[1.]Oluwatosin O. OLOFINTOYE, [2.] Babatunde F. BAKARE

PROBABILITY DISTRIBUTION MODELS FOR ESTIMATING ANNUAL PEAK FLOW IN THE MKHOMAZI RIVER, SOUTH AFRICA

[1.]Department of Water Resources and Environmental Engineering, University of Ilorin, Ilorin, NIGERIA.
[2.]Department of Chemical Engineering, Mangosuthu University of Technology, Durban, SOUTH AFRICA

Abstract: Accurate computation of peak flow in rivers is vital in preventing environmental degradation. This study developed stochastic models for computing annual peak flows in the Mkhomazi River, South Africa. Six curves of selected functions (Normal, Log–Normal, Pearson III, Log–Pearson III, Gumbel and Log–Gumbel) were fitted to the annual duration series of the at–site annual peak flow of the river. Chi–square tests for goodness–of–fit were performed at 10% significance level to determine the appropriateness of all models in describing the observed data. The non–parametric significant differences in the performances of applicable models found was examined by applying the Kruskal–Wallis (H) test. A system of ranking based on aggregate of scores from statistical performance measures was adopted to suggest a most suitable model. The Normal, Pearson, Log–Pearson and Gumbel models were found suitable for application in the study area. These models are recommended for use in modeling peak flow in the river as they were found appropriate based on standard statistical test. Thus, the outcome of this work provides additional inputs for the improvement of hydrological researches in South Africa.
Keywords: annual peak flow, probability function, environmental management, mkhomazi river, test score

1. INTRODUCTION

Rivers are a major part of freshwater supply in the hydrologic cycle. While adequate flows in rivers are beneficial in terms of water supply, poorly managed high flows are detrimental to the environment. This often results in a myriad of environmental degradation problems like flooding and erosion.

Floods are catastrophic events characterized by the inundation of dry lands. These events are sometimes attributed to improper channeling of peak flows in river catchments (Loucks and Bee, 2005). Erosion on another hand is a severe environmental problem worldwide. This phenomenon leads to a series of soil degradation processes, such as nutrient loss, compaction, and a loss in infiltration and water storage capacity, which eventually results in loss of fertile arable land (Bernini et al., 2021). Environmental degradation problems resulting from mismanagement of peak flow in rivers are oftentimes accompanied by severe damages to infrastructures and may also result in loss of lives and livelihood in the floodplain (Sudmeier–Rieux et al., 2019; Bernini et al., 2021).

The calamitous impacts of environmental problems resulting from peak flows in rivers located in the U–drainage regions, (comprising Mvoti, Mgeni and Mkhomazi Rivers) in the KwaZulu–Natal province in South Africa have been reported in recent times. For example, Olofintoye et al. (2023) reports a disastrous flood that hit the province in the month of April of year 2022. The flood is considered the most catastrophic yet recorded in the region in terms of infrastructures destroyed, lives and livelihood lost, and economic losses. Incessant geo–environmental problems like gully erosion resulting in severe loss of fertile soil have also been reported (Bernini et al., 2021).

Increase in the frequency and severity of intense hydrological events is anticipated worldwide, and due to anthropogenic global warming, this trend is expected to continue into the future (Olofintoye et al., 2023). Therefore, it is expedient to establish river peak flow models for accurate computation of high flows to aid in effective environmental management and protection. Hence, developing flow models and propounding measures to mitigate the debilitating effects of mismanaged peak flow events has been the subject of several studies (Bilewu and Sule, 2015; Komolafe et al., 2021; Olofintoye et al., 2023).

Studies in search of accurate stochastic models for computing peak flows in rivers have been reported in the literature. For instance, Olofintoye et al. (2023) established best–fit stochastic models for predicting annual peak flows in the Mgeni River in South Africa. The study fitted six probability distribution curves to the annual duration series of peak river flow. It was found that the Log–Normal, Log–Pearson III and Pearson III models were suitable for peak flow computation along the river.

Kjeldsen et al. (2002) carried out regional frequency analyses of flood for relatively unregulated rivers in the KwaZulu–Natal province of South Africa. The study found that the Generalized Pareto, Pearson III and Normal distributions were suitable for modeling floods in the West and North–Western part of the province. However, no suitable regional model was found for the coastal and midlands area where the Mkhomazi River is situated. The occurrence of a few flood events of extreme magnitude at stations across this area suggests that the assumption of homogeneity made in regional frequency analysis may present

ISSN 1584 – 2665 (printed version); ISSN 2601 – 2332 (online); ISSN-L 1584 – 2665

gross oversimplification in some instances. Thus, it is expedient that at–site investigation of flow along the Mkhomazi River be explored as the river is well gauged and has historical record of considerable length.
This study analyzes the frequency distribution of peak annual flows in the Mkhomazi River in South Africa. It uses methods of statistical frequency analysis to develop models for computing peak flow in the river. Establishing applicable models for predicting peak flow in rivers is germane in articulating proactive measures towards addressing the challenges associated with environmental degradation in watersheds. Attaining a sustainable environment is in line with goal 11 of the sustainable development goals (SDGs) of the United Nations.

2. DESCRIPTION OF THE STUDY AREA

The Nation of South Africa is endowed with a good network of rivers that are vital in sustaining life and livelihood within her various ecosystems. Among these is the Mkhomazi River which is an important watercourse that represents a major source of freshwater supply.
The Mkomazi River (Figure 1) is the third largest river in the KwaZulu–Natal province of South Africa. It originates around the Thabana Ntlenyana, the highest mountain in southern Africa, in the Maloti mountain range that crests Lesotho, a small landlocked country within South Africa. The river flows in a southeastwards direction from a height of 3,300 metres above sea level to discharge into the Indian Ocean through a navigable estuary at Umkomaas located at 30°12′1″S, 30°48′4″E. The watercourse which is about 160 km long drains an area of about 4,400 square–kilometers. Major sources of livelihood and activities within the watershed include commercial afforestation, irrigation farming, paper production and tourism.
The climate shows high seasonal variability in which the winters are dry while rain falls during the summer. High rainfalls are experienced in the peak of the summer months (January and February) while minimum rainfalls are observed between June and July. The mean annual precipitation that ranges between 700 – 1200 mm produces high intra– and inter–seasonal River flows. Rainfalls are generally higher in the upstream than in the downstream sections of the river. Thus, a significant portion of the catchment runoff is generated in the upper part of the watershed. The mean annual temperature varies from 15°C in the highland areas around the source to 24°C in the central area of the watershed (Oyebode et al., 2014).

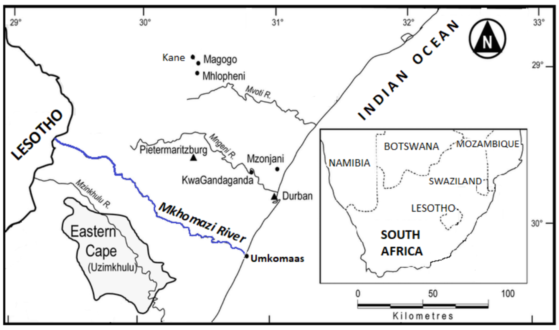

Figure 1: Location of River Mkhomazi in South Africa. Adapted from Greenfield et al. (2005)

3. METHODOLOGY

Six probability distribution curves were fitted to the annual maximum flow series of Mkhomazi River, South Africa. The Chi–square test was applied to validate the goodness–of–fit of the functions while the Kruskal–Wallis (H) test was applied to determine if there were substantial differences in the performances of models that were found appropriate. Statistical performance measures were evaluated to compute a test score statistics. The test score was further employed to suggest a model that is most suitable for application. The suitable models were used to estimate flows for different return periods. Graphs were also plotted to visualize the fitness of the curve models. Details on the procedures followed herein are elaborated on in the relevant sub–sections.

Data Collection, Cleaning and Statistics

Peak annual river flow data for Mkhomazi River was collected at the Department of Water and Sanitation (DWS), Pretoria, South Africa. This government department collects and manages hydrologic information of the study area. The data is measured in cubic metres per second. The data collected spanned Sixty–four (64) years (1960 – 2023). Check for outliers were carried out following the procedure outlined by the United States Water Resource Council (WRC) (Mays, 2011). Statistics of the cleaned and raw data were computed to estimate measures of dispersion, skewness and central tendency of the data.

Probability Function Fitting

Peak flows per year were selected to compose an annual duration series. This series was ranked using Weibull's plotting position and the return periods corresponding to the ranks were computed. The series was further evaluated using six probability functions viz. (Log–Normal (LN), Normal (N), Log–Pearson III (LP), Pearson III (P), Log–Gumbel (LG) and Gumbel (G)). These models have been applied successfully in recent hydrological researches in South Africa and have been recommended for application in all the

ISSN 1584 – 2665 (printed version); ISSN 2601 – 2332 (online); ISSN-L 1584 – 2665

watersheds in the country (Adeyemo and Olofintoye, 2014; Olofintoye et al., 2023). Flow frequency analysis and curve herein were performed in accordance to standard procedure (Viessman et al., 1989; Mustapha and Yusuf, 1999; Topaloglu, 2002; Reddy, 2014; Olofintoye et al., 2023).

Evaluating the Appropriateness of the Fitted Models

The applicability of the fitted distribution curves were inspected using the Chi–square (χ^2) test. This test checks whether the difference between theory and experiment is as a result of the inadequacy of the theory to fit the observed data or is due to chance. The null hypothesis, states that there is good fit between theory and experiment. If the critical (tabulated) value of the statistics is less than the calculated value (or the ratio $\chi^2_{Computed}/\chi^2_{Critical}$ exceeds unity), the null hypothesis is rejected in favour of the alternative at the level of significance (Gupta, 2013). A 10% significance level was adopted herein following the recommendations from previous studies (Haktanir et al., 2012; Adeyemo and Olofintoye, 2014; Olofintoye et al., 2023).

Checking for Discrepancy in the Performances of Models Found Appropriate

When two or more models are found appropriate for estimating peak flow (following a goodness–of–fit test), it is suggested that further test be carried out to determine if there are considerable discrepancy if the performances of the models. The Kruskal–Wallis (H) test was applied herein following the advice of Olofintoye et al. (2023). The test was performed at a 5% level of significance as suggested by Gupta (2013). Graphs were also plotted to aid in visual appraisal of the curve models.

Evaluation of Performance Measures and Scoring of Models

While it is accepted that there are rationale for the use of a specific model over others in a group of appropriate models, it is often desirable to suggest a model for use in computation. To suggest such, a system of scoring and ranking suggested in previous studies was adopted herein (Khudri and Sadia, 2013; Paul et al., 2014; Olofintoye et al., 2023).

The method entails ranking all models based on the aggregate score from a number of performance measures. Scoring of models herein is based on two performance measures viz. the Mean Relative Deviation (MRD) and Mean Square Relative Deviation (MSRD) statistics (Jou et al., 2009). These performance measure statistics have been applied in appraising the performance of river flow models in South Africa and have been recommended for use in all watersheds in the nation (Adeyemo and Olofintoye, 2014; Olofintoye et al., 2023). Lower values of these measures suggest better performance.

In the scheme, a function is ranked based on aggregate of scores obtained on all measures. Since six models are appraised, a highest score of 6 is awarded a function best supported by a measure while the least performing model scores one (1). A zero score is awarded on all measures where the goodness–of–fit test indicates that a model is not appropriate. A model that aggregates the highest total score is suggested as the most suitable model.

4. RESULTS

Models for computing peak annual flows in Mkhomazi River were evaluated. Goodness–of–fit test was performed and appraisals of models performances were made. Table 1 presents statistics of the raw and cleaned data. The result of the H–test for multiple model comparison is presented in Table 2. The evaluated model equations and results of performance measures, goodness–of–fit test and significance status of each model are presented in Table 3. The score obtained on each performance measure is emphasized in italics and enclosed in parentheses. Table 4 presents the annual peak flow quantiles computed using the models. Models exhibiting poor fits are not included in this table. Graphs of the models are presented in Figure 2.

Table 1: Summary of Statistics of Peak Annual Flows in the Mkhomazi River (1960 – 2023)

	Mean value $\bar{x}$, (m^3/s)	Median (m^3/s)	Standard deviation σ, (m^3/s)	Skewness coefficient, G	Minimum (m^3/s)	Maximum (m^3/s)
Raw Data	337.22	330.67	158.16	0.21	8.31	637.84
*Cleaned Data	342.44	334.95	153.77	0.29	59.51	637.84
*Outlier(s) found and removed.						

Table 2: Kruskal–Wallis (H) Test for Multiple Model Comparison

Number of Samples	Size of combined samples	Degree of Freedom	Hstat	χ2Critical	Significance Status
4	252	3	0.1814	7.8147	Not Significant
If ($H_{stat} > \chi^2_{Critical}$):– Test is significant.					

ISSN 1584 – 2665 (printed version); ISSN 2601 – 2332 (online); ISSN-L 1584 – 2665

Table 3: Summary of Developed Probability Distribution Models and Statistical Tests

Model	Model Equation	Performance Measure Statistics			Goodness–of–fit Test	
		MRD	MSRD	Total Score	$\frac{\chi^2_{Computed}}{\chi^2_{Critical}}$	Significant ?
N	Qp = 342.44 + 153.77K	6.33 (4)	155.97 (4)	8	0.0442	No
LN	Qp = Antilog (2.48 + 0.23K)	11.71 (0)	362.36 (0)	0	1.1745	Yes
P	Qp = 342.44 + 153.77K'	5.33 (6)	65.43 (5)	11	0.1762	No
LP	Qp = Antilog (2.48 + 0.23K')	5.75 (5)	63.44 (6)	11	0.2376	No
G	Qp = 273.24 + 119.94Y_T	11.57 (3)	348.28 (3)	6	0.8300	No
LG	Qp = Antilog (2.38 + 0.18Y_T)	22.52 (0)	1447.7 (0)	0	2.1782	Yes

Q_p – Peak flow estimated using a model
K – Normal variate obtained from Normal distribution tables for a given return period
K' – Variate obtained from Pearson distribution tables for a given return period and skew coefficient
Y_T – Reduced variate computed as $Y_T = -\ln[-\ln(1 - {}^1/_T)]$ for a given return period T
(*) – Test score on a specific test

Table 4: Annual Peak Flow Quantile Estimates for Return Periods (m^3/s)

Model	Recurrence interval in years						
	5	10	20	50	100	200	500
Normal	471.86	539.51	595.38	658.25	700.17	738.54	785.03
Pearson III	469.21	543.55	607.20	681.32	732.22	779.86	838.96
Log–Pearson III	479.36	559.78	624.14	692.09	733.79	768.94	807.30
Gumbel	453.15	543.16	629.50	741.26	825.00	908.44	1018.53

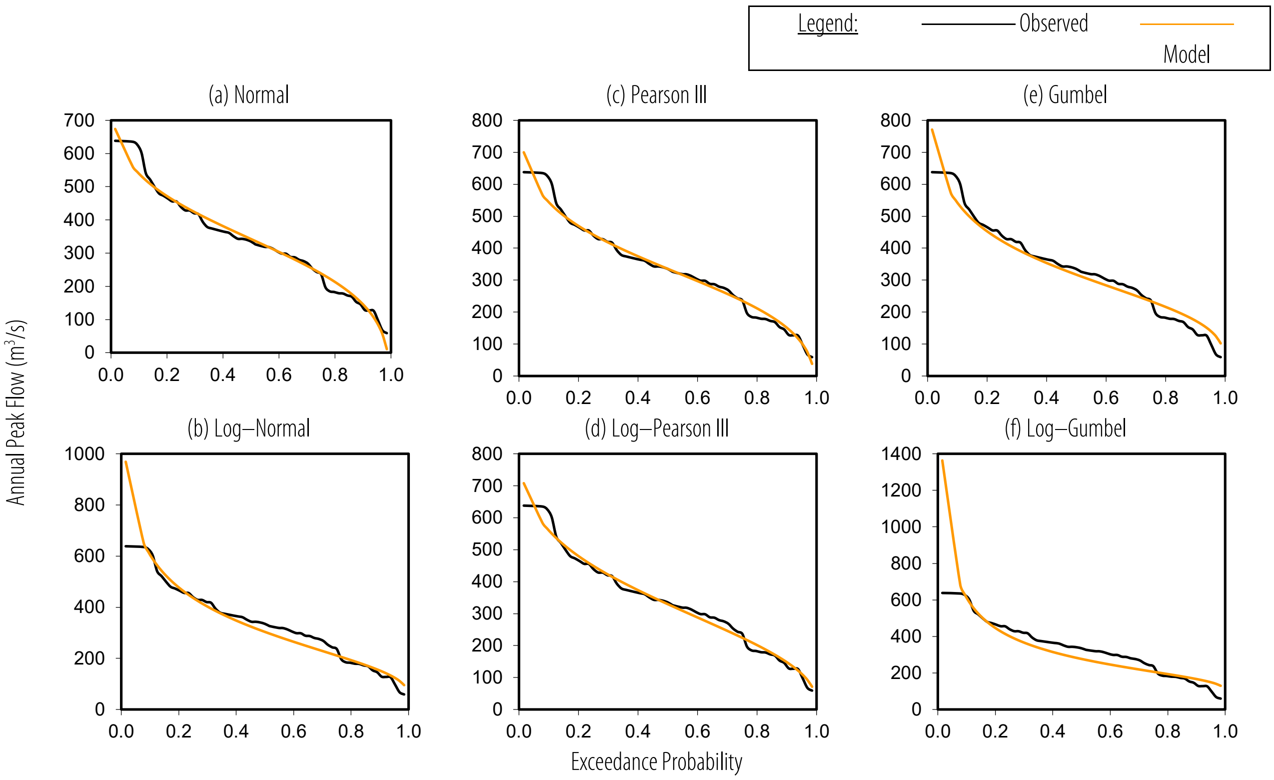

Figure 2: Graphical Plots of Models

5. DISCUSSION OF RESULTS

Analysis of statistics of raw and cleaned data of peak annual river flow in Table 1 shows that the mean values of the series are greater than the median and the skewness coefficients are positive. This shows that the distribution of the peak flows in rivers the region are positively skewed. This is in agreement with the observations of Olofintoye et al. (2023), Viessman et al. (1989) and other studies that have found that river flows are often positively skewed due in part to the impact of natural phenomena. A regional study by (Kjeldsen et al., 2002) noted the occurrence of a few flood events of extreme magnitude in the coastal and midlands area of Kwazulu–Natal. Hence, it is necessary to perform outlier check on the data. A low outlier (the minimum value of the raw data in Table 1) was identified and was thus removed. It is suggested that low outliers be removed from analysis of peak flows to avoid attenuations of high events (Mays, 2011). Result of the H–Test (Table 2) shows that there are no considerable differences in the performances of the models that were found appropriate for estimating peak flow in the river. Four models were found

ISSN 1584 – 2665 (printed version); ISSN 2601 – 2332 (online); ISSN-L 1584 – 2665

appropriate in this study, these are Normal, Pearson III, Log–Pearson III and Gumbel distributions (see Table 3 and Figure 2). The findings here are consistent with the findings of Olofintoye et al. (2023) that Pearson III and Log–Pearson III are suitable for estimating peak flows in River Mgeni which is situated around the same region in South Africa. However, while the previous study on Mgeni River found Log–Normal distribution appropriate, this study indicates that Log–Normal is not appropriate for flow in the Mkhomazi River. Further, while the Gumbel model is suggested as appropriate in this study, it was found not appropriate in the Mgeni catchment. These findings indicate that the coastal and midlands area in Kwazulu–Natal cannot be said to be truly homogeneous.

A regional frequency analyses of flood for relatively unregulated rivers in the KwaZulu–Natal province of South Africa was done by Kjeldsen et al. (2002). The study found that the Pearson III and Normal distributions were suitable for modeling floods in the West and North–Western part of the province. This is consistent with the findings in this study. However, no suitable regional model was found for the coastal and midlands area where the Mkhomazi River is situated. The data series across this region all have high coefficients of variation and coefficients of skew and all include high outliers. This further corroborates the fact that the region is not highly homogeneous. Thus, the assumption of homogeneity of the regions may present a gross over–simplification.

Kjeldsen et al. (2002) further indicated that the regional model in its current form is inappropriate for modeling floods in the region and the method should not be applied until fundamental modelling problems have been solved. Therefore, exploration of single site analysis where historical data of considerable length is available, as is the case in the Mkhomazi River, is expedient. In contrast to the previous regional study by Kjeldsen et al. (2002), four suitable models were found for modelling peak river flow in this study while applying the single site analysis. Therefore, at–site frequency analysis of river flow in South Africa, remains a subject open for further studies.

The Pearson and Log–Pearson models were identified as the best overall in this study (Table 3). Normal came next than Gumbel distribution. The Pearson III model is found consistent for peak flow modeling in all the studies discussed in this section and therefore may be generally adopted. This study found two new appropriate models (Log–Pearson and Gumbel) which were not included in the previous regional study mentioned. Although flows in rivers are often skewed, the Normal probability function was found appropriate this study. This corroborates the postulation of several studies (Viessman et al., 1989; Mustapha and Yusuf, 1999; Mays, 2011) that exists no rationale for the selection of a model over others but various functions should be investigated in frequency analysis of hydrological data. The quantile estimates in Table 4 are accurate and appropriate for estimating peak river flow in the Mkhomazi River.

6. CONCLUSION

The need for accurate peak flow studies cannot be overstated. Establishing models apposite in efficient management of extreme flows in rivers is of great practical value in flood and erosion and control. This is necessary in preventing environmental degradation while facilitating the attainment of a sustainable environment in consonance with the sustainable development goals of the United Nations.

This study presents the frequency analysis of peak annual flow of River Mkhomazi in South Africa. Six stochastic models were evaluated. It was found that the Normal, Pearson III, Log–Pearson III and Gumbel functions were suitable for modeling peak flow in the river. At–site analysis presents models that are more precise in computing river flow where historical data is available. Estimation must be accurate not only for preventing catastrophes, but also for avoiding exorbitant costs in design, arising from overestimating magnitude of peak flows.

The findings from this study can inform stakeholders and decision–makers in proper environmental and water resources management towards achieving a sustainable environment. Thus, the outcome of this work provides additional inputs for the improvement of hydrological studies in South Africa. Further studies may investigate the application of other models and methods not applied in this study.

Acknowledgement

The authors acknowledge the Department of Water and Sanitation, South Africa for making river flow data available for this study.

References

[1] Adeyemo, J. and Olofintoye, O.: Optimized Fourier Approximation Models for Estimating Monthly Streamflow in the Vanderkloof Dam, South Africa, Switzerland, Springer International Publishing, 2014.

[2] Bernini, A.; Bosino, A.; Botha, G. and Maerker, M.: Evaluation of Gully Erosion Susceptibility Using a Maximum Entropy Model in the Upper Mkhomazi River Basin in South Africa, International Journal of Geo–Information, 10 (11), 729–748, 2021.

[3] Bilewu, S. O. and Sule, B. F.: Discharge generation using combined thornthwaite water balance and ihacres models in central Nigeria: a case study of the Asa river basin, Annals of the Faculty of Engineering Hunedoara, International Journal of Engineering, 8 (3), 89–94, 2015.

[4] Greenfield, H. J.; Fowler, K. D. and Schalkwyk, L. O. V.: Where are the gardens? Early Iron Age horticulture in the Thukela River Basin of South Africa, Azania Archaeological Research in Africa, 37 (2), 307–328, 2005.

[5] Gupta, S. C.: Fundamentals of Statistics, Seventh edition, Delhi, Himalaya Publishing House, 2013.

[6] Haktanir, T.; Bajabaa, S. and Masoud, M.: Stochastic analyses of maximum daily rainfall series recorded at two stations across the Mediterranean Sea, Arabian Journal of Geosciences, 6 (2013), 3943–3958, 2012.

[7] Jou, P. H.; Akhoond–Ali, A. M.; Behnia, A. and Chinipardaz, R.: A Comparison of Parametric and Nonparametric Density Functions for Estimating Annual Precipitation in Iran, Research Journal of Environmental Sciences, 3 (1), 62–70, 2009.

[8] Khudri, M. M. and Sadia, F.: Determination of the Best Fit Probability Distribution for Annual Extreme Precipitation in Bangladesh, European Journal of Scientific Research, 103 (3), 391–404, 2013.

[9] Kjeldsen, T. R.; Smithers, J. C. and Schulze, R. E.: Regional flood frequency analysis in the KwaZulu–Natal province, South Africa, using the index–flood method, Journal of Hydrology, 255 (1–4), 194–211, 2002.

[10] Komolafe, A.; Olorunfemi, I.; Akinluyi, F.; Adeyemi, M. and Ajayi, J.: Enhanced flood hazard modelling using hydraulic, analytical hierarchical process and height above nearest drainage models in Ogunpa river basin, Ibadan, Southwestern Nigeria, Modeling Earth Systems and Environment, 7 (1), 967–981, 2021.

[11] Loucks, D. P. and Bee, E. V.: Water Resources Systems Planning and Management: An Introduction to Methods, Models and Applications, First edition, The Netherlands, United Nations Educational Scientific and Cultural Organization, 2005.

[12] Mays, L. W.: Water Resources Engineering, First edition, New York, USA, John Wiley & Sons, Inc., 2011.

[13] Mustapha, S. and Yusuf, M. I.: A Textbook of Hydrology and Water Resources, First edition, Abuja, Nigeria, Jenas Prints & Publishing Co. , 1999.

[14] Olofintoye, O. O.; Bakare, B. F.; Salami, A. W.; Ayanshola, A. M.; Bilewu, S. O.; Abdulkadir, T. S.; Jimoh, M. O. and Oyebode, O.: Probabilistic Modelling of Flood Flow Along Mgeni River at Table Mountain, South Africa, 1st Faculty of Engineering and Technology Conference (FETiCON 2023), 502–507, 2023.

[15] Oyebode, O.; Adeyemo, J. and Otieno, F.: Monthly streamflow prediction with limited hydro–climatic variables in the upper Mkomazi River, South Africa using genetic programming, Fresenius Environmental Bulletin, 23 (3), 707–718, 2014.

[16] Paul, S.; Hasan, M. A. and Shopan, A. A.: An Assessment of Best Fitted Rainfall Distribution for Projected Climate Change over Bangladesh using Statistical Downscaling Method, 2nd International Conference on Civil Engineering for Sustainable Development, 483–491, 2014.

[17] Reddy, P. J. R.: A Textbook of Hydrology, Third edition, Delhi, India, University Science Press, 2014.

[18] Sudmeier–Rieux, K.; Nehren, U.; Sandholz, S. and Doswald, N.: Disasters and Ecosystems, Resilience in a Changing Climate – Source Book, First edition, Geneva, UNEP and Cologne: TH Köln – University of Applied Sciences, 2019.

[19] Topaloglu, F.: Determining Suitable Probability Distribution Models for Flow and Precipitation Series of the Seyhan River Basin, Turk Journal of Agric, 26 189–194., 2002.

[20] Viessman, W.; Krapp, J. W. and Harbough, T. E.: Introduction to Hydrology, Third edition, New York, Harper and Row Publishers Inc., 1989.

ISSN 1584 – 2665 (printed version); ISSN 2601 – 2332 (online); ISSN–L 1584 – 2665