# MODELLING AND FORECASTING OF FLOOD EVENTS FOR DECISION SUPPORT IN INTEGRATED WATER RESOURCES MANAGEMENT IN GUCHA-MIGORI RIVER BASIN, KENYA

**EDITH AUMA ONYANGO**

**A Thesis Submitted to the Graduate School in Partial Fulfillment of the Requirements for Master of Science Degree in Soil and Water Engineering of Egerton University**

**EGERTON UNIVERSITY**

**OCTOBER 2024**

## DECLARATION AND RECOMMENDATION

### Declaration

This thesis is my original work and it has not been wholly or in part presented for the award of a degree or diploma in any other institution known to me.

Signature…… Date ……7/10/2024……

**Edith Auma Onyango**

**BM14/13718/19**

### Recommendation

This thesis has been submitted with our approval as University supervisors

Signature…… Date ……10/10/2024……

**Dr. Raphael M. Wambua, PhD.**

Department of Agricultural and Biosystem Engineering

South Eastern Kenya University

Signature…… Date ……10/10/2024..…

**Dr. Hesbon Otieno, PhD.**

Department of Agricultural and Biosystem Engineering

South Eastern Kenya University

## COPYRIGHT

## DEDICATION

This work is dedicated to God my creator. To my dearest husband Douglas Musa Machage, and mother Cecilia Akinyi for continuous prayers, love, support, motivation, encouragement, guidance, and mentorship. To my children Laurlyn Moraa, Zacharia Machage, Morna Kwamboka, Cecilia Bwari and Tirzah Nyambeki, may this inspire you to work hard.

## ACKNOWLEDGEMENTS

I take this privilege to express my humble gratitude to God my creator for His constant love and grace. With a lot of thanksgiving in my heart, I extend my acknowledgment to Egerton University for providing the necessary learning resources and many opportunities in the Department of Agricultural Engineering to accomplish this academic objective. I take this prerogative to express my deepest appreciation to my supervisors, Dr. Raphael M. Wambua and Dr. Hesbon Otieno for enthusiastic mentorship, a marvellous sense of humour, enthusiastic mentorship, and dedication, throughout the completion of this work.

Additionally, I sincerely appreciate the immense support I got from the faculty of Engineering and Technology staff. My heartfelt indebtedness also goes to my husband, children, mother, brother, and sister for their support, prayers, and encouragement for the entire period, I am forever grateful. Finally, yet importantly, my sincere thanks go to all my friends and colleagues for continuous backing, consultations, different opportunities, and the push to attain this higher academic honour, their support will forever be in my heart.

## ABSTRACT

Floods are devastating natural disasters frequently striking many river basins such as Gucha-Migori River Basin nearly every year. However, non-structural countermeasures have not been fully explored and implemented to reduce damage, risk, and vulnerabilities of flood events. In light of this, the main objective of the study was to model and forecast flood events in the Gucha-Migori River Basin for decision support in integrated water resource management. Modelling the response of flood events to precipitation variability using the HEC-HMS model for the period (1973-2015) was the first activity conducted. Then a relationship between flood magnitude and frequency was developed using probability distribution model. Coefficient of Determination ($R^2$), Goodness of Fit test, and Best-Fit Distribution Curve guided selection of the suitable distribution model. Finally, flood events and their respective annual exceedance probabilities were forecasted using Artificial Neural Networks (ANNs) and the Gumbel distribution model for the period (2015-2052). From the results, independent flood events were detected for the years 1969, 1971, 1974, 1977, 1969, 1981, 1982, 1983, 1985, 1990, 1996, 2004, 2006, 2007, 2010, 2011, and 2013. The $R^2$ and Nash values for HEC-HMS model calibration and validation of daily river discharge at sub-basin 1 were 0.52 and 0.36, and, 0.42 and 0.31 respectively. The correlation coefficients between daily precipitations for JFM, AMJ, JAS, and OND, and the corresponding river discharges had a significant positive correlation at the significant level of 0.05. It was revealed that the probability of a flood event of a magnitude equal to or exceeding the lowest (240.8 $m^3/s$) and the highest (423.90 $m^3/s$) occurring for a particular were 0.98 and 0.02 respectively. Considering the predictions from Gumbel, Log P3, Log N, and Normal models, for the return periods 1.05 and 200 years, the flood magnitude estimated for G-M River Basin would be 221.38 and 466.65, 233 and 478, 178.19 and 367.38, and 209 and 402 $m^3/s$. Gumbel distribution curve was selected to be the best fit for the Gucha-Migori River Basin. Using the comparative analysis, the network topology of 1-20-1 was adopted as the best NAR model for forecasting because of its minimum values of R for training (9.20e-1), validation (9.27e-1), and testing (9.15e-1). The maximum forecasted flood magnitudes were 336, 391, 433, 299, 389, 502, and 483 $m^3/s$ for the periods 2014–2018, 2019–2023, 2024–2028, 2029–2033, 2034–2038, 2039–2043, and 2048–2052, respectively. Their respective annual non-exceedance probabilities from the Gumbel's curve and derived equation were 84.6, 96, 98, 64, 96, 99.6, and 99 %. The information from the research offers prerequisite flood response plans and provides early flood detection and forecasting for planning for the management of flood risks and for preparedness in the Gucha-Migori River Basin.

**TABLE OF CONTENT**

## LIST OF TABLES

## LIST OF FIGURES

## LIST OF SYMBOLS

| Symbol | Description |
|---|---|
| b | Bias at the cell body |
| $f(x)$ | Gamma function |
| $g(x)$ | Gamma distribution |
| $m$ | Number of hidden neurons |
| $n$ | Number of input neurons |
| $Wi$ | Weight attached to the input signal $i$ |
| $Xi$ | An input vector |
| $R^2$ | Coefficient of Determination |
| α | Significance level |
| $p$ | Probability values |
| $r\_$ | Substitution of a parameter by a value |
| $v\_$ | Relative change in a parameter |
| f (.) | Activation function |
| NSE | Nash Sutcliffe model efficiency |
| Sim | Simulation |
| CV(t-p) | Feedback delay |
| $Z_c$ | Calculated standard normal distribution test statistics |

## LIST OF ABBREVIATIONS AND ACRONYMS

| | |
|---|---|
| AEP | Annual Exceedance Probability |
| ANN | Artificial Neural Network |
| AMJ | April, May, June |
| CS | Climatic Scenario |
| DEM | Digital Elevation Model |
| FAO | Food and Agriculture Organization of the United Nations |
| FNN | Feedforward Neural Network |
| GUM | Gumbel |
| HEC-HMS | Hydrologic Engineering Centre - Hydrologic Modeling System |
| HEC-GeoHMS | Hydrologic Engineering Centre -Geospatial Hydrologic Modeling |
| HRU | Hydrologic Response Unit |
| IPCC | Intergovernmental Panel on Climate Change |
| JFM | January, February, March |
| JAS | July, August, September |
| LP3 | Log Pearson type III |
| LULC | Land Use-Land Cover |
| MSS | Multi-Spectral Scanner |
| OND | October, November, December |
| PD | Probability Distribution |
| SCS-CN | Soil Conservation Service Curve Number |
| SRTM | Shuttle Radar Topography Mission |
| SUFI-2 | Sequential Uncertainty Fitting 2 |
| SDG | Sustainable Development Goals of the United Nations |
| SWAT | Soil and Water Assessment Tool |
| TM | Thematic Mapper |
| USGS | United States Geological Survey |
| UTM | Universal Transverse Mercator |
| UNEP | United Nations Environmental Programme |
| UNFCCC | United Nations Framework Convention on Climate Change |
| WGEN | Weather Generator |
| WMO | World Meteorological Organisation |

# CHAPTER ONE
# INTRODUCTION

## 1.1 Background Information

Climate change and growing socio-economic developments continue to alter hydrological regimes and the response of several watersheds globally, thus threatening water energy, and food security (Archer et al., 2010; IPCC, 2014; UNFCCC, 2005; Wambua et al., 2017). In Kenya, there is a study recognition of ecosystem deterioration, spatial and temporal uncertainties in water input of various basins, and changes in watershed storage, which further aggravate extreme hydrological events like floods (Juma et al., 2020; Opere, 2013). The aforementioned challenges accelerate flood scenarios because such extreme hydrologic events depend entirely on parameters that include soil antecedent moisture, land use system, catchment area, and rainfall patterns (Gichere et al., 2013). Furthermore, flooding is usually caused by highly intense and continuous rainfall occurring for an extended duration, which then exceeds the infiltration capacity of soil or flow capacity of coastal regions, rivers, and streams resulting in runoff (Thavhana, 2018).

Surface runoff can be through Hortonian or saturation overland flow. Hortonian overland flow is generated when the intensity of rainfall exceeds the infiltration capacity of the soil, while saturation overland flow is produced when subsurface flow saturates the soil near the bottom of hill slopes and near stream banks. The runoff produced may then cause the catchment to respond in a manner that results in flooding (Singo et al., 2012; Warburton et al., 2010). Adero (2017) explains that flood occurrence in the Gucha-Migori River Basin is a result of heavy rains upstream that lead to heavy surface run-off flow into both the Gucha and Migori river systems. The surface run-off erodes soil as it flows into the river leading to heavy sedimentation of the river channel as the water flows downstream.

On the other hand, the flood-plain area also experiences flash floods from the neighbouring hills within the Lower Gucha Migori Sub-catchment and this leads to the loss of crops and disruption of agricultural activities (Muiruri et al., 2014; Ogembo, 2018). The flood hazard events prevalence rate in Kenya stands at 27% and affects more than 5% of the population (Akello, 2014). Besides, the United Nations Environmental Programme (UNEP, 2010) points out that flood-related fatalities constitute over 60 % of disaster victims in Kenya. Several studies report that flood hazards have increased in number, frequency, and complexity in many parts of Kenya including the Gucha-Migori River Basin (Adero, 2017; Gaya, 2020; Ogembo,

2018; Opere, 2013; Quandt & Kimathi, 2017). The level of destruction has also become more severe and poses consequences on agricultural activities, water resources management, human health, development of engineering projects, wildlife, hydropower generation, and forestry (Akello, 2014; Gaya, 2020). It then implies that strategies addressing flood-related challenges are critical for the realisation of Kenyan's big four agenda, Sustainable Development Goal (SDG) number 1, 2, 3, 13, and 16, and Vision 2030.

The common strategies and mitigation measures for coping with floods involve warnings as well as the construction of flood protection structures especially dykes and dams (Gaya, 2020). It encompasses relevant attempts for mitigation against the impacts of flood damage, which include disruption of the daily livelihoods, and enlighten schemes that make it easier and faster life-skills in recovery from the flood damage. However, WMO (2011a, b) illustrates that in the course of drawing up the flood management plans, the appropriate combination of structural and non-structural measures should be considered. Non-structural flood mitigation measures encompass all activities focused on land-use regulations, hydrologic modelling, disaster prevention, flood-proofing, continuous forecasting, and flood frequency analysis. Moreover, their implementation requires the application of modern techniques, which should describe, analyse spatially and temporally, simulate, and integrate forecasted flood magnitudes with other related information.

Non-structural methods make it possible to monitor the progress of floods in near real-time and to produce accurate, rapid, and cost-effective mapping of floods (Gaya, 2020). The coupling of hydrologic models and flood frequency distribution techniques would therefore provide a hybrid tool for the simulation of flood events. Thus, this study focuses on modelling flood discharge using a hydrologic model (HEC-HMS), and the integration of the PD model with an artificial neural network for frequency analysis and forecasting of the events in the Gucha-Migori River Basin. The study seeks to provide useful insights, knowledge, and data products, which may enhance the probability of establishing that a flood of a given magnitude will hit during any particular year. Further, such a study is necessary for basin rehabilitation, and monitoring as well as the design of suitable approaches for integrated water resource management (Birkholz et al., 2014; Hansson et al., 2008).

### 1.2 Statement of the Problem

Floods are the most devastating natural disaster frequently occurring in catchments within the Gucha-Migori River Basin nearly every year according to the JICA draft report on integrated

flood management plan (2014) and the local water resource authority. Previous studies such as Adero (2017), Gaya (2020), and Ogembo (2018) reveal that trends in flood magnitude, frequency of occurrence, and damages have progressively increased during the last decades. Although flood risk resilience has been conducted by the Water Resources Authority using structural developments and consensus-building approaches, these methods alone have proven insufficient in effectively mitigating the adverse impacts of floods.

The increasing frequency and severity of floods necessitate a shift towards non-structural interventions, which are essential for comprehensive flood management. Non-structural methods include land-use regulations, hydrologic modelling, disaster prevention, flood-proofing, continuous forecasting, and flood frequency analysis. These approaches provide more flexible, adaptive, and sustainable solutions by emphasizing preparedness, early warning systems, and real-time flood monitoring. Integrating non-structural methods with structural measures enhances the overall resilience of the flood management strategy and supports the sustainable development goals.

This study explores the need for non-structural interventions by focusing on hydrologic modelling, quantification of the response to flood events, and prediction of flood occurrence probabilities using probability distribution functions and artificial neural networks. The implementation of these advanced techniques aims to address the knowledge gap regarding future flood scenarios in the basin, thereby improving flood management strategies and enhancing community resilience within the Gucha-Migori River Basin. This integrated approach is critical for basin rehabilitation, monitoring, and the design of effective water resource management plans.

## 1.3 Objectives

### 1.3.1 Broad Objective

The broad objective was to model and forecast flood events in the Gucha-Migori River Basin for decision support in integrated water resource management.

### 1.3.2 Specific Objectives

The specific objectives were to:

i) Determine the response of river discharge to precipitation variability for the period 1969 to 2015 using the HEC-HMS model
ii) Determine the relationship between flood discharge magnitude and frequency using the appropriate probability distribution model

iii) Forecast flood magnitudes and P-percent annual exceedance probabilities for the period (2015 – 2052) using an artificial neural network and probability distribution model

### 1.4 Research Questions

i) How does precipitation variability influence river discharge over the period 1969 to 2015, as simulated by the HEC-HMS model?
ii) How do flood magnitude and frequency relate based on probability distribution models?
iii) How will flood magnitudes and probabilities of exceedance vary for the period between 2015 and 2052?

### 1.5 Justification

The impact of flood events adversely affects livelihoods, its' complex and highly uncertain and has the potential to influence relatively stable socio-economic activities to a precarious situation and jeopardize basin management plans. Therefore, modelling the response of flood events to precipitation variability is important for basin rehabilitation and the design of strategic approaches for integrated flood implementation plans. Analysis of flood frequency and probability distribution curves provide data products for the design of hydraulic structures, floodplain zoning, and the creation of system operational plans. Flood forecasting creates a reasonable design criterion for water resources projects, thus, reducing projects' cost, and rationally saving lives, properties, and socio-economic activities.

The selection of the Hydrologic Engineering Center's Hydrologic Modeling System (HEC-HMS) for this study is based on its proven capabilities and widespread application in hydrologic modeling, particularly in simulating river discharge responses to precipitation events. HEC-HMS has been extensively used in various hydrologic studies globally and has demonstrated high efficacy in simulating hydrologic processes in river basins. Its application is well-documented in numerous studies, including those conducted in Kenya's Nyando, Ruiru, Kuja, Mkurumudzi, and Umba river basins (Ismael et al., 2017; Ogembo 2018; Olang & Fürst 2011; Ouédraogo et al., 2018; Tesfamariam et al., 2021),

Furthermore, HEC-HMS is preferred over other models due to its balance between being physically-based and less demanding on input data compared to fully distributed models (Bhuiyan et al., 2017; Gyawali & Watkins, 2013; Thu et al., 2019). This makes it a practical choice for regions where comprehensive data collection is challenging. HEC-HMS's ability to model various hydrologic processes, including precipitation, infiltration, surface runoff, and

channel routing, ensures that all significant aspects of the hydrologic cycle are captured. This makes it highly effective for flood forecasting and water resources management.

Additionally, probability distribution curves are employed in this study to statistically analyse and predict extreme hydrologic events. These curves are vital for understanding the frequency and magnitude of flood events, thereby providing essential insights for flood risk management. Probability distribution curves help in quantifying the likelihood of various discharge levels occurring, which is crucial for designing infrastructure and implementing mitigation measures. The use of probability distributions, such as the Log-Pearson Type III distribution, is standard practice in hydrology for modeling flood frequency and assessing return periods of extreme events (Engeland et al., 2004; Hussain et al., 2018; Kumar, 2019).

Moreover, artificial neural networks are integrated into this study to enhance the prediction accuracy of hydrologic responses to precipitation variability. ANNs are particularly effective in capturing non-linear relationships between input variables and hydrologic responses, which traditional models may not adequately represent. The flexibility and adaptability of ANNs make them powerful tools for forecasting river discharge, especially when dealing with complex and highly variable hydrologic data (Dawson & Wilby, 2001; Maier & Dandy, 2000). By training the neural network with historical precipitation and discharge data, the model can learn patterns and improve its predictive performance, thus offering a complementary approach to traditional hydrologic models.

In conclusion, the choice of HEC-HMS, probability distribution curves, and artificial neural networks for this study is justified by their extensive application in similar studies, their capability to handle complex hydrologic processes with relatively moderate data requirements, and their robust frameworks for simulation, calibration, validation, and prediction. These methods and models collectively provide a comprehensive toolkit for investigating river discharge responses to precipitation variability and contribute valuable insights into flood water resources planning and management.

Therefore, employing advanced modeling techniques such as the HEC-HMS model, artificial neural networks, and probability distribution models, this research contributes to global efforts in flood risk management and disaster preparedness. Accurate flood predictions will benefit local communities in Kenya while also providing insights applicable to other flood-prone regions worldwide. The findings support global initiatives in IWRM, helping to achieve

Sustainable Development Goal (SDG) 13 on climate action and SDG 6 on ensuring the availability and sustainable management of water resources.

### 1.6 Scope and Limitations of the Study

The study area covers only the Gucha-Migori River Basin. HEC-HMS model was used for flood simulation for the period (1969-2015). For the spatial data (Digital Elevation Model, Land Cover, and Soil Maps), it was assumed that the processed datasets have the same characteristics as the real physical features. Further, the study focused on the integration of the Probability Distribution model with the Artificial Neural Network Model to forecast flood magnitudes and their respective P-percent annual exceedance probabilities.

### 1.7 Definition of terms

| | |
|---|---|
| Artificial Neural Network | A data-centric procedure employing a flexible mathematical algorithm or a mimicry of the biological nervous system that possesses the ability to address intricate nonlinear associations. |
| Climate | The statistical representation in terms of the variability and the mean of relevant variables or average weather over a duration ranging from weeks to millions of years, i.e. it is the weather conditions prevailing in an area over a long period of time. |
| Climate variability impact | The enduring impacts or outcomes of climate fluctuations that extend beyond singular weather occurrences. |
| Precipitation variability | Describes how precipitation may fluctuate over time and tells us how big the difference between wet and dry conditions in a particular place could be. |
| Flood Event | The inundation of a dry place caused by overflowing water in an existing waterway such as a stream, river, or drainage ditch. |
| Forecasting | Forecasting future states of a phenomenon, system, or quantity by leveraging information about its current and past states, frequently achieved through the use of a model. |
| HEC-HMS Model | A model developed by the US Army Corps of Engineers of the Hydrologic Engineering Center (HEC), which contains an integrated tool for simulating hydrologic variables of the dendritic river basin. |

| | |
|---|---|
| Hydrologic Modelling | The description of actual hydrological attributes and systems through the utilization of computer simulations, mathematical counterparts, and miniature physical replicas. |
| Land cover | The ground's outer layer, encompassing elements such as vegetation, urban structures, bodies of water, exposed soil, and more. |
| Land use | The function that the land fulfils, such as providing opportunities for recreation, serving as a habitat for wildlife, or supporting agricultural activities. |
| River Discharge | The volume of water flowing through a river channel; measured at any given point in cubic meters per second |

# CHAPTER TWO
# LITERATURE REVIEW

## 2.1 Flood

According to Bourenane et al. (2019), a flood is an overflow of water that submerges land, which is usually dry, or an abnormal progressive rise in the water levels of rivers or channels that may result in an overflow. The overflow is mainly due to runoff response to changes in upstream land use, heavy precipitation, a storm surge from a tropical cyclone, and continuously increasing population and assets in flood-prone areas (Sriwongsitanon & Taesombat, 2011; Viglione et al., 2009). In hydrology, such response of runoff to human induced land use change and precipitation characteristics can further be illustrated by flood hydrograph as shown in Figure 2.1

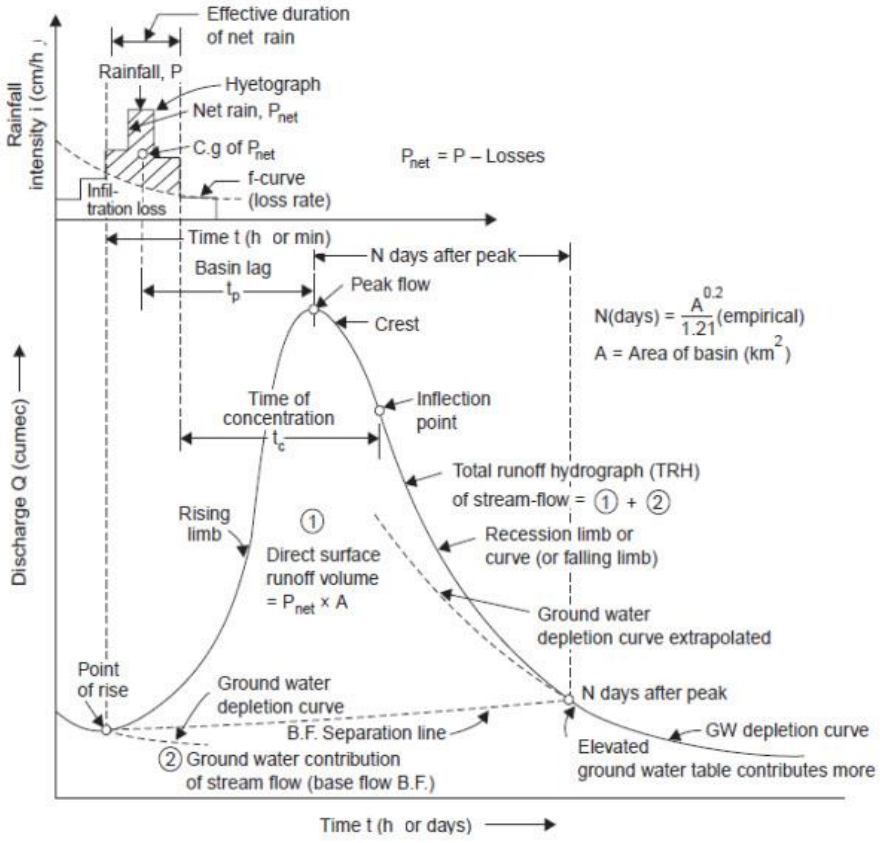



Figure 2.1: Flood Hydrograph

Source: Raghunath, (2006)

In a broader perspective, the flood hydrograph (Figure 2.1) shows how a drainage basin responds to a period of rainfall. Rising limb, excess rainfall, peak discharge, and falling limb respectively indicate an increase in river discharge, excess runoff, time when the river reaches

its highest flow, and runoff is still reaching the river but in decreasing amounts (Albishi et al., 2017; Raghunath, 2006). It implies that a gentle hydrograph shows the river is at low risk of flooding, which means that it takes longer for the peak rainfall to reach the river channel, so the river discharge is increasingly slow. Floods can be categorised into river floods, coastal floods, storm surges, inland flooding, and flash flood (Jonkman & Vrijling, 2008; Perry, 2000). However, USGS illustrates that there exist two basic types of floods, namely, widespread river floods and flash floods. Some researchers like Booij (2005) and Douben (2006) describe river flooding to be a common natural disaster for large rivers in areas with heavy precipitation, in the event where excessive runoff from longer-lasting rainstorms causes a water-level rise for a relatively longer duration over a larger catchment.

Other studies (Dottori et al., 2018; Jongman et al., 2012) point out that river floods can occur slowly or abruptly, where, sudden runoff events happen more frequently on normally dry channels, rivers that flow for much of their length over impermeable terrain, and rivers with steep valleys. On the other hand, flash floods occur when runoff from excessive precipitation causes a rapid rise in the water height of a normally dry channel or stream in a short amount of time sometimes limiting the time available to warn and protect the public (Davis, 2001).

The assertion by Collier (2007) and Hapuarachchi et al. (2011) further demonstrate that flash flood is accompanied by other phenomena like psychological harm to people, damage to properties and businesses, bridge collapse, landslides and mudflows, and, in exceptional circumstances, fatalities. Exacerbated anthropogenic activities and climate change indicate an increase in watershed degradation and the intensity of rainfall posing a potential influence on runoff generation in many parts of the world including Kenya, which may lead to more severe flash flooding (Mishra et al., 2018; Ogembo, 2018; Rwigi, 2014).

### 2.2 Response of Flood Events to Precipitation Variability

Changes in climate and land-use systems alter the hydrologic regime of a basin by changing precipitation partitioning, which brings about extreme variations of subsurface flows, groundwater flows, and surface runoff (Amisi et al., 2020; Schipper, 2017). Such changes bring different responses of runoff, which may then contribute to flooding events because precipitation variability and changes in the Land-use system occur at different spatial and temporal scales. Besides, the impact of changes in precipitation and LULC might compensate for or strengthen each other. On the other hand, some researchers like Renner et al. (2014) explain that they both might occur in parallel.

For instance, Mwetu (2019) analysed the influence of land cover changes and climate variability on the discharge regime of Njoro River Catchment, Kenya. Results showed a negative trend of annual discharge that corresponded to increased deforestation, open fields, bare land, and grassland. Further analysis pointed out that human activity mainly deforestation accounted for 75% of the reduction of discharge while precipitation variability accounted for 25% of the reduction of discharge. This implies that land-use changes were largely responsible for runoff, which in turn would then potentially contribute to flooding events.

Although there have been several studies conducted on the surface runoff response to changes in LULC, the evidence from some studies is still contradictory. For example, other researchers elaborate that an increase in the annual discharge is caused by a reduction of forest cover (Dias et al., 2015; Farley et al., 2005; Hayhoe et al., 2011). On the other hand, studies like Githui et al., (2009), Kundu (2007), Mwetu (2019), and Rodriguez et al. (2010) postulate that land-use changes from forest to mixed rural built-up lands and subsistence agriculture causes the increase of runoff but deteriorates subsurface as well as groundwater flows.

However, the aforementioned responses of surface water circular are based on the fact that LULC influences soil physical properties, available water content, infiltration capacity, and saturated hydraulic conductivity. This would mean that the changes in the land-use system might significantly influence flood magnitude, time, and spatial variation.

For example, Olang and Fürst's (2011), analyses on the effect of LULC on flood peak discharges and runoff volumes in Nyando River Basin, Kenya using HEC-HMS successfully outlined the hydrologic consequences of the eminent land cover changes. The model results further demonstrated that with higher rates of upstream deforestation, the runoff and flood peak discharges and volumes increased by at least 10 % in the entire basin. Such findings are consistent with many studies like Brody et al. (2014), Hussein et al. (2020), and Szwagrzyk et al. (2018). Conversely, Kabeja et al. (2020) evaluated the impact of reforestation-induced LULC on flood peak discharge using HEC-HMS. Their results pointed out that the reforestation resulted in a decrease in flood peak discharge ranging from 14-16%.

Although changes in LULC influence runoff and flood magnitudes, the main dominant factor in the basin affecting the spatial and temporal distribution of flood routing processes is precipitation variability (Davenport et al., 2021). For instance, Tabari (2020) quantified the response of flood intensities to changes in extreme precipitation and found that extreme storm events intensified flood volumes with the seasonal cycle of water availability. Similarly,

Chegwidden et al. (2020), Fang et al. (2021), and Gandini et al. (2020) have also reported a stronger relationship between precipitation and flood events. Also, for the runoff analysis, Ogembo (2018) carried out hydrologic modelling and climate change studies on River Kuja Basin using HEC-HEC-GeoHMS and reported that both Runoff and precipitation showed a downward trend analysis.

The analysis by Ogembo (2018) revealed a significant decline in runoff in drier months compared to wetter months and concluded that there will be a high risk of flash floods in the future but consequently very low discharges during dry seasons. A similar response has also been reported by Amisi et al. (2020), Mango et al. (2010), Mwangi et al. (2016), and Rwigi (2014). It implies that direct runoff is a result of the spikes caused by a rainstorm, which then contributes to peak flow, hence river or flash floods.

Therefore, it can be inferred that precipitation variability influences the variation in spatial redistribution, time, frequency, and magnitude of floods. Therefore, it is imperative to note that the response of flood routing events to precipitation variability is complex, uncertain, and varies spatio-temporally, and hence may only be simulated in most cases by hydrologic models (Devia et al., 2015; Mwangi et al., 2016; Mwetu, 2010).

### 2.2.1 Hydrologic Modelling

Hydrologic modelling is the characterization of real hydrologic features and systems by the use of computer simulations, mathematical analogues, and small-scale physical models (Devia et al., 2015). Hydrologic models can broadly be classified into stochastic and deterministic models. Stochastic models possess inherent randomness, where the same set of parameter values and initial conditions will lead to an ensemble of different outputs, while in deterministic models, the output is fully determined by the parameter values and the initial conditions (Farmer & Vogel, 2016).

Marhaento et al. (2016) elucidate that hydrologic models are vital to the optimization and operation of water resources, and this explains why many statistical, empirical, and conceptual discharge prediction models have been developed for decision support in water management. However, statistical techniques, such as linear regression-based approaches are constrained, simplistic, and have limited capacity to handle non-linear relationships (Amisi et al., 2020).

Conceptual hydrologic models are usually considered the best alternative because they take into account hydrologic processes through mathematical formulations (Devia et al., 2015). The

law of conservation of mass and momentum energy guides the development of conceptual hydrologic models since most physical modelling includes storage of water, the spatial and temporal distribution of flows, and occurrence. These laws can be expressed by the concept of continuity and momentum Equations (2.1 and 2.2). The equations are applied in modelling under the assumption that the flow is unidirectional, the fluid is incompressible, vertical accelerations are negligible and hydrostatic pressure prevails.

$$\frac{\partial \mathrm{A}}{\partial t}+\frac{\partial\left(v\mathrm{A}\right)}{\partial x}-q=0 \tag{2.1}$$

$$\frac{\partial Q}{\partial t}+\frac{\partial\left(vQ\right)}{\partial x}+g\mathrm{A}\left(\frac{\partial y}{\partial x}-s_0+s_f\right)=0 \tag{2.2}$$

where, $q$ is the lateral inflow ($m^3$), $Q$ is the discharge in the channel ($m^3/s$), A is the area of flow in the channel ($m^2$), n is Manning's coefficient for the channel, $S_0$ is the bed slope (m/m), and $S_f$ is the friction slope of the channel (m/m).

Many researchers and organisations affirm the importance of conceptual hydrologic models such as SWAT, HRDROTEL, HEC-HMS, WATFLOOD, MIKE11/SHE, SPHY, WEAP, and TOPMODEL in the vital role they play in integrated water resources management solutions. Out of all the conceptual models, HEC-HMS has been demonstrated to be the most preferred model for flood routing processes simulation (Fleming, 2004; Natarajan & Radhakrishnan, 2019; Oleyiblo & Li, 2010; Thakur et al., 2017). HEC-HMS has been widely used because it is more physically based than the structure of lumped models, and they are less demanding on input data than fully distributed models (Bhuiyan et al., 2017; Gyawali & Watkins, 2013; Thu et al., 2019). Besides, it explains why the HEC-HMS model has been applied successfully in many River basins in Kenya which include Nyando (Olang & Fürst, 2011), Ruiru (Ismael et al., 2017), Kuja (Ogembo, 2018), Mkurumudzi (Ouédraogo et al., 2018), and Umba (Tesfamariam et al., 2021).

### 2.2.2 Hydrologic Engineering Centre - Hydrologic Modeling System (HEC-HMS)

Hydrologic Engineering Center (HEC-HMS) is a model developed by the US Army Corps of Engineers of the Hydrologic Engineering Center (HEC), which contains an integrated tool for simulating hydrologic variables of dendritic river basin systems. It is a physically-based semi-distributed, event-scale hydrologic model used to represent hydrological processes (Chu & Steinman, 2009; Du et al., 2019). As further illustrated by Fleming (2004), the construction of

the HEC-HMS model for simulation involves dividing an entire river basin into homogeneous sub-basins based on the defined drainage area threshold. The mass flux and energy balances within the hydrologic cycle are then modelled using mathematical equations.

HEC-HMS model requires spatial data (digital elevation model, LULC, and Soil maps), meteorological, and hydrologic datasets (Ouédraogo et al., 2018). The spatial data is prepared and processed from HEC-GeoHMS and imported to the HEC-HMS model to simulate flow, stage, and timing for the river basin. Therefore, the HEC-HMS model has three major components namely, control specification, meteorological model, and basin model (Ramly & Tahir, 2016). The meteorological model consists of the precipitation and evaporation datasets, the control specification contains calculation intervals for the run, while the basin model contains elements of the sub-basin, the connectivity, and the runoff parameter. HEC-HMS model consists of the following components for the simulation of the runoff response to precipitation (Gebre, 2015; Ogembo, 2018). Precipitation specification option, which describes the historical data at a given location.

The second component is the loss models that approximate the runoff magnitudes given precipitation datasets and river basin characteristics. Other components include direct runoff models that account for overland flow, energy losses, and storage, and the hydrologic routing models, which account for energy flux and storage during the period water flows through the stream channels. Some of the additional models are those for naturally occurring confluences and bifurcations, and models of water control measures. The HEC-HMS model contains eight elements, which include the diversions, reservoir, sink, source, junction, reach, and sub-basins (Ogembo, 2018; Ramly & Tahir, 2016).

Every sub-basin has corresponding precipitation thus the outflow for the element is computed through the subtraction of the precipitation losses. Surface runoff is then calculated and the base flow is added. The reach element conveys the river discharge to the basin model and the inflow into the reach is obtained from the upstream elements. Translation and attenuation should be accounted for then outflow from the reach is calculated. Figure 2.2 shows an overview of the rainfall-runoff process simulation in HEC-HMS.

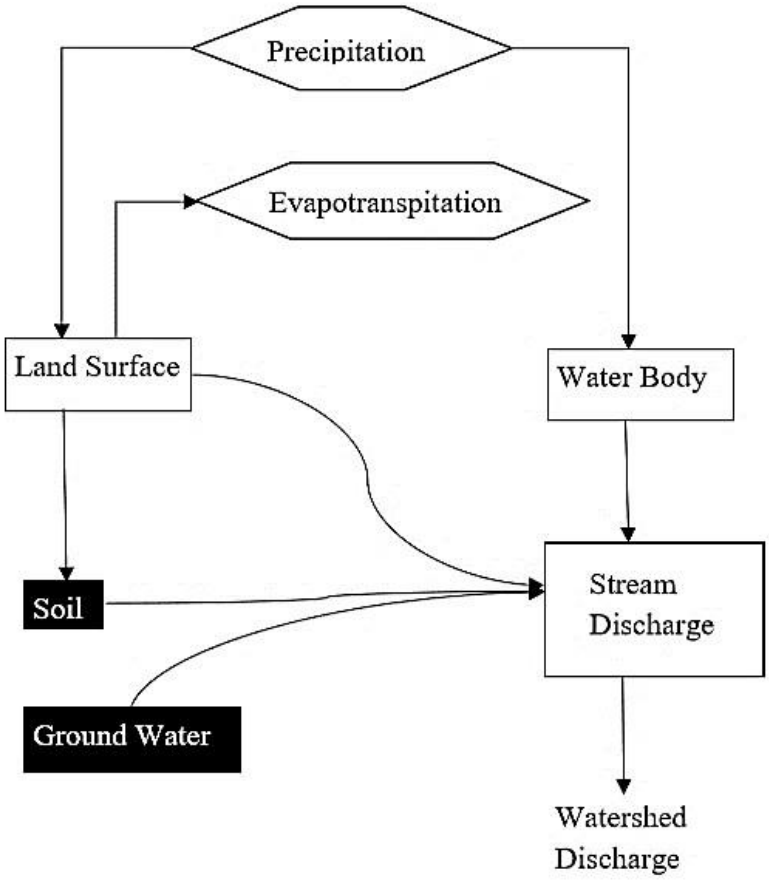



Figure 2.2: Schematic Diagram of the Rainfall-Runoff Process in HEC-HMS

Source: Ogembo, (2018)

All the inflow into the junction must be summed up to obtain the outflow, and the source element introduces inflow into the basin model. However, the outflow of the source element may be defined by the user, and hence the sink element is used to represent the outlet of the basin. The reservoir element models the hydrograph detention and attenuation as a result of the detention pond. Computation of the outflow from the reservoir can be achieved by applying any of the routing approaches in HEC-HMS (Fleming, 2004; Ogembo, 2018).

The diversion element models the discharge leaving the main channel, and its inflow can be obtained from many upstream channels, it should be noted that computation of the diverted flow can be done using the user's input. Finally, the non-diverted and diverted can be connected to other hydrologic elements located downstream of the diversion elements (Scharffenberg & Fleming, 2005).

### 2.2.3 Hydrologic Engineering Centre -Geospatial Hydrologic Modeling Extension

HEC-GeoHMS is an extension for ArcGIS released by the US Army Corps of Engineers, Hydrologic Engineering Center (HEC) (Fleming, 2004; Tesfamariam et al., 2021). It is a toolkit of geospatial hydrology, which allows operators to process and create basin parameters based on the topographic data for hydrologic modelling (USACE, 2010). In addition, HEC-GeoHMS processes geospatial data and creates their input files in ArcGIS for HEC-HMS. Figure 2.3 adapted from USACE (2013) shows the relationship between GIS, HEC-GeoHMS, and HEC-HMS.

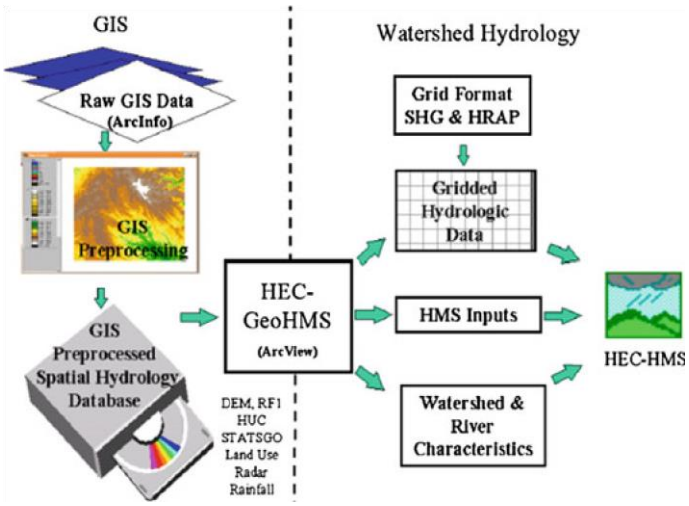



Figure 2.3: Overview of HEC-GeoHMS Program

Source: USACE, (2010)

HEC-GeoHMS is applied to derive a river network of the basin from the digital elevation data. To obtain a digital elevation model to demarcate various components of the catchment, some of the major steps undertaken include filling sinks, flow direction derivation, demarcation of basin and sub-basin, calculation of flow accumulation, and stream network definition (USACE, 2013). HEC-GeoHMS creates the drainage network by analysing the digital terrain data and transforming the drainage paths and basin boundaries into a hydrologic data structure.

### 2.2.4 HEC-HMS Model Calibration and Validation

The successful application of hydrologic models including HEC-HMS depends on how well calibration and validation are achieved. Calibration is the process of minimizing the model output uncertainties before running simulations by adjusting parameters to best capture the local conditions (Abbaspour et al., 2007; Arnold et al., 2011; Kamali et al., 2013). However, validation is the running of the model with calibrated parameters in an independent data mass

so that the applicability of the model can be assessed through statistical tests. For the HEC-HMS model, manual calibration is performed by visual inspection and trial and error-based procedures. (Kamali et al., 2013). Some studies such as Dariane et al. (2016) and Odyuo (2020) point out that manual calibration can be tedious, time-consuming, and may require experience and experimental information. Auto-calibration is based on systematic techniques to find the best parameter values based on pre-defined objective functions, thus, more likely to produce better results (Dariane et al., 2016; Kamali et al., 2013; Sardoii et al., 2012).

Some of the widely applied auto-calibration techniques for the HEC-HMS model entail Sequential Uncertainty Fitting 2 (SUFI-2), Genetic Algorithm (GA), Nelder-Mead (NM), and Univariate-Gradient (UG), Melody Search Algorithm (MeS), Self-Adaptive Global Harmony Search (SGHS). Several studies have highlighted various aspects and challenges inherent in the aforementioned auto-calibration approaches such as parameterization, non-uniqueness of parameter values (Abbaspour et al., 2004), and data quantity and quality (Perrin et al., 2007). Also, the type and the number of objective functions and their characteristics (indifference to parameter values, non-convexity, and multi-modality) (Kamali et al., 2013), and the performance of the optimization algorithms (Ndiritu, 2009). The methodology for addressing uncertainty, sensitivity, calibration, and validation has been detailed. Specific techniques used, parameters tested, and evaluation criteria are explained to ensure robustness and reliability of the model.

For instance, studies by Kamali et al. (2013) preferred the use of a multi-objective to Particle Swarm Optimization (PSO) algorithm for automatic calibration of HEC-HMS. According to Dariane et al. (2016), the Genetic Algorithm is preferred to Nelder-Mead and Univariate-Gradient when using a snowmelt module in a continuous model for enhancement of efficiency. Many researchers such as Abbaspour et al. (2015), Kumar et al. (2020), and Mousavi et al. (2012) recommend the SUFI-2 algorithm due to its higher performance in terms of adjusting model parameters to mimic real hydrologic conditions.

### 2.2.5 Hydrologic Model Performance Assessment

A wide range of statistics has been used to evaluate HEC-HMS hydrologic predictions. By far the most widely used statistics reported for hydrologic calibration and validation are Coefficient of Determination ($R^2$), and the Nash Sutcliffe model efficiency (Nash). The R-squared measures how the observed versus simulated regression line approaches an ideal match (Fernandez et al., 2005). $R^2$ values range from zero to one, with a value closest to zero

indicating no correlation while a value closest to one indicating relative perfect linear covariation between the two datasets. It implies that a correlation measurement reveals the high accuracy of the model when the value is approaching one.

Nash Sutcliffe model efficiency ranges from $-\infty$ to 1 and measures how well the observed and simulated data match the 1:1 line (Moriasi et al., 2007). The values of Nash closer to zero or equal to 0 indicate that the mean of the observed data is a better predictor than the model, and the values closer to one or 1 imply a perfect fit between the observed and simulated. Considering various time steps of modelling, the performance evaluation of the hydrologic model under the defined values of Nash and $R^2$ can be rated as per the recommendations by Fernandez et al. (2005) and Moriasi et al. (2007) (Table 2.1). Practical recommendations for future research and water resource management have been provided. These include specific areas for further investigation and actionable advice for policymakers to improve flood management and mitigation strategies.

Table 2.1: General performance ratings for recommended statistics

| Nash | $R^2$ | Rating |
|---|---|---|
| 75.0% Nash ≤ 100% | 75.0 % < $R^2$ ≤ 100 % | Very good |
| 60.0 % < Nash ≤ 75.0% | 60.0% <$R^2$ ≤ 75.0 % | Good |
| 36.0 % <Nash ≤ 60.0 % | 50.0% <$R^2$ ≤ 6 0.0 % | Satisfactory |
| 0.0 % <Nash ≤ 36.0 % | 25.0 % < $R^2$ ≤ 50.0 % | Poor |
| Nash ≤ 0.0% | $R^2$ ≤ 25.0 % | Inappropriate |

*Note.* Modified from Van Liew et al. (2003) and Singh et al. (2004).

This study discusses the use of the HEC-HMS model for flood simulation (Ikhwali et al., 2022; Ismael et al., 2017). While HEC-HMS is a widely used tool, the review does not sufficiently critique its limitations, such as its performance in different geographic settings or its dependency on high-quality input data, which may not always be available. For example, the HEC-HMS model, though superior in its physical basis compared to other lumped models, relies heavily on detailed and accurate input data, including digital elevation models (DEM), land use/land cover (LULC), and soil maps. In regions where such data is scarce or outdated, the model's accuracy can be significantly compromised.

However, the application of the HEC-HMS model in different hydrological and climatic contexts has revealed varying degrees of effectiveness. Studies have shown that the model's calibration and validation processes are essential to ensure reliable simulation results (Dariane

et al., 2016; Odyuo, 2020). However, manual calibration, often employed in these studies, is not only labor-intensive but also subjective, leading to potential inconsistencies in the results. Auto-calibration methods, while more systematic, may still fall short in capturing the unique hydrological characteristics of diverse river basins.

Despite these challenges, HEC-HMS remains a popular choice for flood modeling due to its balance between physical representation and data requirements. Its integration with GIS through the HEC-GeoHMS extension further enhances its utility by allowing for detailed spatial analysis and visualization. However, further studies need to be done to address the need for continuous updates and improvements in input data and the calibration process to maintain the model's accuracy and applicability across different settings.

### 2.3 Flood Frequency and Probability Distribution Models

Flood frequency analysis and probability distribution models involve the concept of modelling the relationship between magnitude and recurrence interval from the available data, using various techniques to estimate at least roughly, how big the projected flood is, in a decade, century, or a millennium (Lang et al., 2010; Rizwan et al., 2018). According to Şen (2018) and Zischg et al. (2018), a flood can be characterized by magnitude, how severe or large a flood event is. Discharge is the common measure of flood magnitude or how much water in the streambed is flowing past a certain point in a given duration (Ghosh, 2013; Zischg & Bermúdez, 2020). To estimate the flood magnitude, total runoff volumes are measured for the flood hydrograph above the baseline discharge using a threshold discharge of a specific recurrence interval flow.

The calculated magnitude value becomes the runoff volume accumulated, compared to the flood volume of the record. To obtain total runoff volumes, daily runoff values are summed once discharge exceeds the threshold, and the sum is multiplied by contributing watershed area to measure total flood runoff volume. Therefore, it implies that annual peak-flow data collected or simulated provides an estimate of the flood magnitude as well as the frequency at a particular basin outlet. A longer-duration flood exhibits a higher flood magnitude, which would mean that flood volumes would be larger from the larger watersheds if other factors are constant (Gotvald et al., 2012). Nonetheless, Ghosh (2013) and Şen (2018) hold the view that river and watershed size does not affect the magnitude number.

### 2.3.1 Flood Frequency

The terminology associated with flood frequency estimations is shifting away from the T-year recurrence interval flood to the P-percent annual exceedance probability (AEP) flood (Gotvald et al., 2012). AEP is currently preferred because it conveys the probability, or odds, of a flood of a given magnitude being equalled or exceeded in any given year. For instance, a 1-percent AEP flood usually known as the 100-year flood corresponds to the flow magnitude that has a 0.01 probability of being equalled or exceeded in any particular year. The P-percent annual exceedance probability is calculated as the reciprocal of the recurrence interval (T) multiplied by 100.

Recurrence intervals (flood frequencies) or return period which means the average number of years between one flood of a given magnitude and the next flood that is as big or bigger can be described statically as follows.

Taking ($X$) as a random variable that has a cumulative distribution function $F_x(x)$. The probability that ($X$) is less than equal to a given event ($x_p$) is given as:

$$F_x(x) = P\left(X \leq x_p\right) = p \tag{2.3}$$

The probability that this event will be exceeded is then equal to 1-$p$ and the percent exceedance is denoted as 100(1-$p$). For such an event $x_p$, the return period corresponding to this exceedance probability is denoted by $T$.

Here, $$T = \frac{1}{(1-P)} \tag{2.4}$$

From this statistical description, the 100-year return period of the flood may then be understood as an event with a probability of exceedance 1-$p$ = (0.01) or a non-exceedance probability ($p$ = 0.99). Moreover, it would also imply that there is a 99 % chance that this event will not be exceeded within a particular year.

### 2.3.2 Probability Distribution Functions (PDFs)

Table 2.2 shows the probability distribution models and their parameters, which are usually applied in flood forecasting.

Table 2.2: The parameters and PD functions

| Number | Distribution | Probability Density Function | Parameters |
|---|---|---|---|
| 1 | GAM | $F(x) = \frac{x^{\alpha-1}\exp\left(\frac{-x}{\beta}\right)}{\beta^{\alpha}\Gamma(\alpha)}$ <br> Γ is the gamma function | shape parameter = α <br> scale parameter = β |
| 2 | GEV | $F(x) = \exp\{-\exp(-y)\}$ <br> $y = \kappa^{-1}\log\{1 - \kappa(x - \xi)/\alpha\}$ | shape parameter = κ <br> scale parameter = α <br> location parameter = ξ |
| 3 | GLO | $(x) = \frac{1}{\left\{1+\exp\left(-y\right)\right\}}$ <br> $y = -\kappa^{-1}\log\{1 - k(x - \xi)/\alpha\}$ | shape parameter = κ <br> scale parameter = α <br> location parameter = ξ |
| 4 | GNO | $F(x) = \varphi(y)$ <br> $y = -\kappa^{-1}\log\{1 - k(x - \xi)/\alpha\}$ | shape parameter = κ <br> scale parameter = α <br> location parameter = ξ |
| 5 | GPA | $F(x) = 1 - \exp(-y)$ <br> $y = -k^{-1}\log\left\{1 - \frac{k(x-\xi)}{\alpha}\right\}$ | shape parameter = κ <br> scale parameter = α <br> location parameter = ξ |
| 6 | GUM | $F(x) = \exp[-\exp\{-(x - \xi)/\alpha\}]$ | scale parameter = α <br> location parameter = ξ |
| 7 | P3 | $F(x) = \frac{\lvert x-\xi\rvert^{\alpha-1}\exp\left(-\lvert x-\xi\rvert/\beta\right)}{\beta^{\alpha}\Gamma(\alpha)}$ <br> If γ 6= 0, then let $\alpha = \frac{4}{\gamma^2}/, \beta = \frac{1}{2}\sigma\lvert\gamma\rvert$, <br> and ξ = μ− 2σ/γ | scale parameter = σ <br> shape parameter = γ <br> location parameter = μ |

*Note.* Modified from Rizwan et al. (2018).

The major challenge in flood forecasting using the statistical distribution models is the selection of the relevant probability distribution (Rizwan et al., 2018). Furthermore, researchers recommend the user selection of a distribution, which works best for a specific site after a trial of multiple probability distribution curves (Rizwan et al., 2018; Xiong et al., 2018). In the recent past, some commonly applied PDs entail generalized extreme value (GEV), Pearson type III (P3), log Pearson type III (LP3), generalized Pareto (GPA), generalized logistic (GLO), generalized normal (GNO), exponential (EXP), gamma (GAM), Weibull (WEI), and Gumbel (GUM) (Vivekanandan, 2015).

Some countries have adopted specific standard PD for flood frequency analysis. For instance, the United States of America uses log Pearson type III (Stedinger & Griffis, 2008). China adopted Pearson type III, and Europe prefers generalized extreme value distribution (Rizwan et al., 2018). Therefore, a lack of global and regional standard probability distribution models has restrained hydrologists from using a generic distribution globally. The procedure of identifying the most reliable probability distribution model requires a goodness-of-fit test since various flood characteristics in different rivers and the availability of a wide range of selection criteria demand the selection of PDs from a wide range of available distribution.

The general framework for estimating flood magnitude, which accounts for climate variability and land-use change effects is shown in Figure 2.4.

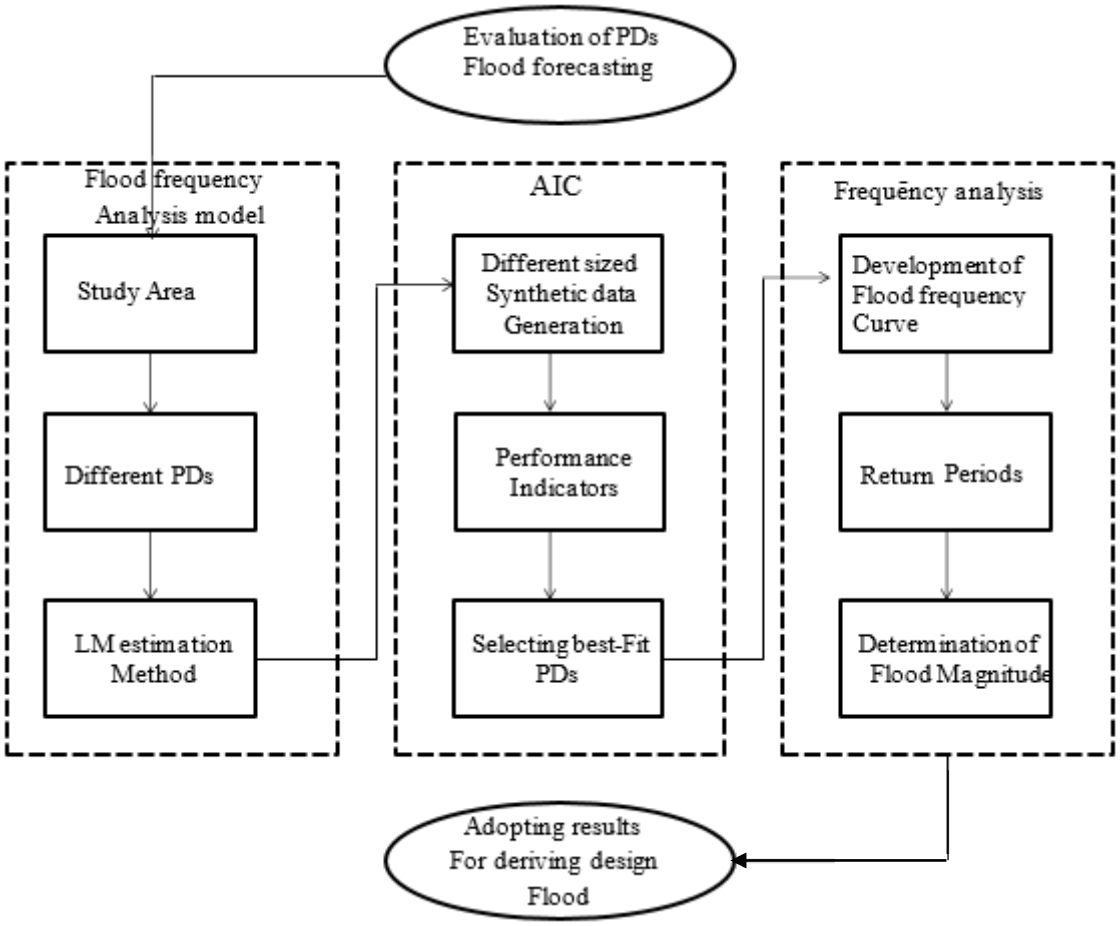



Figure 2.4: Flowchart of the procedure for flood magnitude estimation using the PDFs

While existing literature extensively discusses the methodologies and applications of flood frequency and probability distribution models, there are key areas that require further exploration. Many studies do not adequately address how the performance of these models varies across different geographic settings. This is crucial as factors such as climate,

topography, and land use significantly impact flood behavior and model accuracy. The dependency of these models on high-quality input data is often understated. In regions where data is sparse or of poor quality, the reliability of flood predictions can be severely compromised.

The potential benefits of integrating flood frequency models with other hydrological or climate models are not fully explored. Such integrations could enhance predictive capabilities and provide a more comprehensive understanding of flood dynamics. The reviewed literature provides a solid foundation for understanding flood frequency and probability distribution models. However, Future research should focus on the gaps highlighted above to enhance reliability and applicability.

### 2.4 Forecasting of Flood Events

The first activity in flood forecasting is the use of models or statistical tools to predict future flood magnitudes or peak discharges. Then the relevant PD curve selected was used for estimation of frequencies of the forecasted flood magnitudes. Although many techniques have been proposed and applied to predict hydrologic events, artificial neural networks have been preferred and most used (Das & Ghosh, 2018). ANN models may be a better option because of their power of adaptive learning, real-time operation, and fault tolerance, and they are used in modelling nonlinear and complex phenomena (Lee & Tuan, 2016; Ozoegwu, 2019; Somvanshi et al., 2006). Their capability comes from pattern capturing and statistical parallel processing of historical data (Ozoegwu, 2019).

Artificial neural network models have been widely applied in hydrological studies including flood predictions (Mitra et al., 2016; Mukerji et al., 2009; Rezaeianzadeh et al., 2014; Tiwari & Chatterjee, 2010;). These studies have indicated that ANN models perform better than other statistical modelling techniques because of their robust performance in dealing with noisy input patterns, and the ability to generalize from the input data, which make them a reliable option for probability predictions.

#### 2.4.1 Artificial Neural Network

An artificial neural network can be described as a mathematical structure capable of representing the arbitrary, complex, and non-linear process correlating the input and output of any system. The motivation for the development of neural network technology stemmed from the desire to develop an artificial system that could perform "intelligent" tasks similar to those performed by the human brain (Herreyre et al., 2023; Wambua, 2020).

ANN is a black box-type lumped model, which obtains knowledge through a pragmatic approach, which involves identifying a set of weights for the connections and boundary values referred to as biases, for the neurons (Sharghi et al., 2018; Wambua, 2019). Figure 2.5 shows typical artificial nodes with inputs ($X_1$, $X_2$, …, $X_n$) connected to neuron j with weights ($W_{1j}$, $W_{2j}$, …, $W_{nj}$) on each connection.

The neurons sum all the signals they receive, with each signal being multiplied by its associated weights on the connection. ANN output (Q) is then processed through an activation function f (.), which is usually non-linear to give the final output (Y).

As explained by Barua (2010), the input and output signal relationship within the neural networks can be expressed as expressed in the following function:

$$Y = f(Q) = f\left(\sum_{i}^{n} W_i X_i + b_k\right) \tag{2.14}$$

where, $X_i$ is the input signal $i$, $W_i$ is the weight attached to the input signal $i$, n is the number of input signals, $b_k$ is the bias at the cell of the body, $f$ is the activation function, and $Y$ is the output.

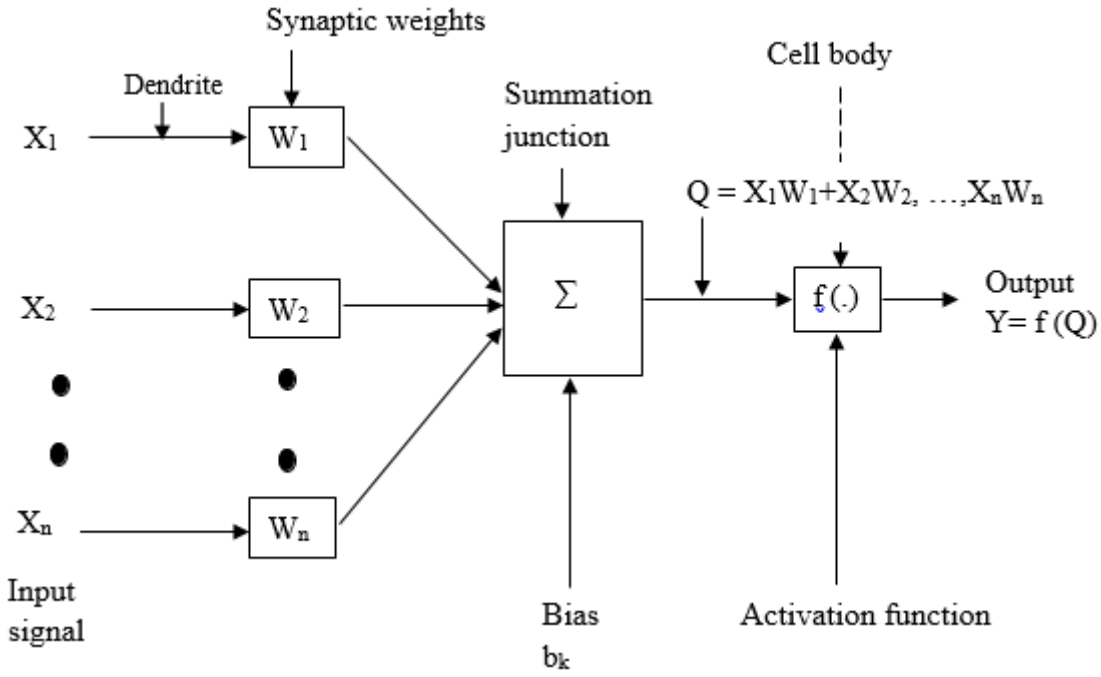



Figure 2.5: A typical Artificial Neurons Architecture

Source: Gurney, (2020).

The neurons are arranged in layers and are the components, which maps input to output. Weight (w) is part of what determines how a network behaves. Neural network knowledge is stored

within interneuron connection strengths known as synaptic weights. The activation function used in ANN is dependent on the type, learning algorithm (Levenberg-Marquardt, Bayesian Regularization, BFGS Quasi-Newton, Gradient Descent with Momentum, and Gradient Descent), and scaling approach used. The most important (and probably most used) transfer function is the sigmoid (logistic) function because of its easily differentiable properties, which is very convenient when the back-propagation algorithm is applied (Özbay & Tezel, 2010).

Figure 2.6 shows the most common transfer (activation) functions of artificial neural networks as illustrated by Bennett et al. (2013) and Mishra and Desai (2006).

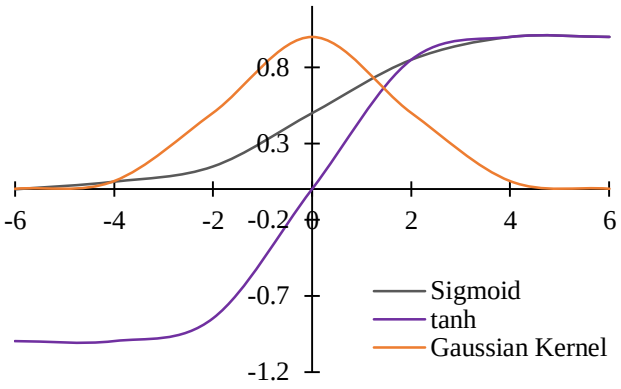



Figure 2.6: Types of Artificial Neural Network activation (transfer) functions

Artificial neural network models may be classified into feed-forward networks and recurrent neurons (Mishra & Desai, 2006). The recurrent network contains loops because of the feedback connections and they include neural network time series applications for future scenario predictions. Unlike recurrent neuron networks, the feed-forward networks have neurons organized into layers having unidirectional connections, thus characterized by neither cycles nor loops in the network. They are used in pattern recognition, curve fitting, forecasting, classifications, and data clustering. Feed-forward neural networks have been applied widely in solving real and complex hydrologic problems with great accuracy specifically the feedforward backpropagation models (Kuok et al., 2010; Peng et al., 2017).

### 2.4.2 Coupling ANN with the Probability Distribution Models

Flood forecasting is normally done by the use of statistical probability distribution curves that encompass the prediction of recurrence interval corresponding to specific magnitude (Rizwan et al., 2018). The best probability distribution curve is selected from the existing statistical distributions such as Gumbel, Normal, Log-normal, Exponential, Weibull, Pearson, and Log-Pearson. These graphs are then used to estimate the design flow values corresponding to

specific return periods, which can be used for hydrologic planning purposes. Therefore, relevant PD curves can then be applied to estimate the P-percent annual exceedance probability of the flood magnitudes predicted by artificial neural networks. The option of coupling models in hydrology has been observed to improve prediction efficiency (Chen et al., 2011; Guo et al., 2021; Khashei et al., 2010).

# CHAPTER THREE

# MATERIALS AND METHODS

## 3.1 The Study Area

The Gucha-Migori River Basin is an extensive basin cutting across five counties in Kenya namely, Narok, Homabay, Migori, Nyamira, and Kisii. It is located between longitudes $34^0$ 7' 48.08'' E and $35^0$ 1' 36.6'' E and latitudes $0^0$ 21' 27.09'' S and $0^0$ 32' 59.24" S and has a total length of 147 km (Figure 3.1).

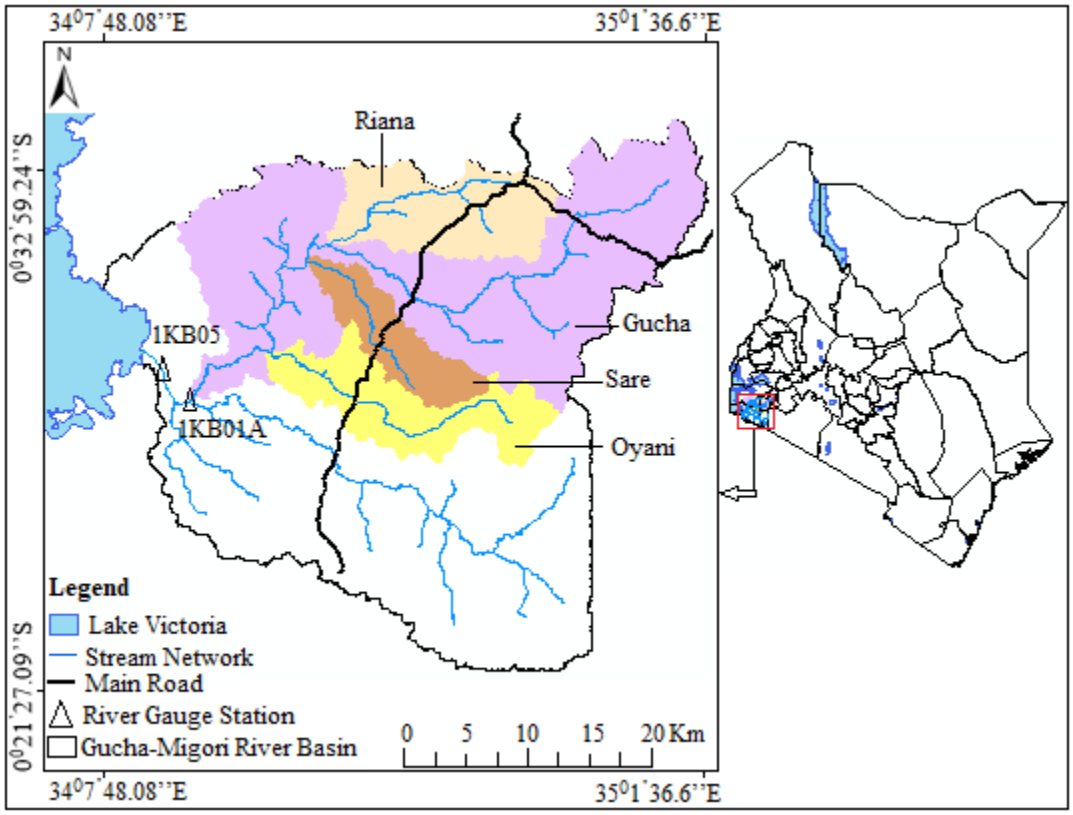



Figure 3.1: Gucha-Migori River Basin

Gucha-Migori River Basin consists of a two-river system, which is R. Gucha and R. Migori. These two rivers merge at a confluence in the Sango area in Central Kadem Location in Nyatike district to become the wide R. Gucha-Migori. The river thereafter flows and pours its water into Lake Victoria. The basin is average 2000 m above sea level but 3000 m above sea level at its source in Nyamira. R. Gucha has its source in Nyamira County in the Kisii highlands while R. Migori has its source in Narok County specifically in the Emuria Dikiri district. The climate in the Gucha River catchment is influenced to some extent by the south-easterly and north-easterly monsoon and can be classified as a modified equatorial climate. Gucha-Migori River Basin experiences two rainy seasons with the highest rainfall being between March and May.

The highlands area has rainfall of up to 2000 mm, while the lower Gucha region's rainfall is as low as 700 mm annually. Temperatures range from 15 $^{o}$C to 20 $^{o}$C within the highlands of Nyamira and 21 $^{o}$C to 30 $^{o}$C in the lowlands towards Lake Victoria. There are three River Gauging Station in the catchment; 1KB07, 1KB05, and 1KB01A. Sloping system of the basin is divided into three major categories i.e. upstream slope system of 25 % to over 40 % slope, midstream slope system of 10 % to 20 % slope, and downstream slope system of 0 % to 10 % slope. Kisii and Nyamira highlands consist of reddish fertile volcanic loamy soils while the downstream of the South Nyanza region consists of greyish fertile alluvial soils (Ogembo, 2018).

### 3.1.1 Spatial Data Collection

The satellite data and other geospatial data products that were downloaded and used for hydrologic modelling using the HEC-HMS model are presented in Table 3.1.

Table 3.1: Satellite remote sensing and Geospatial data products

| Data | Feature values & Source |
|---|---|
| Digital Elevation Model | Earth Explorer website (https://earthexplorer.usgs.gov/). The 30 m spatial resolution (SRTM - DEM) |
| Soil | Food and Agriculture Organisation (FAO) data |
| Landsat imagery (1976) Scene size (185 x 185 Km) | Earth Explorer website (https://earthexplorer.usgs.gov/). Satellite (L1-5), Sensor (MSS), Grid cell size reflective (60 m), Map projection level – 1 (UTM), Datum (WGS 84), WRS Path/WRS row (181/60). For raster MTL.text 4Bands (Green, Red, Near Infrared 1, Infrared 2) |

### 3.1.2 Climate Variables, Hydrologic, and Field Data Acquisition

The following data were obtained from the Water Resources Authority and the fieldwork:

i) Daily historical meteorological data: Precipitation, mean daily surface air temperatures (from 1969 to 2015) (Figure 3B, APPENDIX B).
ii) Daily historical hydrologic data: River discharge and water level data (from 1969 to 2015) (Figures 1B & 2B, APPENDIX B).

iii) Coordinates of streams, reach, and the location of river gauging stations using Global Positioning System (GPS) at different points.
iv) Coordinates of different land-use systems for supervised and unsupervised classification.

### 3.2 Modelling Response of River Discharge to Precipitation Variability

#### 3.2.1 Precipitation Variability

The study began with an assessment of seasonal precipitation anomalies (January to March (JFM), April to June (AMJ), July to September (JAS), and October to December (OND)) in the Gucha-Migori River Basin spanning the years 1969 to 2015. Precipitation anomalies were computed using Equation 3.1, following established climatological methods (Coelho et al., 2012; Raible et al., 2003; WMO, 2011).

$$\text{Anomaly} = X - \text{Xbar} \quad (3.1)$$

where, X is actual annual value of a climate variable, and Xbar is the long-term average value of the variable (30~40 years).

#### 3.2.2 Modelling of River Discharge Using HEC-HMS

Hydrologic modeling of flood events was conducted using the HEC-HMS model, chosen for its superior physical basis compared to other lumped models (Ismail et al., 2022; Natarajan & Radhakrishnan, 2020). The aim was to simulate the response of river discharge to precipitation variability, making the HEC-HMS model particularly relevant. The methodology involved preparing input data encompassing geospatial, hydrologic, and meteorological data. Subsequently, multiple model runs were conducted to compute runoff, base flow, and channel flow. Parametrization, calibration, validation, and model performance assessment were then performed to simulate and analyze flood events effectively.

#### 3.2.3 Processing of Digital Elevation Model Data

The digital elevation model (DEM) processing was implemented using HEC-GeoHMS and the ArcMap platform. The Shuttle Radar Topography Mission (SRTM) DEM with a spatial resolution of 30 meters was clipped and projected onto Universal Transverse Mercator (UTM) based on the WGS 84/UTM zone 36S (Figure 4B, APPENDIX B).

Using the arc hydro and GeoHMS tools, morphometric and topological characteristics of the basin that were derived entail stream network, sub-basin boundaries, flow direction, flow accumulation, and outlets (Figure 5B, APPENDIX B) (Table 5A, APPENDIX A). The

derivation of the aforementioned attributes was accomplished by the use of standard procedures based on the D8 (eight-pour-point) algorithm as further explained by Olang and Fürst (2011).

The digital elevation model was crucial for accurately modeling hydrological processes within the Gucha-Migori River Basin. It provided the foundational datasets necessary for delineating watershed boundaries and stream networks, which was integral for flood modeling. The DEM processing facilitated the identification of critical basin attributes, enabling a more precise analysis of flood dynamics and potential intervention points.

### 3.2.4 Processing of Soil Data and Land Use - Land Cover Maps Datasets

In this study, soil data obtained from FAO was also clipped to cover the Gucha-Migori River Basin, then projected to WGS 84/UTM zone 36s, and finally rasterized in ArcMap applications (Figure 6B, APPENDIX B). The soil dataset was used to derive the major hydrological soil groups (HSG) with the help of the FAO/UNESCO revised manual for soil maps of the world as demonstrated by Batjes and Gicheru (2004).

The obtained satellite imagery of 1976 was clipped, resampled, and projected (WGS 84/UTM zone 36s) to cover the Gucha-Migori River Basin (Figure 7B, APPENDIX B).

The process of classifying land use and land cover involved using satellite imagery obtained from the Landsat 1 MSS sensor, which had four channels, including two infrared and two visible channels (Table 1A, 2A, & 3A, APPENDIX A; Figure 8B, APPENDIX B). The downloaded satellite images underwent initial pre-processing steps, such as clipping and resampling. Subsequently, they were projected to the WGS 84/UTM zone 36S coordinate system. Following this, various procedures were carried out in ArcMap version 10.4 to achieve supervised classification for the land cover, as illustrated in Figure 3.3.

These procedures included layer stacking, sub-setting, geometric correction, and ground-truthing activities as detailed in Table 4A, APPENDIX A), image enhancement, and Maximum Likelihood classification. The ground truthing of satellite imagery obtained from the Landsat 1 MSS sensor was conducted to validate the accuracy of LULC classification. Representative sample locations were selected across the study area to cover diverse and major land cover types still existing from 1976, including forest, water bodies, agricultural land, and settlements. Further references were made to the existing literature on LULC in the Gucha-Migori River Basin, such as the analysis by Muiruri et al. (2014) and Ogembo (2018), along with field

observations. GPS devices were used to record the coordinates of each location, and the land cover type was documented, supported by photographs.

The collected data was imported into GIS software and overlaid on the Landsat imagery for comparison with the classified land cover. A confusion matrix was generated to assess classification accuracy, calculating metrics such as overall accuracy, producer's accuracy, and user's accuracy. Discrepancies between the field data and the classified imagery addressed by refining the classification method and training classes. This procedure ensured the LULC classification aligned with actual ground conditions. The LULC categories that were classified included Cropland/Natural Vegetation, Tree Cover, Settlement/Bare Land, and Water.

### 3.2.5 Preparation of Hydrologic and Meteorological Datasets

Simulation of flood routing processes using the HEC-HMS model used daily precipitation and discharge datasets. Besides flood events simulation, river discharge datasets were also used for model calibration and validation, thus, a comparison was done between the acquired precipitation and runoff data by use of graphical plots for the test of uncertainties in the recordings. The daily precipitation and discharge data records for the period 1969 to 2015 were prepared in a Notepad as required for HEC-HMS applications and as guided by the HEC-HMS manual and other related studies (Halwatura & Najim, 2013; Nyaupane et al., 2018).

### 3.2.6 HEC-HMS Set-up and Run

The whole process of river discharge simulation in this study using HEC-GeoHMS and HEC-HMS is presented in a flow chart (Figure 3.6). For the simulation, the Natural Resource Conservation Service-Curve Number (NRCS-CN) model, Clark's Unit Hydrograph model, Exponential recession base flow model, and Muskingum–Cunge flow routing model were used. As explained by Doan (2000), the model elements for the basin were derived using Geo-HMS, a GIS extension to support the HEC-HMS. The Soil Conservation Service-Curve Number as expressed in Equation 3.2 was used to model the transformation of precipitation excess into direct surface runoff.

$$Q = \frac{\left(P - I_a\right)^2}{\left(P - I_a\right) + S} \tag{3.2}$$

where, $P$ is the total precipitation (mm), $Q$ is the actual Runoff (mm), $S$ is the potential maximum retention of soil moisture after surface runoff begins and $I_a$ is the initial abstraction (mm) which represents all losses before the runoff begins.

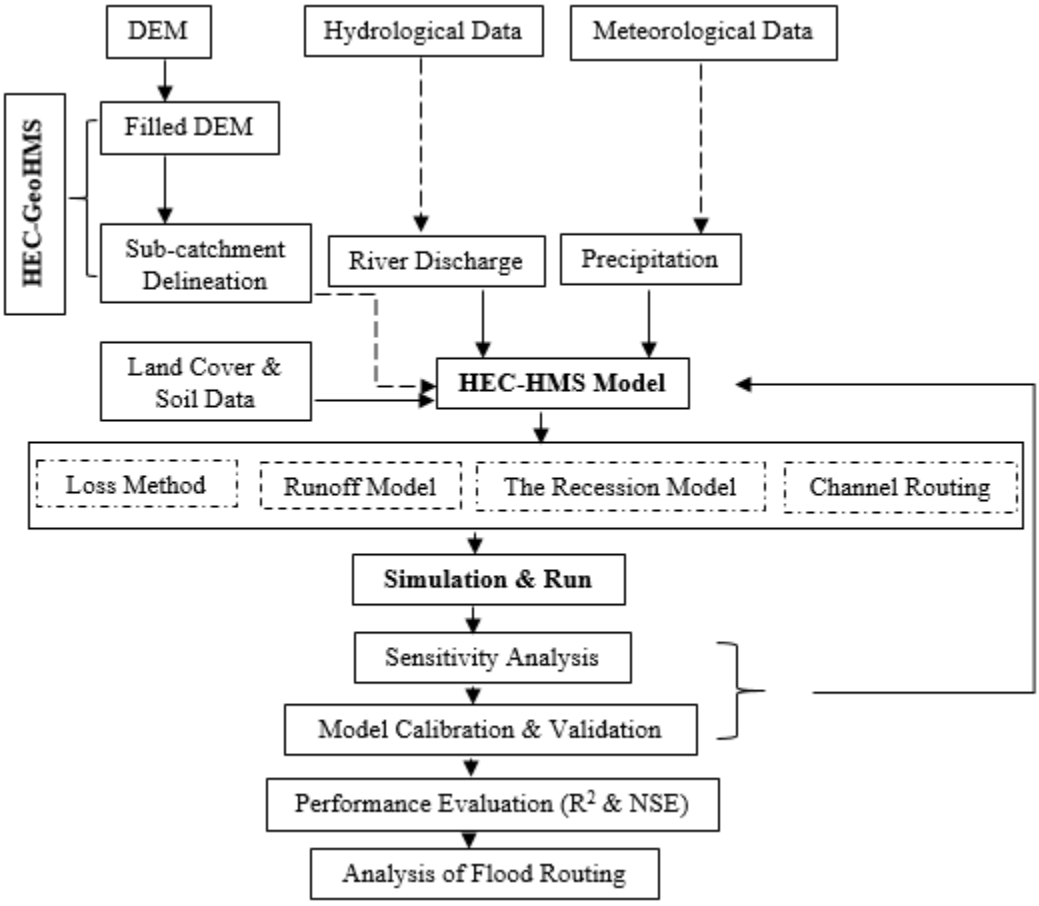



Figure 3.2: The Flow chart showing Modelling of Flood Routing by HEC-HMS

The Curve Number and potential maximum retention of soil moisture after surface runoff begins (*S*) are related by;

$$S = \frac{1000}{CN} - 10 \qquad (3.3)$$

SCS-CN approximates runoff volumes using the physically-based dimensionless curve number (CN) parameter, which will be derived from the LULC and soil maps. Thus, CN was defined such that; and represents impervious and water surface curve numbers, and represents natural surface curve numbers. A lookup table for retrieving CN approximation was created for normal antecedent moisture conditions (AMC II) with the help of standard tables as a guide pointed out by Olang and Fürst (2011). The assumptions for initial abstractions of the maximum potential retention of the soil assumptions were carried out according to the recommendations by Doan (2000). Practical recommendations for future research and water resource management have been provided. These include specific areas for further investigation and actionable advice for policymakers to improve flood management and mitigation strategies.

Transformation of the generated runoffs into corresponding hydrographs was achieved using Clark's Unit Hydrograph concept. The application of Clark's UH requires the determination of the times of concentration (Equation 3.4) and the storage coefficient. These parameters were established from the physical catchment conditions based on procedures for the lag time (Equation 3.5) and storage coefficient.

$$T_c = K L^{0.77} S^{-0.385} \tag{3.4}$$

where, $T_c$ is the time of concentration (day), $S$ is the mean slope of the main river (m/m), $K$ is 0.0078, and $L$ is the length of main river (m).

Basin lag time can be given as:

$$t_{lag} = \frac{1}{1900 Y^{0.5}} \left( L^{0.8} \times \left( \frac{1000}{CN} - 10 + 1 \right)^{0.7} \right) \tag{3.5}$$

where, $L$ is the flow length (m), $CN$ is the average curve number, and $Y$ is the average slope gradient (%) of the basin.

The exponential recession model was applied to represent base flow processes. A similar approach outlined by Olang and Fürst (2011) was used to estimate the recession constant based on the relative sizes of the sub-catchments. The relationship between base flow $Q_t$ at any time $t$ to an initial value $Q_0$ ,where, k is an exponential recession constant can be expressed as follows (De Silva et al., 2014):

$$Q_t = Q_0 k^t \tag{3.6}$$

Flood and channel flow routing was performed using the Muskingum–Cunge hydrologic approach (Fleming, 2004). The Muskingum-Cunge equation adopted for this study is presented in Equation (3.7).

$$Q_2 = C_0 I_2 + C_1 I_1 + C_2 Q_1 \tag{3.7}$$

$$C_0 = \frac{0.5\Delta t - Kx}{0.5\Delta t + K - Kx} \tag{3.8a}$$

$$C_1 = \frac{0.5\Delta t + Kx}{0.5\Delta t + K - Kx} \tag{3.8b}$$

$$C_2 = \frac{-0.5\Delta t + K - Kx}{0.5\Delta t + K - Kx} \tag{3.8c}$$

where, $K$ is the time for the flood wave to pass through the river, $x$ is the flow-specific gravity factor, the value range is 0 to 0.5, and $I_1$ and $I_2$ are the initial and final flow rates of the upper section of the river channel. $Q_1$ and $Q_2$ are the initial and final flow rates of the lower section of the river channel.

### 3.2.7 HEC-HMS Model Sensitivity Analysis

After the trial runs, sensitivity analyses of the parameters were based on the influence of initial loss abstraction (Ia), lag time (tlag), recession constant (k), flood travel time (K), and Muskingum weighting factor (X) on the peak river discharge. The aforementioned models as illustrated in Equation 3.2, 3.4, 3.6, and 3.7 were also selected for uncertainty analysis, validation, and calibration as per the Du et al. (2012) outline. Sensitivity analysis was a fundamental component of HEC-HMS calibration that enhanced the identification of influential parameters. Each of the optimized parameters was used as an input into the model to obtain the desired simulation. The most sensitive parameters of the model needed to be precisely estimated because a minor change in the value led to a significant variation in the simulated output. The methodology for addressing uncertainty, sensitivity, calibration, and validation has been detailed. Specific techniques used, parameters tested, and evaluation criteria are explained to ensure robustness and reliability of the model.

For the accomplishment of sensitivity analysis, the model was first run using the optimized model parameters obtained after calibration and validation. Then, a one-parameter-at-a-time approach was applied, where the value of each parameter was varied from -25% to +25% in increments of 10%, while keeping all other parameters constant. The output values, particularly the river discharge data, peak flows, and base flows, were analyzed to determine variations with respect to the initial estimates of the parameters. The elasticity ratio, as expressed in Equation 3.9, was used to rank the parameters, where a greater ratio indicates a more highly sensitive variable, and a lower ratio reveals a less sensitive variable.

$$e_e = \frac{\Delta O / O_t}{\Delta I / I_p} = \frac{\%\ \text{change}_{\text{output}}}{\%\ \text{change}_{\text{input}}} \tag{3.9}$$

where, O and I are the input and output variables respectively.

### 3.2.8 HEC-HMS Model Calibration and Validation

HEC-HMS model calibration was to find the best set of parameter values that produce the best fit between simulated and observed values. The calibration and validation of the HEC-HMS

model were conducted with a split sample procedure and discharge measured records collected at the outlet of the basin (1KB05). Calibration and validation of the HMS-HEC model were achieved between the periods (January 1990 and June 1992) and (July 1992 and August 1994) respectively. The model was calibrated both manually and HEC-HMS auto-calibration process because the manual process only provides the range of the parameters while the auto-calibration process optimizes the results (Thu et al., 2019).

### 3.2.9 HEC-HMS Model Performance Assessment

The model performance was evaluated using two test statistics in the HEC-HMS model. The first one is the Coefficient of Determination ($R^2$) while the second one is Nash Sutcliffe model efficiency (Nash) and are expressed in equations (3.10) and (3.11) respectively. The performance assessment was carried out as per the guidelines, outlines, and recommendations provided by Fernandez et al. (2005) and Moriasi et al. (2007). Practical recommendations for future research and water resource management have been provided. These include specific areas for further investigation and actionable advice for policymakers to improve flood management and mitigation strategies.

$$R^2 = \frac{\left[\sum_{i=1}^{n}\left(q_i^{obs}-q_i^{avg}\right)\left(Q_i^{sim}-Q_i^{avg}\right)\right]^2}{\sum_{i=1}^{n}\left(q_i^{obs}-q_i^{avg}\right)^2\sum_{i=1}^{n}\left(Q_i^{sim}-Q_i^{avg}\right)^2} \tag{3.10}$$

$$Nash = 1-\frac{\sum_{i=1}^{n}(q_i^{obs}-Q_i^{sim})^2}{\sum_{i=1}^{n}(q_i^{obs}-Q_i^{sim})^2} \tag{3.11}$$

where, $n$ is the total number of observations, $q_i^{obs}$ is the observed discharge ($m^3/s$), $q_i^{avg}$ is the average of the observed data, $Q_i^{sim}$ is the simulated discharge, and $Q_i^{avg}$ is the average of the simulated data.

### 3.2.10 Discharge Hydrometric Indices and Precipitation Variability Correlation

The correlation between simulated river discharge and precipitation was done by regression analysis. Regression analysis was applied to relate the response of a quantitative dependent variable (seasonal discharges (JFM, AMJ, JAS, and OND)) to a quantitative independent variable (precipitation) in an equation form or model.

The general regression model is given as:

$$Q_1 = a_0 + a_1 h_1 + \varepsilon \tag{3.12}$$

where, $Q_1$ is the dependent variable, $a_0$ and $a_1$ are regression constants, $h_1$ is the independent variable, and $\varepsilon$ is the average value of the random error.

Using the method of least squares, regression constants ( $a_0$ and $a_1$ ), in such a way that the sum of squares of the residuals is minimal.

$$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(Q_1 - a_0 - a_1 h_1\right)^2 \tag{3.13}$$

Differentiating SSE with respect to $a_0$ and $a_1$ to obtain:

$$\frac{\partial (SSE)}{\partial a_0} = -2 \sum_{i=1}^{n} \left(Q_1 - a_0 + a_1 h_1\right) \tag{3.14}$$

$$\frac{\partial (SSE)}{\partial a_1} = -2 \sum_{i=1}^{n} \left(Q_1 - a_0 + a_1 h_1\right) h_1 \tag{3.15}$$

Setting the partial derivatives equal to zero and rearranging the terms to obtain the following normal equations:

$$n a_0 + a_1 \sum_{i=1}^{n} h_1 = \sum_{i=1}^{n} Q_1 \tag{3.16}$$

$$a_0 \sum_{i=1}^{n} h_1 + a_1 \sum_{i=1}^{n} h_1^2 = \sum_{i=1}^{n} h_1 Q_1 \tag{3.17}$$

$$a_0 = \frac{\sum_{i=1}^{n} Q_1 - a_1 \sum_{i=1}^{n} h_1}{n} \tag{3.18}$$

$$a_0 = \overline{Q_1} - a_1 \bar{h}_1 \tag{3.19}$$

$$a_1 = \frac{n \sum_{i=1}^{n} h_i Q_i - \left(\sum_{i=1}^{n} h_i\right)\left(\sum_{i=1}^{n} Q_i\right)}{n \sum_{i=1}^{n} h_i^2 - \left(\sum_{i=1}^{n} h_i\right)^2} \tag{3.20}$$

Using the following notations:

$$S_{hh} = \sum_{i=1}^{n} \left(h_1 - \bar{h}_1\right)^2 = \sum_{i=1}^{n} h_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} h_i\right)^2 \tag{3.21}$$

$$S_{QQ} = \sum_{i=1}^{n} \left(Q_1 - \overline{Q}_1\right)^2 = \sum_{i=1}^{n} Q_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} Q_i\right)^2 \tag{3.22}$$

$$S_{hQ} = \sum_{i=1}^{n}\left(h_1 - \bar{h}_1\right)\left(Q_1 - \bar{Q}_1\right) = \sum_{i=1}^{n} h_i Q_i - \frac{1}{n}\left(\sum_{i=1}^{n} h_1\right)\left(\sum_{i=1}^{n} Q_i\right) \quad (3.23)$$

Then; $a_1 = \frac{S_{hQ}}{S_{hh}}$ (3.24)

For the model or the developed equation, the correlation coefficient (R) was then calculated as;

$$R = \frac{S_{hQ}}{\left(S_{hh}S_{QQ}\right)^{\frac{1}{2}}} \quad (3.25)$$

Computation of inferences related to linear regression and correlation. The errors for $a_0$ and $a_1$, in the simple linear regression, was then be determined as:

$$S_{a_o} = S_{\varepsilon}\left(\frac{\sum h_1^2}{nS_{hh}}\right)^{\frac{1}{2}} \quad (3.26)$$

$$S_{a_1} = \frac{S_{\varepsilon}}{\left(S_{hh}\right)^{\frac{1}{2}}} \quad (3.27)$$

Sum of squares for error (*SSE*) was also calculated as:

$$SSE = S_{QQ} - a_1 S_{hQ} \quad (3.28)$$

In addition, the mean squares for error:

$$MSE = S_{\varepsilon}^2 = \frac{SSE}{df_{\varepsilon}} \quad (3.29)$$

Confidence intervals for $a_0$ and $a_1$, is given by:

$$CI_{a_0} = a_0 \pm t_{\alpha/2, df_{\varepsilon}} S_{\varepsilon}\left(\frac{\sum h_1^2}{nS_{hh}}\right)^{\frac{1}{2}} \quad (3.30)$$

$$CI_{a_1} = a_1 \pm t_{\alpha/2,\, df_{\varepsilon}} \frac{MSE}{\left(S_{hh}\right)^{\frac{1}{2}}} \quad (3.31)$$

Choosing the level of statistical significance at 95% ($\alpha = 0.05$), the hypothesis testing was carried out about $a_0$ and $a_1$, using t-distribution statistical test. The rejection criterion for the null hypothesis ($H_o$) was when the calculated t-value ($t_{cal}$) is equal to, or greater than the critical t-value ($t_{crit}$).

$$t_{cal} = \frac{a_0}{S_\varepsilon}\left(\frac{nS_{hh}}{\sum h_1^2}\right)^{\frac{1}{2}} \quad (3.32)$$

The computation of reference durations was based on the standardized index of annual simulated river discharge, which is a robust integrator of the water balance at the basin scale. It is a method that provides the deviation from a long-term average and can translated into extreme hydrological events when the index is greater than +0.5 or -0.5, following Balme et al. (2006) criteria and illustrations. The hydrometric index was computed for hydrological seasons (JFM, AMJ, JAS, and OND) for the period between 1969 and 2015 (Equation 3.33).

$$SHI = \frac{RD_s - \overline{RD_s}}{\sigma} \quad (3.33)$$

where, $RD_s$ is the average discharge for the season, and $\overline{RD_s}$ is the long-term interannual mean

### 3.3 Developing the Relationship between Flood Magnitude and Frequency

#### 3.3.1 Flood Frequency Analysis

Observed river discharge data was used to analyse the temporal variation of independent flood events from 1969 to 2015. Then, the next procedure was the ranking of flood magnitudes from the biggest to the tiniest event. After ranking had been achieved, return periods or recurrence intervals (frequencies) were calculated using Equation 3.34 as outlined by other studies like Bhagat (2017), Hamed and Rao (2019), and Mujumdar and Kumar (2012).

$$T = \frac{(n+1)}{r} \quad (3.34)$$

where, *T* is the return period, *n* is the number of years in record, and *r* is the rank of a particular magnitude of an event.

To estimate the P-percent Annual Exceedance Probability (AEP) for the flood events in the Gucha-Migori River Basin, the formula described by Olson et al. (2014) was applied (Equation, 3.35).

$$AEP = \left(\frac{1}{T}\right) \times 100 \quad (3.35)$$

### 3.3.2 Modelling Using Probability Distribution Models

For this study, the relationship between flood magnitudes and their respective frequencies was achieved by probability distribution function modelling. Therefore, the whole process was based on the testing of various PD curves and then the selection of the relevant model that provides a better relationship between flood magnitudes and their respective frequencies. The four tested PDMs include Normal Distribution, Log-Normal Distribution, Gumbel Distribution, and Log-person Type III Distribution.

The normal cumulative distribution function applied can be expressed as:

$$X_{N,T} = \overline{X} + K_{N,T}S \tag{3.36}$$

where, $X_{N,T}$ is the predicted discharge, at return period T, $\overline{X}$ is the average annual peak discharge, $K_{N,T}$ is the normal deviation (Z) for the standard normal curve, where area = 0.50 – (1/T), and *S* is the standard deviation, of annual peak discharge.

The lognormal distribution function used can be given as;

$$X_{LN,T} = \overline{X}_1 + K_{LN,T}S_1 \tag{3.37}$$

where, $X_{LN,T}$ is the logarithm of predicted discharge, at return period, $\overline{X}_1$ is the average of annual peak discharge logarithms, $K_{LN,T}$ is the normal deviate (Z), of logarithms for the standard normal curve, where area = 0.50 – (1/T), and $S_1$ is the standard deviation, of logarithms of annual peak discharge.

The following express-guided Gumbel extreme value distribution Peak discharges to be developed:

$$X_{G,T} = \overline{X} + K_{G,T}S \tag{3.38}$$

where, $X_{G,T}$ is the predicted discharge, at return period (T), $\overline{X}$ is the average annual peak discharge, $K_{G,T}$ is a function of return period and sample size, and *S* is the standard deviation of annual peak discharge.

The log-Pearson type III distribution, which applies to nearly all series of natural floods was developed as guided by the expression:

$$X_{LP,T} = \overline{X} + K_{LP,T}S_1 \tag{3.39}$$

where, $X_{LP,T}$ is the logarithm of predicted discharge, at return period (T), $\overline{X}$ is the average of annual peak discharge logarithms, $K_{LP,T}$ is a function of return period and skew coefficient, and $S_1$ is the standard deviation of logarithms of annual peak discharge.

The basic paramters that were calculated include the following, mean (Eq. 3.40), Standard deviation (Eq. 3.41), and Skew coefficient (Eq. 3.42).

$$\overline{X} = \frac{\sum x}{n} \tag{3.40}$$

$$S = \frac{1}{n-1}\left(\sum\left(X - \overline{X}\right)^2\right)^{0.5} \tag{3.41}$$

$$G = \frac{n\sum\left(X - \overline{X}\right)^3}{(n-1)(n-2)S^3} \tag{3.42}$$

where, X is annual peak flow ($m^3/s$) or logarithm of annual peak flow, and n is the length of dataset.

### 3.3.3 Selection of Probability Distribution Model

### 3.3.4 Coefficient of Determination ($R^2$)

A wide range of statistics has been used to evaluate hydrologic event predictions. By far the most widely preferred statistic reported for hydrologic variables is the Coefficient of Determination ($R^2$) (Anupam & Pani, 2020; Kia et al., 2012; Tunas et al., 2022). The R-squared statistical performance matric was applied to measure how the observed versus simulated regression line approaches an ideal match as elucidated by Deo and Şahin (2016), Fairchild et al. (2009), and Moriasi et al. (2015). $R^2$ values range from 0 to 1, with a value closer to 0 or 0 indicating no correlation. In addition, a value closer to 1 indicates relative perfect linear covariation between the two datasets (similar spatial or temporal patterns). It implies that a correlation measurement would suggest a higher accuracy of the model when the value is approaching one. Using $R^2$ (Equation 3.9), the performance assessment of the four test PDCs was rated according to Fernandez et al. (2005) and Moriasi et al. (2007) recommendations. Practical recommendations for future research and water resource management have been provided. These include specific areas for further investigation and actionable advice for policymakers to improve flood management and mitigation strategies.

### 3.3.5 Goodness of Fit (GoF) Test

The concept of GoF test statistics is to evaluate the performance of the probability distribution curve. It was preferred and accepted for checking the reliability of PDMs to observe river discharge or hydrologic variables datasets based on Langat et al. (2019) elaborations. It can be illustrated by $\chi^2_{rs}$ $K_s$ statistics as follows:

$$\chi^2_{rs} = \sum_{f=1}^{N} \frac{\left(O_f(Ql_f) - F_f(Q_f)\right)^2}{F_f(Q_f)} \tag{3.43}$$

where, $O_f(Ql_f)$ is the observed frequency value of the jth class, $N$ is the number of frequency classes, and $F_f(Q_f)$ represents the expected frequency value of the jth class.

$$K_s = \underset{f=1}{\overset{N}{Max}}\left(E_e(Q_f) - E_d(Q_f)\right) \tag{3.44}$$

where, $E_e(Q_f)$ is the empirical CDF of $Q_f$ and $E_d(Q_f)$ is the computed CDF of $Q_f$ as given by Langat et al. (2019). If the calculated values of the GoF test statistics were lower than those of the theoretical values at the chosen significance level, then the model distribution was taken to be acceptable for estimation. The distribution was considered acceptable for the estimation of extremal floods when the computed values of the GoF test statistics were lower than those of the theoretical values at a defined level of significance. Some of the tests that were used for the comparison of the observed and theoretical CDF are discussed below.

### 3.3.6 Anderson-Darling (AD) Test

This test compares an observed PDF to a theoretical PDF and gives more emphasis to the tail of the distribution in the $\chi^2_{rs}$ statistics and thus was useful for the detection of outliers. If the AD statistics were more than a critical value of 2.5018 at a 95 % significance level ($\alpha$=0.05), then the test hypothesis was rejected. The AD test statistics,

$$A^2 = -N - \frac{1}{N}\sum_{i=1}^{N}(2i-1)\times(\ln \mathrm{F}_e(Q_f) + (\ln(1-\mathrm{F}_D(Q_f)) \tag{3.45}$$

where, $A^2$ is the Anderson-Darling test statistics, $\mathrm{F}_e$ is the cumulative distribution function of the specified distribution, and $Q_f$ is the ordered observed data.

### 3.3.7 Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov (KS) test (Langat et al., 2019) statistics is based on the greatest vertical distance from the empirical and theoretical PDFs, and similar to the AD test statistics, the rejection of a hypothesis was done if the KS statistic was more than the critical value 0.20517 at a 95 % significance level of confidence (Equation 3.46):

$$D = Max(\mathrm{F}(Q_l) - \frac{l-1}{N},\ \frac{l}{N}(\mathrm{F}(Q_l)) \tag{3.46}$$

where, $\mathrm{F}(Q_l)$ is the number of points is less than $Q_l$, and they $Q_l$ are ordered from the smallest to the largest value. This represents a step function that increases by 1/N at the value of each ordered data point. F is the theoretical cumulative distribution of the distribution being tested, which must be continuous, and it must be fully specified.

### 3.3.8 Cramer-Von Mises Test

This test statistic, unlike the Anderson-Darling and Kolmogorov-Smirnov tests, considers an observed hydrological time series in increasing order. The Cramer-von Mises statistic $\gamma^2$ is computed as in Equation (3.47) (Langat et al., 2019).

$$\gamma^2 = \sum_{l=1}^{N}\left(\mathrm{F}(Q_l) - \frac{l-0.5}{N}\right)^2 + \frac{1}{12N} \tag{3.47}$$

If the test statistic $\gamma^2$ is more than a critical value of 0.221 at a 95 % significance level ($\alpha$ = 0.05), the test hypothesis is rejected.

### 3.3.9 Best-Fit Distribution Curve

Firstly, classical descriptive statistics of the external data were processed, and a skewness-kurtosis plot for the empirical distribution showing values for common distribution was made to assist in the initial visualization and choice of distribution to fit the data. Firstly, goodness-of-fit (GoF) statistics were computed. The quality of fit was assessed using classical goodness-of-fit statistics (Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling statistics. Model estimation and selection of the best-fit distribution for the site were accomplished using information criteria (AIC) and Bayesian Information Criterion (BIC). Where both model estimation and selection could be simultaneously accomplished.

$$AIC_c = -2LL + 2k + \frac{2k(k-1)}{n-k-1} \tag{3.48}$$

$$AIC_c = n\ln\left(\frac{SSE}{n}\right) + 2k + \frac{2k(k+1)}{n-k-1} + n\ln(2\pi) + n \tag{3.49}$$

where, *AIC* is the information criterion, *n* is the number of observations, and *k* is the number of parameters of the probability distribution model.

$$BIC = -2LL + k\ln n \tag{3.50}$$

$$BIC = n\ln\left(\frac{SSE}{n}\right) + k\ln n + n\ln(2\pi) + 2 \tag{3.51}$$

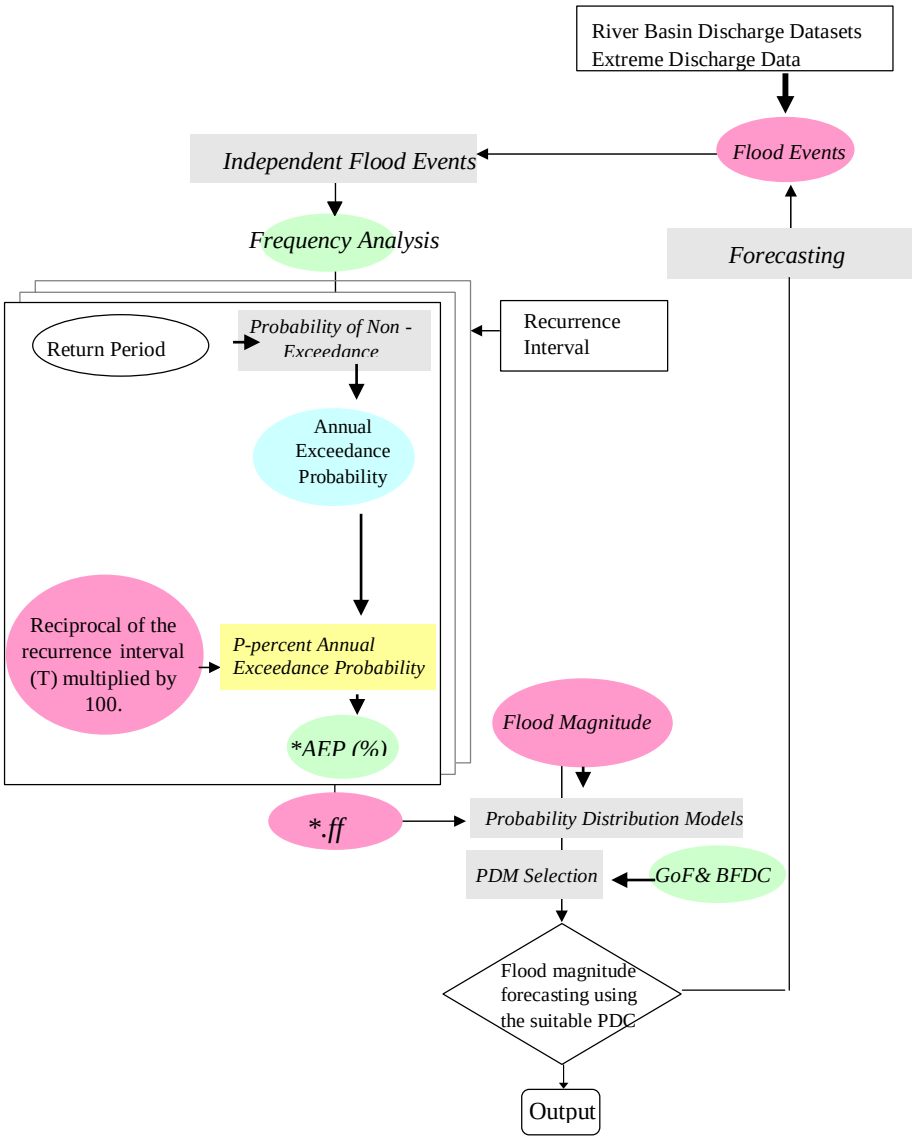
where, *BIC* is the Bayesian Information Criterion, and *SSE* is the residual sum of squares

$$SSE = \sum_{i=1}^{n}\left(\gamma_i - Y_i\right)^2 \tag{3.52}$$

where $\gamma_i$ stands for the i-th data samples, $Y_i$ is the estimated values from the probability distribution curves for the value of $Y_i$ and *n* the upper limit of summation.

### 3.3.10 Estimation of Flood Magnitudes and Probability of Exceedance

The whole procedure and framework used in this evaluation are presented in Figure 3.7.




This figure shows the link between observed extreme river discharges, independent flood events, flood frequency (ff), return period, and the forecasting of the flood magnitudes. Forecasting using suitable PDF.

Figure 3.3: General schematic framework for flood magnitude forecasting

Usually, flood forecasting is done by the use of statistical probability distribution curves that encompass the prediction of recurrence interval corresponding to specific magnitude (Rizwan et al., 2018). In this study, after modelling the relationship between flood magnitude and frequency from the available data using the selected probability distribution model, the curve was then used to estimate the design flow values corresponding to specific return periods. Thus, the identified suitable PDM out of the tested four (Normal, Lognormal, Gumbel, and Log-Pearson Type III Distribution) was then applied to predict the P-percent annual exceedance probability of the flood magnitudes. Some of the periods considered for forecasting include 50,

100, and 200 years, which can provide insights to be applied for hydrologic planning purposes and flood mitigation measures.

### 3.4 Forecasting of Flood Events and their Respective AET

After modelling the relationship between magnitude and frequency from the available data using the selected probability distribution curve, the next procedure followed the forecasting for the period between 2015 and 2052 using NAR and PDF. A nonlinear autoregressive neural network as shown in Figure 3.8 was used for the time series forecasting of the flood magnitudes while the selected probability distribution function was used for the forecasting of the respective recurrence intervals.

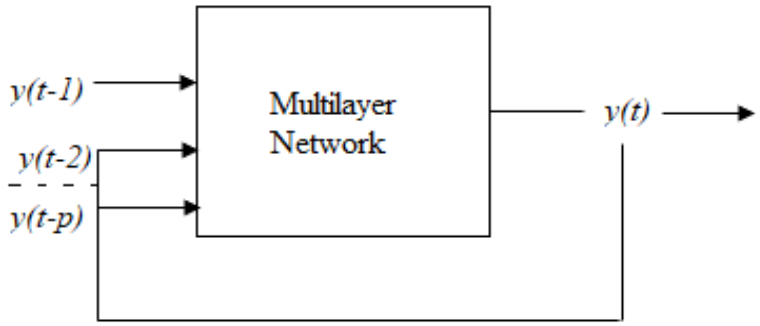



Figure 3.4: The NAR neural network for forecasting flood magnitudes

The forecasting model for flood magnitudes was developed by the past values of flood magnitudes, and this was guided by Equation (3.53). Levenberg-Marquardt backpropagation was adopted as the training algorithm because it is the most preferred learning technique for the Non-linear autoregressive neural network (Alwakeel & Shaaban, 2010).

$$(y(t)) = h\left(y(t-1), y(t-2), \ldots, y(t-p)\right) + \varepsilon(t) \tag{3.53}$$

The p features $y(t\text{-}1), y(t\text{-}2),\ldots,y(t\text{-}p)$, are the feedback delays

Equation (3.53) explains how NAR was applied in forecasting. The value of a data series of a given flood event (y) at time t, y (t), using the p past values of the series. h (·) function was not known in advance. It was approximated by optimization of the network neurons and weight bias. And, ε(t) represents the error of the approximation of the series y at time t. Mean Square Error was used as the objective function for synaptic weight adjustments.

$$MSE = \frac{SSE}{n} \tag{3.54}$$

$$SSE = \sum_{i=1}^{n} \left( \gamma_i - Y_i \right)^2 \tag{3.55}$$

where $\gamma_i$ stands for the i-th data samples, $Y_i$ is the estimated values from the network for the value of $Y_i$ and $n$ the number of data samples for neural network training.

Weights adjustments will be achieved as per Equation 3.56.

$$w_{ij}(n+1) = w_{ij}(n) + \Delta w_{ij}(n) \tag{3.56}$$

where, $\Delta w_{ij}(n)$ is the corresponding weight change, achieved through the jacobian matrix, identity matrix, and scalar factor.

All the modelling computations and programming of the forecasting models development, training, testing, and validation were carried out in MATlab LaBolatory (MATLAB) version 2019 platform (e.g. Figure 10B, 11B, & 12B, APPENDIX B). The model performance of comparing the trained model predictions with the actual values was carried out by the use of statistical error indices. They include Coefficient of Determination and Mean Square Error (MSE) as described in Equations (3.10) and (3.53) respectively. Considering the daily time step forecasting for peak discharge in March, April, May, and June, the performance evaluation of the models developed was rated according to Moustris *et al*., (2010) recommendations. Practical recommendations for future research and water resource management have been provided. These include specific areas for further investigation and actionable advice for policymakers to improve flood management and mitigation strategies.

# CHAPTER FOUR
# RESULTS AND DISCUSSION

## 4.1 Modelling Response of River Discharge to Precipitation Variability

### 4.1.1 Precipitation Variability

The precipitation annual anomalies over the Gucha-Migori River Basin for January, February, and March (JFM) are presented in Figure 4.1.

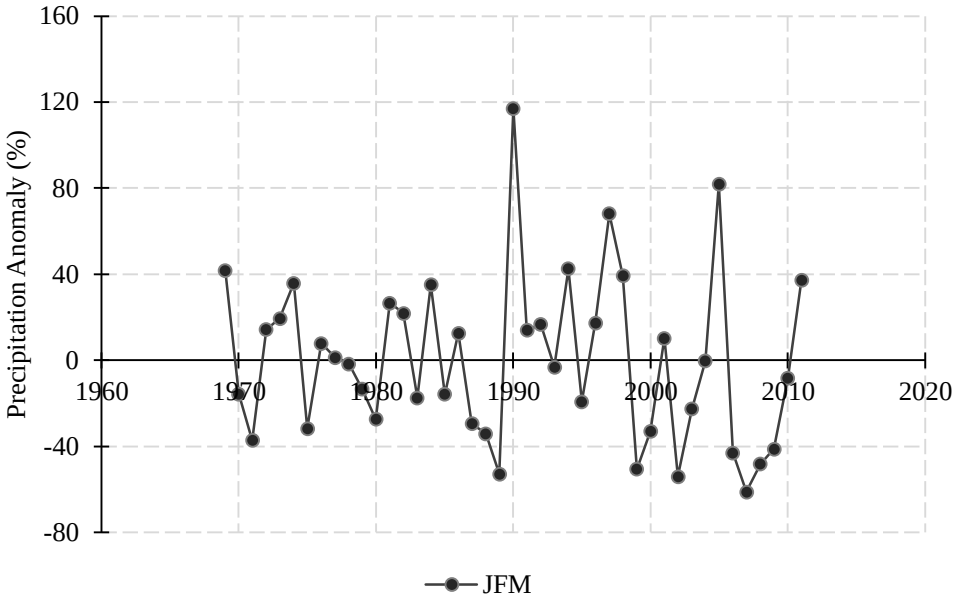



Figure 4.1: Precipitation Anomaly for the Months of January, February, and March

From the results of the average JFM precipitation datasets, the anomalies for the 43 years varied between -61.21 (2007) and 117.15 % (1990) of the long-term average. Values beyond – 35 and + 35 % anomalies threshold were considered as extremely dry and wet climatic conditions respectively. These anomaly ratings were based on the findings of IPCC (2007) and Rwigi et al. (2016) recommendations. However, the precipitation anomalies above 40 % indicated periods associated with river discharges that cause flooding events. Practical recommendations for future research and water resource management have been provided. These include specific areas for further investigation and actionable advice for policymakers to improve flood management and mitigation strategies.

Daily mean precipitation for January, February, and March ranged from 8.23 to 1.47 mm for the period between 1969 and 2015. Out of all the annually recorded precipitation events, 11.63 % (events of above +40 anomalies) represented extremely wet climatic conditions. Eight

independent flood events occurred in January (1973 and 2007), February (2007 and 2010), and four events in March (1969 and 1985) for the period between 1969 and 2015.

The precipitation annual anomalies over the Gucha-Migori River Basin for the months of April, May, and June (AMJ) are shown in Figure 4.2.

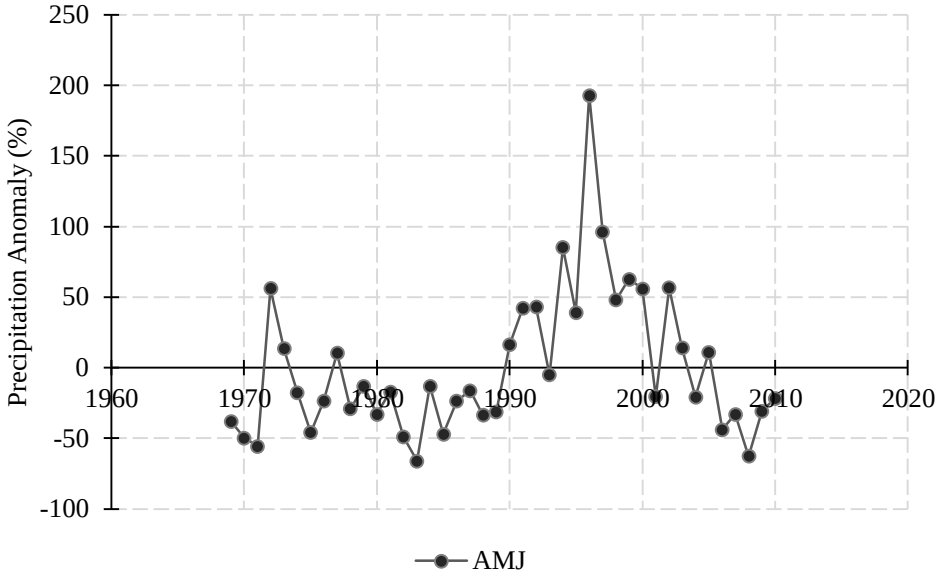



Figure 4.2: Precipitation Anomaly for the Months of April, May, and June

Based on the mean daily precipitation data of the period 1969 and 2015 for April, May, and June, the anomalies ranged between -65.97 (1983) and 192.89 % (1996) of the long-term average. The year 1983 had the lowest (1.15 mm) recorded mean daily precipitation for the three months while 1986 had the highest (9.9 mm). Out of all the annually recorded precipitation events, 23.81 % (events of above +40 anomalies) represented extremely wet climatic conditions. Independent flood events experienced in April include 1(1971), 2(1974), 1(1977), 2(1969), 2(1979), 1(1981), 1(1982), 1(1985), 1(1990), 1(2004), 1(2006), 1 (2011), and 4 (2013) in the study area.

In the case of May, the number of independent flood events recorded include one, three, five, and two occurring in the years 1977, 2012, and 2013, respectively while only 2 in June 1982. Precipitation trends in any of the periods, anomalies, and temporal variation for the months (April, May, and June) showed the manifestation of variability. The variability would affect the spatial as well as temporal response of inflow, storage, and outflow in a catchment hence changes in quantity and dynamics of stream flow (Amisi, 2021; Kundu, 2007; Mwetu, 2019; Ongoma & Chen, 2017; Rwigi, 2016).

The precipitation annual anomalies over the Gucha-Migori River Basin for July, August, and September (JAS) are presented in Figure 4.2.

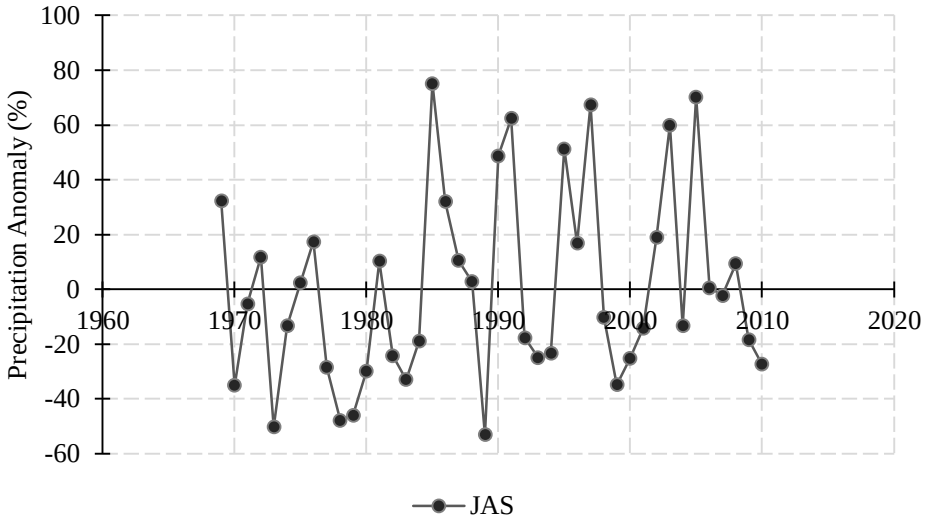



Figure 4.3: Precipitation Anomaly for the Months of July, August, and September

As indicated in Figure 4.3, the mean daily precipitation data of the period 1969 and 2013 for July, August, and September, the anomalies ranged between -53.11 (1989) and +75 % (1985) of the long-term average. The year 1989 had the lowest (1.51 mm) mean daily precipitation for the three months while 1985 had the highest (5.64 mm). Out of all the annually recorded precipitation events, 16.28 % (events of above +40 anomalies) represented wet climatic conditions. There were no independent flood events in July, August, and September for the period 1969 and 2015.

The precipitation anomalies for October, November, and December (OND) are shown in Figure 4.4.

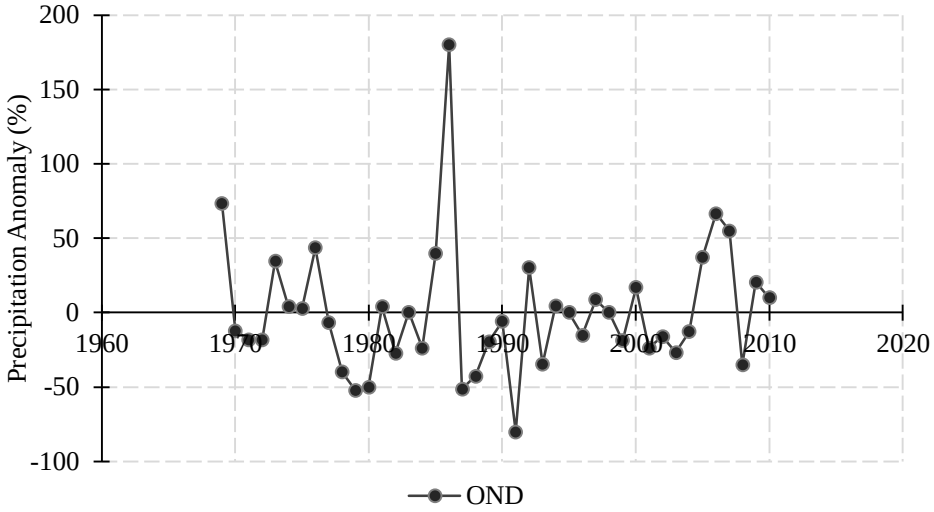


Figure 4.4: Precipitation Anomaly for the Months of October, November, and December

As revealed in Figure 4.4 of the mean daily OND precipitation datasets, the anomalies for the 43 years ranged from -80.20 (1991) and 179.87 % (1986) of the long-term average. Daily mean precipitation for October, November, and December varied from 8.34 to 0.59 mm. Out of all the yearly-recorded precipitation events, 11.63 % (events of above +40 anomalies) represented wet climatic conditions. There was no record of independent flood events in October and November. Nevertheless, six events (3 in 1982 and another 3 in 2011) occurred in December for the period between 1969 and 2015.

The monthly and seasonal precipitation variability at the Gucha-Migori River Basin could be attributed to changes in the Hadley circulation that imply a twice-a-year relocation of the Intertropical Convergence Zone (ITCZ). Dry seasons and months are associated with the effects of low-level north-easterly winds from Egypt and the Arabian Sea over equatorial East Africa in January and February (Camberlin, 2018). The characteristics exhibited by the monthly and seasonal precipitation variability from one year to another (as shown in Figures 4.1, 4.2, 4.3, and 4.4) were sufficient evidence of unreliable changing weather patterns hence the manifestation of extreme hydrologic events. It implies that some months would experience extreme (flooding) or reduced (drought) runoff, subsurface flow, recharge, and return flow, thus river discharge variation. Each figure in the Results and Discussion chapter is accompanied by a detailed explanation, describing what the figure shows, interpreting the data, and explaining its relevance to the study. Figures are well-labeled and referenced in the text.

### 4.1.2 Calibration and Validation of the HEC-HMS Model

Parametrization, calibration, and validation for the river discharge simulation were done at the sub-basin 1 (Figure 4.5 and Figure 4.7). The calibration range and fitted HEC-HMS model parameter values are presented in Table 4.1.

Table 4.1: Initial Parameters from HEC-GeoHMS Used in the HEC-HMS Model

| Parameter Name | Range | Initial Value | Optimized Value |
|---|---|---|---|
| Land Use Curve Number | 35.00~95.00 | 67.10 | 78.00 |
| Lag Time SCS | 0.02~1000.00 | 318.60 | 636.04 |
| Muskingum X-value | 0.17~0.50 | 0.20 | 0.19 |
| Muskingum K-value | 26.77~36.77 | 27.00 | 29.31 |
| Basin Reach | 1.00~14.00 | 1.00 | 2.00 |
| SCS CN – CN ~ Scale Factor | 0.01~100 | 1.00 | 0.01 |

Simulations were accomplished using the same values but the output hydrograph was not a relatively perfect match to that of the observed river discharges (Figures 4.5 and 4.7). Besides, the parameters' ranges (Table 4.1) represent model output, which is an envelope of relatively fair solutions from a lumped process expressed by a certain level of prediction uncertainty. Thus, the disparity could be attributed to uncertainties associated with merging the sub-basins and using the average parameters in the simulation process. The methodology for addressing uncertainty, sensitivity, calibration, and validation has been detailed. Specific techniques used, parameters tested, and evaluation criteria are explained to ensure robustness and reliability of the model. Each figure in the Results and Discussion chapter is accompanied by a detailed explanation, describing what the figure shows, interpreting the data, and explaining its relevance to the study. Figures are well-labeled and referenced in the text.

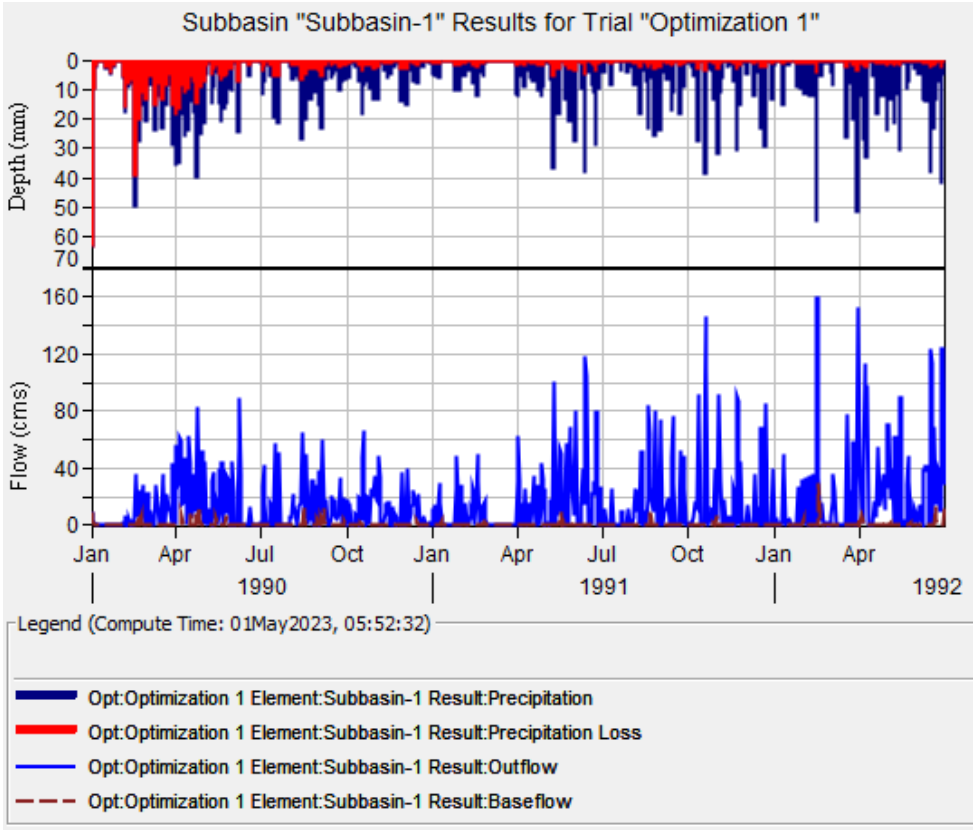



Figure 4.5: Calibration Optimization Simulation Element

For the calibration (Figure 4.6), the optimization trials of the parameterization process yielded discharge outputs that were not acceptable performance according to the recommendations by Moriasi et al. (2007) and Van Liew et al. (2003). As indicated in Figure 4.6, the $R^2$ and NSE values for calibration of daily river discharge at sub-basin 1 were 0.52 and 0.36 respectively. From the hydrographs, the observed values exceeded the simulated values by 9.5 %. This is a margin considered reasonable for analyzing the hydrologic modeling of the basin and not for forecasting purposes (Bajirao et al., 2021; Doherty & Johnston, 2003; Ogembo, 2018). Practical recommendations for future research and water resource management have been provided. These include specific areas for further investigation and actionable advice for policymakers to improve flood management and mitigation strategies.

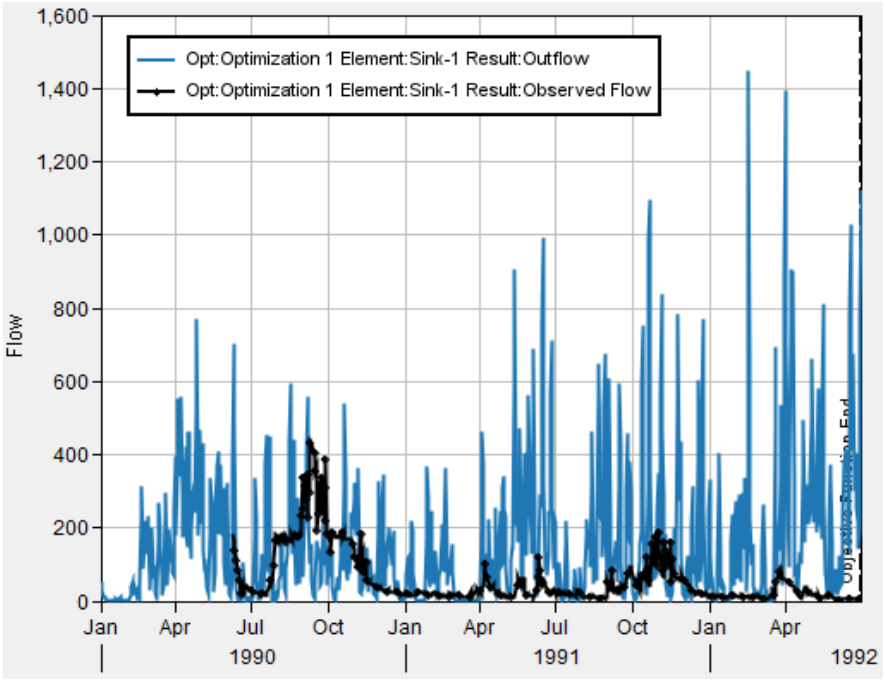



Figure 4.6: Calibration of HEC-HMS using Observed River Discharge Data at 1KB05

Figure 4.7 shows the validation trial simulations for the Gucha-Migori River Basin.

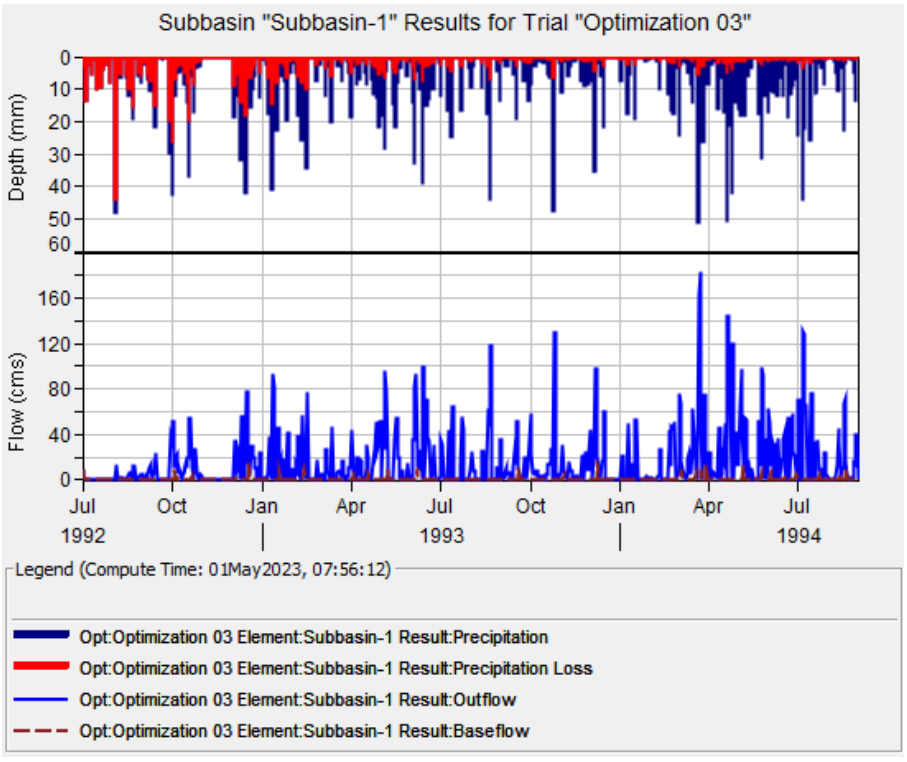



Figure 4.7: Validation Optimization Simulation Element

Simulations were accomplished using the optimized values but the output hydrograph for the validation was not the perfect match to that of the observed river discharges (Figure 4.7).

As given in Figure 4.8, $R^2$ and NSE values for validation of daily river discharge were 0.42 and 0.31 respectively.

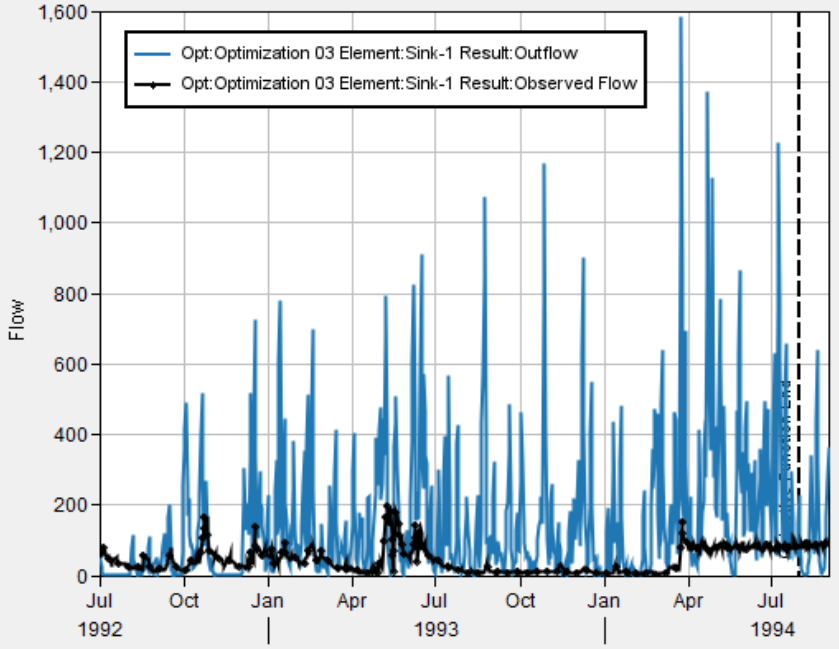



Figure 4.8: Validation of HEC-HMS using Observed River Discharge Data at 1KB05

The reduction in the coefficient of determination value from 0.52 to 0.42 could have been a result of errors in the rainfall and streamflow data due to several missing data. Another factor that could have led to this reduction may be due to the merging of 49 sub-basins to simplify the simulation process. The most likely factor for such a variation may be associated with the fact that the precipitation data did not represent the spatial variability of the entire basin. Besides, in the research carried out by Yassin et al. (2015) in Pakistan, the validation results linked to a single meteorological station were also less than the calibration results with an NSE value of 0.44. They further recommended the use of many weather station datasets to meet international standards and improve the modelling accuracy.

These results (Figures 4.6 and 4.8) are close to those attained by Hashmi (2005) when studying rainfall–runoff modelling from the Kaha hill torrent watershed in Pakistan where a difference of 8.2 % was considered acceptable. Since the difference in the observed and simulated did not exceed 10 %, these results were within the permissible limits for inference of the correlation between the precipitation and river discharge datasets. However, as had been expected, the

HEC-HMS model underestimated most of the peak flows and low flows. Other studies carried out in the same river basin with corresponding HEC-HMS performance include Ogembo (2018).

### 4.1.3 Correlation between River Discharge and Precipitation Variability

The Pearson correlation coefficient (two-tailed) was used to check the relationship between river discharge and precipitation as revealed in Figure 4.9.

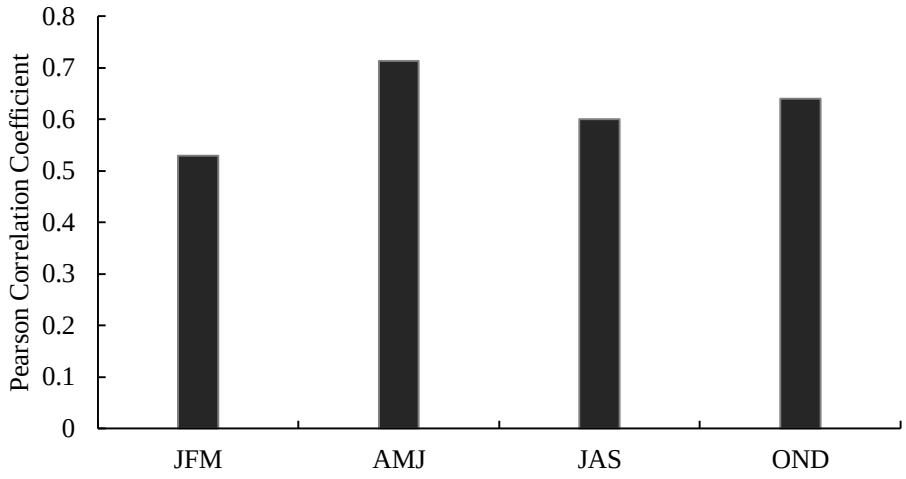



Figure 4.9: Pearson Correlation Coefficient between Precipitation and River discharge

Based on Figure 4.9, the correlation coefficients between daily averaged precipitations for JFM, AMJ, JAS, and OND and the corresponding river discharges had a significant positive correlation at the significance level of 0.05. Out of all the precipitation and simulated river discharge datasets in all the months, the strongest relationship ($r = 0.714$) was found in wetter months (April, May, and June) while the weaker correlation ($r = 0.529$) was realized in dry months (January, February, and March). In practice, it is often necessary to solve problems involving variables, which by experience are known to have an inherent relationship (Su et al., 2012; Weisberg, 2005). Thus, it can be concluded with 95 % confidence that the precipitation data is a key and useful variable for modelling spatial and temporal river discharge variability in Gucha-Migori River Basin. Other studies with corresponding results include (Stephens et al., 2015; Szolgayova et al., 2014).

Figure 4.10 represents standardized hydrometric indices for the season from January to March. The season is characterized by 18 positive and 25 negative indices. From the correlation analysis of the JFM, precipitation values could only account for 28 % of the river discharge

variance, hence, the total number of negative (wet conditions) and positive (dry conditions) indices out of the total number of the standardized hydrometric indices.

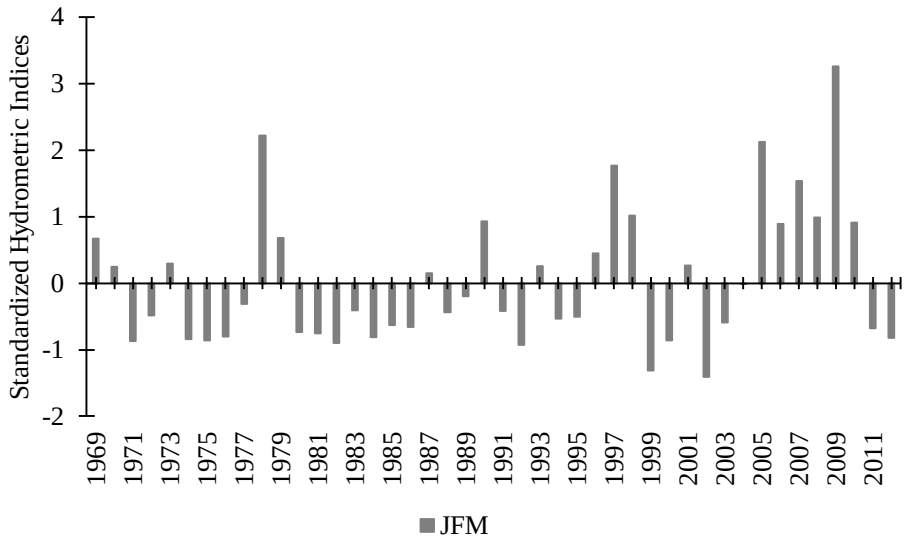



Figure 4.10: Standardized Hydrometric Indices for the January, February, and March

Figure 4.11 shows river discharge variability for the period (1969 – 2012) for the season from April to June. A total of 22 years had positive indices while 20 had negative indices. However, precipitation data could only explain 51 % of the variability or the hydrometric patterns of the simulated river discharge data for the aforementioned season.

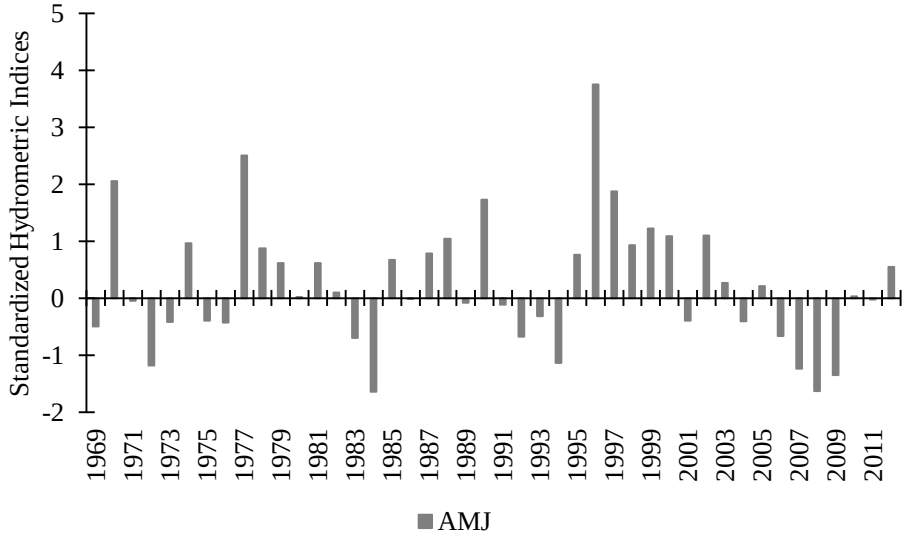



Figure 4.11: Standardized Hydrometric Indices for the April, May, and June

As indicated in Figure 4.12, 23 years had positive indices while 21 had negative indices for river discharge variability for the period (1969 – 2012) for the season from July to September.

Nevertheless, the precipitation dataset could only account for 36 % of the variance of the simulated values and variability or the hydrometric patterns of the simulated river discharge data.

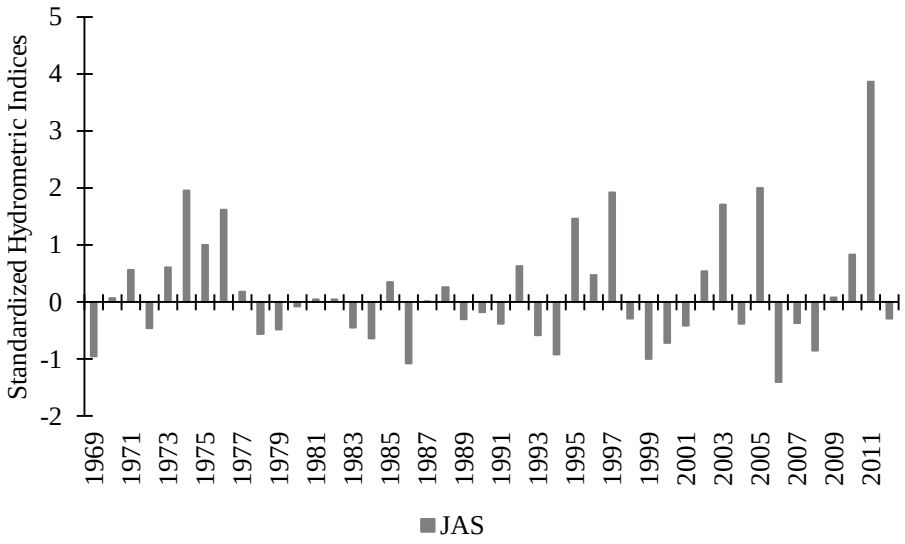



Figure 4.12: Standardized Hydrometric Indices for the July, August, and September

Figure 4.13 represents river discharge variability for the period (1969 – 2012) for the season from October to December. A total of 12 years had positive indices while 30 had negative indices.

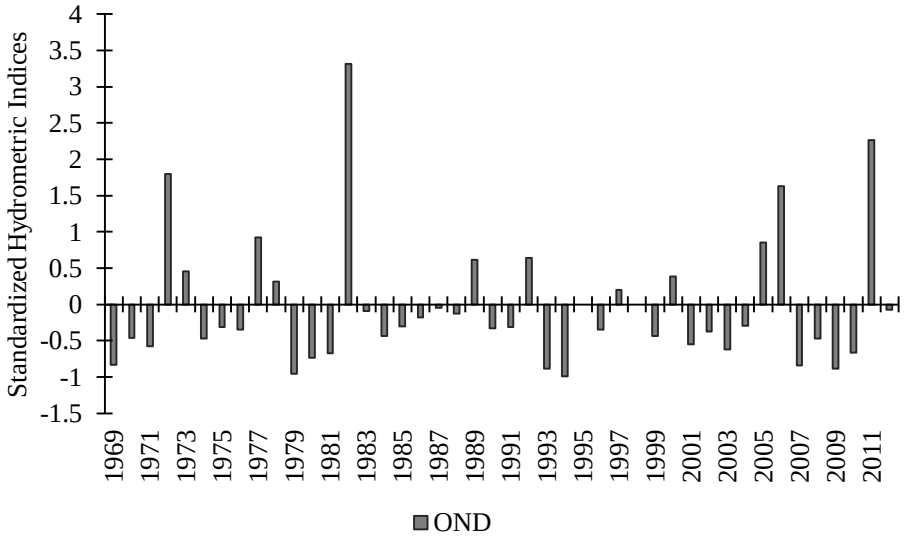



Figure 4.13: Standardized Hydrometric Indices for the October, November, and December

Nonetheless, precipitation values could only explain 41 % of the variability or the hydrometric patterns of the simulated river discharge datasets. Some of the related studies with corresponding analyses on discharge hydrometric indices entail Gao et al. (2011), Jahangir and

Nour et al. (2011), Tremblay et al. (2021), and Yarahmadi (2020). In the aforementioned studies, the +0.5 and -0.5 hydrometric indices represented extreme wet and dry climatic conditions respectively. It implies that for the case of the Gucha-Migori River Basin, flood occurrences were simulated by the HEC-HMS model in 12, 18, 13, and 8 years for the seasons JFM, AMJ, JAS, and OND respectively. There was sufficient evidence of the close relationship between precipitation variability and simulated river discharge temporal variations.

### 4.2 Correlation of Flood Magnitude and Frequency using PDFs

#### 4.2.1 Flood Frequency Analysis

Figure 4.14 represents the frequency analysis of the independent floods of the Gucha-Migori River Basin.

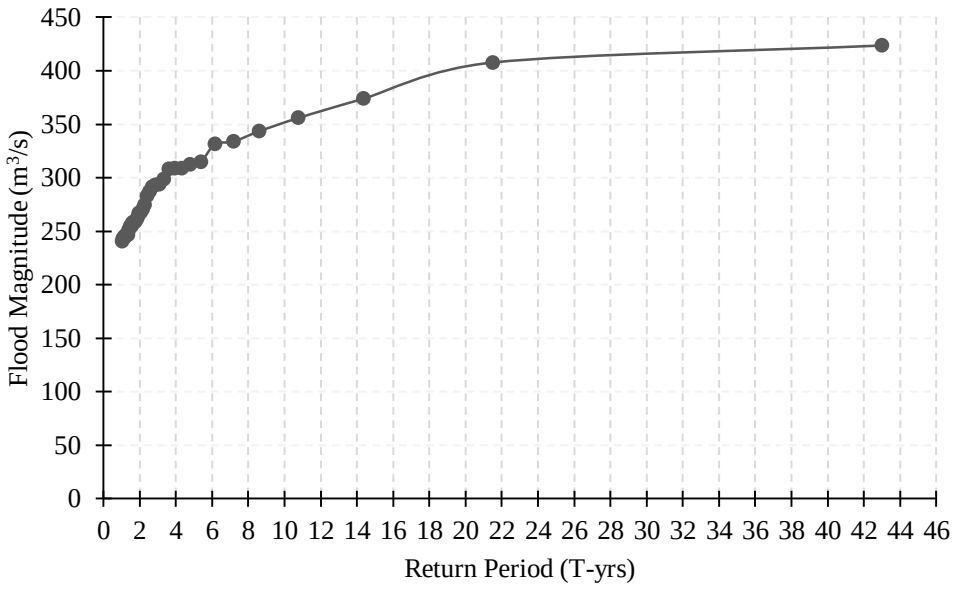



Figure 4.14: Return Period of the Independent Flood Magnitude at G-M River Basin

Out of the 132 flood events that occurred in the period 1969 to 2015, only 42 were identified as independent flood events. From the analysis, the highest and lowest, and the average and standard deviation of the recorded independent flood events were 423.90, 240.8, 284.80 $m^3/s$, and 45.54 respectively (Figure 4.14).

Figure 4.15 presents the percent annual probability of exceedance. The probability of a flood event of a magnitude equal to or exceeding the lowest discharge (240.8 $m^3/s$) and the highest flow (423.90 $m^3/s$) occurring for a particular year were 0.98 and 0.02 respectively. The probability of a flood event ranging between the highest and the lowest discharge occurring was 0.96. It can be inferred that a flood event equal to or which exceeds 423.90 $m^3/s$ in the

Gucha-Migori River Basin has a 1-in-43 chance of occurring in any given year. However, a flood event of magnitude 240.8 $m^3/s$ has a 1-in-1.02 chance of occurring in any particular year. Another way of the statistical inference for Figure 4.14 and Figure 4.15 is to allude that flood events of magnitude 240.8 $m^3/s$ and 423.90 $m^3/s$ are 1-year and 43-year floods respectively. This does not mean, nevertheless, that a flood of discharge 423.90 $m^3/s$ in the G-M River Basin will only occur every 43 years. Besides, 30 or more 45-year floods may easily occur within weeks or months of each other, or the opposite could be true, where there could easily be more than 30 years between two floods of this magnitude. Similarly, Brunner et al. (2016), Salas and Obeysekera (2014), Serinaldi (2015), and Shiau (2003) have expounded on the same observation, which actually agreed with the study's results.

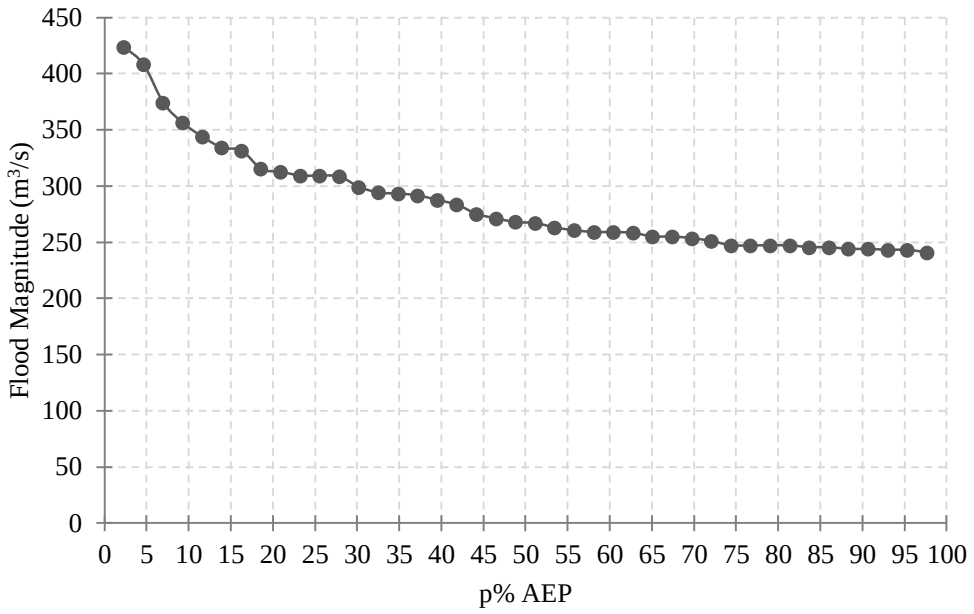



Figure 4.15: The percentage of percentage Annual Exceedance Probability

### 4.2.2 Performance of Probability Distribution Models

The computation of parameters for the derivation of the predicted magnitude for all floods with different return periods (1.05, 1.11, 1.25, 2, 5, 10, 25, 50, 100, and 200) are presented in Table 4.2 (log Pearson 3), Table 4.3 (Log-Normal), and Table 4.4 (Gumbel). For Log Pearson Type 3, the mean value of 2.4, standard deviation (0.064), skewness coefficient (0.0417), and location parameter (C = 1.109) were obtained correspondingly. The calculated mean and standard values for the normal distribution function were 284.8 and 45.54.

Table 4.2: Log Pearson 3 Parameters and Estimated Flood Magnitudes

| C for Log P3 | T-yrs | P | K | Y=log (Q) | Q(T-Log P3) |
|---|---|---|---|---|---|
| 1.109 | 1.05 | 95.2 | -1.277 | 2.368 | 233 |
| | 1.11 | 90.1 | -1.105 | 2.379 | 239 |
| | 1.25 | 80 | -0.848 | 2.395 | 249 |
| | 2 | 50 | -0.181 | 2.438 | 274 |
| | 5 | 20 | 0.744 | 2.497 | 314 |
| | 10 | 10 | 1.340 | 2.536 | 343 |
| | 25 | 4 | 2.067 | 2.582 | 382 |
| | 50 | 2 | 2.588 | 2.616 | 413 |
| | 100 | 1 | 3.091 | 2.648 | 445 |
| | 200 | 0.5 | 3.582 | 2.679 | 478 |

C is the location parameter; K is the frequency factor; P is the probability of occurrence (%); T is the estimated flood magnitude (Theoretical).

As indicated in Table 4.3, the mean value of 2.4, standard deviation (0.064), skewness coefficient (0.0417), and location parameter (C = 0) were determined for the Log-Normal Function.

Table 4.3: Log Normal Parameters and Estimated Flood Magnitudes

| C for LogN | T-yrs | Kz for LogN | Z(T-logN) | Q(T-LogN) |
|---|---|---|---|---|
| 0 | 1.01 | -2.326 | 2.250903 | 178.19824 |
| 0 | 2 | 0 | 2.4 | 251.18864 |
| 0 | 5 | 0.842 | 2.453972 | 284.4279 |
| 0 | 10 | 1.282 | 2.482176 | 303.51223 |
| 0 | 25 | 1.751 | 2.512239 | 325.26632 |
| 0 | 50 | 2.054 | 2.531661 | 340.14289 |
| 0 | 100 | 2.326 | 2.549097 | 354.07609 |
| 0 | 200 | 2.576 | 2.565122 | 367.38515 |

The mean value, reduced mean value, standard deviation, A –factor, and B-factor for the Gumbel distribution model computed were 284.8 $m^3/s$, 0.54 $m^3/s$, 45.54, 38.279, and 263.998 respectively.

Table 4.4: Gumbel's Parameters and Estimated Flood Magnitudes

| T-yrs | P | F=1-p | Y = -ln((-lnF)) | Q(T-Log P3) |
|---|---|---|---|---|
| 1.05 | 95.2 | 0.048 | -1.11 | 221.38 |
| 1.11 | 90.1 | 0.099 | -0.84 | 231.92 |
| 1.25 | 80 | 0.200 | -0.48 | 245.77 |
| 2 | 50 | 0.500 | 0.37 | 278.01 |
| 5 | 20 | 0.800 | 1.49 | 321.39 |
| 10 | 10 | 0.900 | 2.25 | 350.11 |
| 25 | 4 | 0.960 | 3.20 | 386.39 |
| 50 | 2 | 0.980 | 3.90 | 413.31 |
| 100 | 1 | 0.990 | 4.60 | 440.03 |
| 200 | 0.5 | 0.995 | 5.30 | 466.65 |

C is the location parameter; Y is the frequency factor; F is the probability of non-exceedance; T is the estimated flood magnitude (Theoretical).

The theoretical flood magnitudes and the corresponding recurrence intervals of the four Probability distribution functions are shown in Figure 4.16.

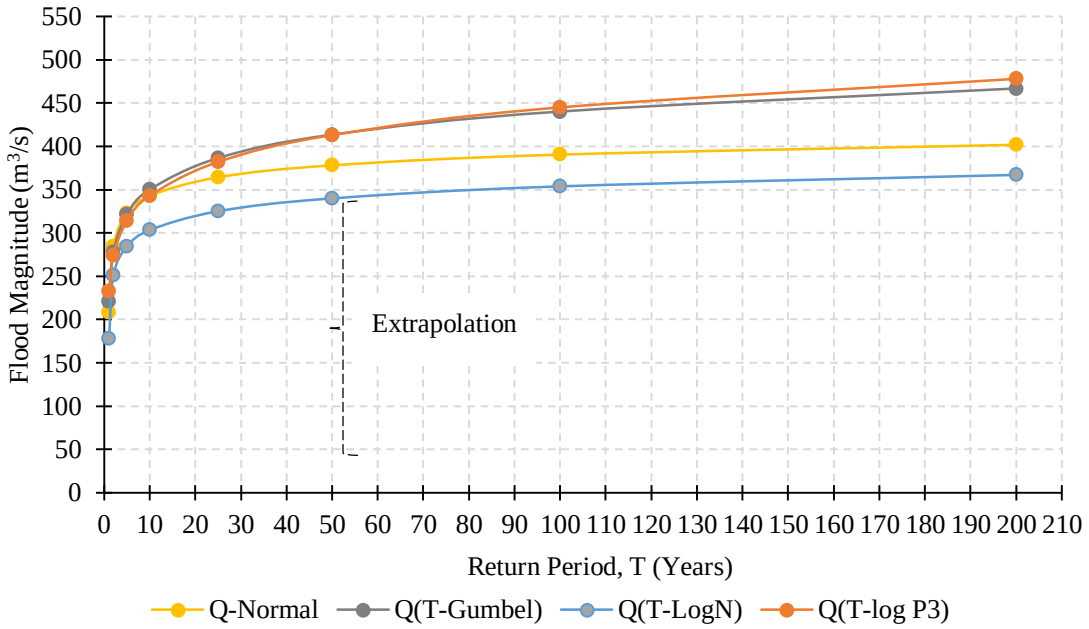



Figure 4.16: Flood Frequency Estimates by the Selected Probability Distribution Curves

Considering the predictions from Gumbel, Log P3, Log N, and Normal models, for the return periods 1.05 and 200 years, the flood magnitude estimated for G-M River Basin would be 221.38 and 466.65, 233 and 478, 178.19 and 367.38, and 209 and 402 $m^3/s$ (Figure 4.16). The

probabilities of the exceedance as predicted by the developed distribution curves are attributed to the recurrence intervals of the recorded actual independent flood events. Besides, estimated flood magnitudes from the four probability distribution curves showed a satisfactory response of peaks related to the return periods, hence, consistency with analyses in Figure 4.14.

### 4.2.3 Suitable Probability Distribution Curve Based on $R^2$

The performance comparison of the theoretical curves with the actual datasets is shown in Figure 4.17.

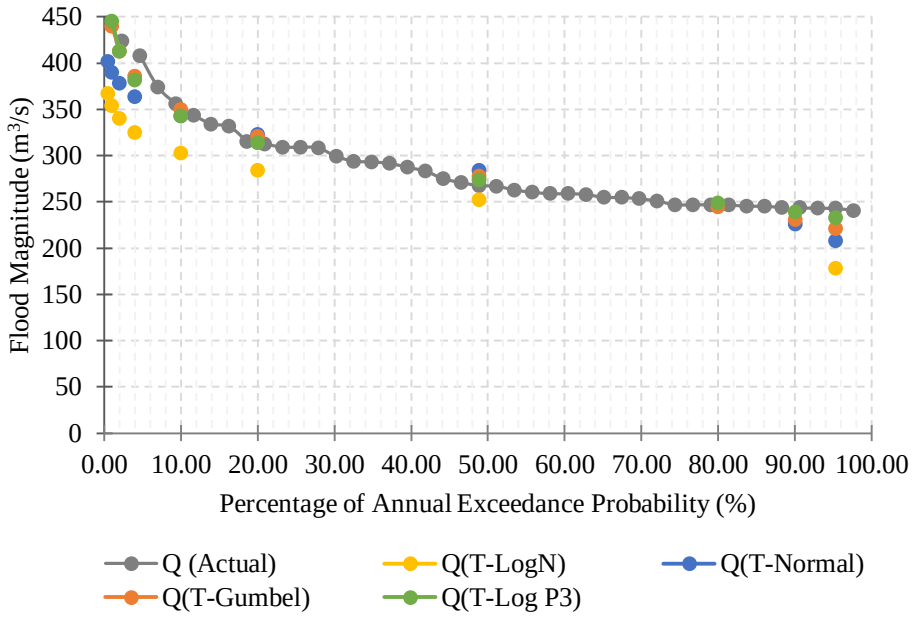



Figure 4.17: Performance comparison of the four Probability Distribution Models

Based on the coefficient of determination, the Gumbel distribution curve appeared to be the best fit of the four distribution models. It is only the Gumbel curve that accounted for at least 95 % of the variance of flood discharges, though, followed closely with the Log Pearson Type 3, which accounted for 92 %. It can then be postulated that there is a 95 % chance that Gumbel's simulated values would have a similar pattern in recurrence interval compared to the actual flood magnitude.

In comparison to Log Pearson 3, Log Normal, and Normal Distribution, the predicted flood magnitudes for the Gucha-Migori River Basin by Gumbel Probability Distribution can be considered the best alternative based on $R^2$ statistics. It explains why in many river basins in Kenya, the Gumbel Distribution Curve has been widely adopted for forecasting extreme hydrologic events. Example, in Narok, Kenya (Houessou-Dossou et al., 2019), Ngong River Basin (Juma et al., 2021), Nzoia Sub-basin (Ngina & Opere, 2017), Nile Equatorial Basins

(Opere et al., 2006). Additionally, it has also been widely adopted for forecasting in River Tana Basin (Maingi & Marsh, 2002), and Lake Victoria Basin (Ndetei, 2006), which is one of the neighbouring basins to the G-M River Basin.

### 4.2.4 GoF Statistics-Based Assessment for Suitable PDC

The Kolmogorov-Smirnov, Cramer-von Mises, and Anderson–Darling tests were used to examine the four probability distributions at a 95 % significance level ($\alpha$ =0.05) (Table 4.5).

Table 4.5: Comparison of Goodness-of-fit Statistics

| Probability Distribution Models | Kolmogorov-Smirnov (Critical value at 0.05 = 0.205) | | Cramer-von Mises (Critical value at 0.05 = 0.221) | | Anderson-Darling (Critical Value at 0.05 = 2.501) | |
|---|---|---|---|---|---|---|
| | Statistic | Rank | Statistic | Rank | Statistic | Rank |
| Gumbel | 0.044 | 1 | 0.036 | 1 | 0.202 | 1 |
| Normal | 0.049 | 2 | 0.038 | 2 | 0.256 | 2 |
| Log Normal | 0.057 | 3 | 0.041 | 3 | 0.423 | 3 |
| Log Pearson type 3 | 0.060 | 4 | 0.059 | 4 | 0.578 | 4 |

As indicated in Table 4.5, the classical goodness-of-fit test statistics for the Kolmogorov–Smirnov, Cramer–von Mises, and Anderson–Darling tests were found to be satisfactory in this context for estimating flood magnitudes based on a 95% significance level. The calculated test statistics values were lower than the theoretical values for the four chosen probability distributions at the selected significance level. Each distribution was assigned a rank from one to four, with one indicating the best fit and four indicating the poorest fit. The results of the goodness-of-fit tests pointed to the Gumbel model as the best-fitting distribution for the river discharge dataset in the Gucha-Migori River Basin.

It's worth noting that the choice of the most suitable probability distribution curve may vary from one basin to another due to temporal and spatial variations in river behavior and the uncertainties associated with model development. For instance, in the Kenyan Tana River Basin, Langat et al. (2019) identified Lognormal and gamma (Pearson type 3) distribution models as the best fits for maximum stream flows. For annual minimum flows, the Weibull, GEV, and Gumbel functions were found to be the best fits, while Lognormal functions were the best fits for annual mean flows. Similar findings were reported by other researchers,

including Norbert et al. (2014) in the Lake Victoria Basin, Kenya, and Maingi and Marsh (2020) in the context of quantifying hydrologic impacts in the Tana River, Kenya.

### 4.2.6 Best-Fit Distribution Curve

The model selection process employs the Akaike Information Criteria (AIC), which utilizes the same response variables across candidate distributions to compare different probability distributions. It doesn't mix null hypothesis testing with information criterion models. The best model is chosen based on the lowest calculated AIC value. Similarly, the Bayesian Information Criterion (BIC) relies on the likelihood function for model selection. Lower BIC values suggest fewer explanatory variables and a better fit. The ideal distribution is the one that best satisfies the criteria of goodness-of-fit statistics and information criterion-based model selection for extreme datasets.

In Table 4.6, you can find the results of the best-fit distribution model selection based on AIC and BIC for the extreme distribution type in the Gucha-Migori River Basin. The Gumbel and Log Pearson type 3 models were identified as the most suitable functions for representing extreme river discharges in the Gucha-Migori River basin, as they had the lowest AIC and BIC values.

Table 4.6: Best fit distribution model selection based on AIC

| Probability Distribution Models | AIC | BIC |
|---|---|---|
| Gumbel | 317.606 | 280.773 |
| Normal | 435.689 | 398.862 |
| Log Normal | 402.744 | 368.341 |
| Log Pearson type 3 | 321.502 | 284.870 |

In other research studies, similar findings have emerged. For instance, Bhat et al. (2019) conducted a flood frequency analysis on the Jhelum River in the Kashmir Basin and found that the Log Pearson Type-III distribution was a better fit than the Gumbel model when analyzing annual peak flow datasets. Likewise, Islam and Sarkar (2020) noted that Log Pearson Type 3 simulated discharges exhibited a closer approximation to a normal distribution compared to the Gumbel distribution, suggesting its greater reliability and suitability for flood forecasting in the Mayurakshi River Basin, India. Other studies that reported analogous results include Ewemoje and Ewemooje (2011) in the Ogun-Oshun River Basin, Nigeria, Farooq et al. (2018) in the River Swat, Pakistan, and Kumar (2019) in the Rapti River Basin.

Additionally, employing the maximum likelihood approach, Eris identified the Weibull function as the most appropriate probability distribution for analyzing eight sites in the Niger Basin. These research findings indicate that distribution functions may be suitable for estimating low, mean, and maximum flood frequencies at the same site. Therefore, the selection of an appropriate model for flood frequency analysis at a site with similar climatic, catchment, and hydrological characteristics depends on the frequency characteristics of the data series.

Consequently, the recommended distributions for the frequency characteristics in this basin were utilized to estimate flood events and their probabilities of occurrence. Using Gumbel's model for forecasting, it can be deduced that the probability of experiencing a flood magnitude falling within the range of 221.38 to 245 m³/s and 413.31 to 450 m³/s in the Gucha-Migori River Basin is 0.95 and 0.02, respectively. As a result, the model or mathematical equation derived from this analysis offers a satisfactory simulation of flood magnitudes and their corresponding annual exceedance probabilities.

### 4.2.7 Probability Distribution Curve for Gucha-Migori River Basin

Figure 4.18 presents the Probability Distribution Curve for Gucha-M River Basin.

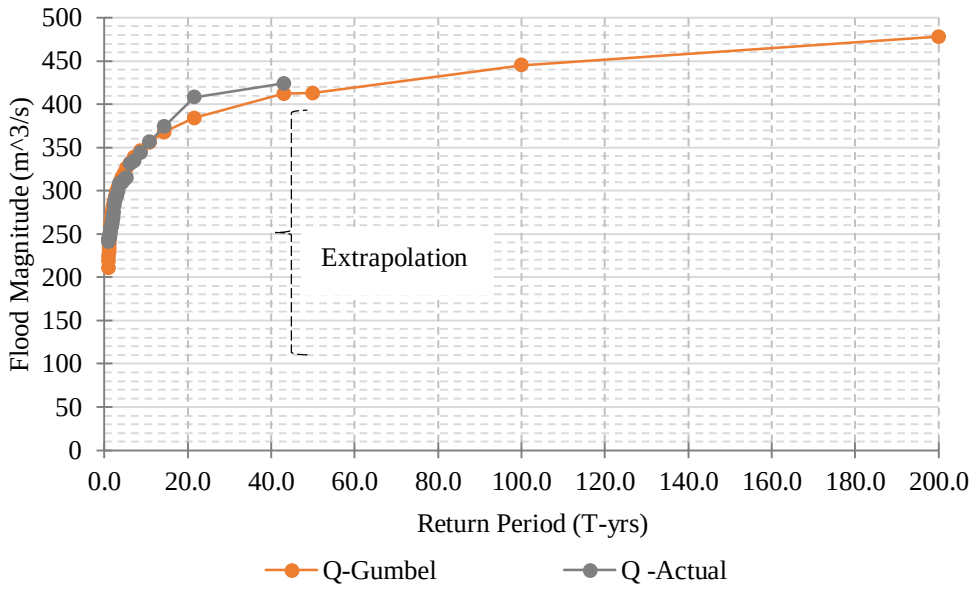



Figure 4.18: Probability Distribution Model for Gucha-Migori River Basin

According to the derived Gumbel's formula (Equation 4.1), the estimated flood magnitude for the return period 1.05, 1.11, 1.25, 2, 5, 10, 25, 50, 100, and 200 years were 221.38, 231.92, 245.77, 278.01, 321.39, 350.11, 386.39, 413.31, 440.03, and 466.65 $m^3/s$ correspondingly.

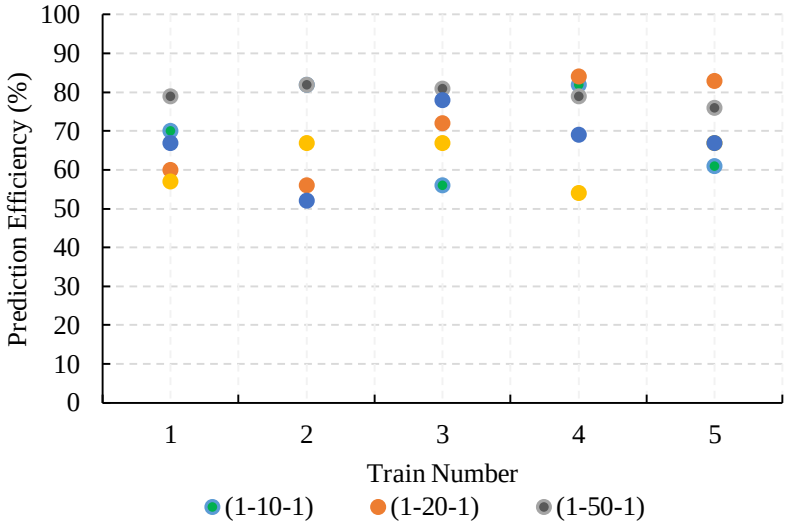
$$Magnitude(Q_T) = A(-\ln(-\ln(1-\frac{1}{\lambda(T_{yrs})}))) + B \quad (4.1)$$

where; $Q_T$ is the flood magnitude, A is equal to 38.279, $T_{yrs}$ is the return period, $\lambda$ is equal to 1, and B is equal to 263.99 $m^3/s$.

### 4.3 Forecasting of the Flood Magnitude and Percent Annual Exceedance

#### 4.3.1 Nonlinear Autoregressive Models Performance

Based on the results presented in Figure 4.19, the best prediction performance was achieved when the number of nodes in the hidden layer was 20. However, in all the trials, the efficiency was also relatively better when neurons in the hidden layer ranged from 20 to 50. From the analyses, it was revealed that the number of hidden layers and neurons per layer is completely flexible, and is optimized through an inbuilt algorithm of obtaining the network topology that provides the best performance (Ibnu Choldun et al., 2020). Nevertheless, increasing the number of neurons makes the system more complex, while a low number of neurons may restrict the generalization capabilities and computing power of ANN (Karsoliya, 2012; Stathakis, 2009; Uzair & Jamil, 2020).




1-10-1 is an architecture of the NAR model consisting of one input vector (historical datasets), and one hidden layer having 10 neurons to forecast a single output (future discharge).

Figure 4.19: Performance Efficiency of Various NAR Models

The performance of various nonlinear autoregressive neural network models in the forecasting of discharge is presented in Table 4.7. Considering all the trials, the highest and the lowest values of coefficient of correction (R= (9.15e-1 & 8.17e-1)) and mean squared error (MSE=

(1613.54e-0 & 779.59e-0)) for testing, indicated a stronger correlation and that the models captured temporal discharge patterns at a very good level.

Using the comparative analysis, the network topology of 1-20-1 was adopted as the best NAR model for forecasting because of its minimum values of R for training (9.20e-1), validation (9.27e-1), and testing (9.15e-1). Its minimum values of MSE were as follows, for training (875.67e-0), validation (815.06e-0), and testing (779.59e-0). The R-values of the aforementioned NAR model indicated that the ANN of network topology (1-20-1) at trial 4 was capable of accounting for at least 84% of the variance of the input variable. It was the best model performance as per Moriasi et al.'s (2007) Van Liew et al.'s (2003), and Nastos et al.'s (2014) recommendations. Practical recommendations for future research and water resource management have been provided. These include specific areas for further investigation and actionable advice for policymakers to improve flood management and mitigation strategies.

Table 4.7: Nonlinear Autoregressive Models Performance

| Network Architecture | Trial | | Target Values | MSE | R |
|---|---|---|---|---|---|
| 1-10 -1 | T2 | Training | 1652 | 840.66e-0 | 9.18e-1 |
| | | Validation | 206 | 998.83e-0 | 9.11e-1 |
| | | Testing | 206 | 971.39e-0 | 9.09e-1 |
| 1-20-1 | T4 | Training | 1652 | 875.67e-0 | 9.20e-1 |
| | | Validation | 206 | 815.06e-0 | 9.27e-1 |
| | | Testing | 206 | 779.59e-0 | 9.15e-1 |
| 1-50-1 | T3 | Training | 1652 | 859.50e-0 | 9.21e-1 |
| | | Validation | 206 | 773.16e-0 | 9.17e-1 |
| | | Testing | 206 | 809.14e-0 | 9.01e-1 |
| 1-100-1 | T5 | Training | 1652 | 769.37e-0 | 9.28e-1 |
| | | Validation | 206 | 1216.57e-0 | 8.96e-1 |
| | | Testing | 206 | 2027.06e-0 | 8.17e-1 |
| 1-200-1 | T4 | Training | 1652 | 838.52e-0 | 9.25e-1 |
| | | Validation | 206 | 1292.36e-0 | 8.83e-1 |
| | | Testing | 206 | 1613.54e-0 | 8.83e-1 |

1-10-1 is an architecture showing that the NAR model consisted of one input vector (historical discharge data), and one hidden layer having 10 neurons to forecast a single output (discharge).

Other NAR models like 1-10-1, 1-50-1, 1-100-1, and 1-200-1 could only account for 82, 81, 67, and 69 % of the variance of the input vectors correspondingly. These results agree with findings from some studies such as Amisi (2021) and Le et al. (2020).

### 4.3.2 Best NAR Performance Indicators

As presented in Figure 4.20, the statistical indicators, the best validation performance for the NAR (1-20-1) adopted for forecasting was 815.0594 at epoch 9.

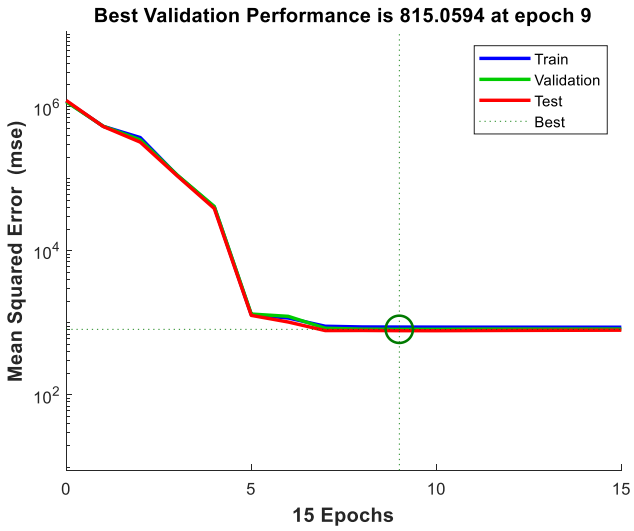



Figure 4.20: L-M algorithm MSE Output

Figure 4.21 shows the error between target values and predicted values after training the ANN model for the prediction of future discharge values for the G-M River Basin.

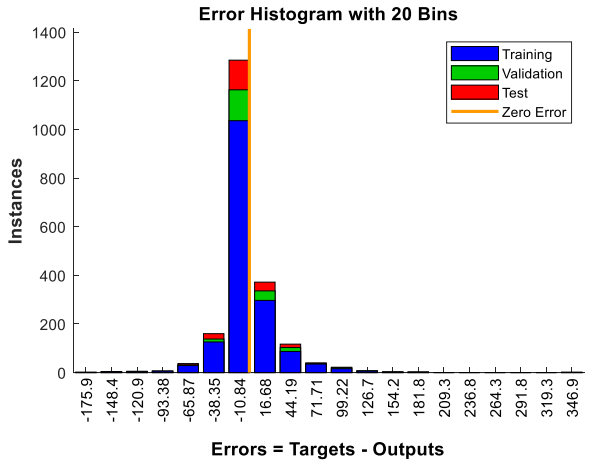



Figure 4.21: Error Histogram for the NAR Model

The number of vertical bars observed in this case was 20, hence, twenty smaller bins (Figure 4.21). In the middle of the plot, there is a bin corresponding to the error of -10.84 and the height of the bin for training datasets lies below but near 1100, and validation and test datasets lie between 1100 and 1300 indicating that many samples had an error lies in that following range. Nonetheless, these error values revealed a satisfactory performance.

Figure 4.22 represents the regression plots of the NAR model. The label of these graphs denotes an equation between the predicted value and the target value, with output as the dependent variable and target as the independent variable.

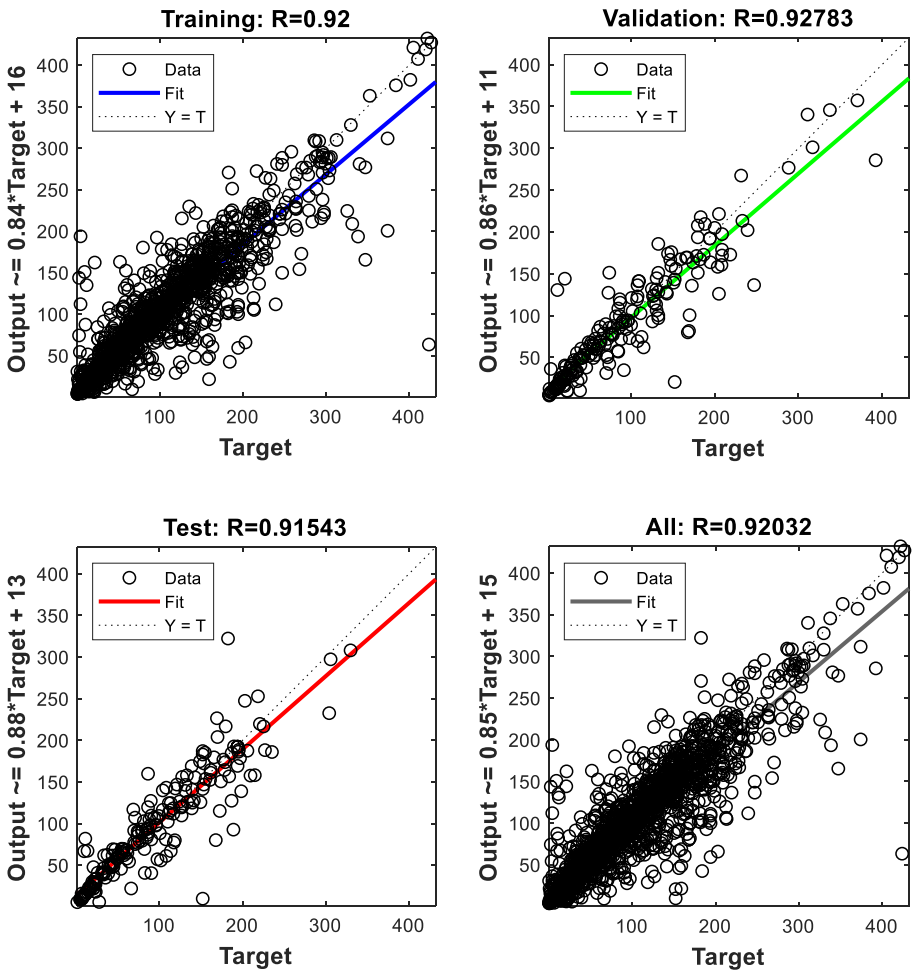



Figure 4.22: Regression of the best Nonlinear Autoregressive Neural Network

These equations clearly show that the nonlinear autoregressive model performance was highly acceptable and very good for forecasting discharge datasets. For example, it can be inferred that the correlation coefficient values for training, validation, and testing (R = 0.92, 0.93, and

0.92), implied that the NAR model was capable of explaining at least 84 % of the variance of observed discharge datasets. The coefficient of the target shows the proportionality between the output and the targets (Karul et al., 2000; Sargent, 2001). Thus, for this scenario, the NAR was selected based on good performance criteria, which was evident that the correlation coefficient was close to a unit as possible, with the negligible disparity between the training and the testing accuracy.

Figure 4.23 shows the frequency-domain analysis for training, validation, and testing of 1-20-1 ANN to understand the stability and performance properties of the model adopted for forecasting.

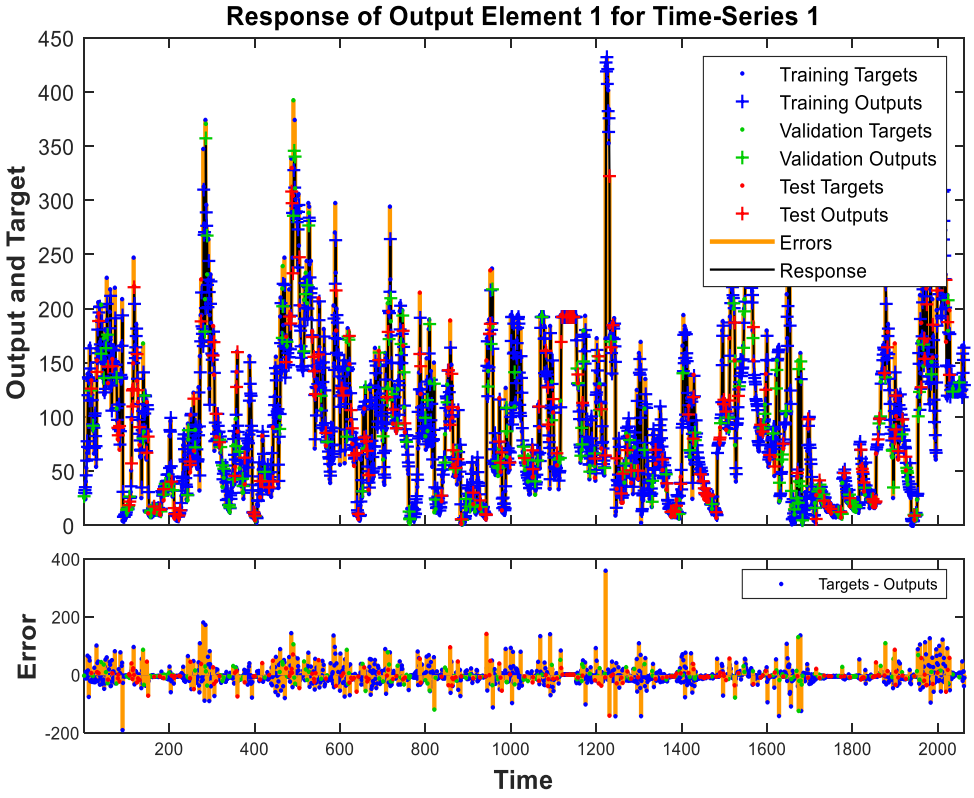



Figure 4.23: Time-Series Response of Nonlinear Autoregressive Neural Network

Validation and testing outputs of the nonlinear autoregressive neural networks showed a satisfactory response to peaks (flood magnitudes) related to extreme occurrences. The statistic indicators of the time series analysis (Figure 4.23) imply a NAR model with a high prediction accuracy and a relatively perfect linear covariation between observed and forecasted datasets. Other studies with similar modeling characteristics based on trained Artificial Neural Network models include Khan and Gupta (2020), Li and Chan (2017), and Wang et al. (2017).

Error autocorrelation is plotted in Figure 4.24. The results from the graph revealed the maximum correlations at zero lag.

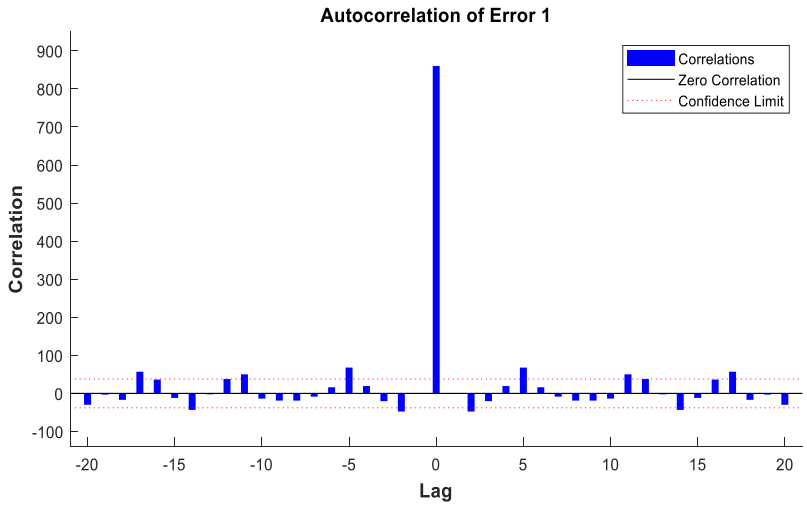



Figure 4.24: Error Autocorrelation of trained Nonlinear Autoregressive Neural Network

### 4.3.3 Time Series of Observed and Predicted Future Flood Events

Figure 4.25 shows a time series of observed and forecasted flood magnitudes with their respective annual non-exceedance probabilities.

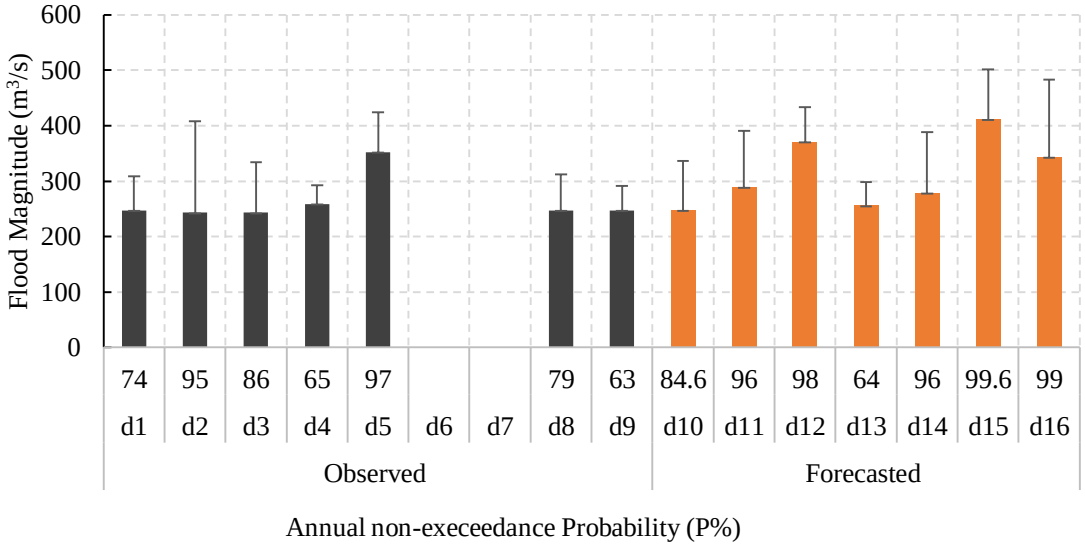



Figure 4.25: Forecasted Flood Magnitudes and their Frequencies

From the analysis, d1, d2, d3, d4, d5, d6, d7, d8, d9, d10, d11, d12, d13, d14, d15, and d16 represent (1969-1973), (1974-1969), (1979-1983), (1984-1988), (1989-1993), (1994-1998), (1999-2003), (2004-2008), (2009-2013), (2014-2018), (2019-2023), (2024-2028), (2029-2033), (2034-2038), (2039-2043, and (2048-2052) respectively. The findings show that the

daily river discharges and flood magnitudes will progressively increase in the periods 2022 and 2052. The positive trend was significant at $p < 0.01$ and the events may be described relatively perfectly by time as the independent variable (Figure 4.25). Daily river discharges positive trends in Gucha-Migori River Basin, will likely continue climatic scenarios following a polynomial curve of the third order.

Forecasted minimum flood magnitudes were 247, 289, 371, 256, 279, 412, and 343 $m^3/s$ for the periods (2014-2018), (2019-2023), (2024-2028), (2029-2033), (2034-2038), (2039-2043, and (2048-2052), respectively. In addition, the maximum forecasted flood magnitudes were 336, 391, 433, 299, 389, 502, and 483 $m^3/s$ for the periods 2014–2018, 2019–2023, 2024–2028, 2029–2033, 2034–2038, 2039–2043, and 2048–2052, respectively. Their respective annual non-exceedance probabilities from Gumbel's curve were 84.6, 96, 98, 64, 96, 99.6, and 99 % respectively. Forecasted flood magnitudes results agreed with flood studies within the catchments neighbouring Gucha-Migori River Basin such as Amisi (2021) as in Njoro, Githui et al. (2009) in Nzoia, Rwigi (2014) in Sondu, Mwangi et al. (2016) in upper Mara River, and Mbote (2016). The aforementioned studies' projections of an increase in water yield in future climatic scenarios are attributed to an increase in precipitation and surface air temperature.

# CHAPTER FIVE

# CONCLUSIONS AND RECOMMENDATIONS

## 5.1 Conclusions

This research successfully modelled and forecasted flood events in the Gucha-Migori River Basin, Kenya, providing valuable insights into flood dynamics and predictive capabilities. The following specific conclusions are drawn from the research findings:

**(i) Model Performance and Predictive Accuracy:**

The HEC-HMS model demonstrated moderate performance in simulating daily river discharge, with $R^2$ values of 0.52 during calibration and 0.42 during validation, and NSE values of 0.36 and 0.31 respectively. The significant positive correlation between daily precipitation and river discharge across different seasonal periods (JFM, AMJ, JAS, OND) indicates that precipitation is a critical driver of river discharge variability in the basin. These results highlight the need for further refinement of the model to improve predictive accuracy.

**(ii) Flood Magnitude Estimation:**

The Gumbel Probability Distribution was identified as the most suitable method for estimating extreme flood events in the Gucha-Migori River Basin. The derived Gumbel's formula provided reliable estimates of flood magnitudes for various return periods, ranging from 1.05 to 200 years. This method offers a robust tool for predicting flood magnitude, aiding in effective flood risk management and planning in the basin.

**(iii) Flood Forecasting Using ANN:**

The use of Artificial Neural Networks (ANN) combined with the Gumbel Probability Distribution achieved the best performance in forecasting flood magnitudes and their respective annual exceedance probabilities for the period between 2022 and 2052. The optimal configuration was found with 20 nodes in the hidden layer. This approach demonstrates the potential of integrating machine learning techniques with traditional hydrological models to enhance flood forecasting accuracy and reliability.

## 5.2 Recommendations

Practical recommendations for future research and water resource management have been provided. These include specific areas for further investigation and actionable advice for policymakers to improve flood management and mitigation strategies:

**(i) Utilization of the HEC-HMS Model with Reliable Data:**

The HEC-HMS model should be employed with reliable data from multiple meteorological stations to better understand the spatial-temporal effects of precipitation variability on river discharge. This approach may help reduce errors and uncertainties during the modeling set-up, parameterization, calibration, and validation processes.

**(ii) Adoption of the Gumbel Probability Distribution Model:**

Based on the data products from this analysis, the study recommends the adoption of the Gumbel probability distribution model for analyzing extreme events, specifically for flood forecasting in the Gucha-Migori River Basin. This model can provide a more accurate representation of the frequency and magnitude of extreme flood events.

**(iii) Hydrological Forecasting with Advanced Techniques:**

Future research should explore hydrological forecasting by integrating various types of Artificial Neural Networks (ANNs) and Probability Distribution Models to predict hydrologic variables under different future scenarios. This technique should also be tested in other catchments to validate its effectiveness. Furthermore, the use of artificial intelligence (AI) and machine learning (ML) techniques should be expanded to enhance predictive accuracy and reliability.

These recommendations aim to improve the understanding and management of flood risks, leveraging advanced modeling techniques and robust data analysis.

## REFERENCES


Adero, C. A. (2017). *Determinants of Sustainability of Community Based Flood Management Projects in Nyatike, Migori County Case: Lower Gucha Migori Water Resource Users Association* (Doctoral dissertation, University of Nairobi).

Akello, S. (2014). *Effects of floods on student's access to secondary education in Nyando District, Kisumu County, Kenya* (Doctoral dissertation, University of Nairobi).

Albishi, M., Bahrawi, J., & Elfeki, A. (2017). Empirical equations for flood analysis in arid zones: the Ari-Zo model. *Arabian Journal of Geosciences*, *10*(3), 51-65. http://dx.doi.org/10.1007/s12517-017-2832-4

Amisi, E. O., Kundu, P. M., & Wambua, R. M. (2020). Modeling Climate Variability Influence on River Regime in Upper Njoro Catchment, Kenya. *Journal of Civil Construction and Environmental Engineering 5(5):126*. http://dx.doi.org/10.11648/j.jccee.20200505.14

Archer, D. R., Forsythe, N., Fowler, H. J., & Shah, S. M. (2010). Sustainability of water resources management in the Indus Basin under changing climatic and socio economic conditions. *Hydrology and Earth System Sciences*, *14*(8), 1669-1680. http://dx.doi.org/10.5194/hess-14-1669-2010

Benmouiza, K., & Cheknane, A. (2013). Forecasting hourly global solar radiation using hybrid k-means and nonlinear autoregressive neural network models. *Energy Conversion and Management*, *75*, 561-569; https://doi.org/10.1016/j.enconman.2013.07.003.

Bergmeir, C. N., & Benítez Sánchez, J. M. (2012). *Neural networks in R using the Stuttgart neural network simulator:* RSNNS. American Statistical Association.

Bhat, M. S., Alam, A., Ahmad, B., Kotlia, B. S., Farooq, H., Taloor, A. K., & Ahmad, S. (2019). Flood frequency analysis of river Jhelum in Kashmir basin. *Quaternary International*, *507*, 288-294. https://doi.org/10.1016/j.quaint.2018.09.039.

Bajirao, T. S., Kumar, P., Kumar, M., Elbeltagi, A., & Kuriqi, A. (2021). Potential of hybrid wavelet-coupled data-driven-based algorithms for daily runoff prediction in complex river basins. *Theoretical and Applied Climatology*, *145*(3-4), 1207-1231. https://doi.org/10.1007/s00704-021-03681-2

Batjes, N. H., & Gicheru, P. (2004). *Soil data derived from SOTER for studies of carbon stocks and change in Kenya (ver. 1.0: GEFSOC Project)* (No. 2004/01). ISRIC-World Soil Information.

Bhagat, N. (2017). Flood frequency analysis using Gumbel's distribution method: a case study of Lower Mahi Basin, India. *Journal of Water Resources and Ocean Science*, *6*(4), 51-54. http://dx.doi.org/10.11648/j.wros.20170604.11.

Bhuiyan, H. A., McNairn, H., Powers, J., & Merzouki, A. (2017). Application of HEC-HMS in a cold region watershed and use of RADARSAT-2 soil moisture in initializing the model. *Hydrology*, *4*(1), 9-22; https://doi.org/10.3390/hydrology4010009.

Birkholz, S., Muro, M., Jeffrey, P., & Smith, H. M. (2014). Rethinking the relationship between flood risk perception and flood management. *Science of the Total Environment*, *478*, 12-20. https://doi.org/10.1016/j.scitotenv.2014.01.06.

Booij, M. J. (2005). Impact of climate change on river flooding assessed with different spatial model resolutions. *Journal of hydrology*, *303*(1-4), 176-198. https://doi.org/10.1016/j.jhydrol.2004.07.013

Bourenane, H., Bouhadad, Y., & Guettouche, M. S. (2019). Flood hazard mapping in urban area using the hydrogeomorphological approach: case study of the Boumerzoug and Rhumel alluvial plains (Constantine city, NE Algeria). *Journal of African Earth Sciences*, *160*, 103602. http://dx.doi.org/10.1016/j.jafrearsci.2019.103602.

Brody, S., Blessing, R., Sebastian, A., & Bedient, P. (2014). Examining the impact of land use/land cover characteristics on flood losses. *Journal of Environmental Planning and Management*, *57*(8), 1252-1265. http://dx.doi.org/10.1080/09640568.2013.802228.

Brunner, M. I., Seibert, J., & Favre, A. C. (2016). Bivariate return periods and their importance for flood peak and volume estimation. *Wiley Interdisciplinary Reviews: Water*, *3*(6), 819-833. http://dx.doi.org/10.1002/wat2.1173.

Calenda, G., Mancini, C. P., & Volpi, E. (2009). Selection of the probabilistic model of extreme floods: The case of the River Tiber in Rome. *Journal of Hydrology*, *371*(1-4), 1-11. http://dx.doi.org/10.1016/j.jhydrol.2009.03.010.

Chegwidden, O. S., Rupp, D. E., & Nijssen, B. (2020). Climate change alters flood magnitudes and mechanisms in climatically-diverse headwaters across the North-western United States. *Environmental Research Letters*, *15*(9), 094048. http://dx.doi.org/10.1088/1748-9326/ab986f.

Chen, F., Crow, W. T., Starks, P. J., & Moriasi, D. N. (2011). Improving hydrologic predictions of a catchment model via assimilation of surface soil moisture. *Advances in Water Resources*, *34*(4), 526-536. http://dx.doi.org/10.1016/j.advwatres.2011.01.011.

Coelho, C. A., Cavalcanti, I. A., Costa, S. M., Freitas, S. R., Ito, E. R., Luz, G., ... & Pezza, A. B. (2012). Climate diagnostics of three major drought events in the Amazon and illustrations of their seasonal precipitation predictions. *Meteorological Applications*, *19*(2), 237-255. http://dx.doi.org/10.1002/met.1324.

Collier, C. G. (2007). Flash flood forecasting: What are the limits of predictability?. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, *133*(622), 3-23. https://doi.org/10.1002/qj.29.

Chu, X., & Steinman, A. (2009). Event and continuous hydrologic modeling with HEC-HMS. *Journal of Irrigation and Drainage Engineering*, *135*(1), 119-124. http://dx.doi.org/10.1061/(ASCE)0733-9437(2009)135:1(119)

Dariane, A. B., Javadianzadeh, M. M., & James, L. D. (2016). Developing an efficient auto-calibration algorithm for HEC-HMS program. *Water Resources Management*, *30*(6), 1923-1937. https://link.springer.com/article/10.1007/s11269-016-1260-7.

Davenport, F. V., Burke, M., & Diffenbaugh, N. S. (2021). Contribution of historical precipitation change to US flood damages. *Proceedings of the National Academy of Sciences*, *118*(4), e2017524118. https://doi.org/10.1073/pnas.2017524118.

Davis, R. S. (2001). Flash flood forecast and detection methods. In *Severe convective storms* (pp. 481-525). American Meteorological Society, Boston, MA.

De Silva, M. M. G. T., Weerakoon, S. B., & Herath, S. (2014). Modeling of event and continuous flow hydrographs with HEC–HMS: case study in the Kelani River Basin, Sri Lanka. *Journal of Hydrologic Engineering*, *19*(4), 800-806. http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000846.

Devia, G. K., Ganasri, B. P. & Dwarakish, G. S. (2015). A review on hydrological models. *Aquatic Procedia*, 4, 1001-1007. https://doi.org/10.1016/j.aqpro.2015.02.126.

Doan, J. (2000). Geospatial hydrologic modeling extension HEC-GeoHMS-user's manual-version 1.0. *US Army Corps of Engineers Hydrologic Engineering Center, Davis, California, USA.*

Doherty, J., & Johnston, J. M. (2003). Methodologies for calibration and predictive analysis of a watershed model 1. *JAWRA Journal of the American Water Resources Association*, *39*(2), 251-265. http://dx.doi.org/10.1111/j.1752-1688.2003.tb04381.x.

Dottori, F., Szewczyk, W., Ciscar, J. C., Zhao, F., Alfieri, L., Hirabayashi, Y., ... & Feyen, L. (2018). Increased human and economic losses from river flooding with anthropogenic

warming. *Nature Climate Change*, *8*(9), 781-786. https://www.nature.com/articles/s41558-018-0257-z.

Douben, K. J. (2006). Characteristics of river floods and flooding: a global overview, 1985–2003. *Irrigation and Drainage: The journal of the International Commission on Irrigation and Drainage*, *55*(1), 9-21. http://dx.doi.org/10.1002/ird.239.

Du, J., Qian, L., Rui, H., Zuo, T., Zheng, D., Xu, Y., & Xu, C. Y. (2012). Assessing the effects of urbanization on annual runoff and flood events using an integrated hydrological modeling system for Qinhuai River basin, China. *Journal of Hydrology*, *464*, 127-139. http://dx.doi.org/10.1016/j.jhydrol.2012.06.057.

Engeland, K., Hisdal, H., & Frigessi, A. (2004). Practical extreme value modelling of hydrological floods and droughts: a case study. *Extremes*, *7*, 5-30. http://dx.doi.org/10.1007/s10687-004-4727-5.

Ewemoje, T. A., & Ewemooje, O. S. (2011). Best distribution and plotting positions of daily maximum flood estimation at Ona River in Ogun-Oshun river basin, Nigeria. *Agricultural Engineering International : The CIGR e-journal 13(3). https://www.researchgate.net/publication/267989991*

Fang, J., Wahl, T., Fang, J., Sun, X., Kong, F., & Liu, M. (2021). Compound flood potential from storm surge and heavy precipitation in coastal China: dependence, drivers, and impacts. *Hydrology and Earth System Sciences*, *25*(8), 4403-4416. https://doi.org/10.5194/hess-25-4403-2021.

Farley, K.A., Jobbagy, E.G., Jackson, R.B. (2005). Effects of afforestation on water yield: a global synthesis with implications for policy. *Glob. Change Biol*. 11 (10), 1565–1576. http://dx.doi.org/10.1111/j.1365-2486.2005.01011.x.

Farmer, W. H., & Vogel, R. M. (2016). On the deterministic and stochastic use of hydrologic models. *Water Resources Research*, *52*(7), 5619-5633. http://dx.doi.org/10.1002/2016WR019129.

Farooq, M., Shafique, M., & Khattak, M. S. (2018). Flood frequency analysis of river swat using Log Pearson type 3, Generalized Extreme Value, Normal, and Gumbel Max distribution methods. *Arabian Journal of Geosciences*, *11*(9), 1-10. http://dx.doi.org/10.1007/s12517-018-3553-z.

Fleming, M. (2004). *Description of the Hydrologic Engineering Center's Hydrologic Modeling System (HEC-HMS) and Application to Watershed Studies*. Engineer Research and Development Center Vicksburg MS.

Gandini, A., Garmendia, L., Prieto, I., Álvarez, I., & San-José, J. T. (2020). A holistic and multi-stakeholder methodology for vulnerability assessment of cities to flooding and extreme precipitation events. *Sustainable Cities and Society*, *63*, 102437. http://dx.doi.org/10.1016/j.scs.2020.102437.

Gao, H., Bohn, T. J., Podest, E., McDonald, K. C., & Lettenmaier, D. P. (2011). On the causes of the shrinking of Lake Chad. *Environmental Research Letters*, *6*(3), 034021. http://dx.doi.org/10.1088/1748-9326/6/3/034021.

Guo, Y., Zhang, Y., Zhang, L., & Wang, Z. (2021). Regionalization of hydrological modeling for predicting streamflow in ungauged catchments: A comprehensive review. *Wiley Interdisciplinary Reviews: Water*, *8*(1), e1487. http://dx.doi.org/10.1002/wat2.1487.

Garba, H., Ismail, A., & Tsoho, U. (2013). Fitting probability distribution functions to discharge variability of Kaduna River. *International Journal of Modern Engineering Research*, *3*(5), 2848-2852.

Gaya, C. O. (2020). *Application of GIS and Remote Sensing in Flood Management in the Lake Victoria Basin* (Doctoral dissertation, JKUAT-COETEC).

Gebre, S. L. (2015). Application of the HEC-HMS model for runoff simulation of Upper Blue Nile River Basin. *Hydrology: Current Research*, *6*(2), 1-14. http://dx.doi.org/10.4172/2157-7587.1000199.

Ghosh, S. (2013). Estimation of flash flood magnitude and flood risk in the lower segment of Damodar River basin, India. *Int J Geol Earth Environ Sci*, *3*(2), 97-114. http://www.cibtech.org/jgee.htm.

Gichere SK, Olado G, Anyona DN, Matano AS, Dida GO, Abuom PO, Amayi J, Ofulla AVO. 2013. Effects of Drought and Floods on Crop and Animal Losses and Socio-economic Status of Households in the Lake Victoria Basin of Kenya. *Journal of Emerging Trends in Economics and Management Sciences 4(1): 31-41*.

Githui, F., Gitau, W., Mutua, F. & Bauwens, W. (2009). Climate change impact on SWAT simulated stream flow in western Kenya. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 29(12), 1823-1834. http://dx.doi.org/10.1002/joc.1828.

Gotvald, A. J., Barth, N. A., Veilleux, A. G., & Parrett, C. (2012). Methods for determining magnitude and frequency of floods in California, based on data through water year 2006. *US Geological Survey Scientific Investigations Report*, *5113*, 38.

Gurney, K. (2018). *An introduction to neural networks*. CRC press.

Gyawali, R., & Watkins, D. W. (2013). Continuous hydrologic modeling of snow-affected watersheds in the Great Lakes basin using HEC-HMS. *Journal of Hydrologic Engineering*, *18*(1), 29-39. http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000591.

Hamed, K., & Rao, A. R. (Eds.). (2019). *Flood Frequency Analysis*. CRC press.

Hansson, K., Danielson, M., & Ekenberg, L. (2008). A framework for evaluation of flood management strategies. *Journal of Environmental Management*, *86*(3), 465-480. https://doi.org/10.1016/j.jenvman.2006.12.037.

Halwatura, D., & Najim, M. M. M. (2013). Application of the HEC-HMS model for runoff simulation in a tropical catchment. *Environmental Modelling & Software*, *46*, 155-162. https://doi.org/10.1016/j.envsoft.2013.03.006.

Hapuarachchi, H. A. P., Wang, Q. J., & Pagano, T. C. (2011). A review of advances in flash flood forecasting. *Hydrological Processes*, *25*(18), 2771-2784. http://dx.doi.org/10.1002/hyp.8040.

Hayhoe, S., Neill, C., Porder, S., McHorney, R., Lefebvre, P., Coe, M., Elsenbeer, H., Krusche, A., 2011. Conversion to soy on the Amazonian agricultural frontier increases streamflow without affecting stormflow dynamics. *Glob. Change Biol. 17 (5), 1821–1833.* http://dx.doi.org/10.1111/j.1365-2486.2011.02392.x.

Herreyre, N., Cormier, A., Hermelin, S., Oberlin, C., Schmitt, A., Thirion-Merle, V., ... & Motto-Ros, V. (2023). Artificial neural network for high-throughput spectral data processing in LIBS imaging: application to archaeological mortar. *Journal of Analytical Atomic Spectrometry*, *38*(3), 730-741. http://dx.doi.org/10.1039/D2JA00389A.

Houessou-Dossou, E. A. Y., Mwangi Gathenya, J., Njuguna, M., & Abiero Gariy, Z. (2019). Flood frequency analysis using participatory GIS and rainfall data for two stations in Narok Town, Kenya. *Hydrology*, *6*(4), 90. http://dx.doi.org/10.3390/hydrology6040090.

Hussein, K., Alkaabi, K., Ghebreyesus, D., Liaqat, M. U., & Sharif, H. O. (2020). Land use/land cover change along the Eastern Coast of the UAE and its impact on flooding risk. *Geomatics, Natural Hazards and Risk*, *11*(1), 112-130. http://dx.doi.org/10.1080/19475705.2019.1707718.

Ibnu Choldun R, M., Santoso, J., & Surendro, K. (2020). Determining the number of hidden layers in neural network by using principal component analysis. In *Intelligent Systems*

*and Applications: Proceedings of the 2019 Intelligent Systems Conference (IntelliSys) Volume 2* (pp. 490-500). Springer International Publishing.

Ikhwali, M. F., Pawattana, C., Nur, S., Azhari, B., Ikhsan, M., Aida, N., & Silvia, C. S. (2022). Reviews, challenges, and prospects of the application of Hydrologic Engineering Center-Hydrologic Modelling System (HEC-HMS) model in Indonesia. *Engineering and Applied Science Research*, *49*(5), 669-680. https://ph01.tci-thaijo.org/index.php/easr/article/view/247833.

Islam, A., & Sarkar, B. (2020). Analysing flood history and simulating the nature of future floods using Gumbel method and Log-Pearson Type III: the case of the Mayurakshi River Basin, India. *Bulletin of Geography. Physical Geography Series*, *19*(1), 43-69. http://dx.doi.org/10.2478/bgeo-2020-0009.

Ismail, H., Kamal, M. R., Mojid, M. A., Abdullah, A. F. B., & Hin, L. S. (2022). Loss methods in HEC-HMS model for streamflow projection under climate change: a review. *International Journal of Hydrology Science and Technology*, *13*(1), 23-42. http://dx.doi.org/10.1504/IJHST.2022.119234.

Ismael, O., Sang Joseph, K., & Home Patrick, G. (2017). HEC-HMS model for runoff simulation in Ruiru reservoir watershed. *American Journal of Engineering Research*, *6*(4), 1-7.

Jahangir, M. H., & Yarahmadi, Y. (2020). Hydrological drought analyzing and monitoring by using Streamflow Drought Index (SDI)(case study: Lorestan, Iran). *Arabian Journal of Geosciences*, *13*, 1-12. http://dx.doi.org/10.1007/s12517-020-5059-8.

Jongman, B., Ward, P. J., & Aerts, J. C. (2012). Global exposure to river and coastal flooding: Long term trends and changes. *Global Environmental Change*, *22*(4), 823-835. https://doi.org/10.1016/j.gloenvcha.2012.07.004.

Jonkman, S. N., & Vrijling, J. K. (2008). Loss of life due to floods. *Journal of Flood Risk Management*, *1*(1), 43-56. http://dx.doi.org/10.1111/j.1753-318X.2008.00006.x

Juma, B., Olang, L. O., Hassan, M., Chasia, S., Bukachi, V., Shiundu, P., & Mulligan, J. (2020). Analysis of rainfall extremes in the Ngong River Basin of Kenya: Towards integrated urban flood risk management. *Physics and Chemistry of the Earth, Parts A/B/C*, 102929. https://doi.org/10.1016/j.pce.2020.102929.

Kabeja, C., Li, R., Guo, J., Rwabuhungu, D., Manyifika, M., Gao, Z., ... & Zhang, Y. (2019, December). The Impact of Reforestation Induced Land Cover Change (1990-2017) on Flash Flood Peak Discharge Using HEC-HMS Hydrological Model and Satellite

Observations: A study in Data Scarce Tributary of Sichuan Basin, China. In *AGU Fall Meeting Abstracts* (Vol. 2019, pp. H13K-1868).

Kamali, B., Mousavi, S. J., & Abbaspour, K. C. (2013). Automatic calibration of HEC-HMS using single-objective and multi-objective PSO algorithms. *Hydrological Processes*, *27*(26), 4028-4042. http://dx.doi.org/10.1002/hyp.9510

Karsoliya, S. (2012). Approximating number of hidden layer neurons in multiple hidden layer BPNN architecture. *International Journal of Engineering Trends and Technology*, *3*(6), 714-717.

Khan, F. M., & Gupta, R. (2020). ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India. *Journal of Safety Science and Resilience*, *1*(1), 12-18. https://doi.org/10.1016/j.jnlssr.2020.06.007

Karul, C., Soyupak, S., Çilesiz, A. F., Akbay, N., & Germen, E. (2000). Case studies on the use of neural networks in eutrophication modeling. *Ecological Modelling*, *134*(2-3), *145-152*. http://dx.doi.org/10.1016/S0304-3800(00)00360-4

Kumar, R. (2019). Flood frequency analysis of the Rapti river basin using log pearson type-III and Gumbel Extreme Value-1 methods. *Journal of the Geological Society of India*, *94*(5), 480-484. http://dx.doi.org/10.1007/s12594-019-1344-0

Kumar, D., & Bhattacharjya, R. K. (2020). Evaluating two GIS-based semi-distributed hydrological models in the Bhagirathi-Alkhnanda River catchment in India. *Water Policy*, *22*(4), 991-1014. http://dx.doi.org/10.2166/wp.2020.159

Kundu, P. M, (2007). *Application of remote sensing and GIS techniques to evaluate the impact of land use and land cover change on stream flows: The case for River Njoro catchment in Eastern Mau-Kenya*. (PhD Thesis. Egerton University, Kenya).

Kundzewicz, Z. W., Krysanova, V., Dankers, R., Hirabayashi, Y., Kanae, S., Hattermann, F. F., ... & Schellnhuber, H. J. (2017). Differences in flood hazard projections in Europe–their causes and consequences for decision making. *Hydrological Sciences Journal*, *62*(1), 1-14. https://doi.org/10.1080/02626667.2016.1241398

Lang, M., Pobanz, K., Renard, B., Renouf, E., & Sauquet, E. (2010). Extrapolation of rating curves by hydraulic modelling, with application to flood frequency analysis. *Hydrological Sciences Journal–Journal des Sciences Hydrologiques*, *55*(6), 883-898. https://doi.org/10.1080/02626667.2010.504186

Langat, P. K., Kumar, L., & Koech, R. (2019). Identification of the most suitable probability distribution models for maximum, minimum, and mean streamflow. *Water, 11*(4), 734. http://dx.doi.org/10.3390/w11040734

Le, T. T., Pham, B. T., Ly, H. B., Shirzadi, A., & Le, L. M. (2020). Development of 48-hour precipitation forecasting model using nonlinear autoregressive neural network. In *CIGOS 2019, Innovation for Sustainable Infrastructure: Proceedings of the 5th International Conference on Geotechnics, Civil Engineering Works and Structures* (pp. 1191-1196). Springer Singapore.

Leauthaud, C., Duvail, S., Hamerlynck, O., Paul, J. L., Cochet, H., Nyunja, J., ... & Grünberger, O. (2013). Floods and livelihoods: The impact of changing water resources on wetland agro-ecological production systems in the Tana River Delta, Kenya. *Global Environmental Change*, *23*(1), 252-263. https://doi.org/10.1016/j.gloenvcha.2012.09.003

Lee, W. K., & Tuan Resdi, T. A. (2016). Simultaneous hydrological prediction at multiple gauging stations using the NARX network for Kemaman catchment, Terengganu, Malaysia. *Hydrological Sciences Journal*, 61(16), 2930-2945. https://doi.org/10.1080/02626667.2016.1174333

Li, Q., & Chan, M. F. (2017). Predictive time-series modeling using artificial neural networks for Linac beam symmetry: an empirical study. *Annals of the New York Academy of Sciences*, *1387*(1), 84-94. https://doi.org/10.1111%2Fnyas.13215

Loveridge, M., & Rahman, A. (2021). Effects of probability-distributed losses on flood estimates using event-based rainfall-runoff models. *Water, 13*(15), 2049. http://dx.doi.org/10.3390/w13152049

Machuke, G. W., Apudo, B. O., Mwita, P. N., Mbugua, L. N., & Kiche, J. (2014). An Application of Extreme Value Theory in Modelling Electricity Production in Kenya.

Maingi, J. K., & Marsh, S. E. (2002). Quantifying hydrologic impacts following dam construction along the Tana River, Kenya. *Journal of Arid Environments*, *50*(1), 53-79. https://doi.org/10.1006/jare.2000.0860

Mango, L. M., Melesse, A. M., McClain, M. E., Gann, D., & Setegen, S. G. (2010). Land use and climate change impacts on the hydrology of the upper Mara River Basin, Kenya: results of a modeling study to support better resource management. *Hydrol. Earth Syst. Sci., 15, 2245–2258, 2011*. https://doi:10.5194/hess-15-2245-2011

Marhaento, H., Booij, M. J., & Hoekstra, A. Y. (2016). Attribution of changes in stream flow to land use change and climate change in a mesoscale tropical catchment in Java, Indonesia. *Hydrology Research*, *48*(4), 1143-1155. https://doi.org/10.2166/nh.2016.110

Mousavi, S. J., Abbaspour, K. C., Kamali, B., Amini, M., & Yang, H. (2012). Uncertainty-based automatic calibration of HEC-HMS model using sequential uncertainty fitting approach. *Journal of Hydroinformatics, 14(2), 286-309*. http://dx.doi.org/10.2166/hydro.2011.071

Mendizabal, M., Sepúlveda, J., & Torp, P. (2014). Climate change impacts on flood events and its consequences on human in Deba River. *International Journal of Environmental Research 8(1):221-230*. https://www.researchgate.net/publication/286739283

Mitra, P., Ray, R., Chatterjee, R., Basu, R., Saha, P., Raha, S., ... & Saha, S. (2016, October). Flood forecasting using Internet of things and artificial neural networks. In *2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)* (pp. 1-5). Institute of Electrical and Electronics Engineers.

Mishra, B. K., Rafiei Emam, A., Masago, Y., Kumar, P., Regmi, R. K., & Fukushi, K. (2018). Assessment of future flood inundations under climate and land use change scenarios in the Ciliwung River Basin, Jakarta. *Journal of Flood Risk Management*, *11*, 1105-1115. https://doi.org/10.1111/jfr3.12311

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50(3), 885-900. http://dx.doi.org/10.13031/2013.23153

Muiruri, P. N., Oonge, Z. I., & Odira, P. M. (2014). *Assessment of Integrated Water Resources Management Practices in Gucha River Catchment* (Doctoral dissertation, University of Nairobi).

Mujumdar, P. P., & Kumar, D. N. (2012). *Floods in a changing climate: hydrologic modeling*. Cambridge University Press.

Mukerji, A., Chatterjee, C. and Raghuwanshi, N.S., 2009. Flood forecasting using ANN, neuro-fuzzy, and neuro-GA models. *Journal of Hydrologic Engineering*, *14*(6), 647-652. http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000040

Mwangi, H. M., Julich, S., Patil, S. D., McDonald, M. A., & Feger, K. H. (2016). Relative contribution of land use change and climate variability on discharge of upper Mara

River, Kenya. *Journal of Hydrology: Regional Studies*, 5, 244-260. https://doi.org/10.1016/j.ejrh.2015.12.059

Natarajan, S., & Radhakrishnan, N. (2019). Simulation of extreme event-based rainfall–runoff process of an urban catchment area using HEC-HMS. *Modeling Earth Systems and Environment*, *5*(2), 1867-1881 https://link.springer.com/article/10.1007/s40808-019-00644-5.

Natarajan, S., & Radhakrishnan, N. (2020). An integrated hydrologic and hydraulic flood modeling study for a medium-sized ungauged urban catchment area: A case study of Tiruchirappalli City Using HEC-HMS and HEC-RAS. *Journal of The Institution of Engineers (India): Series A*, *101*, 381-398. http://dx.doi.org/10.1007/s40030-019-00427-2

Ndetei, C. J. (2006). *A Comparative analysis of the performance of flood frequency distributions in the Lake Victoria Basin in Kenya* (Doctoral dissertation, University of Nairobi).

Ngaina, J. N., & Opere, A. O. (2017). Hydrological characteristics of lower Nzoia Sub-basin in Kenya. http://dx.doi.org/10.4172/2157-7587.1000285

Norbert, J., Mugo, M., & Gadain, H. (2014). Estimation of Design Floods in Un-gauged Catchments using a Regional Index Flood Method. A Case Study of Lake Victoria Basin in Kenya. *Physics and Chemistry of the Earth, Parts A/B/C V, 67–69, 4-11*. https://doi.org/10.1016/j.pce.2014.02.001

Nour, A. M., Vallet-Coulomb, C., Gonçalves, J., Sylvestre, F., & Deschamps, P. (2021). Rainfall-discharge relationship and water balance over the past 60 years within the Chari-Logone sub-basins, Lake Chad basin. *Journal of Hydrology: Regional Studies*, *35*, 100824. https://doi.org/10.1016/j.ejrh.2021.100824

Nyaupane, N., Mote, S. R., Bhandari, M., Kalra, A., & Ahmad, S. (2018, May). Rainfall-runoff simulation using climate change based precipitation prediction in HEC-HMS model for Irwin Creek, Charlotte, North Carolina. In *World Environmental and Water Resources Congress 2018: watershed management, irrigation and drainage, and water resources planning and management* (pp. 352-363). Reston, VA: American Society of Civil Engineers.

Odyuo, N. M. (2020). *Modelling the impact of land use land cover changes on th runoff processes of Chalakudy basin using HEC-HMS model* (Doctoral dissertation, Department of IDE).

Ogembo, V. (2018). *Hydrological Modeling and Climate Change Impacts on River Kuja Basin Using HEC-HMS and HEC-GeoHMS Models* (Master's thesis).

Olang, L. O., & Fürst, J. (2011). Effects of land cover change on flood peak discharges and runoff volumes: model estimates for the Nyando River Basin, Kenya. *Hydrological Processes*, *25*(1), 80-89. http://dx.doi.org/10.1002/hyp.7821

Oleyiblo, J. O., & Li, Z. J. (2010). Application of HEC-HMS for flood forecasting in Misai and Wan'an catchments in China. *Water Science and Engineering*, *3*(1), 14-22. http://dx.doi.org/10.3882/j.issn.1674-2370.2010.01.002

Olson, S. A., & Veilleux, A. G. (2014). *Estimation of flood discharges at selected annual exceedance probabilities for unregulated, rural streams in Vermont*. US Department of the Interior, US Geological Survey.

Opere, A. (2013). Floods in Kenya. In *Developments in Earth Surface Processes* (16, 315-330). https://doi.org/10.1016/B978-0-444-59559-1.00021-9

Opere, A. O., Mkhandi, S., & Willems, P. (2006). At site flood frequency analysis for the Nile Equatorial basins. *Physics and Chemistry of the Earth, Parts A/B/C*, *31*(15-16), 919-927. https://doi.org/10.1016/j.pce.2006.08.018

Ouédraogo, W. A. A., Raude, J. M., & Gathenya, J. M. (2018). Continuous modeling of the Mkurumudzi River catchment in Kenya using the HEC-HMS conceptual model: Calibration, validation, model performance evaluation and sensitivity analysis. *Hydrology*, *5*(3), 44. http://dx.doi.org/10.3390/hydrology5030044

Ozoegwu, C. G. (2019). Artificial neural network forecast of monthly mean daily global solar radiation of selected locations based on time series and month number. *Journal of Cleaner Production*, 216, 1-13. http://dx.doi.org/10.1016/j.jclepro.2019.01.096

Perry, C. A. (2000). *Significant floods in the United States during the 20th century: USGS measures a century of floods*. US Department of the Interior, US Geological Survey. https://doi.org/10.3133/fs02400

Plate, E. J. (2002). Flood risk and flood management. *Journal of Hydrology*, *267*(1-2), 2-11. https://doi.org/10.1016/S0022-1694(02)00135-X

Quandt, A., & Kimathi, Y. A. (2017). Perceptions of the effects of floods and droughts on livelihoods: lessons from arid Kenya. *International Journal of Climate Change Strategies and Management*, 9(3), 337-351. http://dx.doi.org/10.1108/IJCCSM-11-2014-0132

Raghunath, H. M. (2006). *Hydrology: principles, analysis and design*. New Age International.

Raible, C. C., Luksch, U., & Fraedrich, K. (2003). Precipitation and northern hemisphere regimes. *Atmospheric Science Letters*, *5*(1-4), 43-55. http://dx.doi.org/10.1016/j.atmoscilet.2003.12.001

Ramly, S., & Tahir, W. (2016). Application of HEC-GeoHMS and HEC-HMS as rainfall–runoff model for flood simulation. In *ISFRAM 2015* (pp. 181-192). Springer, Singapore.

Rao, A. R., & Hamed, K. H. (2019). *Flood frequency analysis*. CRC press.

Renner, M., Brust, K., Schwärzel, K., Volk, M., & Bernhofer, C. (2014). Separating the effects of changes in land cover and climate: a hydro-meteorological analysis of the past 60 yr in Saxony, Germany. *Hydrology and Earth System Sciences*, *18*(1), 389-405. https://doi.org/10.5194/hess-18-389-2014

Rezaeianzadeh, M., Tabari, H., Yazdi, A. A., Isik, S., & Kalin, L. (2014). Flood flow forecasting using ANN, ANFIS and regression models. *Neural Computing and Applications*, *25*(1), 25-37. http://dx.doi.org/10.1007/s00521-013-1443-6

Rizwan, M., Guo, S., Xiong, F., & Yin, J. (2018). Evaluation of various probability distributions for deriving design flood featuring right-tail events in pakistan. *Water*, *10*(11), 1603. http://dx.doi.org/10.3390/w10111603

Rodriguez, D.A., Tomasella, J., Linhares, C. (2010). Is the forest conversion to pasture affecting the hydrological response of Amazonian catchments? Signals in the Ji-Paraná Basin. *Hydrol. Process*. 24, 1254–1269. http://dx.doi.org/10.1002/hyp.7586.

Salas, J. D., & Obeysekera, J. (2014). Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events. *Journal of Hydrologic Engineering*, *19*(3), 554-568. http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000820

Sardoii, E. R., Rostami, N., Sigaroudi, S. K., & Taheri, S. (2012). Calibration of loss estimation methods in HEC-HMS for simulation of surface runoff (Case Study: Amirkabir Dam Watershed, Iran). *Adv. Environ. Biol*, *6*(1), 343-348. https://www.researchgate.net/publication/289917089

Sargent, D. J. (2001). Comparison of artificial neural networks with other statistical approaches: results from medical data sets. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, *91*(S8), 1636-1642. https://doi.org/10.1002/1097-0142(20010415)91:8+%3C1636::aid-cncr1176%3E3.0.co;2-d

Scharffenberg, W. A., & Fleming, M. J. (2005). US Army Corps of Engineers Hydrologic Engineering Center.

Schipper, T. C. (2017). *The attribution of changes in streamflow to climate and land use change for 472 catchments in the United States and Australia* (Master's thesis, University of Twente).

Şen, Z. (2018). *Flood modeling, prediction and mitigation*. Springer International Publishing.

Serinaldi, F. (2015). Dismissing return periods!. *Stochastic Environmental Research and Risk Assessment*, *29*(4), 1179-1189. http://dx.doi.org/10.1007/s00477-014-0916-1

Shiau, J. T. (2003). Return period of bivariate distributed extreme hydrological events. *Stochastic Environmental Research and Risk Assessment*, *17*(1), 42-57. http://dx.doi.org/10.1007/s00477-003-0125-9

Singh, J., Knapp, H. V., Arnold, J. G., & Demissie, M. (2005). Hydrological modeling of the Iroquois river watershed using HSPF and SWAT 1. *JAWRA Journal of the American Water Resources Association*, *41*(2), 343-360. http://dx.doi.org/10.1111/j.1752-1688.2005.tb03740.x

Singo LR, Kundu PM, Odiyo JO, Mathivha FI, Nkuna TR. 2012. *Flood frequency analysis of annual maximum stream flows for Luvuvhu River catchment, Limpopo Province, South Africa*. In: 16th SANCIAHS National Hydrology Symposium, University of Pretoria, South Africa, 1-3 October. Session 2C.

Somvanshi, V. K., Pandey, O. P., Agrawal, P. K., Kalanker, N. V., Prakash, M. R., & Chand, R. (2006). Modeling and prediction of rainfall using artificial neural network and ARIMA techniques. *J. Ind. Geophys. Union*, 10(2), 141-151.

Stathakis, D. (2009). How many hidden layers and nodes?. *International Journal of Remote Sensing*, *30*(8), 2133-2147. http://dx.doi.org/10.1080/01431160802549278

Stedinger, J. R., & Griffis, V. W. (2008). Flood frequency analysis in the United States: Time to update.

Stephens, E., Day, J. J., Pappenberger, F., & Cloke, H. (2015). Precipitation and floodiness. *Geophysical Research Letters*, *42*(23), 10-316. http://dx.doi.org/10.1002/2015GL066779

Sriwongsitanon, N., & Taesombat, W. (2011). Effects of land cover on runoff coefficient. *Journal of Hydrology*, *410*(3-4), 226-238. https://doi.org/10.1016/j.jhydrol.2011.09.021

Su, X., Yan, X., & Tsai, C. L. (2012). Linear regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, *4*(3), 275-294. http://dx.doi.org/10.1002/wics.1198

Szolgayova, E., Parajka, J., Blöschl, G., & Bucher, C. (2014). Long term variability of the Danube River flow and its relation to precipitation and air temperature. *Journal of Hydrology*, *519*, 871-880. http://dx.doi.org/10.1016/j.jhydrol.2014.07.047

Szwagrzyk, M., Kaim, D., Price, B., Wypych, A., Grabska, E., & Kozak, J. (2018). Impact of forecasted land use changes on flood risk in the Polish Carpathians. *Natural Hazards*, *94*(1), 227-240. https://link.springer.com/article/10.1007/s11069-018-3384-y

Tabari, H. (2020). Climate change impact on flood and extreme precipitation increases with water availability. *Scientific Reports*, *10*(1), 1-10. https://www.nature.com/articles/s41598-020-70816-2

Tesfamariam, E. G., Home, P. G., & Gathenya, J. M. (2021). Rainfall-runoff modelling to determine continuous time series of daily streamflow in the Umba River, Kenya. *African Journal of Rural Development*, *5*(2), 49-68.

Thakur, B., Parajuli, R., Kalra, A., Ahmad, S., & Gupta, R. (2017). Coupling HEC-RAS and HEC-HMS in precipitation runoff modelling and evaluating flood plain inundation map. In *World Environmental and Water Resources Congress 2017* (240-251). http://dx.doi.org/10.1061/9780784480625.022

Thavhana, M. P. (2018). *Runoff simulation using the SWAT model for flood frequency analysis and design flood estimations in the Luvuvhu River catchment, South Africa* (Doctoral dissertation).

Tiwari, M. K., & Chatterjee, C. (2010). Development of an accurate and reliable hourly flood forecasting model using wavelet–bootstrap–ANN (WBANN) hybrid approach. *Journal of Hydrology*, *394*(3-4), 458-470. https://doi.org/10.1016/j.jhydrol.2010.10.001

Thu, K. C. M., Zin, W. W., & Khine, E. E. (2019). *Simulation of Rainfall-runoff process using HEC-HMS model for Chindwin River Basin*. *In Proceedings of National Conference of Science and Engineering, Unpaginated.*

Tramblay, Y., Rouché, N., Paturel, J. E., Mahé, G., Boyer, J. F., Amoussou, E., ... & Lachassagne, P. (2021). ADHI: the African database of hydrometric indices (1950–2018). *Earth System Science Data*, *13*(4), 1547-1560. http://dx.doi.org/10.5194/essd-13-1547-2021

U. S. Army Corps of Engineers, USACE (2010) *HEC-GeoHMS Geospatial Hydrologic Modeling Extension,* Version 5.0 - User's Manual. Davis, CA, pp 1–6.

U.S. Army Corps of Engineers, Hydrologic Engineering Center. U. S. Army Corps of Engineers (USACE). (2013). *Hydrologic Modeling System,* HEC-GeoHMS, Version 10.1, Users Guide.

Uzair, M., & Jamil, N. (2020, November). Effects of hidden layers on the efficiency of neural networks. In *2020 IEEE 23rd international multitopic conference (INMIC)* (pp. 1-6). IEEE.

Van Liew, M. W., Arnold, J. G., & Garbrecht, J. D. (2003). Hydrologic simulation on agricultural watersheds: Choosing between two models. *Transactions of the ASAE*, *46*(6), 1539-1551. http://dx.doi.org/10.13031/2013.15643

Viglione, A., Merz, R., & Blöschl, G. (2009). On the role of the runoff coefficient in the mapping of rainfall to flood return periods. *Hydrology and Earth System Sciences*, *13*(5), 577-593. http://dx.doi.org/10.5194/hess-13-577-2009

Vivekanandan, N. (2015). Flood frequency analysis using method of moments and L-moments of probability distributions. *Cogent Engineering*, *2*(1), 1018704. http://dx.doi.org/10.1080/23311916.2015.1018704

Volpi, E., & Fiori, A. (2012). Design event selection in bivariate hydrological frequency analysis. *Hydrological Sciences Journal*, *57*(8), 1506-1515. http://dx.doi.org/10.1080/02626667.2012.726357

Wambua, R. M. (2019). Hydrological drought forecasting using modified surface water supply index (SWSI) and streamflow drought index (SDI) in conjunction with artificial neural networks (ANNs). *International Journal of Service Science, Management, Engineering, and Technology (IJSSMET), 10*(4), 39-57. http://dx.doi.org/10.4018/IJSSMET.2019100103

Wambua, R. M. (2020). Development of a non-linear integrated drought index (NDI) for managing drought and water resources forecasting in the upper Tana river basin-Kenya. *International Journal of Environmental Sustainability and Green Technologies (IJESGT), 11*(1), 15-33. http://dx.doi.org/10.4018/IJESGT.2020010102

Wambua, R. M., Mutua, B. M., & Raude, J. M. (2017). Analysis of spatial and temporal drought variability in a tropical river basin using Palmer Drought Severity Index (PDSI). *International Journal of Water Resources and Environmental Engineering 9(8):178-190.* http://dx.doi.org/10.5897/IJWREE2017.0723

Wang, K. W., Deng, C., Li, J. P., Zhang, Y. Y., Li, X. Y., & Wu, M. C. (2017). Hybrid methodology for tuberculosis incidence time-series forecasting based on ARIMA and a NAR neural network. *Epidemiology & Infection*, *145*(6), 1118-1129. https://doi.org/10.1017/S0950268816003216

Warburton ML, Schulze RE, Jewitt GPW. 2010. Confirmation of ACRU model results for applications in land use and climate change studies. *Hydrology and Earth System Science 14: 2399–2414*. https://doi.org/10.5194/hess-14-2399-2010

Weisberg, S. (2005). *Applied linear regression*. John Wiley & Sons.

WMO. (2011a). *Manual on Flood Forecasting and Warning:* WMO-No. 1072. Retrieved from http://www.wmo.int/pages/prog/hwrp/publications/flood_forecasting_warning/WM O 1072_en.pdf

WMO. (2011b). WMO Flood Forecasting Initiative. Retrieved April 20, 2019, from https://www. wmo.int/pages/prog/hwrp/FFI-index.php

Xiong, F., Guo, S., Chen, L., Yin, J., & Liu, P. (2018). Flood frequency analysis using Halphen distribution and maximum entropy. *Journal of Hydrologic Engineering, 23*(5), 04018012. http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0001637

Zischg, A. P., & Bermúdez Martínez, M. (2020). Mapping the sensitivity of population exposure to changes in flood magnitude: Prospective application from local to global scale. *Frontiers in Earth Science 8:1-14*. http://dx.doi.org/10.3389/feart.2020.534735

Zischg, A. P., Felder, G., Mosimann, M., Röthlisberger, V., & Weingartner, R. (2018). Extending coupled hydrological-hydraulic model chains with a surrogate model for the estimation of flood losses. *Environmental Modelling & Software, 108*, 174-185. http://dx.doi.org/10.1016/j.envsoft.2018.08.009.

# APPENDICES

## Appendix A: Tables

Table 1A: Attributes of the downloaded Landsat 1-3 Satellite Data

| | Band | Wavelength Spectral Range (µm) | Spatial Resolution (m) |
|---|---|---|---|
| Landsat 1-3 | Band 4 - Green | 0.5-0.6 | 60 |
| | Band 5 - Red | 0.6-0.7 | 60 |
| | Band 6 – Near IR | 0.7-0.8 | 60 |
| | Band 7 – Near IR | 0.8-1.1 | 60 |

Table 2A: Attributes of the downloaded Landsat 4-5 Satellite Data

| | Band | Wavelength Spectral Range (µm) | Spatial Resolution (m) |
|---|---|---|---|
| Landsat 4-5 | Band 1 – Blue | 0.45-0.52 | 30 |
| | Band 2 – Green | 0.52-0.60 | 30 |
| | Band 3 – Red | 0.63-0.69 | 30 |
| | Band 4 – Near IR | 0.77-0.90 | 30 |
| | Band 5 – Short -wave IR | 1.55-1.75 | 30 |
| | Band 6 – Thermal IR | 10.40-12.50 | 120 |
| | Band 7– Shortwave IR | 2.09-2.35 | 30 |

Table 3A: Attributes of the downloaded Landsat 7 Satellite Data

| | Band | Wavelength Spectral Range (µm) | Spatial Resolution (m) |
|---|---|---|---|
| Landsat 7 | Band 1 – Blue | 0.45-0.52 | 30 |
| | Band 2 – Green | 0.52-0.60 | 30 |
| | Band 3 – Red | 0.63-0.69 | 30 |
| | Band 4 – Near IR | 0.77-0.90 | 30 |
| | Band 5– Short wave IR1 | 1.55-1.75 | 30 |
| | Band 6 – Thermal IR | 10.40-12.50 | 30/60 |
| | Band 7– Shortwave IR2 | 2.09-2.35 | 30 |
| | Band 8- Pan | 0.52-0.90 | 15 |

Table 4A: GPS points used for Geo-referencing in image classification

| Latitude | Longitude | Description | Classified LULC |
|---|---|---|---|
| -0.9186886 | 34.1629787 | Kabuto area | Farm land |
| -0.9324261 | 34.1706723 | Relocated to new position due to flood | Kabuki Primary school |
| -0.9200208 | 34.1671365 | River Gucha Migori river banks | Sand Harvesting |
| -0.9329420 | 34.1723661 | Kabuto Area | Residential area |
| -0.9959811 | 34.2261873 | Ogoche area | Farm lands |
| -0.9967397 | 34.2276904 | | Bare land with minimal activity |
| -0.9224677 | 34.2126381 | Nyatike area | Nyatike irrigation scheme |
| -0.9268485 | 34.2274315 | Gunga stream | Rice farming |

Table 5A**:** GPS points used for Geo-referencing of stream digitization

| Latitude | Longitude | Description |
|---|---|---|
| -0.9194701 | 34.1656569 | River Gucha-Migori |
| -0.9298539 | 34.1783116 | River Dike, used to prevent flood water from crossing over to the residential areas |
| -0.9401968 | 34.1825582 | Original River Gucha-Migori route, changed in 2003 |
| -0.9965885 | 34.2273561 | Ogoche Stream, seasonal river on the upstream |
| -0.9500831 | 34.2101668 | 1KB 05 gauging station |
| -0.8985250 | 34.1924344 | Obware Stream |
| -0.9884325 | 34.24976 | Gucha- Migori Sango, where the two rivers meet |
| -0.9709883 | 34.260812 | River Kuja |
| -0.9925320 | 34.2918827 | River Migori |

**Appendix B: Figures**

Each figure in the Results and Discussion chapter is accompanied by a detailed explanation, describing what the figure shows, interpreting the data, and explaining its relevance to the study. Figures are well-labeled and referenced in the text.

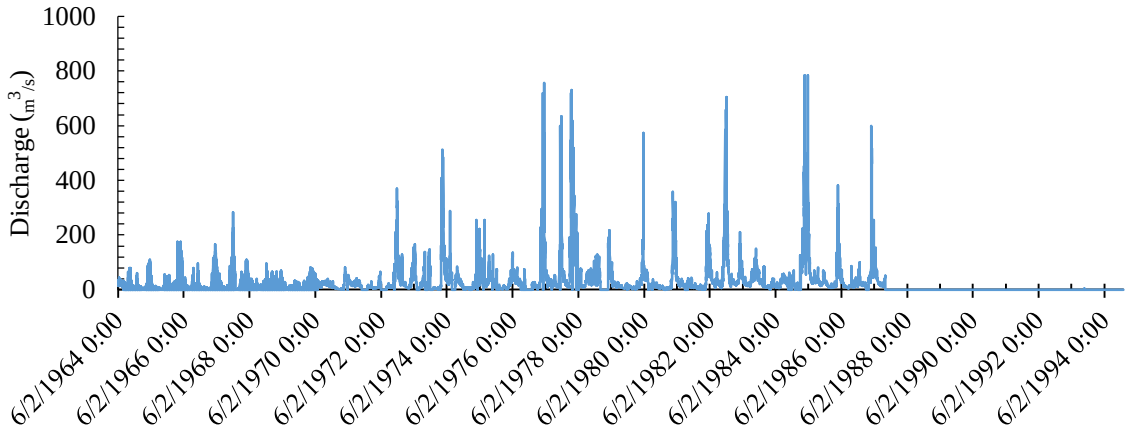



Figure 1B: 1KB01A-Gucha Macalder [$m^3/s$] (Position: 34.2732, -0.9548 Decimal Degrees)

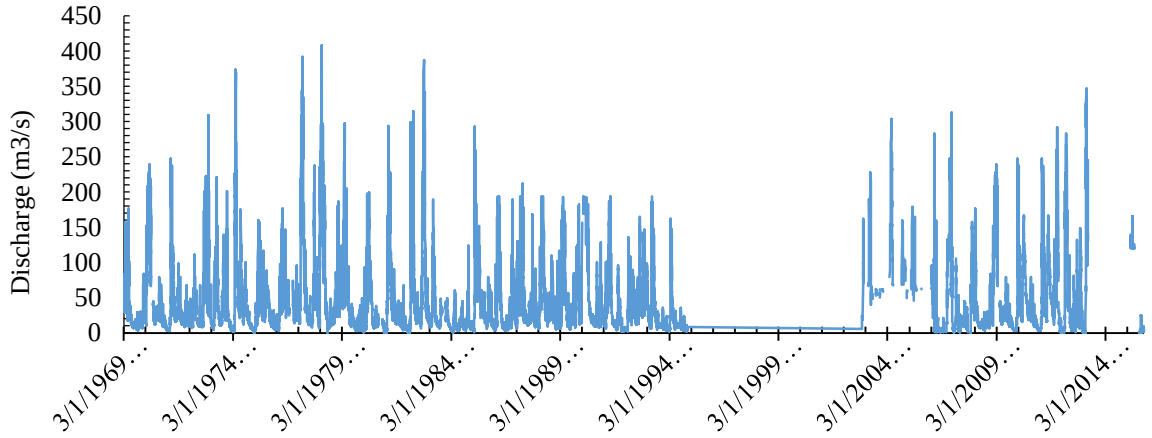



Figure 2B: 1KB05-Gucha Migori [m^3/s] (Position: 34.2096, -0.9492 Decimal Degrees)

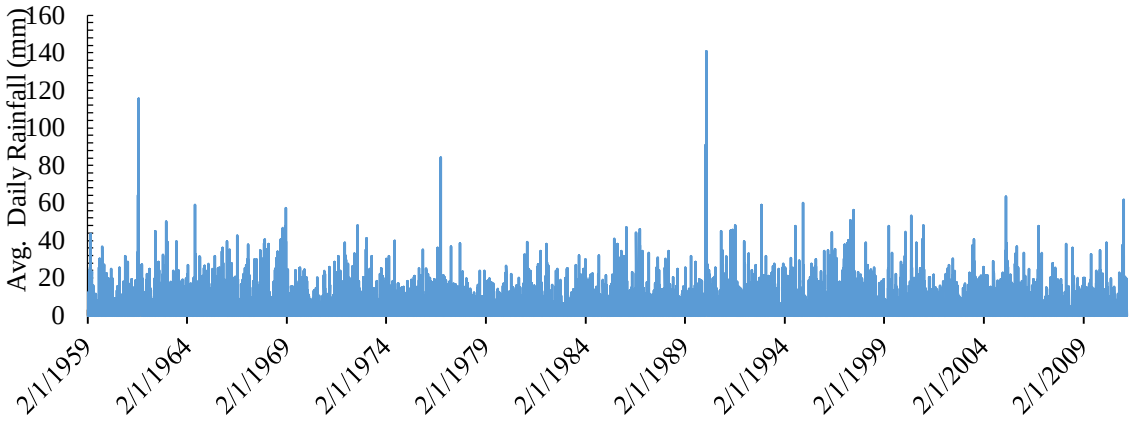



Figure 3B: Gucha-Migori River Basin Average Daily Rainfall (mm)

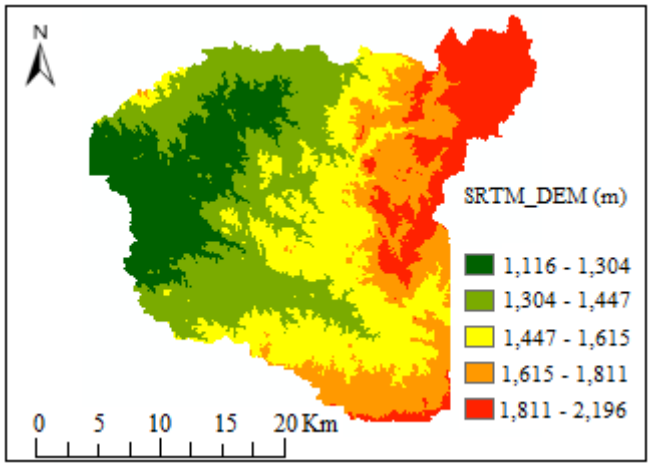



Figure 4B: Shuttle Radar Topography Mission-DEM for Gucha-Migori River Basin

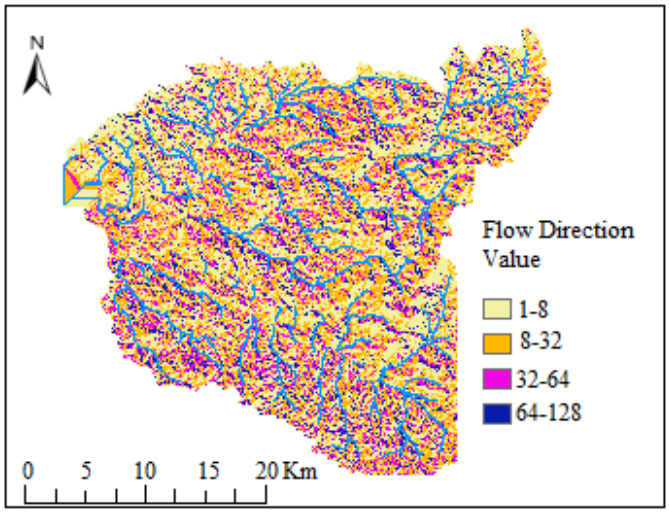



Figure 5B: Flow Direction Processing using GeoHMS and Arc Hydro tools

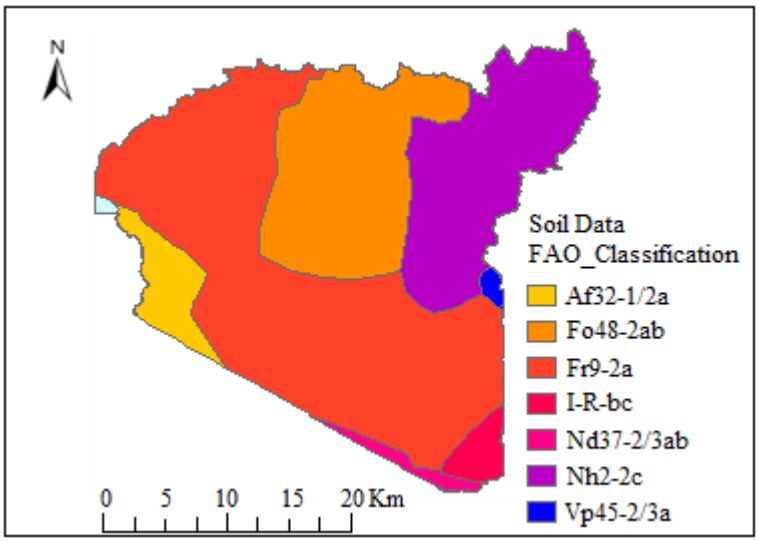



Figure 6B: FAO soil classification at Gucha-Migori River Basin

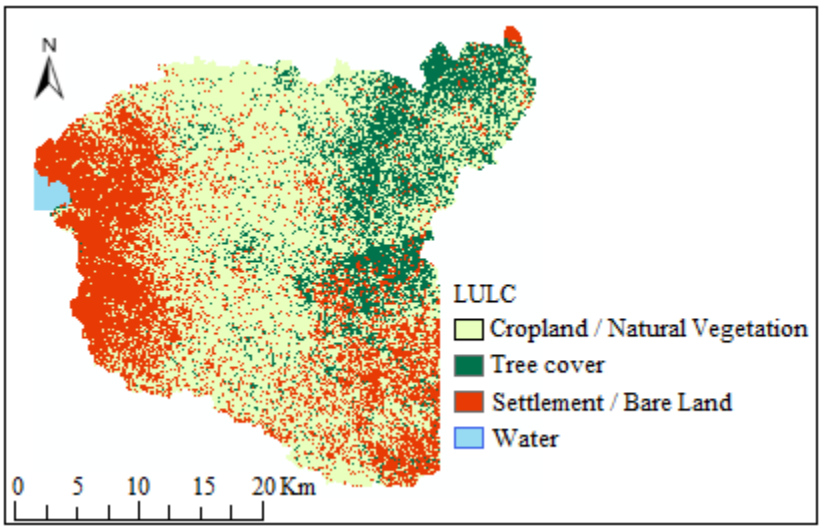



Figure 7B: Classified Landsat MSS image of 1976 for Gucha-Migori River Basin

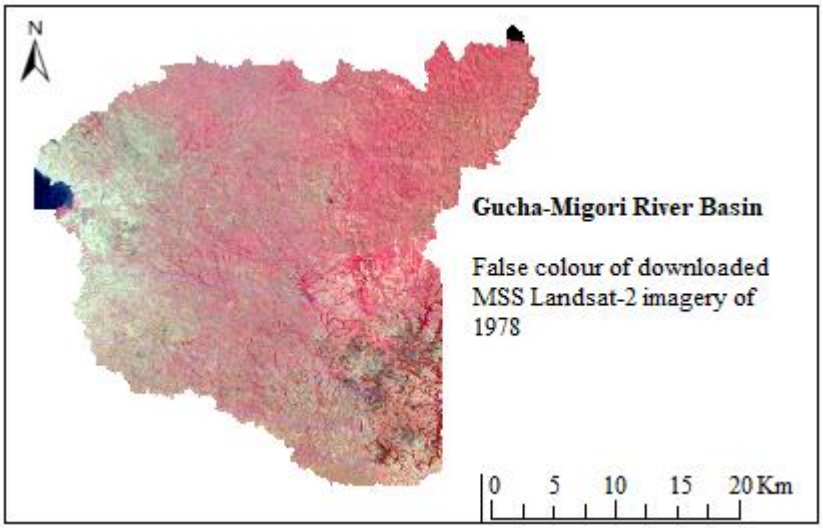



Figure 8B: Shows an imagery for raster MTL.text 4Bands (Green, Red, Near Infrared 1, and Infrared 2) for Gucha-Migori River Basin

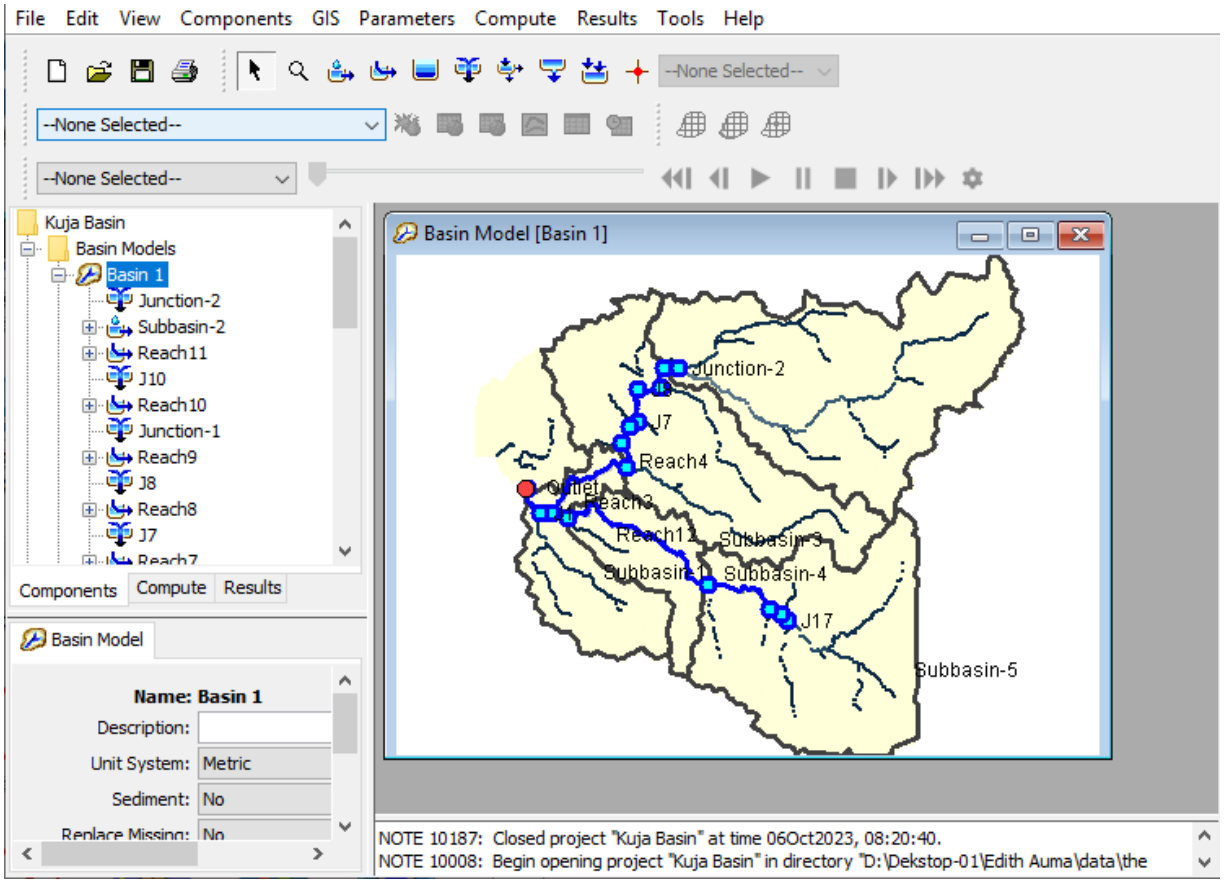


Figure 9B: The HEC-HMS Modelling Process for Gucha-Migori River Basin

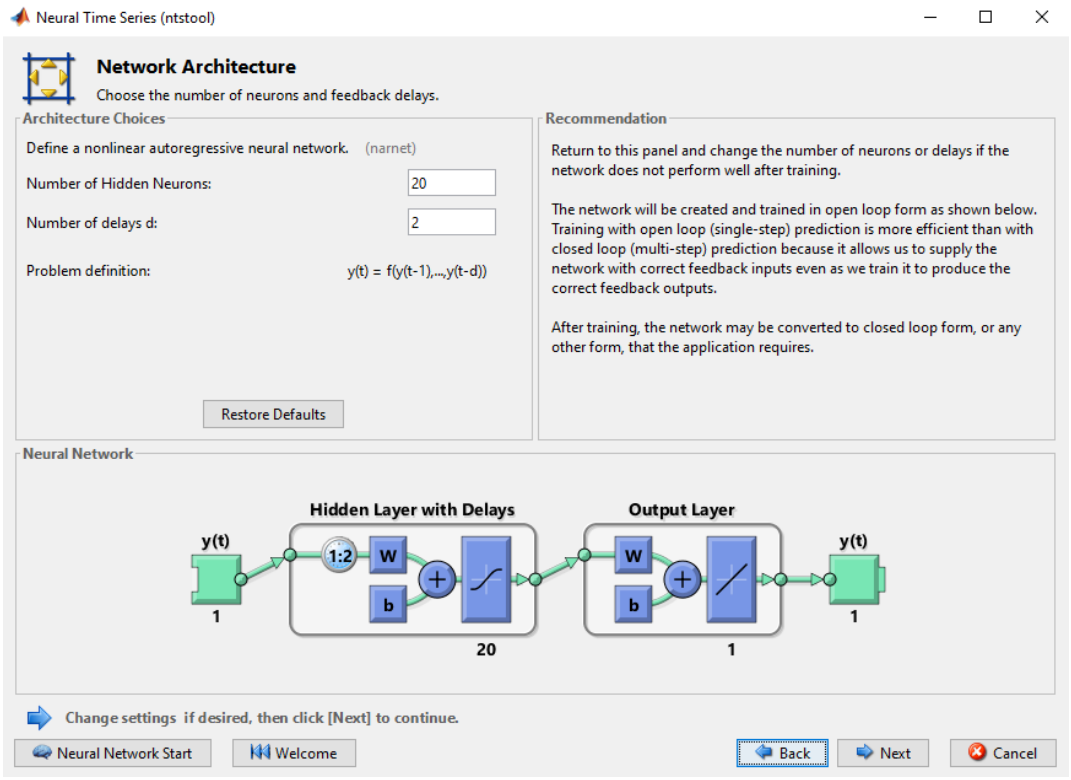



Figure 10B: Selection of the number of neurons and feedback delays

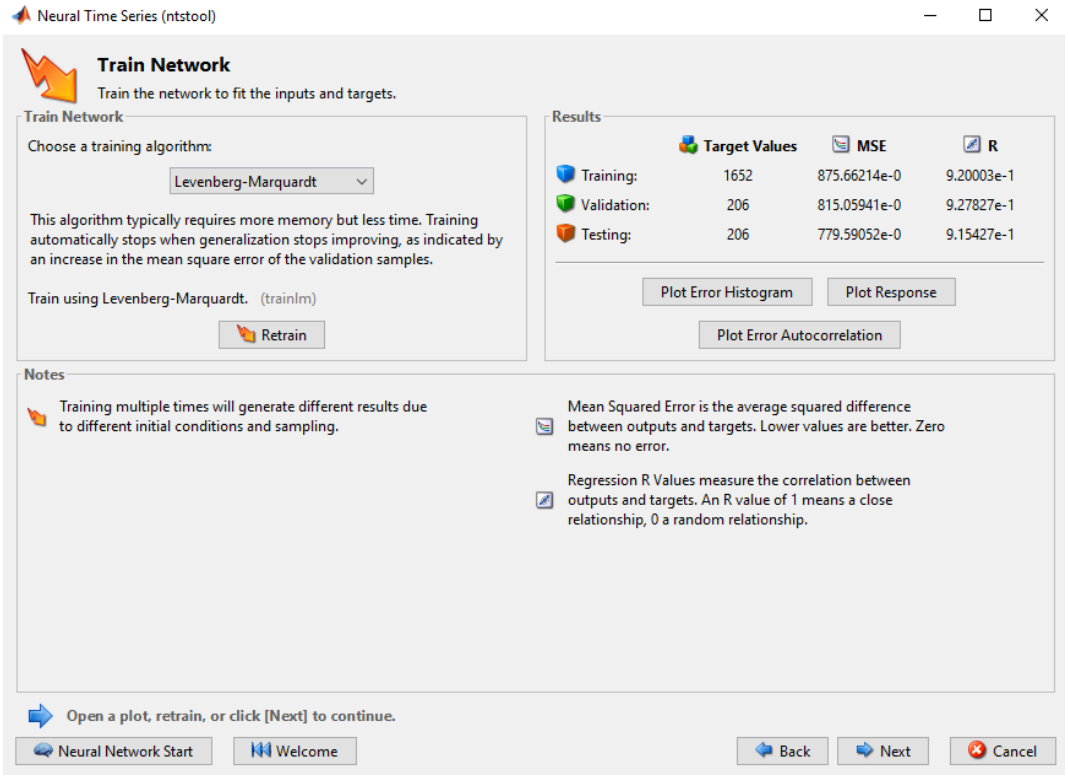



Figure 11B: Training the network to fit the inputs and targets in MATLAB

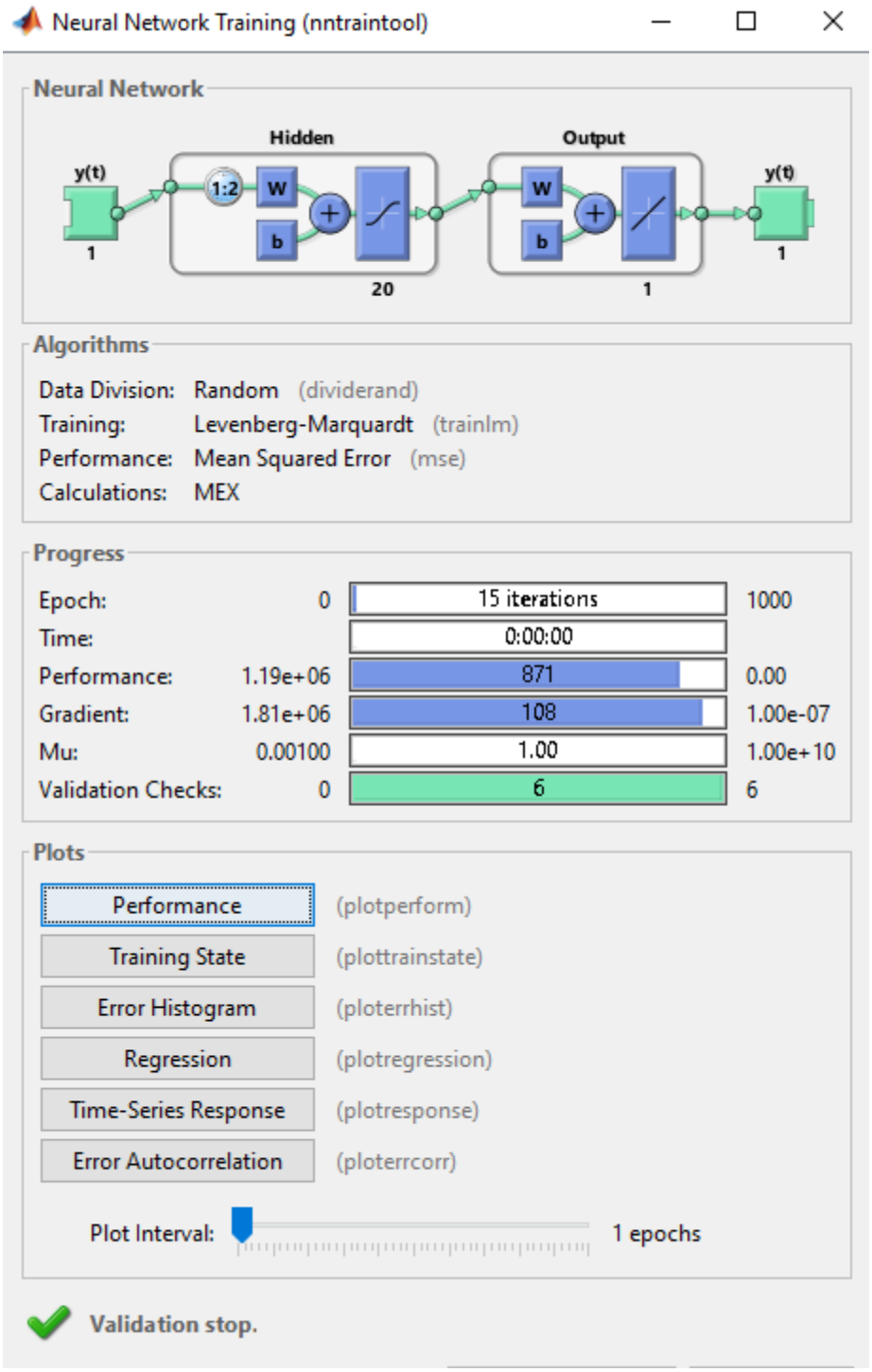



Figure 12B: ANNs Performance, Training State, and Time Series Response

Figure 13B: IKB05 River Gauging Station & Ogoche Stream

Figure 14B: Sand harvesting along the River Kuja Banks

Figure 15B: River Kuja at Obware Stream

Figure 16B: Nyatike Irrigation Scheme

Figure 17B: Gucha Migori Sango, Confluence of the two rivers

## Appendix C: Research License

REPUBLIC OF KENYA

NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Ref No: **878184**

Date of Issue: **18/July/2022**

**RESEARCH LICENSE**

**This is to Certify that Ms.. Edith Auma Onyango of Egerton University, has been licensed to conduct research in Migori on the topic: Modelling and forecasting of flood events in Gucha- Migori River Basin, Kenya for the period ending : 18/July/2023.**

License No: **NACOSTI/P/22/18968**

**878184**

Applicant Identification Number

Director General
NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Verification QR Code

NOTE: This is a computer generated License. To verify the authenticity of this document, Scan the QR Code using QR scanner application.

## Appendix D: Publication

# FORECASTING OF FLOOD EVENTS FOR WATER RESOURCES MANAGEMENT IN GUCHA-MIGORI RIVER BASIN, KENYA

Onyango. E. A. [1*], S. Nyakach [1], R. M. Wambua[2] and H. Otieno[2]

[1]Department of Agricultural Engineering, Faculty of Engineering and Technology, Egerton University, Nakuru County, Kenya.

[2]Department of Agricultural and Biosystem Engineering, South Eastern Kenya University, Kitui County, Kenya.

* Correspondence Author Email: edithauma@gmail.com

DOI: https://doi.org/10.37017/jeae-volume10-no1.2024-1
Publication Date: 18 March 2024

## ABSTRACT

Floods are devastating natural disasters frequently occurring in many river basins such as Gucha-Migori River Basin, Kenya. However, non-structural countermeasures such as hydrologic modelling, flood proofing, and continuous forecasting have not been fully explored and implemented to reduce risk, damage, and vulnerabilities of flood events. The major challenge in flood forecasting is the selection of the relevant probability distribution. The main objective of the study was to forecast flood events in Gucha-Migori River Basin for integrated water resource management. Daily hydrological datasets between 1969 and 2015 were obtained from Water Resources Authority. After the ranking of independent flood events had been achieved, return periods and flood frequencies were computed. The relationship between flood magnitudes and their respective frequencies was through modelling using probability distribution functions (PDFs). The evaluated PDFs include Normal Distribution, Log-Normal Distribution, Gumbel Distribution, and Log-Person Type III Distribution. Coefficient of Determination ($R^2$), Goodness of Fit test, and Best-Fit Distribution Curve guided selection of the suitable distribution model. The suitable probability distribution model out of the tested four was then applied to predict P-percent annual exceedance probabilities corresponding to specific flood magnitudes. From the results, out of the 132 flood events that occurred in the period 1969 to 2015, only 42 were identified as independent flood events. Based on the frequency analysis, the highest and lowest, and the average and standard deviation of the recorded independent flood events were 423.90, 240.8, 284.80 $m^3/s$, and 45.54 respectively. It was revealed that the probability of a flood event of a magnitude equal to or exceeding the lowest (240.8 $m^3/s$) and the highest (423.90 $m^3/s$) occurring for a particular year were 0.98 and 0.02 respectively. Gumbel distribution curve was selected to be the best fit for the Gucha-Migori River Basin. According to the derived Gumbel's formula, the estimated magnitude for the return period 1.05, 1.11, 1.25, 2, 5, 10, 25, 50, 100, and 200 years were 221.38, 231.92, 245.77, 278.01, 321.39, 350.11, 386.39, 413.31, 440.03, and 466.65 $m^3/s$. This research provides useful information for enhanced determination of probability occurrence of flood events and thus proper water resources planning and management strategies in the Gucha-Migori river basin.

**Keywords:** Flood Discharge, Probability Distribution Models, Forecasting, Flood Frequency

## 1.0. INTRODUCTION

Climate change and growing socio-economic developments continue to alter hydrological regimes and the response of several watersheds globally, thus threatening water, energy, and food security (Archer et al., 2010; IPCC, 2014; UNFCCC, 2005; Wambua et al., 2017). In Kenya, there is a study recognition of ecosystem deterioration, and spatio-temporal uncertainties