
1	Kainji	Gumbel (EVI)	$Q_T = 817.118 + 675.28y_T$
2		Log – Normal	$Q_T = 2.8764 + 0.5384K_T$
3		Normal	$Q_T = 1206.718 + K_T865.7371$

Q_T is expected discharge associated with a particular probability of occurrence.
Y_T is the reduced variate.
K_T are selected from the normal distribution table

Table 2: Results of Goodness of Fit Tests and the Selected Best Fit Model for the Inflow

s/no	Best fit distribution model	X^2	R^2	r	Se
1	Gumbel	0.0054	0.99	0.96	0.00876
2	Log-normal	0.31	0.85	0.79	0.02
3	Normal	1376.39	0.68	0.75	0.056

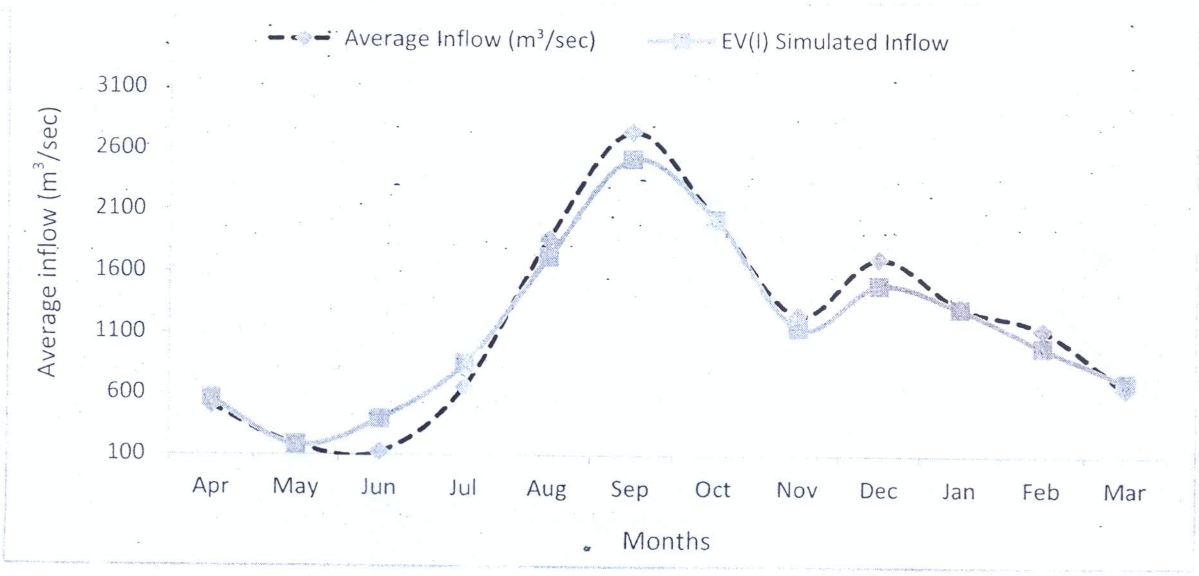

Fig. 2: Relationship between Average Observed Inflows and Gumbel (EVI) Simulated inflow

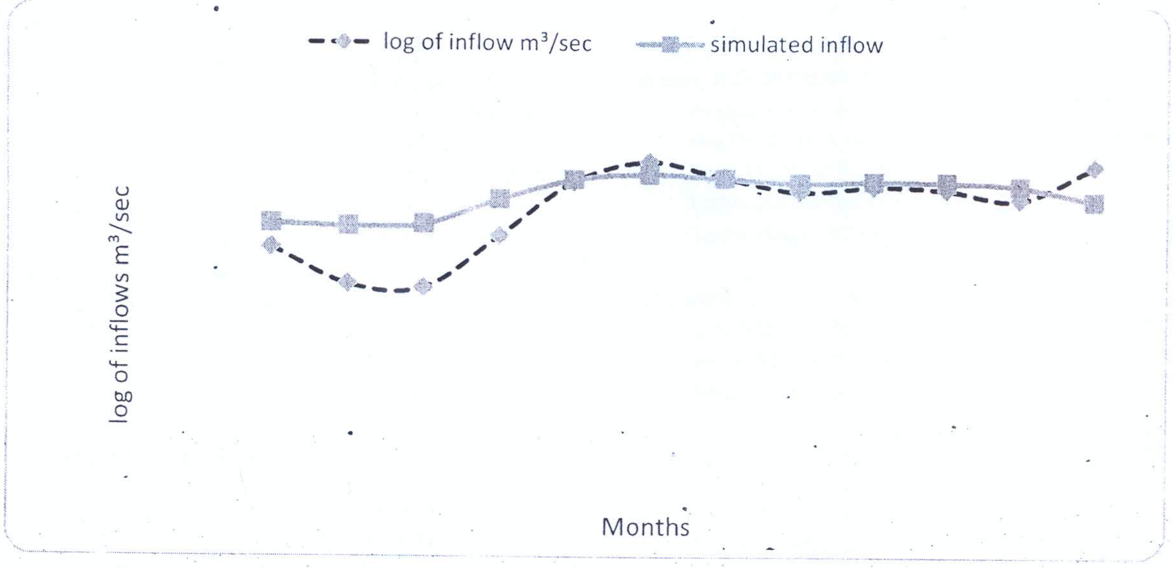

Fig. 3: Relationship between Observed Inflows and Log-normal Simulated Inflow

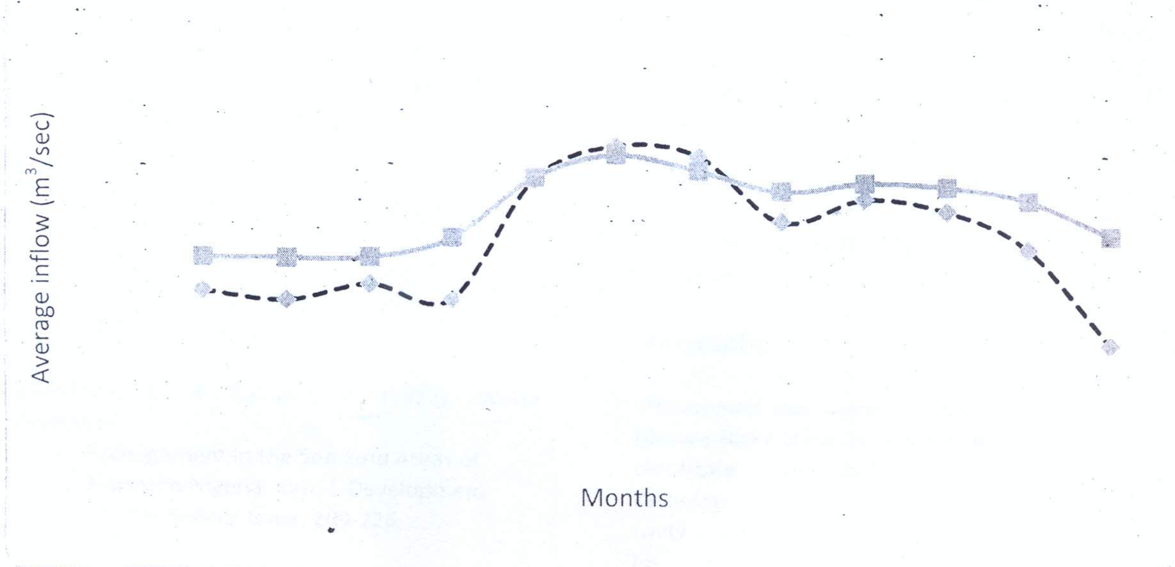

Fig. 4: Relationship between Observed Inflows and Normal Simulated Inflow

Conclusion

Various probability distribution models were fitted to the reservoir inflow records to evaluate the model that is most appropriate for prediction at Kainji hydropower station in Nigeria. Three models were established for the hydropower station and the most suitable model was selected based on the goodness of fit tests. The EVI model was found to be appropriate for Kainji Reservoir.

Recommendations

1. Based on the findings of this research it is recommended that EVI be adopted in the prediction of inflows at Kainji dam.
2. Other distribution functions could be evaluated to determine their fitness to the inflow.

References

Ajayi, A.E., A.A. Olufayo, and T.A. Affinnih. (2007): Historical study of flood events in the lower Niger River Basin, Nigeria. Geophysical Research Abstracts, 9(10883): 4-13.

Binnie, G.M. (1981). Early Victorian Water Engineers, Thomas Telford, London.

Busari, A.O,Mohammed, S. & Ajibola O.(2013). Evaluation of Best Fit Probability Models for the Prediction of Rainfall and Runoff Volume(Case Study of Tagwai Dam,Minna - Nigeria). International Journal of Engineering and Technology. 3(2):4-6

Chowdhury, J.U., and J.R Stedinger. (1991). Goodness of fit tests for regional generalized extreme value flood distributions. Water Resour. Res., 27(7): 1765 – 1776.

Dibike, B.Y., and D.P Solomatine (1999). River flow forecasting using Artificial Neural Networks. Paper Presented at European Geophysical Society (EGS) xxiv General Assembly. The Hague, The Netherlands, 1 – 11.

Dukiya, J.J (2013). Spatial Analysis of the Impacts of Kainji Hydropower Dam on the Down Stream Communities an Overview. A Journal of Geoinformatics and Geostatistics: Vol. 4(4):105-113,

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

Onoz .A and Bayazit .B (1995). Best distributions of largest Available flood samples. Journal of hydrology 167 (1):195-208).

Bayazit, M., and Önöz B. (2000). Conditional Distributions of Ideal Reservoir Storage Variables J. Hydrol. Eng. 5, 52–8.

Olukanni, D.O. and Salami, A.W. (2008). Fitting Probability Distribution Functions to Reservoir Inflow atHydropower Dams in Nigeria. Journal of Environmental Hydrology, 16 (1):6-9

Oyebande, L. & Balogun, I. (1992). Water Resources Management in the Semiarid Areas of Northern Nigeria. Can. J. Development Studies Special Issue, 209-226.

Salami, A.W. (2002). Fitting probability curves for flood levels at Jebba, Belle, Oshin and Yelwa gauging station, along River Niger. Technical Paper presented to the Department of Civil Engineering, University of Ilorin, Nigeria.

Salami, A.W. (2004). Prediction of the annual flow regime along Asa River using probability distribution models. AMSE periodical, Lyon, France. 65 (2), 41-56.

Salaudeen, F.A and Yusuff, O. K (2008). Comparison of Probability Distribution Models for the Prediction of Annual Flow Regime along River Niger and Benue in Nigeria. Journal of Research Information in Civil Engineering, 5(1): 21.

Vogel, R.M., W.O. Thomas, and T.A. Mahon. (1993). Flood flow frequency model selection in South Western United States, Journal of Water Resources Planning Management. ASCE 119 (3): 353 – 366.

Yusuf, K.O and Salami, A.W (2009) Influence of Jebba Hydropower Dam on Statistical Distribution Of Hydro - Meteorological Variables in Niger River Basin, Nigeria. Technical Paper presented to the Department of Civil Engineering, University of Ilorin, Nigeria, at 1st Annual Civil Engineering Conference (1): 34 - 39

Zhang Jingyia and Hall, M.J.(2004). Regional flood frequency analysis for the Gan-Ming River basin in China. Journal of hydrology. 296 (1): 98-117.

Zhou Y. (2000). Statistics and Stochastic Processes in Hydrology, HH327/00/1 (3): 14-26, Second Edition.

Biography.

Mohammed Jiya Mamman was born in 1979 in Mokwa Niger State, he got his first school leaving certificate from Government Primary School Kataeregi, Niger State, attended Government Unity school Afon Kwara State, where he obtained his O level certificate, in 1998. He had his degree in Agricultural and Bioresources Engineering in 2006 at Federal University of Technology Minna. He is currently undergoing his master's degree in soil and water Engineering at Federal University of Technology Minna. He is a lecturer serving in Niger State College of Agriculture Mokwa. He is a researcher in general Agricultural Engineering but specialised in Surface Hydrology. He has also published in research journals. He is married with children.