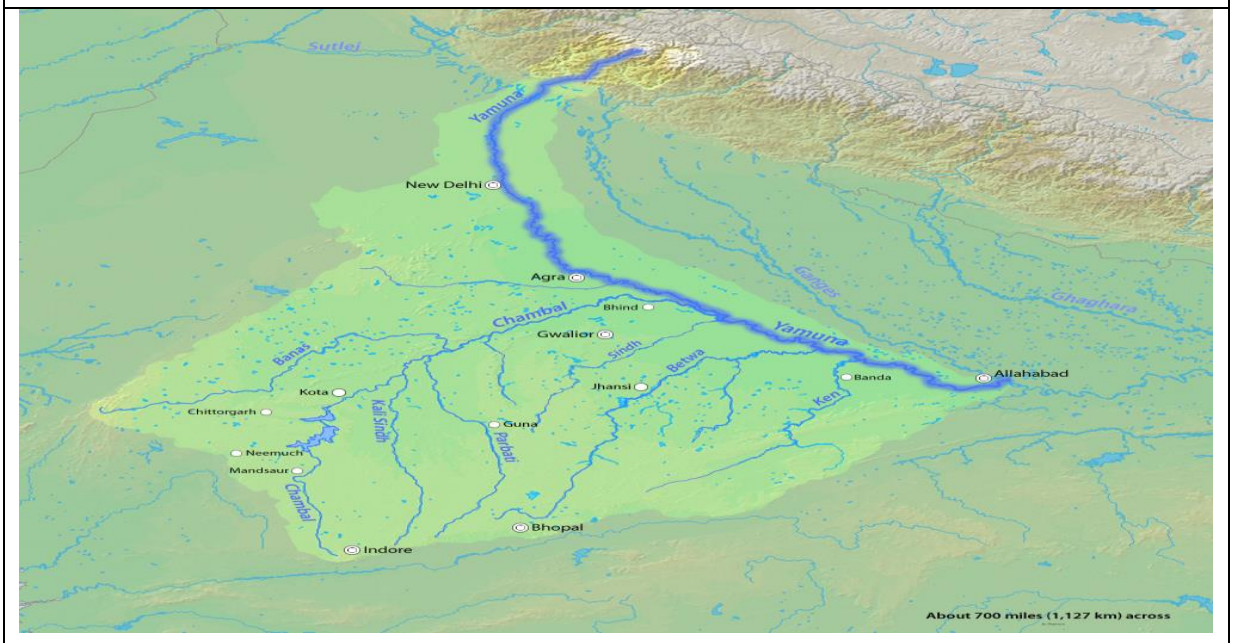
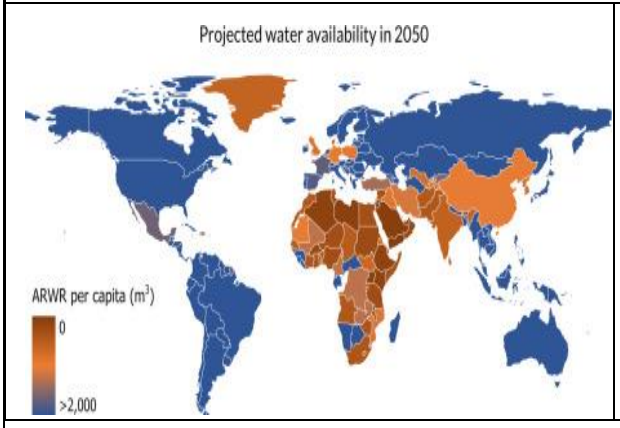
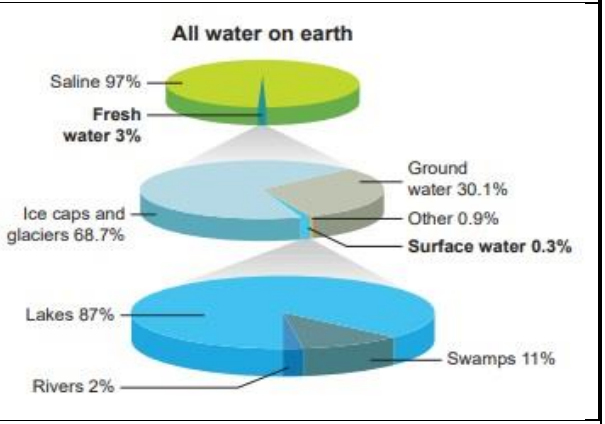
Final Report

Water Availability Assessment for Project Formulation in Sub Basins of Ganga River in Madhya Pradesh







Organization

National Institute of Hydrology, Roorkee(Uttarakhand),
and
Water Resources Department, Madhya Pradesh, India

Water Availability Assessment for Project Formulation in Sub-Basins of Ganga in Madhya Pradesh

FINAL REPORT

By

NATIONAL INSTITUTE OF HYDROLOGY REGIONAL CENTRE, BHOPAL (M.P.)

&

WATER RESOURCE DEPARTMENT, GOVT. OF MADHYA PRADESH

Water Availability Assessment for Project Formulation in Sub Basins of Ganga River in Madhya Pradesh

Duration of the project	**3 Years (Nov 2021 to Oct 2024)**
Lead Organization	**National Institute of Hydrology Central IndiaHydrology Regional Centre, Bhopal**
Study Team from Lead Organization	
PI	Dr. Rahul Kumar Jaiswal, Scientist-F, NIH-Bhopal
Co-PI (RC Bhopal) Co-PI (NIH Roorkee) Project Staff (RC Bhopal)	Dr. Ravi V. Galkate, Scientist-F Dr. A. K. Lohani Scientist-G Ms. Pushpanjali Kumari, Mr. Siddik Ahmed Barbhuiya
Partner Organization PI (State Data Centre, WRD)	Water Resources Department, Govt. of M.P. Dr. Brijendra Baghel, Data Base Administrator
Total Cost of Project (Rs)	**Rs. 30.00 Lakhs**
End Users/Beneficiaries of the Study	Water Resource Department, Govt. of Madhya Pradesh, and other water user agencies

PREFACE

Water is a vital resource for sustaining life and driving socio-economic development. In the state of Madhya Pradesh, India, where agriculture forms the backbone of the economy and industries are flourishing, the prudent management of water resources is of paramount importance. The Ganga River and its sub-basins play a significant role in providing water to the region, supporting diverse ecosystems, and meeting the growing demands of various sectors. However, effective project planning and formulation require accurate and comprehensive knowledge of water availability in these sub-basins. Water availability is one of the prerequisites for the preparation of the Detail Project Report (DPR) of water resource projects. When preparing a DPR, water availability at the dam site is a critical input, as it determines key factors such as gross storage, dam height, command area etc. Currently, in the absence of specific data at the project site, a rainfall-runoff relationship from nearby locations is used to estimate dependable water yield. In this study, sites across the basin have been selected based on homogeneity to develop robust rainfall-runoff relationships. Additionally, machine learning techniques have been employed to generate more precise and compressed regression equations. These equations can be applied by field officers to assess water availability in ungauged catchments of the Ganga River in Madhya Pradesh.

Internal Study titled "Water Availability Assessment for Project Formulation in Sub-basins of Ganga in Madhya Pradesh" was conducted by NIH, Regional Centre Bhopal, and WRD, Govt. of Madhya Pradesh, Bhopal as an internal study with the objectives to statistically analyses the hydroclimatic variables for homogeneity/similarity assessment, development of the rainfall-runoff model for gauged catchments using SCS-CN, linear regression, modified SCS-CN, NAPI, or soft computing techniques, assessment of water availability and its statistical/spatial characteristics in gauged catchments, development of regional relationships for transfer of information in ungauged catchments, and development of GIS/web-based application for the computation of water availability for ungauged catchments.

The results of the study will greatly help WRD MP during the formulation of projects in ungauged catchments as most of the catchments in MP are ungauged and project formulation becomes difficult for field engineers. The developed approach will offer an efficient and accessible tool, enabling quick and accurate water availability assessments, thus facilitating

more informed decision-making and sustainable water resource management in the region. The final report contains the various aspects of hydrological interaction in sub-basin, review of literature, GIS database, existing database and analysis, regression analysis, field data collection, etc. The report was prepared by Dr R. K. Jaiswal, Scientist-F as P.I. and Dr Ravi Galkate, Scientist-F, Miss Pushpanjali Kumari and Mr Siddik Ahmed Barbhuiya as Junior Resource Person from National Institute of Hydrology, Regional Centre Bhopal, Dr A. K. Lohani, Scientist-G as Co-PIs from National Institute of Hydrology, Roorkee and Dr Brijendra Baghel, Data Base Administrator, from State Data Centre, WRD, MP. This report is the results of final research works conducted by both organizations under this Internal Study.

(Dr. M. K. Goel)
Director,
NIH, Roorkee

Table of Content

List of Figures

List of Tables

ABSTRACT

Accurate water availability assessment is vital for sustainable management in India, where agriculture is crucial and industrial growth is rising. Effective management of the Ganga river and its sub-basins is essential for water supply and ecosystem support in Madhya Pradesh. A key challenge is formulation of water resources is to simulate flow in ungauged or poorly gauged catchments. This study addresses this challenge by focusing on six major sub-basins of the Ganga basin in Madhya Pradesh: Chambal, Betwa, Ken, Sindh, Son, and Tons. Key to this study is the application of the GR2M model (a two-parameter, conceptual monthly scale model) coupled with the Inverse Distance Weighting (IDW) for regionalization of parameters, the Soil Conservation Service-Curve Number (SCS-CN) method in Google Earth Engine for universal application, and the machine learning and linear based regression models for computation of monthly water yield from any ungauged basin. The GR2M model was employed to predict runoff across the basins, with calibration efficiencies ranging between 75.4% and 94.1% and validation efficiencies from 61.2% to 98%. The IDW method was used to estimate model parameters for ungauged sites, achieving efficiencies from 58% to 93.4%.

Regression analysis revealed significant variability in runoff predictions across different basins and months, with the Ken basin demonstrating high predictability, with R^2 values as high as 81.75%, while the Chambal basin exhibited notable variability, particularly in October, where R^2 dropped to 17.32%. Machine learning models, including Huber Regression, Decision Tree Regressor, Ridge Regression, and Random Forest, were integrated to enhance predictive accuracy, with performance metrics such as R^2 and RMSE guiding model optimization. For instance, the Betwa basin saw a peak R^2 of 47.69% in September, with the lowest RMSE of 16.59 cumecs in October.

Cluster analysis was employed to zonate the basins, grouping regions with similar hydrological characteristics, thereby improving model accuracy and reliability. This study underscores the effectiveness of integrating traditional hydrological models like GR2M with modern techniques such as machine learning, geospatial analysis, and the IDW method for comprehensive water availability assessment. The findings offer essential insights for water resource management, agricultural planning, and drought mitigation in the Ganga River sub-basins. It is recommended that the GR2M model with suggested regional parameters or regression equations can be used for computation of water availability in the sub-basins of the Ganga basin in Madhya Pradesh.

CHAPTER: 1
INTRODUCTION

1.1 General

The Ganga River basin is India's largest river basin, spanning 11 states and covering an area of over 861,000 square kilometers. As a vital water resource for the region, understanding the water availability and dynamics within the Ganga basin is crucial for effective water resource management and planning. This assessment focuses on evaluating the water availability in the sub-basins of the Ganga River located within the state of Madhya Pradesh. Madhya Pradesh is an important state within the Ganga basin, contributing significantly to the overall water resources. Thus, the continuous discharge data corresponding to the observed gauge can be obtained by developing a stage-discharge relationship and using this relationship to convert the recorded stages into corresponding discharges. This relationship is determined by correlating measurements of discharge with the corresponding observations of stage (Maidment et al., 1996). The availability of freshwater has been recognized as a global issue, and the reliable evaluation and quantification of it within the basin is necessary to bolster the sustainable management of water. Effective water resources management necessitates an understanding of streamflow characteristics, including different flow quantiles, to make informed decisions. This knowledge is crucial for developing systems for hydropower generation, flood control, irrigation, and agricultural purposes. However, many regions around the world suffer from inadequate or nonexistent flow-gauging networks (Kapangaziwiri et al., 2012). Hydrologists have traditionally employed techniques to transfer flow data from gauged (i.e., donor) basins to ungauged ones. These techniques are commonly known as regionalization methods.

As the global population continues to grow, ensuring a sustainable and reliable supply of water has emerged as a critical challenge for governments, communities, and industries. The assessment of water availability plays a pivotal role in managing water resources efficiently and formulating sound policies and projects. This process involves evaluating the quantity and quality of water present in each region and understanding its spatial and temporal distribution based on various factors, including hydrological patterns, precipitation trends, and human demands (Bera et al., 2021; Anand et al., 2018). Water availability assessment provides essential information for developing effective water management strategies, supporting

sustainable development plans, and ensuring the equitable distribution of water resources among competing demands. Such assessments are indispensable for guiding decision-makers in water-related infrastructure projects, disaster preparedness, and climate change adaptation efforts.

Assessing water availability in river basins has been the focus of numerous studies, particularly concerning the Ganga River System and similar large river systems. A large number of regionalization methods have been proposed in the literature since the “decade on the prediction in ungauged basins” of the International Association of the Hydrological Sciences (IAHS) (Sivapalan et al., 2003). Various hydrological models and methods have been employed to predict runoff based on rainfall data, each with its strengths and limitations. Numerous studies have attempted to predict runoff based on rainfall data using various hydrological models and methods. One widely used approach is the Soil Conservation Service - Curve Number (SCS-CN) method, which estimates direct runoff from rainfall events by taking into account land cover and soil characteristics. This method's simplicity and effectiveness have made it a popular choice in hydrological studies. For instance, Ponce and Hawkins (1996) demonstrated the SCS-CN method's applicability in different hydrological settings, highlighting its robustness in estimating runoff for agricultural and urban areas. However, despite its widespread use, the SCS-CN method and other traditional models often face limitations when applied to ungauged or data-scarce regions. Studies by Sivapalan et al. (2003) and Blöschl et al. (2013) have emphasized the challenges associated with transferring model parameters from gauged sites to ungauged sites, often resulting in reduced accuracy and reliability. These challenges underscore the need for innovative approaches that can enhance the transferability and applicability of hydrological models across diverse regions. While the SCS-CN method provides a strong foundation, the need to enhance its applicability in ungauged basins has led researchers to explore additional methodologies. The GR2M model, a monthly water balance model, has been employed in various studies to account for the variability in precipitation and evapotranspiration. The GR2M model accounts for monthly variations in precipitation and evapotranspiration, making it suitable for regions with significant seasonal changes. This model's simplicity, requiring only two parameters, makes it particularly useful for regional applications where data availability might be limited. The use of the GR2M model in conjunction with other methods, such as the SCS-CN, enhances its effectiveness in predicting runoff. For instance, the combination of GR2M and SCS-CN

provides a comprehensive approach to monthly water balance and runoff estimation, addressing both short-term and long-term hydrological processes.

One common approach is the hydrological model parameter regionalization method. This technique enables hydrologists to estimate the model parameters for an ungauged site. By using observed meteorological data, they can then run a hydrological model to produce continuous streamflow data for that site, typically on a daily or monthly basis. Alternatively, hydrologists might estimate hydrological signatures to determine the behavioral characteristics of the ungauged site, allowing them to generate simulations that align with these characteristics (Bárdossy et al., 2007; Yadav et al., 2007). Another method involves directly regionalizing flow quantiles, which is useful for obtaining design metrics (Shu and Ouarda, 2007). To address the challenge of parameter transferability from gauged to ungauged catchments, studies have incorporated spatial interpolation techniques such as Inverse Distance Weighted (IDW). IDW technique has been effectively applied to interpolate rainfall data, enhancing the spatial representation of hydrological models in various regions, including Andean catchments. By integrating IDW with models like GR2M, information can be efficiently transferred to ungauged catchments, thus improving the accuracy of water resource assessments in data-scarce areas. Comparative studies of different interpolation techniques have shown that IDW performs well even when compared to more complex methods such as kriging, particularly in regions with limited data availability. The simplicity and ease of implementation of IDW make it a practical choice for numerous hydrological applications. Beyond rainfall interpolation, IDW has also been successfully used to estimate streamflow and runoff in ungauged catchments. Its application in interpolating streamflow data across river networks has demonstrated that IDW can effectively transfer information from gauged to ungauged sites, thereby enhancing the accuracy of streamflow predictions in areas lacking sufficient data.

The lack of long-term runoff data in the Ganga basin, particularly in Madhya Pradesh, poses a significant challenge to effective water resource management. However, using innovative technologies, community-based monitoring, institutional reforms, and international collaboration, it is possible to overcome this challenge. By improving the availability and accuracy of hydrological data, water resource engineers and policymakers can develop more reliable models, make informed decisions, and ensure the sustainable management of water resources in the region.

The study will commence with an in-depth analysis of historical runoff records in the

region to identify areas with similar hydrological characteristics, leading to the establishment of homogeneous zones. Subsequently, rainfall-runoff models will be developed using various techniques, enabling the understanding of the intricate relationships between precipitation and runoff in these sub-basins. By correlating model parameters with easily accessible topographical and geomorphological properties, the study seeks to enhance the accuracy and applicability of the developed models. The validation process will play a pivotal role in ensuring the robustness and reliability of the models. By rigorously testing the equations against catchments not utilized in model development, the study will demonstrate the models' capacity to generalize and predict water availability in ungauged areas. This step is crucial in extending the methodology's practical utility across the Ganga sub-basins in Madhya Pradesh.

The primary research addressed by this study is the development of reliable and transferable rainfall-runoff models that can be applied to both gauged and ungauged sub-basins within the Ganga River System in Madhya Pradesh. While traditional models like the SCS-CN method provide a foundation for runoff estimation, their reliance on historical data from gauged sites limits their applicability in ungauged regions. Furthermore, existing models often do not fully capture the spatial variability in hydrological responses due to differences in rainfall patterns, land use, and soil properties across sub-basins. The study employs the GR2M model, a robust monthly water balance model, to predict water availability in the Ganga River System. The GR2M model accounts for the variability in precipitation and evapotranspiration monthly, making it particularly suitable for regions with significant seasonal variations. To enhance the model's applicability in ungauged catchments, the Inverse Distance Weighting (IDW) interpolation technique is used. IDW helps in interpolating spatial data, thereby transferring model parameters from gauged (donor) sites to ungauged (receiver) sites. This combination of the GR2M model and IDW technique addresses the challenge of parameter transferability and improves the model's accuracy and reliability in ungauged regions. The regression models developed are rigorously validated against observed runoff data. This validation process ensures that the models accurately represent the hydrological behavior of the region. The use of historical data enhances the model's reliability, while the incorporation of real-time data capabilities ensures its relevance for future applications.

1.2 Origin of Project

One prevalent challenge in this domain is simulating flow in ungauged or poorly gauged

catchments. This issue is particularly pressing in Madhya Pradesh, India, where agriculture forms the backbone of the economy and industries are rapidly flourishing. Effective and prudent management of water resources is paramount in this region, as the Ganga River and its sub-basins are integral to providing water, supporting diverse ecosystems, and meeting the burgeoning demands of various sectors. Effective project planning and formulation in Madhya Pradesh hinge on accurate and comprehensive knowledge of water availability in the sub-basins of the Ganga River. One of the critical challenges faced by water resource engineers and policymakers in this region is the lack of long-term runoff data for many catchments in the Ganga basin. This scarcity of data poses a significant impediment to the formulation of projects that rely on a reliable assessment of water availability. Without this information, the planning process faces significant uncertainties, potentially resulting in inefficient water resource management and heightened risks for water-dependent activities.

Historically, the Ganga River System has served as the lifeline of northern India, sustaining rich agricultural landscapes and dense populations. The river's seasonal flow patterns are heavily influenced by the monsoon rains, which account for most of the annual precipitation. Understanding and predicting the river's response to rainfall events are crucial for effective water resource management, flood control, and agricultural planning. However, the vast size and diverse topography of the basin pose significant challenges for hydrological modeling and water resource assessment. Many areas within the basin are ungauged, meaning there are no direct measurements of streamflow. This absence of data complicates efforts to develop accurate hydrological models, which are essential for predicting runoff, managing floods, and allocating water resources effectively. Overcoming these challenges requires the development and implementation of innovative data collection methods and modeling techniques to ensure sustainable water resource management in the region. One promising approach to address the challenge of data scarcity in the Ganga basin is the use of remote sensing and Geographic Information Systems (GIS). These technologies can provide valuable data on land use, soil type, vegetation cover, and rainfall, which are essential for hydrological modeling. Remote sensing can offer continuous spatial coverage and temporal resolution, making it possible to monitor changes in the basin over time. Coupled with GIS, this data can be used to develop detailed hydrological models that simulate the flow of water in ungauged catchments. Another innovative method is the use of hydrological modeling techniques such as rainfall-runoff models, which can predict streamflow based on rainfall data and watershed

characteristics. These models can be calibrated and validated using data from gauged catchments and then applied to ungauged catchments to estimate streamflow. Citizen science and community-based monitoring are also emerging as valuable tools for data collection in the Ganga basin. Engaging local communities in monitoring efforts can help gather data from remote and inaccessible areas. This participatory approach not only provides valuable data but also raises awareness and fosters a sense of ownership among local populations, which is crucial for the sustainable management of water resources.

This research aims to fill this gap by developing regional models that leverage the power of advanced data processing platforms and incorporate clustering analysis to classify the study area into distinct hydrological zones. By doing so, the study enhances the understanding of spatial variations in water availability and hydrological processes, ultimately improving the accuracy and transferability of the developed models. Developing models that are not only accurate but also transferable across different hydrological settings. Traditional models often struggle with parameter transferability, especially in ungauged basins. By employing clustering analysis and leveraging GEE, this study provides a robust solution to this problem. The integration of modern data processing platforms with traditional hydrological methods marks an innovative step forward in the field

1.3 Objectives

The main objectives envisaged the PDS study are:

- Statistical analysis of hydroclimatic variables for homogeneity/similarity assessment.
- Development of the rainfall-runoff model for gauged catchments using SCS-CN, linear regression, modified SCS-CN, NAPI, or soft computing techniques.
- Assessment of water availability and its statistical/spatial characteristics in gauged catchments.
- Development of regional relationships for the transfer of information in ungauged catchments.
- Development of GIS/web-based application for the computation of water availability for ungauged catchments.

CHAPTER: 2
REVIEW OF LITERATURE

2.1 General

Understanding the hydrological cycle and the complex processes that govern rainfall-runoff relationships is crucial for effective water resource management, especially in ungauged basins where direct measurement data is unavailable. This chapter explores various methodologies and models employed in the estimation of runoff in such ungauged catchments, with a focus on the regionalization of hydrological model parameters. The review discusses regression analysis, inverse distance weighting (IDW), the SCS-CN curve number method, and the application of hydrological models like GR2M, among others, highlighting their utility and effectiveness in enhancing runoff predictions in diverse climatic and geographical contexts. By examining these approaches, the chapter aims to provide a comprehensive overview of current advancements and challenges in hydrological modeling, particularly in regions with limited data availability.

A hydrological system is defined as a spatial structure or volume bounded by a boundary that receives water and other inputs, processes these inputs internally, and produces outputs (Chow et al., 1988). The precipitation is all the form of water either snow fall or rainfall which reach the earth surface from the atmosphere. The flowing of precipitation is known as the from the catchment is known as the Runoff This will satisfy all the surface and subsurface loses (Dubayah et al., 1997). In a basin if runoff data is not available then we can call the basin as Ungauged basin. So, the rainfall runoff model parameter cannot be obtained by the calibration of the direct runoff data hence we need some other methods to obtain the runoff in ungauged basins (Bloschl, 2006).

2.2 SCS-CN Model

The SCS-CN method is a widely recognized hydrological tool for estimating direct runoff or infiltration from rainfall. Ponce and Hawkins (1996) provide a comprehensive review of the method, discussing its development, applications, limitations, and potential areas for improvement, emphasizing how it has evolved since its inception. Mishra and Singh (2003) offer an in-depth analysis of the SCS-CN methodology, exploring its theoretical foundation

and empirical basis, along with modifications and case studies illustrating its application across various landscapes and hydrological conditions. Hawkins (1978) examines the influence of antecedent moisture conditions on curve numbers, highlighting the adaptability of the SCS-CN method to different soil moisture scenarios. Garen and Moore (2005) review the use of the SCS-CN method in water quality modelling, identifying common misapplications and suggesting future research directions to enhance its effectiveness in environmental studies. Bonta (1997) explores statistical methods to derive curve numbers from observed runoff data, providing insights into the spatial variability of hydrological responses within watersheds. Hernandez et al. (2000) investigate the impact of land cover and rainfall variability on runoff predictions using the SCS-CN method, emphasizing its application in semi-arid environments. Rallison and Miller (1981) provide a historical perspective on the development of the SCS-CN method, discussing its theoretical underpinnings and practical applications over the years. Additionally, Pandey, Chowdary and Mal (2009) demonstrate the integration of the SCS-CN method with other hydrological models, such as the WEPP model, to model runoff and sediment yield in agricultural watersheds, highlighting its utility for comprehensive watershed management. These studies collectively underscore the versatility and robustness of the SCS-CN method, making it a critical component in hydrological modelling and water resource planning. Furthermore, the USDA Natural Resources Conservation Service (NRCS) provides technical references and manuals that offer detailed explanations of the SCS-CN method, including guidelines for its application in various hydrological and environmental studies, thereby reinforcing its relevance and practical utility.

2.3 GR2M Model

Makungo *et al.* (2010) conducted a study utilizing the GR2M (a monthly lumped rainfall-runoff) model to simulate streamflow in semi-arid catchments. Their research aimed to assess the model's capability to accurately capture the hydrological variations that occur on a monthly basis within these regions. The GR2M model, known for its simplicity and efficiency, was particularly suited for the semi-arid catchments due to its minimal data requirements and robustness in handling sparse datasets. Through their simulations, GR2M model effectively reproduced the observed streamflow patterns, highlighting its reliability in depicting the monthly fluctuations in water flow. This validation of the model's performance underscored its potential as a valuable tool for hydrological modeling and water resource management in semi-

arid environments.

Ditthakit et al. (2023) provided a comprehensive analysis of the GR2M model's performance, showcasing its effectiveness in producing accurate runoff predictions with limited data. This study highlighted the model's practical utility for regional hydrological studies and water resource management. Study conducted an extensive evaluation of the GR2M model across multiple catchments with varying climatic and geographical conditions. The study emphasized the model's robustness in areas with sparse hydrological data, demonstrating that GR2M could achieve reliable runoff estimates despite minimal input information. The findings underscored the GR2M model's potential to streamline hydrological modelling in regions where data collection is challenging or resources are limited. By effectively predicting runoff with only two calibration parameters, the model proved to be an invaluable tool for water resource managers and researchers working in under-resourced or remote areas. The study also explored the model's sensitivity to parameter changes, revealing its stability and adaptability to different hydrological contexts.

Further validation of the GR2M model came from a comparative study by Oudin et al. (2005), which tested its performance against other hydrological models across European catchments. The study found that GR2M performed comparably well, particularly in catchments with limited data, emphasizing its parsimonious nature and ease of calibration. The model's ability to deliver reliable results with minimal data has made it a popular choice for hydrological simulations in various climatic and geographic settings, reinforcing its relevance and utility in modern hydrological research.

2.3.1 Regional Approach

2.3.1.1 Inverse Distance Weighting (IDW)

By providing a robust framework for runoff prediction, the GR2M model supports effective water resource planning and management, especially in regions facing data scarcity. Its integration with other hydrological techniques and models, such as the IDW interpolation method, further enhances its applicability, making it a critical component of contemporary hydrological studies. IDW is a widely used spatial interpolation technique that estimates unknown values based on the distance between known data points. The principle underlying IDW is that points closer to each other are more likely to have similar values than those farther apart. This method has found extensive applications in various fields, including hydrology, meteorology, and environmental sciences, due to its simplicity and effectiveness in

interpolating spatial data.

Merwade et al. (2008) applied IDW to interpolate streamflow data across river networks in the United States. IDW is a type of spatial interpolation technique where the estimated value at an ungauged site is calculated based on the values at nearby gauged sites, with closer sites having a greater influence on the estimation. This method assigns weights inversely proportional to the distance between the gauged and ungauged sites, meaning that closer sites have a larger impact on the interpolated values than those farther away. They found that IDW could successfully transfer streamflow information from gauged to ungauged sites. The effectiveness of IDW in their study was evident in its ability to enhance the accuracy of streamflow predictions, particularly in regions where direct measurement data were sparse or unavailable. The method allowed for more reliable estimates by leveraging the existing data from nearby gauged locations, thereby filling in the gaps in streamflow information across extensive and complex river networks. Their research highlighted the potential of IDW as a valuable tool in hydrology, especially for improving water resource management and planning in areas with limited gauging infrastructure. Demonstrating the IDW could produce accurate streamflow predictions, contributed to the advancement of techniques for regionalization, ultimately aiding in better decision-making for flood management, irrigation planning, and other water-related needs.

The study by Xu et al. (2015) aimed to improve surface runoff predictions in the Loess Plateau of China, a region characterized by complex topography, variable rainfall patterns, and significant soil erosion. Accurate estimation of surface runoff in such regions is crucial for effective water resource management, soil conservation, and flood control. Traditional hydrological models often face difficulties in accurately predicting runoff due to the spatial variability of rainfall and soil moisture, especially in regions with sparse data. To address these challenges, integrated the Inverse Distance Weighting (IDW) interpolation technique with a hydrological model. The primary objective was to enhance the spatial representation of rainfall and soil moisture data, thereby improving the accuracy of runoff predictions. The interpolated data were integrated into a hydrological model specifically tailored to the characteristics of the Loess Plateau. By combining IDW with the hydrological model, the study was able to account for the spatial variability of rainfall and soil moisture, leading to more accurate runoff predictions. The results demonstrated that the integration of IDW significantly improved the model's performance, showcasing the potential of combining IDW with hydrological models

to enhance water resource assessments. This approach provided a robust framework for predicting surface runoff in data-scarce regions, highlighting the practical benefits of using advanced interpolation techniques in hydrological studies.

With advancements in geospatial technologies and data processing platforms, the application of IDW has become more sophisticated. The integration of IDW with Geographic Information Systems (GIS) and remote sensing data has enhanced its capability to process large datasets and produce high-resolution spatial interpolations. Gao et al. (2017) used IDW within a GIS framework to interpolate rainfall data and assess its impact on flood modeling in the Yangtze River basin. The study demonstrated that GIS-based IDW interpolation significantly improved the spatial resolution of rainfall inputs, leading to more accurate flood predictions.

Studies by Kumar et al. (2015) explored the integration of satellite-derived rainfall estimates with ground-based measurements to enhance the spatial representation of rainfall using the Inverse Distance Weighting (IDW) method. This innovative approach was particularly advantageous in regions with sparse gauging networks, where the limited availability of ground-based data often hampers accurate rainfall analysis and prediction. Satellite-derived rainfall estimates offer a broad spatial coverage and can capture rainfall patterns over large and remote areas, which are typically underserved by traditional ground-based rain gauges. By combining these satellite data with available ground measurements, study aimed to leverage the strengths of both data sources. The satellite data provided a comprehensive spatial overview, while the ground measurements offered localized accuracy and validation points. Using the IDW method, the researchers interpolated rainfall data by assigning weights to the measurements based on their distances from the points of interest. In this context, the closer a ground-based measurement or satellite estimate was to the point of interest, the more influence it had on the interpolated value. This approach allowed for a more nuanced and accurate spatial representation of rainfall, addressing the gaps left by sparse ground networks. Study demonstrated that integrating satellite-derived rainfall estimates with ground-based measurements significantly improved the spatial accuracy of rainfall representation. This hybrid method provided a more reliable basis for various hydrological applications, such as water resource management, flood forecasting, and agricultural planning. By enhancing the precision of rainfall data in regions with limited gauging infrastructure, their work contributed to better decision-making and management of water-related issues in these areas.

Li and Heap (2011) conducted a comprehensive review of spatial interpolation methods, including IDW, kriging, and spline interpolation. Their study highlighted that while IDW is less complex and computationally intensive than kriging, it performs comparably well in many practical applications. The choice of interpolation method often depends on the specific requirements of the study, including data availability, spatial resolution, and the nature of the variable being interpolated.

Buytaert et al. (2005) applied the IDW technique to interpolate rainfall data in Andean catchments, enhancing the spatial representation of hydrological models. The integration of the IDW technique with models like GR2M ensures that information can be effectively transferred to ungauged catchments, improving the accuracy of water resource assessments in data-scarce regions.

Goovaerts (2000) conducted a comparative study of different interpolation techniques, including IDW, for mapping rainfall in a semi-arid region. The study concluded that IDW performed well compared to more complex methods such as kriging, especially in areas with limited data availability. The simplicity and ease of implementation of IDW made it a practical choice for many hydrological applications. Beyond rainfall interpolation, IDW has also been used to estimate streamflow and runoff in ungauged catchments.

2.4 Regression Analysis

Regression analysis is a key approach in regionalization studies aimed at estimating runoff in ungauged catchments. This literature review highlights significant studies that have utilized regression and other regionalization techniques across various geographical contexts. Parajka et al. (2005) worked out on 213 catchments in Australia. the objective of study was to transfer the model parameter from gauged to ungauged basin to produce the daily runoff the hydrological model. Author used HBV ((lumped model but allowing different model states in different elevation) in this study. They have used 5 different Regionalization method which are Arithmetic mean (global and local mean), spatial proximity, Regression, Physical similarity, Multiple linear regression. McIntyre et al. (2005) carried out their study on 127 catchments in UK The objective of their study was Predictions of daily runoff in ungauged catchments The Hydrological model used in this study was Probability distribution model (PDM) of Moore. The regionalization method used in this study was Regression method and Ensemble modelling and similarity weighted averaging (SWA) method. The Ensemble modelling & SWA

method provides the best results and performs significantly.

Young (2006) carried out their work on 260 UK catchments and young (2006) have tried to regionalize the parameter of PDM model. In this study regression based, spatial proximity and physical similarity method have been compared and as per this study regression-based approach performed well among all. Key et al. (2006) worked out on 119 catchments in UK. The author basically compared the performance of two model PDM and TATE. The author has physical similarity method was performed well in case of PDM and regression-based approach performed well in case of TATE also the author has concluded that the selection of regionalization is model dependent.

Oudin et al. (2008) studied on 913 French catchments and the climatic condition of the catchment is warm temperature. Objective of the study was calculating or estimation of daily continuous flow in the ungauged catchment by using GR4J and TOPOMODEL. The different regionalization method used in the study was Spatial proximity, Physical similarity, Multiple regression method According to this study spatial proximity provide the best result. Zhang and chiew (2009) work out on 210 southeast Australian catchments. author have used SIMHYD model with different regionalization technique for estimating runoff in ungauged catchment. The author has concluded that combination of spatial proximity and physical similarities method perform well.

Li et al. (2019) applied the Xinanjiang hydrological model with two regionalization approaches for estimating the runoff in ungauged basin. The author has concluded that the both physical and spatial proximity method give similar results. Makungo et al. (2010) worked out on A quaternary catchment in Nzhelele river catchment in South Africa (92 km 2). The objective of their study was Generating natural streamflow in ungauged catchment for that they have used MIKE 11 NAM and AWBM hydrological method and the regionalization method used in the study was modified nearest neighboring method. The approach is applied in near-real-time modelling.

Samuel et al. (2011) have estimated the daily stream flow of ungauged basin on provenance of Toronto. The author has used MAC- HBV with spatial proximity method, physical similarity method, regression based approached for transferring data from gauged to ungauged basin. Author concluded that the combination of spatial proximity and physical similarities produce best model performance

Swain et al. (2015) worked out on the Koel river basin which is a sub-basin of Brahmani- Baitarni river. Author provides an effective approach which deals with the geomorphologic features of the ungauged drainage basin and derives the GIUH based transfer function and consequently geomorphologic unit hydrograph (GUH) for the study area. Geomorphologic instantaneous unit hydrograph (GIUH) approach can be used as a transfer function for modelling the rainfall – runoff process. Author concluded that GUH outperformed The CWC approach.

Yang et al. (2018) carried out their study on 118 catchments in Norway, which is in northern Europe on the western and northern part of the Scandinavian Peninsula. Author in this study evaluated strengths and limitations of existing regionalization methods in predicting ungauged stream flows in the high latitudes, large climate and geographically diverse, seasonally snow-covered mountainous catchments of Norway. The regionalization methods were evaluated using the water balance model – WASMOD (Water and Snow balance Modelling system). Regionalization method used in this study was Spatial proximity Physical similarity, Kriging and regression-based method. According to this study Spatial proximity and physical similarity perform better than the regression-based method. The combination of spatial proximity and physical similarity slightly improve the simulation in ungauged catchment but classifying the catchments into homogeneous groups did not improve the simulations in ungauged catchments in our study region

Li Q et al. (2019) carried out their study on 8 catchments of the Nansihu Lake is the largest freshwater lake in northern China. The objective of the study is to estimate the continuous runoff in the ungauged catchment by using combined transferring methods which are parameter transferring and area ratio to provide daily monthly continuous runoff in the ungauged catchment. They have discussed the catchment classification and different parameter transfer techniques and how their application will further improve the ability to estimate runoff in ungauged basins. The combined method outperformed the individually applied methods.

Arsenault et al. (2019) carried out a study on 30 catchments in Mexico. Authors predicted the streamflow in the ungauged basins using 3 hydrological models i.e., GR4J, HMETS, and MOHYSE using multiple linear regression (MLR), spatial proximity (SP), and physical similarity (PS) regionalization techniques in the study. Transferring complete parameter sets from a neighboring catchment provided the most robust method to estimate streamflow in semi-arid and humid ungauged basins. The model performance for arid

catchments was worse in the context of regionalization, with GR4J being more robust than the other models due to its simpler structure.

Kanishka et al. (2020) conducted a study on 30 watersheds within the Godavari River basin in India, focusing on ungauged catchments through regionalization methods. They employed the SWAT hydrological model and applied regionalization techniques such as global average, IDW, and physical similarity. Their findings indicate that classifying watersheds using isomap and applying regionalization based on physical similarity enhances the accuracy of streamflow estimation in ungauged basins.

Golian S et al. (2021) applied different models on 44 catchments across Ireland. They used GR4J and GR6J models for discharge simulation and compared the performance of the simulating high, median, and low flow conditions. To regionalize the model parameter the author has used multiple linear regression (MLR), non-linear regression (NL), and random forests (RF). It was concluded that all of the above methods have performed well in the case of average flow condition and the regionalization of the GR6J model parameter using RF has performed well in the case of low flow conditions.

CHAPTER: 3
STUDY AREA

3.1 General

The study area for this research is the Ganga basin located in Madhya Pradesh (MP) comprises several sub-basins, including Betwa, Chambal, Ken-Dhasan, Sindh, Son, and Tons. Madhya Pradesh is abundant in water resources, featuring major rivers such as the Narmada, Tapi, and Mahi, along with key tributaries of the Ganga basin, including Chambal, Sindh, Betwa, Ken, Tons, and Son. In this region, rivers such as Chambal, Betwa, and Ken, along with their tributaries, flow northward through the Bundelkhand region and eventually merge with the Yamuna River. The Son River, on the other hand, flows in an East-North-East direction. Figure 3.1 illustrates the location of the study area. The approximate catchment areas of the Ganga basin sub-basins in MP are as follows: Betwa – 19,365 km², Chambal – 59,940 km², Dasan – 829 km², Ken – 24,785 km², Sindh – 1,235 km², Son – 28,880 km², and Tons – 8,460 km². Geographically, the basin extends between 75°45’ and 85°1’ E longitudes and 26°47’ and 22°27’ N latitudes.

3.2 Basin Description

3.2.1 Betwa Basin

The Betwa basin, extending approximately between 22.87° to 26.02° North latitude and 77.10° to 80.24° East longitude, is a significant geographical and ecological area within Madhya Pradesh, India. The Betwa river, originating from the Vindhya Range in Raisen district, flows through Madhya Pradesh and Uttar Pradesh before merging with the Yamuna River. Key tributaries in Madhya Pradesh include the Jamni, Dhasan, and Bina rivers. The basin's topography is characterized by undulating plains, plateaus, and valleys, creating a diverse landscape. The basin supports rich biodiversity with notable vegetation, including teak, Sal and, various grasses, forming dense forests and scrublands. Human activities in the Betwa basin are predominantly agricultural, with crops like wheat, soybeans, and pulses, sustained by river irrigation. The region also has some industrial activities, particularly related to food processing and agro-industries.

3.2.2 Chambal Basin

The Chambal basin, situated between approximately 22.45° to 26.13° North latitude and 74.75° to 77.64° East longitude, is a significant geographical and ecological area in Madhya

Pradesh, India. The Chambal river, originating from the Vindhya range near Mhow, flows through Madhya Pradesh, Rajasthan, and Uttar Pradesh before joining the Yamuna river. Key tributaries within Madhya Pradesh include the Banas, Kali Sindh, and Parbati rivers. The basin features diverse topography, including plateaus, ravines, and plains, with the Chambal ravines being particularly distinctive due to their deep, steep-sided gullies. The National Chambal Sanctuary, a protected tri-state area, is renowned for its critically endangered species such as the Ganges river dolphin, gharials, and red-crowned roof turtles, alongside other wildlife like mugger crocodiles and Indian skimmers. The surrounding landscapes, featuring riverine forests to grasslands, host various plant species like babul and khair. The primary human activities here are split between agriculture and the conservation efforts aimed at protecting the basin's unique wildlife.

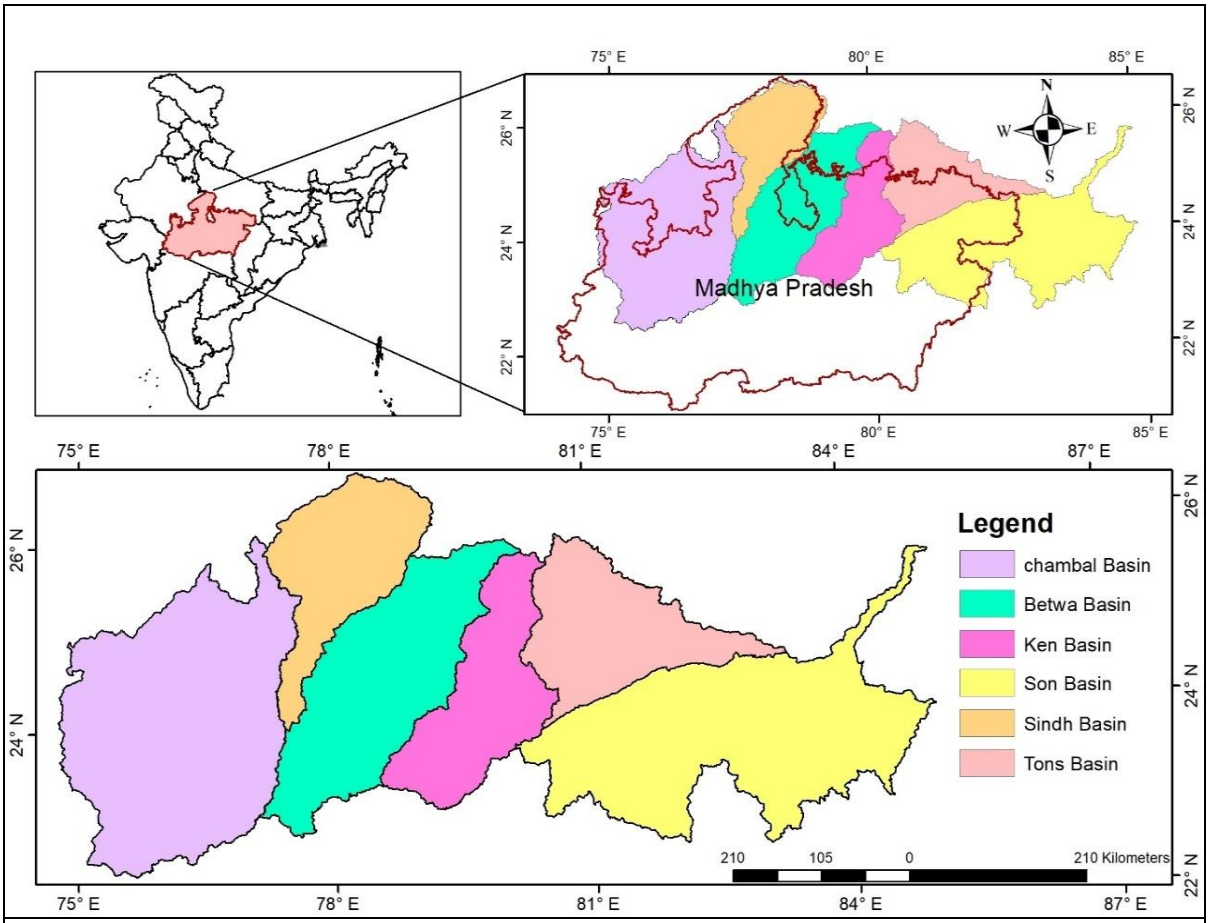


Figure 3. 1 Study area Map, Sub-Basin of Ganga in Madhya Pradesh.

3.2.3 Ken Basin

The Ken basin, extending approximately between 23.13° to 25.88° North latitude and 78.51° to 80.62° East longitude, is a prominent geographical and ecological region within

Madhya Pradesh, India. The Ken River, originating from the Kaimur Range near Ahirgawan in Jabalpur district, flows through Madhya Pradesh and Uttar Pradesh before merging with the Yamuna River. Key tributaries in Madhya Pradesh include the Sonar, Bearma, and Kopra rivers. The basin is marked by a diverse topography, characterized by plateaus, valleys, and gorges, which contribute to its distinct landscape. Rich in biodiversity, it houses mixed deciduous forests dominated by Teak and Sal, which serve as a habitat for a variety of wildlife. The region is also noted for its artisanal industries, such as handloom weaving and handicrafts, adding a unique cultural layer to its ecological significance.

3.2.4 Sindh Basin

The Sindh basin, extending approximately between 24.02° to 26.79° North latitude and 77.18° to 79.22° East longitude, is a significant geographical and ecological region within Madhya Pradesh, India. The Sindh River, originating from the Malwa Plateau in Vidisha district, flows through Madhya Pradesh and Uttar Pradesh before joining the Yamuna River. Key tributaries within Madhya Pradesh include the Mahuar, Pahuj, and Parbati rivers. The basin's water availability is dependent on both surface and groundwater resources, with irrigation systems significantly reliant on monsoon rainfall. The region receives an average annual rainfall ranging from 437.50 mm to 774.98 mm, which is critical for agriculture. Additionally, the Sindh Basin faces water management challenges, including equitable distribution and efficient utilization of available water resources to sustain agricultural activities.

3.2.5 Son Basin

The Son Basin, extending approximately between 22.64° to 25.71° North latitude and 80.10° to 85.02° East longitude, is a prominent geographical and ecological area within Madhya Pradesh, India. The Son River, originating from the Amarkantak Plateau, flows through Madhya Pradesh and Bihar before joining the Ganges River. The key tributaries within Madhya Pradesh include the Rihand, Kanhar, and Banas rivers. The Son Basin supports rich biodiversity with notable vegetation, including Sal, Teak, and mixed deciduous forests, which provide habitat for various wildlife species. Water availability in the basin is influenced by both seasonal rainfall and hydroelectric projects such as the Bansagar Dam. The region receives significant rainfall during the monsoon season, which helps recharge groundwater and sustain irrigation.

3.2.6 Tons Basin

The Tons basin, extending approximately between 23.98° to 26.06° North latitude and

80.18° to 83.30° East longitude, is a significant geographical and ecological region within Madhya Pradesh, India. The Tons River, also known as the Tamsa River, is a tributary of the Ganges River that flows through the Indian states of Madhya Pradesh and Uttar Pradesh, originates from Tamakund in the Kaimur Range of the Vindhyanchal Mountains in Satna, at an elevation of 610 meters. The river's tributaries include the Bihad, Babai, Sonkar, Belaj, Mahan, and Bichiya. The basin's topography is diverse, featuring rugged terrain, deep valleys, and forested hillsides, contributing to its scenic landscape. Rich in biodiversity, the Tons basin supports a variety of vegetation types, including mixed deciduous forests, pine forests, and alpine meadows, which provide habitat for numerous wildlife species. The region has variable groundwater potential, with areas of good, moderate, and poor groundwater availability identified through spatial data analysis (Nigam et al., 2020). The basin's water management practices focus on sustainable usage and conservation, ensuring that both surface and groundwater resources are adequately available for agricultural and domestic needs. There are also some industrial activities related to mining and small-scale manufacturing, though the area remains largely rural.

3.3 Rainfall and climate

The state has a tropical climate. The lowest temperature during the cooler months of December and January is 10°C, and in the summer months of May and June the temperature reaches 29°C. Most parts of the state in summer are hot and humid. Seasonal Average Rainfall of all the 6 sub-basin of ganga are shown in Figure 3.2. displays the Seasonal Average Rainfall for all six sub-basins of the Ganga river. The data indicates that the maximum seasonal average rainfall falls within the range of 1000 to 1300 mm. Conversely, the minimum seasonal average rainfall is recorded in the range of 500 to 700 mm.

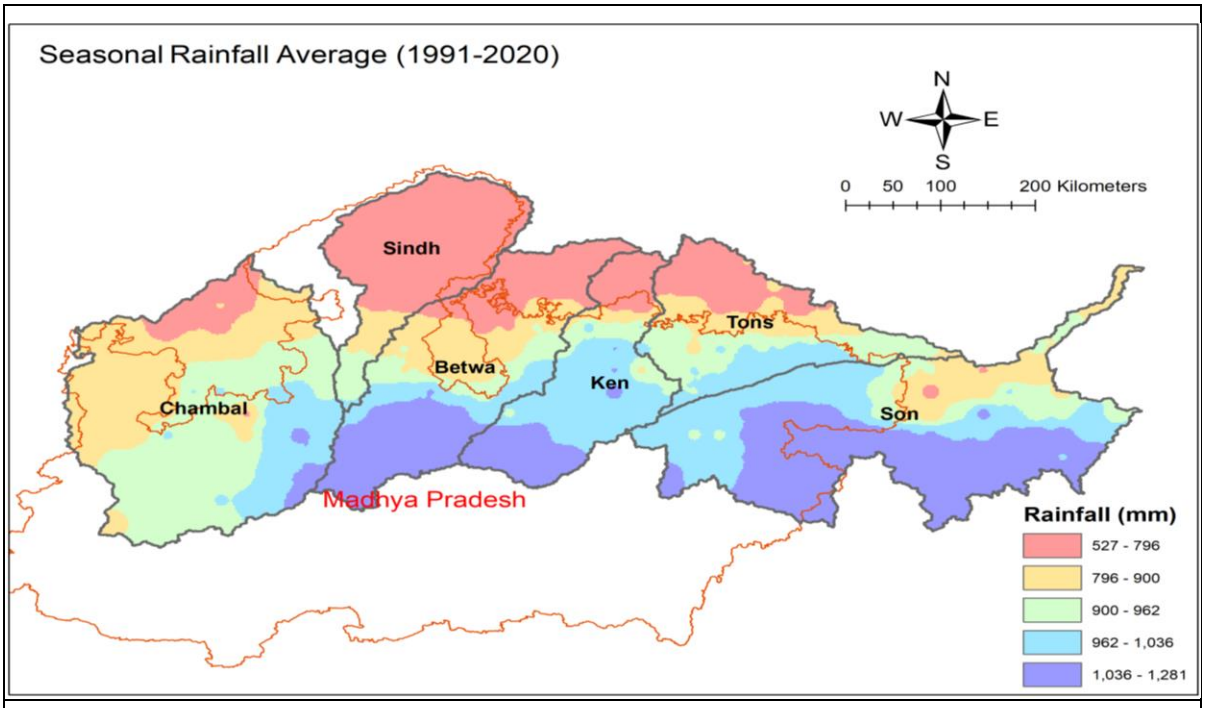


Figure 3.2 Seasonal Rainfall Average

The study found that during the winter months, the annual average minimum temperature tends to be relatively low, signaling the onset of colder periods. These lower temperatures may have implications for various aspects of the region's ecology, agriculture, and water resources. The lowest average minimum temperature is approximately 17 degrees Celsius and highest is 18 degrees Celsius. Understanding the annual average minimum temperature is essential for water resource management, particularly for predicting and mitigating the impacts of extreme weather events, such as frost and freezing temperatures. Colder temperatures during winter can influence snow accumulation in higher altitudes, affecting snowmelt patterns and streamflow in the subsequent months. Figure 3.3 shows the annual average minimum temperature (1991 -2020).

During the summer months, the annual average maximum temperature often reaches its peak, leading to increased evaporation rates and potential water stress. The higher temperatures can exacerbate water scarcity and impact agricultural productivity, especially if not adequately managed. Additionally, elevated temperatures can affect the region's hydrological cycle, altering precipitation patterns and runoff, which, in turn, may impact water availability and ecosystem dynamics. Figure 3.4 displays the annual average maximum temperature, which ranges from 31 to 32 degree Celsius.

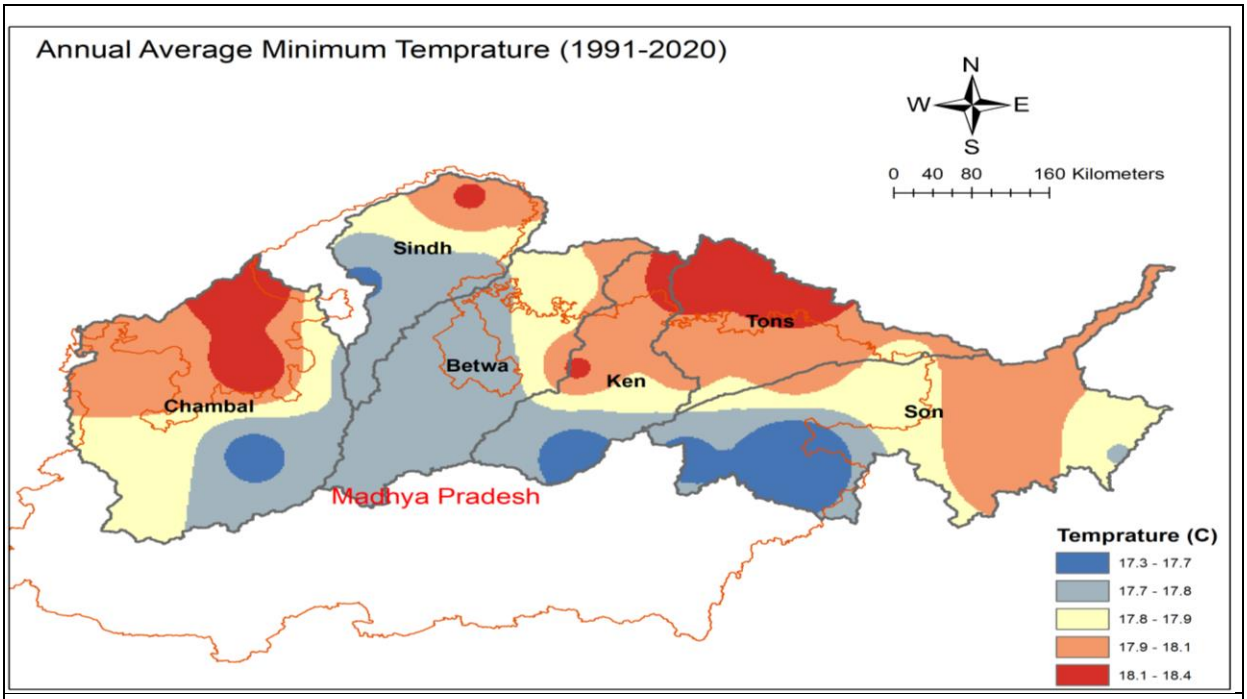


Figure 3. 3 Annual Average Minimum Temperature.

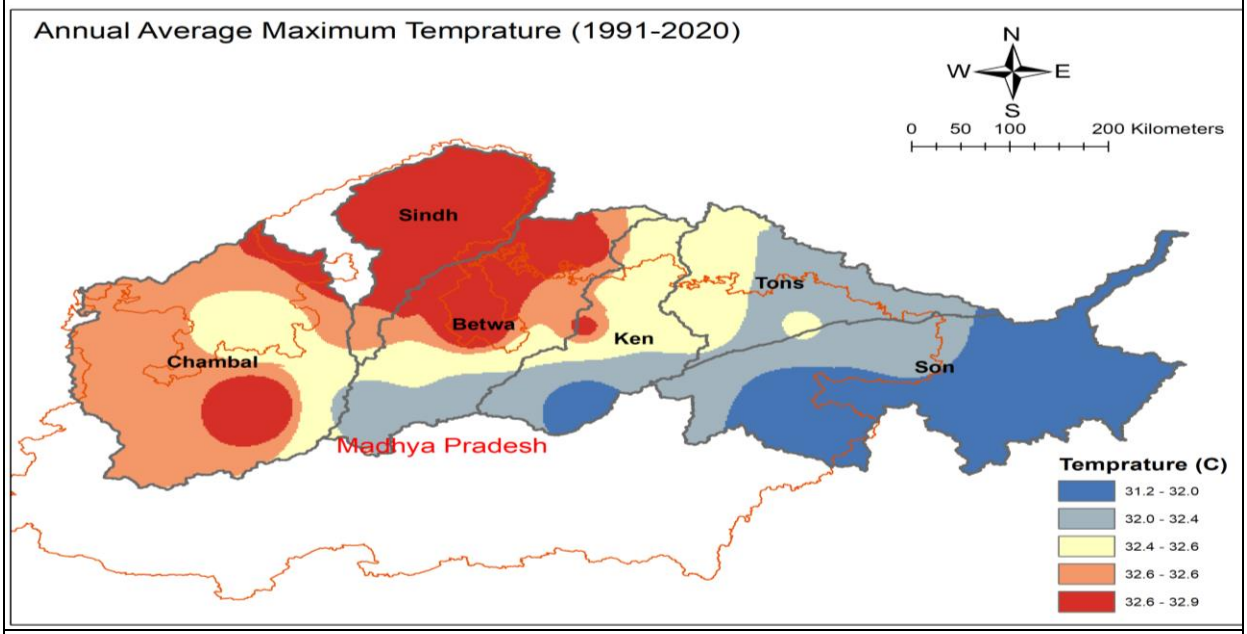


Figure 3. 4 Annual Average Maximum Temperature.

3.4 Geology

The geology of the Madhya Pradesh is influenced by the diverse topography and lithology of the region, which includes a combination of hard rocks, sedimentary rocks, and alluvial deposits.

The Betwa river's basin displays a geological composition of sedimentary rocks, including sandstone and shale from the Vindhyan Supergroup. The river's course has created

alluvial plains enriched by fertile sediment deposits. It has also shaped the landscape through erosion, forming ravines and valleys.

The Chambal basin in Central India exhibits a diverse geological profile. Comprising sedimentary deposits from the Vindhyan Supergroup and Gondwana rocks, it features alluvial plains shaped by the Chambal River and its tributaries. Notably, limestone deposits contribute to the area's mineral wealth. The basin's rich biodiversity supports endangered species like the Gharial and Gangetic Dolphin.

The Ken Basin is rich in mineral resources, particularly diamonds and sandstone. The Panna district, within the basin, is famous for its diamond mines, which have been a significant source of gemstones for centuries. Additionally, the sandstone from the region is widely used in construction. Topography of the Ken Basin is marked by undulating plains interspersed with rocky outcrops and hills. The terrain influences the flow of the Ken River and its tributaries, creating varied microclimates and affecting local hydrology. The river meanders through deep gorges and valleys, especially where it cuts through the Vindhyan ranges.

The Sindh River in central India showcases a geological landscape characterized by sedimentary deposits and alluvial plains. It is surrounded by rock formations from the Vindhyan Supergroup and features fertile alluvial plains formed by sediment deposition.

The Son River basin showcases a diverse geological composition, with sedimentary rocks dominating the region. It has contributed to the formation of fertile alluvial plains, making the area agriculturally productive. The Son river's flow has carved valleys and gorges, adding to the region's geological heritage. Additionally, the river's presence supports biodiversity and serves as a valuable water resource for human activities.

The Tons River in central India is associated with a geological profile comprising diverse sedimentary formations. The river has shaped the region by depositing alluvial soils in the surrounding plains. Its flow has contributed to the formation of valleys and gorges, adding to the area's geological diversity.

3.4.1 Soils

The soils in the sub- basin of Ganga in Madhya Pradesh are diverse and influenced by the geology and topography of the region. The soil types can be broadly classified into alluvial soils, black cotton soils, and red soils.

The alluvial soils Found in the river valleys and plains, alluvial soils are rich in nutrients

and fertility. They are composed of silt, clay, sand, and organic matter deposited by rivers like the Ganga, Yamuna, Narmada, and Tapti. These soils are excellent for agriculture and support a significant portion of the state's crop production.

The black cotton soils are Also known as Regur or "black cotton soil," these soils are widespread in Madhya Pradesh, covering large areas in the central region. They are formed from basaltic rocks and are rich in clay content, with high moisture-retaining capacity. Black soils are well-suited for growing cotton, oilseeds, and certain types of cereals. The red soils are common in the southern and western parts of the state, originating from weathered metamorphic and igneous rocks. Red soils are typically sandy or loamy and are suitable for cultivating millets, pulses, and oilseeds.

3.5 Gauge and Discharge (G/D) sites

G/D site the lack of runoff data at the point of interest is a massive concern for water resource engineers in India and Madhya Pradesh. Figure 3.5 shows the Different G/D site of Water Resource Department in Madhya Pradesh and Table 3.1 displays the Different G/D site of Water Resource Department in Madhya Pradesh and Figure 3.5 showing the Different Operational G/D site of Ganga sub-basin in Madhya Pradesh. Figure 3.6 showing the Different Operational G/D site of Ganga sub-basin in Madhya Pradesh.

Table 3. 1 Different G/D site of Water Resource Department in Madhya Pradesh

Basin	No. of GD Sites	Operational	Obsolete
Chambal	15	11	4
Betwa	16	10	6
Sindh	12	11	1
Ken	6	3	3
Son	12	3	9
Tons	4	4	0
Total	65	42	23

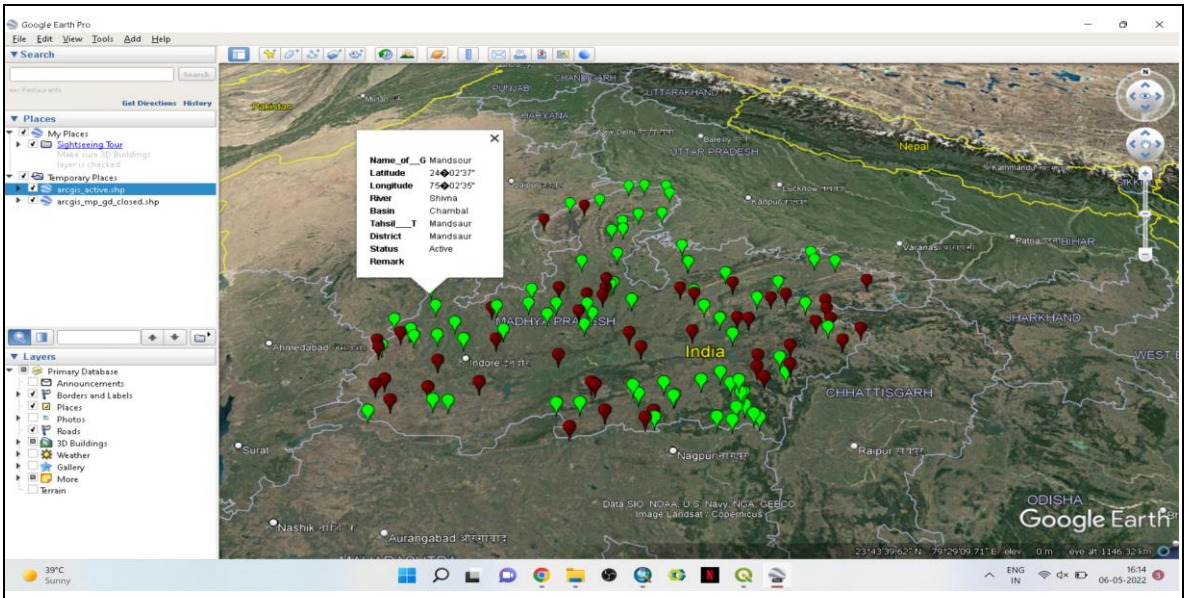


Figure 3. 5 Different G/D site of WRD in Madhya Pradesh.

Figure 3. 6 Different Operational G/D Site of Ganga Sub-Basin in Madhya Pradesh.

The Table 3.2 lists various geographical locations of gauge/discharge (G/D) sites within different river basins in India used in this study, including Betwa, Chambal, Ken, Sindh, Son, and Tons. For each G/D site, the table specifies the name of the site, the river basin it belongs to, and its latitude and longitude coordinates.

In the Betwa basin, there are sites such as Bah, Betwa, Mohanpur, Neemkheda, and Rahatgarh. The Chambal basin includes G/D sites like Sinduriya, Manakhedi, Maksudangar, Mandsour, Badnagar, Kuno, Samaskhedi, Choma, and Sanwar. In the Ken basin, the sites listed are Sonar, Pandwan, Kopra, Niwari, Nohta, and Tigra. The Sindh basin has sites including Daboh, Didi, Amola, Goraghat, and Budhara. The Son basin includes Marsiabad, Amarpur, Bhadora, and Ghatkhirwa. Finally, the Tons basin encompasses sites such as Sansarpur, Patehera, Khatkhari, and Maihar. Each site's specific latitude and longitude coordinates are provided to pinpoint their exact location within their respective basins.

Table 3. 2 lists various geographical locations of gauge/discharge (G/D) sites.

G/D site	Basin	Area (km^2)	Latitude	Longitude
Bah	Betwa	938.58	23° 42' 22.54"	77° 40' 57"
Betwa	Betwa	2666.74	23° 30' 32"	77° 47' 13"
Mohanpur	Betwa	1409.05	24° 50' 11"	78° 4' 36"
Neemkheda	Betwa	1966.12	23° 16' 10"	77° 40' 55"
Rahatgarh	Betwa	1166.45	23° 47' 57"	78° 23' 46"
Sinduriya	Chambal	1402.21	23° 39' 38"	75° 12' 56"
Manakhedi	Chambal	3515.01	23° 42' 36"	77° 11' 3"
Maksudangar	Chambal	4853.72	24° 2' 40"	77° 10' 49"
Mandsour	Chambal	967.55	24° 4' 6"	75° 1' 51"
Badnagar	Chambal	870.01	23° 3' 28"	75° 23' 14"
Kuno	Chambal	4476.61	23° 7' 12"	77° 5' 34"
Samaskhedi	Chambal	1849.09	23° 11' 30"	76° 26' 2"
Choma	Chambal	1371.93	23° 38' 57"	76° 13' 22"
Sanwar	Chambal	410.37	22° 58' 50"	75° 49' 48"
Sonar	Ken	3875.50	24° 0' 26"	79° 24' 4"
Pandwan	Ken	16785.37	24° 24' 33"	79° 24' 26"
Kopra	Ken	867.40	23° 56' 31.88"	79° 24' 26.24"
Niwari	Ken	1305.79	24° 3' 15.01"	79° 10' 32.23"
Nohta	Ken	3422.74	23° 41' 7.01"	79° 33' 59"
Tigra	Ken	4188.80	24° 17' 55"	79° 57' 56"
Daboh	Sindh	1903.60	28° 52' 0.84"	78° 47' 50"
Didi	Sindh	6266.27	26° 37' 5.88"	78° 50' 49"
amola	Sindh	4938.85	25° 24' 35"	77° 55' 41"
Goraghat	Sindh	11880.51	25° 50' 6"	78° 22' 15"
Budhara	Sindh	5300.22	26° 38' 2"	78° 26' 2"
Marsiabad	Son	10623.64	23° 45' 51"	81° 11' 57"
Amarpur	Son	395.20	23° 58'	80° 58'
Bhadora	Son	3139.26	23° 52' 7"	80° 39'
Ghatkhirwa	Son	1012.76	23° 52' 15"	80° 25' 18"
Sansarpur	Tons	720.93	24° 44' 52"	81° 27' 3.6"

Patehera	Tons	8066.44	24° 55' 30"	81° 24' 29"
Khatkhari	Tons	527.38	24° 43' 38"	81° 48' 20"
Maihar	Tons	506.87	24° 15' 42"	80° 48' 22"

Figure 3.7 shows the map of the Betwa basin, showing specific G/D sites used in the analysis. The map highlights the Betwa basin with a red boundary line and marks several locations within this basin. The locations include Mohanpur, Bah, Betwa, Neemkheda, and Rahatgarh, each shaded in different colors to distinguish them.

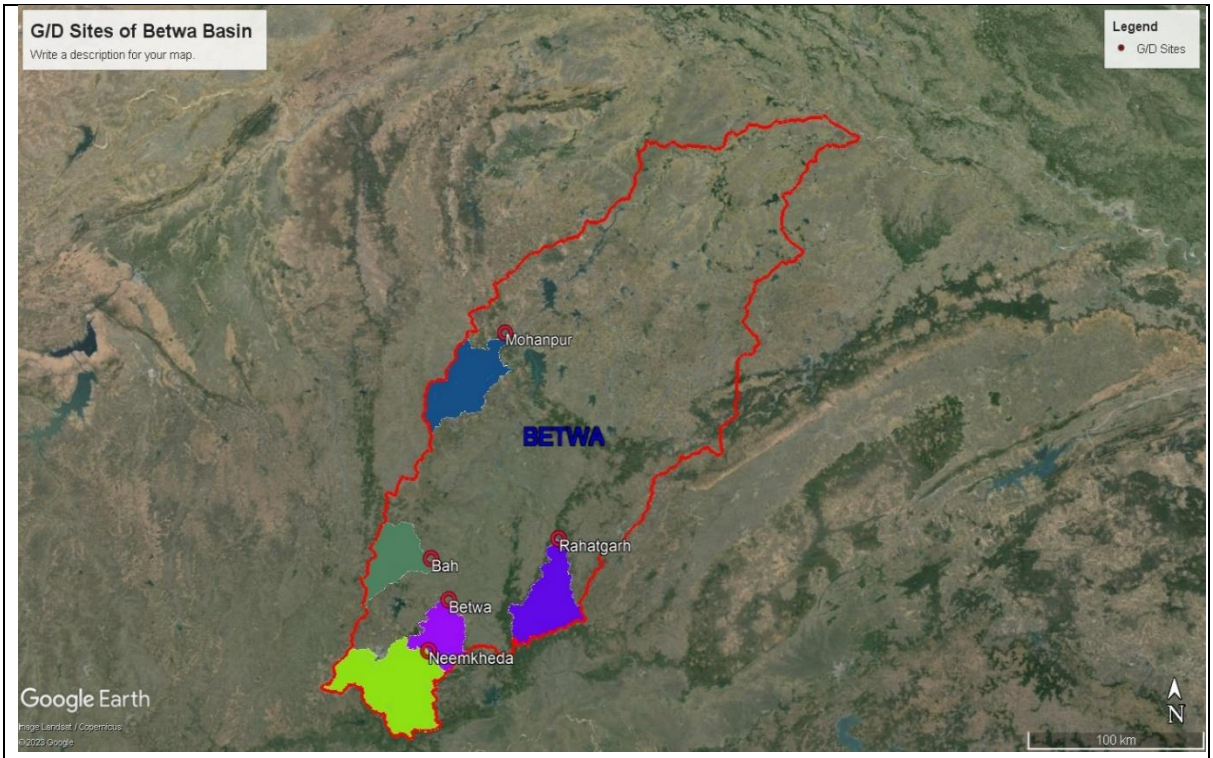


Figure 3. 7 Different Operational G/D site within the Betwa Basin in Madhya Pradesh.

Figure 3.8 presents the Chambal basin, highlighting G/D (geographical and demographic) sites and their respective catchment areas used in the analysis. The map outlines the Chambal basin with a distinct black boundary and marks several key locations, including Kuno, Mandsour, Sinduriya, Badnagar, Sanwar, Choma, Samaskhedi, Maksudangar, and Manakhedi, each represented in different colors for clarity.

The Chambal basin encompasses diverse geographical and demographic sites, each contributing uniquely to the region's landscape. Kuno, situated in the northern part of the basin, is renowned for the Kuno Wildlife Sanctuary and Kuno National Park, both significant for biodiversity conservation and recognized as a proposed site for the reintroduction of Asiatic

lions.

Mandsour, rich in history, is known for its ancient temples and archaeological heritage. Economically, it plays a key role in opium cultivation and slate pencil production. Nearby Sinduriya, primarily an agricultural hub, thrives on the basin's fertile lands and shares cultural and economic similarities with Mandsour.

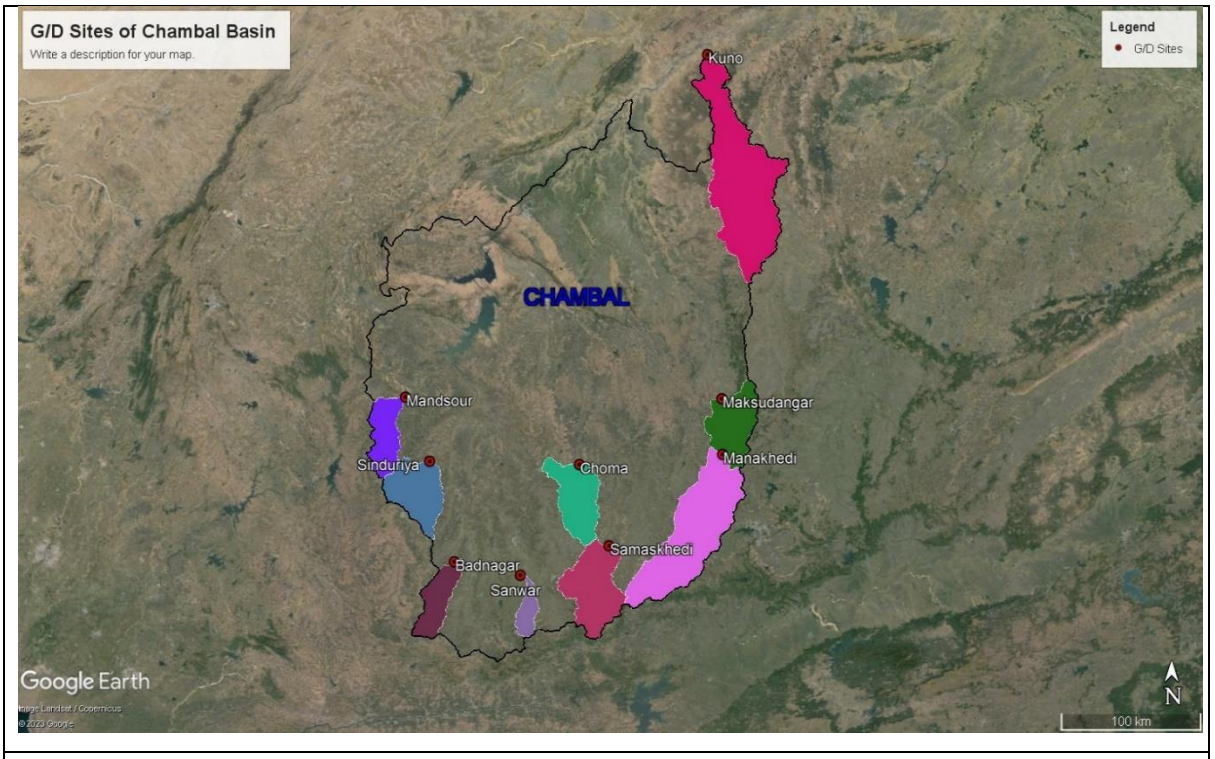


Figure 3. 8 Different Operational G/D site within the Chambal Basin in Madhya Pradesh.

Badnagar serves as a key agricultural hub, primarily cultivating soybeans, wheat, and mustard. It likely supports small-scale industries related to agriculture, such as oil extraction units. Sanwar, a predominantly rural area, is deeply rooted in traditional farming practices, focusing on crops well-suited to the region's climate and soil conditions. Similarly, Choma emphasizes agriculture and may feature small water bodies or forested areas that enhance local biodiversity.

Samaskhedi integrates agriculture with small-scale industrial activities, balancing crop production and livestock rearing. The Maksudangarh G/D site catchment area is distinguished by its sustainable farming practices and vibrant local markets, reflecting traditional agricultural methods. Manakhedi, heavily dependent on agriculture, focuses on cultivating both staple and cash crops while maintaining rich cultural and social traditions tied to its agrarian lifestyle.

Figure 3.9 illustrates the Ken basin, clearly demarcated with a distinct boundary and highlighting key geographical and demographic (G/D) sites. These include Sonar, Pandwan, Kopra, Niwari, Nohta, and Tigra, each marked with specific names and color-coded regions, showcasing their unique contributions to the region's landscape.

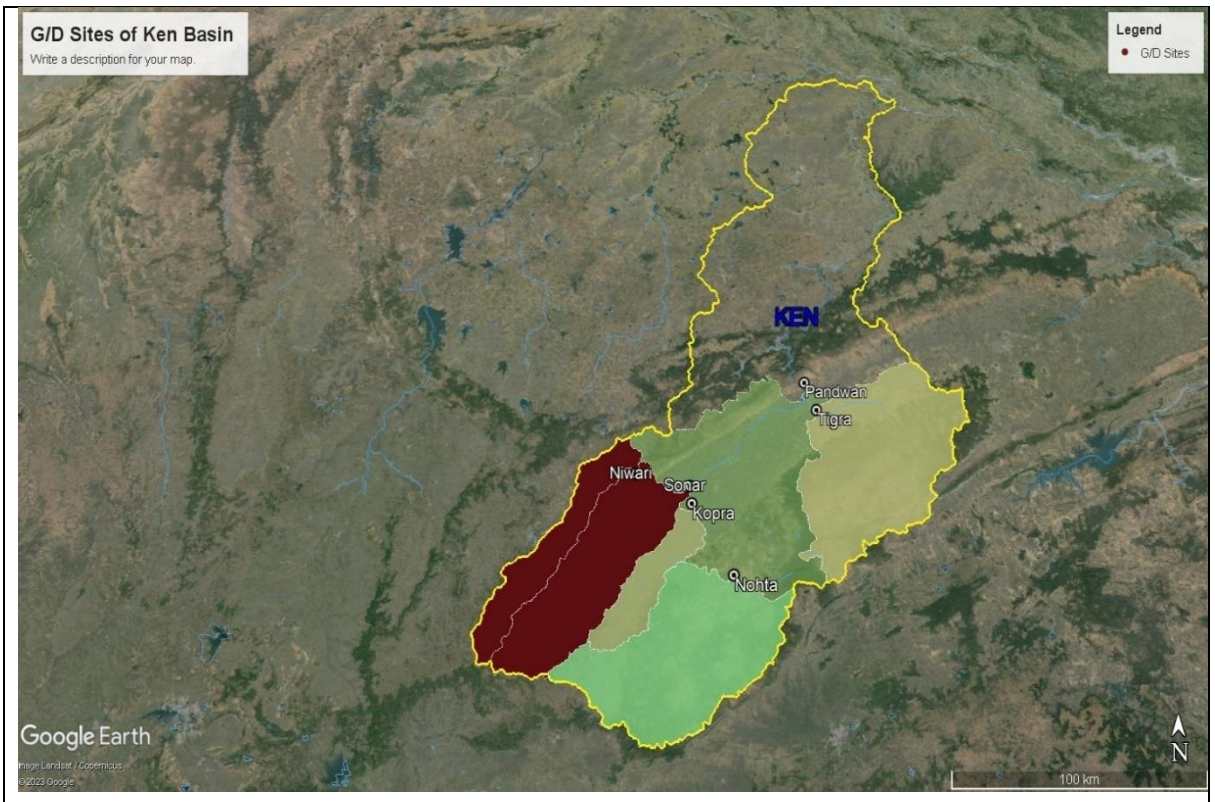


Figure 3. 9 Different Operational G/D site within the Ken Basin in Madhya Pradesh.

Sonar is likely an agricultural hub, supported by small water bodies or irrigation systems, and may also hold historical or cultural significance tied to the region's heritage. Pandwan region focuses on traditional farming practices and serves as an economic center for surrounding rural areas through its local markets. The area near the Kopra G/D site is characterized by its extensive crop production, particularly wheat, pulses, and oilseeds, and possibly features small-scale agro-industries within a robust rural community.

The catchment area surrounding the Niwari G/D site is recognized for its agricultural productivity, with a variety of crops cultivated across its fertile lands. The region may also hold historical significance, featuring ancient structures that add to its cultural heritage. The areas near the Nohta G/D site are similarly rich in agricultural activities and may gain prominence due to their proximity to natural resources or significant historical landmarks, further enhancing

their regional importance. Finally, Tigra showcases a blend of traditional and modern agricultural practices, supporting its rural population through sustainable farming, local markets, and a vibrant community.

Figure 3.10 is a detailed representation of the Sindh basin, highlighting various G/D (gauging/discharge) sites. The basin is outlined in bold blue, with different areas shaded in red, green, and light blue to denote distinct sub-regions. Five specific G/D sites are marked with red dots and labelled as Budhara, Didi, Goghat, Amola, and Daboh.

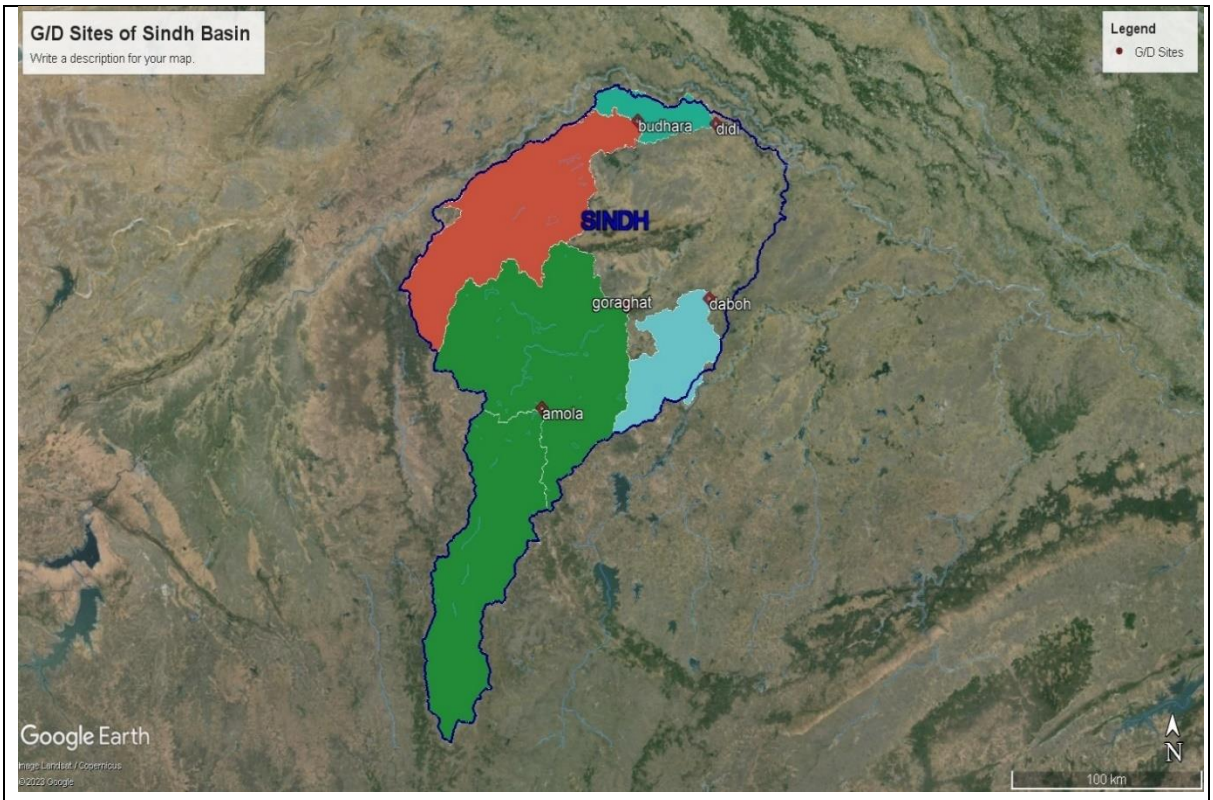


Figure 3. 10 Different Operational G/D site within the Sindh Basin in Madhya Pradesh.

The site Budhara is located in the northern part of the basin, within the red-shaded area. This site plays a crucial role in monitoring water flow and discharge in the upper reaches of the Sindh basin. Moving eastward, Didi is situated within the same red zone and serves a similar purpose in assessing water resources and hydrological patterns.

Further south, in the central green-shaded region, Goraghat is another key G/D site, providing important data on water flow and discharge for the mid-basin area. Continuing southwards, Amola is positioned within the same green region and contributes to the overall understanding of water dynamics in this part of the basin.

Lastly, Daboh is in the southern part of the basin, within the light blue-shaded area. This site is essential for monitoring the lower basin's water discharge and ensuring effective water management across the entire Sindh basin. Each of these sites is integral to maintaining an accurate and comprehensive hydrological profile of the region, facilitating efficient water resource management and planning.

The Figure 3.11 illustrates the Son basin, highlighting various G/D (gauging/discharge) sites within the region. The basin is outlined in bold black within the Madhya Pradesh, and different areas within the basin are shaded in green and pink to represent distinct sub-regions. Five specific G/D sites are marked with red dots and labelled as Ghathkhiwa, Bhadara, Amarpur, Marsiaabad, and one additional site.

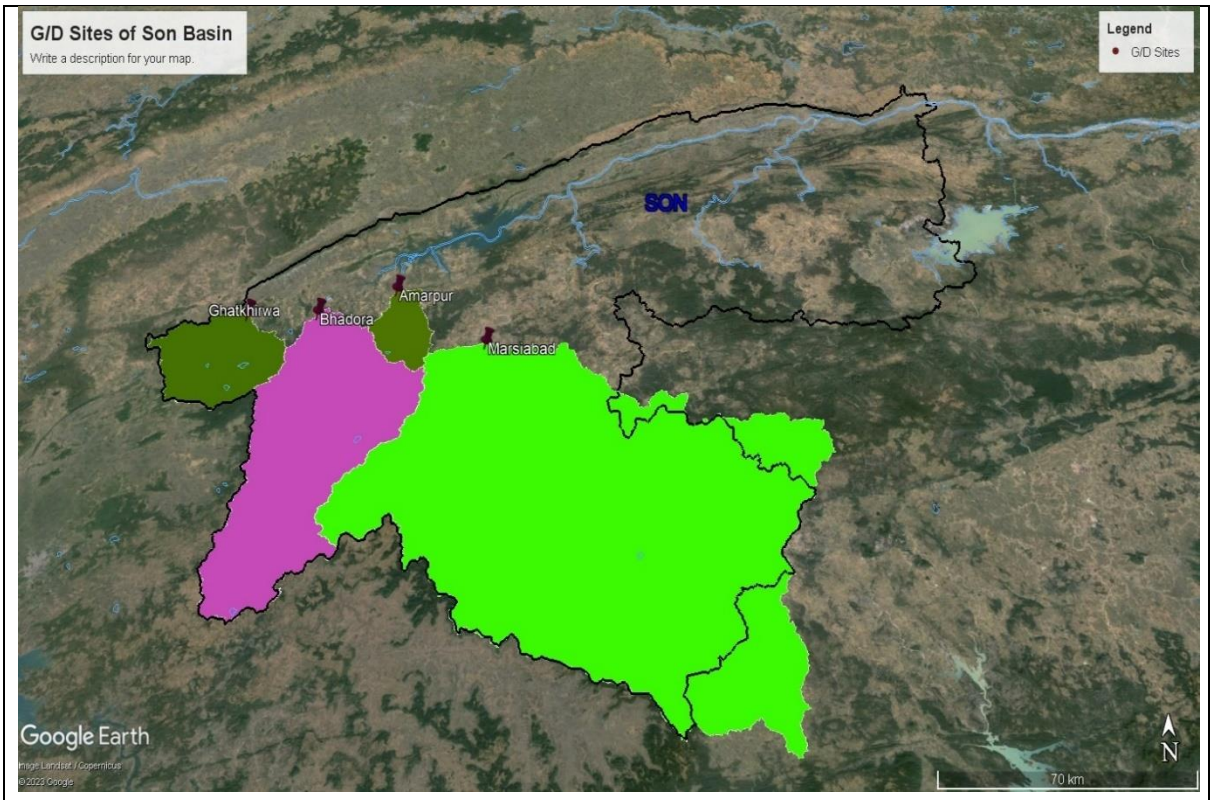


Figure 3. 11 Different Operational G/D site within the Son Basin in Madhya Pradesh.

The site Ghathkhiwa G/D site is in the northwestern part of the basin, within the green-shaded area. This site is crucial for monitoring water flow and discharge in the upper reaches of the Son basin. Just east of it, Bhadara G/D is situated and serves a similar purpose in assessing water resources and hydrological patterns in the basin.

Further east, in the green-shaded region, Amarpur is another key G/D site. It provides

essential data on water flow and discharge for the central part of the Son basin. Moving southward, Marsiaabad G/D site is located within a green-shaded area. This site contributes to the overall understanding of water dynamics in the lower basin area.

Each of these sites is integral to maintaining an accurate and comprehensive hydrological profile of the Son basin, facilitating efficient water resource management and planning across the region. The marked sites help in assessing the basin's water availability, planning for irrigation, flood management, and sustaining ecological balance.

Figure 3.12 illustrates the Tons basin, highlighting various G/D (gauging/discharge) sites within the region. The basin is outlined in bold red, with different areas within the basin shaded in green, blue, red, and pink to denote distinct sub-regions. Several specific G/D sites are marked with marks and labelled as Patehera, Sansarpur, Khatkhari, and Maihar G/D sites.

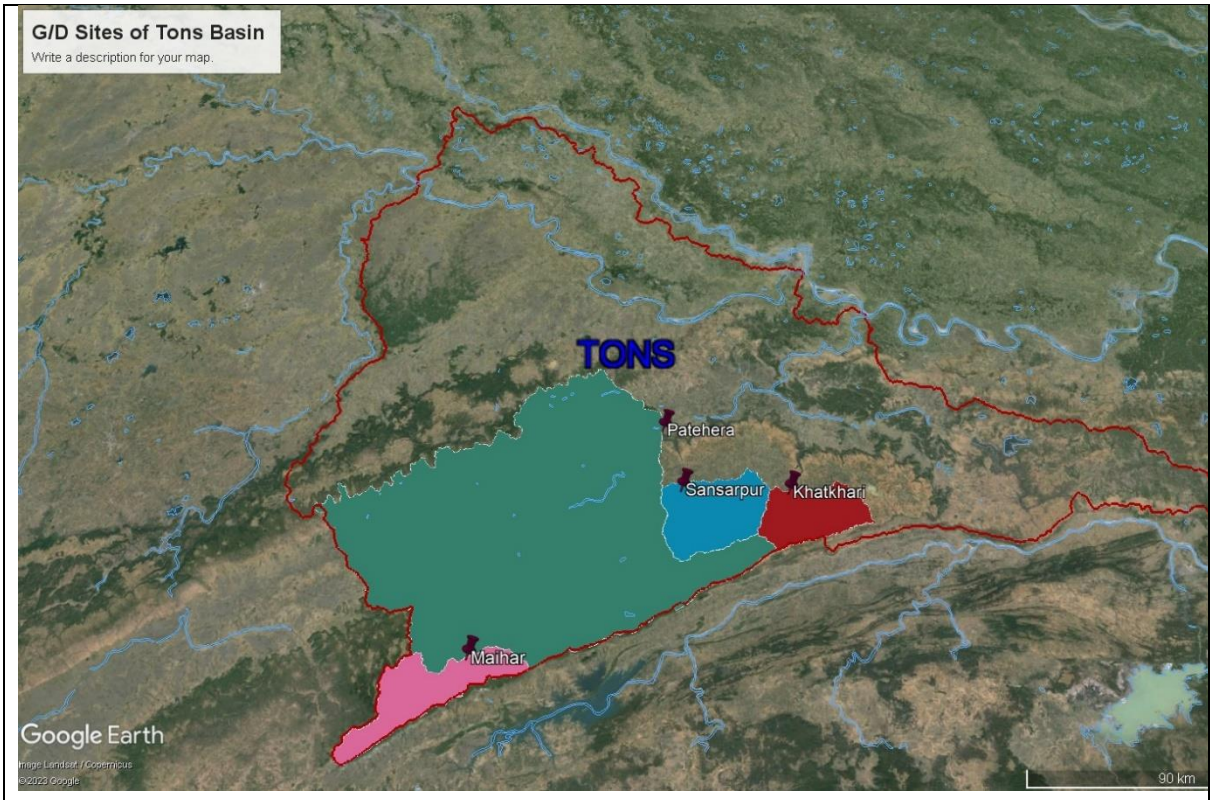


Figure 3. 12 Different Operational G/D site within the Tons Basin in Madhya Pradesh.

The site Patehera is in the northeastern part of the Tons basin, within the dark green-shaded area. This site is essential for monitoring water flow and discharge in the upper reaches of the Tons basin. Moving southward, Sansarpur is situated in a blue-shaded area and serves a similar purpose in assessing water resources and hydrological patterns in this section of the

basin.

Further to the south, in the red-shaded region, Khatkhari is another key G/D site. It provides crucial data on water flow and discharge for the central part of the Tons basin. Lastly, Maihar is in the southwestern part of the basin, within the pink-shaded area. This site contributes to the overall understanding of water dynamics in the lower basin area.

Each of these sites plays a vital role in maintaining an accurate and comprehensive hydrological profile of the Tons basin, facilitating efficient water resource management and planning across the region. The marked sites help in assessing the basin's water availability, planning for irrigation, flood management, and sustaining ecological balance.

3.6 Data Collected/Used

The data collection process for this water availability study involved gathering essential hydrological and meteorological data from various sources. Historical rainfall and temperature data was obtained from IMD (Indian Metrological Department) and global satellite datasets, providing comprehensive coverage of precipitation patterns in Madhya Pradesh. Streamflow or runoff data was collected from stream gauges strategically placed within the study area, allowing for the observation of actual water flow in different basins. Table 3.3 shows the List of collected and created GIS Data.

Other various data sources such as SRTM DEM, Road Layer, Waterbodies Layer, and Population Grid were downloaded from OSM and DIVA GIS, respectively, to create an Elevation Map, Slope Map, Distance from Roads Map, Distance from Water Bodies Map.

Table 3. 3 List of collected and created GIS Data

Data collection	
Data collected	Period
1. Daily Discharge data	
a. Chambal	Jan 1990 – December 2019
b. Sindh	Jan 1990 – December 2019
c. Betwa	Jan 1990 – December 2019
d. Ken	Jan 1990 – December 2019
e. Son	Jan 1990 – December 2019
f. Tons	Jan 1990 – December 2019

GIS data creation	
GIS data	**Description**
1. DEM of catchment and command for slope, elevation map	SRTM
2. Road Layer for Distance from road map	OSM
3. Population Grid for Distance from population map	Diva GIS
4. Waterbodies and drainage layer Distance from Waterbodies map	OSM and Hydroshed
5. Catchment Area Maps	Global Hydroshed
6. Sentinal-2 10m image for LULC preparation	Year-2020

CHAPTER: 4
METHODOLOGY

4.1 Methodology Adopted for Estimation of Water Availability

Water availability is one of the critical factors for sustainable water resource management, especially in regions where water scarcity impacts agriculture, industry, and daily life. In India, and, particularly in Madhya Pradesh, the absence of long-term runoff data at prospective reservoir sites poses a significant challenge for water resource engineers. While transferring data from neighboring locations is a common solution, this method is susceptible to planner bias regarding donor catchment selection and the appropriate length of record. This study aims to mitigate these concerns by developing rainfall-runoff equations tailored for different homogeneous regions in Madhya Pradesh. The goal is to create a robust model for evaluating water availability at any location within medium-sized catchments. To achieve this objective, the study will first employ statistical analysis to identify different homogeneous zones within the selected Ganga basins in Madhya Pradesh. This involves examining long-term runoff records, spanning almost 30 years, to determine patterns and correlations. Following the identification of these zones, rainfall-runoff models will be developed for gauged catchments using a combination of linear regression, the Soil Conservation Service (SCS) Curve Number method, and soft computing techniques. These models aim to accurately represent the relationship between rainfall and runoff in the identified zones.

In the next step, the study will establish relationships between model parameters and readily available topographical or geomorphological properties. This includes factors such as slope, soil type, land use, and other relevant physical characteristics that influence runoff. By understanding these relationships, the study aims to generalize the developed models for broader application across different catchments. The validation process involves applying the equations derived from gauged or donor catchments to ungauged catchments. This step is crucial to ensure that the models are robust and reliable when used in areas without direct runoff data. The final stage of the study will focus on estimating water availability from ungauged catchments, using the validated models. Once the water availability models are developed, a web-based application will be made accessible to field engineers. This tool will enable engineers to determine water availability in any medium basin in Madhya Pradesh efficiently. By providing a reliable and user-friendly method for estimating water availability, the study aims to support better planning and management of water resources in the region. This

approach will not only help in mitigating the issues associated with data scarcity but also improve the accuracy and reliability of water resource assessments in Madhya Pradesh.

4.2 SCS-CN Model

The SCS-CN method is a model, developed by the United States Department of Agriculture's Natural Resources Conservation Service is a widely utilized hydrological method for estimating direct runoff from rainfall events. This model is appreciated for its simplicity and effectiveness, making it a cornerstone in hydrological studies and water resource management. It incorporates three primary components: initial abstraction (Ia), potential maximum retention (S), and actual evapotranspiration (AE). Initial abstraction represents the portion of rainfall that does not contribute to direct runoff, including interception by vegetation, infiltration into the soil, and surface depression storage. Typically, initial abstraction is assumed to be a fixed percentage of the potential maximum retention (S), commonly set at 20%. Figure 4.1 shows the Flow chart for the computation of discharge through the Google Earth engine based on the SCS-CN model. Normally the SCS model computes direct runoff with the help of the following relationship (Handbook of Hydrology, 1972)

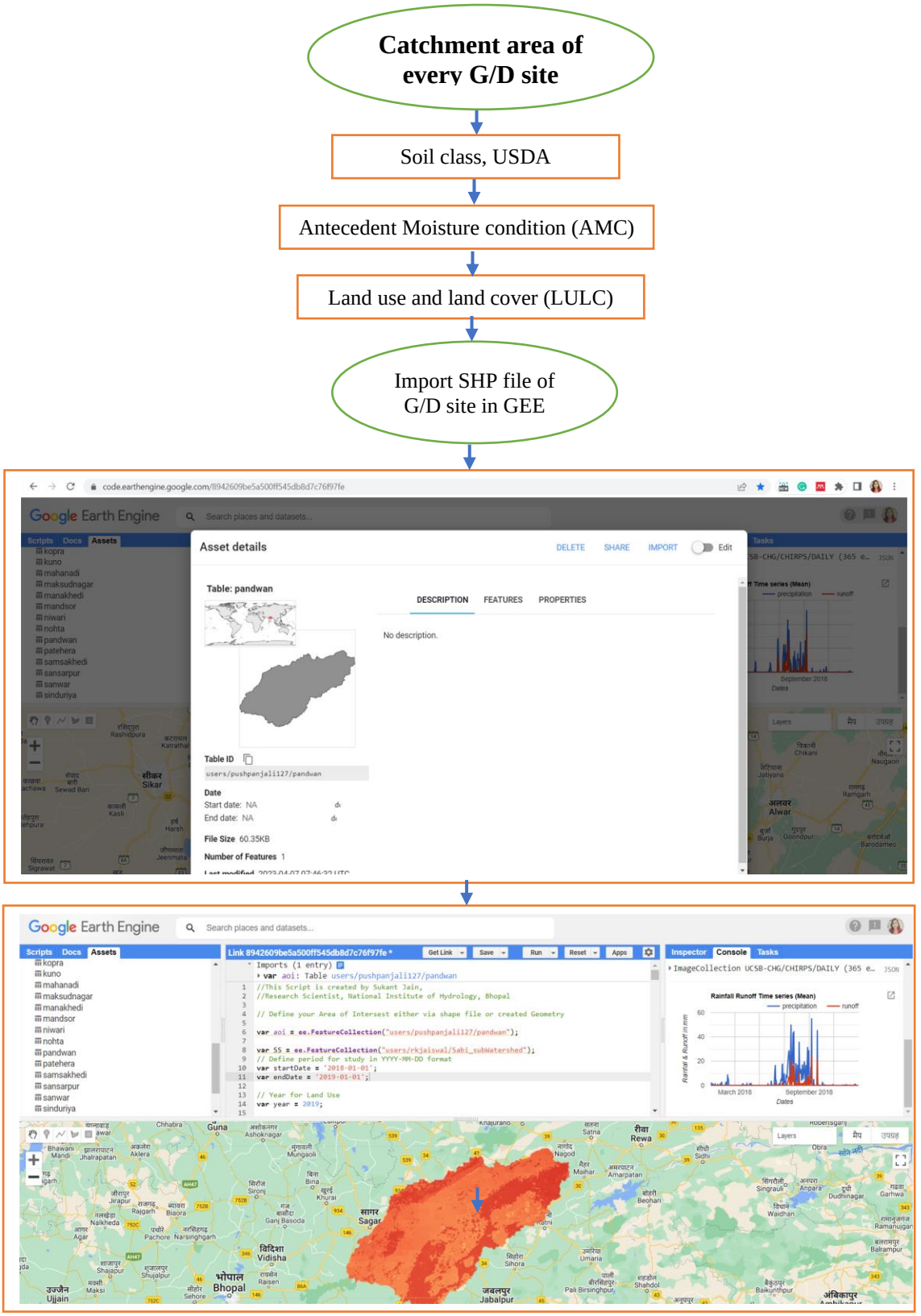


Figure 4. 1 Flow chart for computation of Discharge through Google Earth engine Based on SCS-CN model.

Potential maximum retention (S) reflects the soil's ability to retain water and is

influenced by soil type, land use, and the watershed's hydrological condition. The value of S is inversely related to the Curve Number (CN), which ranges from 0 to 100, with lower CN values indicating higher potential retention and less runoff. The relationship between S and CN is established through a specific formula, highlighting the inverse nature of their relationship. Actual evapotranspiration (AE), which represents water loss due to evaporation and plant transpiration, is implicitly accounted for in the initial abstraction and potential maximum retention components, even though it is not explicitly calculated in the SCS-CN method.

The SCS-CN method estimates direct runoff by considering the total rainfall and the initial abstraction and then applying a relationship involving potential maximum retention. This relationship ensures that runoff is only calculated when the total rainfall exceeds the initial abstraction, thus accounting for the soil's retention capacity before generating runoff. For example, in a watershed with a specific Curve Number, the potential maximum retention can be calculated, and combined with observed rainfall data, the method predicts the direct runoff generated from a rainfall event.

To enhance the computation of discharge using the SCS-CN method, the Google Earth Engine (GEE) platform is utilized. GEE is a cloud-based platform that facilitates large-scale environmental data analysis and visualization. By integrating the SCS-CN model with GEE, runoff estimation can be automated across extensive watersheds. This process entails gathering essential data, including rainfall, land use, soil type, and topography, from the extensive dataset repository provided by GEE. Assigning appropriate CN values based on land use and soil type data follows, and rainfall data for the period of interest is then inputted, derived from satellite data or local weather stations. Finally, the SCS-CN relationship is applied to compute direct runoff, and the results are visualized and analyzed using GEE's powerful tools. By leveraging GEE, the SCS-CN model can be efficiently applied to estimate runoff, supporting better planning and management of water resources in various watersheds. Equation 1, Equation 2, and Equation 3 describe the potential maximum retention after runoff begins, runoff, and curve number respectively.

$$S = \frac{2500}{CN} - 254 \quad (1)$$

$$Q = \frac{(P - I_a)^2}{P - I_a + S} \quad (2)$$

$$CN_{weighted} = \frac{\sum(CN_i \times A_i)}{\sum(A_i)} \quad (3)$$

Where,

$CN_{weighted}$ = weighted curve number.

Q = Runoff (in mm)

P = Total rainfall (in mm)

I_a = Initial abstraction (losses before runoff begins, including surface storage, interception, and infiltration).

S = Potential maximum retention after runoff begins, which is related to the curve number.

CN_i = Curve number from 1 to any no. N.

A_i = Area with curve number CNi

A = the total area of the watershed.

The *CN* is a function of soil type, land cover, and antecedent moisture condition (AMC); Q, actual direct runoff, mm; P, total storm rainfall, mm; and S, the potential maximum retention of water by the soil, mm.

The runoff curve number method is the procedure for hydrological abstraction developed by the USDA Soil Conservation Service. The runoff curve number method was developed based on 24-hour rainfall-runoff data.

$$Q = \frac{R(CN\left(\frac{P}{R}+2\right)-200)^2}{CN(CN\left(\frac{P}{R}-8\right)+800)} \quad (4)$$

Where,

R = 2.54,

P = rainfall (mm), Q runoff (cm),

In this study, a program was developed in the Google Earth Engine (GEE) platform for the computation of runoff through the SCS-CN model.

The SCS-CN model is a valuable tool for estimating runoff and understanding rainfall-runoff relationships in watersheds. Its simplicity and applicability have made it a popular choice for various hydrological applications. For the study, several models including NAM, SWAT, NAPI, GR4J, GR2M, SCS-CN, and RRL models were applied on different basins with observed data. No single model was found suitable for all basins because of various regions.

SCS-CN model developed in GEE provides satisfactory results on a monthly scale.

4.3 GR2M Hydrological Model

Hydrological models are indispensable tools for understanding and predicting the intricate processes that govern the movement of water within a catchment area. These models facilitate the simulation of hydrological phenomena such as rainfall-runoff processes, groundwater flow, and evapotranspiration, providing essential insights for water resource management, flood forecasting, and ecological conservation. Among the various hydrological models, the GR2M (Global Rainfall-Runoff Model with 2 parameters) stands out due to its simplicity and effectiveness in simulating rainfall-runoff processes. The GR2M model was developed in the late 1980s by researchers at CEMAGREF, an institution in France that has since been rebranded as INRAE (National Research Institute for Agriculture, Food and the Environment). This model emerged from the conceptual framework established by the GR4J model, another well-known hydrological model. The GR2M model was designed to offer a more straightforward approach while retaining the essential functionalities needed for accurate rainfall-runoff simulations. Fig 4.2 shows the working methodology of the Gr2M hydrological model.

The GR2M hydrological model operates on a monthly time step and utilizes two parameters to describe the rainfall-runoff relationship within a catchment. These parameters are designed to capture the primary hydrological processes that influence runoff generation. The model incorporates a conceptual representation of soil moisture and routing components, which are critical for understanding how precipitation transforms into streamflow.

One of the key strengths of the GR2M model is its simplicity. With only two parameters, the model is relatively easy to calibrate and apply across different catchments, making it a versatile tool for hydrological studies. Despite its simplicity, the GR2M model is capable of capturing the essential dynamics of the hydrological cycle, providing reliable estimates of runoff under varying climatic and land-use conditions.

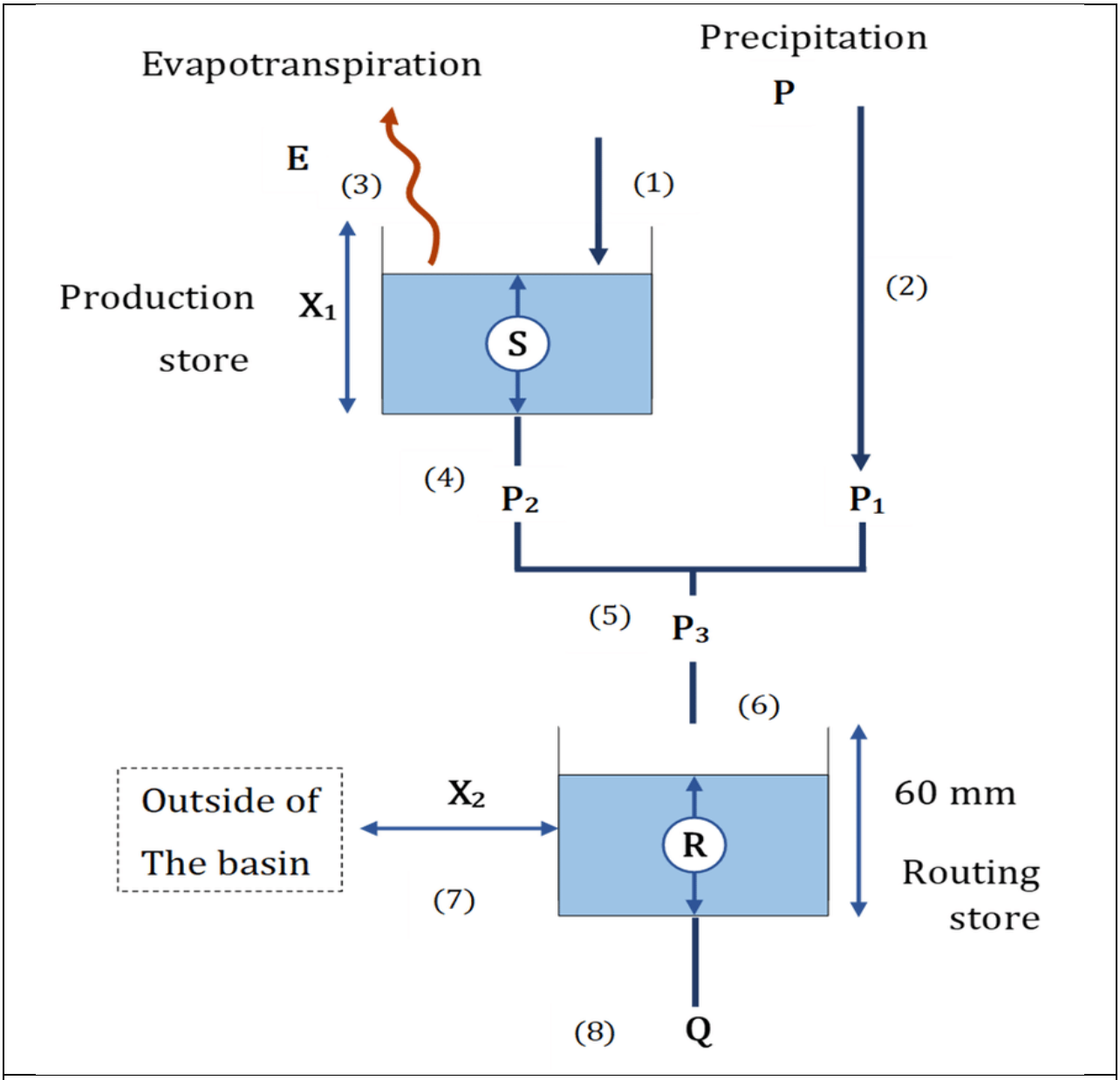


Figure 4. 3 Working methodology of GR2M hydrological model.

The GR2M model structure consists of mathematical equations that describe the relationships between inputs (e.g., precipitation), internal states (e.g., storage capacities), and outputs (e.g., streamflow). Calibration of the model involves estimating a set of parameters that govern the behavior of the catchment. These parameters include storage capacities, recession coefficients, and infiltration parameters, among others. The model is calibrated using observed streamflow data, and the goodness-of-fit between simulated and observed streamflow is evaluated using statistical metrics such as Nash-Sutcliffe Efficiency (NSE) and Root Mean Square Error (RMSE). Once calibrated and validated, the GR2M model can be applied to simulate streamflow under various hydrological scenarios, including historical analysis, flood

forecasting, and climate change impact assessment.

The GR2M hydrological model is a monthly water balance model that simulates the hydrological processes of a catchment using production and routing.

Production

In the production storage, the initial soil moisture is considered as *S*, and it is reduced constantly due to evaporation (*E*) and may become S_1. The S_1 depends on the coefficient of storage (X_1) and can be computed using the following equation (6) to (15)

$$S_1 = \frac{S + X_1 * \varphi}{1 + \varphi \frac{S}{X_1}} \qquad \text{with } \varphi = tanh \frac{P}{X_1} \tag{6}$$

The outflow parts from the production storage may be P_1 and P_2. The P_1 is the excess of water after meeting the soil moisture and can be written as:

$$P_1 = P + S_1 - S \tag{7}$$

In the process of evaporation and change of soil moisture, the soil moisture changes to S_2 and can be presented by the following equation:

$$S_2 = \frac{S_1 * (1 - \varphi)}{1 + \varphi \frac{S_1}{X_1}} \tag{8}$$

The P_2 part of rainfall from production storage can be defined by the following equation represents the excess of soil moisture.

$$P_2 = S_2 - S \tag{9}$$

The soil moisture for the next step (S) can be expressed using the following equation.

$$S = \frac{S_2}{\left[1 + \left(\frac{S_2}{X_1}\right)^3\right]^{\frac{1}{3}}} \tag{10}$$

The combined parts of the precipitation from overland (P_1) and soil moisture change (P_2) reach the routing storage and can be written as below.

$$P_3 = P_1 + P_2 \tag{11}$$

Routing

The initial storage in routing storage is *R*, changed to R_1 and represented by the following equation:

$$R_1 = R + P_3 \tag{12}$$

The R_1 due to the effect of routing may change to R_2 and can be represented by the

following equation:

$$R_2 = X_2 * R_1 \quad (13)$$

The resultant storage after the end of the monthly step will become *R*, and outflow from the basin *Q* can be given by the following equations:

$$Q = \frac{{R_2}^2}{R_2+60} \quad (14)$$

$$R = R_1 - Q \quad (15)$$

The model requires inputs of total precipitation, potential evapotranspiration, and the maximum capacity of the production. The model outputs the net rainfall, percolation, and total runoff at each time step. The GR2M hydrological model has two parameters: X1 and X2. X1 represents the maximum capacity of the production reservoir, while X2 represents the underground exchange coefficient. These parameters are used to simulate the hydrological behavior of a basin, including the production, percolation, routing, and exchange processes with factors external to the atmosphere. The primary output of the model is the runoff at the basin outlet (*Q*). To use the GR2M model in a specific basin, the following data are required: the surface area of the basin in square kilometers, monthly records of rainfall (P) averaged spatially over the basin in millimeters, monthly records of potential evapotranspiration (E) in millimeters, and initial values for maximum capacity of soil storage controlled by parameter X1. The catchment is divided into hydrological response units, each characterized by its own set of parameters representing soil type, land cover, and topography. Equations describing the interactions between precipitation, infiltration, runoff generation, and routing are applied to each Hydrological Response Units (HRUs) to capture spatial variability within the catchment.

4.3 Distance Based Method

Distance-based methods serve as a vital tool for estimating hydrological model parameters in ungauged basins by extrapolating values from donor catchments to the target station or ungauged catchments. Among these methods, two prominent approaches stand out: IDW and Inverse Similarity Weighted (ISW). In the realm of hydrological modeling, the IDW technique relies on the principle of proximity. It assigns weights to neighboring catchments based on their distances from the target location. Simply put, the closer a donor catchment is to the target, the higher the influence of its parameter values on the estimation. This method operates under the assumption that catchments closer in space share more similar hydrological characteristics, hence their parameters are more applicable to the target basin.

On the other hand, ISW methodology takes a nuanced approach by considering the similarity of physical attributes between catchments. Instead of solely relying on distance, ISW evaluates the resemblance in terms of various hydrological characteristics such as land use, soil type, topography, and climate. Consequently, catchments exhibiting greater similarity in these aspects are accorded higher weights in parameter transfer, irrespective of their spatial proximity to the target. This approach acknowledges that basins with similar physical attributes are likely to exhibit comparable hydrological behaviors, making their parameter values more relevant for estimation purposes. In essence, while IDW emphasizes spatial closeness as a determinant factor, ISW broadens the scope by incorporating the likeness of hydrological attributes, thereby offering a more comprehensive and nuanced approach to parameter estimation in ungauged basins.

Equation (16) given below showed the weighted average parameter value (P_w) for the target catchment can be calculated using the IDW formula:

$$P_w = \frac{\sum_{i=1}^{n} \frac{P_i}{d_i^p}}{\sum_{i=1}^{n} \frac{1}{d_i^p}} \quad (16)$$

Where,

P_i = Parameter value of the donor catchment i

d_i = Parameter value of the donor catchment i

P = power parameter that determines the rate of decrease of influence with distance

n = number of donor catchments

The IDW method involves a step-by-step process to achieve spatial interpolation. Initially, a target point, where the parameter needs to be estimated, is identified. The method then selects a set of known points within a specified radius or neighborhood around this target. Each of these known points has associated parameter values and distances from the target point.

The distances are critical in IDW as they determine the weights assigned to each known point. The weight of each known point is calculated as the inverse of its distance to the target point, raised to a power (commonly used 2). This power parameter controls how quickly the influence of a known point diminishes with increasing distance. After calculating the weights, the method computes the weighted average of the known parameter values.

The IDW also ensures that closer points have a more significant impact on the estimated value than those farther away. This method is particularly useful when there is a dense network

of known points, as it can leverage the spatial relationship effectively. Additionally, IDW does not require assumptions about the statistical distribution of the data, making it a non-parametric method. This flexibility and simplicity are part of why IDW is widely used, despite its limitation of not accounting for physical or hydrological similarities beyond distance.

4.4 Regression Analysis

4.4.1 Cluster Analysis for Zonation

K-clustering is a machine learning technique used to group similar data points into clusters, while regression predicts outcomes based on input variables. By using K-clusters, data can be segmented into distinct groups, allowing regression models to be applied more accurately within each cluster, improving prediction precision.

Homogeneity is essential for developing regional relationships in hydrological studies. Grouping regions with similar hydrological behaviors ensures that models and predictions are accurate, leading to more precise analysis and interpretation of hydrological data. This, in turn, improves water resource management and planning. In this study, data from 25 grid stations, including IMD gridded rainfall and minimum and maximum temperatures, were utilized to identify clusters, providing a comprehensive dataset for the analysis.

The machine learning technique employed for identifying homogeneous clusters in this study is K-means clustering, a method well-documented in hydrological research by Peng and Wei (2007), Wu et al. (2021) and Long et al. (2021). K-means clustering is a partitioning method that divides the dataset into K distinct, non-overlapping subsets or clusters. Each cluster is characterized by its centroid, which is the mean of the data points within the cluster. The objective is to minimize the within-cluster sum of squares (WCSS), which measures the variability within each cluster. By doing so, the method ensures that data points within a cluster are as similar as possible, while clusters themselves are as distinct as possible.

In this study, a Python program was developed for K-means clustering using pandas, pycaret, and matplotlib. The optimal number of clusters was determined via the WCSS method's "elbow point." The analysis helps identify homogeneous regions, enhancing spatial variability understanding, modeling accuracy, and water resource management decisions.

4.4.2 Rainfall-Runoff Regression Analysis

Rainfall-runoff regression analysis is an indispensable technique in the field of hydrology, playing a critical role in water resource management, flood forecasting, and

understanding the hydrological behavior of a region. By analyzing historical rainfall and runoff data, regression analysis allows researchers and engineers to develop mathematical models that can predict runoff based on rainfall inputs. This capability is essential for planning and managing water resources, mitigating the impacts of floods, and making informed decisions in various water-related sectors. This report aims to provide an in-depth understanding of rainfall-runoff regression analysis, its significance, methodologies, and applications in water resource management.

The relationship between rainfall and runoff is a fundamental aspect of the hydrological cycle. Understanding this relationship is crucial for several reasons. Firstly, it aids in predicting flood events. Flooding can cause significant damage to infrastructure, agriculture, and human life. By developing accurate rainfall-runoff models, we can forecast floods more reliably and implement measures to mitigate their impact. Secondly, these models are essential for water resource management. They help in planning the allocation and use of water resources, ensuring that there is sufficient water supply for various needs, including domestic, agricultural, industrial, and ecological requirements. Lastly, understanding the rainfall-runoff relationship is vital for designing hydraulic structures such as dams, levees, and drainage systems. Accurate models ensure that these structures are capable of handling the expected runoff and prevent structural failures. The choice of regression model in rainfall-runoff analysis depends on the nature of the relationship between rainfall and runoff. Simple linear regression is often used when a straightforward linear relationship is anticipated between these variables. This method involves fitting a linear equation to the observed data, where the runoff is expressed as a function of rainfall, an intercept, and a slope. The intercept represents the baseline runoff when there is no rainfall, and the slope indicates the rate of increase in runoff with an increase in rainfall. This model is suitable for scenarios where the relationship between rainfall and runoff is relatively simple and direct.

However, in many cases, the relationship between rainfall and runoff is more complex, exhibiting non-linear patterns. In such situations, polynomial regression or multiple linear regression techniques can be employed. Polynomial regression involves fitting a polynomial equation to the data, which allows for the modeling of more complex, curved relationships. This approach is useful when the rate of runoff increase varies at different levels of rainfall. Multiple linear regression, on the other hand, involves multiple independent variables. In addition to rainfall, other factors such as soil moisture, land use, and topography may influence runoff. By incorporating these additional variables, multiple linear regression models can

provide a more comprehensive understanding of the factors affecting runoff.

Equation 17 shows the simple linear regression.

$$Y = \beta_0 - \beta_{1x} + \epsilon \quad (17)$$

Where:

Y is the dependent variable (the variable being predicted or explained).

X is the independent variable (the variable used to predict the dependent variable).

β_0 is the intercept (the value of Y when X is zero).

β_1 is the regression coefficient (the change in Y associated with a one-unit change in X).

ϵ is the error term (representing the random variability or unexplained variation in Y).

Rainfall-runoff regression analysis has numerous applications in water resource management. One of the primary applications is in flood forecasting. Accurate prediction of runoff based on rainfall data allows for the timely warning of potential flood events. This capability is crucial for implementing flood mitigation measures, such as the construction of levees, the creation of floodways, and the development of emergency response plans. Effective flood forecasting can significantly reduce the damage caused by floods and save lives.

Another important application is in the planning and management of water resources. Understanding the rainfall-runoff relationship helps in predicting water availability and planning its allocation for various uses. For example, in agricultural regions, accurate runoff predictions can inform irrigation planning, ensuring that crops receive sufficient water during dry periods. In urban areas, runoff models can aid in the design of drainage systems to prevent urban flooding. Additionally, these models are vital for the operation of reservoirs and dams, helping to balance water storage and release to meet downstream water demands while minimizing the risk of flooding.

Rainfall-runoff regression analysis is also essential for environmental management. It helps in assessing the impact of land use changes on hydrological behaviour. For instance, urbanization typically increases runoff due to the reduction in permeable surfaces. By analysing historical data and modelling the rainfall-runoff relationship, planners can predict the impact of urban development on runoff and implement measures to mitigate adverse effects, such as

the creation of green spaces and the installation of permeable pavements.

4.4.3 Machine Learning for R-R Regression

In the context of hydrological modeling, machine learning, a subset of artificial intelligence (AI), represents a transformative approach that empowers computers to learn from data without explicit programming. This approach involves the development of algorithms capable of analyzing and interpreting patterns within data to make informed predictions or decisions. The key components of machine learning include training data, algorithms, and model evaluation, all of which are integral to developing accurate predictive models. Specifically, regression analysis, a supervised learning technique, is frequently employed to uncover correlations between variables, allowing for the prediction of continuous output variables. This technique is commonly applied in hydrological modeling for tasks such as predictions, forecasting, and time series analysis.

Machine learning techniques are increasingly utilized in hydrological modelling to enhance the accuracy and efficiency of predicting various hydrological parameters. These algorithms are particularly well-suited for handling large datasets and uncovering complex relationships within the data. By leveraging historical data, machine learning models can learn underlying patterns and provide reliable predictions, which are crucial for effective water resource management, flood forecasting, and climate impact assessments.

The ability to predict hydrological parameters accurately is essential for several reasons. First and foremost, accurate predictions of parameters such as runoff, river discharge, and water levels are critical for flood forecasting. Flooding can cause extensive damage to infrastructure, agriculture, and human lives. By developing reliable models that predict these parameters based on rainfall inputs, it becomes possible to forecast flood events more accurately and implement measures to mitigate their impacts. Additionally, these predictions are vital for water resource management. They inform the planning and allocation of water resources to ensure sufficient supply for various needs, including domestic use, agriculture, industry, and maintaining ecological balance.

Moreover, understanding and predicting hydrological parameters are crucial for designing hydraulic structures such as dams, levees, and drainage systems. Accurate models ensure that these structures are capable of handling the expected runoff and prevent structural failures. The insights gained from machine learning models also aid in assessing the impact of land use changes and climate variability on hydrological behaviour, enabling better planning and management of water resources. Table 4.1 describe the nineteen-regression model used in

the study.

Table 4. 1 Name of Regression Model

Sl. No.	Name of Regression Model
1	Bayesian Ridge
2	Lasso Regression
3	Lasso Least Angle Regression
4	Elastic Net
5	Linear Regression
6	Ridge Regression
7	Least Angle Regression
8	Huber Regressor
9	Orthogonal Matching Pursuit
10	Light Gradient Boosting Machine
11	Dummy Regressor
12	Random Forest Regressor
13	Extra Trees Regressor
14	AdaBoost Regressor
15	K Neighbors Regressor
16	Passive Aggressive Regressor
17	Extreme Gradient Boosting
18	Gradient Boosting Regressor
19	Decision Tree Regressor

The choice of regression model in rainfall-runoff analysis and hydrological modelling depends on the nature of the relationship between the variables involved. Simple linear regression is often used when a straightforward linear relationship is anticipated between these variables. This method involves fitting a linear equation to the observed data, where the dependent variable (such as runoff) is expressed as a function of the independent variable (such as rainfall), an intercept, and a slope. The intercept represents the baseline value when there is no input from the independent variable, and the slope indicates the rate of increase in the dependent variable with an increase in the independent variable. This model is suitable for scenarios where the relationship between the variables is relatively simple and direct.

However, in many cases, the relationship between hydrological parameters and their predictors is more complex, exhibiting non-linear patterns. In such situations, polynomial regression or multiple linear regression techniques can be employed. Polynomial regression

involves fitting a polynomial equation to the data, which allows for modelling of more complex, curved relationships. This approach is useful when the rate of change in the dependent variable varies at different levels of the independent variable. Multiple linear regression, on the other hand, involves multiple independent variables. In addition to rainfall, other factors such as soil moisture, land use, and topography may influence hydrological parameters. By incorporating these additional variables, multiple linear regression models can provide a more comprehensive understanding of the factors affecting the dependent variable.

In recent years, advanced machine learning techniques, such as decision trees, random forests, support vector machines (SVMs), and neural networks, have gained prominence in hydrological modelling. These techniques offer greater flexibility and accuracy in modelling complex relationships and handling large datasets. Decision trees and random forests, for instance, are capable of capturing non-linear interactions between variables and are robust against overfitting. SVMs are effective for high-dimensional data and can handle both linear and non-linear relationships by transforming the data using kernel functions. Neural networks, particularly deep learning models, excel at capturing intricate patterns and dependencies in the data, making them highly effective for time series forecasting and other hydrological modelling tasks.

In this study, machine learning models were developed for each month of the kharif season—June, July, August, September, and October—to estimate hydrological parameters in various river basins. The kharif season, characterized by monsoonal rains, is critical for agricultural planning and water resource management in the region. Accurate predictions of hydrological parameters during this period are essential for mitigating flood risks and optimizing irrigation practices. The study involved the application of nineteen different regression models to analyse the hydrological data, each chosen for its unique approach to handling data and making predictions. This diversity allowed for a comprehensive evaluation of the models' performance in estimating hydrological parameters.

The dataset used in the study was divided into training and testing sets. Seventy percent (70%) of the station data was allocated for training the models, while the remaining thirty percent (30%) was reserved for testing. This division ensures that the models are trained on a substantial portion of the data while still being evaluated on an independent subset to assess their predictive performance. The performance of each regression model was evaluated based on its ability to accurately predict hydrological parameters, using key metrics such as the coefficient of determination (R^2), mean absolute error (MAE), and root mean square error (RMSE). These metrics provide insights into the models' accuracy, robustness, and reliability

in predicting continuous variables.

4.4.3.1 Key Metrics for Model Evaluation

The coefficient of determination (R^2) is a statistical measure that indicates how well the regression model explains the variability of the dependent variable. It ranges from 0 to 1, with higher values indicating a better fit of the model to the data. However, R^2 can be misleading if used alone and should be considered alongside other metrics. Mean absolute error (MAE) measures the average absolute errors between predicted and actual values, giving an idea of how close the predictions are to the actual values. Lower MAE values indicate better accuracy. Mean squared error (MSE) measures the average squared differences between predicted and actual values, penalizing larger errors more heavily than smaller errors. While useful for understanding the spread of errors, MSE is less interpretable than MAE since it squares the errors.

Root mean square error (RMSE) is the square root of the MSE, providing a measure of the average magnitude of errors in the same units as the target variable. It is easier to interpret than MSE since it is on the same scale as the original data. Root mean squared logarithmic error (RMSLE) measures the ratio between the actual and predicted values after taking the logarithm of both. This metric is particularly useful when the target variable has a wide range of values and, like RMSE, is easier to interpret than MSE.

Mean absolute percentage error (MAPE) measures the average absolute percentage difference between predicted and actual values, providing a relative measure of accuracy. This metric is often used in forecasting to understand the size of the errors relative to the actual values. By examining these various metrics, researchers can gain a comprehensive understanding of the models' performance and select the most appropriate model for their specific hydrological application.

4.4.3.2 Performance Evaluation and Model Selection

In this study, the analysis revealed that maximizing the coefficient of determination (R^2) and minimizing the root mean square error (RMSE) were key indicators of model performance. The models that achieved these criteria demonstrated high accuracy and robustness in predicting hydrological parameters, making them valuable tools for water resource management during the kharif season. For instance, MAE provided a straightforward measure of how close the predictions were to actual values, while MSE offered insights into the spread

of errors. RMSE, being on the same scale as the original data, was particularly useful for interpreting the average magnitude of errors. R^2, despite its limitations, helped assess how well the model captured the variability in the data. RMSLE and MAPE further contributed to understanding the performance of models in scenarios with a wide range of target values and relative error measures, respectively.

The evaluation of the nineteen different regression models provided a comprehensive understanding of their performance in various contexts. Simple linear regression models were found to be effective in scenarios where the relationship between rainfall and runoff was straightforward and linear. Polynomial regression models, with their ability to fit curved relationships, were better suited for scenarios with non-linear interactions. Multiple linear regression models, which included additional variables such as soil moisture and land use, provided a more nuanced understanding of the factors influencing hydrological parameters.

Advanced machine learning models, such as decision trees, random forests, SVMs, and neural networks, demonstrated superior performance in handling complex relationships and large datasets. Decision trees and random forests, with their ability to model non-linear interactions and robustness against overfitting, showed high accuracy in predicting hydrological parameters. SVMs, with their effectiveness in high-dimensional data and flexibility in handling both linear and non-linear relationships, also performed well. Neural networks, particularly deep learning models, excelled at capturing intricate patterns and dependencies in the data, making them highly effective for time series forecasting and other hydrological modelling tasks.

4.5 Web Based Application for Computation of Water Yield

A web-based application is being developed for the computation of water yield for any ungauged sub-basin of the Ganga river in Madhya Pradesh. This tool is accessible on Google Colab and provides a user-friendly interface for water yield estimation. Users begin by inputting the name of the sub-basin, basin, catchment area, monthly rainfall, and monthly evaporation after which the regional model will compute coefficient values (X1 and X2) that best represent the hydrological characteristics of the area. Following this, users enter the the start and end months and years for the desired yield calculation period.

Monthly rainfall and evaporation data will be computed, allowing the tool to compute the runoff. The application generates a detailed rainfall-runoff time series as part of the results. Additionally, it calculates dependable monthly and yearly runoff at 50%, 60%, 70%, 75%, and

95% confidence levels, offering valuable insights for water resource planning and management. This tool is particularly beneficial for ungauged basins, where direct measurements are unavailable, helping researchers, water managers, and planners make more informed decisions. Currently, this application is under development.

CHAPTER: 5

RESULTS & DISCUSSION

This chapter presents the results of the study and provides an in-depth discussion of their significance in relation to the research objectives and existing literature. The analysis focuses on identifying key trends and patterns in the data, examining their implications, and exploring how these findings contribute to the current understanding of the subject. The discussion also addresses the limitations of the study and suggests avenues for future research. By integrating the results with theoretical and practical perspectives, this chapter aims to offer a comprehensive understanding of the study's contributions to the field.

5.1 Computation of discharge through GEE based SCS- CN model

The computation of discharge through the Google Earth Engine (GEE) based SCS-CN model harnesses the power of large-scale geospatial data analysis to estimate direct runoff in watersheds. The SCS-CN model, developed by the USDA, is widely utilized for this purpose. It relies on rainfall data and land cover characteristics to predict direct runoff, making it an essential tool in hydrological studies and water resource management.

In this study, the GEE-based SCS-CN model was applied to six sub-basins of the Ganga river located in Madhya Pradesh, India: Chambal, Betwa, Sindh, Son, Tons, and Ken. The integration of GEE with the SCS-CN model allows for efficient and scalable processing of geospatial data, facilitating detailed hydrological analysis across large areas.

GEE provides access to various global rainfall datasets, such as CHIRPS (Climate Hazards Group InfraRed Precipitation with Station data) or TRMM (Tropical Rainfall Measuring Mission). These datasets are preprocessed within GEE to ensure consistency and accuracy. GEE also offers access to land cover datasets like MODIS (Moderate Resolution Imaging Spectroradiometer) or the National Land Cover Database (NLCD). This data is crucial for determining the Curve Number (CN) values based on land cover types. The CN is a key parameter in the SCS-CN model, representing the land's ability to infiltrate water. It is influenced by land cover, soil type, and antecedent moisture conditions. Within GEE, land cover data is classified, and corresponding CN values are assigned based on standard SCS-CN tables. Figure 5.1 illustrates the GEE interface for the SCS-CN model, including steps for selecting the basin for modeling. Users can interactively select the basin boundaries, input

parameters, and initiate the modeling process. GEE's interface allows users to visualize the input data, intermediate results, and final runoff estimates, enhancing the understanding and accuracy of the model outputs. Overall, the GEE-based implementation of the SCS-CN model represents a significant advancement in hydrological modeling, leveraging modern technology to enhance our understanding and management of water resources in large and complex basins.

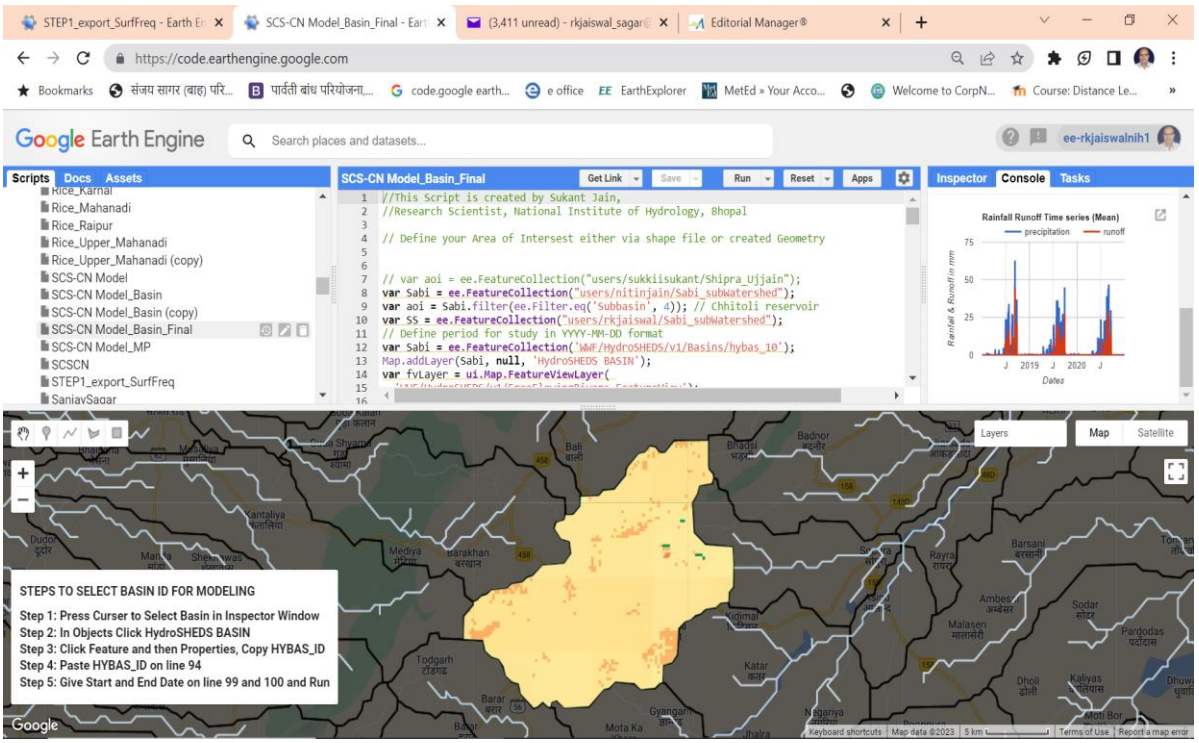


Figure 5. 1 SCS-CN model in GEE.

Comparison of observed and computed runoff is a critical step in hydrological modeling and water resource management. It involves assessing how well a hydrological model's predictions match the actual measured runoff in a watershed. This comparison allows for model evaluation, validation, and the identification of potential model improvements. Figure 5.2 shows the comparison of observed and computed runoff of Bina river which is tributary of Betwa river with 5th order stream, having catchment area of 115 sq km. Rainfall exhibits a clear seasonal pattern with significant peaks occurring at regular intervals, typically around the start of each year. This indicates distinct rainy seasons followed by dry periods. Despite the fluctuating rainfall, the actual volume remains relatively high and stable, with values generally close to 500 MCM. This suggests that the system being measured has a high capacity or effective storage/reservoir mechanisms that maintain the volume despite changes in rainfall. There are occasional drops in volume, but these do not correlate directly with the rainfall spikes. This could imply other factors influencing volume changes, such as water

usage, evaporation, or management practices. The graph indicates that the volume does not respond significantly to the immediate changes in rainfall, pointing towards a buffer or regulation mechanism that moderates the effect of rainfall on volume.

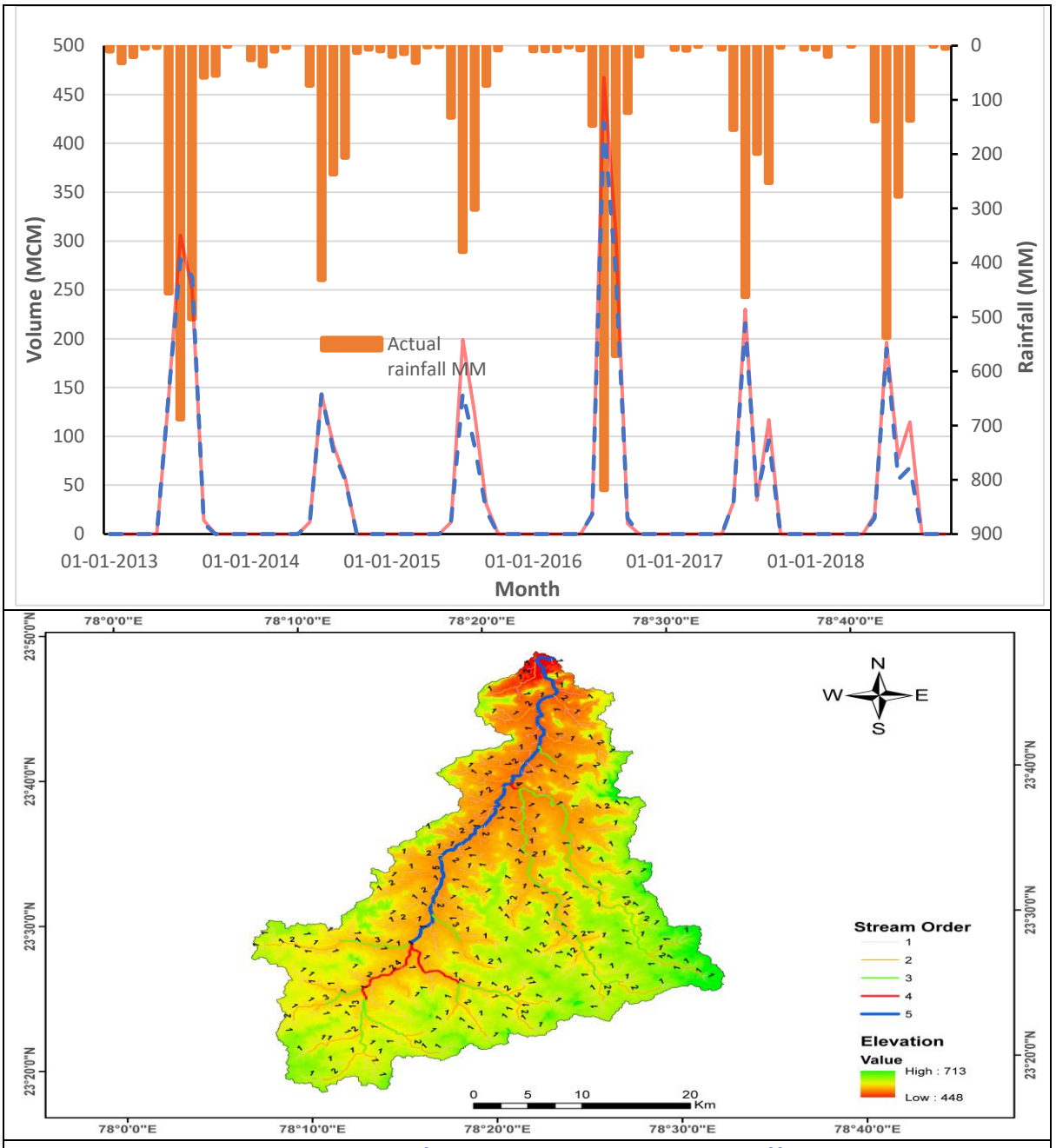


Figure 5. 2 Comparison of observed and Computed runoff: Bina River.

Same as Figure 5.3 shows the comparison of observed and computed runoff of bah river a is tributary of Betwa river with 5th order stream and the catchment area is 918msq km. Coefficient of determination in computed and observed runoff is 0.80. a time series graph and a scatter plot, both related to water discharge. The time series graph compares observed

discharge (depicted by a dashed blue line) with computed discharge (shown by a solid orange line) over the period from January 1996 to September 2003. The y-axis measures discharge in centimeters. Both observed and computed discharge exhibit periodic peaks, indicating seasonal variations in water flow. The computed discharge closely follows the observed data, suggesting that the computational model accurately reflects the actual discharge patterns.

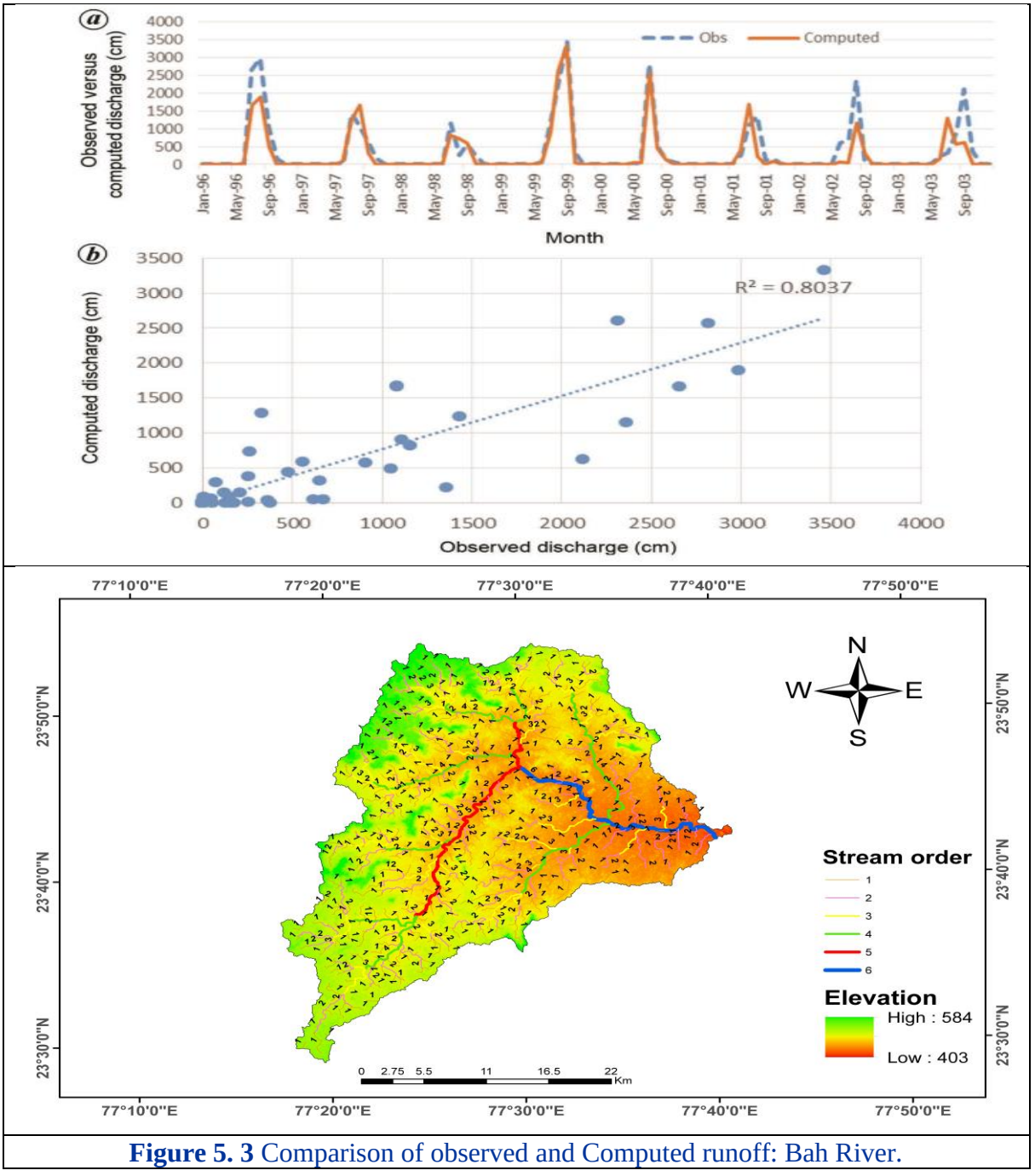


Figure 5. 3 Comparison of observed and Computed runoff: Bah River.

The scatter plot in the bottom part of the first image further supports the model's reliability. It plots observed discharge on the x-axis against computed discharge on the y-axis,

with each point representing a paired observation. The dotted blue line indicates the line of best fit, and the R-squared value of 0.8037 demonstrates a strong positive correlation between observed and computed values. This high correlation underscores the model's effectiveness in predicting discharge. illustrating stream orders and elevation levels within a specific geographic area. Stream orders are color-coded from 1 (yellow-green) to 6 (red), with higher stream orders representing the main river and lower orders representing its tributaries. Elevation is also color-coded, ranging from high elevations (584 meters, shown in red) to low elevations (403 meters, shown in green). The map includes a compass for orientation and a scale for distance measurement. This visualization highlights the relationship between elevation and stream order, showing how water flows from higher elevations through various tributaries into the main stream.

Overall, these images provide valuable insights into the region's hydrological behaviour. The high correlation between observed and computed discharge confirms the accuracy of the computational model used, making it a reliable tool for future predictions and water management decisions. The map further illustrates the topographical and hydrological dynamics of the area, essential for effective watershed management and flood prediction.

5.2 Regression Analysis (Rainfall- Runoff Relationship)

Regression analysis can be used to investigate the relationship between rainfall and runoff. The dependent variable is the runoff, and the independent variable is the rainfall. The relationship between rainfall and runoff is often nonlinear, but it can be approximated by a linear regression model.

5.2.1 Betwa Basin R-R Relationship

The figure 5.4 showing graphs illustrate the relationship between precipitation and observed discharge for the Betwa basin over the months of June, July, August, September, and October. Data for these graphs were collected from five gauging/discharge (G/D) sites within the basin: Bah, Betwa, Mohanpur, Rahatgarh, and Neemkheda. Each graph plots precipitation (in millimeters) on the x-axis against the observed discharge (in millimeters) on the y-axis.

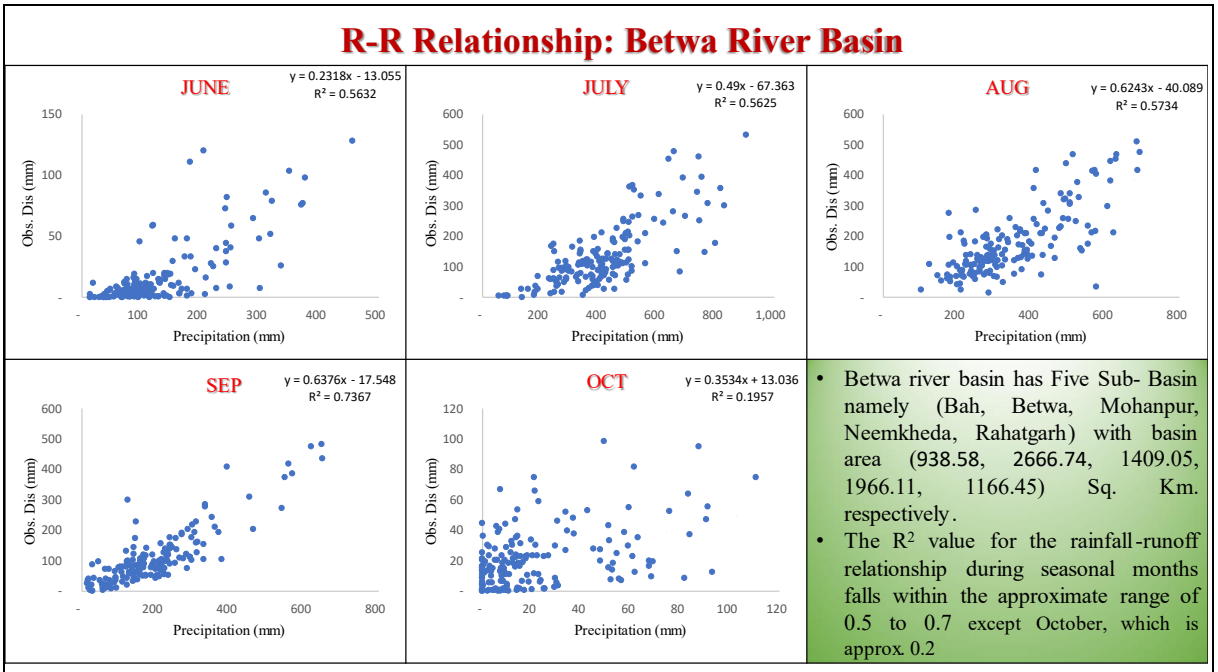


Figure 5. 4 Relationship between precipitation and observed discharge for the Betwa Basin over the months of June, July, August, September, and October.

In June, the scatter plot demonstrates a moderate positive correlation between precipitation and observed discharge. The R^2 value of 0.5632 indicates that approximately 56.32% of the variation in discharge can be explained by the amount of precipitation. This suggests that as precipitation increases, the observed discharge tends to increase as well, although there is still considerable variability.

The July graph shows a similar pattern, with an R^2 value of 0.5625. This means that about 56.25% of the variation in discharge is attributable to changes in precipitation. Like June, this indicates a moderate positive correlation where increased rainfall generally leads to higher discharge levels, but other factors also influence the discharge.

In August, the correlation between precipitation and discharge remains positive and slightly stronger, with an R^2 value of 0.5734. This suggests that 57.34% of the variation in observed discharge can be explained by precipitation levels. The trend shows that higher precipitation usually results in higher discharge, reflecting the basin's response to rainfall during this month.

The September graph reveals the strongest correlation among the months, with an R^2 value of 0.7367. This high value indicates that 73.67% of the discharge variation is explained by precipitation, suggesting a very strong positive relationship. In September, the Betwa basin's discharge is highly responsive to changes in precipitation, with increased rainfall leading to significantly higher discharge levels.

Finally, the October graph depicts a much weaker positive correlation, with an R^2 value of 0.1957. This means that only 19.57% of the variation in discharge can be explained by precipitation. The weak correlation indicates that other factors besides precipitation have a more significant impact on discharge during this month, and the relationship between rainfall and discharge is not as pronounced as in the previous months. This weaker correlation can be attributed to the fact that by October, the monsoon season is winding down, leading to lower overall rainfall. Additionally, the soil and basin may be saturated from previous months, reducing the direct impact of precipitation on discharge. Other factors, such as groundwater contributions and evapotranspiration, may play a more significant role in influencing discharge during this time. (Parmar et al., 2016)

5.2.2 Chambal R-R Relationship

Figure 5.5 describe the R-R relation of the Chambal river basin which consists of nine sub-basins: Sinduriya, Manakhedi, Maksudnagar, Mandsor, Badnagar, Kuno, Samaskhedi, Choma, and Sanwar, with varying basin areas. Each graph plots precipitation (in millimeters) on the x-axis against observed discharge (in millimeters) on the y-axis.

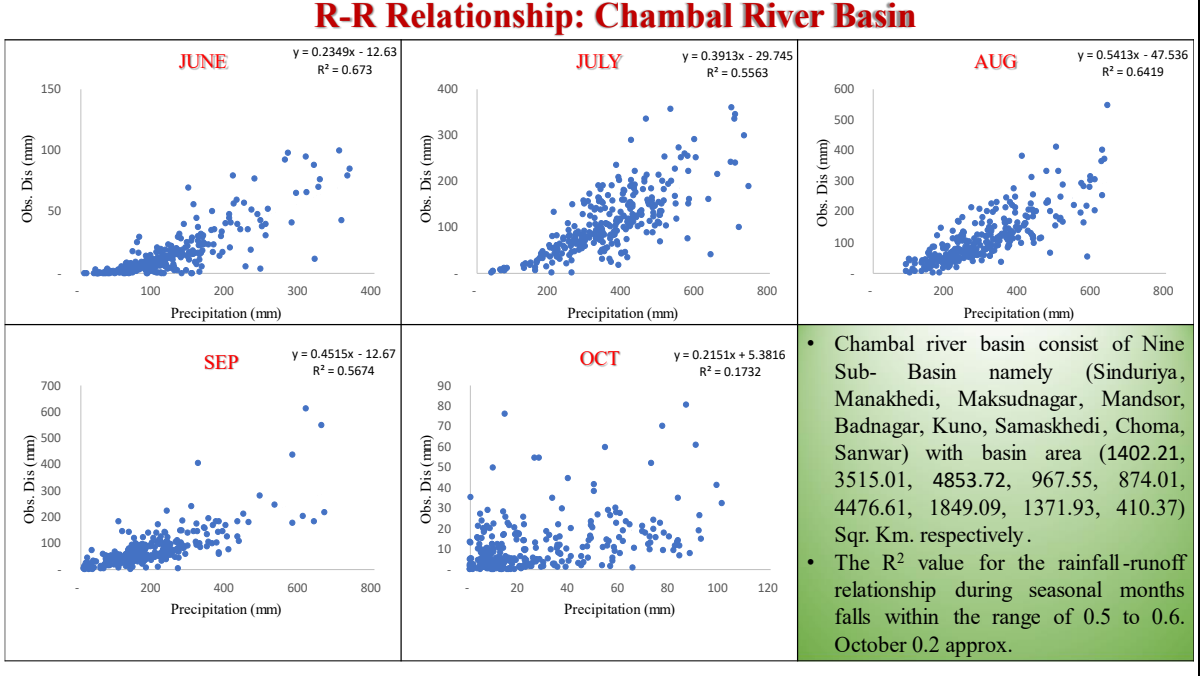


Figure 5. 5 Relationship between precipitation and observed discharge for the Chambal Basin over the months of June, July, August, September, and October.

In June, the graph shows a moderate positive correlation between precipitation and observed discharge, with an R^2 value of 0.673, indicating that approximately 67.3% of the variation in discharge can be explained by the amount of precipitation. This suggests a

relatively strong relationship, where increased rainfall tends to result in higher discharge levels.

The July graph displays a similar moderate correlation, with an R^2 value of 0.5563. This means that about 55.63% of the variation in discharge is attributable to precipitation, showing a consistent pattern of rainfall influencing discharge levels, although other factors also play a role.

In August, the correlation is slightly stronger, with an R^2 value of 0.6419, suggesting that 64.19% of the variation in discharge can be explained by precipitation levels. This indicates a significant positive relationship, where increased precipitation generally leads to higher discharge.

The September graph shows a moderate positive correlation with an R^2 value of 0.5674, meaning that 56.74% of the variation in discharge is due to precipitation. This demonstrates that rainfall continues to be a major factor influencing discharge during this month.

The October graph, similar to the Betwa basin, shows a weaker relationship between precipitation and observed discharge, with an R^2 value of 0.1732, indicating that only 17.32% of the variation in discharge can be attributed to precipitation. The weaker correlation suggests that by October, other factors such as soil saturation, groundwater contributions, and reduced rainfall play a more significant role in influencing discharge, diminishing the direct impact of precipitation. the Chambal river basin reveal that the relationship between precipitation and observed discharge is strongest in June and August, moderate in July and September, and weakest in October.

5.2.3 Ken R-R Relationship

Figure 5.6 describe the R-R relation of the Ken river basin consists of six sub-basins: Sonar, Pandwan, Kopra, Niwari, Nohta, and Tigra, with varying basin areas. Each graph plots precipitation (in millimeters) on the x-axis against observed discharge (in millimeters) on the y-axis.

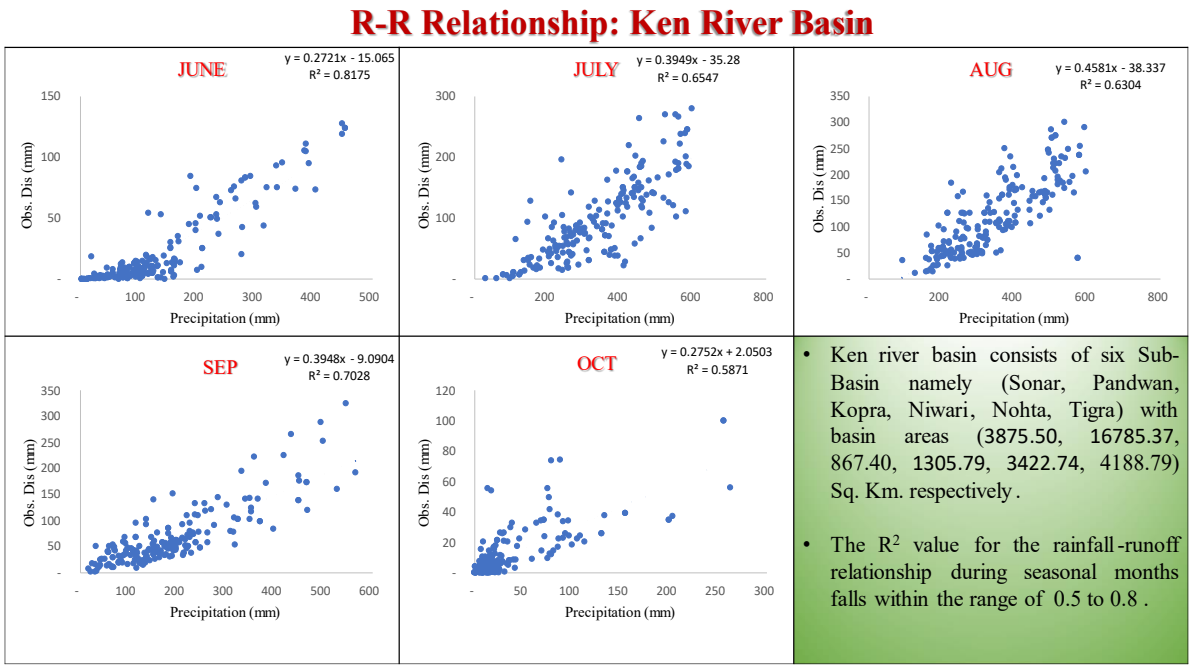


Figure 5. 6 Relationship between precipitation and observed discharge for the Ken Basin over the months of June, July, August, September, and October.

In June, the graph demonstrates a strong positive correlation between precipitation and observed discharge, with an R^2 value of 0.8175. This indicates that approximately 81.75% of the variation in discharge can be explained by the amount of precipitation, suggesting a very strong relationship where increased rainfall leads to higher discharge levels.

The July graph shows a moderate correlation, with an R^2 value of 0.6547. This means that about 65.47% of the variation in discharge is attributable to precipitation, indicating that while increased rainfall generally leads to higher discharge, other factors also influence the discharge.

In August, the correlation remains moderately strong, with an R^2 value of 0.6304, suggesting that 63.04% of the variation in discharge can be explained by precipitation levels. This demonstrates a significant positive relationship, where increased precipitation typically results in higher discharge.

The September graph shows a strong positive correlation, with an R^2 value of 0.7028, indicating that 70.28% of the variation in discharge is due to precipitation. This highlights that rainfall is a major factor influencing discharge during this month, with increased precipitation leading to significantly higher discharge levels.

The October graph depicts a moderate positive correlation, with an R^2 value of 0.5871, meaning that 58.71% of the variation in discharge can be explained by precipitation. Although this is weaker than the previous months, it still suggests that rainfall has a notable impact on discharge levels, but other factors may also play a significant role.

5.2.4 Sindh R-R Relationship

Figure 5.7 describe the R-R relation the Sindh river basin comprises five sub-basins: Didi, Amola, Goraghat, Budhara, and Daboh, each with distinct basin areas. Precipitation (in millimeters) is plotted on the x-axis, while observed discharge (in millimeters) is plotted on the y-axis.

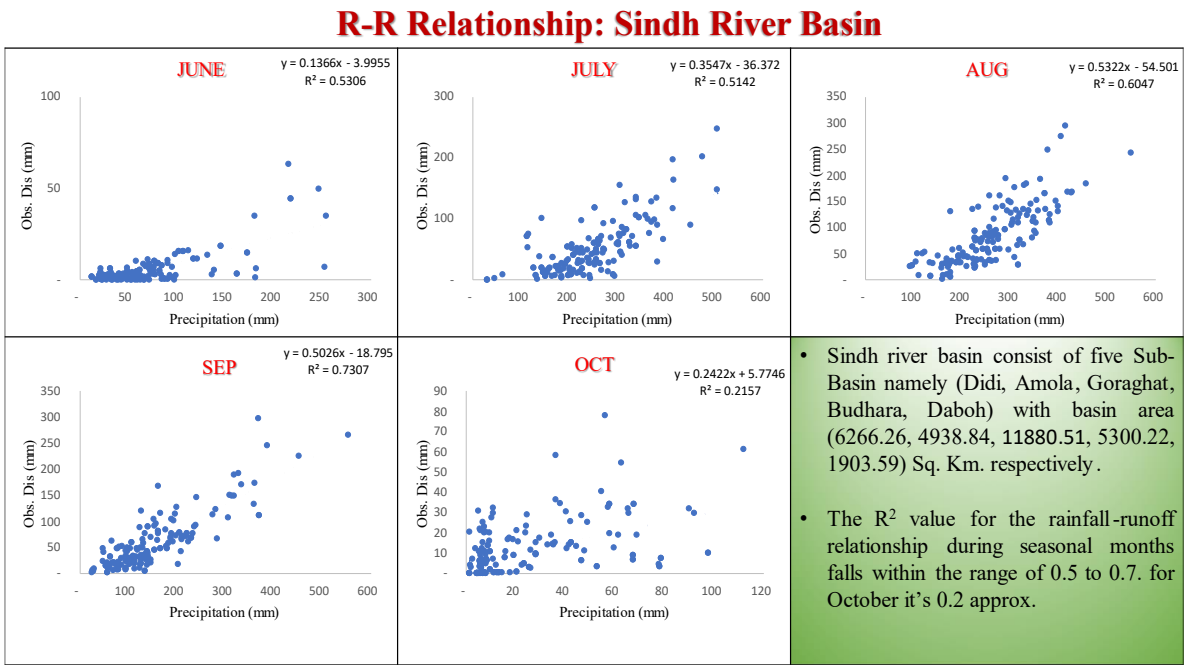


Figure 5. 7 Relationship between precipitation and observed discharge for the Sindh Basin over the months of June, July, August, September, and October.

In June, the graph indicates a moderate positive correlation between precipitation and observed discharge, with an R^2 value of 0.5306. This suggests that approximately 53.06% of the variation in discharge can be explained by the amount of precipitation, showing a fairly strong relationship where increased rainfall leads to higher discharge levels.

The graph for July shows a similar moderate correlation, with an R^2 value of 0.5142. This means that 51.42% of the variation in discharge is due to precipitation, indicating that rainfall generally impacts discharge, although other factors may also influence the outcomes.

August displays a stronger positive correlation, with an R^2 value of 0.6047. This suggests that 60.47% of the variation in discharge can be attributed to precipitation, highlighting a significant relationship where increased rainfall typically results in higher discharge. In September, the correlation strengthens further, with an R^2 value of 0.7307. This

indicates that 73.07% of the variation in discharge is due to precipitation, showing a robust relationship where increased precipitation significantly impacts discharge levels.

The October graph shows a weaker correlation between precipitation and observed discharge, with an R^2 value of 0.2157, meaning only 21.57% of the variation in discharge can be explained by precipitation. This weaker relationship suggests that by October, factors such as soil saturation, groundwater contributions, and reduced rainfall play a more significant role in influencing discharge, reducing the direct impact of precipitation.

5.2.5 Son R-R Relationship

The Figure 5.8 displays a series of scatter plots demonstrating the rainfall-runoff relationship for different months in the Son river basin. Each scatter plot represents the observed discharge (in millimeters) against precipitation (in millimeters) for June, July, August, September, and October. The plots include data from four sub-basins: Marsiabada, Amarpur, Bhadara, and Ghathkirwa, with basin areas of 10,623.64 km², 395.19 km², 3,139.26 km², and 1,012.76 km² respectively.

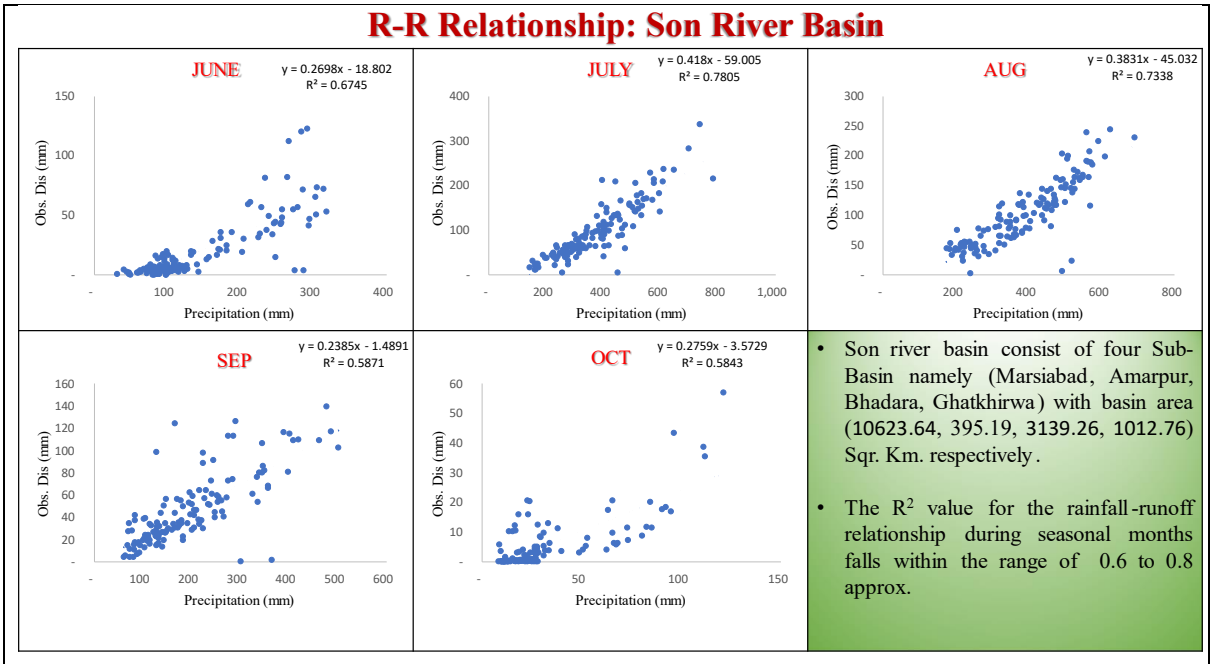


Figure 5. 8 Relationship between precipitation and observed discharge for the Son Basin over the months of June, July, August, September, and October.

For each month, a trend line and corresponding equation are provided to show the linear relationship between precipitation and discharge. The R^2 values indicate the goodness of fit for each month's regression model, with values ranging approximately from 0.6 to 0.8. This suggests a moderate to strong correlation between rainfall and runoff in these sub-basins during the seasonal months.

- **June:** The scatter plot for June shows a positive correlation with an R^2 value of 0.6745, indicating that around 67.45% of the variance in observed discharge can be explained by precipitation.
- **July:** The plot for July has a higher R^2 value of 0.7805, suggesting a stronger relationship where 78.05% of the variance is explained by precipitation.
- **August:** The August plot also demonstrates a strong correlation with an R^2 value of 0.7338.
- **September:** In September, the R^2 value is 0.5871, indicating a moderate correlation.
- **October:** The October plot shows an R^2 value of 0.5843, similar to September, suggesting a moderate relationship between precipitation and discharge.

5.2.6 Tons R-R Relationship

The Figure 5.9 displays The Tons river basin R-R relation which comprises four sub-basins: Sansarpur, Patehera, Khatkhari, and Maihar, each with distinct basin areas. Precipitation (in millimeters) is plotted on the x-axis, while observed discharge (in millimeters) is plotted on the y-axis.

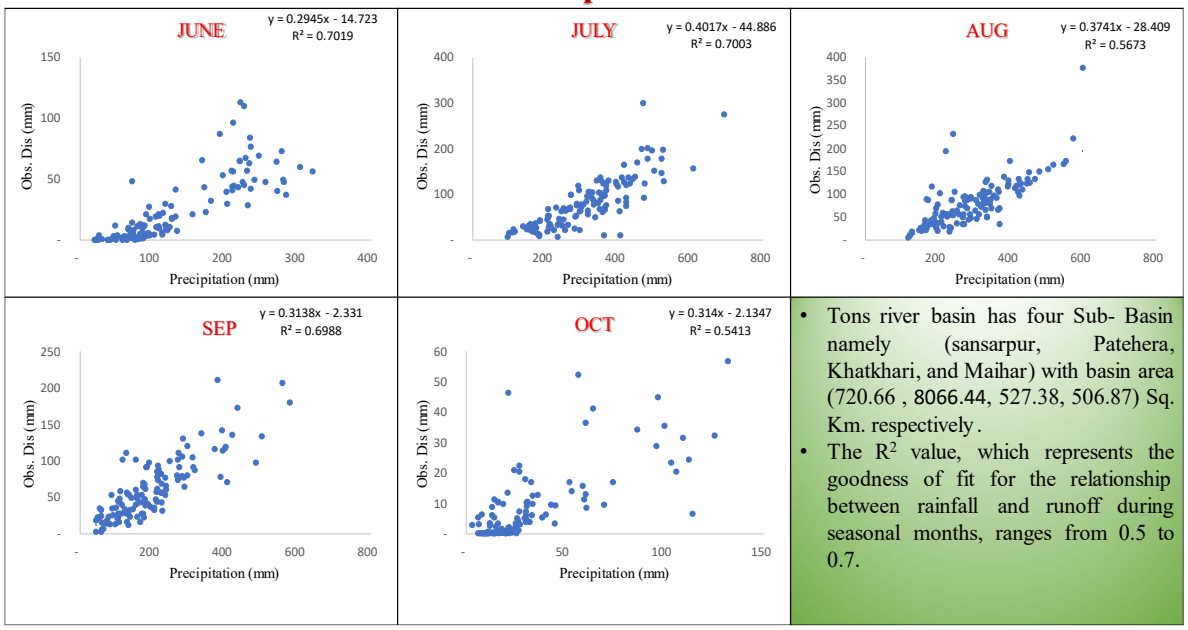


Figure 5. 9 Relationship between precipitation and observed discharge for the Tons Basin over the months of June, July, August, September, and October.

In June, the graph indicates a strong positive correlation between precipitation and observed discharge, with an R^2 value of 0.7019. This suggests that approximately 70.19% of the variation in discharge can be explained by the amount of precipitation, showing a significant relationship where increased rainfall leads to higher discharge levels.

The graph for July shows a similar strong correlation, with an R^2 value of 0.7003. This means that 70.03% of the variation in discharge is due to precipitation, indicating that rainfall generally impacts discharge, although other factors may also influence the outcomes.

August displays a moderate positive correlation, with an R^2 value of 0.5673. This suggests that 56.73% of the variation in discharge can be attributed to precipitation, highlighting a substantial relationship where increased rainfall typically results in higher discharge.

In September, the correlation remains strong, with an R^2 value of 0.6988. This indicates that 69.88% of the variation in discharge is due to precipitation, showing a robust relationship where increased precipitation significantly impacts discharge levels.

The October graph shows a moderate correlation between precipitation and observed discharge, with an R^2 value of 0.5413, meaning 54.13% of the variation in discharge can be

explained by precipitation. This moderate relationship suggests that in October, factors such as soil saturation, groundwater contributions, and reduced rainfall play a more significant role in influencing discharge, slightly reducing the direct impact of precipitation.

The moderate correlation indicates that the direct influence of rainfall on streamflow might vary across the sub-basins, implying that other factors, such as geological characteristics, land use, and soil properties, might also play important roles in determining the runoff response to rainfall events. These findings underscore the complexity of the hydrological dynamics in the Ganga river sub-basins within Madhya Pradesh. Accurate and site-specific water resource management strategies should consider these varying correlations to ensure sustainable utilization and conservation of water resources in the region. Further research and modeling efforts will be crucial for a deeper understanding of these sub-basins' hydrological behavior and for developing effective water management policies to meet the region's growing water demands.

The Table 5.1 provides a detailed comparative analysis of six river basins—Chambal, Betwa, Ken, Sindh, Son, and Tons—by presenting the R^2 values and linear regression equations for each month from June to October. These statistics help to understand the strength of the linear relationships and variations in the data for each basin over the months, shedding light on seasonal patterns and potential influencing factors.

Table 5. 1 Basin R2 Value for Rainfall- Runoff Relationship.

Sl. No.	Basin	June	July	Aug	Sep	Oct
1.	Chambal	R^2 = 0.673	R^2 = 0.5563	R^2 = 0.6419	R^2 = 0.5674	R^2 = 0.1732
		y = 0.2349x - 12.63	y = 0.391x - 29.745	y = 0.5413x - 47.536	y = 0.4515x - 12.67	y = 0.215x + 5.3816
2.	Betwa	R^2 = 0.5468	R^2 = 0.5625	R^2 = 0.5734	R^2 = 0.7367	R^2 = 0.1957
		y = 0.252x - 15.154	y = 0.49x - 67.363	y = 0.6243x - 40.089	y = 0.6376x - 17.548	y = 0.353x + 13.036
3.	Ken	R^2 = 0.8175	R^2 = 0.6547	R^2 = 0.6304	R^2 = 0.7028	R^2 = 0.5871
		y = 0.272x - 15.065	y = 0.3949x - 35.28	y = 0.4581x - 38.337	y = 0.3948x - 9.0904	y = 0.275x + 2.0503
4.	Sindh	R^2 = 0.5306	R^2 = 0.5142	R^2 = 0.6047	R^2 = 0.7307	R^2 = 0.2157
		y = 0.136x - 3.9955	y = 0.354x - 36.372	y = 0.5322x - 54.501	y = 0.502x - 18.795	y = 0.242x + 5.7746
5.	Son	R^2 = 0.6745	R^2 = 0.7805	R^2 = 0.7338	R^2 = 0.5871	R^2 = 0.5843
		y = 0.269x - 18.802	y = 0.418x - 59.005	y = 0.3831x - 45.032	y = 0.2385x - 1.4891	y = 0.2759x - 3.5729
6.	Tons	R^2 = 0.7019	R^2 = 0.7003	R^2 = 0.5673	R^2 = 0.6988	R^2 = 0.5413
		y = 0.294x - 14.723	y = 0.401x - 44.886	y = 0.3741x - 28.409	y = 0.3138x - 2.331	y = 0.314x - 2.1347

In **June**, the Ken basin shows the highest R^2 value at 0.8175, indicating a strong correlation, which suggests a well-defined relationship in the data for that month. This might

be due to consistent rainfall patterns or specific hydrological characteristics of the Ken basin that respond predictably to climatic variables. On the other hand, the Sindh basin has the lowest R^2 value at 0.5306, implying a weaker correlation which might be due to more variable rainfall or diverse land-use patterns affecting the runoff.

Moving to **July**, the Son basin demonstrates the highest R^2 value at 0.7805, reflecting a strong linear relationship. This could be attributed to the peak monsoon period where the rainfall-runoff relationship is more pronounced. Conversely, the Sindh basin remains the lowest at 0.5142, which might suggest ongoing variability in rainfall or other influencing factors such as soil saturation levels and agricultural activities.

In **August**, the Son basin again shows a high R^2 value of 0.7338, indicating a consistent and strong correlation during the late monsoon period. The lower R^2 value for Sindh (0.6047) still reflects moderate variability, possibly due to regional differences in monsoon intensity or catchment characteristics.

September reveals a significant increase in the R^2 value for the Betwa basin at 0.7367, suggesting an improved data correlation which might be due to stabilized hydrological conditions post-peak monsoon. The Sindh basin, leading with the highest R^2 value of 0.7307, indicates that by this time, the basin's hydrological responses have become more predictable, possibly due to soil saturation and reduced variability in rainfall patterns.

By **October**, the Ken basin displays a relatively high R^2 value at 0.5871, indicating a reasonably good fit, which might be due to receding monsoon impacts and more stable hydrological conditions. In contrast, the Chambal basin has the lowest R^2 value at 0.1732, suggesting a very weak correlation, possibly due to reduced rainfall and increased influence of other factors like evapotranspiration, groundwater contributions, or human activities such as irrigation withdrawals.

Significance and Insights

- **Ken Basin**: Shows consistently high R^2 values, indicating strong and stable linear relationships, especially in June and October. This suggests a reliable hydrological response possibly due to less variable rainfall patterns and consistent land use.
- **Sindh Basin**: Generally lower R^2 values with a notable increase in September, indicating variability in its hydrological response potentially due to diverse rainfall patterns, soil types, and land use practices.
- **Son Basin**: Exhibits strong correlations in July and August, reflecting the impact of

peak monsoon on hydrological responses. This basin might have well-defined catchment characteristics that respond predictably during heavy rainfall.

- **Chambal Basin**: Displays significant variability with a sharp drop in October, suggesting that this basin might be influenced by other factors like reduced rainfall and increased human intervention during the dry post-monsoon period.
- **Betwa Basin**: Improved correlation in September indicates a stabilizing hydrological regime post-monsoon, suggesting that early and late monsoon periods have different impacts on its hydrological behaviour.
- **Tons Basin**: Maintains relatively high R^2 values throughout, indicating consistent and predictable hydrological responses, likely due to stable climatic conditions and uniform catchment characteristics.

Understanding these patterns is crucial for effective water resource management, planning for agricultural activities, and mitigating the impacts of floods and droughts in these basins. The variability and strength of correlations highlight the need for basin-specific strategies to manage water resources efficiently.

5.3 Machine learning for R-R Regression

A recent Python program for regression analysis has introduced modifications to improve the accuracy and robustness of the model. In this enhanced analysis, 70% of the station data is allocated for training, while the remaining 30% is reserved for testing. Nineteen different regression models have been applied, including Huber Regression (huber), Decision Tree Regressor (DTR), Ridge Regression (RR), and Random Forest (RF).

The performance of these models is evaluated using several metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Coefficient of Determination (R^2), Root Mean Squared Logarithmic Error (RMSLE), and Mean Absolute Percentage Error (MAPE). The analysis concludes that the best models are those that maximize the coefficient of determination (R^2) and minimize the root mean square error (RMSE), indicating a good fit and high predictive accuracy. The Table 5.2 describes the performance of various regression models for predicting runoff in different basins, evaluated across the months of June to October. For the Betwa basin, the best models included Huber Regressor, Orthogonal Matching Pursuit, and Bayesian Ridge. The mean rainfall ranged from 127.19 mm in June to 411.9 mm in July, with mean ET0 ranging from 111.54 mm in August to 180.09 mm in June. The mean runoff varied from 16.98 mm in June to 175.81 mm in August, with R^2 values peaking in September (0.4769) and the lowest in

October (0.0363), while RMSE values ranged from 16.597 in October to 66.9791 in August.

Table 5. 2 Best fit regression model for monsoon month of ever basin.

Basin	Basin (Betwa)	June	July	Aug	Sep	Oct
Betwa	Best Model	Huber Regressor	Huber Regressor	Orthogonal Matching Pursuit	Huber Regressor	Bayesian Ridge
	Mean Rainfall	127.19	411.9	345.84	192.97	22.2
	Mean ET0	180.09	128.3	111.54	116.35	131.3
	Mean Runoff	16.98	134.4	175.81	105.49	20.9
	R^2	0.2554	0.45	0.3749	0.4769	0.0363
	RMSE	18.2125	64.26	66.9791	42.5647	16.597
Chambal	Best Model	Gradient Boosting Regressor	Huber Regressor	Huber Regressor	Huber Regressor	Huber Regressor
	Mean Rainfall	122.38	360.430	299.89	189.056	27.203
	Mean ET0	178.96	131.529	114.104	122.303	141.3835
	Mean Runoff	16.116	111.289	114.79	72.680	11.233
	R^2	0.756	0.564	0.7236	0.535	0.216
	RMSE	8.499	40.685	40.436	40.811	10.872
Ken	Best Model	Gradient Boosting Regressor	Huber Regressor	Elastic Net	Huber Regressor	Huber Regressor
	Mean Rainfall	132.252	338.0736	344.157	189.869	35.4296
	Mean ET0	177.55	130.810	144.236	117.938	133.326
	Mean Runoff	20.91	98.237	119.315	65.863	11.799
	R^2	0.765	0.538	0.5309	0.547	0.501
	RMSE	9.927	42.018	40.7155	31.819	10.164
Sindh	Best Model	Huber Regressor	Light Gradient Boosting Machine	Extra Trees Regressor	Extreme Gradient Boosting	Random Forest Regressor
	Mean Rainfall	71.015	243.404	258.991	149.716	25.404
	Mean ET0	191.58	147.967	129.163	128.427	140.94
	Mean Runoff	5.705	49.955	83.343	56.44	11.928
	R^2	0.210	0.493	0.702	0.683	0.325

	RMSE	6.051	24.621	26.794	23.767	8.929
Son	Best Model	Huber Regressor	Least Angle Regression	Ridge Regression	Huber Regressor	Huber Regressor
	Mean Rainfall	82.703	188.211	193.52	107.626	32.114
	Mean ET0	156.217	136.373	129.76	130.139	135.59
	Mean Runoff	15.738	51.255	53.67	27.484	9.062
	R^2	0.4857	0.849	0.774	0.639	0.419
	RMSE	13.647	20.171	21.399	15.356	8.939
Tons	Best Model	Huber Regressor	Ridge Regression	Huber Regressor	Huber Regressor	Huber Regressor
	Mean Rainfall	74.929	158.439	151.080	107.040	32.033
	Mean ET0	160.26	140.5133	133.318	131.410	136.842
	Mean Runoff	16.8933	42.9169	43.726	33.673	10.255
	R^2	0.6403	0.777	0.756	0.7686	0.4181
	RMSE	10.643	21.745	24.0434	16.6526	9.6928

In the Chambal basin, Gradient Boosting Regressor and Huber Regressor were the top models. The mean rainfall ranged from 122.38 mm in June to 360.430 mm in July, and mean ET0 from 114.104 mm in August to 178.96 mm in June. Mean runoff varied between 11.233 mm in October and 114.79 mm in August, with R^2 highest in June (0.756) and lowest in October (0.216), and RMSE ranging from 8.499 in June to 40.811 in September.

The Ken basin saw the Gradient Boosting Regressor, Huber Regressor, and Elastic Net models perform best. Mean rainfall ranged from 132.252 mm in June to 344.157 mm in August, mean ET0 from 117.938 mm in September to 177.55 mm in June, and mean runoff from 11.799 mm in October to 119.315 mm in August. R^2 values were highest in June (0.765) and lowest in July (0.538), while RMSE ranged from 9.927 in June to 42.018 in July.

For the Sindh basin, models like Huber Regressor, Light Gradient Boosting Machine, Extra Trees Regressor, Extreme Gradient Boosting, and Random Forest Regressor were most effective. Mean rainfall varied from 71.015 mm in June to 258.991 mm in August, mean ET0 from 128.427 mm in September to 191.58 mm in June, and mean runoff from 5.705 mm in June to 83.343 mm in August. R^2 peaked in August (0.702) and was lowest in June (0.210), with RMSE values ranging from 6.051 in June to 26.794 in August.

In the Son basin, the top models included Huber Regressor, Least Angle Regression, and Ridge Regression. Mean rainfall ranged from 82.703 mm in June to 193.52 mm in August, mean ET0 from 129.76 mm in August to 156.217 mm in June, and mean runoff from 9.062 mm in October to 53.67 mm in August. R^2 was highest in July (0.849) and lowest in October (0.419), while RMSE ranged from 8.939 in October to 21.399 in August.

Lastly, the Tons basin's best models were Huber Regressor and Ridge Regression. Mean rainfall varied from 74.929 mm in June to 158.439 mm in July, mean ET0 from 131.410 mm in September to 160.26 mm in June, and mean runoff from 10.255 mm in October to 43.726 mm in August. R^2 values peaked in September (0.7686) and were lowest in October (0.4181), with RMSE ranging from 9.6928 in October to 24.0434 in August.

Overall, this study demonstrates the application of machine learning models in hydrological prediction, highlighting the variability in rainfall, evapotranspiration, and runoff across different river basins and months. The performance metrics (R^2 and RMSE) indicate the models' accuracy, which is essential for effective water resource management and planning in these regions.

5.4 Cluster Analysis for Zonation

Cluster analysis is a statistical technique that is used to group data points together based on their similarity. This technique can be used in hydrology to zonate watersheds, which is the process of dividing a watershed into smaller areas with similar hydrological characteristics. There are many different clustering algorithms available, but the most common ones are hierarchical clustering and k-means clustering. Hierarchical clustering starts by assigning each data point to its own cluster. Then, the algorithm repeatedly merges the two most similar clusters until there is only one cluster left. K-means clustering, on the other hand, starts by randomly assigning k clusters to the data points. Then, the algorithm iteratively assigns each data point to the cluster with the closest mean.

The result shows that the region can be identify with 3 clusters which is shown in Figure 5.10, and Figure 5.11

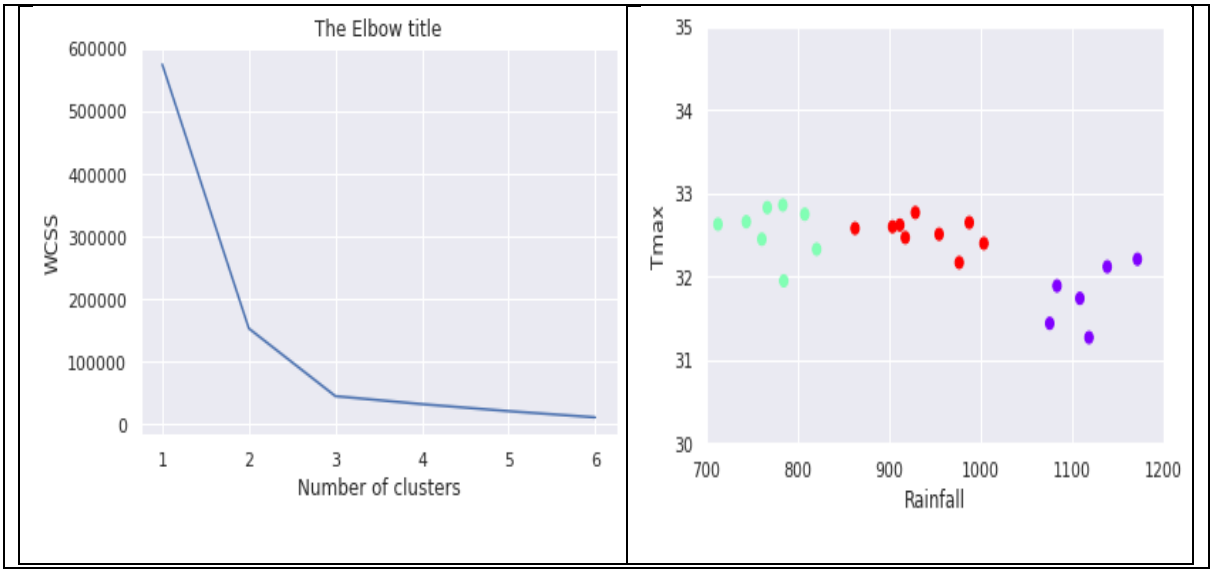


Figure 5. 10 From Elbow Chart of WCSS, the region can be identified with 3 clusters.

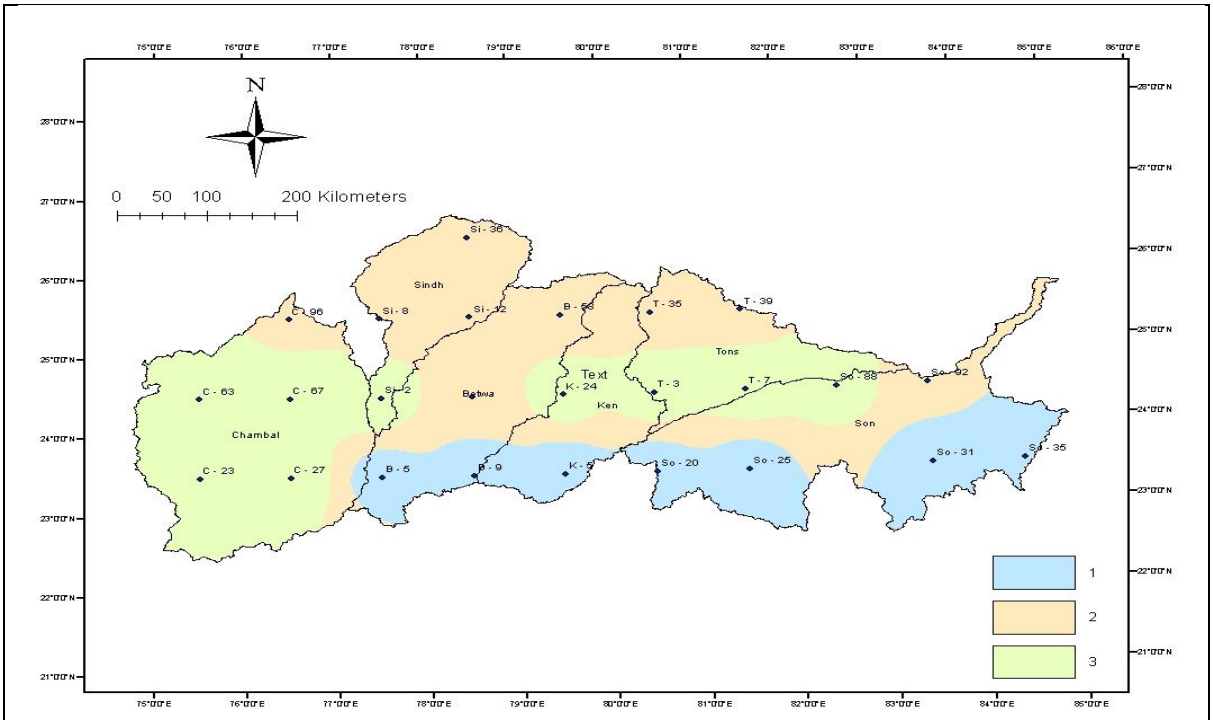


Figure 5. 11 Clusters in different basins.

The results of the cluster analysis revealed that the study area can be effectively divided into three distinct clusters. Each cluster represents regions with similar patterns of rainfall and temperature, which have a direct impact on water availability and hydrological processes.

*5.4.1 **Cluster 1:*** (High Rainfall, Moderate Temperature).

This cluster includes areas with relatively high levels of rainfall and moderate temperatures. These regions are likely to experience abundant water resources due to frequent and significant precipitation. The combination of high rainfall and moderate temperatures supports lush vegetation and contributes to the presence of rivers, lakes, and other water bodies. Such areas may be suitable for agricultural activities and ecosystem preservation. Figure 5.12 showing the R-R relation of Cluster-1 includes river basins such as Bah, Neemkheda, Rahatgarh, Bhadora, Ghatkhirwa, and a significant portion of the Ken river. Each scatter plot depicts observed discharge (in millimeters) against precipitation (in millimeters) for June, July, August, September, and October.

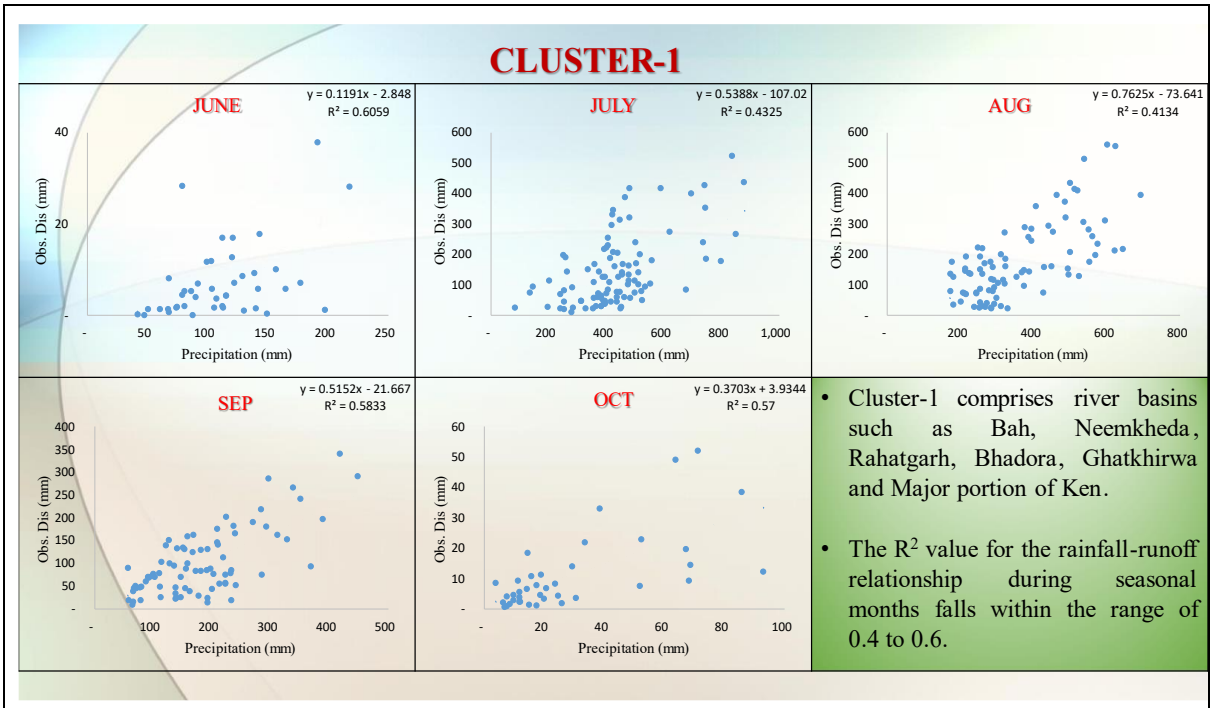


Figure 5. 12 R-R relation of Cluster-1.

In June, the graph shows a moderate positive correlation between precipitation and observed discharge, with an R^2 value of 0.5059. This indicates that approximately 50.59% of the variation in discharge can be explained by the amount of precipitation, suggesting a fairly strong relationship where increased rainfall typically leads to higher discharge levels.

The graph for July reveals a similar positive correlation, with an R^2 value of 0.4325. This means that 43.25% of the variation in discharge is due to precipitation, indicating that

while rainfall impacts discharge, other factors also influence the outcomes.

In August, the plot displays a moderate positive correlation with an R^2 value of 0.4134. This suggests that 41.34% of the variation in discharge can be attributed to precipitation, demonstrating a substantial relationship where increased rainfall generally results in higher discharge.

September's graph indicates a stronger positive correlation, with an R^2 value of 0.5833. This means that 58.33% of the variation in discharge is due to precipitation, highlighting a significant relationship between rainfall and runoff during this month.

The October graph shows a moderate correlation between precipitation and observed discharge, with an R^2 value of 0.57, meaning 57% of the variation in discharge can be explained by precipitation. This moderate relationship suggests that, by October, factors such as soil saturation, groundwater contributions, and reduced rainfall play a more substantial role in influencing discharge, slightly diminishing the direct impact of precipitation.

5.4.2 Cluster 2: (Moderate Rainfall, High Temperature)

comprises regions with moderate levels of rainfall but higher temperatures. These areas may face water stress during dry periods, as evaporation rates are likely to be higher than precipitation. Water resource management in this cluster should focus on efficient water use practices, water conservation, and sustainable agricultural practices to ensure water availability during periods of lower rainfall. The Figure 5.13 presents a comprehensive analysis of the rainfall-runoff relationships for Cluster-2, encompassing river basins such as Sindh, Mohanpur, Marisabad, Amarpur, Pandwan, and Kopra. The analysis is divided by months, with scatter plots for June, July, August, September, and October, each displaying the observed discharge (Obs. Dis) in millimeters against precipitation in millimeters.

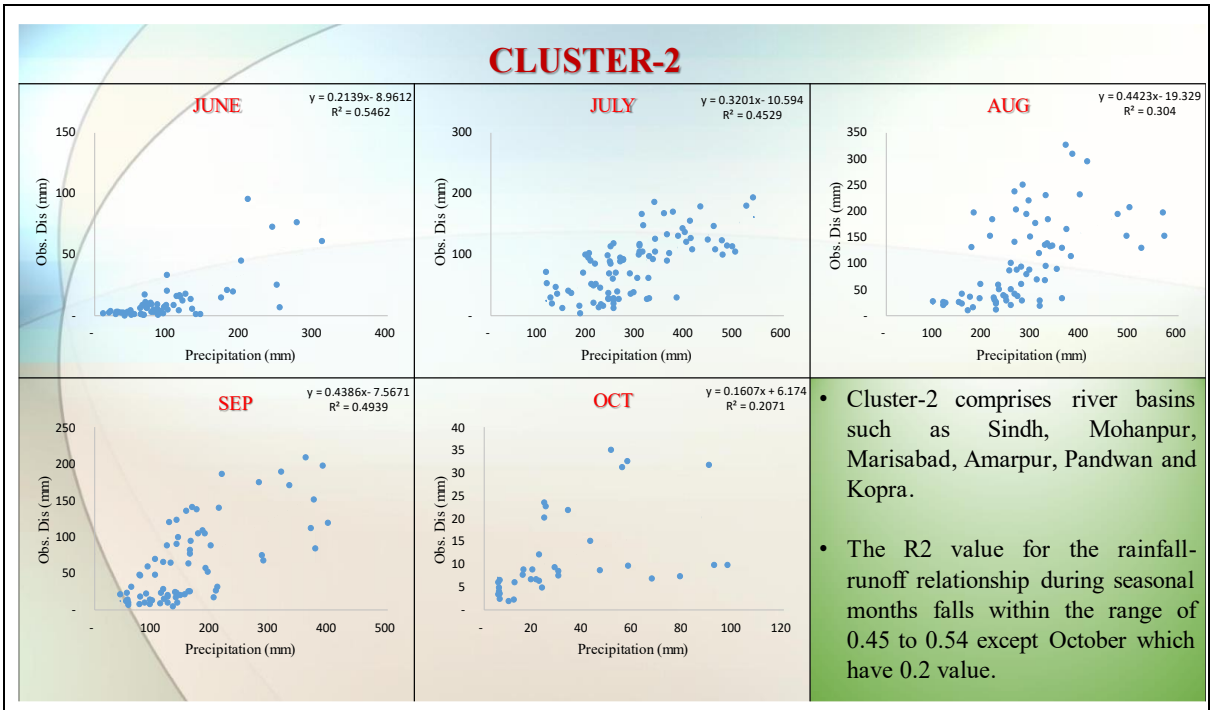


Figure 5. 13 R-R relation of Cluster-2.

For June, the scatter plot illustrates a moderate positive correlation between precipitation and observed discharge, with an R^2 value of 0.5059, suggesting that approximately 50.59% of the variability in discharge can be explained by the variability in precipitation.

In July, the relationship shows a weaker positive correlation with an R^2 value of 0.4325, indicating that 43.25% of the variability in discharge is explained by precipitation. This suggests a slightly less predictable runoff response to rainfall compared to June.

August's scatter plot also demonstrates a moderate correlation, with an R^2 value of 0.4134. This indicates that 41.34% of the variability in discharge can be attributed to changes in precipitation, reflecting a similar trend to July but with a marginally lower predictability.

September exhibits an improved correlation with an R^2 value of 0.5833, showing that 58.33% of the variability in discharge is due to precipitation. This higher R^2 value suggests a more consistent runoff response to rainfall during this month.

October presents the strongest correlation among the months, with an R^2 value of 0.57, meaning that 57% of the variability in observed discharge can be explained by precipitation. This indicates a relatively high level of predictability in the rainfall-runoff relationship for October.

*5.4.3 **Cluster 3:*** (Low Rainfall, Varying Temperature)

This cluster represents areas with comparatively low rainfall and varying temperature patterns. Water scarcity is a significant concern in these regions, particularly during prolonged dry spells. Effective water management strategies, such as rainwater harvesting and groundwater recharge, are vital to augment water supplies and support local communities' water needs. The Figure 5.14 presents a detailed analysis of the rainfall-runoff relationship for Cluster-3, which encompasses regions such as Tons, a significant portion of Chambal, and specific sites including Sinduria, Manakhedi, Maksudanganj, Madsour, Badnagar, Kuno, Samsakhedi, Choma, and Sanwar. The analysis is visualized through scatter plots for each month from June to October, illustrating the observed discharge in millimeters against precipitation in millimeters.

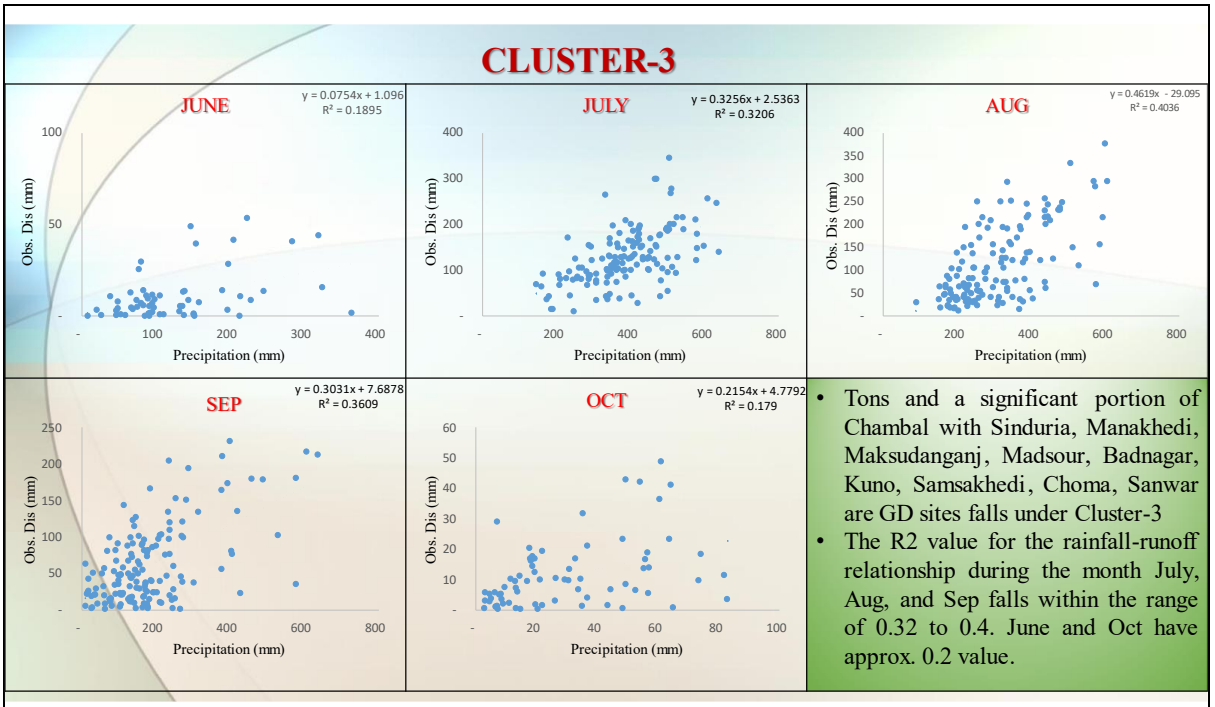


Figure 5. 14 R-R relation of Cluster-3.

In June, the scatter plot reveals a moderate correlation between precipitation and observed discharge. This is indicated by an R^2 value of 0.6059, suggesting that over half of the variability in discharge can be explained by precipitation levels during this month.

July's scatter plot shows a slight decrease in correlation with an R^2 value of 0.4325. This indicates that precipitation still explains a significant portion of the variability in discharge, though the relationship is not as strong as in June.

August continues this trend with an R^2 value of 0.4134, demonstrating a moderate correlation. The scatter plot suggests that while precipitation is an important factor in influencing discharge, other variables may also be playing a significant role during this month.

In September, the correlation strengthens again, with an R^2 value of 0.5833. The scatter plot for this month shows a clearer relationship between precipitation and discharge, indicating that precipitation levels are a more reliable predictor of discharge during this period.

October's scatter plot reveals the weakest correlation among the months analyzed, with an R^2 value of 0.57. Despite this, the relationship between precipitation and discharge remains notable, although it suggests increasing influence from other factors not accounted for by precipitation alone.

The zonation based on cluster analysis provides valuable insights into the spatial variations in climatic conditions and water availability across the study area. Understanding these hydrological zones is crucial for devising targeted water resource management plans, optimizing agricultural practices, and addressing potential water challenges in each cluster.

5.5 Efficiency of GR2M model and values of parameter

The Table 5.3 provides a comprehensive overview of the hydrological characteristics and model performance metrics for 33 gauging station sites across different river basins in India, including the Betwa, Chambal, Ken, Sindh, Son, and Tons basins. Each station is described by a set of physical attributes, including catchment area, mean elevation, and mean slope, as well as the GR2M model parameters (X1 and X2), and Nash-Sutcliffe efficiencies for both the calibration and validation periods.

Table 5. 3 Efficiency of GR2M model and values of parameter.

Name of the station		Basin	Area (Km2)	Mean Elevation(m)	Mean slope(%)	parameter		Efficiency			
Sl. No.	GD Sites					X1	X2	calibration		validation	
1	Bah	Betwa	938.58	471.34	2.1	6.00	0.9	1992-2007	**82.2**	2008-2013	**76.5**
2	Betwa	Betwa	2666.74	478.64	3.02	5.77	0.83	1990-2007	**88.9**	2008-2017	**74.4**
3	Mohanpur	Betwa	1409.05	463.36	2.36	5.16	0.84	1990-2001	**77.1**	2002-2008	**84.7**
4	Neemkheda	Betwa	1966.12	485.94	2.85	5.97	0.74	1990-2008	**79.7**	2009-2018	**61.2**
5	Rahatgarh	Betwa	1166.45	535.87	2.57	5.53	0.92	1990-2007	**87.3**	2008-2017	**88.5**
6	Sinduriya	Chambal	1402.21	492.01	1.78	4.51	0.54	1990-2009	**89.20**	2010-2019	**82.20**

7	Manakhedi	Chambal	3515.01	493.14	2.35	5.13	0.54	1990-2009	**87.00**	2010-2019	**87.10**
8	Maksudangar	Chambal	4853.72	470.39	2.43	5.80	0.67	1990-2009	**90.00**	2010-2019	**87.30**
9	Mandsour	Chambal	967.55	498.22	1.65	5.00	0.70	1995-2012	**92.50**	2013-2019	**89.20**
10	Badnagar	Chambal	870.01	534.94	1.69	5.89	0.35	1990-2009	**87.60**	2010-2019	**89.50**
11	Kuno	Chambal	4476.61	388.29	3.79	5.05	1.01	1990-2009	**94.10**	2010-2019	**94.80**
12	Samaskhedi	Chambal	1849.09	510.23	2.43	4.56	0.46	1990-2009	**77.90**	2010-2019	**87.90**
13	Choma	Chambal	1371.93	466.44	2.44	5.42	0.82	1990-2009	**84.60**	2010-2019	**76.60**
14	Sanwar	Chambal	410.37	543.28	1.71	3.70	0.59	1990-2009	**87.40**	2010-2019	**86.50**
15	Sonar	Ken	3875.50	444.768	2.55	4.87	0.61	1990-2009	**84.5**	2010-2019	**91.7**
16	Pandwan	Ken	16785.37	368.671	3.28	5.33	0.6	1990-2009	**80.2**	2010-2019	**89.2**
17	Kopra	Ken	867.40	396.027	2.09	2.47	0.42	1990-2009	**86.3**	2010-2019	**84.7**
18	Niwari	Ken	1305.79	516.245	2.56	0.11	0.54	1990-2009	**92.1**	2010-2019	**93**
19	Nohta	Ken	3422.74	428.169	3.05	3.27	0.43	1990-2009	**83.5**	2010-2019	**84.2**
20	Tigra	Ken	4188.80	406.203	4.4	3.76	0.54	1990-2009	**82.4**	2010-2019	**82.5**
21	Daboh	Sindh	1903.60	216.583	2.74	5.71	1.09	1990-2009	**87.40**	2010-2019	**98.00**
22	Didi	Sindh	6266.27	148.595	2.96	5.13	0.73	1990-2009	**75.40**	2010-2019	**81.50**
23	Amola	Sindh	4938.85	474.998	1.84	5.13	0.63	1990-2007	**78**	2008-2017	**84.3**
24	Goraghat	Sindh	11880.51	340.512	2.33	5.59	0.73	1990-2008	**75.6**	2009-2018	**90.2**
25	Budhara	Sindh	5300.22	242.113	3.6	4.81	0.77	1990-2009	**80.7**	2010-2019	**92.8**
26	Marsiabad	Son	10623.64	566.340	4.81	4.05	0.41	1990-2009	**93.1**	2010-2019	**86**
27	Amarpur	Son	395.20	414.663	3.4	3.71	0.37	1990-2009	**88.8**	2010-2019	**93.3**
28	Bhadora	Son	3139.26	446.977	4.57	4.28	0.42	1990-2009	**93.2**	2010-2019	**88**
29	Ghatkhirwa	Son	1012.76	426.652	3.44	2.2	0.36	1990-2009	**81.3**	2010-2019	**94.6**
30	Sansarpur	Tons	720.93	316.425	1.8	2.72	0.37	1990-2009	**87.9**	2010-2019	**76.6**
31	Patehera	Tons	8066.44	344.801	3.3	1.73	0.37	1990-2009	**84.3**	2010-2019	**71.1**
32	Khatkhari	Tons	527.38	364.424	1.99	2.01	0.47	1990-2009	**86.3**	2010-2019	**85.6**
33	Maihar	Tons	506.87	447.386	5.64	2.83	0.38	1990-2009	**91.6**	2010-2019	**79.8**

For each station, crucial geographical attributes are provided:

1. Area (km^2): This parameter gives an indication of the catchment size for each station, ranging from relatively small areas like 260.93 km^2 for Sansarpur in the Tons basin to large catchments like 16785.37 km^2 for Pandwan in the Ken basin. The variation in catchment sizes allows for the assessment of the model's performance across different scales.

2. Mean Elevation (m): This data point provides insight into the topographical characteristics of each station's location. Elevations range from as low as 148.595 m for Didi in the Sindh basin to as high as 543.28 m for Sanwar in the Chambal basin. The elevation data is crucial for understanding the potential impacts of topography on hydrological processes.

3. Mean Slope (%): The mean slope offers information about the terrain's steepness in each catchment area. Values range from relatively flat terrains with slopes of 1.65% (Mandsour, Chambal basin) to steeper areas with slopes up to 5.64% (Maihar, Tons basin). This parameter is significant as it influences runoff characteristics and other hydrological processes.

The table includes two key parameters of the GR2M model: X1 and X2. These parameters are essential for understanding how the model has been calibrated for each specific location. X1, this parameter typically represents the maximum capacity of the production store in the GR2M model. Values in the table range from 1.73 (Patehera, Tons basin) to 6.57 (Neemkheda, Betwa basin). X2, This parameter often relates to the groundwater exchange coefficient in the GR2M model. The values in the table vary from 0.35 (Badnagar, Chambal basin) to 1.09 (Daboh, Sindh basin). The variation in these parameters across different stations indicates the model's flexibility in adapting to diverse hydrological conditions.

One of the most critical aspects of the table is the presentation of the model's efficiency during two distinct periods. Calibration Period, typically spanning from 1990 to 2009 for most stations, represents the timeframe during which the model parameters were adjusted to best fit the observed data. Some stations have shorter calibration periods, such as Mohanpur (1990-2001) in the Betwa basin. Validation Period, generally covering 2010 to 2019, this period is used to test the calibrated model's performance on a separate dataset. It's crucial for assessing the model's predictive capabilities and long-term reliability.

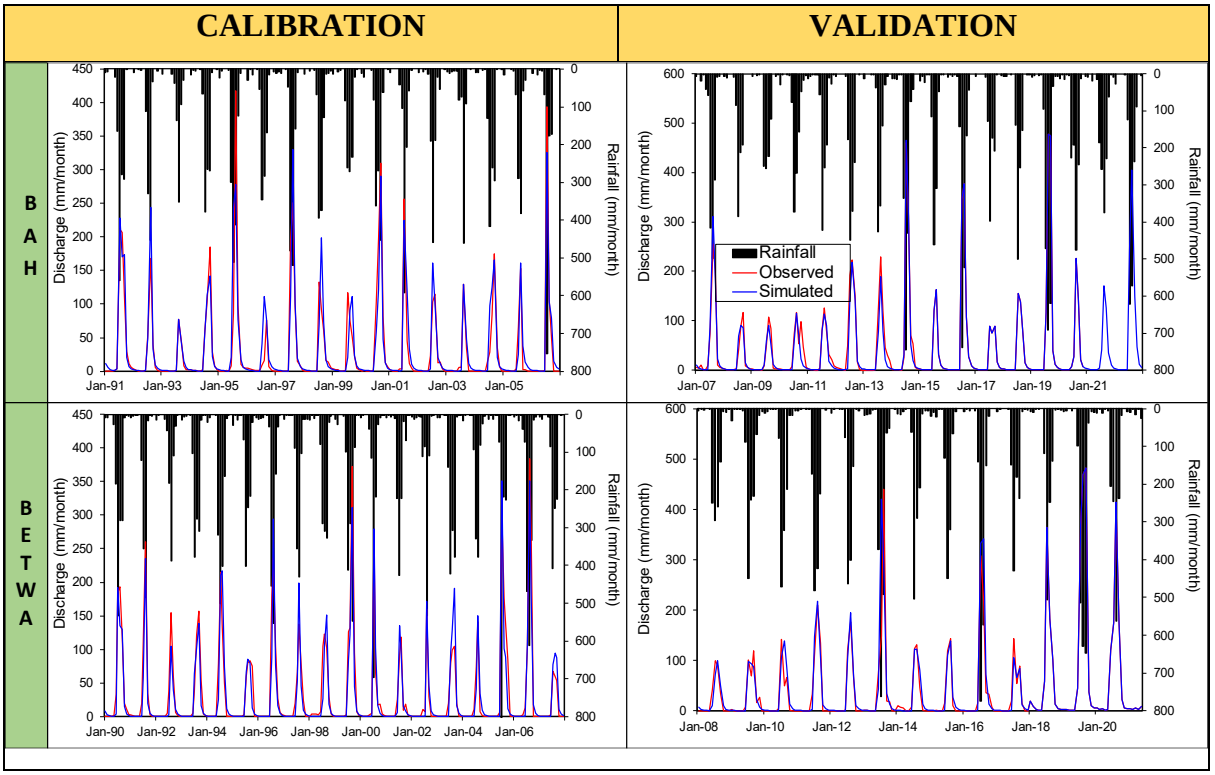
The efficiency of the GR2M model is presented as percentages for both the calibration and validation periods. This allows for a direct comparison of the model's performance across different temporal scales and geographical locations. Calibration efficiency, values range from 75.4% (Didi, Sindh basin) to 94.10% (Kuno, Chambal basin). The majority of stations show calibration efficiencies above 80%, indicating generally good model performance during the fitting phase. Validation Efficiency, these values range from 61.2% (Neemkheda, Betwa basin) to 98.00% (Daboh, Sindh basin). The variation in validation efficiencies provides crucial

information on the model's ability to maintain performance when applied to new data.

The stations are grouped by major river basins, allowing for regional comparisons and analysis:

1. **Betwa Basin**: Five stations are listed, showing generally high calibration efficiencies (77.1% to 88.9%) but more variable validation efficiencies (61.2% to 88.5%). The lower validation efficiency for Neemkheda (61.2%) might indicate challenges in long-term predictions for this particular location.

In both calibration and validation (Figure 5.15) phases, rainfall is represented by black bars on the right y-axis in millimeters per month, exhibiting distinct seasonal patterns with high rainfall during monsoon months and minimal precipitation during the dry season. Discharge, plotted on the left y-axis in millimeters per month, reflects surface runoff trends, with its peaks closely following rainfall events. In the calibration phase, the simulated and observed discharge values generally align well for different locations, indicating that the model effectively captured the rainfall-discharge relationship. However, minor deviations suggest potential variability in parameters or limitations in the input data.



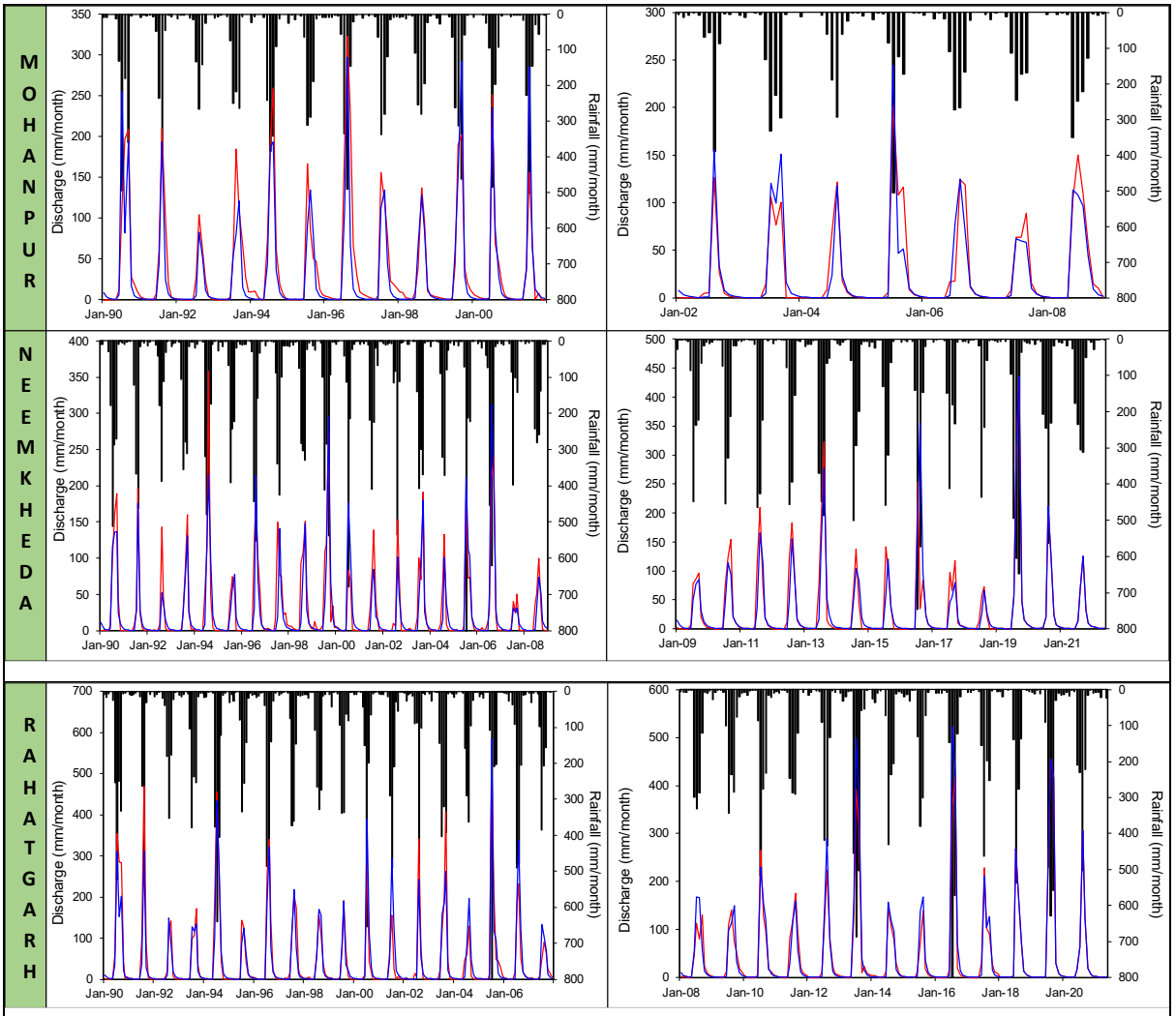
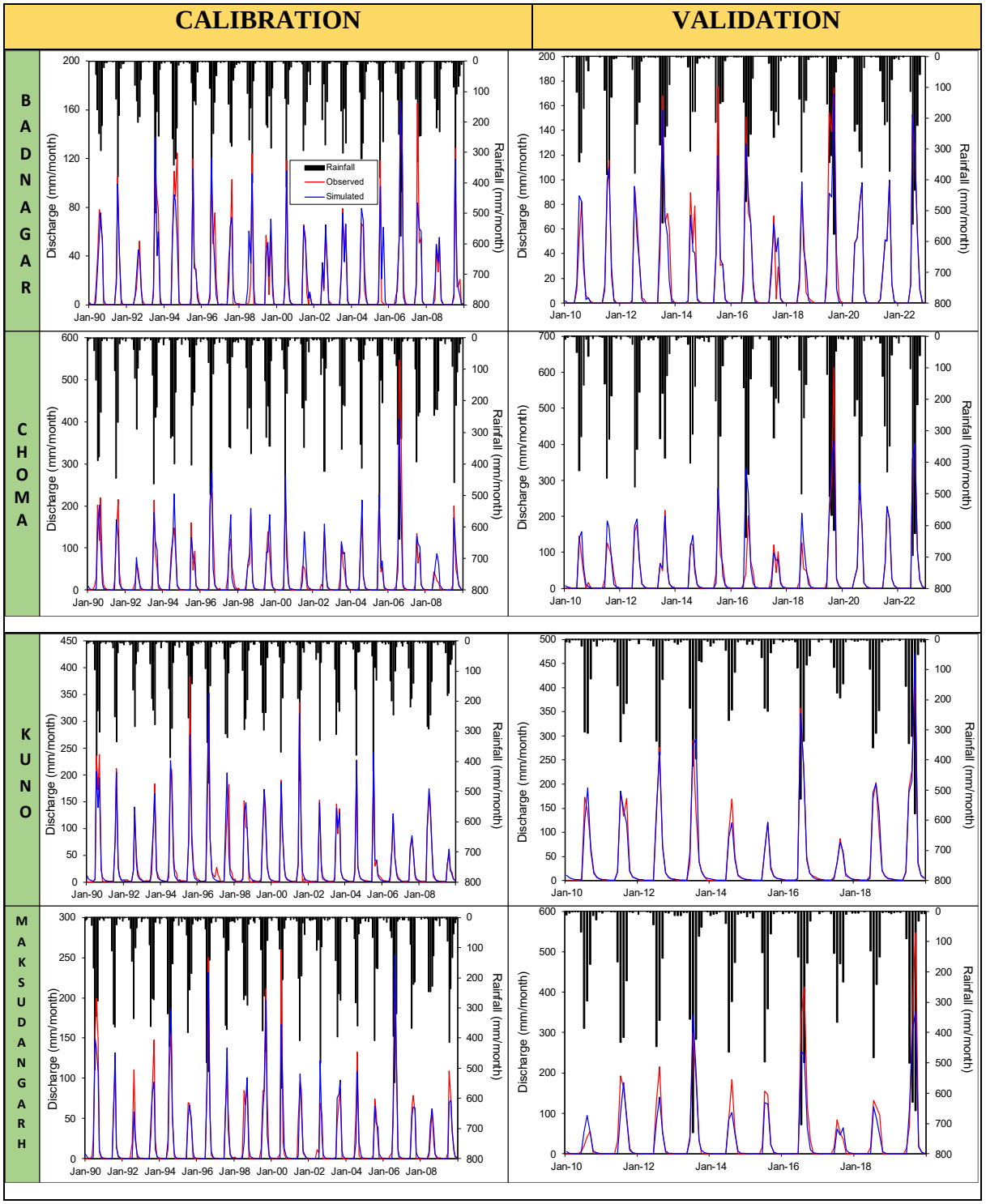


Figure 5. 15 Graph of calibration and validation of different G/D sites of Betwa basin.

The graphs are significant as they demonstrate the model's ability to simulate the hydrological behavior of different locations, by comparing observed and simulated discharge against rainfall patterns. They highlight the model's reliability during calibration and validation, ensuring its applicability for water resource management and prediction in similar basins.

2. **Chambal Basin:** With nine stations, this basin shows consistently high efficiencies in both calibration (84.60% to 94.10%) and validation (76.60% to 94.80%) periods. The high performance across both periods suggests the GR2M model is particularly well-suited for this basin's hydrological characteristics. Figure 5.16 showing the calibration and validation graph of different G/D sites of Chambal basin.


CALIBRATION
VALIDATION
BADNAGAR
CHOMA
KUNO
MAKSUDANGARH
Discharge (mm/month)
Rainfall (mm/month)
Rainfall
Observed
Simulated
Jan-90 Jan-92 Jan-94 Jan-96 Jan-98 Jan-00 Jan-02 Jan-04 Jan-06 Jan-08
Jan-10 Jan-12 Jan-14 Jan-16 Jan-18 Jan-20 Jan-22

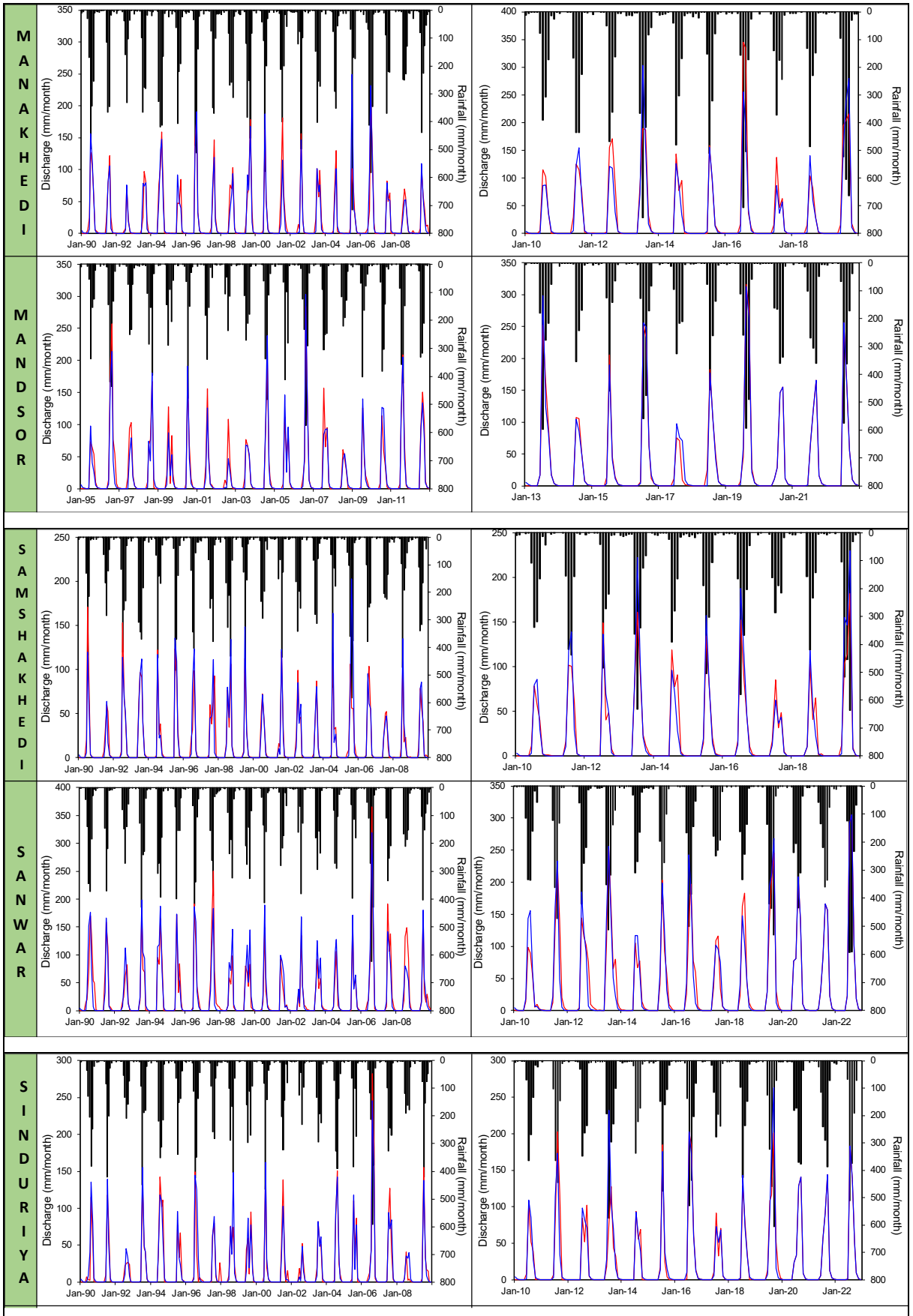
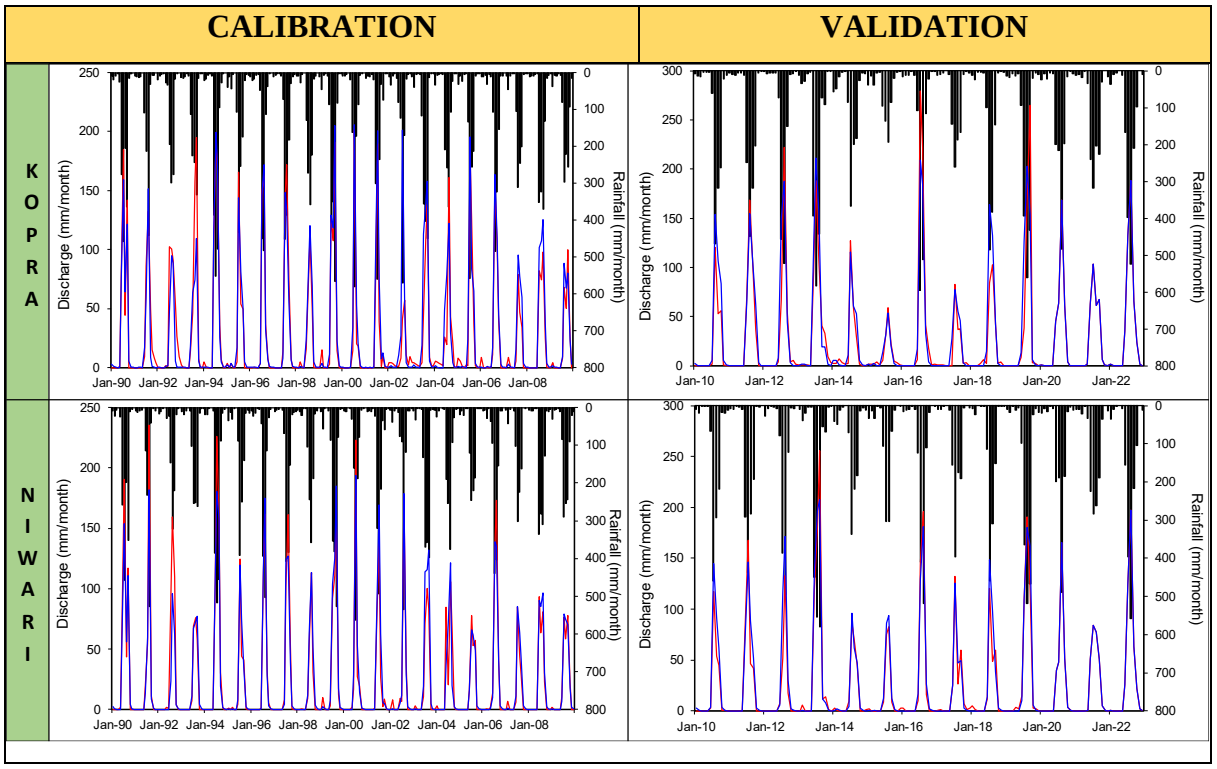


Figure 5. 16 Graph of calibration and validation of different G/D sites of Chambal Basin.

3. Ken Basin: The six stations in this basin demonstrate good model performance, with calibration efficiencies ranging from 80.2% to 92.1% and validation efficiencies from 82.5% to 93%. The consistent performance across periods indicates the model's robustness in this region. Figure 5.17 showing the calibration and validation graph of different G/D sites of Ken basin.

During the validation phase, the graphs reveal the model's ability to replicate discharge trends under new conditions. For Kopra, the alignment between observed and simulated discharge remains consistent, with minor discrepancies during high-flow events. Similarly, for Niwari, the model captures the seasonal dynamics well, though slightly higher variability is observed, potentially due to external factors such as changes in land use, catchment characteristics, or data uncertainty.



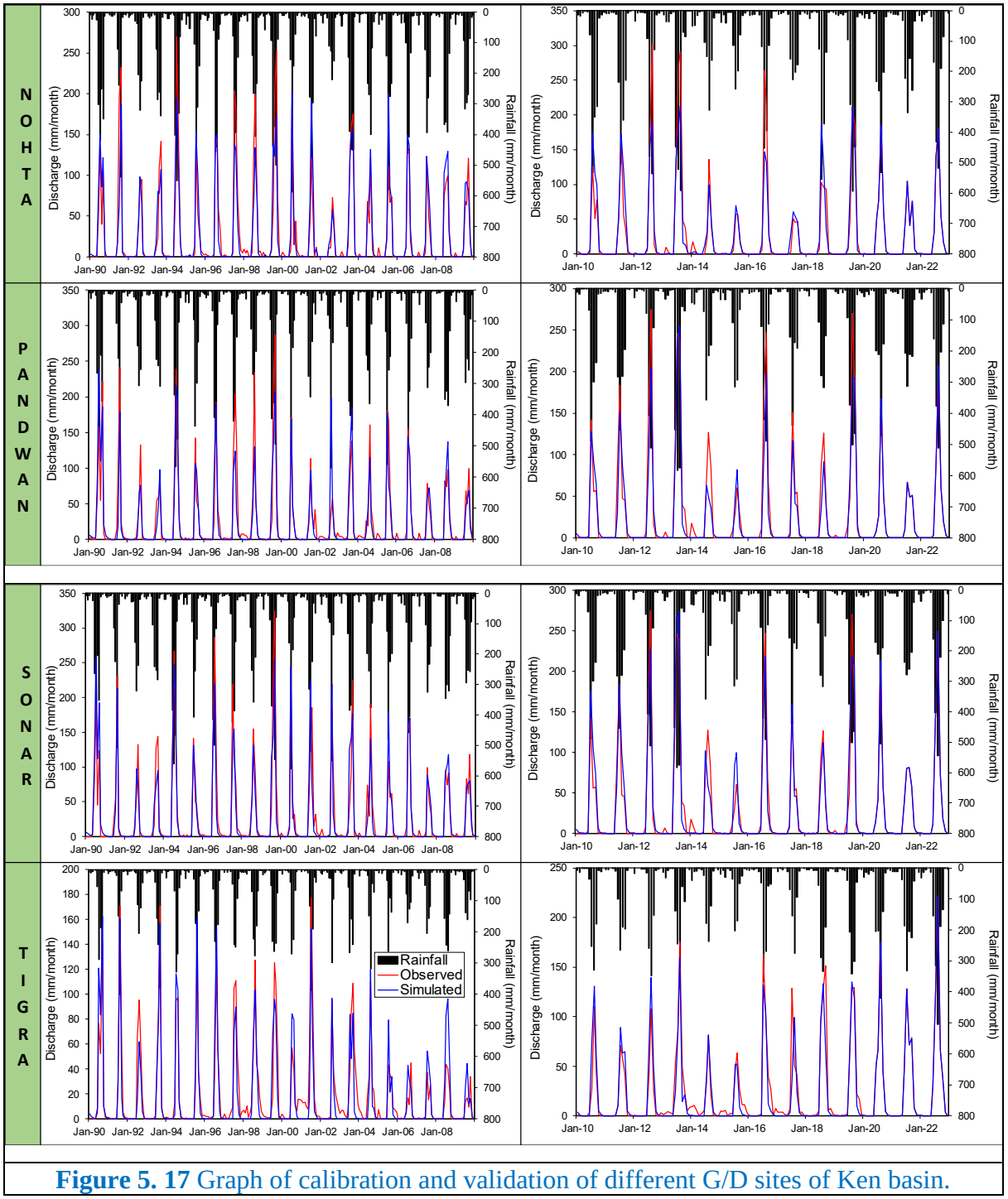
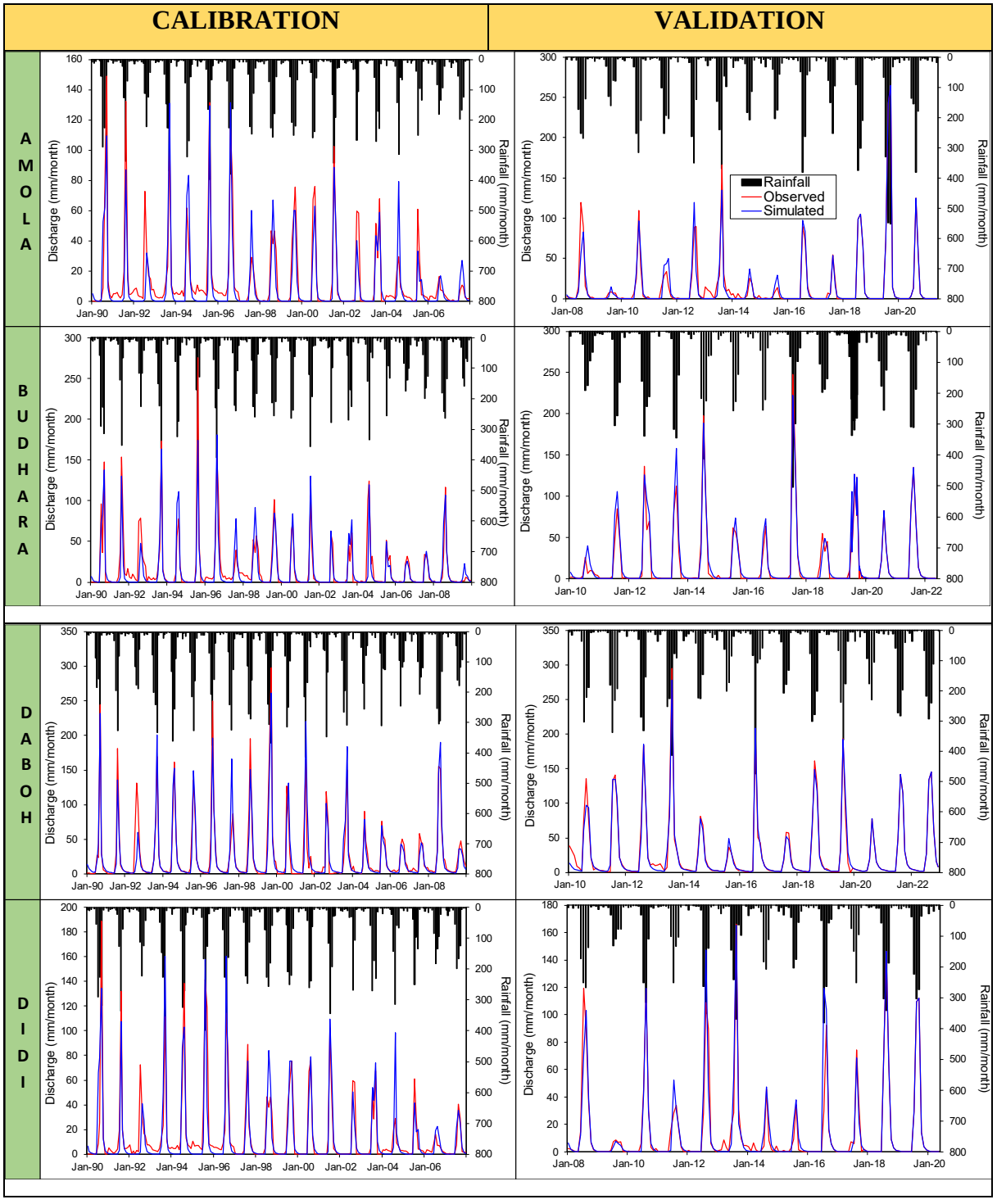


Figure 5. 17 Graph of calibration and validation of different G/D sites of Ken basin.

Overall, the calibration results demonstrate the model's ability to reproduce historical hydrological behaviour, while the validation results highlight its predictive reliability. Despite slight inconsistencies, the model proves to be a valuable tool for understanding rainfall-discharge relationships and assessing water availability. Further refinement, especially for extreme events or site-specific anomalies, could enhance its applicability in ungauged basins.

4. Sindh Basin: Five stations are listed, showing a wider range of efficiencies. Calibration efficiencies vary from 75.4% to 87.40%, while validation efficiencies range from 81.50% to 98.00%. The notable improvement in validation efficiency for some stations (e.g., Daboh and Goraghat) is particularly interesting and may warrant further investigation. Figure 5.18 showing the calibration and validation graph of different G/D sites of Sindh basin.



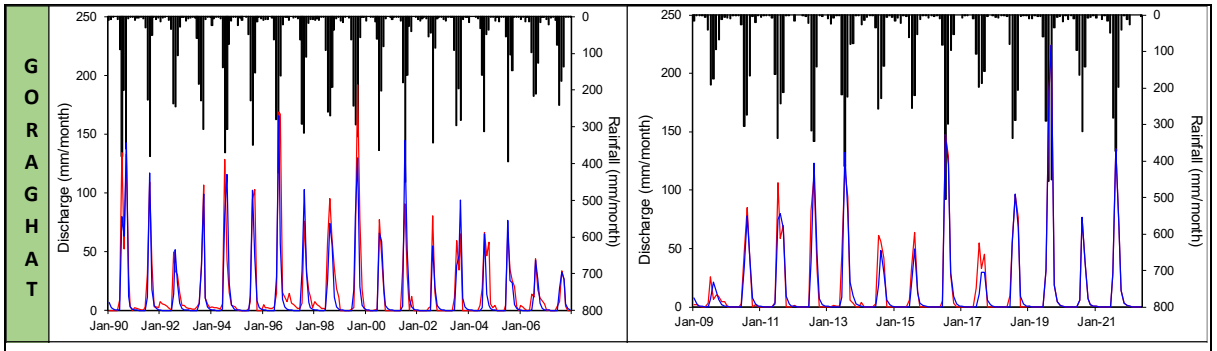
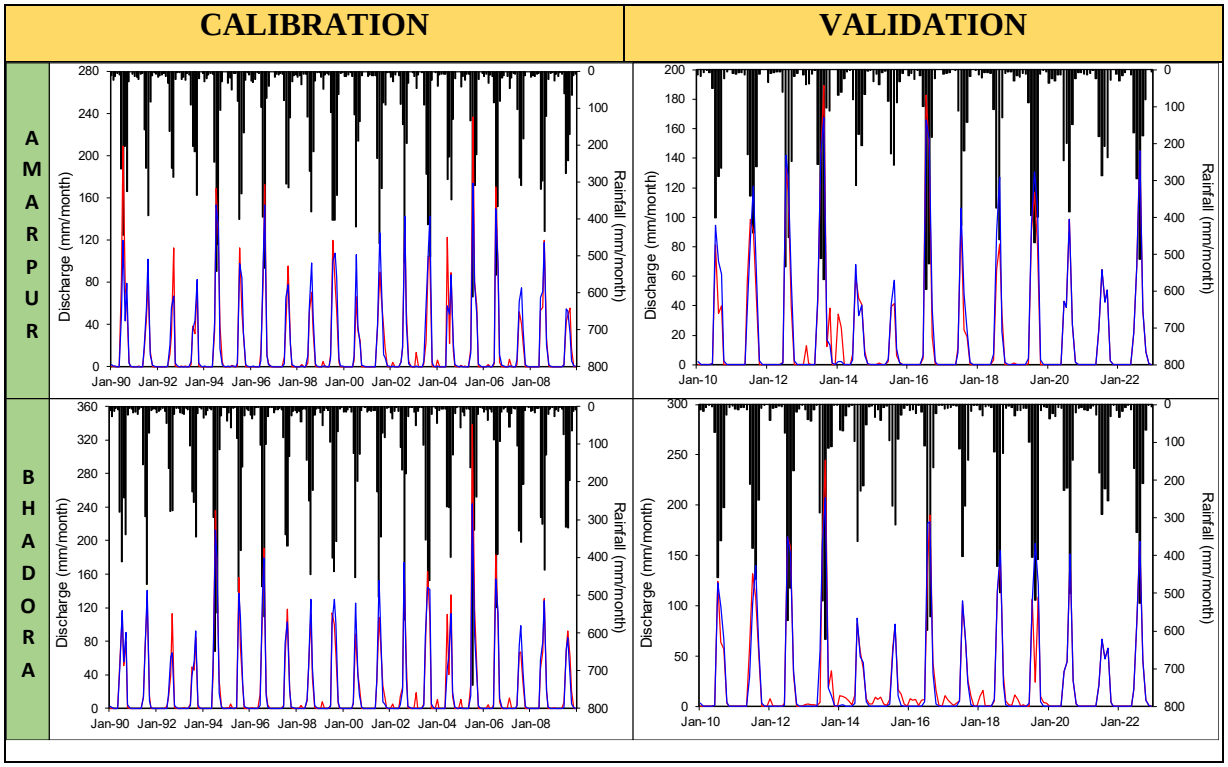


Figure 5. 18 Graph of calibration and validation of different G/D sites of Sindh basin.

5. Son Basin: The four stations in this basin show high efficiencies in both periods, with calibration ranging from 81.3% to 93.2% and validation from 86% to 94.6%. This consistent performance suggests the model is well-suited to the Son basin's hydrological regime. Figure 5.19 showing the calibration and validation graph of different G/D sites of Son basin.



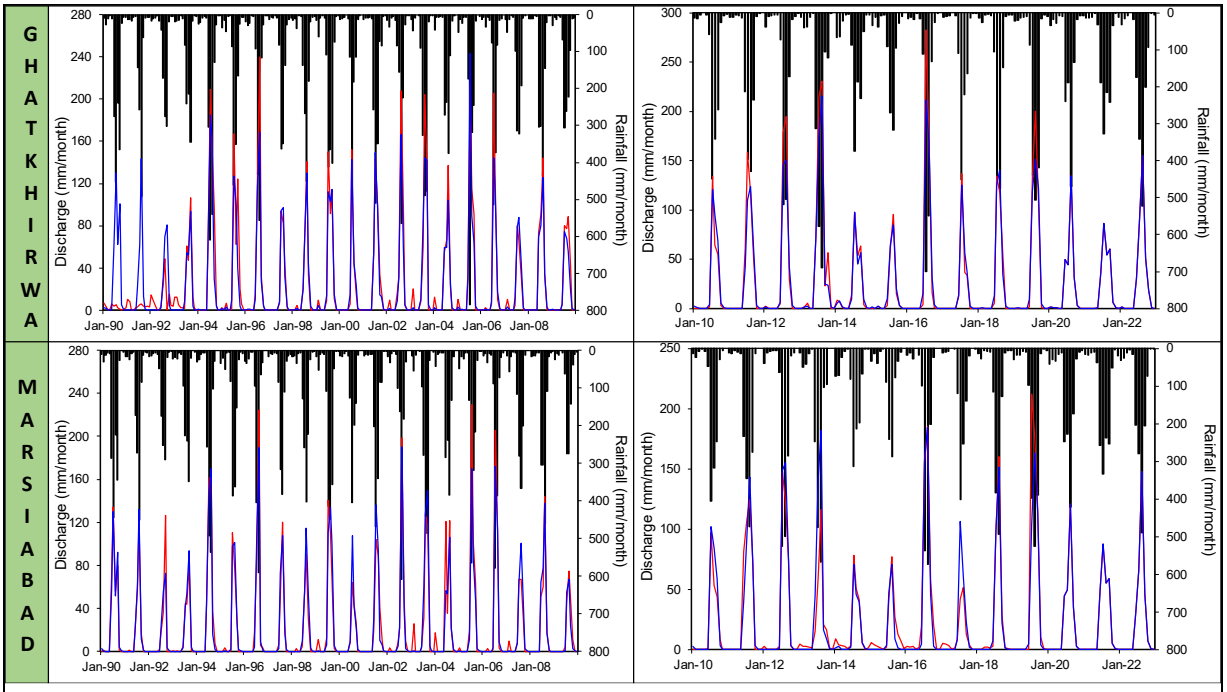
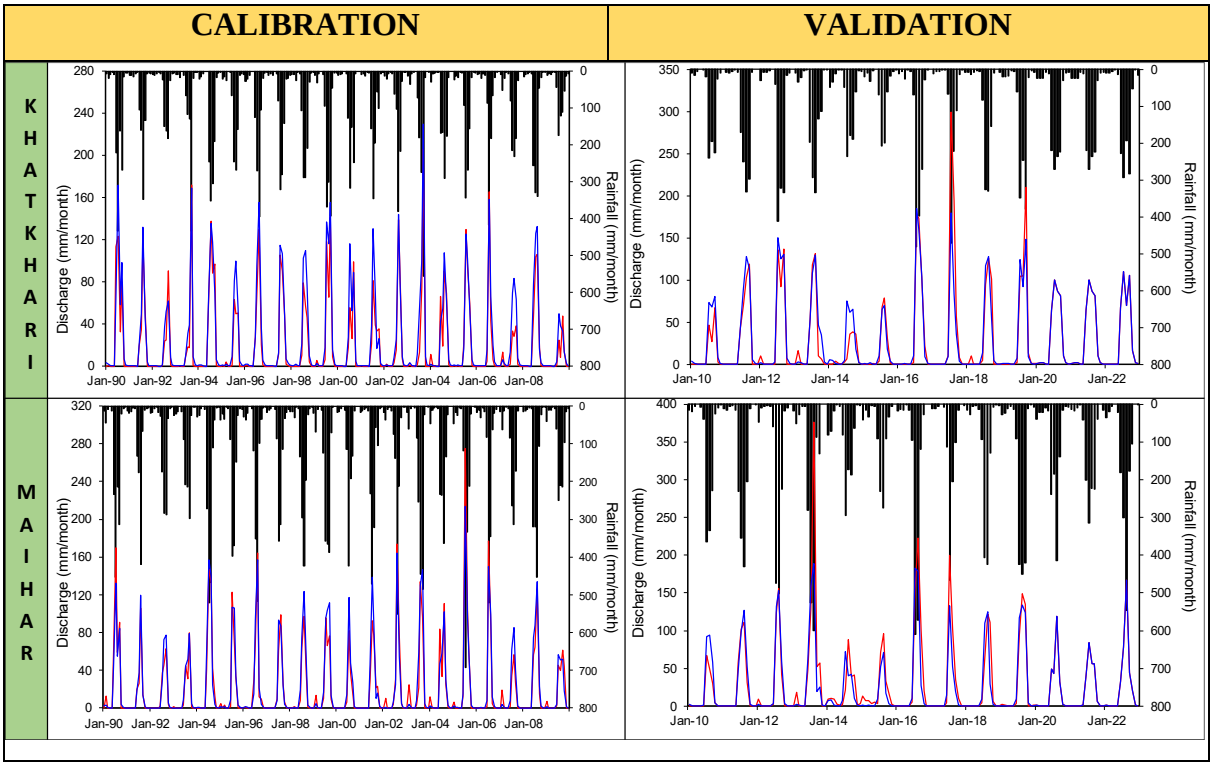


Figure 5. 19 Graph of calibration and validation of different G/D sites of Son Basin.

6. Tons Basin: The four stations in this basin show good calibration efficiencies (84.3% to 91.6%) but more variable validation efficiencies (71.1% to 85.6%). The lower validation efficiency for Patehera (71.1%) might indicate some challenges in long-term predictions for this location. Figure 5.20 showing the calibration and validation graph of different G/D sites of Tons basin.



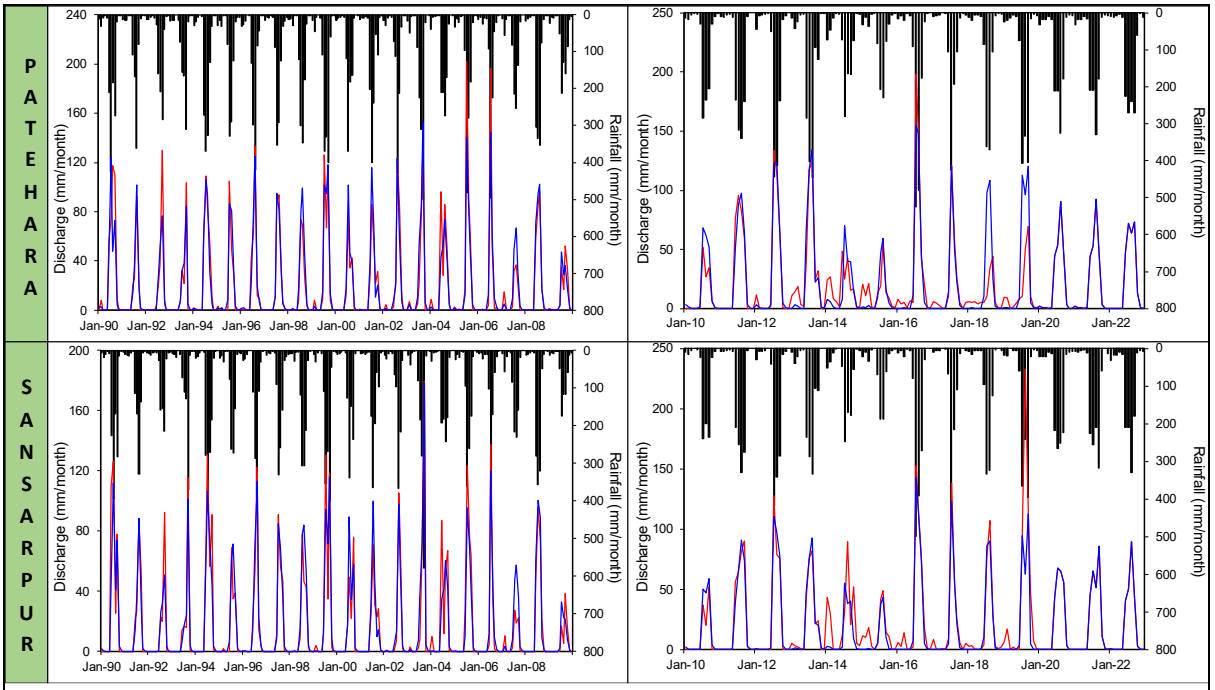


Figure 5. 20 Graph of calibration and validation of different G/D sites of Tons Basin.

The efficiency values provide a clear indication of the GR2M model's performance across diverse geographical and hydrological settings. The generally high efficiencies suggest that the model is well-suited for Indian river basins, though there are variations that warrant further investigation.

5.6 Regionalization using Inverse Distance Weighting (IDW)

The inverse distance weighting (IDW) method was employed to estimate the parameters of the GR2M model for various ungauged sites. The parameters were then compared with observed runoff, yielding efficiencies that ranged from 58% (Sinduriya) to 93.4% (Mandsour), as detailed in the accompanying Table 5.4.

For the Betwa basin, the efficiencies varied from 75.9% at Neemkheda to 87% at Bah. In the Chambal basin, the efficiencies ranged from 58% at Sinduriya to 86% at Sanwar. The Ken basin showed efficiencies between 69.7% at Pandwan and 88.1% at Niwari. The Sindh basin had efficiencies spanning from 52.2% at Budhara to 74.4% at Amola. For the Son basin, the efficiencies ranged from 72.2% at Marsiabad to 88.8% at Amarpur. Lastly, the Tons basin efficiencies were between 70.2% at Sansarpur and 82.5% at Patehra.

The at-site efficiency for GR2M model for all the stations were computed using calibrated parameters for the period of 1991-2021 and compared this at-site model efficiency

with regional efficiency for all the stations and given in Table 5.4. The regional approach for GR2M model for sub-basins in Ganga River System found reasonably appropriate results and has promising capability in application for ungauged catchments. The study suggested that the GR2M model, along with the IDW interpolation technique, has the capability to effectively transfer information to ungauged catchments, thereby enhancing water resource planning and management in regions lacking direct measurement data. This method supports informed decision-making in water resource management, which is crucial in regions lacking extensive hydrological data. Figure 5.21 showing the Comparison of At-site and Regional Efficiency Across the G/D Sites.

Table 5. 4 GR2M parameter efficiency using IDW method.

Sl. No.	GD Sites	Basin	At-site parameters		At-site efficiency	Regional parameters		Regional Efficiency
			X1	X2		X1	X2	
1	Bah	Betwa	6.00	0.9	87.7	5.77	0.82	87.0
2	Betwa	Betwa	5.77	0.83	84.6	5.95	0.83	84.1
3	Mohanpur	Betwa	5.16	0.84	70.9	5.78	0.87	63.8
4	Neemkheda	Betwa	5.97	0.74	81.8	5.8	0.85	75.9
5	Rahatgarh	Betwa	5.53	0.92	79.6	5.81	0.83	81.2
6	Sinduriya	Chambal	4.51	0.54	76.4	4.22	0.61	58.0
7	Manakhedi	Chambal	5.13	0.54	79.3	5.41	0.65	76.9
8	Maksudangar	Chambal	5.80	0.67	90.1	4.98	0.57	83.4
9	Mandsour	Chambal	5.00	0.70	95.0	4.32	0.55	93.4
10	Badnagar	Chambal	5.89	0.35	63.2	4.29	0.6	64.7
11	Kuno	Chambal	5.05	1.01	94.5	4.79	0.61	72.7
12	Samaskhedi	Chambal	4.56	0.46	83.0	4.57	0.66	73.9
13	Choma	Chambal	5.42	0.82	77.9	4.38	0.53	73.8
14	Sanwar	Chambal	3.70	0.59	87.3	3.52	0.49	86.1
15	Sonar	Ken	4.87	0.61	84.5	2.34	0.43	83.7
16	Pandwan	Ken	5.33	0.6	80.1	3.67	0.54	69.7
17	Kopra	Ken	2.47	0.42	86.6	4.49	0.6	75.0
18	Niwari	Ken	0.11	0.54	93.6	3.83	0.52	88.1
19	Nohta	Ken	3.27	0.43	83.5	3.11	0.51	73.9
20	Tigra	Ken	3.76	0.54	82.2	5.07	0.59	72.2
21	Daboh	Sindh	5.71	1.09	63.5	5.37	0.72	66.0
22	Didi	Sindh	5.13	0.73	77.7	5.06	0.81	66.9
23	Amola	Sindh	5.13	0.63	80.5	5.49	0.82	74.4
24	Goraghat	Sindh	5.59	0.73	60.0	5.4	0.91	61.8
25	Budhara	Sindh	4.81	0.77	26.7	5.27	0.77	52.2
26	Marsiabad	Son	4.05	0.41	73.0	3.67	0.38	72.2
27	Amarpur	Son	3.71	0.37	91.3	3.87	0.41	88.8
28	Bhadora	Son	4.28	0.42	82.2	2.82	0.37	81.5
29	Ghatkhirwa	Son	2.2	0.36	86.4	4.19	0.41	86.1
30	Sansarpur	Tons	2.72	0.37	79.9	1.85	0.4	70.2

31	Patehera	**Tons**	1.73	0.37	83.2	2.6	0.39	82.5
32	Khatkhari	**Tons**	2.01	0.47	79.4	2.39	0.37	82.2
33	Maihar	**Tons**	2.83	0.38	89.1	2.22	0.39	85.0

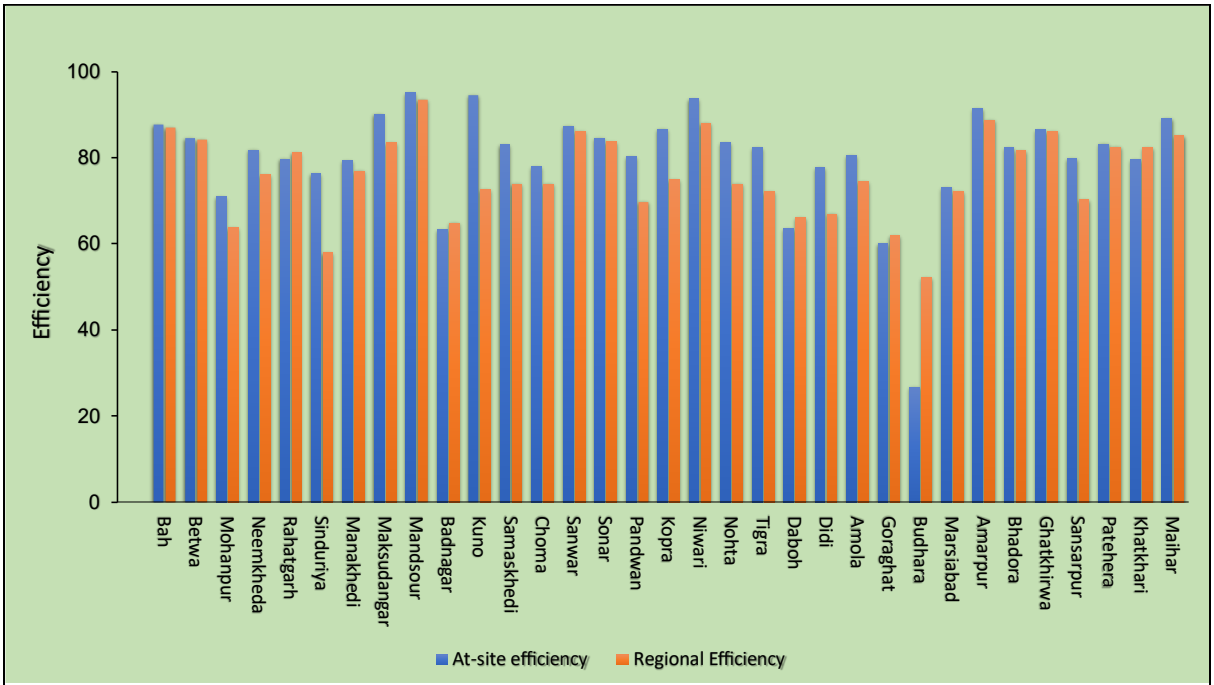


Figure 5. 21 Comparison of At-site and Regional Efficiency Across the G/D Sites.

5.6.1 Betwa Basin IDW Regionalization

In the Betwa Basin, the parameter X1 derived from the IDW equation ranges between 5.77 and 5.95, while X2 varies from 0.82 to 0.87. These parameters represent site-specific hydrological characteristics and play a significant role in runoff estimation. The corresponding efficiency percentages for the basin range from 63.8% at Mohanpur to 87.0% at Bah, indicating a performance level that varies from moderate to high across the GD sites.

The simulated runoff for the basin was calculated using the GR2M model. The strong correlation between observed and simulated runoff, reflected in the efficiency values, demonstrates the reliability of the IDW-derived parameters and the GR2M model in capturing the hydrological behavior of the Betwa Basin. This highlights the applicability of such methodologies in assessing runoff in diverse hydrological conditions across different sites within the basin. For Bah, the highest observed runoff occurred in 2007 with a peak discharge of approximately 500 mm/month, while the simulated peak also aligns closely, reflecting the model's accuracy in capturing extreme events. Similarly, for Betwa, the highest observed runoff was recorded in 2011, peaking at around 450 mm/month, with simulated peaks closely

mirroring the observed data. Figure 5.22 showing the graph of observed vs. simulated monthly discharge at different G/D sites of the Betwa basin.

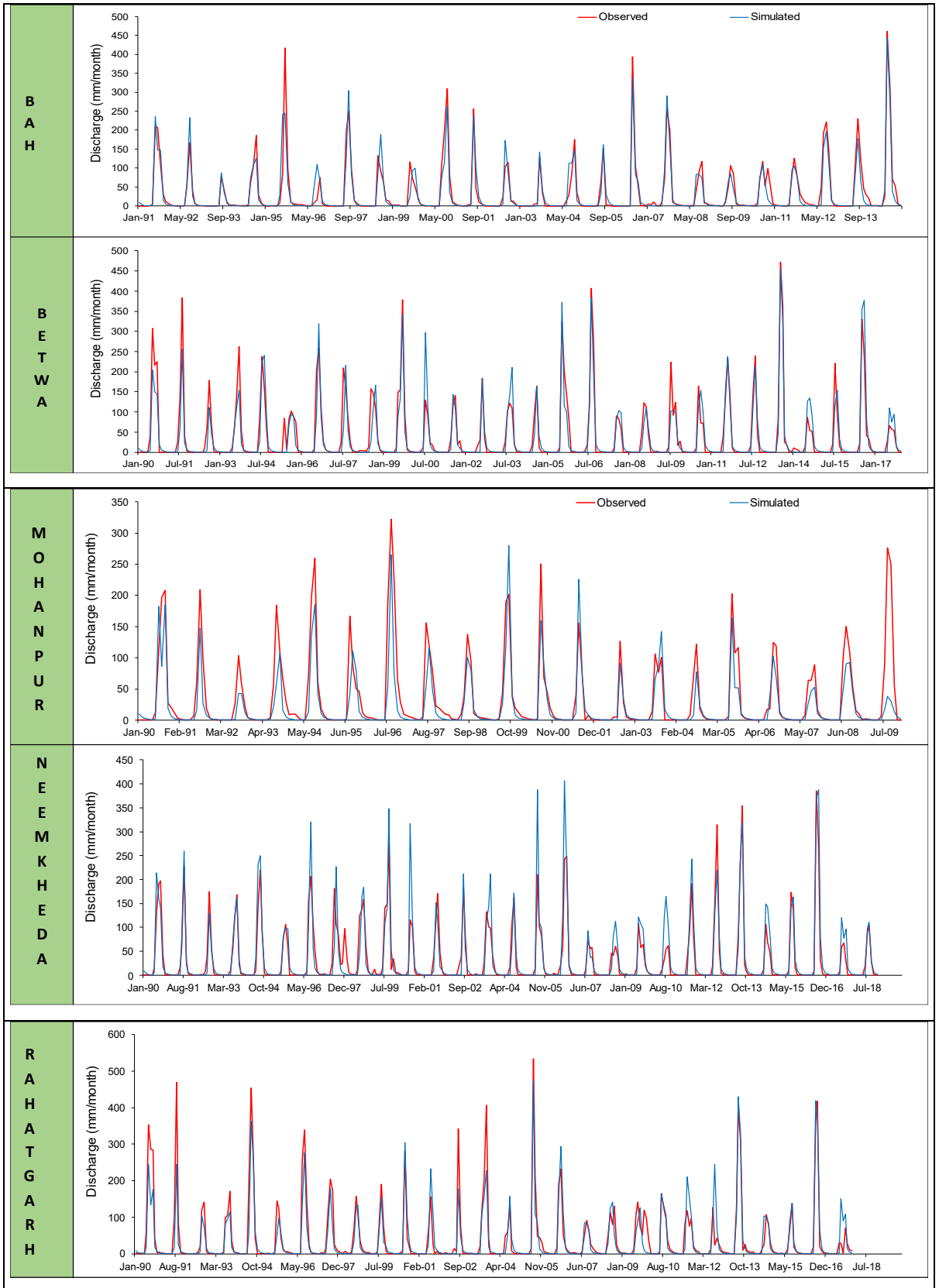
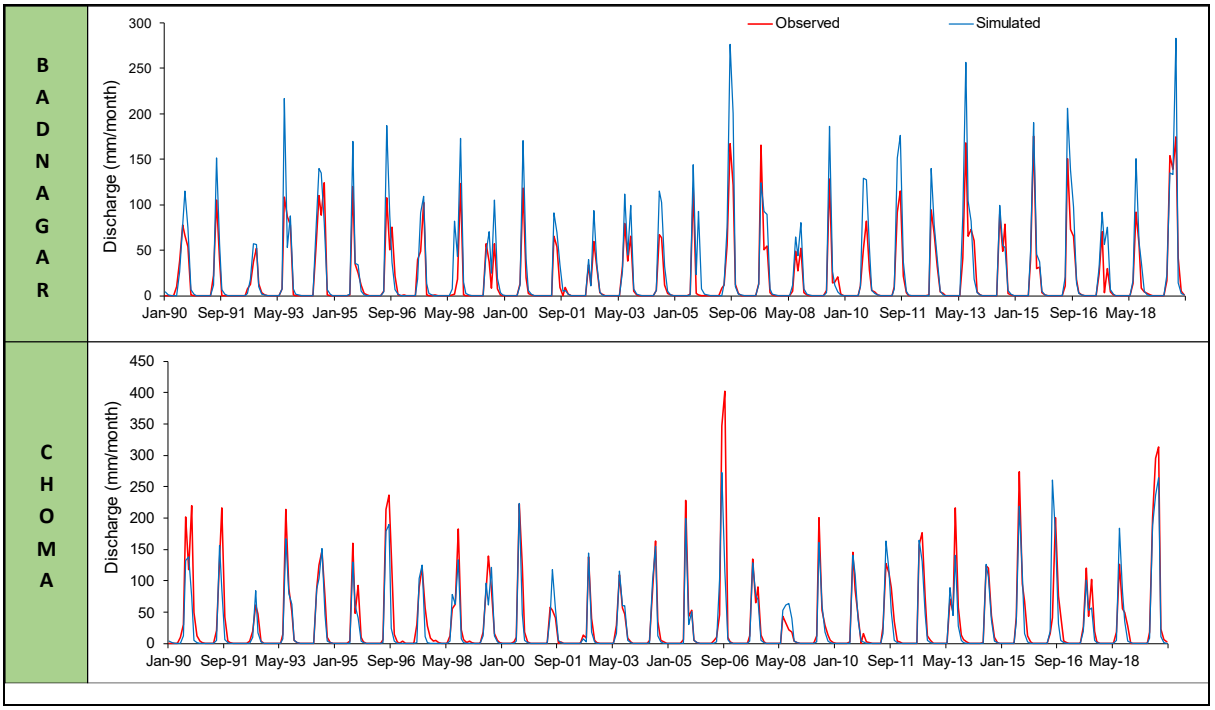


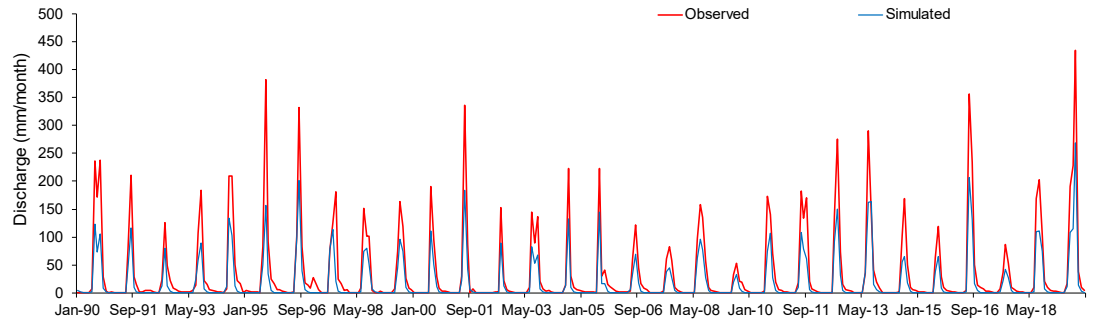
Figure 5. 22 Observed vs. Simulated Discharge at Different G/D Sites of the Betwa Basin.

5.6.2 Chambal Basin IDW Regionalization

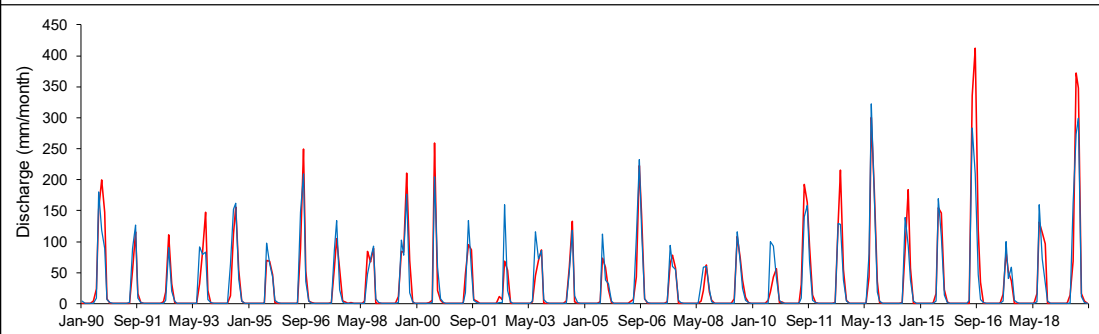
The Chambal Basin exhibits greater variability in its parameters. X1 ranges from 3.52 (Sanwar) to 5.41 (Manakhedi), and X2 spans from 0.49 (Sanwar) to 0.66 (Samaskhedi). Efficiency values in the Chambal Basin range widely from 58.0% (Sinduriya) to 93.4% (Mandsour), suggesting significant disparities in site performance. Figure 5.23 showing the graph of observed vs. simulated monthly discharge at different G/D sites of the Chambal basin.



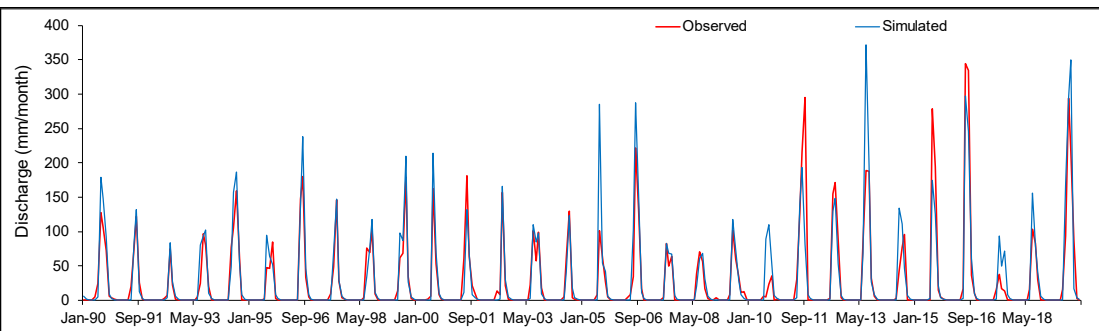
KUNO



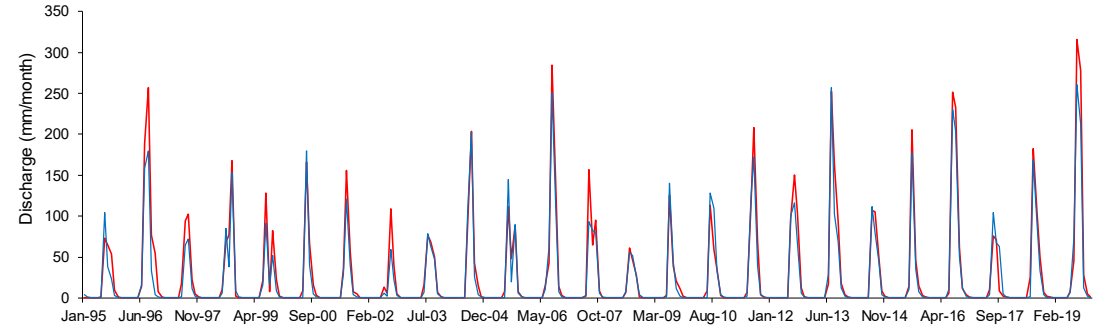
MAKSUDANGARH



MANAKHEDI



MANDSOR



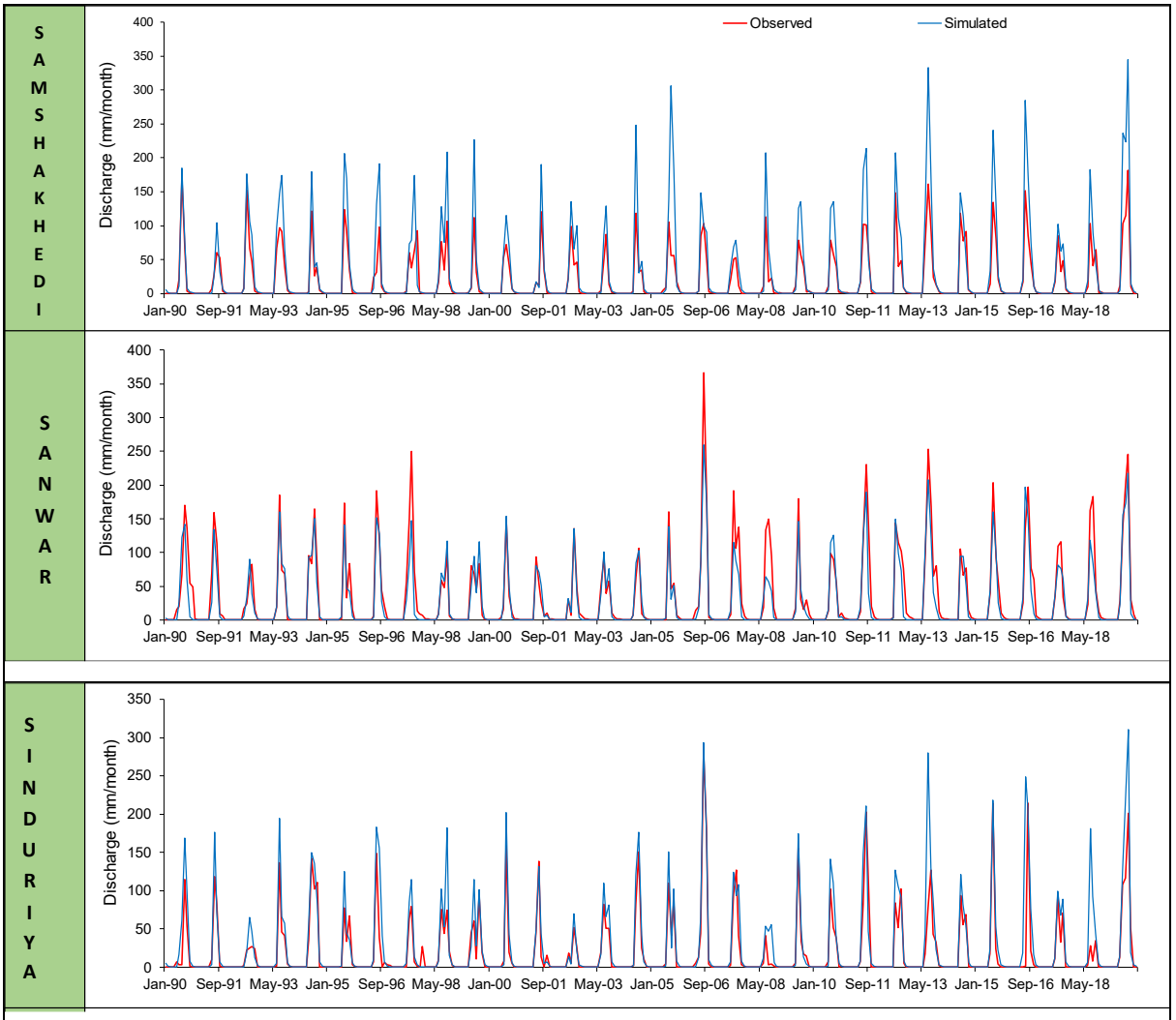


Figure 5. 23 Observed vs. Simulated Monthly Discharge at Different G/D Sites of the Chambal Basin.

5.6.3 Ken Basin IDW Regionalization

In the Ken Basin, X1 values range from 2.34 (Sonar) to 5.07 (Tigra), while X2 varies from 0.43 (Sonar) to 0.60 (Kopra). Efficiencies in the Ken Basin are relatively high, ranging from 69.7% (Pandwan) to 88.1% (Niwari), reflecting effective water utilization across most sites. Figure 5.24 showing the graph of observed vs. simulated monthly discharge at different G/D sites of the Ken basin.


KOPRA
NIWARI
NOHTA
PANDWAN
observed
Simulated
Discharge (mm/month)
0
50
100
150
200
250
300
350
Jan-90
Jan-92
Jan-94
Jan-96
Jan-98
Jan-00
Jan-02
Jan-04
Jan-06
Jan-08
Jan-10
Jan-12
Jan-14
Jan-16
Jan-18

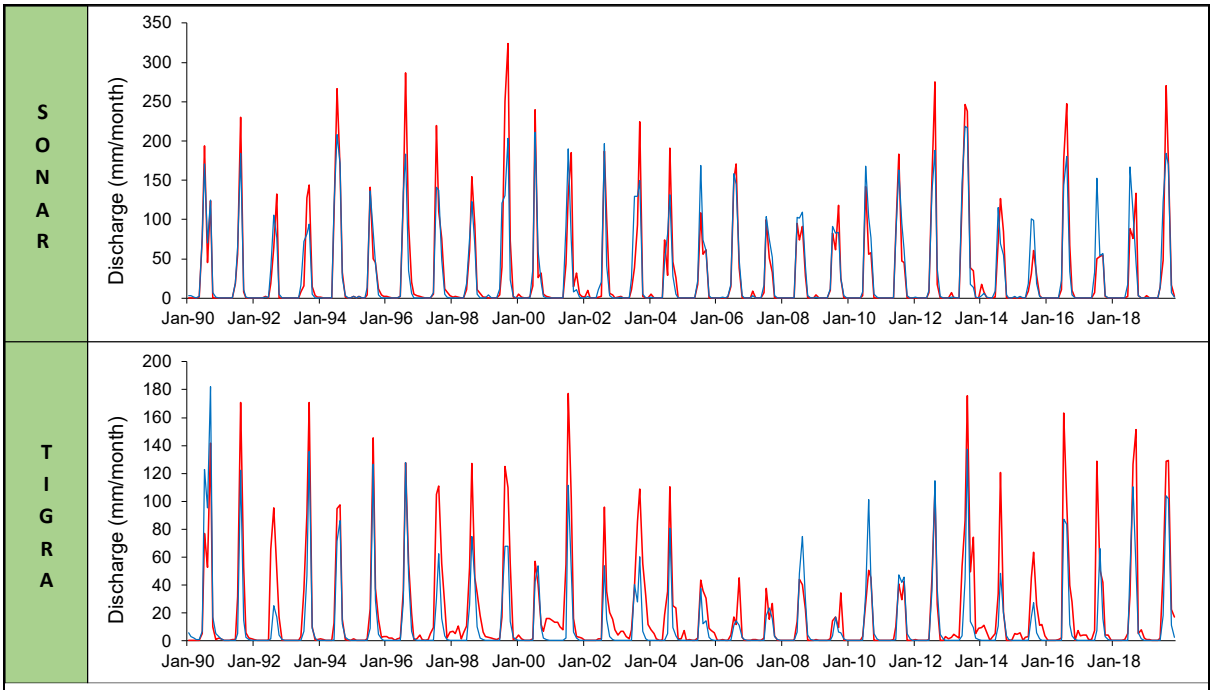


Figure 5. 24 Observed vs. Simulated Discharge at Different G/D Sites of the Ken Basin**.**

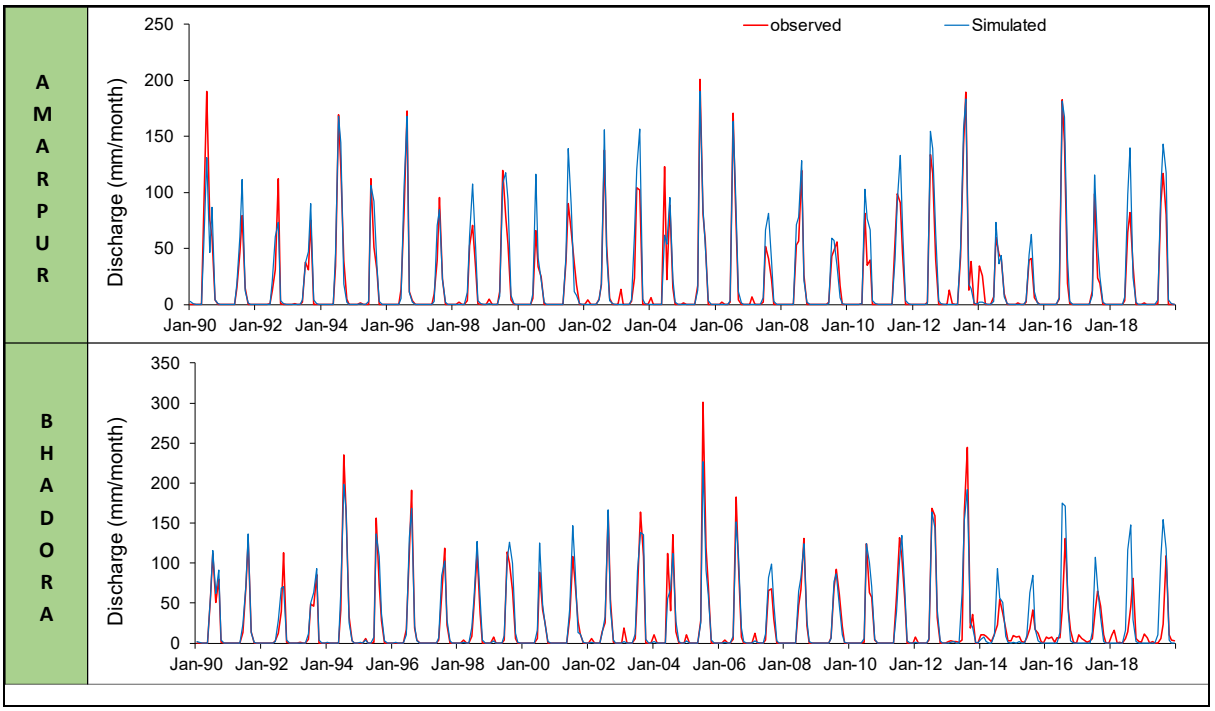
5.6.4 Sindh Basin IDW Regionalization

The Sindh Basin shows a narrower range for X1 and X2, with X1 spanning from 5.06 (Didi) to 5.49 (Amola) and X2 from 0.72 (Daboh) to 0.91 (Goraghat). Efficiency values in the Sindh Basin, however, range from a low of 52.2% (Budhara) to 74.4% (Amola), indicating some inefficiencies at specific sites. Figure 5.25 showing the graph of observed vs. simulated monthly discharge at different G/D sites of the Sindh basin.

Figure 5. 25 Observed vs. Simulated Discharge at Different G/D Sites of the Sindh Basin**.**

5.6.5 Son Basin IDW Regionalization

For the Son Basin, X1 values range from 2.82 (Bhadora) to 4.19 (Ghatkhirwa), while X2 remains consistent between 0.37 (Bhadora and Khatkhari) and 0.41 (Amarpur and Ghatkhirwa). Efficiencies in the Son Basin are generally high, ranging from 72.2% (Marsiabad) to 88.8% (Amarpur). The discharge graphs for GD sites in the Son basin demonstrate strong seasonal peaks corresponding to monsoon months, with observed and simulated values showing good agreement.At Amarpur, discharge peaks around 200 mm/month, while Bhadora exhibits higher variability, reaching up to 350 mm/month during extreme rainfall events. At Ghathirwa, discharge reaches up to 300 mm/month during high-flow periods, indicating substantial runoff during intense rainfall. Marsiabad exhibits moderate discharge levels, peaking at around 250 mm/month. Figure 5.26 showing the graph of observed vs. simulated monthly discharge at different G/D sites of the Son basin.



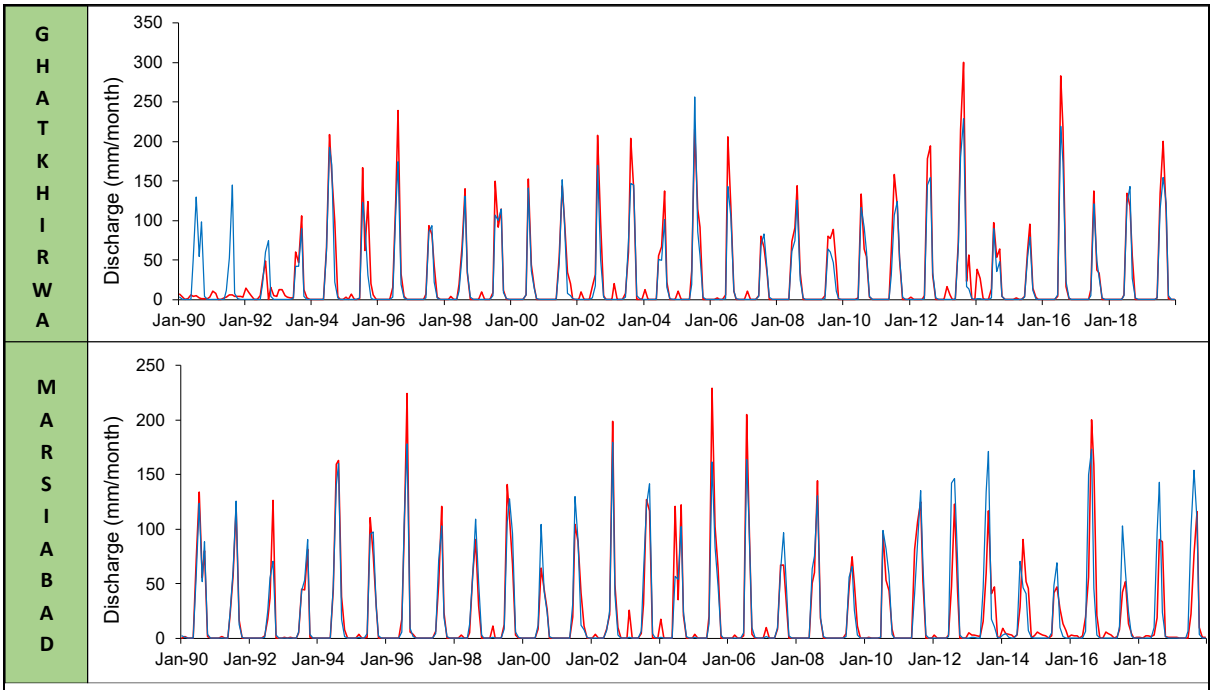


Figure 5. 26 Observed vs. Simulated Discharge at Different G/D Sites of the Son Basin**.**

5.6.6 Tons Basin IDW Regionalization

In the Tons Basin, X1 values are the lowest among all basins, ranging from 1.85 (Sansarpur) to 2.6 (Patehera). X2 is also lower, spanning from 0.37 (Khatkhari) to 0.40 (Sansarpur). Efficiency values for this basin are moderate to high, ranging from 70.2% (Sansarpur) to 85.0% (Maihar). Figure 5.27 showing the graph of observed vs. simulated monthly discharge at different G/D sites of the Tons basin.

Figure 5. 27 Observed vs. Simulated Discharge at Different G/D Sites of the Tons Basin**.**

Figure 5.28 showing the IDW Map of X1 and X2 parameter for Betwa and Chambal Basin. In the Chambal Basin, for Parameter X1, higher values are predominantly concentrated in the northern and central regions, while lower values are observed in the southeastern areas. For Parameter X2, the highest values are located in the central parts of the basin, with lower values distributed towards the northern and southeastern regions. Similarly, in the Betwa Basin, Parameter X1 exhibits the highest values in the western region, while lower values are observed

in the eastern parts. For Parameter X2, higher values are concentrated in the central regions, with lower values extending towards the northern and southern boundaries.

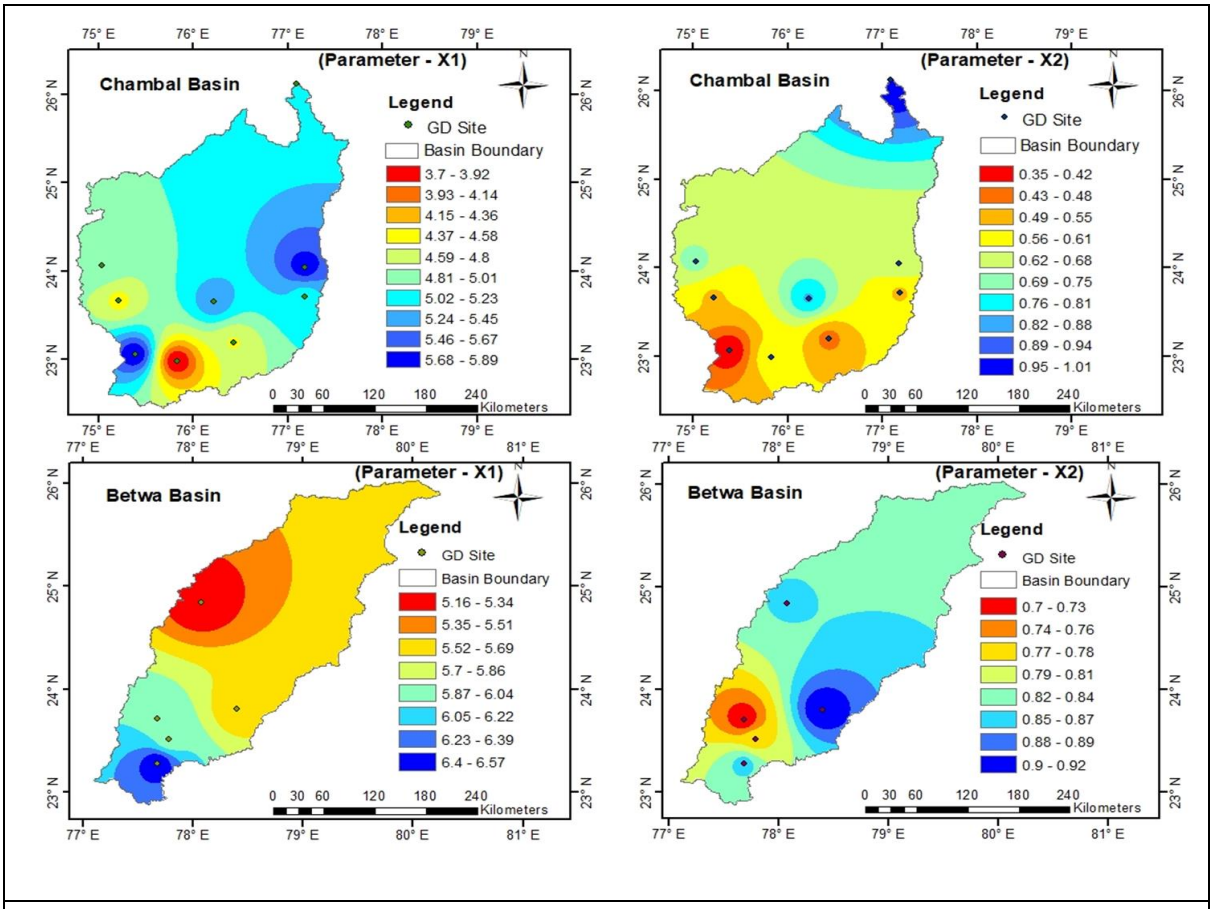


Figure 5. 28 IDW Map of X1 and X2 parameter for Betwa and Chambal Basin.

Figure 5.29 showing the IDW Map of X1 and X2 parameter for Ken and Sindh Basin. In the Ken Basin, higher X1 values (4.25–5.90) dominate the northern-central region, while higher X2 values (0.46–0.54) are localized. In the Sindh Basin, higher X1 values (5.13–5.23) and X2 values (0.65–0.83) are concentrated in the northern regions.

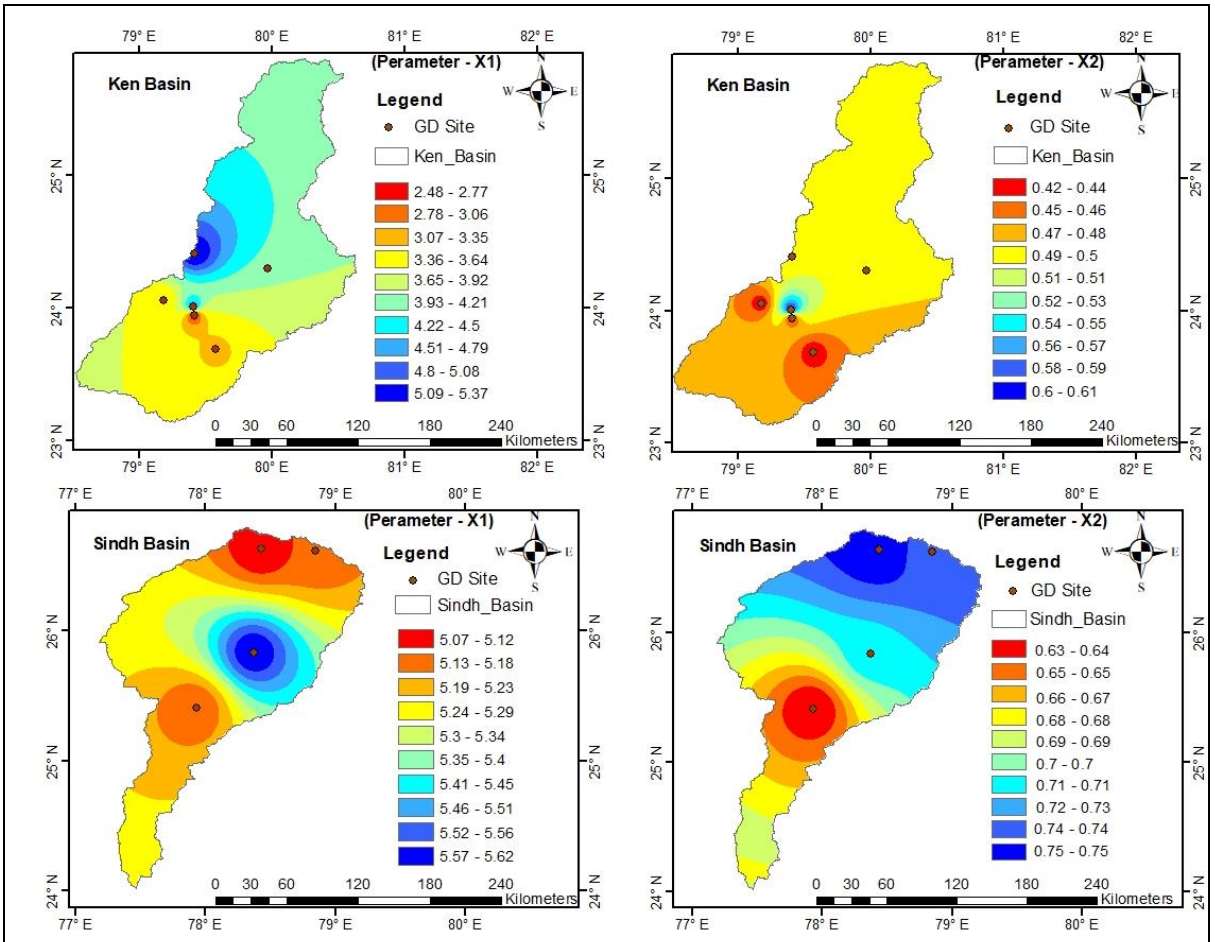


Figure 5. 29 IDW Map of X1 and X2 parameter for Ken and Sindh Basin.

Similarly Figure 5.30 showing the IDW Map of X1 and X2 parameter for Ken and Sindh Basin. In the Son Basin, higher X1 values (3.94–4.28) are concentrated in the northwestern region, while lower values (2.2–3.59) dominate the central and southeastern areas. Similarly, for X2, higher values (0.42–0.42) are localized in the northwest, with lower values (0.36–0.41) covering most of the basin. In the Tons Basin, higher X1 values (2.66–2.83) are observed in the southeastern region, with lower values (1.73–2.46) spread across the central and northern areas. For X2, the higher values (0.46–0.47) are in the southeastern part, while lower values (0.37–0.44) dominate the rest of the basin.

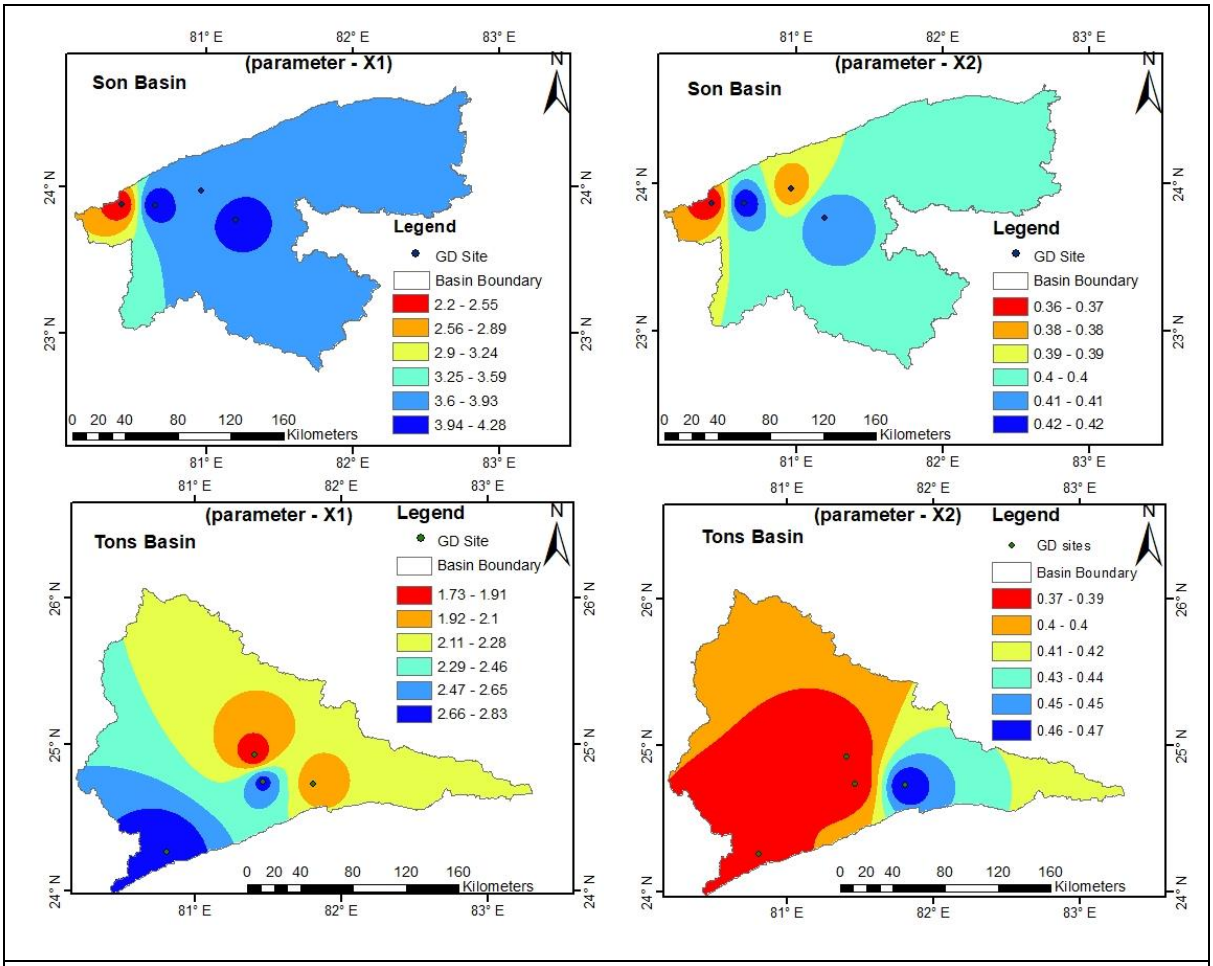


Figure 5. 30 IDW Map of X1 and X2 parameter for Son and Tons Basin.

CHAPTER: 6
SUMMARY & CONCLUSIONS

6.1 Conclusion

This study successfully applied the GEE-based SCS-CN model to evaluate runoff in six sub-basins of the Ganga river in Madhya Pradesh, India: Chambal, Betwa, Sindh, Son, Tons, and Ken. The integration of GEE with the SCS-CN model, along with rainfall data from the IMD, facilitated efficient and scalable hydrological analysis. This approach enabled accurate runoff predictions and provided insights into the hydrological behaviour of these diverse landscapes.

- **Hydrological Model Performance and Regression Analysis**

The regression analysis of runoff for the monsoon months (June to October) revealed significant variability in model performance across the basins, as reflected by the coefficient of determination (R^2). The Ken basin consistently demonstrated stable hydrological responses, with R^2 values peaking at 81.75% in June and 58.71% in October, while the Sindh basin exhibited greater variability, with R^2 ranging from 53.06% in June to 73.07% in September, highlighting the basin's sensitivity to rainfall variability and land use practices. The Son basin showed strong correlations during peak monsoon months, with R^2 values of 78.05% in July and 73.38% in August, underlining the impact of heavy rainfall on runoff. The Chambal basin demonstrated considerable variability, particularly in October, with R^2 dropping to 17.32%, indicating the effects of reduced rainfall and anthropogenic influences. The Betwa basin exhibited improved correlations post-monsoon, with R^2 peaking at 73.67% in September, reflecting stable hydrological conditions. Meanwhile, the Tons basin maintained consistently high R^2 values, peaking at 76.98% in September, attributed to its relatively uniform catchment characteristics and stable climatic conditions.

- **Machine Learning for Enhanced Prediction**

Machine learning models, including Huber Regression, Decision Tree Regressor, Ridge Regression, and Random Forest, were employed to improve runoff prediction accuracy. Evaluated using metrics such as MAE, RMSE, R^2, RMSLE, and MAPE, these models provided basin-specific insights. For example, in the Betwa basin, R^2 peaked at 47.69% in September, while the lowest RMSE of 16.597 was observed in October. These results underscore the necessity of tailored modeling approaches for hydrological prediction, as different machine

learning models demonstrated varying degrees of success across basins and temporal scales.

- **Regionalization with IDW**

The Inverse Distance Weighting (IDW) technique played a pivotal role in regionalizing hydrological parameters for ungauged basins. IDW interpolates values based on the proximity of gauged sites, assuming greater similarity between spatially closer points. In this study, IDW was used to estimate the parameters X1 and X2 for the GR2M model, resulting in high predictive accuracy for runoff. For example, in the Betwa basin, X1 ranged from 5.77 to 5.95, while X2 varied from 0.82 to 0.87, with corresponding runoff prediction efficiencies ranging from 63.8% (Mohanpur) to 87.0% (Bah). This highlights the robustness of IDW in capturing spatial heterogeneity and improving parameter estimation for ungauged sites, particularly in basins with diverse climatic and land use conditions.

- **Cluster Analysis for Watershed Zonation**

To further enhance the understanding of watershed characteristics, cluster analysis was applied to group areas with similar hydrological behavior. This zonation technique identified regions with consistent runoff responses and enabled the delineation of hydrologically homogeneous areas. By integrating the results of cluster analysis with IDW and GR2M, this study effectively reduced modeling uncertainties and provided a deeper understanding of spatial variations in runoff generation. For instance, the clustering process revealed distinct hydrological zones within the Chambal and Sindh basins, which helped refine parameter interpolation for ungauged sites. The synergy between cluster analysis and regionalization techniques ensured that both localized and broader-scale hydrological behaviors were adequately captured, leading to improved model performance.

- **GR2M Model Performance**

The GR2M model, a lumped monthly hydrological model, demonstrated strong performance in simulating runoff across the six sub-basins. Calibration efficiencies ranged from 75.4% in the Sindh basin to 94.10% in the Chambal basin, while validation efficiencies ranged from 61.2% in the Betwa basin to 98.00% in the Sindh basin. The GR2M model's simplicity and adaptability allowed it to capture essential hydrological processes effectively. When combined with IDW-derived parameters, the GR2M model achieved high efficiencies even in ungauged regions, further demonstrating its suitability for data-scarce environments.

The integration of GR2M with IDW resulted in validation efficiencies as high as 93.4%

in the Chambal basin and 58% in the Sindh basin, showcasing the method's potential for enhancing runoff predictions in regions lacking direct measurements. The synergy between GR2M and IDW offers a practical framework for regionalizing hydrological parameters and simulating monthly runoff with significant accuracy.

- **Key Insights and Future Directions**

This study highlights the importance of regionalization techniques like IDW and hydrological models like GR2M in addressing challenges associated with ungauged basins. The IDW method effectively bridges the gap between gauged and ungauged sites, while GR2M provides a robust platform for simulating runoff at a monthly scale. Together, these tools enabled accurate runoff prediction, with R^2 values exceeding 80% in several basins and minimized RMSE, confirming their utility for water resource management.

Future research should explore refining the IDW technique by integrating additional variables such as topography, soil characteristics, and land use changes to further enhance parameter interpolation accuracy. Extending the GR2M model to incorporate climate change scenarios and socioeconomic factors will provide deeper insights into the impacts of future climatic and anthropogenic changes on water resources. Expanding the application of these methodologies to other regions with varied climatic and hydrological conditions will validate their generalizability and contribute to the development of more resilient and sustainable water resource management strategies.

By demonstrating the efficacy of integrated approaches like GEE-based SCS-CN modeling, IDW regionalization, and GR2M simulation, this study provides a robust framework for addressing the challenges of hydrological modeling in data-scarce regions and lays the groundwork for future advancements in water resource management.

REFERENCES

- Anand, J., Gosain, A. K., Khosa, R., & Srinivasan, R. (2018). Regional scale hydrologic modeling for prediction of water balance, analysis of trends in streamflow and variations in streamflow: The case study of the Ganga river basin. Journal of Hydrology: Regional Studies, 16, 32-53.
- Arsenault, R., Breton-Dufour, M., Poulin, A., Dallaire, G., & Romero-Lopez, R. (2019). Streamflow prediction in ungauged basins: analysis of regionalization methods in a hydrologically heterogeneous region of Mexico. Hydrological Sciences Journal, 64(11). 1297-1311. doi: 10.1080/02626667 2019.1639716
- Bárdossy, A., 2007. Calibration of hydrological model parameters for ungauged catchments. *Hydrology and Earth System Sciences Discussions*, 11 (2), 703–710. https://doi.org/10.5194/hess-11-703-2007.
- Bera, S. and Maiti, R., 2021. Assessment of water availability with SWAT model: A study on Ganga river. *Journal of the Geological Society of India*, *97*(7), pp.781-788. https://doi.org/10.1007/s12594-021-1760-9
- Blöschl, G. (2006). Rainfall-runoff modeling of ungauged catchments. Encyclopedia of hydrological sciences.
- Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A. and Savenije, H. eds., 2013. *Runoff prediction in ungauged basins: synthesis across processes, places and scales*. Cambridge University Press.
- Bonta, J. V. (1997). Determination of Watershed Curve Number Using Derived Distributions. *Journal of Irrigation and Drainage Engineering*, *123*(1), 28-36. doi:10.1061/(ASCE)0733-9437(1997)123:1(28)
- Boscarello, L., Ravazzani, G., Cislaghi, A., and Mancini, M. (2015). "Regionalization of-duration curves through catchment classification with streamflow signatures and physiographic-climate indices. J. Hydrol. Eng., 10.1061/(ASCE)HE 1943-5584.0001307, 05015027.
- Böttcher, S., Merz, C., Lischeid, G., and Dannowski, R. (2014). "Using Isomap to differentiate between anthropogenic and natural effects on groundwater dynamics in a complex geological setting." J. Hydrol., 519(B), 1634-1641.
- Burn, D. H., and Boorman, D. B. (1993). "Estimation of hydrological parameters at ungauged catchments. J. Hydrol., 143(3-4), 429-454
- Buytaert, W., Celleri, R., Willems, P., De Bievre, B., and Wyseure, G. (2006). “Spatial and

temporal rainfall variability in mountainous areas: A case study from the south Ecuadorian Andes." *J. Hydrol. (Amsterdam)*, 329(3-4), 413–421.

- Chiang, S. M., Tsay, T. K., and Nix, S. J. (2002b). "Hydrologic regionalization of watersheds.11: Applications." J. Water Resour. Plann. Manage, 10.1061/(ASCE)0733-9496(2002)128:1(12), 12-20.
- Chiang, S. M., Tsay, T.K., and Nix, S. J. (2002a). "Hydrologic regionalization of watersheds. 1: Methodology development." J. Water Resour. Plann. Manage, 10.1061/(ASCE)0733-9496(2002)128:1(3), 3-11
- Chow, V.T., Maidment, D.R. and Mays, L.W. (1988) Applied Hydrology. International Edition, McGraw-Hill Book Company, New York.
- Ditthakit, P., Pinthong, S., Salaeh, N., Weekaew, J., Tran, T.T. and Pham, Q.B., 2023. Comparative study of machine learning methods and GR2M model for monthly runoff prediction. *Ain Shams Engineering Journal*, *14*(4), p.101941.
- Dubayah, R and Lerrenmaier D. (1997). Combing Remote Sensing and Hydrological Modeling for Applied Water and Energy Balance Studies. NASA EOS Interdisciplinary Working Group Meeting, San Diego, CA.
- Gao, Z., Long, D., Tang, G., Zeng, C., Huang, J. and Hong, Y., 2017. Assessing the potential of satellite-based precipitation estimates for flood frequency analysis in ungauged or poorly gauged tributaries of China's Yangtze River basin. *Journal of hydrology*, *550*, pp.478-496.
- Garen, D. C., & Moore, D. S. (2005). Curve Number Hydrology in Water Quality Modeling: Uses, Abuses, and Future Directions. *Journal of the American Water Resources Association*, *41*(2), 377-388. doi:10.1111/j.1752-1688.2005.tb03742.x
- Golian, S., Murphy, C. and Meresa, H., 2021. Regionalization of hydrological models for flow estimation in ungauged catchments in Ireland. *Journal of Hydrology: Regional Studies*, *36*, p.100859.
- Goovaerts, P., 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of hydrology*, *228*(1-2), pp.113-129.
- Hawkins, R. H. (1978). Runoff Curve Numbers with Varying Site Moisture. *Journal of the Irrigation and Drainage Division, 104*(4), 389-398. doi:10.1061/JRCEA4.0001507
- Hawkins, R.H., Steenhuis, T.S., Frankenberger, J.R., Winchell, M., Zollweg, J.A. and Walter, M.F., 1996. Discussion and Closure: SCS Runoff Equation Revisited for Variable-Source Runoff Areas. *Journal of Irrigation and Drainage Engineering*, *122*(5), pp.319-320.
- Hernandez, M., et al. (2000). Modeling runoff response to land cover and rainfall spatial

variability in semi-arid watersheds. *Environmental Monitoring and Assessment*, *64*(1), 285-298. doi:10.1023/A:1006486229452

- Kanishka, G., & Eldho, T. (2017), Watershed Classification Using Isomap Technique and Hydrometeorological Attributes, Journal Of Hydrologic Engineering, 22(10), 04017040 doi: 10.1061/(asce)he: 1943-5584.0001562
- Kapangaziwiri, E., Hughes, D.A., and Wagener, T., 2012. Incorporating uncertainty in hydrological predictions for gauged and ungauged basins in southern Africa. *Hydrological Sciences Journal*, 57 (5), 1000–1019. https://doi.org/10.1080/02626667.2012.690881.
- Kay, A.L., Jones, D.A., Crooks, S.M., Calver, A. and Reynard, N.S., 2006. A comparison of three approaches to spatial generalization of rainfall–runoff models. *Hydrological Processes: An International Journal*, *20*(18), pp.3953-3973.
- Kumar, S., Gil, G.S. and Santosh, S., 2015. Spatial Distribution of Rainfall with Elevation in Satluj River Basin: 1986-2010, Himachal Pradesh, India. *World Scientific News*, *19*, pp.1-15.
- Li, J. and Heap, A.D., 2011. A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. *Ecological Informatics*, *6*(3-4), pp.228-241.
- Li, Q., Peng, Y., Wang, G., Wang, H., Xue, B., & Hu, X. (2019). A Combined Method for Estimating Continuous Runoff by Parameter Transfer and Drainage Area Ratio Method in Ungauged Catchments Water. 11(5), 1104. doi: 10.3390/w11051104
- Maidment, D.R., 1996, January. GIS and hydrologic modeling-an assessment of progress. In *Third International Conference on GIS and Environmental Modeling, Santa Fe, New Mexico*.
- Makungo, R., Odiyo, J.O., Ndiritu, J.G. and Mwaka, B., 2010. Rainfall–runoff modelling approach for ungauged catchments: A case study of Nzhelele River sub-quaternary catchment. *Physics and Chemistry of the Earth, Parts a/b/c*, *35*(13-14), pp.596-607.
- McIntyre, N., Lee, H., Wheater, H. S., Young, A., and Wagener, T. (2005). "Ensemble predictions of runoff in ungauged catchments." Water Resour. Res., 41(12), W12434.
- Merwade, V., Cook, A. and Coonrod, J., 2008. GIS techniques for creating river terrain models for hydrodynamic modeling and flood inundation mapping. *Environmental Modelling & Software*, *23*(10-11), pp.1300-1311.
- Merz, R., and Blöschl, G. (2004). "Regionalisation of catchment model parameters." J. Hydrol., 287(1-4), 95-123.
- Mishra, S. K., & Singh, V. P. (2003). *Soil Conservation Service Curve Number (SCS-CN) Methodology*. Springer Science & Business Media.

- Nigam, A., Awasthi, M.K. and Bunkar, N., 2020. Assessment of groundwater potential zones of tons basin using spatial data. *International Journal of Agriculture, Environment and Biotechnology*, *13*(3), pp.261-268.
- Oudin, L. Andréassian, V., Perrin, C., Michel, C., and Le Moine, N. (2008). "Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments." Water Resour. Res., 44(3), W03413
- Pandey, A., Chowdary, V. M., & Mal, B. C. (2009). Runoff and sediment yield modeling from a small agricultural watershed in India using the WEPP model. *Journal of Hydrology*, *348*(3-4), 305-319. doi:10.1016/j.jhydrol.2007.10.005
- Parajka, J., Merz, R. and Blöschl, G., 2005. A comparison of regionalisation methods for catchment model parameters. *Hydrology and earth system sciences*, *9*(3), pp.157-171.
- Parmar, H.V., Mashru, H.H., Vekarıya, P.B., Rank, H.D., Kelaıya, J.H., Pardava, D.M., Patel, R.J. and Vadar, H.R., 2016. Establishment of rainfall-runoff relationshıp for the estimation runoff in semi-arid catchment. *AGRES–An International e-Journal*, *5*(1), pp.60-67.
- Ponce, V. M., & Hawkins, R. H. (1996). Runoff curve number: Has it reached maturity? *Journal of Hydrologic Engineering, 1*(1), 11-19. doi:10.1061/(ASCE)1084-0699(1996)1:1(11)
- Rallison, R. E., & Miller, N. (1981). Past, Present, and Future SCS Runoff Procedure. *Journal of the Hydraulics Division, 107*(HY4), 503-514. doi:10.1061/JYCEAJ.0005655
- Rao, A. R.. and Srinivas, V. V. (2006). "Regionalization of watersheds by hybrid-cluster analysis. J. Hydrol., 318(1-4), 37-56.
- Razavi, T., and Coulibaly, P. (2013a). "Classification of Ontario watersheds based on physical attributes and streamflow series." J. Hydrol., 493, 81-94.
- Samuel, J., Coulibaly, P., and Metcalfe, R. A. (2011). "Estimation of continuous streamflow in Ontario ungauged basins: Comparison of regionalization methods." J. Hydrol. Eng. 10.1061/(ASCE)HE 1943-5584 0000338, 447-459
- Shu, C. and Ouarda, T.B.M.J., 2007. Flood frequency analysis at ungauged sites using artificial neural networks in canonical correlation analysis physiographic space. *Water Resources Research*, 43 (7). https://doi.org/10.1029/2006WR005142.
- Sivapalan, M., et al. (2003). "IAHS decade on predictions in ungauged basins (PUB), 2003-2012: Shaping an exciting future for the hydrological sciences." Hydrol. Sci. J., 48(6), 857-880
- Sivapalan, M., et al., 2003. IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: shaping an exciting future for the hydrological sciences. *Hydrological Sciences Journal*, 48 (6), 857–880. doi:https://doi.org/10.1623/hysj.48.6.857.

- Swain, J.B., Jha, R. and Patra, K.C., 2015. Stream flow prediction in a typical ungauged catchment using GIUH approach. *Aquatic Procedia*, *4*, pp.993-1000.
- Xu, W., Zou, Y., Zhang, G. and Linderman, M., 2015. A comparison among spatial interpolation techniques for daily rainfall data in Sichuan Province, China. *International Journal of Climatology*, *35*(10).
- Yadav, M., Wagener, T., and Gupta, H., 2007. Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. *Advances in Water Resources*, 30 (8), 1756–1774. https://doi.org/10.1016/j.advwatres.2007.01.005.
- Yang, X., Magnusson, J., Rizzi, J, & Xu, C. (2017) Runoff prediction in ungauged catchments Norway: comparison of regionalization approaches. Hydrology Research, 49(2), 487-505. doi: 10.2166/nh.2017
- Zhang, Y. and Chiew, F.H., 2009. Relative merits of different methods for runoff predictions in ungauged catchments. *Water Resources Research*, *45*(7)