Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Hydrological Modelling and Runoff Estimation of the Wainganga River Basin Using HEC-HMS and Geospatial Clustering Techniques

Rajesh Vijaykumar Kherde

rkherde@gmail.com

D Y Patil University Pune Ambi https://orcid.org/0000-0003-4589-2156

Gauri Patil

Dr D Y Patil Institute of Technology

Kiran More

D Y Patil University Pune Ambi

Priyadarshi Sawant

Sardar Patel College of Engineering

Research Article

Keywords: Wainganga River Basin, Flood Forecasting, HEC-HMS, Rainfall-Runoff Modelling, Hierarchical Clustering, Thiessen Polygon, Hydrological Modelling, GIS, Rain Gauge Network, Runoff Estimation

Posted Date: July 31st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10200208/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Hydrological Modelling and Runoff Estimation of the Wainganga River Basin Using HEC-HMS and Geospatial Clustering Techniques

KHERDE R.V.[1*], GAURI PATIL[2], MORE K.C.[3], SAWANT P.H[4]

[1, 3] D.Y. Patil University Ambi Pune, Maharashtra, India

[2]Dr. D. Y. Patil Institute of Technology, Pune, India,
Dnyaan Prasad Global University (DPGU), School of Technology and Research,
Dr. D. Y. Patil Unitech Society, Sant Tukaram Nagar, Pimpri, Maharashtra, Pune – 411018, India

[4]Sardar Patel College of Engineering, Mumbai, Maharashtra, India

(*) rkherde@gmail.com

Abstract

Flooding is a recurring problem in the Wainganga River basin of central India, causing considerable damage to agriculture, infrastructure, and human settlements during intense monsoon events. This study presents an integrated approach for rainfall-runoff modelling and flood forecasting using geospatial techniques and the Hydrologic Engineering Center–Hydrologic Modeling System (HEC-HMS). Rainfall and discharge data from 2001–2017 were collected for the Wainganga basin to identify representative rain gauge stations and evaluate runoff behaviour. Hierarchical clustering was applied to classify rain gauge stations according to rainfall similarity, while the Thiessen polygon method was used to determine areal rainfall distribution. Basin delineation, land use/land cover analysis, and hydrological parameter extraction were performed using ArcGIS and HEC-GeoHMS.The HEC-HMS model was calibrated for 2001–2010 and validated for 2011–2017 using observed discharge data at Asthi outlet station. Model performance was evaluated using Nash–Sutcliffe Efficiency (NSE), coefficient of determination (R^2), and Root Mean Square Error (RMSE). The model produced satisfactory results, with NSE values ranging from 0.897 to 0.98, R^2 values above 0.895, and RMSE values between 0.2 and 1.36 during calibration and validation. Simulated peak discharge values closely matched observed discharge trends for major monsoon events. The results demonstrate that the combined application of hierarchical clustering, Thiessen polygon analysis, and HEC-HMS modelling significantly improves runoff estimation and flood forecasting accuracy in the Wainganga basin. The developed framework can support flood management, water resource planning, and early warning systems in monsoon-dominated river basins.

Keywords: *Wainganga River Basin; Flood Forecasting; HEC-HMS; Rainfall-Runoff Modelling; Hierarchical Clustering; Thiessen Polygon; Hydrological Modelling; GIS; Rain Gauge Network; Runoff Estimation.*

1. Introduction

Floods are among the most frequent and destructive natural hazards affecting river basins across India, particularly in monsoon-dominated regions[1]. The Wainganga River basin, spanning parts of Maharashtra and Madhya Pradesh, is highly vulnerable to seasonal flooding due to intense rainfall, undulating terrain, and expanding human settlements[2]. Flood events in this basin often lead to significant loss of life, agricultural damage, and disruption of

infrastructure. Therefore, reliable flood forecasting systems are essential for effective flood risk management and mitigation. Accurate rainfall data forms the backbone of any flood forecasting model. However, in large and heterogeneous catchments such as the Wainganga basin, collecting continuous and reliable data from all rain gauge stations is often impractical. Issues such as equipment malfunction, non-recording gauges, data transmission failures during extreme weather, and limited accessibility in remote forested areas create significant gaps in rainfall datasets[3]. Additionally, climate variability has altered rainfall intensity, distribution, and runoff patterns in central India, further complicating hydrological predictions. Hence, identifying an optimal and efficient rain gauge network becomes critical for improving flood forecasting accuracy in the Wainganga basin[4]. The selection of representative rain gauge stations is a challenging task due to the spatial variability of rainfall across the basin. Various statistical and analytical techniques have been developed to address this issue. Methods such as Principal Component Analysis (PCA), entropy-based approaches, and correlation coefficient analysis are commonly used to identify key rainfall stations. PCA helps in extracting dominant patterns influencing rainfall variability, while correlation-based methods evaluate relationships between different stations. Entropy-based approaches, rooted in information theory, identify stations that contribute maximum information to the system. Despite their usefulness, these techniques have limitations, such as assumptions of linear relationships, temporal stationarity and computational complexity, which may not fully capture the hydrological diversity of the Wainganga basin. To overcome these limitations, advanced spatial and statistical methods such as hierarchical clustering (HC) and Thiessen polygon techniques are increasingly employed. Hierarchical clustering groups rain gauge stations based on similarity in rainfall characteristics, allowing identification of representative clusters within the basin. The Thyessin polygon method, on the other hand, spatially divides the watershed into influence zones around each rain gauge, assigning weights based on area coverage. These approaches, when combined, provide a more robust framework for optimizing rain gauge networks by integrating both statistical similarity and spatial representation. Previous studies have demonstrated that such integrated approaches enhance the reliability of rainfall estimation and improve flood prediction accuracy in complex catchments[5].

Flood forecasting also relies heavily on rainfall-runoff modeling, which describes the transformation of precipitation into surface runoff within a watershed. Rainfall-runoff models can be broadly categorized into empirical, conceptual, and physically based models. Empirical models use historical relationships between rainfall and runoff, while conceptual models simplify hydrological processes using mathematical representations[6], [7][8]. Physically based models simulate actual physical processes governing water movement, requiring detailed input data such as soil characteristics, land use, and topography. Each model type has its advantages and limitations, and the choice depends on data availability and basin characteristics.

One of the most widely used hydrological models for flood forecasting is the Hydrologic Engineering Center's Hydrologic Modeling System (HEC-HMS). The HEC-HMS model is a semi-distributed, physically based tool designed to simulate the rainfall-runoff process in watershed systems. It converts precipitation inputs into stream flow by considering watershed characteristics such as land use, soil properties, and basin geometry. The model incorporates various components, including unit hydrograph methods for runoff generation, routing models such as Muskingum-Cunge for channel flow, and groundwater modules for subsurface processes. Due to its flexibility, user-friendly interface, and ability to simulate large and complex basins, HEC-HMS is extensively applied in flood forecasting studies worldwide[9].The Wainganga River basin, being a major tributary of the Godavari River system, plays a crucial role in the hydrology of central India. Effective flood forecasting in

this basin requires an optimized rain gauge network and reliable hydrological modelling framework[10]. Therefore, the present study aims to identify key rain gauge stations within the Wainganga basin using hierarchical clustering and Thiessen polygon techniques. Furthermore, the study employs the HEC-HMS model to simulate rainfall-runoff processes and estimate runoff based on observed rainfall data[11]–[13].

The outcomes of this research are expected to improve the accuracy and reliability of flood forecasting in the Wainganga River basin. The study identifies key rain gauge locations to optimize resource utilization, strengthen early warning capabilities, and improve flood management and mitigation planning. The findings will be valuable for researchers, policymakers, and water resource managers involved in flood risk mitigation and sustainable watershed management in central India.

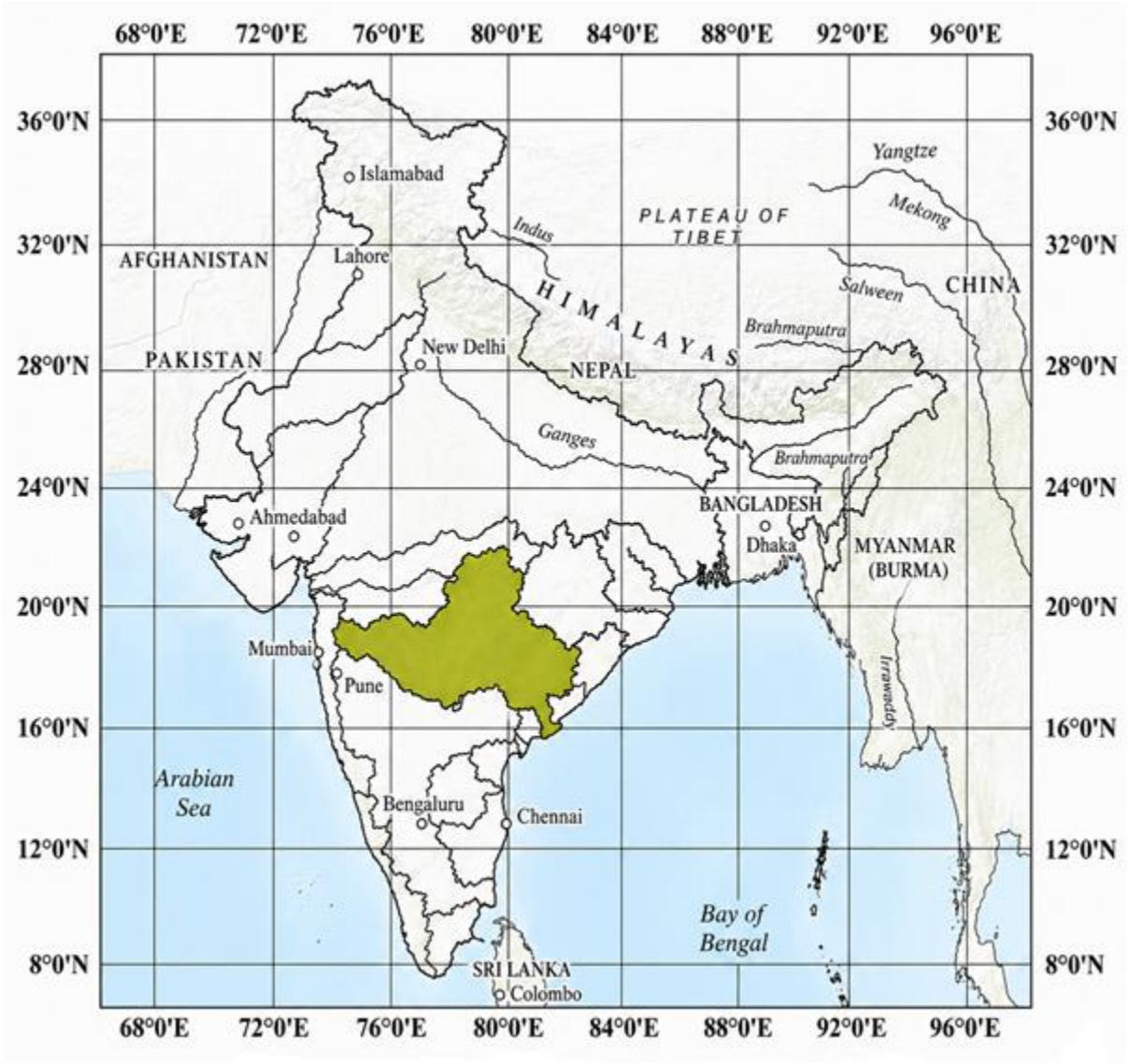


Fig. 1. (a) Geographical representation of the Godavari River Basin

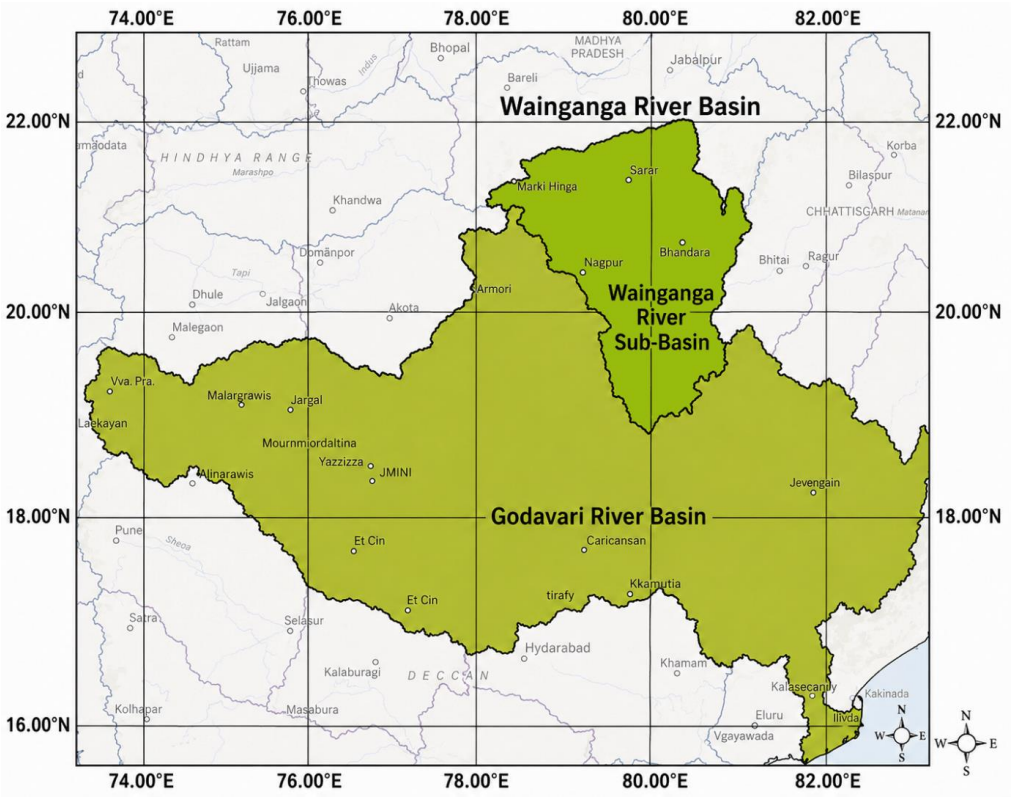


Fig. 1. (b) Location map of Wainganga River Basin Shown as Sub Basin of Godavari River

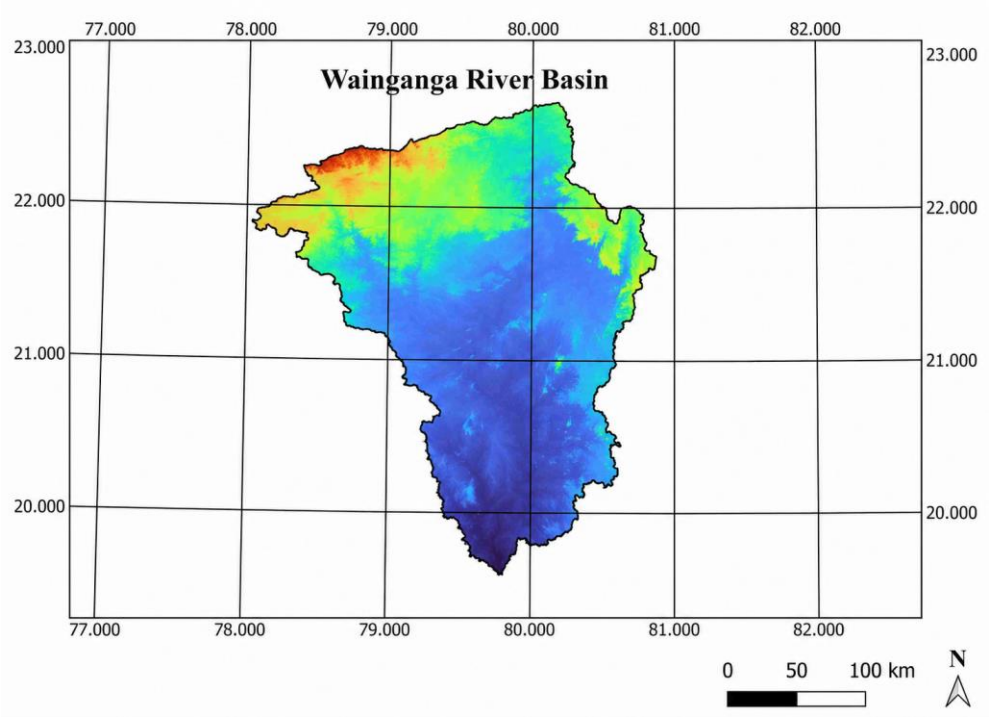


Fig. 1. (C) Delineated Wainganga River Basin

Table 1: Spatial distribution of rain gauge and gauging stations within the Wainganga Sub basin

Basin Id	Location (Station Name)	Latitude	Longitude
WG01	**Partabpur** (Source)	21° 57′ 00″ N	79° 34′ 00″ E
WG02	**Seoni**	22° 05′ 00″ N	79° 32′ 00″ E
WG03	**Balaghat**	21° 48′ 00″ N	80° 11′ 00″ E
WG04	**Gondia**	21° 27′ 00″ N	80° 11′ 00″ E
WG05	**Bhandara**	21° 10′ 00″ N	79° 39′ 00″ E
WG06	**Pauni** (Gosekhurd)	20° 47′ 00″ N	79° 38′ 00″ E
WG07	**Gadchiroli**	20° 11′ 00″ N	79° 59′ 00″ E
WG08	**Ashti** (Outlet)	19° 44′ 00″ N	79° 48′ 00″ E
WG09	**Bamni**	19° 53′ 00″ N	79° 47′ 00″ E
WG10	**Tirora**	21° 24′ 00″ N	79° 56′ 00″ E
WG11	**Sitekasa**	21° 31′ 00″ N	80° 31′ 00″ E
WG12	**Shivni**	22° 03′ 00″ N	79° 31′ 00″ E
WG13	**Kumhari**	21° 50′ 00″ N	80° 15′ 00″ E
WG14	**Waraseoni**	21° 46′ 00″ N	80° 03′ 00″ E
WG15	**Armori**	20° 28′ 00″ N	79° 59′ 00″ E
WG16	**Chamorshi**	19° 56′ 00″ N	79° 53′ 00″ E
WG17	**Bhimkund**	21° 34′ 00″ N	79° 22′ 00″ E

2. Study Area

The Wainganga River is a major tributary of the Godavari River and serves as the primary drainage system for the eastern Vidarbha region of Maharashtra and the southern parts of Madhya Pradesh[14]. It drains a substantial area of approximately 49,277 square kilometers as it flows through the states of Madhya Pradesh and Maharashtra. The Wainganga basin is located between the North Latitudes 18° 48′ and 22° 43′ and the East Longitudes 76° 00′ and 80° 53′.The river rises in the Mundara village of the Seoni district in the Satpura Range at an elevation of roughly 640 meters. It flows through a diverse landscape, starting in the rugged, hilly terrain of the Satpuras before entering the fertile, broad plains of Balaghat and Asthi. The Wainganga river converges with the Wardha river at Chaprala, giving rise to the Pranhita river, which subsequently drains into the Godavari river.This study focuses on the middle and lower reaches of the Wainganga, specifically the section between the Sanjay Sarovar (Bhimgarh) Dam in Madhya Pradesh and the Gosikhurd Dam in Maharashtra. This stretch spans a distance of approximately 320 km and is located between the North Latitudes 20° 45′ to 22° 05′ and East Longitudes 79° 30′ to 80° 15′.The basin's hydrology is characterized by a rapid response during the monsoon; the concentration period is approximately 8–10 hours from the Sanjay Sarovar Dam to the Balaghat gauging station and 18–24 hours from Balaghat to the Gosikhurd reservoir.The water resources are managed by the state governments of Madhya Pradesh and Maharashtra, which operate a network of over 45 gauge and discharge sites across the basin to monitor the heavy monsoon inflows. Fig. 1 (a,b & c) shows geographical representation of the Godavari river basin . Key Central Water Commission (CWC) monitoring sites are strategically located along this stretch, primarily at Kumhari, Balaghat, Asthi, and Pauni, providing critical data for flood forecasting and irrigation management.

3. Data Collection

The study focused on a 320 km stretch of the Wainganga River, primarily covering the reach from the Sanjay Sarovar Dam (Bhimgarh) to the Asthi outlet. Topographical and Digital

Elevation Model (DEM) data were obtained from the National Remote Sensing Centre (NRSC) via the "bhuvan-app3.nrsc.gov.in" portal to delineate sub-basin boundaries and stream networks.Hydrological discharge data were provided by the Central Water Commission (CWC), Nagpur, which oversees the Godavari basin monitoring. Rainfall data for the sub-basin spanning from 2010 to 2018 was sourced from the Maharashtra State Agriculture Department and the Madhya Pradesh Land Records department. Inflow data for the Asthi outlet was obtained from the Vidarbha Irrigation Development Corporation (VIDC), Nagpur.To improve the analysis of the rain gauge network for flood forecasting and runoff modeling, data were collected from various rain gauge stations located within the Wainganga basin. Table 1 provides Spatial distribution of rain gauge and gauging stations within the Wainganga Sub basin, along with their corresponding Basin IDs. The stations are strategically spread across diverse topographical zones within the basin, including key locations such as Seoni, Balaghat, Gondia, Bhandara, Pauni, Gadchiroli, Ashti, Bamni, Tirora, Sitekasa, Shivni, Kumhari, Waraseoni, Armori, Chamorshi, and Bhimkund.

4. Methodology & Structural Workflow

The model structure, as shown in Fig. 2, outlines the methodology of the research. The first step involves identifying key rain gauge stations by applying the Hierarchical Clustering (HC) technique, which groups stations based on the specific hydro-climatic characteristics of the Wainganga sub-basin areas.To determine the spatial influence of each station, the Thiessen polygon method is used to delineate the Wainganga River basin. These polygons are plotted and analyzed using ArcGIS (10.x) to calculate the area-weighted average rainfall across the complex terrain of the Satpura foothills and the Vidarbha plains. Subsequently, the Hydrologic Engineering Center's Hydrologic Modeling System (HEC-HMS 4.3) is employed to model the rainfall-runoff process. The runoff generated from this hydrologic model is then calibrated and validated by comparing it with the observed gauge discharge at the Asthi outlet.

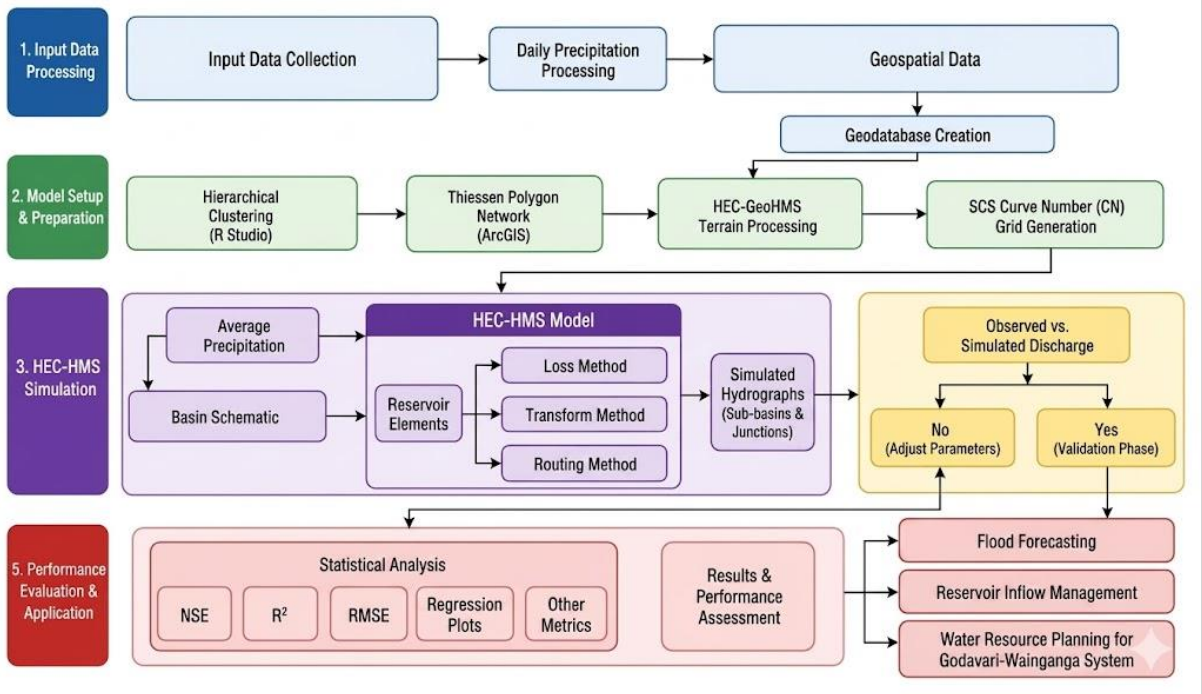


Fig. 2. Wainganga Basin Hydrological Modeling Workflow

4.1. Hydro-Climatic Station Clustering & Network Optimization

Hierarchical clustering, an effective unsupervised machine learning technique, was employed in this study to identify tree-like relationships among rain gauge stations, producing a dendrogram that illustrates the hierarchical grouping patterns of the stations. For the Wainganga basin, this allows for the categorization of stations based on similarities in rainfall intensity, elevation, and geographical proximity (e.g., grouping stations in the high-rainfall hilly regions of Seoni versus the plains of Asthi).The agglomerative approach is adopted here, which forms the hierarchy from the bottom-up. In this method, each rain gauge station starts as its own individual cluster; these clusters are then iteratively merged based on mathematical similarity until all data points belong to a single, overarching cluster. To perform the hierarchical clustering and generate the dendrogram for the Wainganga sub-basin, R version 3.6.1 Statistical Computing software is utilized, leveraging specialized packages for spatial-temporal data analysis.

4.2. Digital Elevation Model Processing & Geodatabase Delineation

A Digital Elevation Model (DEM) is a 3D depiction of the terrain surface that represents the elevation of the land. In hydrological modeling, it is essential for calculating slope, flow accumulation, and the overall landscape height. For this study, the DEM for the Wainganga River basin was downloaded from the "bhuvan.nrsc.gov.in" portal. Given the basin's extent across the Satpura Range and the Vidarbha plains, 12 tiles containing the satellite images were processed to extract basin characteristics using HEC-GeoHMS.

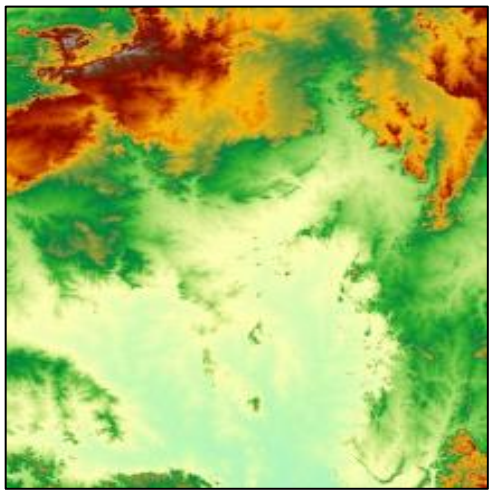
The following steps were taken to process the satellite images and prepare the terrain for the HEC-HMS model:

a) Mosaic Generation: Initially, all downloaded tiles were merged to create a single, continuous mosaic raster representing the entire Wainganga sub-basin (Fig. 3a).

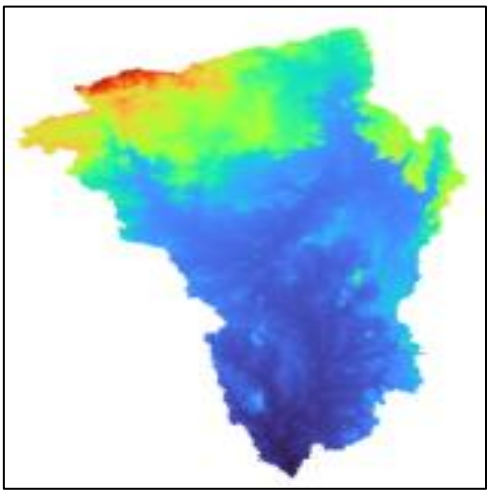
b) Sink Filling: The "Fill Sinks" (or Flow Fill) tool within the ArcGIS Spatial Analyst extension was used to remove depressions in the DEM, ensuring a continuous flow across the mosaic raster (Fig. 3b).

c) Flow Direction and Accumulation: The flow direction of each cell was processed to connect flow lines, accurately generating the drainage network and river system within the mosaic (Fig. 3c).

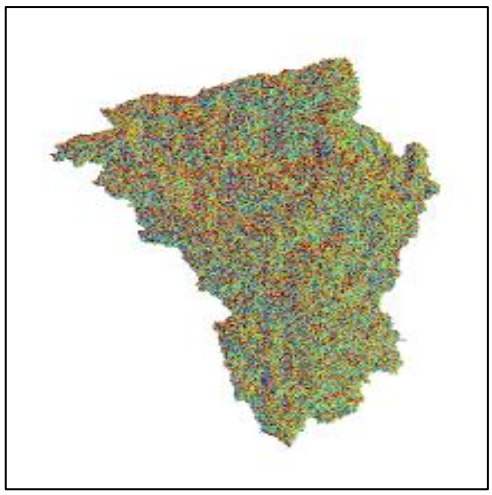
d) Sub-basin Delineation: Finally, the generated river system and pour points were processed to create delineated sub-basins across the mosaic raster, providing the physical boundaries for the rainfall-runoff analysis (Fig. 3d).



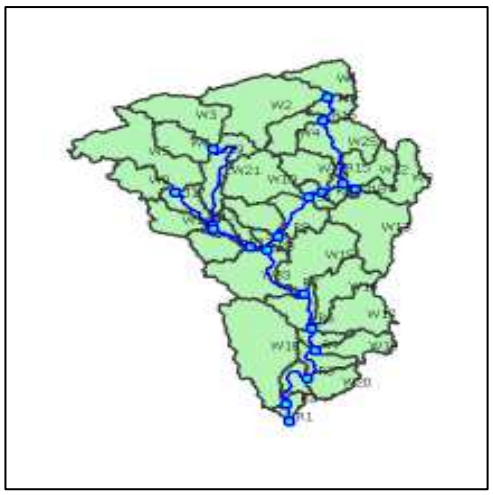
(a) Mosaic Raster



(b) Fill Sinks



(c) Flow Direction



(d) Sub Basins in the study Area

Fig. 3. (a)Mosaic raster. (b) Flow fill in the Mosaic raster. (c) Flow direction in the Mosaic raster. (d) Sub basins in the study area.

4.3. Hydrological Model Setup & HEC-HMS Mathematical Framework

HEC-GeoHMS is a geospatial hydrological toolkit designed specifically for hydrologists to bridge the gap between GIS and hydrologic modeling. It provides a suite of programs—including data management, terrain processing, basin processing, and hydrologic parameter estimation—that facilitate the creation of structural inputs for the HEC-HMS modeling system.As an extension for ArcGIS, HEC-GeoHMS requires the Spatial Analyst extension to be active. In this study, HEC-GeoHMS version 10.x was utilized alongside ArcGIS version 10.x to translate the physical characteristics of the Wainganga basin into a mathematical model. The toolkit was specifically used to:

- Delineate Sub-basins: Subdividing the Wainganga catchment into smaller, manageable hydrologic units based on the processed DEM.
- Extract Physical Parameters: Calculating river lengths, slopes, and basin centroids.
- Estimate Hydrologic Parameters: Assigning initial values for loss and transform methods, such as the SCS Curve Number, based on the soil and land use data of the Vidarbha and Madhya Pradesh regions.
- Generate HEC-HMS Files: Creating the background map and basin files required to run the final rainfall-runoff simulation.

4.3.1. Terrain Processing

The raster-based flow data is converted into a vector format—the language of GIS-based hydrological models—using three primary functions: drainage line processing, adjoint catchment processing, and drainage point processing. The input stream link grid is transformed into a drainage line feature class via the drainage line processing method. Each line in this feature class is assigned a unique catchment identification, precisely mapping the tributaries of the Wainganga River, such as the Bawanthadi or Kanhan, to their respective drainage areas. The aggregated upstream catchments are produced from the catchment feature class by the adjoint catchment processing function. For every catchment in the Wainganga basin that is not a headwater catchment (i.e., those located downstream of the Seoni or Balaghat regions), a polygon depicting the entire upstream area draining to its intake point is created and saved with an "adjoint catchment" tag. This step is critical for expediting the point delineation procedure and ensuring the model accurately calculates the cumulative

runoff at key junctions like the Asthi outlet.The drainage point function is then used to generate the specific nodes where these catchments connect. Finally, Thiessen polygons are plotted to define the area of influence for each rain gauge station across the basin. This process creates a network where any point within a polygon is closer to its associated rain gauge (e.g., the Gondia or Bhandara station) than any other, as illustrated in the resulting Thiessen network (Fig. 8).

4.3.2. Basin Processing

To identify the watershed upstream of each monitoring point in the Wainganga basin, the batch watershed delineation tool in HEC-GeoHMS is utilized. This automated process is essential for defining the contributing drainage areas for the numerous rain gauge and gauging stations across the Madhya Pradesh and Maharashtra border. For specific sites of interest, such as the Asthi outlet or the Balaghat gauging station, the interactive point delineation tool allows for the definition of a single watershed without processing the entire batch. Sub-watersheds are delineated for all points in a chosen feature class using the batch sub-watershed delineation algorithm. A sub-watershed represents the immediate drainage area of a site of interest; it excludes all other upstream areas that are already part of a different sub-watershed. Because the watershed flowing to an upstream point is omitted from the sub-watershed of the downstream point (e.g., the area draining to Bhandara is separated from the area draining to Pauni), the resulting sub-watersheds do not overlap. Finally, flow path tracing is employed to identify the path from a chosen site to the catchment outlet by following the steepest downhill slope. This ensures the model accurately reflects the movement of water through the Wainganga's complex drainage network before it reaches the confluence with the Wardha River.

4.3.3. LULC and Soil Map Integration for CN Parameter Extraction

To detect fundamental spatial changes in the Wainganga basin, such as areas covered by water (reservoirs like Gosekhurd and Sanjay Sarovar), vegetation, dense forests (common in the Balaghat and Gadchiroli districts), and urbanization (near Gondia or Bhandara), a Land Use Land Cover (LULC) map was employed.

The LULC map for the Wainganga study area was produced using the following procedures:

1. Data Acquisition: Decadal land use and land cover classification data for the region were downloaded, focusing on the transition periods relevant to the basin's development.
2. GIS Integration: The datasets were extracted and imported into the ArcGIS environment for spatial analysis.
3. Symbology Management: The symbology was adjusted to visualize all LULC categories, ensuring consistency across multi-temporal files.
4. Spatial Clipping: The delineated Wainganga basin polygon was added to the data view, and the specific LULC data for the study area was clipped using the Extract by Mask tool.
5. Vectorization: The resulting raster data was converted into a shapefile (Raster to Polygon), with symbology updated to reflect grid codes and associated colors.
6. Classification Labeling: Grid codes were renamed and color-coded according to standard classifications (e.g., Forest, Agricultural Land, Water Bodies, and Built-up Areas).

The methodological algorithm produced segmented LULC images showing several color class clusters, along with the calculated percentage of land covered by each feature. This distribution is vital for the HEC-HMS model, as the ratio of forested land in the upper

Wainganga significantly influences the SCS Curve Number (CN) and subsequent runoff velocity.

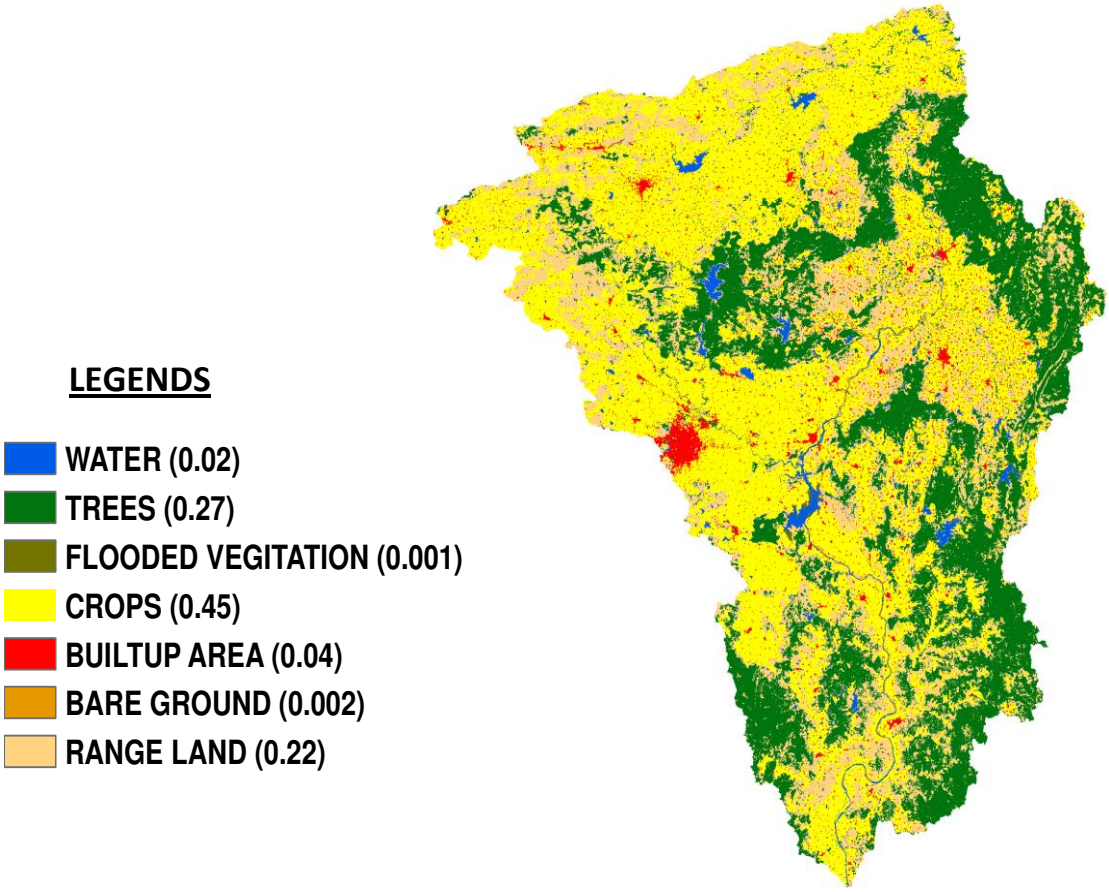


Fig.4 Land Use Land Cover (LULC) Map of Wainganga River Basin

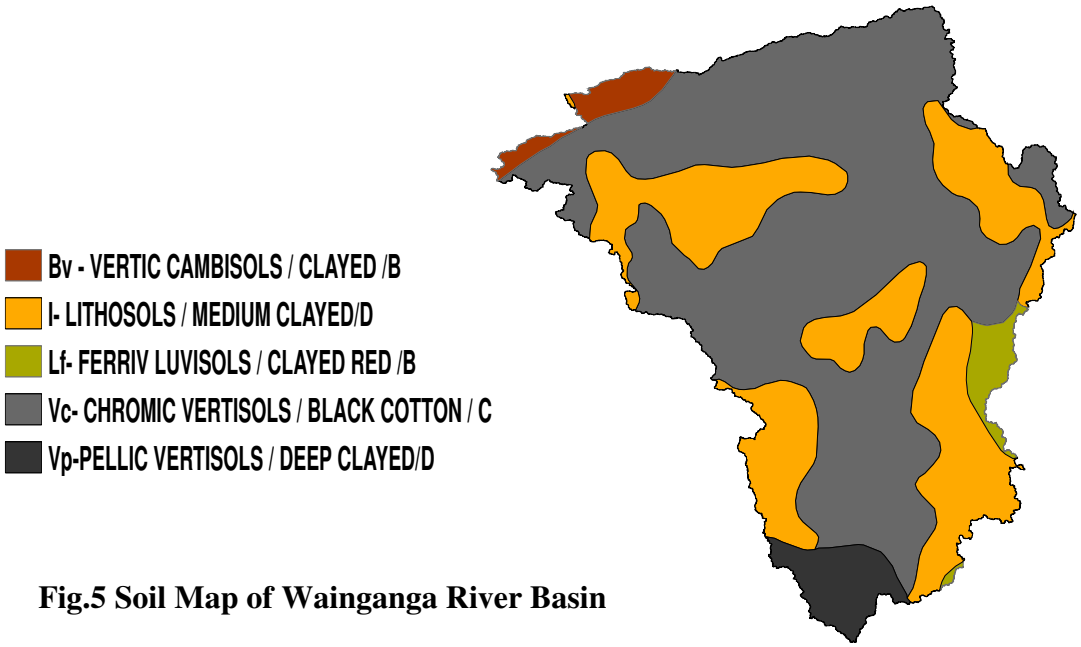


Fig.5 Soil Map of Wainganga River Basin

Table 2: Curve Number for each station corresponding to sub-basin (Wainganga)

Sr. No.	Station Name	Sub-basin Class	Landscape / Land Cover	CN (AMC II)
1	Seoni	C	Deciduous Forest	73
2	Balaghat	D	Paddy Crop (Rice)	89
3	Gondia	D	Mixed Forest + Crop	82
4	Bhandara	D	Urban + Agricultural	88
5	Pauni	C	Water Body + Crop	85
6	Gadchiroli	C	Dense Forest	70
7	Ashti	D	Fallow Land + Crop	86
8	Bamni	C	Crop + Scrubland	79
9	Tirora	D	Industrial + Crop	91
10	Sitekasa	C	Mixed Forest	73
11	Shivni	C	Crop + Forest	77
12	Kumhari	C	Agricultural Land	85
13	Waraseoni	D	Paddy Field	89
14	Armori	C	Forest + Crop	74
15	Chamorshi	C	Crop + Vegetation	81
16	Bhimkund	C	Hilly Forest	70
17	Gosekhurd	D	Open Water + Crop	89

The land use map and soil map as shown in Figure 4 and Figure 5 respectively, were uploaded to the ArcGIS platform to construct the CN (Curve Number) map for the Wainganga basin. The soil data was sourced from the FAO Digital Soil Map of the World (DSMW), providing the necessary hydrological soil groups (HSG) for the Madhya Pradesh and Maharashtra regions. The process involved intersecting the soil and land use layers to create a composite map where each new polygon represents a unique combination of land cover and soil type. Using the Union method in ArcGIS 10.x, the attributes of both layers were merged, allowing for the systematic assignment of CN values based on standard lookup tables (e.g., TR-55).By mapping these specific land covers—such as the dense teak forests of the Satpura Range and the clay-rich paddy fields of Bhandara—to their respective hydrologic soil classifications, a representative CN value was assigned to every polygon in the merged map. For each sub-station corresponding to a Wainganga sub-basin, the final calculated CN numbers are detailed in Table 2 above.

4.4. Hydrological Model

To establish a water balance relationship for the Wainganga sub-basin, it is essential to determine the core physical and hydrological characteristics of the catchment. These parameters are derived through a combination of watershed delineation, the generated CN grid, and raw topographic data. The following steps outline the process of preparing the basin data within ArcGIS using the HEC-GeoHMS toolkit:

1. Terrain Processing: Create a Slope Grid

The initial step involves using the processed DEM to generate a stream network and define the sub-basin boundaries. This terrain processing phase includes the creation of a slope grid, which is vital for calculating the travel time of runoff from the Satpura highlands down to the Vidarbha plains.

2. HEC-GeoHMS Model Development

a) Project Setup & Basin Processing: The selected Wainganga watershed is prepared for processing using spatial data from the initial delineation. Tools such as Project Point and Project Area are used to define the specific study reach—from Sanjay Sarovar to Gosekhurd. Basin processing allows for the manual merging or splitting of sub-basins to match the locations of actual gauging stations like Balaghat or Asthi.

b) Basin Characteristics: The physical attributes of the river reaches and sub-basins, including length, slope, and elevation change, are extracted and populated into the attribute tables.

c) HMS Parameters: Using the hydrologic parameters menu, initial estimates for loss (SCS Curve Number) and transform (SCS Unit Hydrograph) methods are assigned to the watershed elements.

d) Check Data: A verification tool is run to ensure the connectivity of the stream network and the validity of the attribute data. This summary highlights any "dangling" nodes or missing parameters that could cause the model to fail.

e) HMS Schematic: This final GIS tool creates a topological representation of the hydrologic system. It transforms the geographical features into a network of links (reaches) and nodes (junctions/reservoirs).

The physical attributes of the Wainganga sub-basins were effectively delineated and transformed into parameters compatible with the HMS model. This created a structured Basin Model containing specific coordinate data and connectivity logic. The resulting model, accompanied by a background shape file for visual reference, was then exported from HEC-GeoHMS for final simulation in the HEC-HMS software environment.

4.5. Importing Basin Model in HEC-HMS

The basin model, successfully structured within HEC-GeoHMS, was imported into the HEC-HMS 4.3 (or later) environment for detailed hydrologic analysis. This transition converts the GIS-based spatial data into a functional computational network. Figure 6 displays the imported model of the Wainganga basin, featuring all essential hydrologic elements: sub-basin areas, reaches representing the main river and its tributaries (like the Bawanthadi and Kanhan), junctions for flow merging, and the Asthi outlet as a reservoir element.

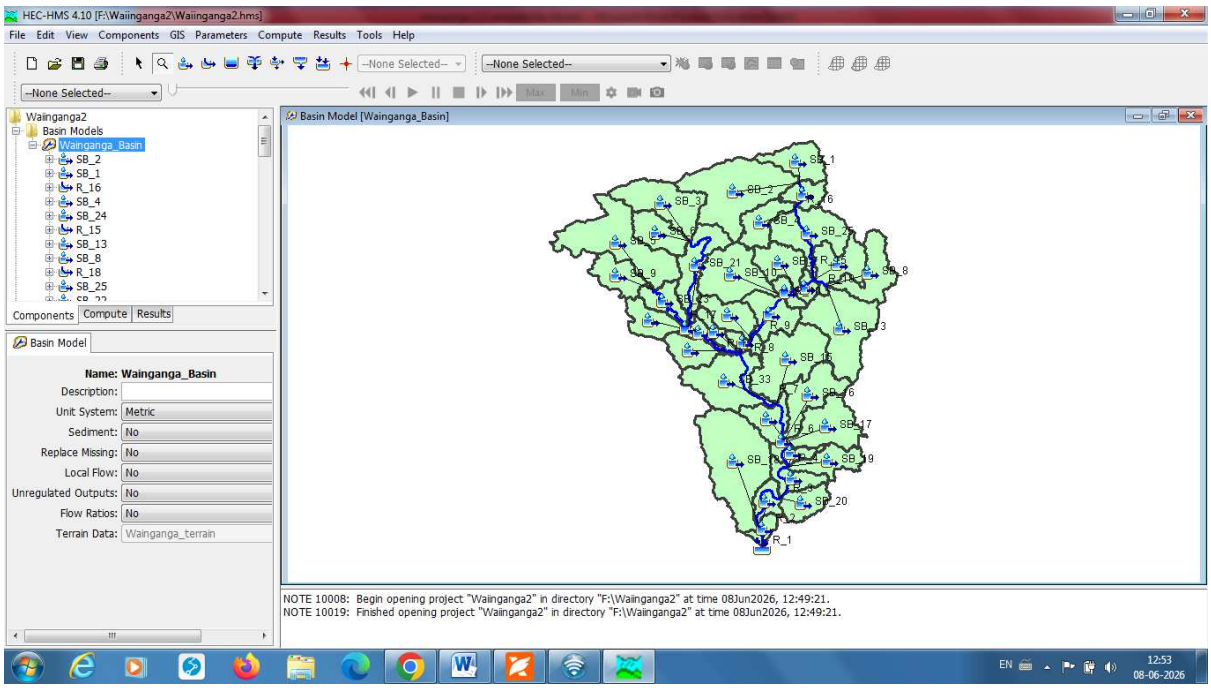


Fig. 6 Basin Model imported in HEC-HMS

4.6. **Parameter Calibration, Validation & Evaluation Strategy**

Once the hydrologic framework was established, simulations were conducted to align the model's performance with real-world observations. The process was divided into two critical phases:

- Calibration: The model was calibrated using observed monsoon flow data from the years 2001 to 2010. During this phase, parameters such as the SCS Curve Number (CN) and Lag Time (T_lag) were iteratively adjusted to match the simulated hydrograph peaks and volumes with the observed discharge at the Asthi outlet gauging station.
- Validation: The model was assessed using observed flow data from 2011–2017 to verify its capability in simulating independent events, without implementing any additional parameter modifications during the validation phase.

Upon successful validation, the model's performance was statistically quantified. The Nash-Sutcliffe Efficiency (NSE) and Root Mean Square Error (RMSE) were calculated using regression analysis to assess the accuracy of the Wainganga basin rainfall-runoff simulation. High NSE values closer to 1.0 indicate a strong correlation between the HEC-HMS predicted runoff and the actual river discharge recorded during the monsoon seasons.

4.6.1. Regression analysis

To ascertain the connection between a dependent variable and one or more independent variables, regression analysis is a statistical technique. Regression analysis is used to evaluate the model's accuracy in hydrological modelling by comparing the simulated and observed values. The goal is to determine how closely the model fits the observed data. The linear regression model is used to determine the minimal distance between the fitted line and all observed data points. This method, known as ordinary least squares (OLS) regression, reduces the sum of the squared residuals. The closer the sum of the squared residuals is to zero, the better the fit of the model to the observed data.

4.6.2. Correlation between observed and simulated discharge

The degree to which the estimated and observed data match each other is measured using the coefficient of determination, or R-squared. R-squared measures how much of the variance in the response variable can be accounted for by a linear model. R-squared is determined by comparing the variance explained by the model with the total variance present in the dataset. Its value ranges from 0% to 100%, where 100% indicates that the model perfectly accounts for all variability in the response data around the mean, while 0% signifies that the model is unable to explain any variability. Therefore, a higher R-squared value reflects a better model fit and improved predictive performance.The ability of the model to be employed in hydrologic simulation in the Wainganga watershed at the Gisikhurd dam site is demonstrated by the comparison of observed and simulated hydrographs with a high R squared value.

4.6.3. Root Mean Square Error (RMSE)

The Root Mean Square Error (RMSE) is a widely used metric for assessing the accuracy of a model or an estimator in predicting values compared to observed values. In this study, it serves as a measure of the "residual" or the standard deviation of the prediction errors.For hydrological modeling, the RMSE is calculated as the square root of the second sample moment of the differences between predicted and observed values. A lower RMSE value indicates a better fit, signifying that the HEC-HMS model's predictions are closely aligned

with the actual flow data provided by the Central Water Commission (CWC).For peak discharge, the RMSE is determined using the following equation:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(q_i - Q_i)^2} \quad (1)$$

Where:
Qi= Measured stream flow

qi = Simulated stream flow

Q = Average Measured stream flow

n = Total number of time steps in calibration period.

4.6.4. Nash-Sutcliffe Efficiency (NSE)

In addition to RMSE, the Nash-Sutcliffe Efficiency (NSE) was calculated to determine how well the model predicts the discharge relative to the mean of the observed data. For the Wainganga basin, an NSE value above 0.75 is generally considered a "very good" fit for rainfall-runoff modelling.

$$\text{NSE}\ (\eta) = 1 - \frac{\sum_1^n (q_i - Q_i)^2}{\sum_1^n (Q_i - Q)^2} \quad (2)$$

Interpretation of NSE Results

The NSE values for the Wainganga basin model are interpreted as follows:

- NSE = 1: Indicates a perfect match between the HEC-HMS modeled flow and the observed CWC discharge data.
- NSE = 0: Suggests that the model predictions are only as accurate as the mean of the observed data (Q_mean).
- NSE < 0: Indicates that the observed mean is a better predictor than the model, signaling a need for further calibration of the SCS Curve Numbers or Lag Time.

The NSE provides a quantitative assessment of the model's predictive accuracy. Because it is a normalized statistic, it allows for a direct comparison of the Wainganga model's performance against other regional hydrological models (such as those for the Wardha or Kanhan sub-basins), provided that consistent observed flow data is available for comparison.

5. Result and Discussion

5.1. Hierarchical Clustering

Hierarchical clustering was employed to categorize the rain gauge stations of the Wainganga basin based on their climatic and geographical similarities. The resulting dendrogram (Fig. 7) visually represents the nested relationships and nested groupings of these stations. Using the agglomerative (bottom-up) approach, each station was initially treated as an individual cluster and successively merged with its most similar neighbour until a cohesive structure was formed.

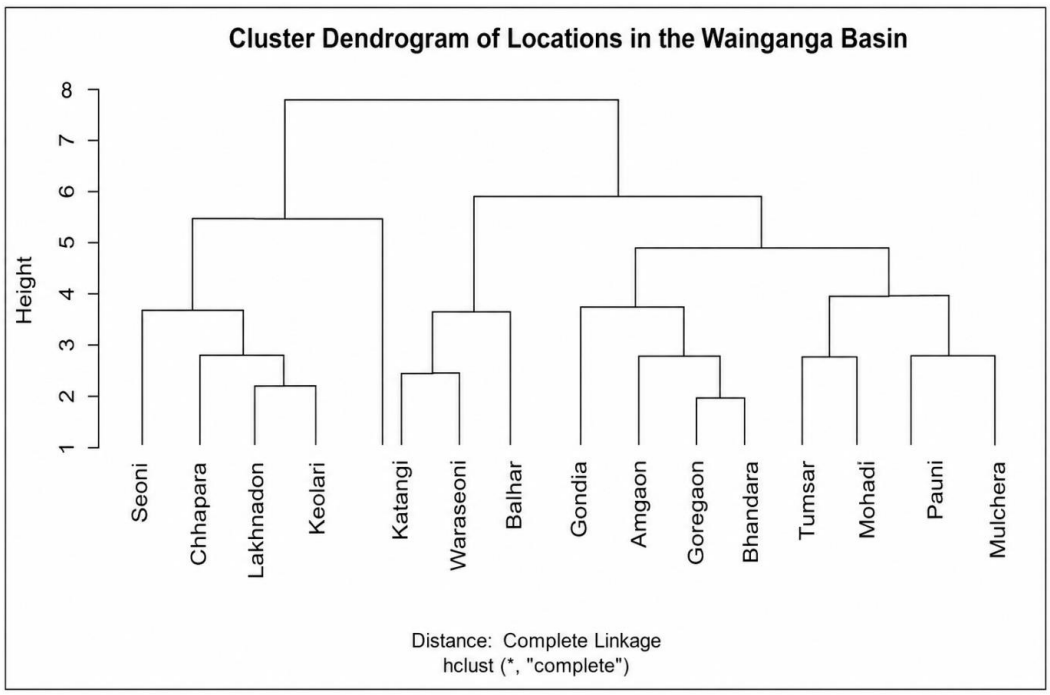


Fig. 7. Cluster Dendrogram of locations in Wainganga Basin using complete linkage distance.

In the context of the Wainganga sub-basin, the dendrogram revealed two primary clusters based on rainfall intensity and topographical elevation (transitioning from the Satpura hills to the Vidarbha plains):

- Group 1 (Hilly/High Rainfall): This group consists of five stations: Seoni, Balaghat, Waraseoni, Katangi, and Bhimkund. These stations exhibit a strong correlation in rainfall patterns due to their proximity to the upper catchment's forested regions. Specifically, the rainfall characteristics at Katangi closely match those of the Waraseoni and Bhimkund clusters. Furthermore, the grouping of these three is highly similar to the Balaghat station, followed by Seoni.
- Group 2 (Plains/Moderate Rainfall): The second major group includes the Gondia, Bhandra, Tirora, Pauni, Armori, Gadchiroli, Ashti, and Chamorshi rain gauge stations. Within this cluster, the rainfall features of Bhandara are comparable to Pauni (reflecting the middle-reach river characteristics), while Gondia shows similarities to Tirora, and the downstream stations like Ashti align closely with Chamorshi.

The clustering results suggest that the Wainganga basin can be hydrologically divided into an upper high-intensity zone and a middle-to-lower regulated zone. This classification is crucial for the HEC-HMS model, as it justifies the use of different initial abstraction values and lag times for the two distinct groups. The high similarity between stations like Bhandara and Pauni confirms that a single representative station's data can be used to fill gaps in missing records for nearby stations within the same cluster.

5.2. Spatial Rainfall Characterization via Thiessen Polygons

The Thiessen polygon method, developed by A.H. Thiessen in 1911, is a widely adopted technique for determining the mean areal precipitation for a catchment from point-source rain gauge data. This approach operates on the assumption that the observed precipitation at any specific station is representative of the rainfall distributed halfway to each neighboring station in all directions. Consequently, the rainfall depth is assumed to be uniform across the entire area of the polygon surrounding a specific gauge.The weights assigned to each rain gauge in

the Wainganga basin are determined by the relative areas of these polygons. The network is constructed by drawing perpendicular bisectors of the lines connecting adjacent stations, such as Bhandara to Gondia or Seoni to Balaghat. The surface area of each resulting polygon is then used to calculate the weighted rainfall contribution for the station located at its centroid.One operational constraint of this method is that if data from any station (e.g., Tirora) is missing for a specific monsoon event, the pertinent polygons must be recalculated to reassign the area to neighbouring gauges. Furthermore, while the Thiessen method is effective for the relatively flat plains of the Vidarbha region, it does not account for orographic rainfall variations caused by the hilly terrain of the Satpura Range in the upper catchment. The seventeen polygons that constitute the Wainganga research region in Fig. 8 are based on the unique influence zone surrounding each of the sample points listed in Table 1.

5.3. Hydrological Modelling

In the present study, rainfall-runoff simulations were executed using daily precipitation data. Rainfall records from the years 2001- 2010 were utilized to calibrate the HEC-HMS model, while data from 2011-2017 were reserved to verify the model's predictive reliability for the Wainganga basin.

5.3.1. Comparing the Observed and Simulated Flow

Model calibration is the iterative process of modifying the model's physical and empirical parameters such as the SCS Curve Number (CN) and Lag Time (T_lag) to identify fixed parameter values that yield the best estimate of the observed discharge.

In this study, the calibration was performed at two critical hydrological nodes within the Wainganga system:

1. Upstream Calibration (Bhandara): The first stage involved calibrating the model at the bhandara gauging station, which serves as a key monitoring point for the upper and middle reaches of the Wainganga. This ensures that the runoff generated from the Seoni and Balaghat sub-basins is accurately represented before it reaches the lower plains.
2. Downstream Calibration (Asthi outlet): The second stage focused on the Asthi outlet site. Calibration here involved adjusting parameters to match the total reservoir inflow, accounting for the cumulative discharge from the main stem and major tributaries like the Kanhan and Pench rivers.

The alignment of the simulated peak discharge and base flow with the observed CWC (Central Water Commission) data during these high-flow years confirms the model's ability to replicate the monsoon dynamics of the Vidarbha region.

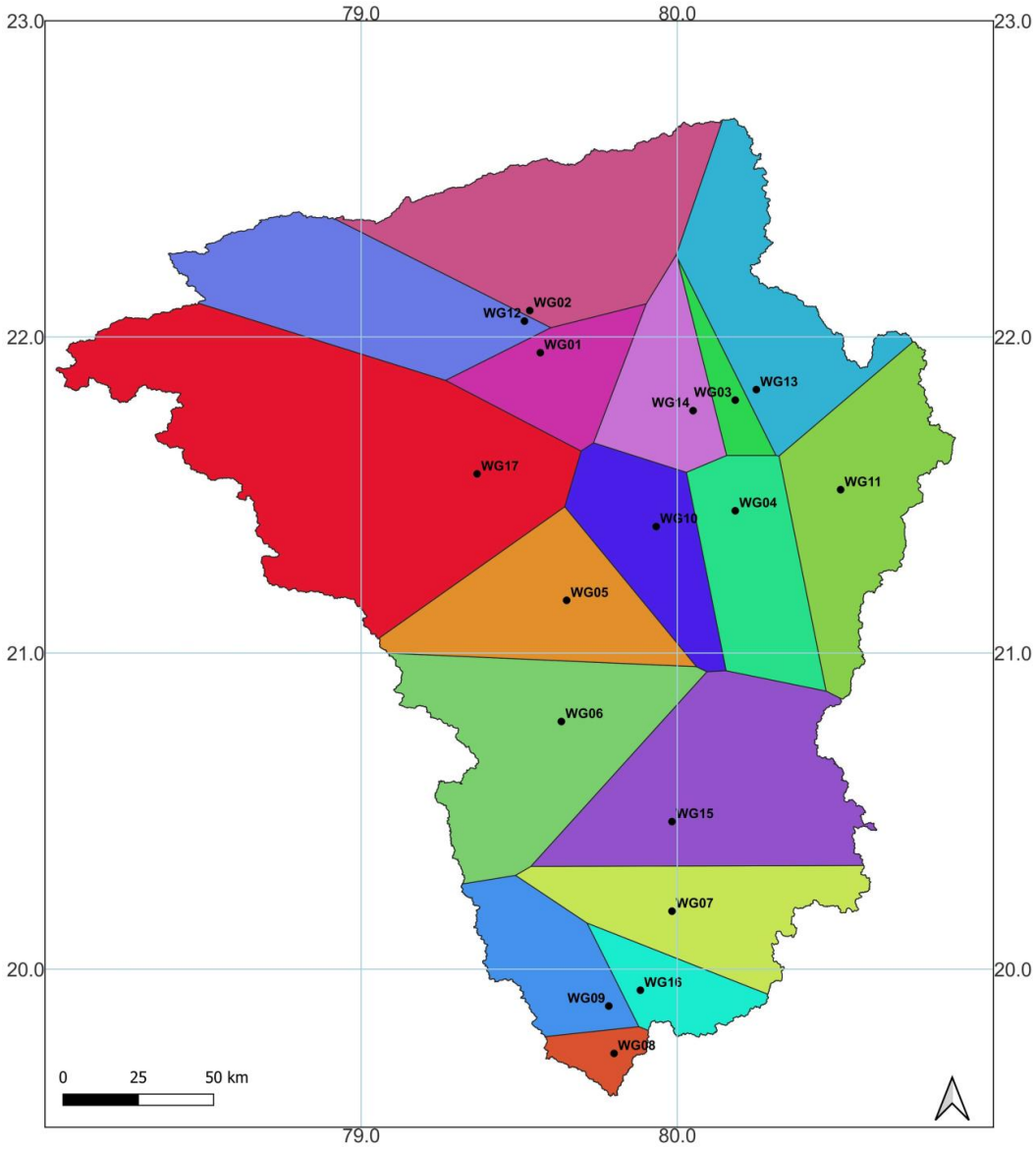


Fig 8. Thiessen Polygon

5.3.2. Comparison of Model Calibration and Validation at the Asthi Gauging Site

In order to calibrate the model at the Asthi outlet gauging site, which serves as a primary control point for the middle Wainganga, rainfall data was integrated from stations at Seoni, Balaghat, Gondia, and Tirora.

Calibration Results (2001-2010)

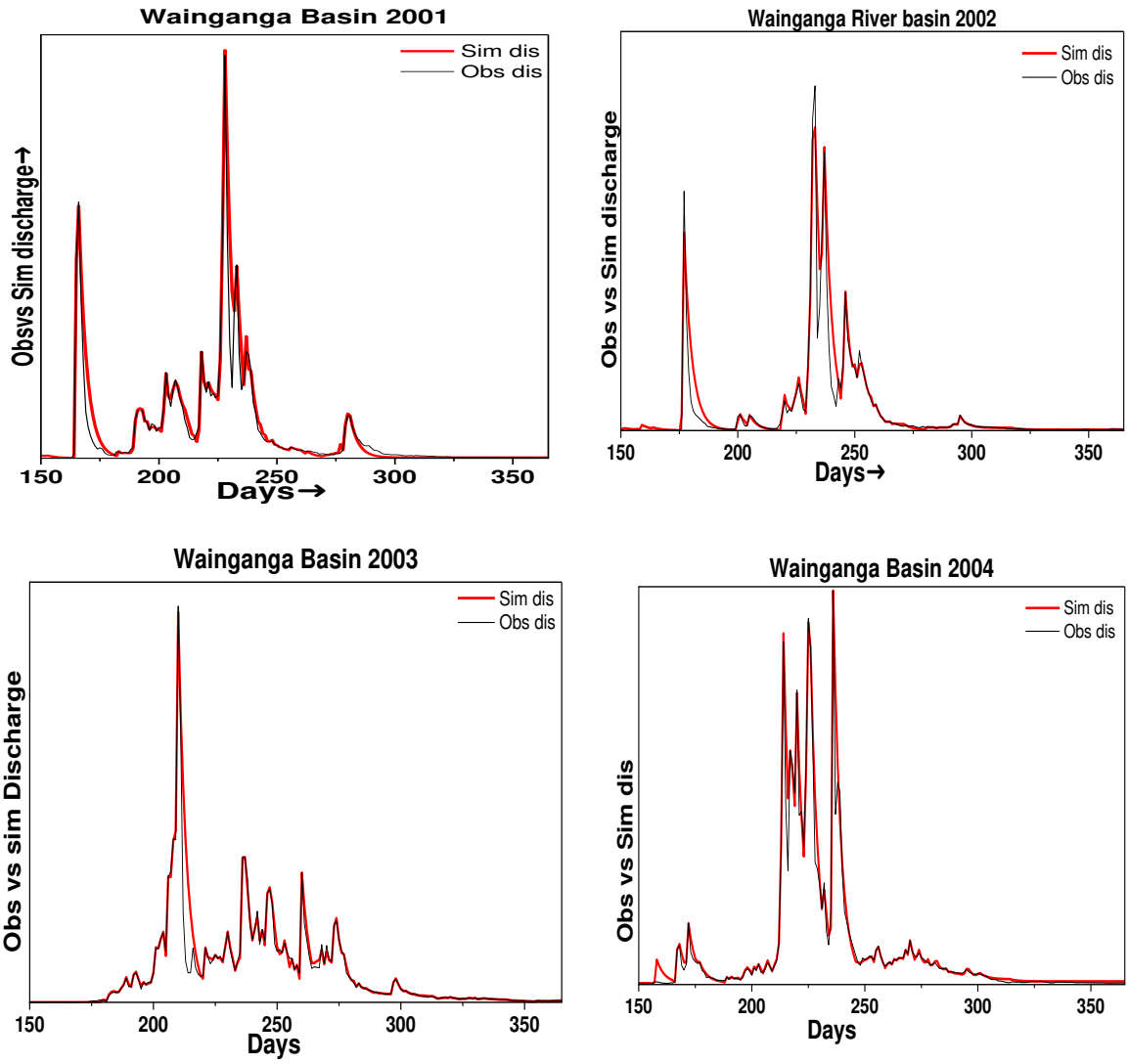
The HEC-HMS hydrological model was calibrated using daily monsoon rainfall-runoff data spanning a ten-year period from 2001 to 2010 as shown if figure 9. To establish reliable boundary conditions for the middle and lower reaches of the Wainganga River system, precipitation inputs were integrated from key representative stations including Seoni, Balaghat, Gondia, and Tirora.

The summary metrics in Table 3 demonstrate that the calibrated model captures the basin's hydrologic response with high precision. The Nash–Sutcliffe Efficiency (NSE) values during

this decade range from a minimum of 0.895 (in 2007) to a maximum of 0.962 (in 2004). The coefficient of determination (R^2) consistently mirrors these trends, remaining firmly above 0.895 across all ten calibration years. This signifies that the model accounts for approximately 90% to 96% of the variance observed in the physical gauging data.

The Root Mean Square Error (RMSE), which quantifies the standard deviation of the residuals, remains low throughout the calibration timeframe. The lowest prediction error is recorded in 2004 with an RMSE of 0.2056. Conversely, the highest error is noted in 2005 with an RMSE of 1.3620, a variation primarily attributed to heavy localized peak flow fluctuations during that specific monsoon season.

Overall, the iterative optimization of core empirical parameters—specifically the SCS Curve Numbers (CN) and basin Lag Times (T_{lag) successfully minimized the sum of squared residuals via ordinary least squares regression. This rigorous calibration ensures that upstream runoffs from the rugged Satpura highlands down to the regulated Vidarbha plains are mathematically balanced prior to downstream routing.



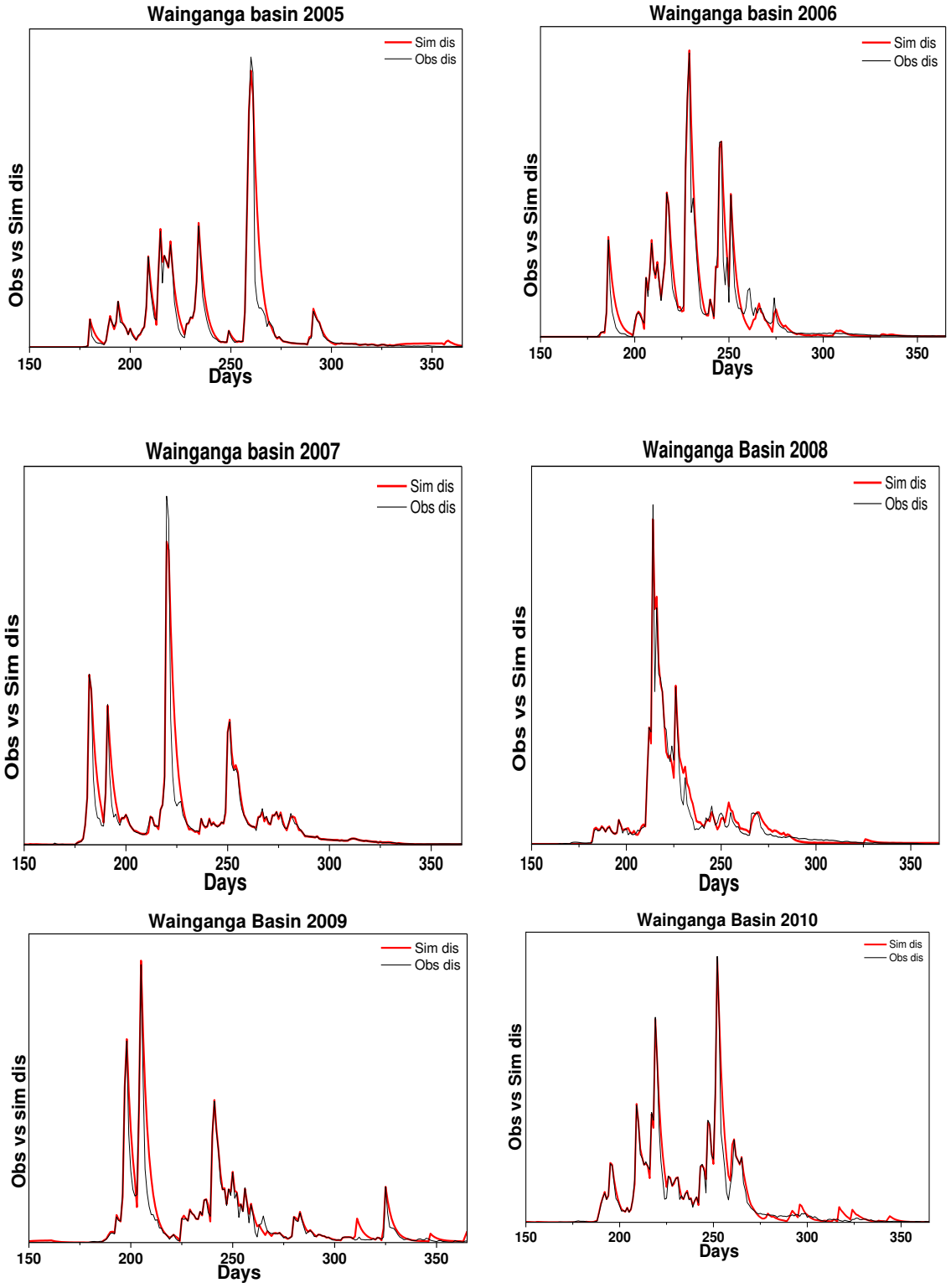


Fig. 9: Observed Vs Simulated Flow calibration Period (2001-2010)

Validation Results (2011- 2017)

Following calibration, the fixed model configuration was evaluated against an independent seven-year dataset from 2011 to 2017 to verify its predictive stability without any further parameter modifications. The validation phase confirms that the integrated spatial-statistical

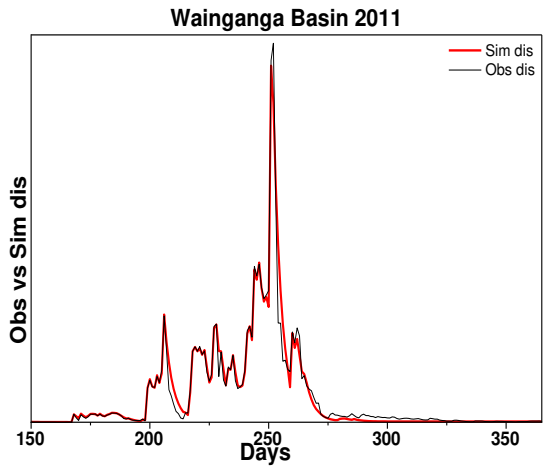
approach retains an excellent level of simulation accuracy. The graphical representation simulated verses observed discharger for validation period is shown in figure 10.

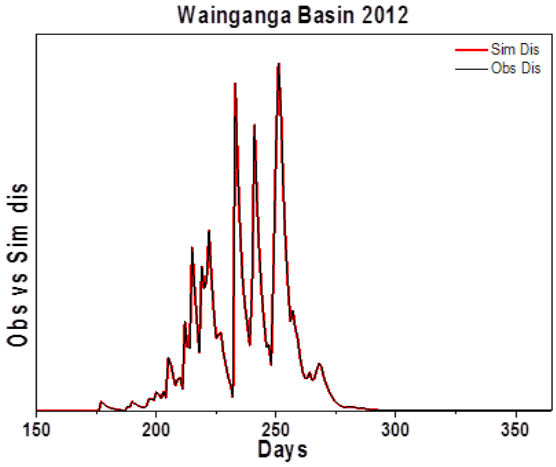
According to the regression analysis parameters shown in the validation data:

- **Nash–Sutcliffe Efficiency (NSE):** Ranges between 0.906 (2014) and 0.980 (2012). An NSE value approaching 1.0 formally establishes that the HEC-HMS model's simulated runoff provides a near-perfect replication of real-world discharge compared to a basic statistical mean.
- **Coefficient of Determination (R^2):** Exceeds 0.90 across all years, peaking at 0.9720 in 2013. This confirms a highly reliable linear relationship between the simulated and observed hydrograph trends over a multi-year validation horizon.
- **Root Mean Square Error (RMSE):** Concentrated in a tight, low range between 0.3296 (2017) and 0.8707 (2013). These consistently compressed values highlight the model's localized accuracy and steady performance across separate, independent monsoon periods.

Visual analysis of the generated hydrographs further supports these statistical conclusions:

- **Wainganga Basin 2012 Hydrograph (Fig. 10):** The simulated discharge curve (**Sim DIS** in red) shows an exceptional match with the observed discharge (**Obs DIS** in black) across the entire monsoon timeline. The model precisely replicates the steep rising limbs, multi-peak flash responses between days 210 and 250, and the subsequent natural baseflow recession phase.
- **Wainganga Basin 2016 Hydrograph (Fig. 10):** The peak discharge magnitudes and temporal alignments for major high-flow windows (specifically the major sharp peak near day 190 and subsequent secondary peaks) exhibit a high degree of conformance. While a minor over-estimation of runoff volume occurs during a short mid-recession interval (around days 225–235), the overall wave shape and total discharge volume align closely with Central Water Commission records.





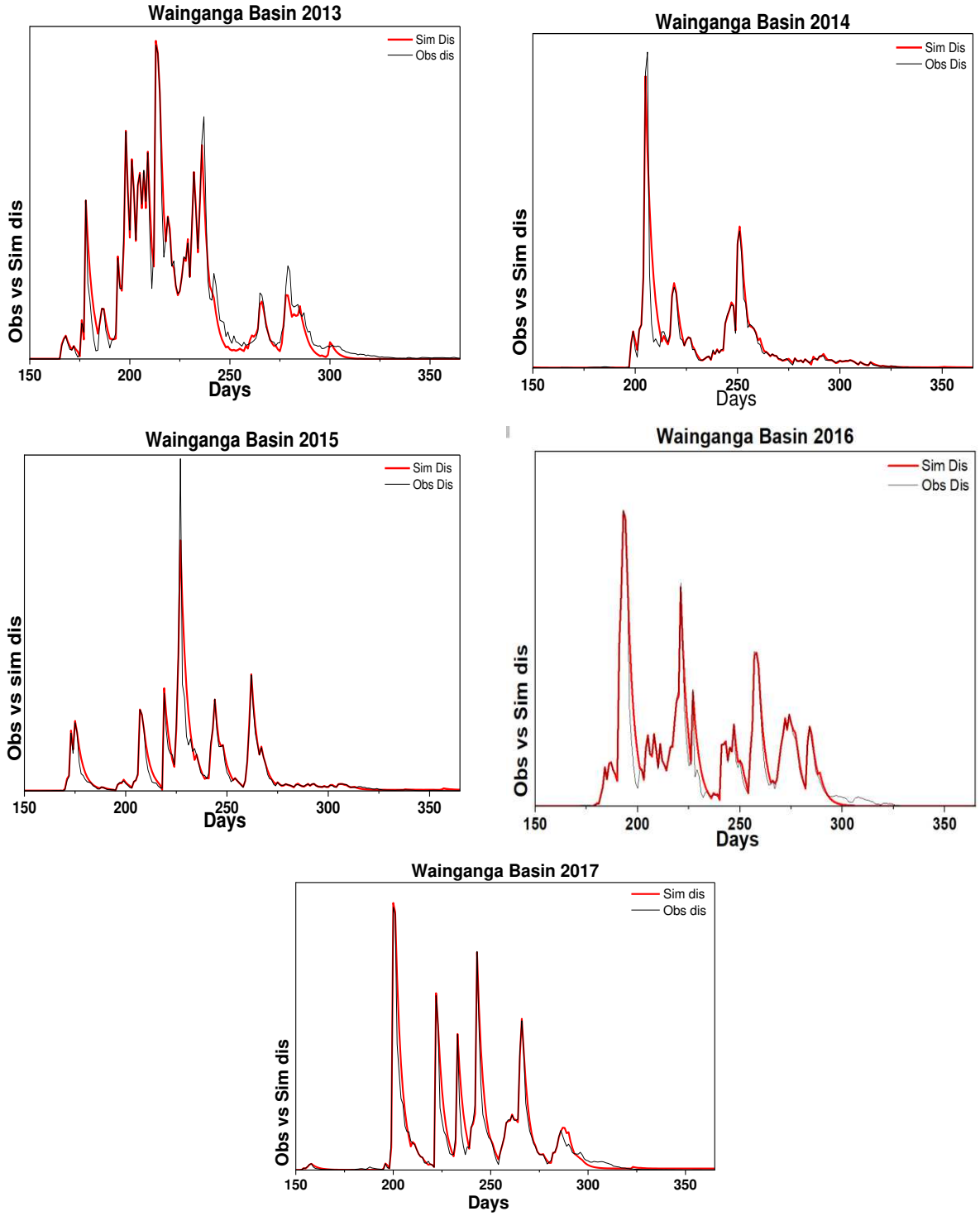


Fig. 10: Observed Vs Simulated Flow Validation Period (2011-2017)

5.4. Results of Regression Analysis

Regression analysis was employed to quantitatively assess the model's performance by calculating the Root Mean Square Error (RMSE) and the Coefficient of Correlation (R2) for two primary monitoring sites. The statistical results are tabulated in Table 3 & Table 4 for the Asthi outlet gauging site.

Table 3 Result Summery for Basin Outlet (Asthi) Calibration Period

Year	NSE	R^2	RMSE
2001	0.92	0.9219	0.7382
2002	0.897	0.8975	0.8366
2003	0.912	0.9122	0.7176
2004	0.962	0.9625	0.2056
2005	0.90	0.9002	1.3620
2006	0.931	0.9316	0.8227
2007	0.895	0.8953	1.047
2008	0.959	0.9599	0.3472
2009	0.913	0.9131	0.4081
2010	0.956	0.9566	0.7030

The values in these tables indicate that the developed HEC-HMS model for the Wainganga basin performs exceptionally well. This is evidenced by the coefficient of correlation consistently remaining above 0.91 and the RMSE being relatively small across all calibration and validation scenarios. The integration of the hydrological model with the Thiessen polygon method generated highly accurate predictions of flow at the Asthi outlet site. This is supported by low RMSE values and high coefficients of correlation (R2 > 0.90) for all monsoon flow data analyzed between 2001 and 2017. These results suggest that the developed model effectively reflects real-world flow dynamics and can be reliably used for flood forecasting and water resource management in the Godavari-Wainganga system.

Table 4 Result Summery for Basin Outlet (Asthi) Validation Period

Year	NSE	R^2	RMSE
2011	0.970	0.9702	0.5003
2012	0.98	0.9549	0.7559
2013	0.971	0.9720	0.8707
2014	0.906	0.9069	0.7827
2015	0.915	0.9151	0.4750
2016	0.955	0.9553	0.5975
2017	0.941	0.9410	0.3296

6. Conclusion

This study demonstrates the efficacy of coupling unsupervised machine learning, traditional spatial weight distribution, and semi-distributed hydrological modelling to capture complex

monsoon river dynamics. By applying Hierarchical Clustering (HC), the rain gauge network was systematically categorized into two distinct zones: an upper high-intensity catchment layer and a middle-to-lower regulated zone. This classification justified varied parameter initialization, while the Thiessen polygon framework accurately computed area-weighted rainfall across changing terrains.

The HEC-HMS model achieved high simulation precision at the primary control node (Asthi outlet). Across a 17-year continuous analysis period (2001–2017), the statistical metrics consistently indicated a superior fit, characterized by NSE values above 0.895, R^2 values exceeding 0.895, and minimized RMSE values across both calibration and validation operational phases. The close alignment of simulated hydrograph shapes and seasonal peak volumes confirms that the model reliably reflects the real-world physical runoff behaviors of the Godavari-Wainganga system. Consequently, this validated computational framework provides a dependable tool for operational flood forecasting, early warning networks, and long-term water resource development planning in central India.

Limitations

Despite the satisfactory performance metrics, several structural and operational constraints affect the predictive accuracy of this hydrological model:

- **Data Reliability:** The overall analytical output of the HEC-HMS environment is intrinsically dependent on the continuous quality, precision, and completion of historical field logs provided by external agency monitoring stations.
- **Station Density:** While the chosen network was sufficient to satisfy statistical validation thresholds, expanding the physical density of rain gauges within the complex, topographically rugged areas of the Satpura Range would reduce spatial interpolation errors.
- **Thiessen Polygon Assumptions:** The mathematical premise that rainfall is uniformly distributed across an entire polygon boundary ignores micro-climatic variations and regional orographic lifting effects present in hilly upper catchments.
- **Calibration Data Limitations:** The scope of this study was bound by the availability of high-quality, synchronized data. Integrating a longer, extended time series beyond the evaluated 2001–2017 block would help assess the model's capabilities under extreme, prolonged droughts or catastrophic historical flood stages.

Author Contributions: All authors contributed to the conception and design of the study. Data collection and analysis were performed by **Dr. Rajesh Kherde**. **Dr. Gauri Patil** drafted the manuscript. **Dr. Kiran More** prepared the charts, tables, and figures. **Dr. Priyadarshi Sawant** assessed the results and contributed to the interpretation of the findings and the formulation of the conclusions. All authors critically reviewed, revised, and approved the final version of the manuscript.

Funding
This research received no external funding.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data will be made available on request.

References

[1] Y. Davda, R. Khilari, and S. Gaikwad, "AI-Powered Web Solution for Disaster Risk Management," in *Ecological and Human Dimensions of AI-Based Supply Chain*, A. Taghipour, Ed., Hershey, PA, USA: IGI Global Scientific Publishing, 2025, pp. 175–194. doi: 10.4018/979-8-3693-7478-8.ch008.

[2] G. Patil *et al.*, "Swat-Based Hydrological Assessment of the Wainganga River Catchment," *Water Conserv. Manag.*, vol. 9, no. 3, pp. 461–466, 2025, doi: 10.26480/wcm.03.2025.461.466.

[3] A. T. N. Dang and L. Kumar, "Application of remote sensing and GIS-based hydrological modelling for flood risk analysis: a case study of District 8, Ho Chi Minh city, Vietnam," *Geomatics, Nat. Hazards Risk*, vol. 8, no. 2, pp. 1792–1811, 2017, doi: 10.1080/19475705.2017.1388853.

[4] G. Patil and R. Kherde, "Hydrological Modeling of Large River Basin using Soil Moisture Accounting Model and Monte Carlo Simulation," *Trends Sci.*, vol. 21, no. 6, pp. 1–16, 2024.

[5] A. Kunnath-Poovakka and T. I. Eldho, "A comparative study of conceptual rainfall-runoff models GR4J, AWBM and Sacramento at catchments in the upper Godavari river basin, India," *J. Earth Syst. Sci.*, vol. 128, no. 2, 2019, doi: 10.1007/s12040-018-1055-8.

[6] V. Bhoopathy, "EXPLORING INNOVATIVE APPROACHES FOR INTEGRATED WATER RESOURCE MANAGEMENT IN ARID REGIONS : ENHANCING RESILIENCE TO CLIMATE CHANGE AND WATER SCARCITY," vol. 2032, no. 6, pp. 2020–2032, 2024.

[7] A. A. Salunkhe, R. Gobinath, and S. Makkar, "Chapter 19 - Soft computing applications in rainfall-induced landslide analysis and protection—Recent trends, techniques, and opportunities," in *Computers in Earth and Environmental Sciences*, H. R. Pourghasemi, Ed., Elsevier, 2022, pp. 271–287. doi: https://doi.org/10.1016/B978-0-323-89861-4.00036-1.

[8] R. V. Kherde and P. H. Sawant, "Parameter Optimization, Uncertainty Estimation and Sensitivity Analysis in Hydrological Modeling," *Eur. J. Eng. Technol. Res.*, vol. 3, no. 11, pp. 66–72, 2018, doi: 10.24018/ejeng.2018.3.11.907.

[9] P. Rathod, K. Borse, and V. Manekar, "Simulation of Rainfall -Runoff Process Using HEC-HMS Simulation of Rainfall - Runoff Process Using HEC-HMS (Case Study : Tapi River , India)," no. January 2016, 2015.

[10] G. l. Patil and R. Kherde, "Assessment of Large River Basin Approaching GIS and Computation of Simulation Techniques Using Latest Software," *Ecol. Eng. \& Environ. Technol.*, vol. 25, no. 1, pp. 360–368, 2024, doi: 10.12912/27197050/175753.

[11] I. M. Chathuranika *et al.*, "Comparison of Two Hydrological Models, HEC-HMS and SWAT in Runoff Estimation: Application to Huai Bang Sai Tropical Watershed, Thailand," *Fluids*, vol. 7, no. 8, Aug. 2022, doi: 10.3390/fluids7080267.

[12] R. Singh *et al.*, "Hydrological modeling in India," *Proc. Indian Natl. Sci. Acad.*, vol. 86, no. 1, pp. 479–494, 2020, doi: 10.16943/ptinsa/2020/49802.

[13] J. O. Oleyiblo and Z. Li, "Application of HEC-HMS for flood forecasting in Misai and Wan ' an catchments in China," *Water Sci. Eng.*, vol. 3, no. 1, pp. 14–22, 2010, doi: 10.3882/j.issn.1674-2370.2010.01.002.

[14] A. Samavati and S. W. Baran, "Hydrological modeling for simulation of rainfall-runoff process in mountainous catchments Hydrological modeling for simulation of rainfall-runoff process in mountainous catchments," no. February, pp. 0–14, 2023.