Assessing Hydrologic Uncertainty with an Event-Based Conceptual Model for Ensemble Flood Forecasting

Submitted in partial fulfilment of the requirement for the award of the degree of

Doctor of Philosophy

By

Velpuri Manikanta

Roll No. 701911

Supervisor

Prof. N.V. Umamahesh

DEPARTMENT OF CIVIL ENGINEERING
NATIONAL INSTITUTE OF TECHNOLOGY
WARANGAL-506004, INDIA
August 2023

APPROVAL SHEET

This dissertation work entitled "**Assessing Hydrologic Uncertainty with an Event-Based Conceptual Model for Ensemble Flood Forecasting**" by **Mr. Velpuri Manikanta** is approved for the degree of Doctor of Philosophy.

Examiners

_____

_____

_____

Supervisor

Prof. N.V. Umamahesh

Department of Civil Engineering, NIT Warangal

Chairman

Professor & Head

Department of Civil Engineering, NIT Warangal

Date:

NATIONAL INSTITUTE OF TECHNOLOGY
WARANGAL

CERTIFICATE

This is certify that the thesis entitled "**Assessing Hydrologic Uncertainty with an Event-Based Conceptual Model for Ensemble Flood Forecasting**" being submitted by **Mr. Velpuri Manikanta** for award of the degree of Doctor of Philosophy to the Faculty of Engineering and Technology of National Institute of Technology Warangal is a record of bonafide research work carried out by him under my supervision and this thesis has not been submitted elsewhere for award for any other degree.

(**N.V. Umamahesh**)

Thesis supervisor

Professor

Department of Civil Engineering

National Institute of Technology Warangal

Declaration

I hereby declare that the work presented in the thesis entitled "**Assessing Hydrologic Uncertainty with an Event-Based Conceptual Model for Ensemble Flood Forecasting**" is my original and bona fide research work conducted under the guidance and supervision of **Prof. N. V. Umamahesh**. This work has not been previously submitted elsewhere to obtain any degree.

I affirm that this written submission reflects my ideas and thoughts, expressed in my own words. In cases where I have included ideas, concepts, or direct quotations from other sources, I have provided appropriate citations and references to acknowledge the original authors and sources. I have also adhered to the principles of academic honesty and integrity throughout this research, ensuring that I have not misrepresented or fabricated any information, data, facts, or sources.

I acknowledge that any violation of the aforementioned principles may result in disciplinary actions by the Institute. Moreover, it is understood that improper citation or failure to obtain necessary permissions for the use of external sources can lead to legal consequences.

_____

(Name of the Student: **Velpuri Manikanta**)

(Roll No. 701332)

Date:_____

Abstract

Floods are widespread natural disasters that can severely impact communities, disrupting day-to-day activities. Unlike other natural disasters, floods can be predicted in advance, allowing for preparedness measures. Forecasting flood events relies on observed precipitation as input, but longer lead-time forecasts require transforming quantitative precipitation forecasts obtained from Numerical Weather Prediction models (NWPs) into flood hydrographs using hydrological models. The integration of hydrological and meteorological models forms the basis of a hydro-meteorological forecasting system. Traditional flood forecasting systems operate on deterministic forecast values, providing a single value for each forecast lead time. However, this approach lacks a conclusive estimate of forecast uncertainty, limiting its effectiveness. In response, operational flood forecasting systems have shifted towards the adoption of ensemble forecasts, generating multiple plausible future weather variables states. Ensemble flood forecasts offer probabilistic information that outperforms deterministic forecasts, particularly for longer lead times. Despite the increasing adoption of ensemble weather prediction systems, India's current flood forecasting systems predominantly rely on deterministic approaches and overlook the inherent uncertainties in flood forecasts, particularly concerning hydrologic prediction. These hydrologic uncertainties encompass various aspects, such as model structural uncertainty, parameter uncertainty, uncertainty in the spatial resolution of models, and uncertainty in estimating initial hydrologic conditions. It is imperative to address these uncertainties to establish a robust ensemble flood forecasting system that enhances flood risk management and facilitates more effective mitigation strategies.

This thesis focuses on developing an ensemble flood forecasting framework for the Godavari River Basin, incorporating hydrologic uncertainties within the forecasting process using an event-based conceptual model. The primary objective includes identifying the most suitable ensemble weather forecast products and post-processing methods for the GRB through a systematic verification study, facilitating the creation of a dependable ensemble flood forecasting system. Furthermore, the research aims to investigate the impact of model resolution on flood peak simulation by employing event-based semi-distributed models, thereby enhancing the accuracy of flood predictions. Additionally, the thesis examines the compatibility of reanalysis-based and continuous model-simulated soil moisture products with conceptual models and evaluates the accuracy of methods used for estimating initial hydrologic conditions in event-based models. By addressing

these objectives, the thesis contributes to the advancement of ensemble flood forecasting techniques, ultimately improving flood risk management strategies in the Godavari River Basin.

The initial part of this thesis focuses on the verification study, which sheds light on critical aspects concerning ensemble precipitation forecast products and post-processing methods for the Godavari River Basin (GRB). The analysis reveals that both raw National Centers for Environmental Prediction (NCEP) and European Centre for Medium-Range Weather Forecasts (ECMWF) forecasts exhibit poor skill in capturing extreme precipitation events across all lead times, and the applied statistical post-processing methods prove ineffective in addressing this issue. These findings emphasize the necessity to enhance the underlying physics of Numerical Weather Prediction (NWP) models to achieve accurate forecasts of extreme precipitation events in the GRB. The correlation between ensemble mean and observed precipitation declines with increasing lead time, while the Root Mean Square Error (RME) remains unaffected by lead time variations. Notably, the post-processed forecasts utilizing the Quantile Regression Forest (QRF) method demonstrate superior performance compared to other forecast types. The ensemble mean of QRF post-processed NCEP and Multi-Model Ensemble (MME) forecasts outperforms additional forecasts regarding correlation coefficient and RME across all subbasins and lead times.

Additionally, the post-processed forecasts exhibit an improved ensemble spread-error relationship compared to the raw forecasts. The analysis indicates that QRF is more effective than Quantile Mapping (QM) in preserving the ensemble spread-error relation. Rank histograms show underdispersion and bias in raw NCEP and ECMWF forecasts for all subbasins, but applying post-processing techniques helps mitigate these issues. Reliability diagrams suggest that the raw NCEP and ECMWF forecasts are overconfident, while the post-processed forecasts perform well at a 1-day lead time. However, as the lead time increases, the reliability of the forecasts declines due to overconfidence. The Area Under the Curve (AUC) values consistently exceed 0.75 for all lead times and subbasins, indicating the usefulness of the forecasts. However, the discrimination ability of the forecasts, as measured by AUC, diminishes with lead time, indicating a higher false alarm rate. Comparing the performance measures employed, the raw MME forecasts exhibit better overall performance than the raw NCEP and ECMWF forecasts. However, both QRF post-processed NCEP and MME forecasts demonstrate similar performance. Considering the computational cost, the thesis recommends utilizing 20-member QRF post-processed NCEP forecasts for hydrologic forecasting applications in the GRB. Furthermore, the study reveals that

the overall performance of NCEP and MME forecasts surpasses that of ECMWF forecasts, and the QRF post-processed forecasts outperform both QM post-processed and raw forecasts. The QRF post-processed NCEP and MME forecasts exhibit satisfactory performance in various subbasins, including Lower Godavari, Middle Godavari, Indravati, Manjira, and Weinganga, as assessed using deterministic and probabilistic measures.

The second part of the thesis focuses on analyzing the impact of model resolution on the simulation of flood peaks using event-based semi-distributed models. The study utilizes the semi-distributed and semi-lumped GR4J model setup at three different spatial resolutions to simulate streamflow at the Jagdalpur and Wardha basins. The results of the study demonstrate very good performance of all models in simulating streamflow during the calibration period. The Nash-Sutcliffe Efficiency (NSE) values for all models exceed 0.76, indicating a high level of accuracy. Among the models, the semi-distributed models perform the best in both the Jagdalpur and Wardha basins. Overall, the discretization-based models prove effective in capturing peak flows. The validation results of the calibrated models during the validation period also indicate a high level of accuracy in simulating streamflow. The streamflow simulations obtained from the calibrated models exhibit strong performance based on NSE.

Furthermore, the study evaluates the performance of the lumped and discretization-based GR4J models in simulating historical flood events. The median NSE values of the lumped models during the calibration period exceed 0.68 at both Jagdalpur and Wardha, indicating good performance. However, during the validation period, the median NSE values of the lumped models at Jagdalpur (0.56) and Wardha (0.269) indicate their limited ability to account for spatial variability. In contrast, the semi-distributed and semi-lumped models demonstrate strong performance in capturing flood peaks during calibration. The median NSE values for flood events exceed 0.84 (calibration) and 0.71 (validation) for all semi-distributed and semi-lumped models at Jagdalpur. Similarly, the median NSE values at Wardha surpass 0.79 (calibration) and 0.67 (validation). These results indicate the models' ability to accurately capture flood peaks in both basins.

In summary, analysing model resolution's impact on flood peak simulation using event-based semi-distributed models showcases the superiority of semi-distributed models over lumped models in accurately representing streamflow and flood peaks. The discretization-based models effectively capture peak flows and keep a balanced water budget. Notably, the difference between median NSE values in the calibration and validation periods for the semi-lumped models is relatively lower than that of the semi-distributed models, indicating efficient parameter transferability. Moreover,

the semi-lumped models display improved performance with increased discretization, underscoring their efficient parameter transferability. This study provides valuable insights into the performance and suitability of different model resolutions for simulating flood events in the Jagdalpur and Wardha basins.

The third part of the thesis aims to assess the accuracy of assimilating observed Initial Hydrologic Conditions (IHC) and continuous model outputs into event-based rainfall-runoff models. This evaluation focuses on the performance of lumped and semi-distributed event-based hydrological models, considering the assimilation of soil moisture and streamflow data. For the lumped event-based models, incorporating Data Assimilation (DA) using soil moisture (SM-DA) and streamflow (Q-DA) exhibit satisfactory performance during the calibration period, with median NSE values exceeding 0.5. However, the continuous simulations from the lumped models based on DA show lower performance. In the validation period, the lumped models based on DA do not perform satisfactorily, indicating poor temporal transferability.

Conversely, the event-based and continuous lumped models utilizing Q-DA demonstrate good performance, as indicated by lower values of PEPF (Percent Error in Peak Flow), PBIAS (Percent Bias), and PETP (Percent Error in Time to Peak) when compared to other models. The results suggest that the predictive ability of both continuous and event-based semi-distributed models, incorporating soil moisture and streamflow data assimilation, significantly improves compared to their lumped counterparts. The event-based semi-distributed model with SM-DA exhibits the best performance during the calibration period, with a median NSE value of 0.82. This highlights the advantage of considering spatial variability in the model. However, the temporal transferability of the semi-distributed models in the validation period is poor. The event-based semi-distributed model with Q-DA performs satisfactorily, with median NSE values of 0.64 and 0.57 in the calibration and validation periods, respectively.

Both the continuous lumped and semi-distributed models are calibrated using NSE, logNSE, Kling-Gupta Efficiency (KGE), and Fourth root of the mean quadrupled error (R4MS4E), and the resulting model states are used as IHC in their corresponding event-based models. The event-based lumped model, calibrated based on NSE, KGE, and R4MS4E, exhibits an excellent median NSE value (>0.65) during the calibration period, except for the logNSEssss calibrated model. The (Percentage Error in Peak Flow) PEPF and (Percentage Error in Timing to Peak) PETP values demonstrate satisfactory performance of all lumped models in capturing flood peak magnitude and timing during the calibration period. In the validation period, a decline in performance is observed

for all models based on the selected evaluation statistics. However, the NSE values indicate satisfactory performance for all models during validation. In the case of semi-distributed models, the simulated flood hydrographs accurately represent the observed flood hydrographs. The median NSE values for all semi-distributed models are above 0.77 during the calibration period. Similarly, the performance in the validation period, based on KGE, R4MS4E, and NSE-calibrated continuous models, is also good (NSE > 0.65). The PEPF values remain below 30% in calibration and validation periods, indicating good performance in capturing observed flood magnitudes. In conclusion, the study demonstrates that the event-based and continuous lumped models based on Q-DA perform well, exhibiting lower PEPF, PBIAS, and PETP values. The performance of continuous and event-based semi-distributed models, incorporating soil moisture and streamflow data assimilation, surpasses that of their lumped counterparts in terms of NSE, PEPF, PETP, and PBIAS.

The final part of the thesis focuses on generating ensemble flood forecasts using the calibrated lumped and semi-distributed GR4J models. Both raw and post-processed ensemble precipitation forecasts are utilized as inputs to generate ensemble streamflow forecasts. The short- and medium-range ensemble flood forecasts are evaluated for seven historical flood events at Bamni (Wardha). The performance of the generated Ensemble Flood Forecasts (EFF) is deemed satisfactory for shorter lead times, ranging from 1 to 3 days, as indicated by PEPF and PETP values consistently below 30% during the calibration period of the post-processor. However, the performance of the post-processed forecasts deteriorates during the validation period. As the lead time increases, there is a noticeable discrepancy in the prediction of the timing of flood peaks. Comparing the performance of the semi-distributed model-based EFF with its lumped counterparts, it is evident that the semi-distributed model yields better results. The semi-distributed model exhibits improved accuracy in predicting flood events compared to the lumped models. The findings of this study contribute to the understanding of ensemble flood forecasting techniques and their applicability in the study area. Further research and improvements in post-processing methods are recommended to enhance the accuracy and reliability of long-range flood forecasts.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 From Deterministic to Probabilistic: The Evolution of Flood Forecasting Systems

Flood is a common natural disaster worldwide that can cause catastrophic impacts on day-to-day operations. Unlike other natural disasters, floods can be forecast in advance as a preparedness measure (H. L. Cloke and Pappenberger, 2009). There are several methods to forecast flood events for different time scales, and many of them primarily rely on observed precipitation as input from the upstream of the catchment for forecasting at much shorter time scales. However, to forecast for longer lead time, quantitative precipitation forecasts of longer lead times obtained from Numerical Weather Prediction models (NWPs) are transformed into flood hydrographs by employing hydrological models. Hence, hydrological and meteorological models are integrated to forecast flood hydrographs, and the ensemble system is called a hydro-meteorological forecasting system. The traditional flood forecasting system often operates based on deterministic forecast value. Deterministic forecast systems yield a single variable value for each forecast lead time but do not provide a conclusive estimate of the uncertainty in the forecast, which is a major limitation. Therefore, most operational flood forecasting systems are shifting towards adopting ensemble forecasts over the past two decades. The idea of ensemble forecasting is developed in the atmospheric community to overcome the limitations inherent in deterministic forecasts (Jain et al., 2018). In an ensemble weather forecast, two significant sources of uncertainty must be accounted for; initial condition uncertainty and model uncertainty.

Many operational Ensemble Prediction Systems employ the Monte Carlo approach in NWPs to perturb the initial conditions concerning a control forecast. Generally, combinations of singular vectors and ensemble data assimilation methods create initial condition uncertainty(Cloke and Schaake, 2018). Model uncertainty is mainly represented by three schemes namely perturbed parameterization, stochastic parameterization and multi-model ensembles (Palmer et al., 2009). In contrast to single-valued deterministic forecasts, many forecast centres generate an ensemble of 10 to 50 plausible future states of the weather variables. The probabilistic information about the ensemble of forecasted flood events provides an added advantage in comparison to deterministic

forecasts at much longer lead times scaling from short (up to 3 days) to medium (3-15 days) range forecasts (H. L. Cloke and Pappenberger, 2009; Cloke and Schaake, 2018; Wu et al., 2020a). The forecasted ensemble flood values do not result in actual probabilities as uncertainties associated with the forecast system are not treated thoroughly. Thus, disseminating the probabilistic forecasts to the end users isn't easy. Despite these challenges related to ensemble prediction systems, the advantages of ensemble flood forecasts over deterministic forecasts have been reported many times in the literature so far (Bischiniotis et al., 2019; Emerton et al., 2016; Gomez et al., 2019; Siqueira et al., 2020; Thiemig et al., 2015).

1.2 Uncertainties in Ensemble Flood Forecasting Systems

The emergence of ensemble forecasting approaches can be traced back to the discovery of the chaotic nature of the atmosphere by Edward Lorenz (1963). Lorenz proposed that uncertainty associated with the atmospheric system's inherent nonlinearity and initial conditions can be addressed through ensemble forecasting. Epstein (1969) accounted for the intrinsic uncertainties by proposing a theoretical Stochastic-Dynamic approach. However, this approach was impractical in operational applications as there were a plethora of model equations to be implemented with limited computational facilities at that time. Leith (1974) proposed the Monte Carlo framework as a practical implementation of Epstein's Stochastic-Dynamic approach, representing the first ensemble forecasting attempt. Due to the computational constraints in the 1970s, only a few ensemble members were used to describe the uncertainty in initial conditions. Over the years, the amplification of computational power has facilitated addressing the initial condition uncertainty along with model structure, model dynamics and even model parameters. European Centre for Medium-Range Weather Forecasts (ECMWF) and the National Centers for Environmental Prediction (NCEP) started using ensemble forecasts in 1992. The World Meteorological Organization (WMO), recognizing the significance of ensemble forecasting, launched a 10-year international program in 2003 named The Observing System Research and Predictability Experiment (THORPEX). The main objective of THORPEX is to accelerate the usage of ensemble meteorological forecasting for lead times spanning 1 day to 2 weeks. Around ten numerical weather prediction centres worldwide have provided their medium-range ensemble weather forecasts to THORPEX Interactive Grand Global Ensemble (TIGGE) database since the end of 2006. Since the 1970s, ensemble forecasting has been used in the hydrological perspective. The United States National Weather Service (NWS) generated Climatological Extended Streamflow Predictions

(ESP) for application in reservoir and water supply operations (Day, 1985). An initiative for the global adoption of hydrological ensemble flood forecasting named Hydrological Ensemble Prediction Experiment (HEPEX) started in 2004.

Ensemble hydro-meteorological forecasting systems are complex and embedded with various levels of uncertainty. A comprehensive flow chart showing multiple components of a hydro-meteorological ensemble flood forecasting system is shown in Figure 1.1.

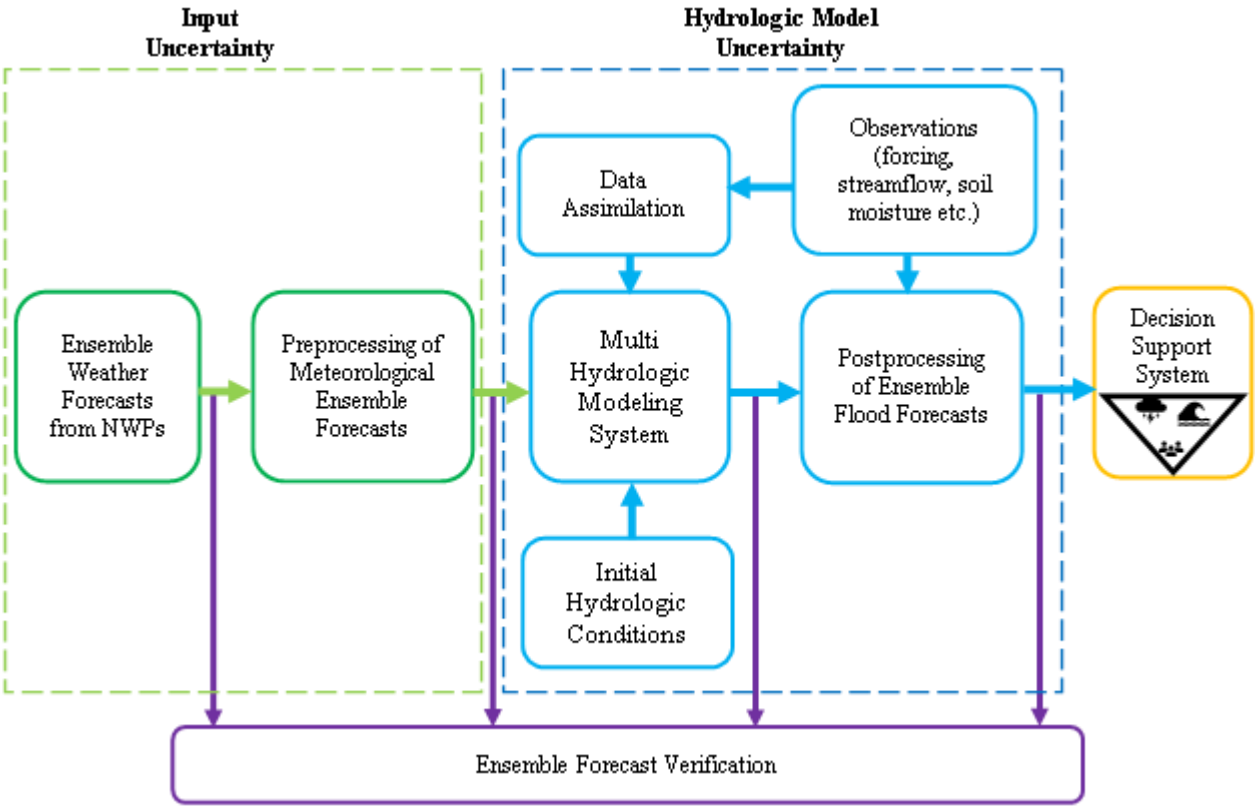

Figure 1.1 Flow chart of hydro-meteorological ensemble flood forecasting system with all possible components.

The fundamental principle of predictability for the hydro-meteorological system is twofold; predictability of the atmosphere that drives the hydrologic cycle and estimation of the initial hydrologic conditions of the watershed. The significant uncertainty contributors in the forecasting system are the input data, hydrological model structure, model parameters and observed data used for model evaluation (Wu et al., 2020a). The uncertainty in the input is addressed by generating the ensemble weather forecasts by perturbing the initial atmospheric conditions of the numerical

weather prediction models. Many operational hydro-meteorological ensemble flood forecasting systems do not explicitly assess the uncertainty associated with the hydrological modelling, typically leading to an under-dispersive forecast (Wang et al., 2009). The full range of uncertainty, including uncertainty involved in hydrological modelling, is not accounted for in the modelling chain, making it an active area of current research.

1.3 Evaluating Ensemble Precipitation Forecasts in the Godavari River Basin

Reliable and skilful ensemble hydrologic forecasts play a crucial role in reservoir operations, hydropower management, and the prediction of hydrological extremes like droughts and floods. However, the accuracy of such forecasts heavily relies on uncertainties associated with meteorological and hydrological models. Precipitation forecasts are essential among the various factors contributing to these uncertainties as they drive the hydrological cycle. The inherent uncertainty in precipitation forecasts directly impacts the quality of hydrologic forecasts. To address this challenge, the hydrological community has increasingly embraced ensemble forecasting methods to quantify uncertainties in forecasting systems. Ensemble weather forecasts from Numerical Weather Prediction (NWP) models provide inputs for calibrated hydrological models to generate ensemble streamflow forecasts. While substantial efforts have been made to improve NWP predictions, accurately predicting precipitation, storm intensity, and location remains a critical challenge (Novak et al., 2014; Ridwan Siddique et al., 2015). A systematic forecast verification framework is crucial to assess and quantify the skill of ensemble forecasts. Although numerous verification studies have been conducted worldwide for various forecasting systems, limited research has focused on ensemble precipitation forecasts in India's Godavari River Basin (GRB). Apart from the advantages of ensemble precipitation forecasts (Liu et al., 2012; Park et al., 2008), the raw ensemble forecasts are unsuitable as direct input for hydrological applications due to their inherent biases. Hence, various statistical post-processing methods are generally employed for spatial downscaling and correcting forecast biases regarding mean and ensemble spread (Hamill et al., 2004). In this regard, a systematic verification study is necessary to assist the hydrological community in choosing the best suitable forecasts and post-processing methods to develop a reliable ensemble flood forecasting system.

1.4 Spatial Scale and Calibration Approaches in Hydrological Modelling

Accurate flood prediction is crucial for effectively mitigating flood impacts, and selecting an appropriate hydrological model plays a pivotal role in achieving accurate predictions. Hydrological models are categorized based on spatial scale and process description, with lumped, distributed, and semi-distributed models representing different spatial scales and data-driven, physics-based, and conceptual models reflecting various process descriptions. Among these models, conceptual models, known as 'grey-box' models, have gained popularity due to their simplicity and computational efficiency. Semi-distributed models, in particular, offer advantages by considering spatial heterogeneity and simulating streamflow even in areas where observed data may be limited. Previous studies have explored the influence of spatial variability and heterogeneity using semi-distributed models by dividing catchments into sub-basins of approximately uniform size based on topography (Booij, 2005; Das et al., 2008; Lobligeois et al., 2014). However, for a semi-distributed conceptual model, it is desirable to maintain homogeneity within sub-basins, aligning with the assumptions of a lumped model.

Additionally, a threshold exists for subdivision levels beyond which no further improvements in model performance can be achieved (Das et al., 2008). Balancing the finite model resolution obtained through landscape discretization with computational burden is essential. Although the implementation of event-based lumped conceptual models has been reported in previous studies, the impact of nested discretization and calibration strategies on the hydrologic response of event-based conceptual models remains largely unexplored. This study aims to investigate the effects of varying spatial scales and calibration strategies on streamflow simulations using the modèle du Génie Rural à 4 paramètres Journalier (GR4J) model in the Jagdalpur and Wardha catchments in India, shedding light on the performance of event-based models in flood simulations under different conditions.

1.5 Uncertainty in the estimation of IHC

Event-based models are often preferred over continuous models within the flood modelling framework due to their ease of calibration and reliance on event-scale data (Tramblay et al., 2012). However, continuous and event-based hydrologic predictions depend on accurately simulating initial hydrological conditions, particularly soil moisture within the watershed. To address this challenge, numerous modelling efforts have focused on minimizing uncertainties associated with the estimation of initial hydrologic conditions, employing various predictors such as piezometric

levels, baseflow, antecedent discharge index, outputs from continuous models, and in situ or remote-sensing observables (Bahramian et al., 2021; Coustau et al., 2012; Franchini et al., 1996; Hegdahl et al., 2020; Huang et al., 2016; Longobardi and Villani, 2003; Meng et al., 2017). To reduce uncertainties associated with input data, model states, and output variables, Data Assimilation (DA) techniques are widely employed, optimizing the combination of observations and model simulations to enhance initial state estimates of hydrological models. However, the choice of model structure and the trade-off between input data, model complexity, and computational costs are critical in real-time flood forecasting applications. While fully distributed physics-based models offer better simulations by accounting for spatial heterogeneity, they require extensive data and computational resources.

In contrast, conceptual hydrologic models' computational efficiency and simplicity (CHMs) have made them effective in operational streamflow forecasting. Although DA studies on CHMs have primarily focused on lumped versions, understanding the relevance of DA-estimated initial states in improving the performance of event-based conceptual hydrological models is crucial. Moreover, the choice of calibration metrics in scenarios where initial conditions are estimated through continuous models impacts the model states these models simulate. Therefore, the present study aims to assess the performance of a lumped and semi-distributed event-based conceptual model, utilizing DA-estimated initial conditions and corresponding continuous models calibrated using different objective functions for improved flood forecasting.

1.6 Objectives of the study

The objectives of the study are listed as follows:

- Identification of the best suitable ensemble weather forecast products and post-processing methods over the Godavari River Basin
- Analysing the impact of model resolution on simulation of flood peaks through event-based semi-distributed models.
- Accuracy evaluation of assimilation of observed initial hydrologic states/conditions and IHC obtained from continuous models to initialize event-based rainfall-runoff models.

1.7 Contributions from the study

With the understanding of different sources of uncertainties in the Ensemble Flood Forecasting framework, the present study is mainly designed to evaluate the hydrological uncertainties associated with flood prediction. Initially, the study aims to identify the most suitable ensemble precipitation forecast products and their statistical post-processing methods over a large river basin (Godavari River Basin) in India. The aim is to improve the reliability of flood forecasting by selecting the optimal ensemble forecast products by correcting their inherent biases through post-processing techniques. Subsequently, the thesis examines the impact of model resolution and calibration strategies on the simulation of flood peaks using event-based semi-distributed conceptual hydrological models. The study determines the optimal finite model resolution through landscape discretization by comparing different spatial resolutions. The findings improve flood predictions by identifying the most accurate representation of streamflow and flood peaks. In addition, the accuracy of assimilating observed initial hydrologic states and continuous model outputs into event-based rainfall-runoff models was evaluated. The study focuses on lumped and semi-distributed event-based models and considers the assimilation of soil moisture and streamflow data. The findings highlight the benefits of incorporating assimilation techniques in improving flood event simulation. Finally, the raw and post-processed ensemble precipitation forecasts were forced into the calibrated lumped and semi-distributed event-based GR4J model to generate short and medium-range ensemble streamflow forecasts at Wardha Basin.

At the very beginning, the thesis extensively evaluates the skill of ensemble precipitation forecasts obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF) and the National Centers for Environmental Prediction (NCEP) within "The Observing System Research and Predictability Experiment" (THORPEX) Interactive Grand Global Ensemble (TIGGE) database. The evaluation is conducted over the eight subbasins of the Godavari River Basin in India. These forecasting systems are chosen due to their operational nature, availability of multiyear datasets, and ability to capture scenarios of interest to forecasters. Additionally, the study verifies the skill of the multi-model grand ensemble (MME) created by integrating NCEP and ECMWF forecasts. Two statistical post-processing methods address conditional biases in the raw forecasts. The verification process includes daily accumulations for various lead times (ranging from 1 to 15 days) in hindcast mode. Deterministic and probabilistic measures such as forecast error box plots, correlation coefficient, Relative Mean Error (RME), mean Continuous Ranked Probability Score (CRPS), spread-skill relationship, rank histograms, reliability diagram, and Area

under Relative Operating Characteristic Curve (AUC) are utilized to assess the skill of basin-averaged raw and post-processed forecasts in comparison to the observed data from 2016 to 2020.

Subsequently, the thesis explores the trade-offs between increasing model resolution of a conceptual model through a nested discretization scheme versus its capability to simulate accurate streamflow. The catchments are discretized iteratively by assessing the spatial heterogeneity of the Natural Resources Conservation Service (NRCS) Curve Number (CN) using spatial moments computed from the catchment outlet. Two different calibration strategies (semi-distributed and semi-lumped) were employed to calibrate the spatially discretized GR4J model. Lumped, semi-lumped and semi-distributed GR4J models were used to simulate the observed streamflow at Jagdalpur and Wardha basins in India. Along with a continuous model, an event-based model was set up in this study to evaluate the model's performance in capturing the observed flood events. A diagnostic performance evaluation strategy was adopted to assess the accuracy of the simulated streamflow in matching the different attributes of the observed hydrograph, such as high flows, low flows and overall water balance. In addition, the performance of event-based models is assessed concerning their ability to capture the timing and magnitude of their flood peaks.

Further, the performance of different methods to estimate initial hydrologic conditions for event-based rainfall-runoff modelling was assessed. The first method involves assimilating observed variables, including streamflow and soil moisture, using the Ensemble Kalman Filter. Root zone soil moisture data from the Global Land Data Assimilation System (GLDAS) is assimilated for soil moisture assimilation. The IHCs estimated through assimilating root zone soil moisture (SM-DA) and streamflow (Q-DA) into the conceptual hydrological model GR4J, at both lumped and semi-distributed spatial resolutions, are used as the initial conditions in the event-based models. The second method utilizes states obtained from the continuous model, calibrated using four different calibration metrics. The parameters of the continuous models are automatically calibrated using the Genetic Algorithm and four calibration objective functions: Nash-Sutcliffe Efficiency (NSE), NSE of logarithmically transformed flows (logNSE), Kling-Gupta Efficiency (KGE), and Fourth root Mean Quadruple Error (R4MS4E). The observed flood events in the Jagdalpur catchment are simulated using a conceptual hydrologic model at two spatial resolutions: lumped and semi-distributed. To evaluate the performance of the event-based GR4J models in simulating observed flood hydrographs in the Jagdalpur basin, various metrics such as NSE, PETP, PEPF, and PBIAS are utilized.

The final part of the thesis focuses on generating ensemble flood forecasts (EFF) using the calibrated lumped and semi-distributed GR4J models. Both raw and post-processed ensemble precipitation forecasts are utilized as inputs to generate ensemble streamflow forecasts. The short- and medium-range ensemble flood forecasts are evaluated for seven historical flood events at Bamni (Wardha). The first four events are included in the calibration period of the post-processor, while the next three events fall within the validation period. The performance of semi-distributed model-based EFF is compared to EFF generated through its lumped counterparts.

1.8 Outline of the thesis

Within Chapter 2, a thorough literature review is conducted, focusing on crucial areas pertinent to the research. It investigates forecast verification studies, statistical post-processing methods, discretization schemes, hydrological modelling, and the estimation of Initial Hydrologic Conditions (IHCs). This chapter constructs a robust theoretical foundation by critically analysing relevant literature and providing essential contextual background for the study.

Chapter 3 is dedicated to the verification of ensemble precipitation forecasts, with the objective of determining the most effective forecast products and statistical post-processing techniques to improve the reliability of precipitation predictions in the study area. This chapter employs a comprehensive evaluation and comparison approach to identify ensemble forecast products and post-processing methods that exhibit superior performance and accuracy in capturing the complex precipitation patterns and variability observed over the Godavari River Basin.

Chapter 4 analyses the impact of spatial discretization and calibration strategy on event-based models for flood simulation. This chapter aims to assess how different levels of model resolution, specifically through event-based semi-distributed models, affect the simulation of flood peaks. A nested discretization scheme is proposed to delineate a catchment into homogeneous subbasins iteratively. These subbasins are then utilized to establish semi-distributed and semi-lumped conceptual models operating in continuous and event-based modes.

Chapter 5 presents a comprehensive performance assessment of methods to estimate initial hydrologic conditions for event-based rainfall-runoff modelling. The objective is to evaluate the accuracy and effectiveness of these approaches, including the assimilation of observed variables and utilization of continuous model states. The assessment covers different spatial resolutions and includes the simulation of observed flood events in the Jagdalpur catchment.

Chapter 6 focuses on generating ensemble flood forecasts by forcing raw and post-processed ensemble precipitation forecasts as inputs into the calibrated lumped and semi-distributed GR4J model. The chapter examines the generation of short and medium-range ensemble flood forecasts for seven historical flood events at Bamni (Wardha).

Chapter 7 presents the summary and conclusions of the work described in the thesis.

Chapter 2

Literature Review

2.1 Introduction

Literature related to forecast verification studies, statistical post-processing methods, discretization schemes, hydrological modelling, and the estimation of Initial Hydrologic Conditions (IHCs) is concisely reviewed in this chapter. Specifically, the literature review focuses on different forecast verification studies conducted globally to gain insights into the inherent biases in the forecasts, various statistical post-processing methods, and other forecast verification metrics. Furthermore, it explores the implementation of different discretization schemes to incorporate spatial variability of catchment characteristics and rainfall into hydrological models, as well as various studies that assess the performance of hydrological models considering different spatial resolutions and calibration strategies. Additionally, the literature review explores various techniques for estimating Initial Hydrologic Conditions (IHCs) and investigates data assimilation methods that assimilate observed variables to estimate IHCs in conceptual models. By addressing these specific topics, the literature review offers valuable insights into advancements, challenges, and potential areas for further research.

2.2 Forecast Verification and Statistical Post-Processing

Due to varying climatic, physiographic, hydrographic, and demographic factors across the length and breadth of the country, India experiences different types of flooding, including fluvial, coastal and urban flooding (Singh & Kumar, 2017). With a large flood-prone area and millions of hectares affected annually, India is one of the most flood-prone countries in the world (National Disaster Management Plan, 2019; Singh and Kumar, 2017). The escalating population and urbanization further exacerbate flood risks, emphasizing the need for timely flood risk reduction and mitigation measures. Flood forecasting, as a low-cost non-structural measure, is crucial in developing flood control measures and management strategies with limited resources. However, the accuracy of flood forecasts heavily relies on addressing the uncertainties associated with meteorological and hydrological models to improve their reliability and skillfulness for reservoir operations,

hydropower, and the prediction of hydrological extremes (H L Cloke and Pappenberger, 2009; Cloke and Schaake, 2018).

Precipitation plays a crucial role in flood forecasting as it drives the hydrological cycle, and the uncertainties associated with precipitation forecasts significantly contribute to uncertainties in flood forecasts. Accurate precipitation forecasts are essential for generating reliable and skilful flood forecasts, leading to a shift towards adopting ensemble forecasts in the hydrological community to quantify uncertainties in forecasting systems (Cloke and Pappenberger, 2009). Ensemble weather forecasts, generated by Numerical Weather Prediction (NWP) models, provide ensemble simulations of meteorological variables such as precipitation and temperature for different lead times. These ensemble forecasts are then used as inputs to calibrated hydrological models to generate ensemble streamflow forecasts. Despite progressive efforts in improving NWP models, accurate prediction of precipitation, storm intensity, and location remains a critical challenge (Novak et al., 2014; Ridwan Siddique et al., 2015). To address this, a systematic forecast verification framework is crucial for monitoring, verifying, and quantifying the skill of ensemble forecasts obtained from NWPs. By comparing hindcasts with observations using deterministic and probabilistic verification measures, this framework helps assess the forecasts' quality, errors, and biases, providing valuable insights for policymakers, researchers, and forecast users.

Several verification studies have been conducted globally for various forecasting systems, including the National Centers for Environmental Prediction (NCEP) Global Ensemble Forecast System Reforecast (GEFSRv2) (Baxter et al., 2014; Brown et al., 2014; Kim et al., 2018; Ridwan Siddique et al., 2015) and the Short Range Ensemble Forecast (SREF) (Brown et al., 2012; Ridwan Siddique et al., 2015), European Centre for Medium-Range Weather Forecasts (ECMWF) (Hamill, 2012; Leonardo and Colle, 2017; Medina et al., 2019), United Kingdom Met Office Unified Model (UKMO) (Anderson et al., 2019; Leonardo and Colle, 2017; Ran et al., 2018).

Siddique et al. (2015) assessed the quality of precipitation forecasts in the Middle Atlantic Region of the USA, utilizing two ensemble forecasting systems, GEFSRv2 and SREF. Various verification metrics and conditional analysis based on precipitation amounts, basin size, forecast lead times, accumulation periods, and seasonality were employed to gain insights into the accuracy of these forecasting systems. The findings revealed that GEFSRv2 and SREF tended to overestimate light to moderate precipitation while underestimating heavy precipitation. Moreover, the accuracy of precipitation forecasts was found to vary depending on factors such as basin size and forecast lead times. Notably, the study emphasized the significance of accounting for predictive uncertainty in precipitation forecasts to enhance the accuracy of hydrologic modelling.

Sharma et al. (2017) examined the quality of ensemble precipitation forecasts across the eastern United States, with a focus on the NCEP Global Ensemble Forecast System Reforecast (GEFSRv2), Short Range Ensemble Forecast (SREF) system, and NCEP's forecast systems. The study area covered 12 NOAA/National Weather Service River Forecast Centers (RFCs), spanning from Maine to Florida and from the Atlantic coast to the Mississippi River. A verification strategy was employed, which included a common period of analysis (2012-2013) for lead times ranging from 1 to 3 days and a longer period (2004-2013) specifically for GEFSRv2, with lead times from 1 to 16 days. The study found that the GEFSRv2 and SREF systems generally performed well in precipitation forecasts, with the GEFSRv2 demonstrating higher skill scores for longer lead times. However, regional and seasonal variations in forecast quality were observed, with the SREF system performing better in the Northeast and the GEFSRv2 performing better in the Southeast. The study also suggested modifying the weather forecasting systems and guidance could improve flood forecasts. Additionally, it was noted that statistical postprocessing techniques had the potential to remove systematic biases and enhance the skill and reliability of forecasts.

Hamill et al. (2004) investigated on enhancing medium-range weather forecasts by integrating ensemble reforecasting and model output statistics (MOS) techniques. The study highlights the significance of statistical approaches, particularly MOS, in improving forecast accuracy by addressing model bias and distinguishing between predictable and unpredictable elements. Using retrospective forecasts from the NCEP Medium-Range Forecast (MRF) model spanning 23 years, the researchers demonstrate significant improvements in forecast skill by statistically adjusting current forecasts based on prior forecasts. The findings reveal that the adjusted forecasts outperformed operational forecasts, emphasizing the potential of ensemble reforecasting and MOS-like techniques in enhancing forecast accuracy and reliability.

Taillardat et al. (2016) conducted a study to enhance the accuracy and reliability of ensemble weather forecasts through the application of Quantile Regression Forests (QRF) and Ensemble Model Output Statistics (EMOS). The authors proposed a statistical method that combines QRF models for estimating conditional quantiles and EMOS for generating calibrated probabilistic forecasts. Through extensive experimentation, the research demonstrated that integrating QRF and EMOS significantly improved forecast reliability and sharpness. The calibrated forecasts exhibited enhanced skill and better representation of uncertainty compared to the original ensembles.

Nandita and Mishra (2021) stressed the importance of incorporating regional dimensions in flood forecasting infrastructure and establishing a comprehensive flood forecast and warning system.

They emphasized that India's higher expenditure on post-flood recovery than pre-flood mitigation schemes underscored the potential benefits of investing in an efficient flood early warning system. Further, they highlighted the current status and future requirements to strengthen the flood early warning systems in India. They emphasized the need to translate the ensemble weather and climate forecast to a Hydrologic Ensemble Prediction (HEP) system by integrating improved meteorological forecast, hydrologic and hydraulic modelling, data assimilation, and post-processing. The authors also called for revamping guidelines for flood forecasting and developing and enhancing the ensemble forecast system, utilizing the existing ensemble precipitation forecast systems. They argued that the spatial and temporal resolutions and lead time should be improved for precipitation forecast in smaller catchments and urban areas prone to flash flooding. Overall, the study highlighted the importance of an integrated approach at various levels to enhance the operational flood forecast in India, which was essential for developing an effective flood early warning system.

The adoption of ensemble weather prediction systems in India since 2018 has not fully addressed the uncertainties in flood forecasts, as the existing flood forecasting systems remain deterministic (Nanditha and Mishra, 2021; Teja and Umamahesh, 2020). No studies have explored ensemble flood forecasts in a catchment with a regional outlook in India. Therefore, as s precursory study for ensemble flood forecasting, verifying and quantifying the skill of raw and post-processed ensemble precipitation forecasts from two NWP models in the Godavari River Basin (GRB), India is quintessential. Further, it is also vital to assist the hydrological community in selecting the most suitable forecasts and post-processing methods to develop a reliable ensemble flood forecasting system.

2.3 Landscape Discretization and Calibration Strategies in Hydrological Modelling

Hydrological models are simplified representations of a real-world watershed that aims to model the interactions between the input meteorological forcing and catchment using distinct mathematical expressions (Wagener et al., 2001; Mostafaei et al., 2018; Kamamia et al., 2019). Hydrological models can be classified differently based on spatial scale and process description. In spatial-scale-based classification, the models can be divided into lumped, distributed and semi-distributed models. Based on the process description, hydrological models can be broadly classified into data-driven, physics-based and conceptual models (Acero Triana et al., 2019; Ghimire et al., 2020; Madsen, 2000; Vu Van Nghi et al., 2020). Data-driven models employ mathematical

equations to analyse and establish the relation between input and output time series. These models do not incorporate any physical process in the catchment; hence they are known as ‘black-box’ models. Recent advancements in computation intelligence boosted the usage of data-driven hydrological modelling based on machine learning and artificial neural networks (Bafitlhile and Li, 2019). Physics-based distributed models are based on understanding the physics of the processes involved in water circulation. In these models, physics-based partial differential equations describe the catchment processes, such as mass transfer, momentum, and energy. The rationale behind using physics-based distributed approaches is that by accounting for the spatial heterogeneity of meteorological forcing and physical features within the basin, better simulations can be achieved at the basin outlet. However, these models are data intensive and demand more computational resources. Conceptual models, also known as ‘grey-box’ models, consist of very few components (non-linear reservoirs), simplified representations of elements in hydrological system. Nonlinearities associated with the catchment responses such as determining excess rainfall, surface/subsurface runoff and movement of soil moisture are taken care of by the thresholds of different storage reservoirs.

The utilization of conceptual-based hydrological models has been mainly amplified in the past few years, attributable to their simplicity and computational efficiency (Tran et al., 2018). Lumped and semi-distributed conceptual models are well adapted for water resources management and flood forecasting due to their parsimonious model structures (Perrin et al., 2001). Many studies have reported that calibration of lumped and semi-distributed models produces similar or higher overall model performance than the more complex distributed models—when evaluated at the outlet stations.

Bourqui et al. (2006) analysed the impact of considering spatial variability of the catchment into account by dividing the basin into sub-basins on the performance of streamflow simulations. 212 French basins were characterized using approximately 50 indices related to pedology, geology, morphology, and land use. The performance of lumped and semi-distributed rainfall-runoff model was compared using 3300 Chimaera basins (virtually aggregated from two real basins). The findings indicate that incorporating spatial heterogeneity showed improved performance; meanwhile, integrating "useful" spatial data into a lumped model can enhance its performance without compromising its parsimonious structure. Certain indices demonstrated correlations with rainfall, affirming that the semi-distributed approach is particularly advantageous for basins characterized by high spatial variability of precipitation.

Viney et al. (2005) assessed the performance of 10 hydrological models for a catchment located in Germany. Lumped to fully distributed models representing different levels of complexity were used, and they found that simpler models performed well compared to more complex models. However, during the less extreme validation period, some more complex models exhibit greater improvement in performance.

Das et al. (2008) evaluated the performance of lumped, semi-lumped, semi-distributed and fully-distributed models for a European mesoscale catchment. They found that semi-lumped and semi-distributed models outperformed the lumped and fully-distributed models. They identified some plausible reasons for the underperformance of distributed models especially poor estimation of input areal precipitation through spatial interpolation techniques and parameter optimization ending in a local optimum due to sub-grid parameterization. The calibration procedure compensates for bias in the precipitation observations and other model input data in the semi-distributed and semi-lumped model structures.

Generally, it is expected that the performance of any numerical model, in principle, should improve with increasing model resolutions. But the scarcity of finer-scale spatially resolved data poses a significant challenge to increasing the model resolution. Hence, the characteristics of both distributed and lumped models can be found by opting for semi-distributed models and integrating sub-grid parameterization schemes in hydrologically homogenous areas.

Organization of spatial patterns of various physical characteristics of catchment, such as soil characteristics, soil moisture and vegetation type, significantly influence catchment runoff. Splitting the landscape into hydrologically homogeneous regions is called ‘landscape discretization’. In congruence with the dominant hydrological processes, the topography is the most influential factor in the discretization process, whereas soil and vegetation are determining factors in the lower hierarchy (Pilz et al., 2017). However, there is no universally accepted definition of discretization; it typically consists of a stratification scheme of partitioning a hydrological basin into sub-basins, which are sequentially divided into irregularly shaped hydrologically uniform entities (Krysanova et al., 1998). Conceptual models should have a lower limit of splitting the landscapes simply because they represent average spatial behaviour (Das et al., 2008). However, there is a threshold of subdivision level above which no more improvements can be achieved (Haghnegahdar et al., 2015; Han et al., 2014; Wood et al., 1988). It is essential to determine an optimum finite model resolution through landscape discretization.

Several methods have been developed to address the discretization issues, for which no generic solution has been found so far (Beven, 2006). Wood et al. (1988) investigated the scale-dependent nature of the variability of catchment characteristics and proposed the concept of 'Representative Elementary Area (REA)' with the assumption of an implicit continuum. They employed a modified version of the top model to simulate the hydrologic response of the catchment, considering both infiltration excess and saturation excess runoff and incorporating the spatial variability of soils, topography, and rainfall. The study revealed that the size of the proposed REA is predominantly influenced by topography. It was observed that there exists a relationship between the REA and catchment hydrologic responses, with the topography exerting a strong influence on the REA. Furthermore, the findings indicated that the length scale of rainfall plays a secondary role in determining the size of the REA, while increased variability in rainfall and soils between sub-catchments contributes to enhanced variability in runoff generation.

Another prominent discretization scheme, namely hydrological response units (HRUs), was introduced by Leavesley et al. (1983) for their Precipitation–Runoff Modelling System (PRMS) and further elaborated by Flügel (1995). HRUs are formed by unique soil, land use and slope combinations, which are assumed to respond similarly to input meteorological data. This scheme has been incorporated in many models such as SWAT (Manguerra and Engel, 1998), SWIM (Krysanova et al., 1998), GSFLOW (Markstrom et al., 2008) etc. However, the HRU approach cumulates the generated streamflow over all HRUs without representing the water flow pathways. Merz and Plate (2009) investigated the threshold where the effects of spatial heterogeneity decrease to a negligible size.

Kouwen et al. (1993) defined grouped response units (GRUs) that involve grouping hydrologic response units based on similar response characteristics derived from classified land-cover maps. Unique model parameters are assigned to each land-cover class, reducing the need for extensive model calibration and facilitating the transfer of parameters across both time and space. The model's performance was evaluated in four watersheds in Southern Ontario, where it was initially calibrated for the Grand River watershed. Subsequently, the model was applied to the three other watersheds without requiring further calibration of the hydrologic parameters, with only the river roughness being adjusted. The results degrade as the parameters are applied to the watersheds located in a different physiographic area and further away from the calibration watershed (Humber and eastern metro watersheds, respectively).

Several methods of discretization of hydrological models have been acknowledged in multitudes of studies (Euser et al., 2015; González et al., 2016; Kumar et al., 2010). However, this influence

has been assessed extensively in grid-based models (Melsen et al., 2016; Molnar and Julien, 2000; Sulis et al., 2011) as it is relatively easy to change the grid resolution. However, a systematic analysis of the effect of varying scales (discretization) on simulated flood response from semi-distributed models is necessary. It is also essential to frame a computationally efficient landscape discretization method to determine the optimum model resolution. Further, previous studies have focused on event-based lumped conceptual models, but the influence of nested discretization on the hydrologic response of event-based models has not been thoroughly investigated. The analysis of streamflow response at various spatial scales and the comparison of different model representations contribute to the understanding of improving streamflow predictions in hydrological modelling to achieve greater accuracy.

2.4 Estimation of Initial Hydrologic Conditions

In addition to proper landscape discretization, the choice of hydrological models plays a crucial role in capturing the updating scheme of initial hydrologic conditions. Depending on their approach, these models can be classified as continuous or event-based. Continuous models require a warm-up period to minimize the impact of arbitrarily selected initial values and often rely on long-term continuous data, which can be challenging for operational forecasting. On the other hand, event-based models, widely used for flood forecasting, are easier to calibrate and only require data at the event scale. However, accurately simulating flood events with event-based models necessitates estimating the Initial Hydrologic conditions or Antecedent Wetness Conditions before an event (Tramblay et al., 2010). Previous studies emphasize the significance of estimating the antecedent wetness conditions of a catchment in event-based rainfall-runoff models to improve event predictions. Soil moisture, a necessary antecedent condition influencing the interaction between the atmosphere and land surface, strongly impacts flood magnitudes. Estimating a catchment's initial soil moisture states can be done through two main approaches: 1) simulating soil moisture using continuous rainfall-runoff modelling and 2) deriving soil moisture from reanalysis products/satellite-based imagery.

Berthet et al. (2009) compared event-based and continuous rainfall-runoff modelling approaches for flood forecasting river flows over 178 French catchments. The influence of antecedent soil moisture over the performance of flood forecasting was evaluated, along with possible compensations with soil moisture updating techniques. They reported that the continuous approach

is the best to ensure good forecasting performances. However, assimilation of the previously observed flow considerably reduced the differences in performance.

Tramblay et al. (2010) compared four soil moisture indicators antecedent precipitation, baseflow, in-situ soil moisture and soil moisture simulated from Safran–Isba–Modcou (SIM) model. They calibrated the event-based SCS-CN method and found that in-situ measurement is the best among all, after which SIM simulated soil moisture, and logarithmic baseflow are good predictors.

Khalki et al. (2020) compared five satellite-based soil moisture products: ESA-CCI, SMOS, SMOS-IC, ASCAT and ERA5 reanalysis with in-situ measurements and soil moisture simulated from continuous SMA model. They reported that the SMOS-IC and the ERA5 reanalysis agree with in-situ soil moisture and SMA model outputs. The results suggested using ERA5 reanalysis data available at an hourly resolution due to the quick deprecation of soil moisture after rainfall events in semi-arid regions.

Beck et al. (2021) evaluated 18 contemporary global near-surface soil moisture products, including satellite-based products, the model simulated soil moisture with and without data assimilation. These datasets were compared with in-situ measurements at 5cm depth acquired from 826 sensors spread across the USA. They found that SMAPL3E, SMOS, AMSR2, and ASCAT are performing well with the L-band-based SMAPL3E.

Zhuo and Han (2016) investigated the compatibility of satellite-based soil moisture products with the operational hydrologic model. The study's findings revealed that effective hydrological application of soil moisture data necessitates two interconnected components: 1) relevant soil moisture data for hydrology and 2) a compatible hydrological model structure. The authors recommended a comprehensive evaluation of satellite soil moisture observations in hydrological modelling to address these requirements. Additionally, they suggested modifying the representation of soil moisture in hydrological models to better align with field variations. The study also proposed the development of a soil moisture product tailored explicitly for hydrological modellings, such as a soil moisture deficit measure, to enhance the models' ability to accurately capture soil moisture dynamics.

Data assimilation techniques have shown potential in enhancing the estimation of initial states in hydrological models by incorporating observational data to obtain the most accurate representation of the current model states. Various observed datasets have been utilized to update the states of hydrological models, including streamflows (Seo et al., 2003), soil moisture (Luca Brocca et al., 2010), snow-covered area and snow water equivalent (Andreadis and Lettenmaier, 2006; Clark et al., 2006), and satellite observations of soil moisture and discharge (Andreadis et al., 2007).

Discharge data, in particular, is frequently assimilated as it encompasses information from all other hydrological states. However, due to the challenges of obtaining real-time observations, many studies have turned to satellite-based soil moisture for data assimilation.

In hydrology, three main types of data assimilation methods are commonly employed: Kalman filters and their variants, particle filters, and variational methods. Among these alternatives, Kalman filters are the most widely used. The standard Kalman Filter assumes linearity in the model and Gaussian error statistics in the measurement process. However, the Extended Kalman Filter was introduced to address nonlinear dynamics, although it can become unstable when strong nonlinearities are present. In contrast, the Ensemble Kalman Filter (EnKF) replaces the linearized model with an ensemble of model realizations. A challenge in data assimilation arises from the time lag between the catchment state and streamflow, known as the time of concentration, particularly in large catchments. This means that updating the model states simultaneously as the discharge observations may not be physically realistic, leading to under- or over-estimation in flow correction at later timesteps (Mendoza et al., 2012). McMillan et al. (2013) proposed a recursive ensemble Kalman filter that allowed for natural lag time of catchment and showed an improvement in the results compared to traditional EnKF. Wanders et al. (2014) used Ensemble Kalman Filter to assimilate three satellite-derived soil moisture from ASCAT, AMSR-E and SMOS and discharge observations into the LISFLOOD model. Their study reported enhanced accuracy in flood forecasts when incorporating discharge observations, and further improvement was achieved by assimilation of satellite-based soil moisture. Consequently, for ensemble prediction systems, ensemble data assimilation methods like EnKF and recursive EnKF are well-suited for updating the model states.

The selection of calibration metrics is crucial when estimating the initial states of event-based models through their continuous counterparts. This is because commonly used calibration objective functions tend to prioritize specific flow segments in a hydrograph, leading to a bias reflected in the simulated model states of continuous models (Mizukami et al., 2019).

Various products are accessible for estimating the initial states of hydrological models, including satellite-based products and model-derived soil moisture products. However, the compatibility between these products and the soil moisture simulated by conceptual models raises concerns when attempting direct assimilation. To address this challenge, ensemble data assimilation filters such as EnKF are employed to update the soil moisture of conceptual models using soil moisture products.

2.5 Concluding remarks

This chapter provides a comprehensive overview of the literature related to forecast verification studies, statistical post-processing methods, discretization schemes, hydrological modelling, and the estimation of Initial Hydrologic Conditions (IHCs). The literature review encompasses several critical aspects of flood forecasting and hydrological modelling, focusing on India's specific context of the Godavari River Basin (GRB). It begins by highlighting the limitations of deterministic flood forecasting systems in addressing uncertainties, particularly in the presence of ensemble weather prediction systems in India. The need for ensemble flood forecasts in the GRB is emphasized, along with the importance of verifying and quantifying the skill of raw and post-processed ensemble precipitation forecasts from numerical weather prediction (NWP) models.

The review also acknowledges the growing popularity of conceptual-based hydrological models, specifically lumped and semi-distributed models, due to their simplicity and computational efficiency. These models have shown promising results in water resources management and flood forecasting, often performing as well as or better than more complex distributed models, mainly when evaluated at outlet stations. Specifically, the semi-distributed model is considered a suitable compromise between the characteristics of lumped and distributed models, particularly when integrated with sub-grid parameterization schemes in hydrologically homogeneous areas. For setting up a semi-distributed model, landscape discretization or partitioning hydrological models into spatial units is a critical component. While various discretization methods have been explored in grid-based models, the literature review highlights the need for a systematic analysis of the impact of varying scales on simulated flood response in semi-distributed models. Additionally, a computationally efficient landscape discretization method is required to determine the optimal model resolution. The influence of nested discretization on the hydrologic response of event-based models is investigated in the present study.

Further, the review underscores the significance of soil moisture as a necessary antecedent condition influencing flood magnitudes and the interaction between the atmosphere and land surface. The compatibility between satellite-based soil moisture products, model-derived soil moisture products, and the soil moisture simulated by conceptual models is a challenge for direct assimilation. To overcome this challenge, ensemble data assimilation filters, particularly the Ensemble Kalman Filter (EnKF), are used to update the soil moisture of conceptual models using reanalysis-based soil moisture products and observed streamflow. Hence, two approaches are used in the present study for estimating initial soil moisture states: simulating soil moisture using

continuous rainfall-runoff modelling and deriving soil moisture from reanalysis products or satellite-based imagery.

Chapter 3

Forecast Verification and Statistical Post-processing

3.1 Introduction

Reliable and skilful ensemble hydrologic forecasts are needed for reservoir operations, hydropower and forecasting of hydrological extremes such as droughts and floods (H L Cloke and Pappenberger, 2009; Das et al., 2022; Madadgar et al., 2014). However, the accuracy of hydrologic forecasts usually depends upon the uncertainties associated with meteorological and hydrological models (H L Cloke and Pappenberger, 2009; Cloke and Schaake, 2018).

Precipitation is the crucial variable of interest in hydrologic forecasting as it is the driving force in the hydrological cycle. The uncertainty associated with the precipitation forecasts is a significant contributor to the uncertainties in the hydrologic forecasts (Wu et al., 2020b). Accurate precipitation forecasts are pivotal for generating reliable and skilful hydrologic forecasts. Hence, the hydrological community is shifting towards adopting ensemble forecasts to quantify the uncertainties in forecasting systems. Ensemble weather forecasts for meteorological variables such as precipitation and temperature are simulated by different Numerical Weather Prediction (NWP) models for the lead times ranging from short (0-3 days) to medium range (3-15 days). The obtained ensemble forecasts are then inputs to the calibrated hydrological models to generate ensemble streamflow forecasts.

Despite the progressive efforts in the evolution of NWPs by understanding the underlying atmospheric physics, accurate prediction of precipitation along with storm intensity and location still remains a critical challenge (Novak et al., 2014; Ridwan Siddique et al., 2015). However, a systematic forecast verification framework is quintessential to monitor, verify and quantify the skill of ensemble forecasts obtained from NWPs. The quality of the forecasts from NWPs is generally assessed by reconstructing retrospective forecasts for the previous period (also known as ‘hindcasts’) (Ratri et al., 2019). Various deterministic and probabilistic verification measures assess the degree of correspondence between the hindcasts and the observations. Hence, a

systematic forecast verification framework assists policymakers, researchers and various forecast users in understanding the inherent errors and biases of the forecasts (Sharma et al., 2017).

Multitudes of verification studies spanning across the globe were reported so far for various forecasting systems such as National Centers for Environmental Prediction (NCEP) Global Ensemble Forecast System Reforecast (GEFSRv2) (Baxter et al., 2014; Brown et al., 2014; Kim et al., 2018; Ridwan Siddique et al., 2015) and the Short Range Ensemble Forecast (SREF) (Brown et al., 2012; Ridwan Siddique et al., 2015), European Centre for Medium-Range Weather Forecasts (ECMWF) (Hamill, 2012; Leonardo and Colle, 2017; Medina et al., 2019), United Kingdom Met Office Unified Model (UKMO) (Anderson et al., 2019; Leonardo and Colle, 2017; Ran et al., 2018). Apart from the advantages of ensemble precipitation forecasts (Liu et al., 2012; Park et al., 2008), the raw ensemble forecasts are unsuitable as direct input for hydrological applications due to their inherent biases. Hence, various statistical post-processing methods are generally employed for spatial downscaling and correction of forecast biases in terms of mean and ensemble spread (Hamill et al., 2004).

Despite of the adoption of ensemble weather prediction system in India since 2018, the existing flood forecasting systems are still deterministic and unable to address the uncertainties in flood forecasts (Nanditha and Mishra, 2021; Teja and Umamahesh, 2020). No studies have been reported so far in India that look into the ensemble flood forecasts in a catchment with a regional outlook. The primary objective of this study is to verify and quantify the skill of raw and post-processed ensemble precipitation forecasts from two NWP models in the Godavari River Basin (GRB), India. The main motive behind this verification study is to assist the hydrological community in choosing the best suitable forecasts and post-processing method to develop a reliable ensemble flood forecasting system. Verifying the forecasts is essential as GRB is often prone to floods that can cause devastating damage (Garg and Mishra, 2019; Rakhecha and Singh, 2017). Limited studies have been documented so far verifying the skill of ensemble precipitation forecasts for GRB (Chakraborty et al., 2021; Dube et al., 2017; Durai and Bhardwaj, 2014), and no studies have been reported so far comparing the skill of raw and post-processed forecasts. In this context, the following questions are explored and addressed: How to compare the skillfulness of various forecasting systems? How does the predictive skill of the ensemble precipitation forecasts fluctuate among and within multiple forecasting systems and quality attributes? How does the quality of ensemble forecasts differ over varying lead times? Does statistical post-processing of precipitation forecasts lead to enhanced forecasts? For this purpose, the quantitative ensemble precipitation forecasts from two NWP models, i.e., 50-member European Centre for Medium-Range Weather

Forecasts (ECMWF), 20-member National Centers for Environmental Prediction (NCEP) forecasts and 70-member multi-model grand ensemble (MME) were verified. The skill of raw and post-processed NCEP, ECMWF and MME forecasts were verified against the observations as a preliminary analysis for hydrologic forecasting.

3.2 Study Area and Datasets

3.2.1 Study area

The Godavari River Basin (GRB) is one of the largest river basins in the Indian subcontinent, with a catchment area of 3,13,000 km^2 which is about 10% of India's geographical area (CWC, 2018a). GRB lies between 73°24' to 83°40' East longitudes and 16°19' to 22°34' North latitudes (CWC and NRSC, 2014). The major subbasins are Manjira, Wardha, Weinganga, Godavari Upper, Godavari Middle, Godavari Lower, Indravati, and Pranahita. The average annual rainfall of this basin is about 1100mm (CWC, 2018a). However, the GRB receives 85% of its yearly rainfall during the South-West monsoon (June to September) (Kulkarni et al., 2010). The location map of the GRB, along with its subbasins, is shown in Figure 3.1.

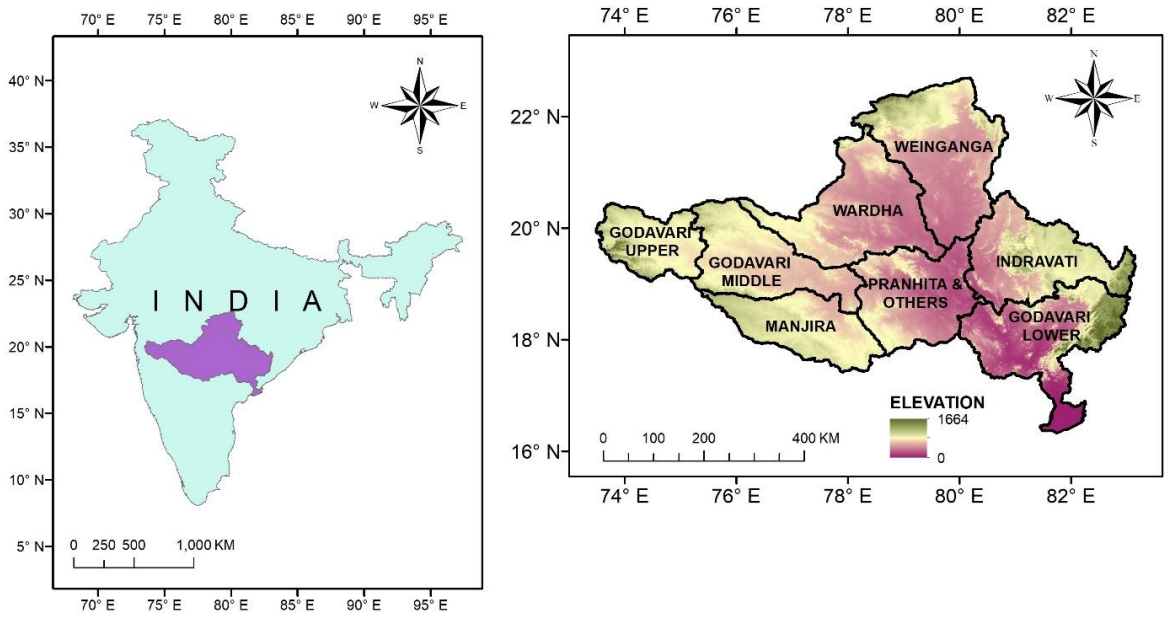

Figure 3. 1 Location map of Godavari River Basin (GRB) and its subbasins.

3.2.2 Datasets

Precipitation ensembles from two NWP models were obtained at a spatial resolution of 0.25^0 x 0.25^0, from "The Observing System Research and Predictability Experiment" (THORPEX) Interactive Grand Global Ensemble (TIGGE) database over the study area. The skill of ensemble precipitation forecasts of two NWP models, i.e., NCEP, a 20-member ensemble, and ECMWF, a 50-member ensemble, was verified in this study against precipitation records. The above NWP models generate forecasts up to 15-day lead-time at 6-hourly accumulations generated (or initiated) 4 times a day at 0-, 6-, 12-, and 18-hr time-steps. This study considers retrospective precipitation forecasts commenced at the 0th hour for forecast verification at varying lead times (1 to 15 days) for 24-hr accumulations. To assess the performance of a multi-model grand ensemble (MME), this study has also used a 70-member ensemble forecast by integrating NCEP and ECMWF forecasts. The NCEP, ECMWF and MME forecasts are collected from 2009 to 2020 for 1-day to 15-day lead times. The selected NWP ensemble forecasts were verified against the observed daily precipitation data available from Indian Meteorological Department (IMD) at 0.25^0 x 0.25^0 spatial resolution (Pai et al., 2014). The Indian Meteorological Department (IMD) dataset used in this study is indeed a gridded dataset, created by Pai et al. (2014). through a process involving weather station data and subsequent post-processing. Specifically, the dataset referred to as IMD4 is a high-resolution gridded rainfall dataset with a spatial resolution of 0.25° × 0.25°, covering a span of 110 years from 1901 to 2010. It is compiled from a comprehensive record of daily rainfall obtained from 6995 rain gauge stations. These stations represent the highest number used in such studies to date and have undergone stringent quality control measures. When compared to existing datasets, IMD demonstrates accurate representation of climatological and variability features of rainfall over India. Its higher spatial resolution and dense station coverage contribute to a more realistic depiction of rainfall patterns in specific regions. This data set is currently the highest quality collection of observations on the Indian subcontinent.

3.3 Methodology

3.3.1 Verification Measures

A diverse pool of verification measures that include deterministic and probabilistic metrics is used in this study for forecast verification. Deterministic metrics assess the quality of the ensemble

mean, whereas probabilistic metrics evaluate the errors in the forecast probabilities.

3.3.1.1 Deterministic metrics

The linear Correlation Coefficient and Relative Mean Error (RME) are used in this study as the deterministic metrics to measure the match with the observed temporal dynamics and under/overestimation biases, respectively (Sharma et al., 2017; Ridwan Siddique et al., 2015). The equation of RME is given in Eq. 1.

$$RME = \frac{\sum_{i=1}^{n}(\overline{E}_i - O_i)}{\sum_{i=1}^{n} O_i} \qquad \text{where, } \overline{E}_i = \frac{1}{m}\sum_{k=1}^{m} E_{i,k} \tag{3.1}$$

where, m is the number of ensemble members, $E_{i,k}$ is the forecast for k^{th} member at time i, O_i denotes the corresponding observation at time i, and n represents the total number of pairs of observed and forecast values.

3.3.1.2 Probabilistic Metrics

As per Murphy and Winkler (1987), the quality of ensemble forecasts is determined by analyzing the following attributes: reliability, sharpness, resolution, discrimination, bias, accuracy and skill. Probabilistic measures are generally employed to analyze the aforementioned attributes in the forecasts.

3.3.1.2.1 Mean Continuous Ranked Probability Score (CRPS)

The continuous ranked probability score (CRPS) is a widely used probabilistic performance measure of ensemble forecasts (Wilks, 2011). CRPS is a quadratic measure that evaluates the difference between the cumulative distribution functions (CDF) of probabilistic forecasts and observed values (Zamo and Naveau, 2018). CRPS is similar to the Mean Absolute Error (MAE) of deterministic forecasts, which can be easily interpreted. The equation of CRPS is given in Eq. 3.2.

$$\text{CRPS} = \int_{-\infty}^{\infty} [F_E(h) - F_O(h)]^2 dh \tag{3.2}$$

where, F_E and F_O are the forecast and observed CDFs, respectively, and h represents every possible threshold. CRPS is averaged over the total pairs of forecast and observed values to verify a set of forecasts. The lower the mean CRPS, the better the forecasts are.

3.3.1.2.2 Spread-Skill relationship

In the case of an ideal ensemble forecast, the mean of the ensemble spread should be equal to the skill score (e.g. root mean square error (RMSE)) over the same period (Wang et al., 2013; Whitaker and Louche, 1998; Zhu and Toth, 2008). To assess this feature, the relationship between ensemble spread and skill is often investigated. In this study, RMSE is used as the skill score, and the standard deviation of ensemble members is used to represent the ensemble spread. RMSE and standard deviation are normalized with observed standard deviation to compare the skill of forecasts concerning a climatological value. When the skill score (RMSE) line is above and far from the spread line, the ensemble is considered over-confident, i.e. under dispersive. If RMSE is lower than ensemble spread, it indicates an under-confident (over dispersive) forecast. Ideally, RMSE should be equal to the ensemble spread for a forecast to be consistent and reliable.

3.3.1.2.3 Rank Histograms

Equiprobability of ensemble members is an ideal quality of ensemble forecasts. The ensemble members should be equally probable in principle as all perturbed initial conditions, varying physics and model structures could possibly be true (Bellier et al., 2017; Hamill, 2001). To check the equiprobability of ensemble members, rank histograms are used. They are generated by repeatedly accumulating the rank of the observed values concerning the ensemble forecasts sorted from lowest to highest. A uniform histogram is desirable, representing an equal likelihood of each ensemble member. Since rank histograms do not assess the resolution of forecasts, they are used in conjunction with other evaluation measures, such as CRPS and area under the Relative operating characteristic (ROC) curve, to understand more about the quality of ensemble forecasts (Hamill, 2001).

3.3.1.2.4 Reliability Diagrams

Reliability diagrams are frequently used graphical tools to visualize the joint distribution of forecasts and dichotomous events (observations) to assess the reliability of probabilistic forecasts (Casati et al., 2008; Dube et al., 2017; Ratri et al., 2019). Quantitative precipitation forecasts are converted to probabilistic forecasts based on an exceedance threshold h and segregated into k bins. The average probability of the forecasts corresponding to the kth bin, B_k, is given by

$$\bar{F}_{E_k}(h) = \frac{1}{|I_k|}\sum_{I_k} F_{E_i}(h), \text{ where } I_k = \{i : i \in B_k\} \quad (3.3)$$

The corresponding average probability of the observations is given by

$$\bar{F}_{O_k}(h) = \frac{1}{|I_k|}\sum_{I_k} F_{O_i}(h), \text{ where } F_{O_i}(h) = \begin{cases} 1, & O_i > h; \\ 0, & \text{otherwise.} \end{cases} \quad (3.4)$$

The forecast probabilities ($\bar{F}_{E_k}(h)$) of an event are plotted against the corresponding observed frequencies given forecast probability ($\bar{F}_{O_k}(h)$) for a total number of forecasts $|I_k|$ in each bin B_k. In this study, an exceedance threshold of 1 mm is chosen to plot reliability diagrams to assess the reliability of NWPs in forecasting precipitation events.

3.3.1.2.5 Area under ROC curve (AUC)

Relative operating characteristic (ROC) curves are the graphical representation of the trade-off between the fraction of hits (true alarm) and the fraction of false alarms to measure event discrimination (Casati et al., 2008; Gill and Buchanan, 2014; Wilks, 2011). For a given threshold h, the hit rate and false alarm rate are given by

$$hit\ rate = \frac{\sum_{i=1}^{n} I_{E_i}(F_{E_i}(h)>d|O_i>h)}{\sum_{i=1}^{n} I_{O_i}(O_i>h)}, \quad (3.5)$$

$$false\ alarm\ rate = \frac{\sum_{i=1}^{n} I_{E_i}(F_{E_i}(h)>d|O_i\leq h)}{\sum_{i=1}^{n} I_{O_i}(O_i\leq h)} \quad (3.6)$$

where d is the probability threshold at which the event causes some action (i.e., the forecast is considered an occurrence), and I is the indicator function. The ROC curve plots a false alarm rate against the hit rate using a set of decreasing forecast probability threshold decreases for possible values of d [0, 1]. Ideally, with decreasing forecast probability thresholds, the hit rate should increase faster than the false alarm rate. Area under the ROC curve (AUC) is a scalar entity that summarizes discrimination between events and non-events. An AUC of 0.5 indicates random forecasts, while an AUC value of 1 reflects perfect forecasts. The value of the AUC of a probabilistic forecast greater than 0.7 is considered 'useful', whereas the AUC above 0.8 is regarded as a 'good' prediction (Ben Bouallègue and Theis, 2014; Mullen and Buizza, 2002). A threshold of 1 mm is chosen in this study to compute the AUC values.

3.3.2 Statistical post-processing Methods

This study used two statistical post-processing techniques to post-process the raw ensemble forecasts obtained from NCEP and ECMWF. The forecast data from 2009 to 2015 is used as the training dataset for post-processing models. The fitted models post-process the raw daily forecasts from 2016 to 2020. It should be noted that the forecast verification was performed from 2016 to 2020. Quantile mapping and quantile regression forests (QRF) are both methods used in in this study for post-processing ensemble precipitation forecasts. Quantile mapping involves adjusting the forecasted distribution to match the observed distribution by fitting a transfer function, often a linear regression, to their cumulative distribution functions. This method assumes a linear relationship between forecast and observed quantiles. On the other hand, quantile regression forests utilize an ensemble of decision trees to estimate conditional quantiles, without assuming any specific distributional form. This makes QRF more flexible in capturing non-linear relationships between variables. Therefore, while quantile mapping is effective for linear relationships, QRF is particularly useful when the relationship between forecasted and observed quantiles is complex and non-linear. A brief description of the selected post-processing methods is given in the succeeding sections.

3.3.2.1 Quantile Mapping

The ensemble precipitation forecasts obtained from the NWP models comprise inherent biases due to the chaotic nature of atmospheric circulation and their regulating dynamic feedback (Christensen et al., 2001). These biases should be corrected to improve the forecasts' quality. Quantile mapping, also known as 'histogram equalization', is a simple and most widely used bias correction method (Acharya et al., 2013). In this method, the distribution function of modelled variable is transformed such that it matches the distribution of observed variable. Based on the assumption that the internal errors in the NWP models and scale relationship are time-invariant, the correction factor obtained during the calibration period is valid for future periods. Gudmundsson et al. (2012) reported that nonparametric quantile mapping performs best among different bias correction methods. The bias correction has been performed on each ensemble member of the raw ensemble forecasts at each grid point (for all 428 grids falling in Godavari River Basin) using Eq. (3.7), considering 2009-2015 as the calibration period. Then the correction factor obtained from the calibration period is used to correct the bias during the validation period.

$$P_o = {F_o}^{-1}\left(F_m(P_m)\right) \tag{3.7}$$

where P_o and P_m are observed and forecasted precipitation, F_m represents the cumulative

distribution function (CDF) of P_m and ${F_o}^{-1}$ is the inverse CDF corresponds to P_o. For performing, non-parametric quantile mapping, the 'qmap' package in R has been used that can be downloaded from https://cran.r-project.org/web/packages/qmap/index.html.

3.3.2.2 Quantile Regression Forests

Quantile Regression Forests (QRF) is a nonparametric postprocessing method developed by (Meinshausen, 2006). QRF provides a nonparametric way to evaluate conditional quantiles for high-dimensional predictors of variables. This technique has been used for meteorological data such as precipitation by Bhuiyan et al. (2018) and Taillardat et al. (2016). Given a set of predictors, the Quantile Regression technique helps to estimate the conditional median and any quantile values of the response variable, whereas the classical regression techniques are helpful in estimating the conditional mean of a response variable. The conditional distribution function ($\hat{F}(y|X = x)$) is expressed as given in Eq. (3.8)

$$\hat{F}(y|X = x) = P\,(Y \leq y|X = x) = E\left(1_{\{Y \leq y\}}|X = x\right) \quad (3.8)$$

Where, X, Y represents the covariate and the observations of the response variable respectively, and E represents the conditional mean, $E\left(1_{\{Y \leq y\}}|X = x\right)$.

In QRF, random forests are generated from binary regression trees called classification and regression trees (CART) that provide robust estimates of conditional quantiles. After building the random forests, a new vector of predictors and its associated leaf in each tree will be determined using binary splitting. This iterative binary splitting on predictors aggregates the observations according to their forecasts. Therefore, for every precipitation event, an ensemble of observations is restored to create an empirical cumulative distribution function (CDF). The weighted average of the CDFs from all the trees is the final forecast, which can be used to obtain the predictive quantiles. The weighted mean over the observation of $1_{\{Y \leq y\}}$ ($1_{\{Y \leq y\}}$ is a step function with a value of 1 when $Y \leq y$ and with a value of 0 when $Y > y$) is used to estimate the conditional mean in Eq. (3.9). Then the conditional distribution function in Eq. (3.9) can be defined as

$$\hat{F}(y|X = x) = \sum_{i=1}^{n} w_i(x) 1_{\{Y \leq y\}} \quad (3.9)$$

$$w_i(x) = k^{-1} \sum_{t=1}^{k} w_i(x, \theta_t) \quad (3.10)$$

where, $w_i(x)$ is the weight vector estimating using random forest regression, k represents the number of trees, and θ is an independent and identically distributed random parameter vector that

determines the growth of the tree (e.g., variables that determine the split points at each node).

The major advantage of QRF is that it not only retains the mean but also keeps all the observation values in the nodes to calculate the conditional distribution. The final CDF obtained from the weighted average CDF of all trees is bounded between the lowest and highest value of the learning sample and can forecast only the quantiles within these limits. The algorithm for computing the estimate of the conditional distribution function $\hat{F}(y|X = x)$ can be summarized as:

Step 1: Grow k trees as in random forests. The corresponding tree is denoted by $T(\theta_t), t = 1, \ldots\ldots, k$. All the observations should be retained for every leaf of every tree.

Step 2: For a given X = x, drop x down all trees. Compute weight $w_i(x)$ for every observation $i \in \{1, \ldots, \mathrm{n}\}$ as an average over $w_i(x, \theta_t)$ $t = 1, \ldots\ldots, k$, as in (3.3).

Step 3: Compute the estimate of the conditional distribution function as in Eq. (3.2) for all $\mathrm{y} \in R$, using the weights from Step 2.

For a quantile of order α, the probability of the random variable being less than y is α, i.e. (CDF value for y). Thus, the quantiles are estimated by Eq. (3.4)

$$\widehat{Q_\alpha} = P\,(Y \leq y|X = x) = \alpha \qquad (4)$$

Instead of using the raw ensemble forecasts, this method can potentially use other predictors and has the advantage of maintaining non-linearity. For instance, Bhuiyan et al.(2018) used global reanalysis precipitation and air temperature datasets along with satellite near-surface soil moisture and elevation datasets as predictors to post-process multiple global precipitation datasets. However, QRF needs a large training dataset which is a major drawback. In this study, all the ensemble members (20 members for NCEP forecasts and 50 members for ECMWF forecasts) were used as predictors for post-processing. The tree size is set to 1000 with the maximum size of terminal nodes as 20. In this study, the post-processing of raw forecasts using Quantile Regression Forests was performed using the 'quantregForest' package in R (Meinshausen, 2017).

3.3.4 Results and Discussion

The mean areal precipitation of ensemble forecasts (raw and post-processed NCEP and ECMWF) and observed precipitation from IMD over each sub-basin of the Godavari River Basin is computed. The skill of basin-averaged forecasts was verified against basin-averaged IMD precipitation records using a pool of deterministic and probabilistic evaluation measures.

3.3.4.1 Rank histograms

Rank histograms measure the frequency of observation to fall between each pair in an ordered set of ensemble values. The under-dispersiveness in the ensemble forecasts leads to higher frequencies of the extreme ranks. Asymmetric histograms represent the bias in the forecasts, and the histograms should be uniform for ideal forecasts. The rank histograms of the NCEP and ECMWF forecasts at 1 day lead time are plotted in Figure 3.2. For brevity of plots, the rank histograms of MME forecasts were not plotted in this section. It can be noticed that both NCEP and ECMWF raw forecasts overpopulate the bins on the very left at all subbasins indicating biased and under-dispersive forecasts. However, it should be noted that ECMWF forecasts are highly biased compared to NCEP forecasts. The higher frequencies on the very left side of the histograms of raw NCEP and ECMWF forecasts are decreasing with increasing lead time, and the frequencies on the very right side of the histograms were found to be increasing with increasing lead time, leading to U-shaped distribution (refer to Figures 3.3 and 3.4). This shows that the raw forecasts are becoming more under-dispersive with respect to lead time. Statistical post-processing has alleviated the problem of bias in the raw forecasts to some extent. The rank histograms of QRF post-processed forecasts were comparatively uniform than QM post-processed forecasts. The U-shaped distribution of QM post-processed forecasts denotes that the forecasts are still under-dispersive. It should be noted that the rank histograms of NCEP-QRF forecasts are slightly biased at lower Godavari and Indravati, whereas ECMWF-QRF is biased at lower Godavari, Upper Godavari, Indravati and Pranahita subbasins.

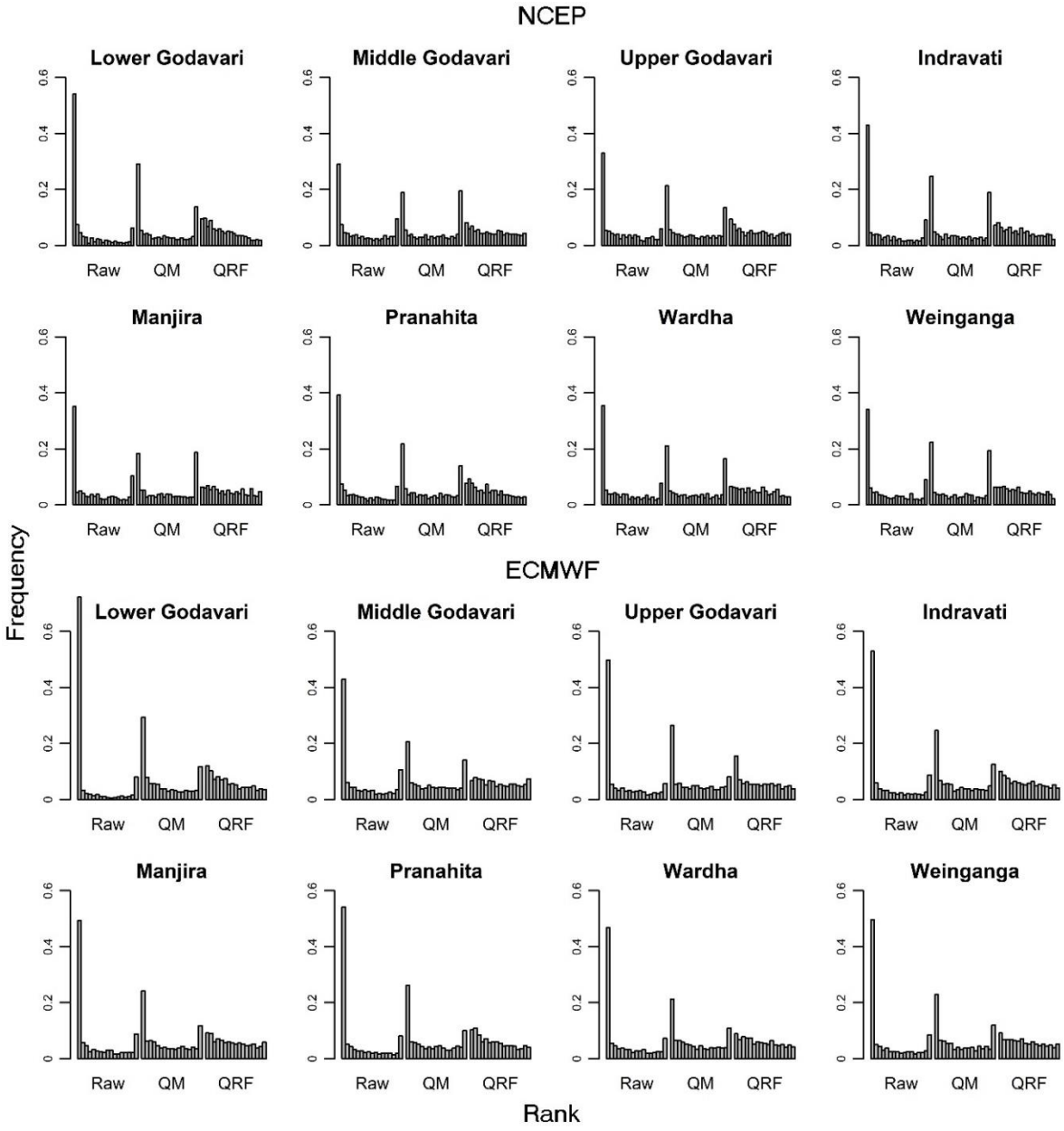

Figure 3.2 Rank histograms of raw and post-processed National Center for Environmental Prediction (NCEP) and European Center for Medium-term Weather Forecasts (ECMWF) forecasts over all subbasins for 24 hours lead time.

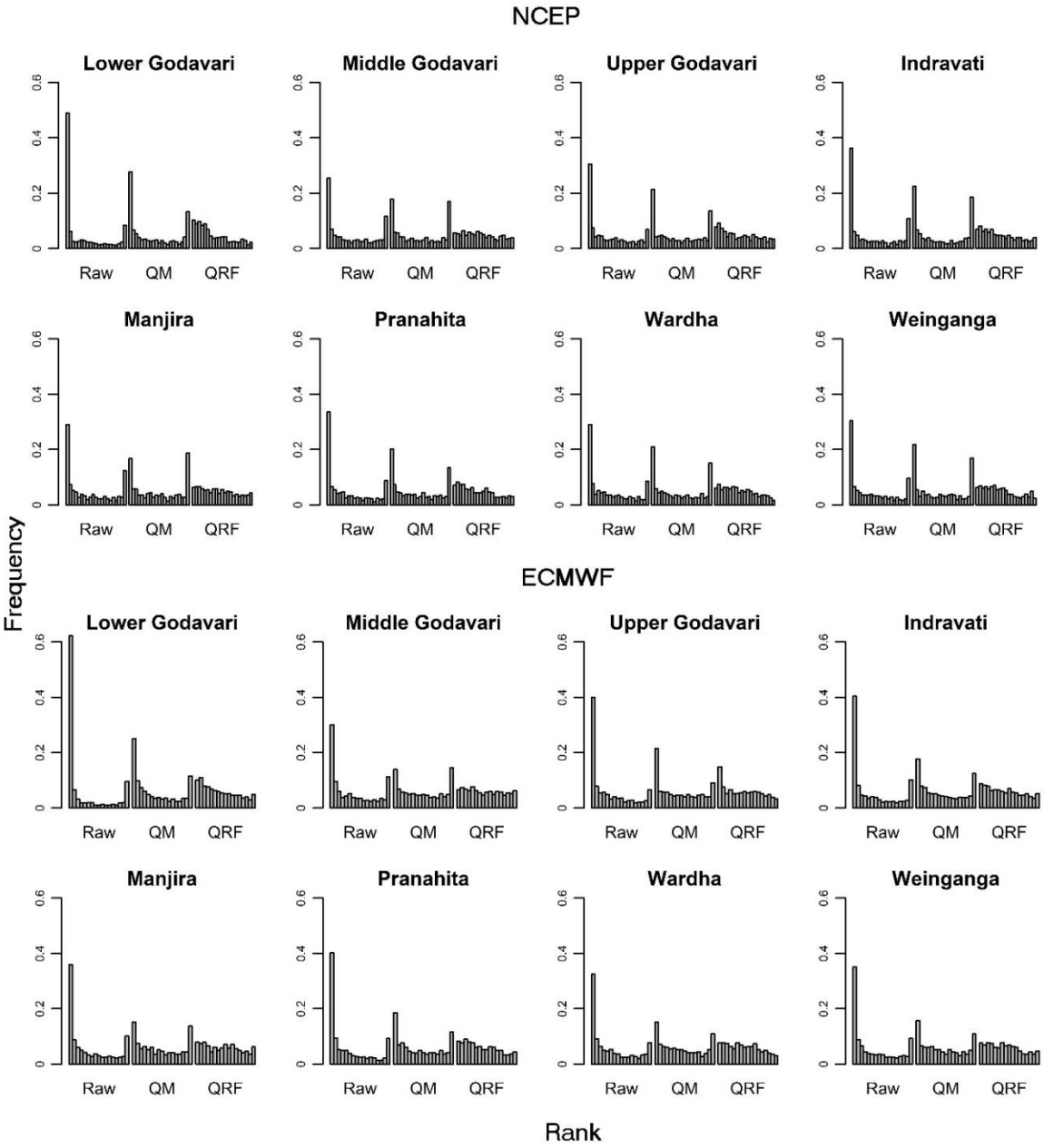

Figure 3.3 Rank histograms of raw and post-processed National Center for Environmental Prediction (NCEP) and European Center for Medium-term Weather Forecasts (ECMWF) forecasts over all subbasins for 5-day lead time.

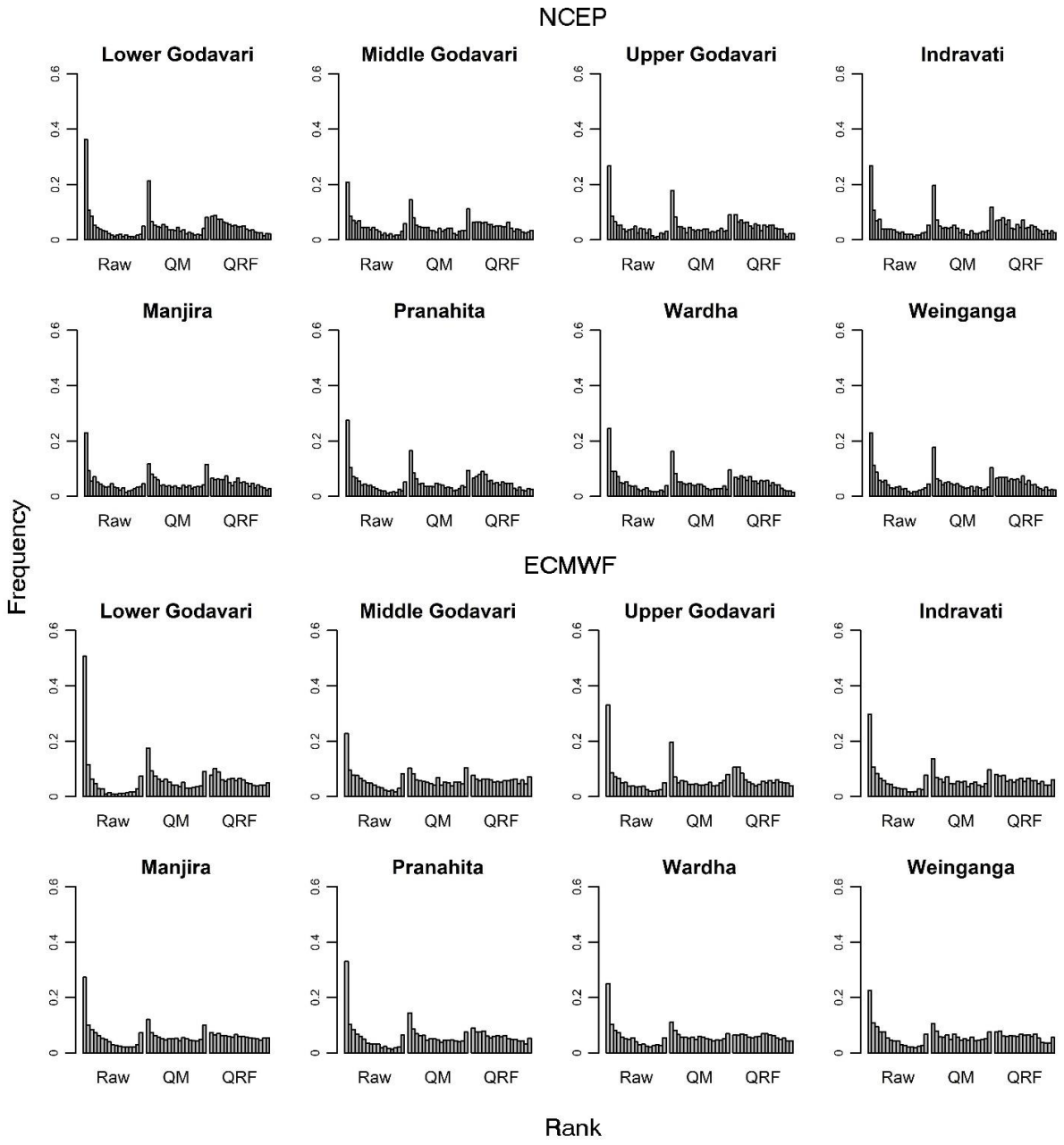

Figure 3.4 Rank histograms of raw and post-processed National Center for Environmental Prediction (NCEP) and European Center for Medium-term Weather Forecasts (ECMWF) forecasts over all subbasins for 15-day lead time.

3.3.4.2 Boxplots of Forecast Error

The performance evaluation measures described in the preceding section were used to assess the skill of raw and post-processed ensemble precipitation forecasts. The forecast error boxplots of the basin averaged 24 hourly accumulations of raw and post-processed (QM and QRF) forecasts of

NCEP and ECMWF, which are plotted against increasing amounts of observed precipitation in Figures 3.5 and 3.6, respectively, for 1 day (short) lead time. For brevity of plots, forecast boxplots of MME were excluded in this subsection. Similarly, the forecast error boxplots for higher lead times (5-day and 15-day lead times) were also provided in Figures 3.7 to 3.10. It can be observed from Figures 3.5 and 3.6 that both NCEP and ECMWF forecasts (raw and post-processed) constantly underestimate the extreme precipitation events in all subbasins.

In contrast, light to moderate precipitation events (below 20 mm and above 1 mm/day) are slightly overestimated. However, QRF post-processed forecasts are marginally better when compared to that raw forecasts at all lead times for light to moderate precipitation events. In agreement with the results of previous studies (Ridwan Siddique et al., 2015), the raw forecasts from NWPs tend to underestimate extreme precipitation events due to conditional biases (i.e., the forecast error is correlated to the observed precipitation). The forecast boxplots at higher lead times (Figures 3.7 to 3.10) show that the interquartile range (IQR) of boxplot of forecast errors is decreasing from 1 to 15-day lead times and is approximately equal to that of the magnitude of observed precipitation, indicating the inability of NWPs to forecast extreme events at higher lead times. This shows that the ensemble spread is smaller at longer lead times, indicating an over-confident forecast. This is plausibly due to the nudging of NWPs towards climatological values (Wilks, 2006). This analysis suggests that further research is necessary with a target to improve the ability of NWPs to simulate extreme precipitation events. The ECMWF forecast errors (raw and post-processed) in Lower Godavari, Indravati, Manjira and Weinganga are higher at larger precipitation events when compared to other subbasins at all lead times. The error associated with NCEP forecasts (raw and post-processed) also follows a similar pattern as that of ECMWF.

To summarize, both QM and QRF methods effectively remove the inherent biases in raw forecasts. However, they differ in their abilities to correct forecast spread and provide reliable forecasts. It is evident that the forecast error boxplots of QM-postprocessed NCEP forecasts are overly narrow at higher precipitation amounts indicating smaller ensemble spread.

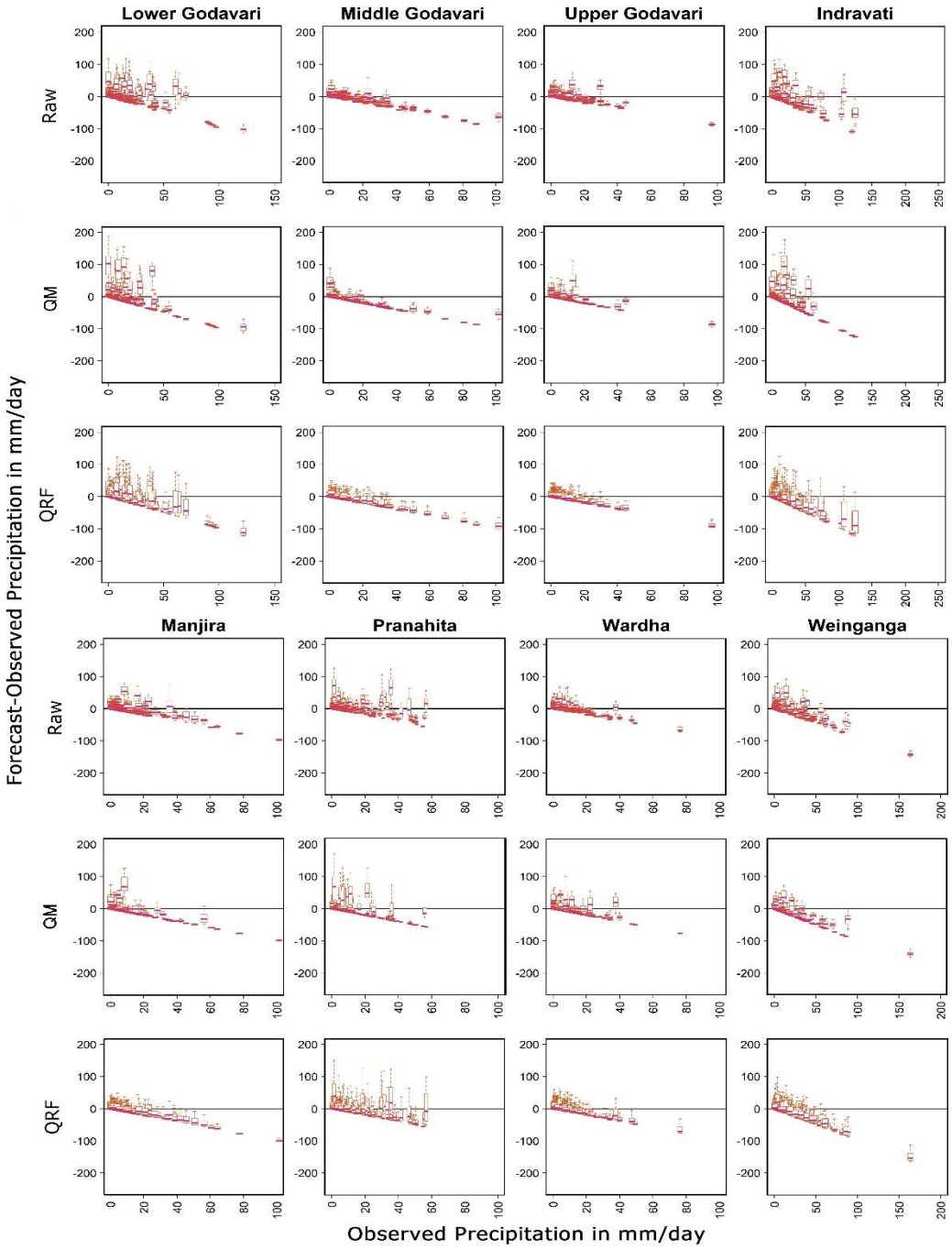

Figure 3.5 Boxplots of forecast errors of raw and post-processed National Center for Environmental Prediction (NCEP) forecasts for 1 day lead time. The boxplot closer to zero (solid black line) with a smaller spread indicates a reliable and sharp forecast.

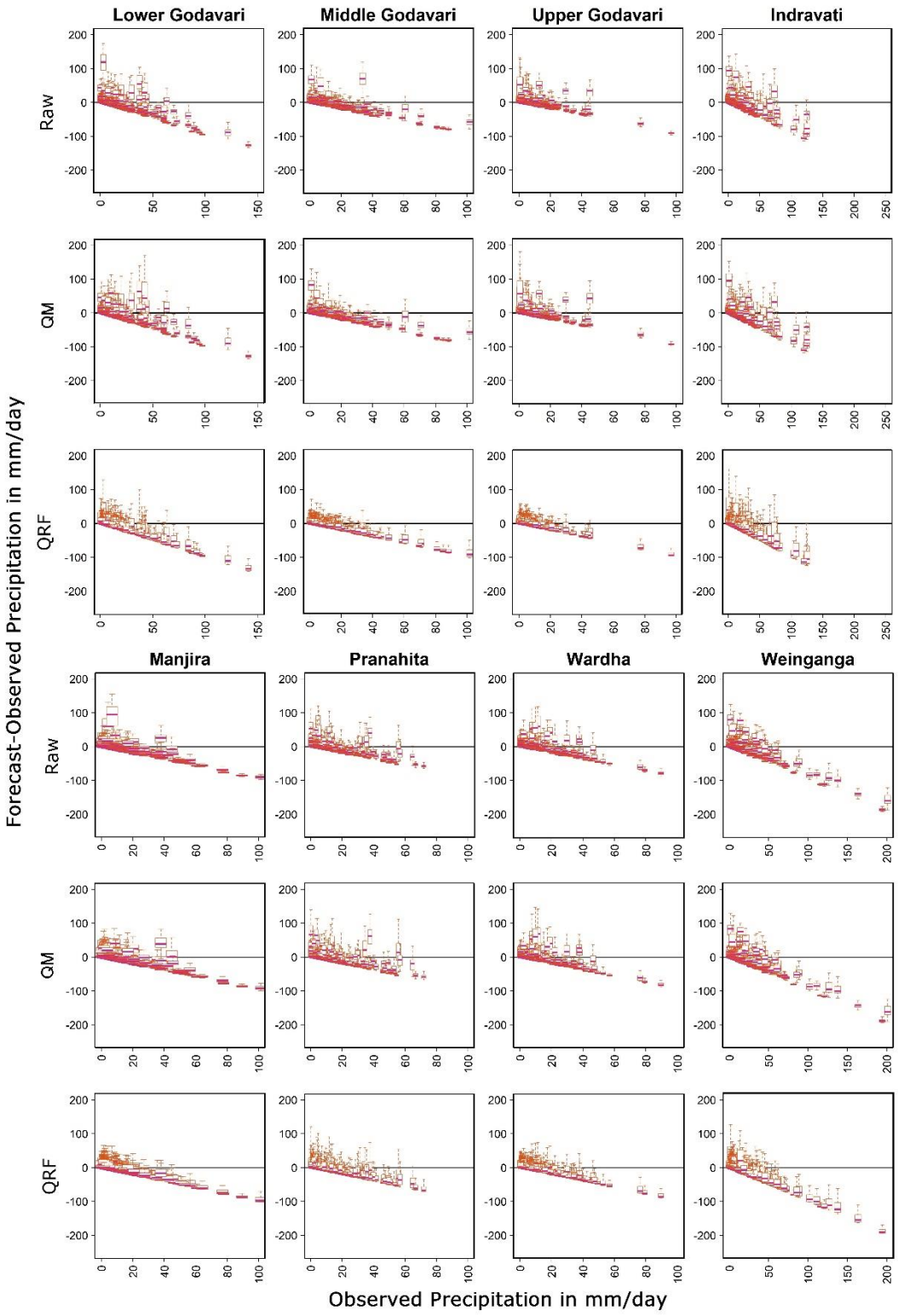

Figure 3.6 Boxplots of forecast errors of raw and post-processed European Center for Medium-term Weather Forecasts (ECMWF) forecasts for 1 day lead time. The boxplot closer to zero (solid black line) with a smaller spread indicates a reliable and sharp forecast.

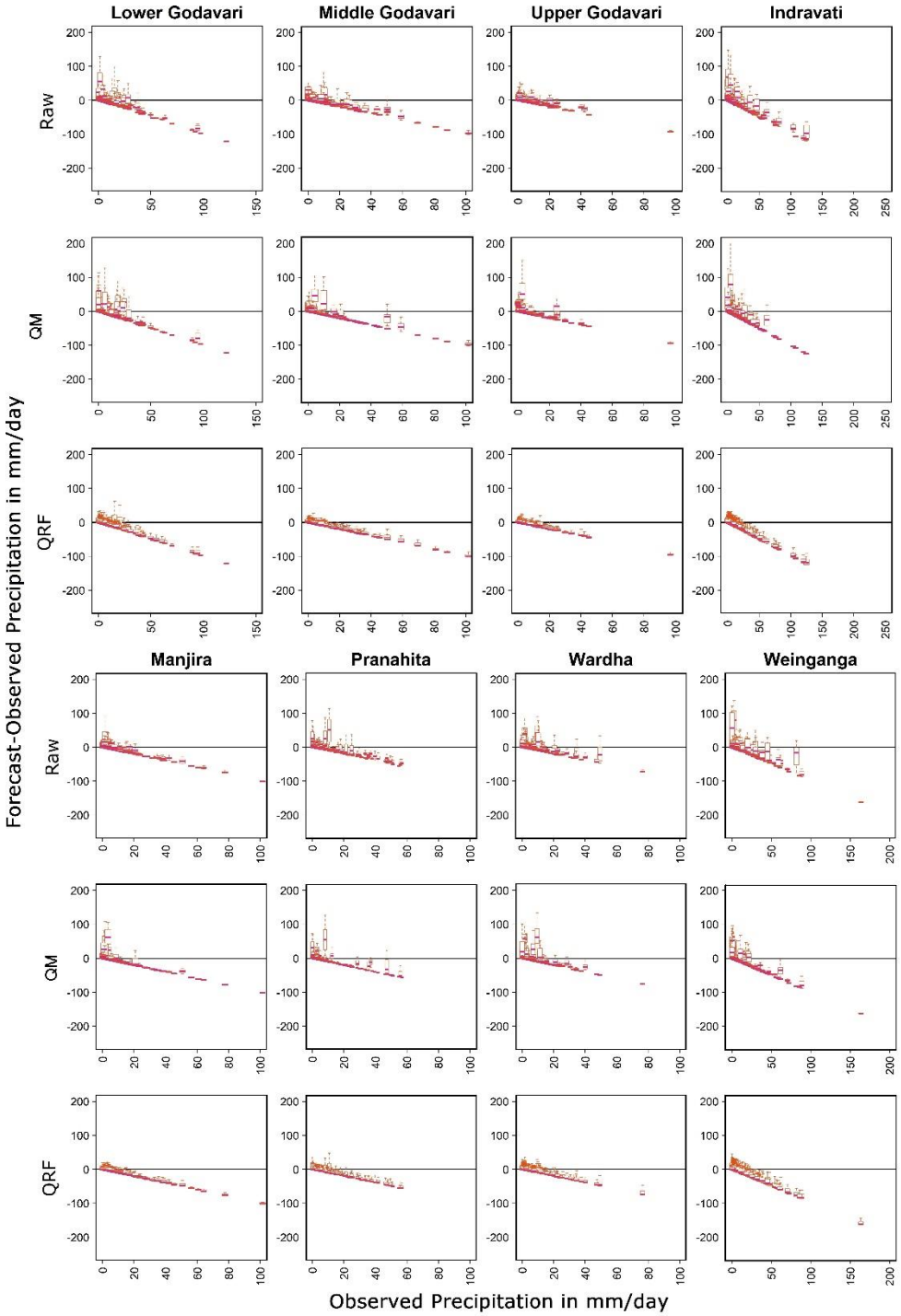

Figure 3.7 Boxplots of forecast errors of raw and post-processed National Center for Environmental Prediction (NCEP) forecasts for 5-day lead time. The boxplot closer to zero (solid black line) with a smaller spread indicates a reliable and sharp forecast.

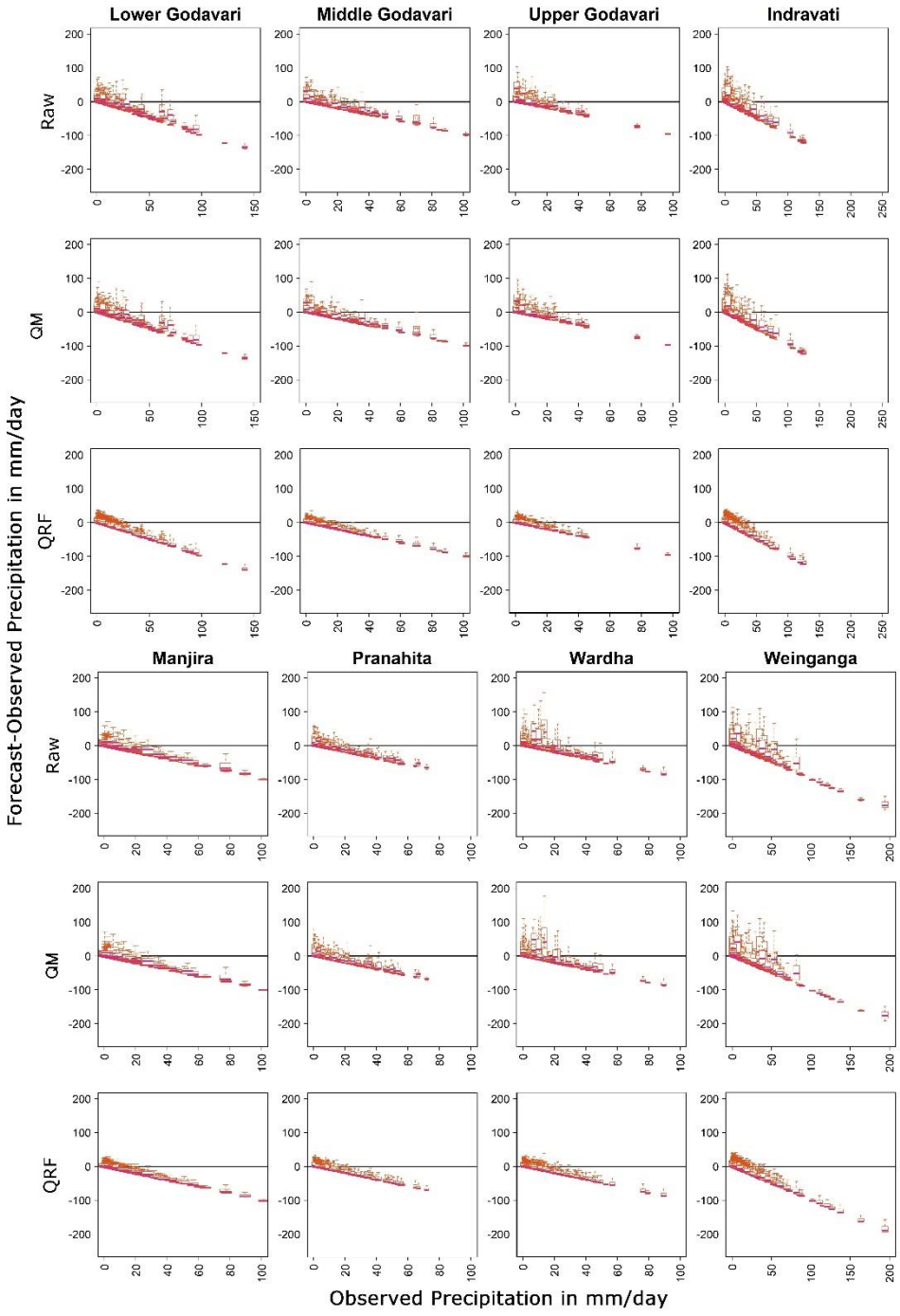

Figure 3. 8 Boxplots of forecast errors of raw and post-processed European Center for Medium-term Weather Forecasts (ECMWF) forecasts for 5-day lead time. The boxplot closer to zero (solid black line) with a smaller spread indicates a reliable and sharp forecast.

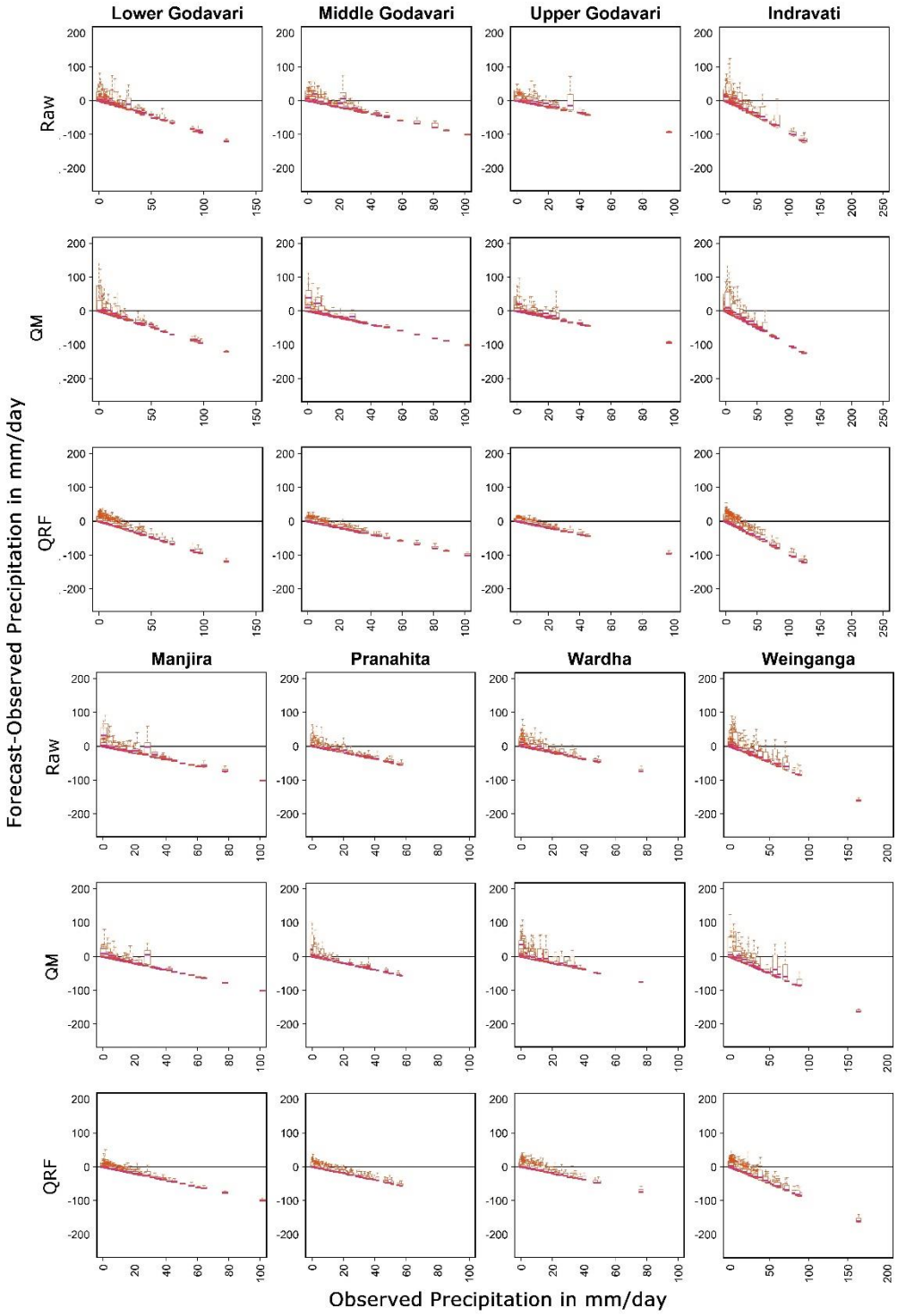

Figure 3.9 Boxplots of forecast errors of raw and post-processed National Center for Environmental Prediction (NCEP) forecasts for 15-day lead time. The boxplot closer to zero (solid black line) with a smaller spread indicates a reliable and sharp forecast.

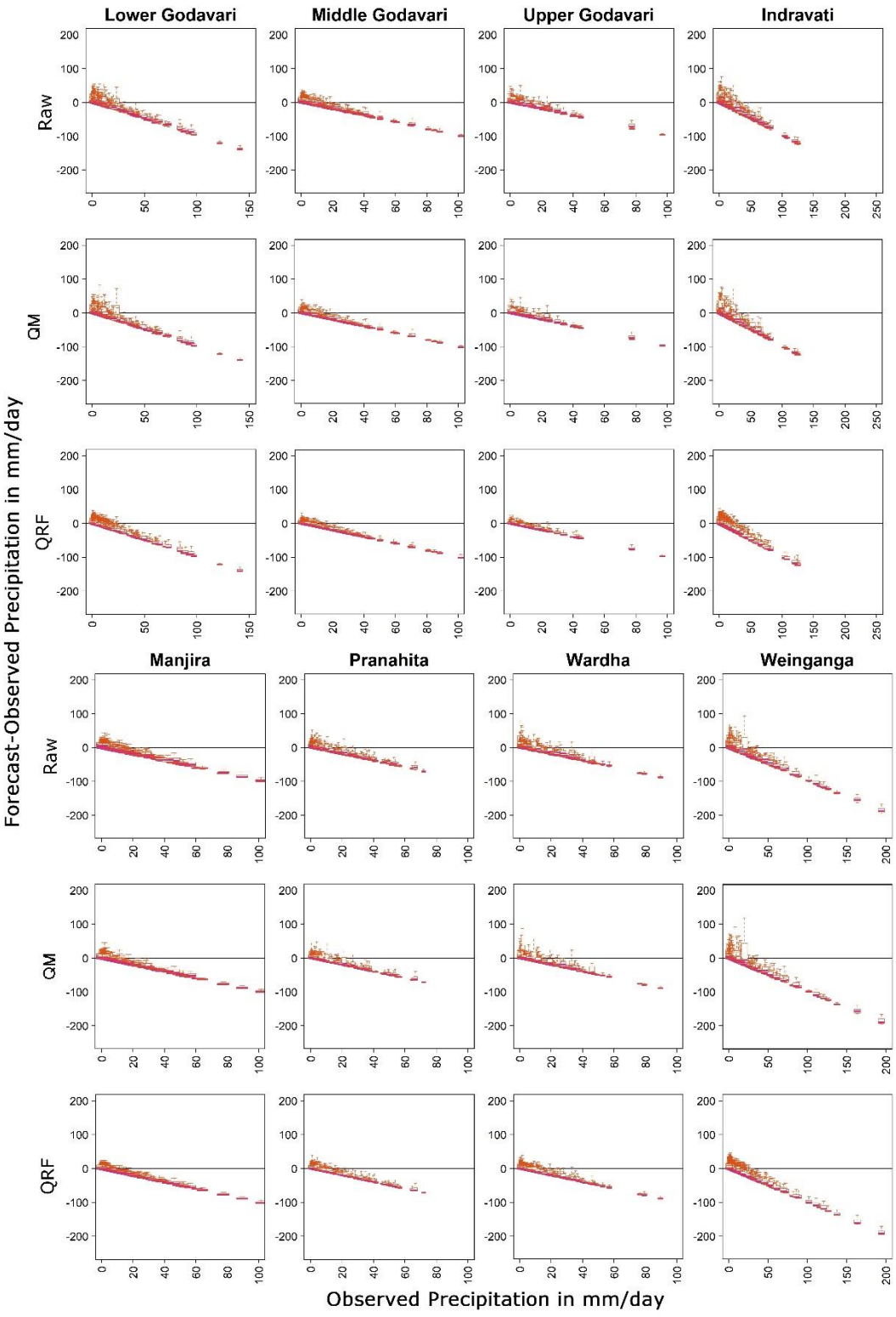

Figure 3.10 Boxplots of forecast errors of raw and post-processed European Center for Medium-term Weather Forecasts (ECMWF) forecasts for 15-day lead time. The boxplot closer to zero (solid black line) with a smaller spread indicates a reliable and sharp forecast.

3.3.4.3 Performance Evaluation of ensemble mean

The ability of the ensemble mean to capture the temporal dynamics of observed precipitation events is assessed by computing the correlation coefficient. Figure 3.11 shows the correlation coefficient between the ensemble mean of forecasted precipitation and observed precipitation for lead times varying from 1 to 15 days. It can be noticed that the correlation coefficient is high at a lead time of 1 day and tends to decrease with an increase in lead time, denoting that forecasts are unable to capture the observed events at higher lead times. Despite the fact that the correlation tends to decrease with forecast lead time, an increase in the correlation can be observed in NCEP forecasts (raw and post-processed) can be observed after a lead time of 6 days in Middle Godavari, Upper Godavari, Indravati, Wardha and Weinganga subbasins. Similar correlation patterns were observed in the study of Sharma et al. (2017), where the correlation is found to be increasing at higher lead times as the basin size increases. Furthermore, the basin averaging of precipitation plausibly alleviates the spatial forecast errors leading to improved correlation. Additionally, future research could investigate the impact of incorporating watershed-specific characteristics and topographic attributes in ensemble precipitation forecasting models to better understand and mitigate spatial forecast errors, potentially leading to further improvements in forecast performance. The correlation of raw MME forecasts is found to be higher than that of raw forecasts from individual NWP models, and is found to be as skillful as the post-processed forecasts. It is worth mentioning that the QRF post-processed forecasts were found to be better than that of their corresponding raw forecasts in terms of correlation coefficient in all the subbasins. However, QRF post-processed NCEP and MME forecasts are outperforming the remaining forecasts in all the subbasins. The mean of QM post-processed forecasts were found to be less correlated with the observed precipitation when compared with their corresponding raw forecasts. QM method can produce both positively (for ECMWF and MME) and negatively (for NCEP) skillful forecasts on the grounds that it neglects the association between the observations and raw forecasts. The skill of NCEP forecasts is found to be good in Indravati and Weinganga subbasins with a correlation greater than 0.4 event at a higher lead times (>7 days). Similarly, the correlation of ensemble mean (except QRF post-processed forecasts) and observed values is low at Lower Godavari, Manjira and Pranahita at higher lead times.

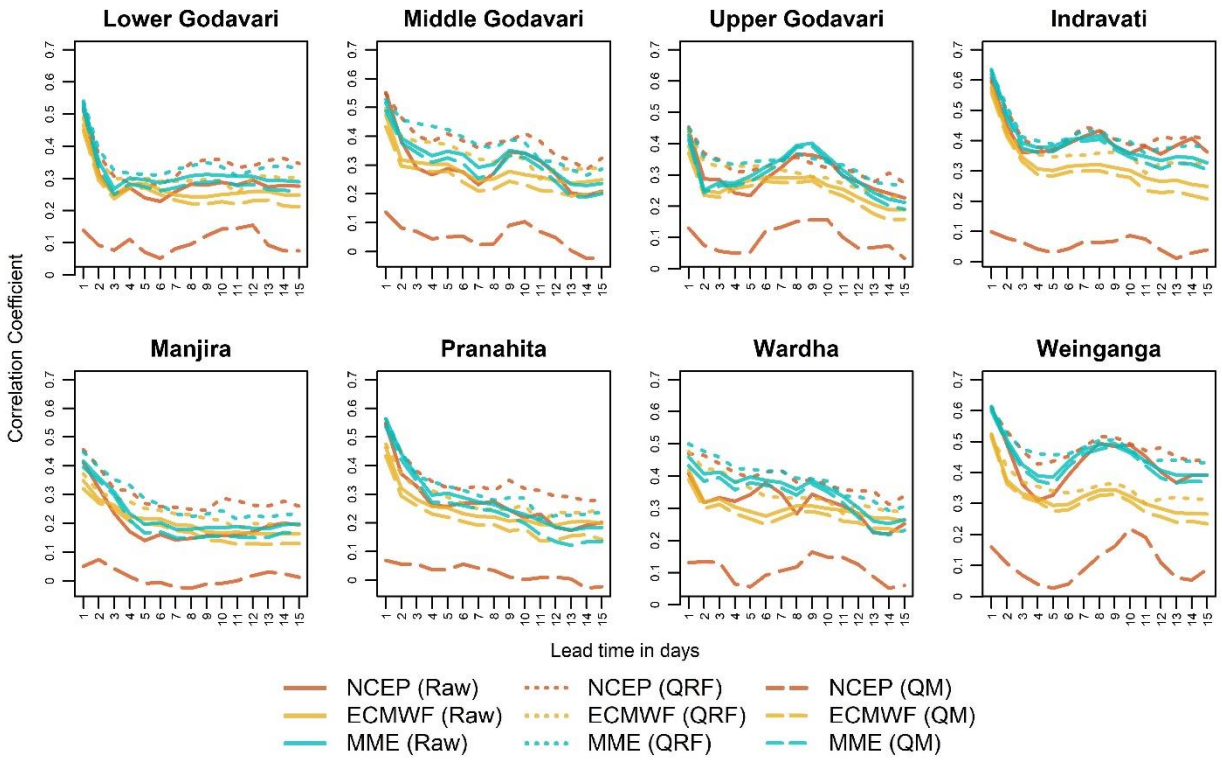

Figure 3.11 Correlation coefficient between the ensemble mean and observed precipitation for different lead times.

To understand the under/over estimation bias of the forecasts, RME between mean of the ensemble forecasts and observed precipitation is computed for varying lead times ranging from 1 day to 15 days (Figure 3.12). It can be observed that the raw forecasts are being overestimated in Upper Godavari and Wardha subbasins at almost all lead times due to the inherent biases. With increasing lead times, the post-processed ECMWF forecasts were underestimated in all subbasins. The NCEP-QRF forecasts were found to be less biased in all subbasins except Middle Godavari and Wardha subbasins. The performance of raw and post-processed MME forecasts was found to be satisfactory regarding RME at all the sub-basins indicating the multi-model ensembling alleviates the biases. The skill of post-processed ECMWF forecasts is superior in terms of RME in the Wardha subbasin. The superior performance of post-processed ECMWF forecasts in terms of Relative Mean Error (RME) in the Wardha subbasin may be attributed to several factors. It's possible that the ECMWF model, in its raw form, exhibits certain biases or uncertainties that are effectively mitigated through the post-processing technique employed. This could be related to the model's representation of local meteorological conditions, topographical features, or other dynamic factors specific to the Wardha subbasin. Additionally, the post-processing method itself may have been particularly effective in calibrating and refining the ECMWF forecasts in this

specific region. The verification results in terms of correlation coefficient, and RME were found to be in line with the findings of Siddique et al. (2015) and Sharma et al. (2017).

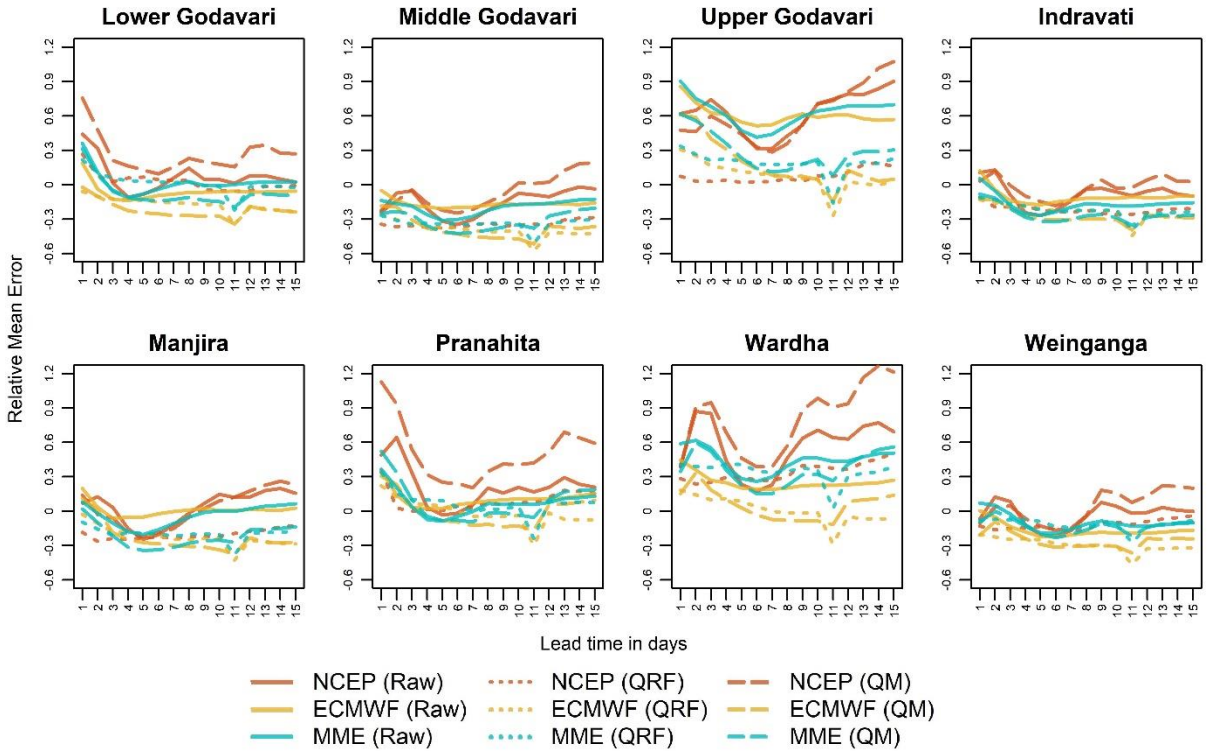

Figure 3.12 RME between the ensemble mean and observed precipitation for different lead times.

3.3.4.4 Spread-skill relationship

The spread-skill relationship of the NCEP, ECMWF and MME forecasts are evaluated and plotted in Figure 3.13 for lead times ranging from 1 to 15 days. The skill score (RMSE in this study) should be equal to the mean of ensemble spread for consistent and reliable ensemble forecast. RMSE values higher than the spread indicates under-dispersive (over confident) forecasts and vice versa. From Figure 3.13, it can be noticed that NCEP, ECMWF and MME forecasts (raw and post-processed) are under dispersive at all subbasins at all lead times. The ensemble spread of raw forecasts is low at smaller lead times and is increasing with increasing lead time. The ensemble spread of post-processed forecasts is higher than that of raw forecasts at smaller lead times alleviating the under-dispersion of raw forecasts. However, the spread of both raw and post-processed forecasts converges at lead times from 11 to 15 days.

It should be noted that the RMSE at Upper Godavari, Pranahita and Wardha subbasins are considerably greater (>1) than that of RMSE at remaining subbasins. It can also be observed that the RMSE of QM post-processed forecasts are high, indicating poor skill compared to QRF post-

processed forecasts. A good linear relation is observed between RMSE and spread of QRF post-processed forecasts compared to other forecasts. It shows the ability of QRF post-processed forecasts to preserve the spread-skill relationship.

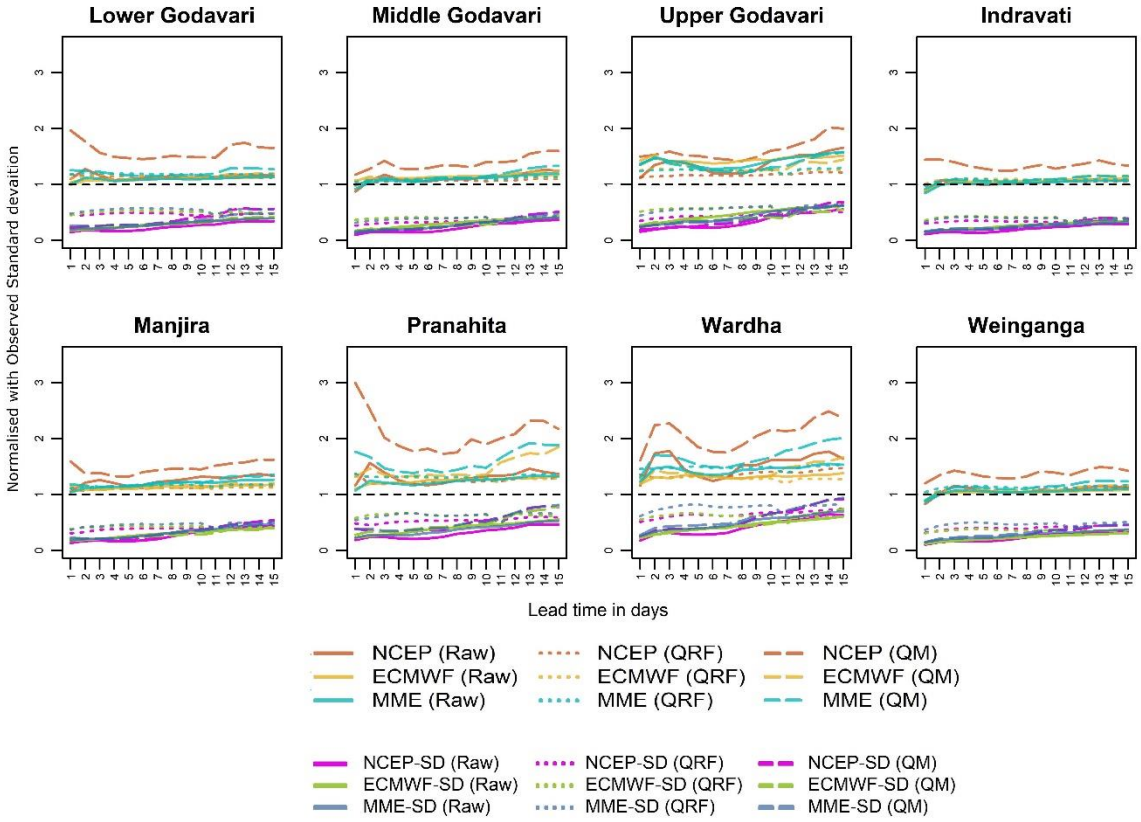

Figure 3.13 Ensemble spread-skill relationship of the National Centre for Environmental Prediction (NCEP) and European Centre for Medium-term Weather Forecasts (ECMWF) forecasts for different lead times over all subbasins.

3.3.4.5 Continuous Ranked Probability Score (CRPS)

CRPS generalizes the Mean Absolute Error (MAE) to the case of probabilistic forecasts. Lesser the mean CRPS value, the more accurate the forecasts are. The mean CRPS value of ensemble forecasts for various lead times ranging from 1 day to 15 days is plotted in Figure 3.14. The general tendency of the mean CRPS to increase with respect to lead time. The mean CRPS value of the raw and QM post-processed forecasts is high at shorter lead times (1 to 3 days) at most subbasins. However, the mean CRPS values are low in all subbasins except Lower Godavari, Indravati and Weinganga. The higher values of mean CRPS and correlation coefficient in Indravati and Weinganga subbasins indicates that the ensemble spread is comparatively higher in comparison to the ensemble spread in other basins. However, the good correlation between ensemble mean and observations shows the skill of ensemble spread in capturing the observed events. This problem of high/low ensemble spread leading to under/over confident forecasts can be alleviated by employing post-processing methods (Raftery et al., 2005). It can be noticed from Figure 3.14 that the performance efficiency in terms of CRPS has increased with the application of QRF to NCEP, ECMWF and MME forecasts in all subbasins. However, the MME forecasts are comparatively better than those from individual NWPs in all subbasins. The QRF post-processed MME and NCEP forecasts were more skilful in all sub-basins at all lead times.

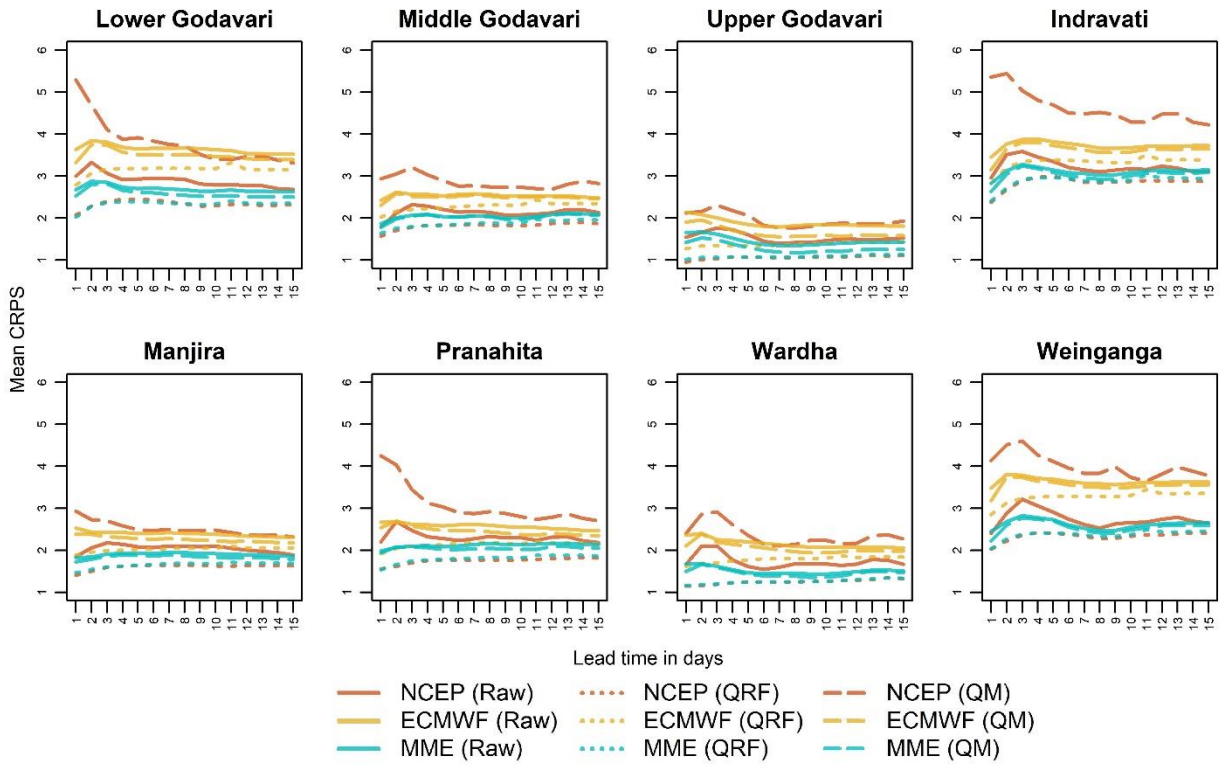

Figure 3.14 Mean CRPS value of National Center for Environmental Prediction (NCEP), European Center for Medium-term Weather Forecasts (ECMWF) and MME forecasts as a function of lead time over all subbasins.

3.3.4.6 Reliability Diagrams

With the intent to evaluate the reliability of NCEP and ECMWF forecasts, reliability diagrams are plotted for 1 day (Figure 3.15), 5 days (Figure 3.16) and 15 days (Figure 3.17) lead times. In an ideal case, the curve of observed frequency for given forecast probabilities should be close to the diagonal. The log-transformed curves in the inset of reliability plot indicates the sharpness of the forecast i.e., how frequently each probability was issued. Figure 3.13 shows that the raw NCEP, ECMWF and MME forecast are under dispersive (overconfident) for forecast probabilities greater than 0.4 overall subbasins at 1 day lead time. However, it can be observed that the reliability has increased in the post-processed forecasts at 1 day lead time. It is also essential to notice that QM post-processed forecasts are slightly more reliable (closer to the diagonal line) than QRF post-processed forecasts. However, the QM post-processed forecasts were under-confident for forecast

probabilities less than 0.8 for Middle Godavari, Indravati, Manjira, Wardha and Weinganga. The QRF post-processed forecasts are slightly over-confident for forecast probabilities less than 0.8 in most subbasins and are more reliable than QM post-processed forecasts for higher forecast probabilities (>0.8). No significant difference has been observed between the performance of NCEP, ECMWF and MME forecasts post-processed with the QRF method. However, it can be noticed from Figures 3.16 and 3.17 that the reliability of both raw and post-processed forecasts declines with increasing lead times.

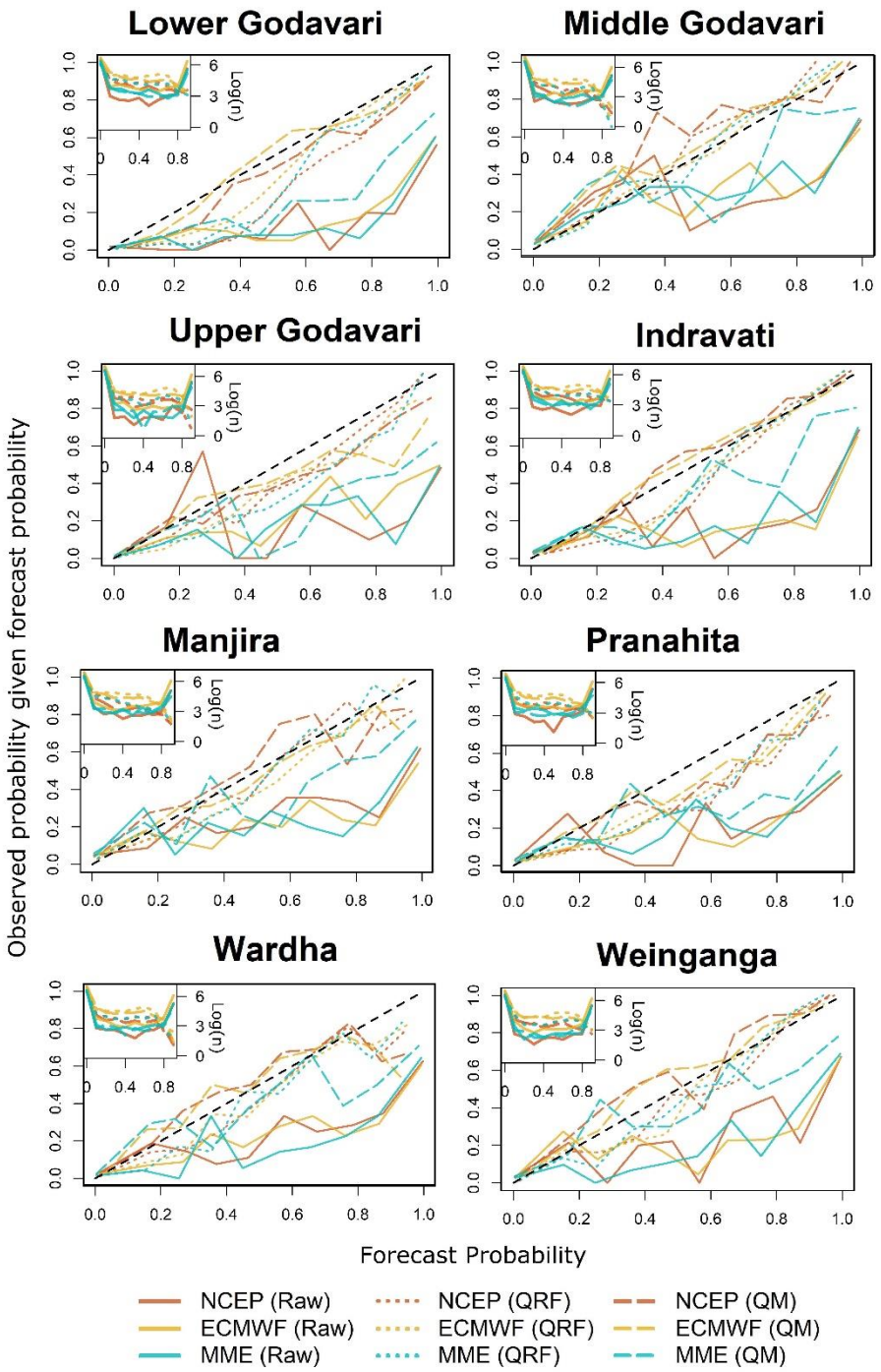

Figure 3.15 Reliability diagram of National Centre for Environmental Prediction (NCEP), European Centre for Medium-term Weather Forecasts (ECMWF) and MME forecasts for 1 day lead time over all subbasins.

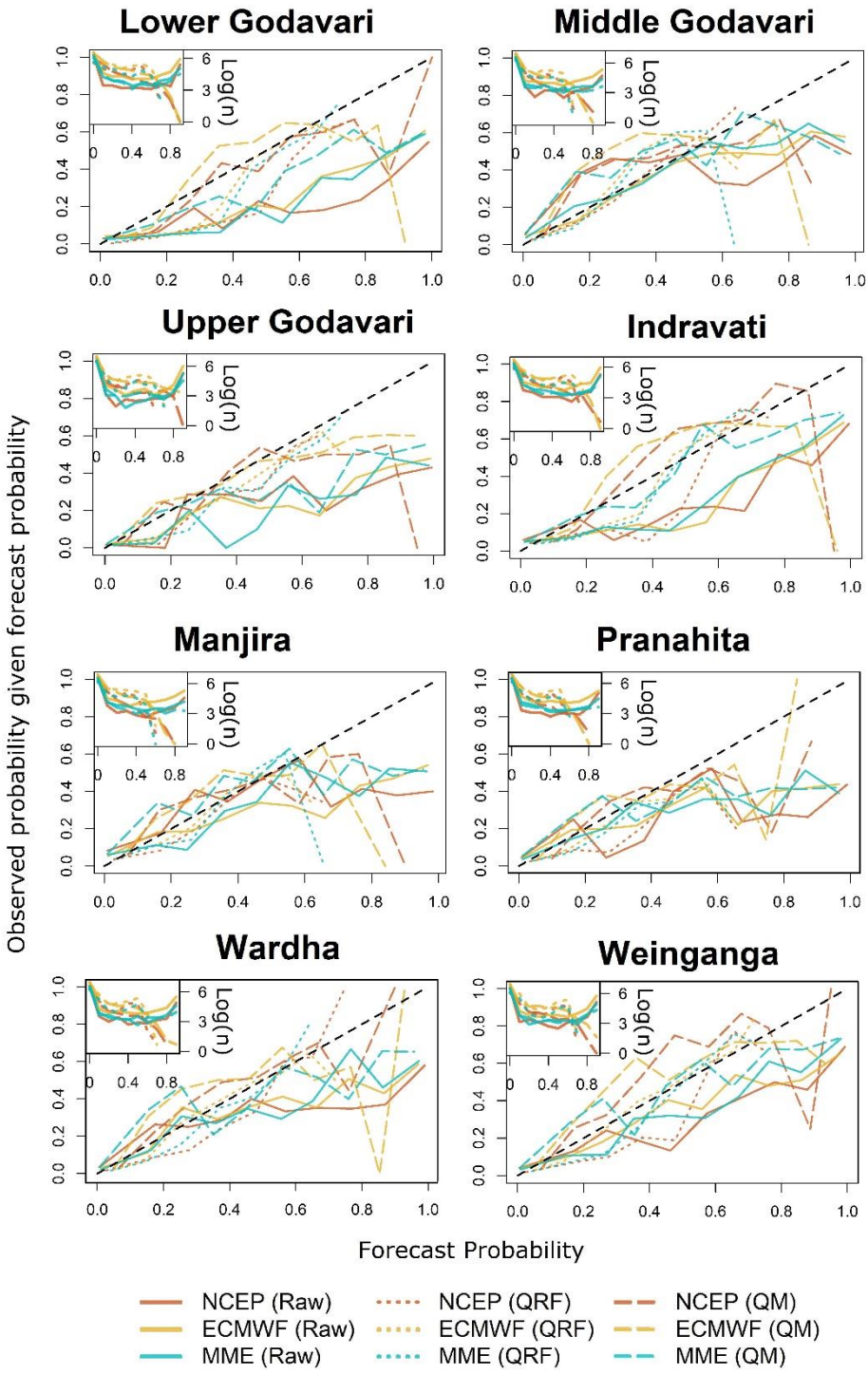

Figure 3.16 Reliability diagram of National Center for Environmental Prediction (NCEP) and European Center for Medium-term Weather Forecasts (ECMWF) forecasts for 5-day lead time over all subbasins.

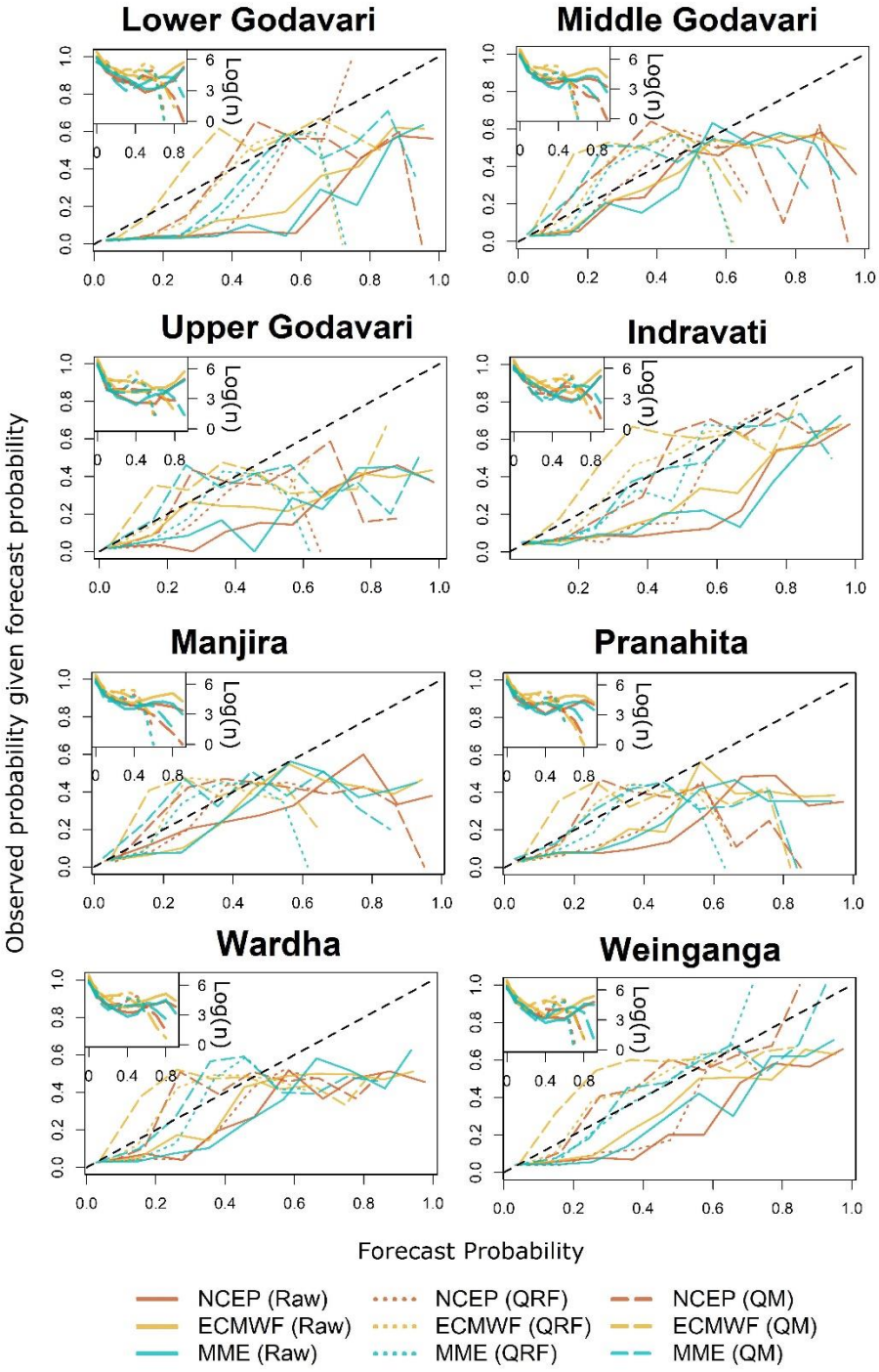

Figure 3.17 Reliability diagram of National Center for Environmental Prediction (NCEP), European Center for Medium-term Weather Forecasts (ECMWF) and MME forecasts for 15-day lead time over all subbasins.

3.3.4.7 Area under ROC curve (AUC)

The area under the ROC curve (AUC) is a scalar measure that quantifies the discrimination ability of forecasts between occurrences and non-occurrences. The ability of event discrimination of NCEP, ECMWF and MME forecasts in terms of area under the ROC curve (AUC) is plotted for different lead times across all subbasins in Figure 3.18. From the figure, it can be observed that AUC at all subbasins at all lead times is greater than 0.75, indicating that the forecasts are 'useful' (as mentioned in the methodology section). It can also be observed that the value of AUC is decreasing with increasing lead time at all subbasins. This indicates that the discrimination ability of forecasts declines with lead time, suggesting an increased false alarm rate. The AUC values of raw NCEP forecasts at 3- to 5 days lead time are deteriorating at Middle Godavari, Manjira and Wardha subbasins, which is efficiently alleviated by post-processing. The overall performance of post-processed forecasts is better than raw forecasts highlighting the potential of post-processing. QRF post-processed NCEP and MME forecasts outperform other forecasts at all subbasins for all lead times. It should also be noted that the AUC values at Manjira and Pranahita are lower than other subbasins.

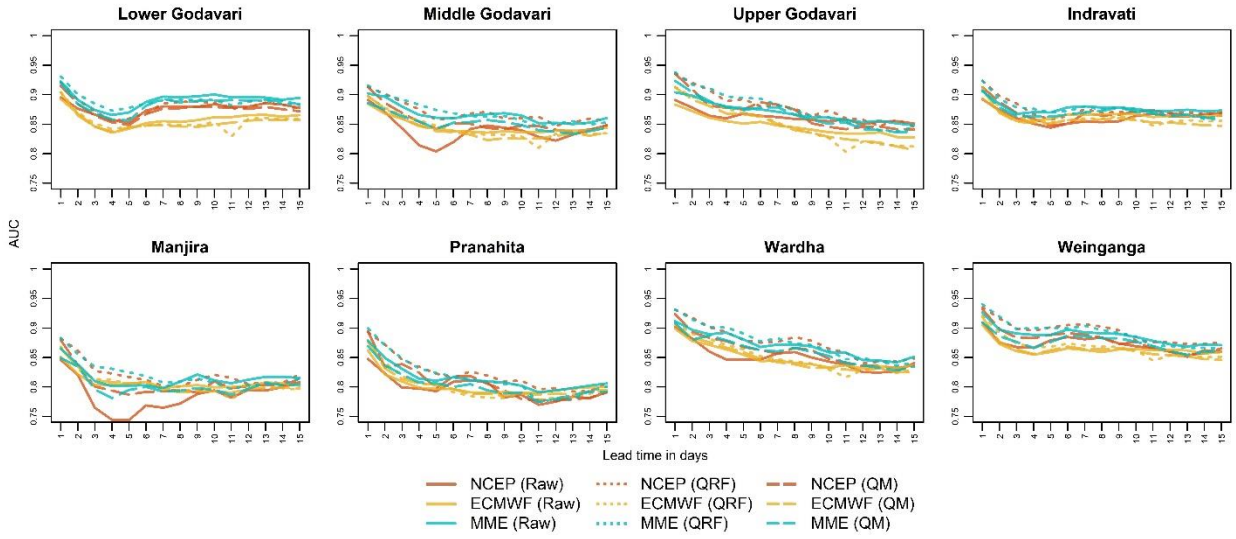

Figure 3.18 AUC values of the National Center for Environmental Prediction (NCEP), European Center for Medium-term Weather Forecasts (ECMWF) and MME forecasts as a function of lead time over all subbasins**.**

3.3.5 Summary and Conclusions

In this study, the skill of ensemble precipitation forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF), National Centers for Environmental Prediction (NCEP) from "The Observing System Research and Predictability Experiment" (THORPEX) Interactive Grand Global Ensemble (TIGGE) database is evaluated over the eight subbasins of Godavari River Basin, India. These two forecasting systems are preferred as they are operational, have multiyear datasets, and/or capture scenarios that forecasters are interested in (Sharma et al. (2017)). Along with them, the skill of the multi-model grand ensemble (MME) generated by integrating NCEP and ECMWF forecasts is also verified in this study. Two statistical post-processing methods emphasize the ability of post-processing methods to alleviate the conditional biases in the raw forecasts. The verification was conducted for daily accumulations for different lead times (1 to 15 days) in hindcast mode. The deterministic and probabilistic measures (forecast error box plots, correlation coefficient, RME, mean CRPS, spread-skill relationship, rank histograms, reliability diagram, AUC) are employed to verify the skill of basin averaged raw and post-processed forecasts in comparison to the observed data during the period from 2016 to 2020. The key findings of this verification study are mentioned below.

- The skill of both NCEP and ECMWF raw forecasts in capturing the observed extreme precipitation events is poor across all lead times and statistical post-processing also could not alleviate this problem. This highlights the need to develop the underlying physics of NWPs to accurately forecast extreme precipitation events.
- The correlation between ensemble mean and observed precipitation declines with increasing lead time, whereas RME does not depend upon lead time. The ensemble mean of QRF post-processed NCEP and MME forecasts outperforms other forecasts regarding correlation coefficient and RME at all lead times in all subbasins.
- The ensemble spread-error relationship in the post-processed forecasts is improved in comparison to raw forecasts. It is also found that QRF outperforms QM in preserving the ensemble spread-error relation.
- The rank histograms suggest that NCEP and ECMWF raw forecasts are under-dispersive and biased at all subbasins. However, post-processing of raw forecasts has alleviated the problem of bias.
- Based on the analysis of reliability diagrams, the raw NCEP and ECMWF forecasts tend to be overconfident, whereas the post-processed forecasts are performing well at 1 day lead

time. However, the reliability declines with increasing lead times due to overconfident forecasts.

- The AUC values are greater than 0.75 at all lead times and all subbasins, showing that the forecasts are 'useful'. The discrimination ability of forecasts in terms of AUC is found to be decreasing with lead time, indicating a higher false alarm rate.
- The results of the analysis suggest that the performance of raw MME is comparatively better than raw NCEP and ECMWF forecasts in terms of the employed verification measures. However, the QRF post-processed NCEP and MME are performing equally well. Hence, keeping in mind of the computational cost, 20-member QRF post-processed NCEP forecasts are suggested for hydrologic forecasting applications over the present study area.
- The overall performance of NCEP and MME forecasts is better than ECMWF forecasts. The performance of QRF post-processed forecasts is outperforming QM post-processed and raw forecasts. The underperformance of QM is because it does not account for the under/over spread in the raw forecasts, and it does not consider the correlation between raw forecasts and observations, leading to unreliable forecasts.
- The performance of post-processed forecasts in Lower Godavari, Middle Godavari, Indravati, Manjira and Weinganga is satisfactory regarding both deterministic and probabilistic measures. The performance of QRF post-processed NCEP and MME forecasts was satisfactory at most of the subbasins for all verification measures.

Even though the aforementioned verification results provide diagnostic information about the skill of ensemble precipitation forecasts, they do not give any information on how to improve the NWP models. More weather factors than only precipitation are needed to properly understand the physical and environmental factors linked to forecast errors and skill (Moore et al., 2015). In the future, we plan to explore and evaluate various forecasting scenarios to assess the benefits of integrating the outputs from different precipitation forecasting systems, hydrologic model structures, and hydrological post-processing techniques to potentially improve flood forecasting across spatiotemporal scales.

Chapter 4

Impact of model resolution on simulation of flood peaks

4.1 Introduction

Flood is a common natural disaster worldwide that can cause catastrophic impacts on day-to-day operations (Hirabayashi et al., 2013). Unlike other natural disasters, floods can be forecast in advance as a preparedness measure (H. L. Cloke and Pappenberger, 2009). The hydrological and meteorological models are often integrated to forecast flood hydrographs, and the ensemble system is termed as hydro-meteorological forecasting system (Das et al., 2022). Hence, accurate prediction of floods through various hydrological models is the need of the hour for mitigating the effects of floods. This can be done by appropriately selecting hydrological model structure along with its spatial and temporal resolution. In general, hydrological models are mainly classified based on spatial-scale and process description. In spatial-scale based classification, the models can be divided into lumped, distributed and semi-distributed models. Based on the process description, hydrological models can be broadly classified into data-driven, physics-based and conceptual models (Acero Triana et al., 2019; Ghimire et al., 2020; Madsen, 2000). Data driven models, as they establish empirical relation between input and output time series, do not incorporate any physical process in the watershed. Physics-based distributed models are based on understanding of the physics of the processes involved in the water circulation, such as the transfer of mass, momentum, and energy. However, these models are data intensive and demand more computational resources. Conceptual models, also known as ‘grey-box’ models, consist of very few components (non-linear reservoirs) which are simplified representations of elements in a hydrological system. The utilization of conceptual-based hydrological models has been amplified to a large extent in the past few years, attributable to their simplicity and computational efficiency (Tran et al., 2018).

Theoretically, the fully distributed models are generally expected to perform better than the lumped models. However, contradictory results were reported stating that increasing the model complexity might not lead to better simulations when evaluated at the outlet stations (Das et al.,

2008; Reed et al., 2004; Viney et al., 2005). Hence, due to their parsimonious model structures, lumped and semi-distributed conceptual models are well adapted for water resources management and flood forecasting (Perrin et al., 2001). Apart from their ability to consider spatial heterogeneity, an important advantage of semi-distributed models is their ability to simulate streamflow at interior locations where observed streamflow is unavailable for model calibration (Khakbaz et al., 2012; Sharma and Regonda, 2021). In the semi-distributed approach, the first step is to divide the basin into sub-basins based on the density of the flow gauges and the second step consists of subdivision of the sub-basins into homogenous elements with distinct hydrologic responses. Generally, it is expected that the performance of any numerical model, in principle, should improve with increasing model resolutions. But the scarcity of finer-scale spatially resolved data poses a major challenge to increase the model resolution. Hence, the characteristics of both distributed and lumped models can be found by opting for semi-distributed models and integrating sub-grid parameterization schemes in hydrologically homogenous areas.

Organization of spatial patterns of various physical characteristics of catchment, such as soil characteristics, soil moisture and vegetation type, into homogeneous units have a significant influence on catchment runoff. This process of splitting the landscape in hydrologically homogeneous regions is called 'landscape discretization'. In congruence with the dominant hydrological processes, the topography is the most influential factor in the discretization process, where are soil and vegetation are determining factors in the lower hierarchy (Pilz et al., 2017). However, there is no universally accepted definition of discretization, it typically consists of a stratification scheme of partitioning a hydrological basin into sub-basins, which are sequentially divided into irregularly shaped hydrologically uniform entities (Krysanova et al., 1998). Several methods have been developed to address the discretization issues, for which no generic solution has been found so far (Beven, 2006). Some of the widely used watershed discretization methods are Representative Elementary Area (REA) (Wood et al., 1988) Wood et al. (1988), Grouped Response Units (GRUs) (Kouwen et al., 1993) and Hydrological Response Units (HRUs) (Flügel, 1995; Leavesley et al., 1983). However, these discretization schemes are not often used for conceptual models as REA is primarily influenced by topography, GRUs consider a range of land cover characteristics with uniform meteorological forcing, and the HRU approach cumulates the generated streamflow over all HRUs without representing the water flow pathways.

The influence of different discretisation methods on hydrologic responses has been acknowledged in multitudes of studies (Caldeira et al., 2019; Euser et al., 2015; González et al., 2016; Kumar et al., 2010). However, this influence has been assessed extensively in grid-based models (Euser et

al., 2015; Melsen et al., 2016; Molnar and Julien, 2000; Sulis et al., 2011) as is it is relatively easy to change the grid resolution. Previous studies have investigated the role of spatial variability of rainfall and spatial heterogeneity of catchment on the hydrologic responses of a catchment using semi-distributed models by dividing the catchments into sub-basins of roughly uniform size based on topography (Booij, 2005; Das et al., 2008; Lobligeois et al., 2014). However, for a semi-distributed conceptual model, it is desirable that the catchment properties of a sub-basin should be homogenous, satisfying the assumption of a lumped model. Further, there is a threshold of subdivision level above which no more improvements can be achieved (Haghnegahdar et al., 2015; Han et al., 2014; Wood et al., 1988). For conceptual models there should be a lower limit of splitting the landscapes simply because they represent spatial average behaviour (Das et al., 2008). Hence, finding a compromise between the optimum finite model resolution through landscape discretization and computational burden is quintessential. For this purpose, a nested discretization approach would be more appropriate, where the catchment is divided into sub-catchments only if there is significant landscape heterogeneity.

In addition to proper landscape discretization, it is also important to choose among the hydrological models that are classified based on the updating scheme of initial hydrologic conditions. Based on this classification, hydrological models are divided into continuous and event-based models. In continuous models, a warm-up period is required to mitigate the effect of arbitrarily selected values of initial hydrologic conditions. These models often require long-term continuous data which is a major drawback in operational forecasting perspective. Event-based hydrological models are extensively used for operational flood forecasting purposes since they are easier to calibrate and require only data at the event scale. Previous studies reported the performance of event-based lumped conceptual models (Berthet et al., 2009; Bournas and Baltas, 2021; Kalantari et al., 2015). However, as per the authors' best of knowledge, no studies have been reported so far on the influence of nested discretization on the hydrologic response of an event-based conceptual model.

This study addresses the following research question: Does a nested discretization scheme lead to improved streamflow simulations for a semi-distributed conceptual hydrologic model? Does the predictive performance of a model increase with increasing model resolution (level of discretization)? What is the role of calibration strategies on the performance of model simulations at different resolutions? To address these questions, the current study attempts to analyse the effect of varying spatial scales (based on nested discretization) on the streamflow response simulated by continuous and event-based semi-distributed conceptual models. For this purpose, a conceptual rainfall-runoff model, modèle du Génie Rural à 4 paramètres Journalier (GR4J) has been selected.

The predictive performance at different spatial representations using lumped, semi-distributed and semi-lumped models was assessed using different performance evaluation measures. The performance of these models in predicting the flood hydrographs of the catchments is evaluated by setting up event-based models. These models are applied to the Jagdalpur and Wardha catchments located in India.

4.2 Study Watersheds and Hydrometeorological Data

4.2.1 Study Area

The Jagdalpur and Wardha catchments were chosen as study watersheds in this study. They are subbasins in Godavari River Basin, India. The catchments of Jagdalpur and Wardha were delineated from the flood forecasting stations Jagdalpur (for Jagdalpur basin) and Bamni (for Wardha basin), respectively. The delineated catchments have an area of 7382 and 46282 km^2 for Jagdalpur and Wardha, respectively. The location map of the study watershed, along with their Digital Elevation Map, is shown in Figure 4.1.

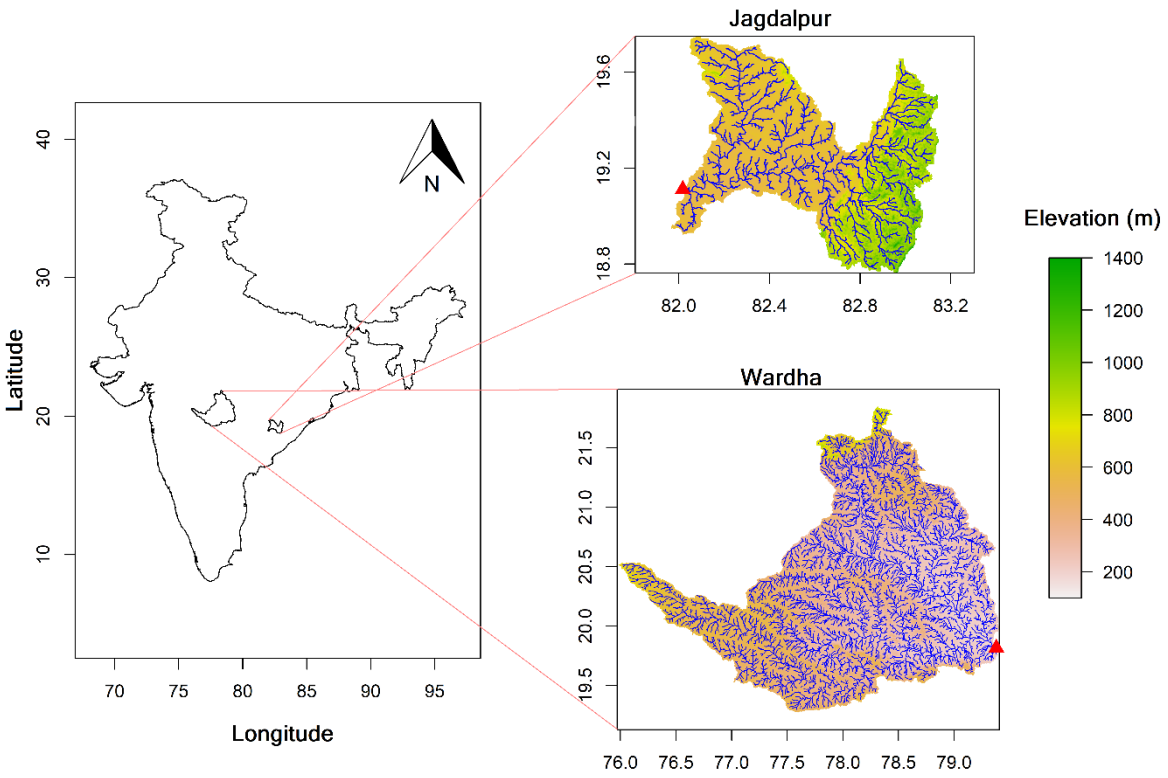

Figure 4.1 Location map of Jagdalpur and Wardha catchments with detailing of river streams, elevation, and flood forecasting stations.

4.2.2 Hydrometeorological Data

The daily streamflow data at Bamni (for Wardha basin) and Jagdalpur (for Jagdalpur basin) are available at the India-WRIS portal from 1965-2018. The observed flood events are separated from the streamflow data using the warning levels obtained from the appraisal report on Flood Forecasting and Warning Network Performance by Central Water Commission (CWC, 2018b). In this study, the daily gridded datasets from the India Meteorological Department (IMD) at a spatial resolution of 0.25^{o}x0.25^{o} for Precipitation and 1^{o}x1^{o} for maximum and minimum temperature were used as meteorological forcing data (Pai et al., 2014; Srivastava et al., 2009). The elevation data in the form of a Shuttle Radar Topography Mission (SRTM) based Digital Elevation Model (DEM) with a spatial resolution of 90m was obtained to delineate the basins and extract the stream network.

In this study, the Natural Resources Conservation Service (NRCS) Curve Number (CN) is used to assess the landscape heterogeneity. Curve Number is a empirical parameter that indicates the runoff response considering the soils, land use, land cover and antecedent wetness conditions. The

Global Gridded Curve Number dataset (GCN250), developed by Jaafar et al. (2019), is a global gridded curve number dataset available at a spatial resolution of 250 m × 250 m. GCN250 dataset is developed using a 300 m × 300 m ESA-CCI land use land cover database and a 250 m × 250 m Hydrologic Soil Group (HYSOG) global soil database. GCN250 dataset has been used in hydrological applications such as flash flood susceptibility prediction (Ekmekcioğlu et al., 2022), prediction in ungauged basins (Estacio et al., 2021), rainfall-runoff modelling (Ogarekpe et al., 2022; Rahimi et al., 2021) and satisfactory results were reported.

4.3 Landscape Discretization

Runoff generation in a catchment is highly influenced by the spatio-temporal characteristics of rainfall and physical features of catchments, such as topography, vegetation, and soil properties. A nested discretization approach is employed in this study to split the watershed into homogeneous spatial units. For this purpose, the watershed is divided into sub-basins based on stream gaging stations available in the interior locations and are again delineated into finer sub-basins based on topography. To assess the homogeneity of characteristics of the delineated sub-basins, Curve Number (CN) has been employed in this study as an indicator to address the vegetation and soil characteristics of the catchment. CN is a widely used conceptual parameter in the field of hydrology that captures the catchment physical characteristics that affect runoff generation in a single value. The value of CN ranges from 1–100 (a larger value indicates a higher runoff ratio). The value of CN is determined from NRCS lookup tables accounting for the combinations of various land use/land cover characteristics, four hydrological soil types (A, B, C and D), which are further categorized as good, fair, and poor based on their hydrological condition, and antecedent soil moisture (AMC) conditions (Type I - dry, Type II - moderate, Type III - wet). The reference CN values in the NRCS lookup tables have been extracted experimentally from rainfall and runoff measurements over various geographic, soil and land management conditions.

GCN250, a global gridded curve number dataset, has been used in this study to determine the spatial distribution of the curve numbers in the delineated sub-basins. A set of spatial moments (δ_1 and δ_2) are used to quantify the spatial heterogeneity of the CN in the sub-basins. These spatial moments proposed by Zoccatelli et al. (2011) are used to quantify the spatial variability of rainfall in previous studies (Douinot et al., 2016; Emmanuel et al., 2015; Garambois et al., 2014). These indices are computed as the spatial moments of a catchment characteristic (i.e., CN in this study),

which describes the interactions between spatial patterns of the corresponding characteristic and basin morphology. The equations of the spatial moments are given in Eq. 4.1 and 4.2.

$$\delta_1 = \frac{\frac{1}{C}\sum c_i \eta_i}{\frac{1}{A}\sum \eta_i} \quad (4.1)$$

$$\delta_2 = \frac{\frac{1}{C}\sum c_i \eta_i^2 - \left(\frac{1}{C}\sum c_i \eta_i\right)^2}{\frac{1}{A}\sum \eta_i^2 - \left(\frac{1}{A}\sum \eta_i\right)^2} \quad (4.2)$$

where, c_i is the CN value at the i^{th} cell, C represents the cumulative value of CN over the catchment, η_i is the flow distance from the i^{th} cell to the catchment outlet, and N is the total number of catchment cells. The flow distance from each grid to the catchment outlet is determined using an R package called *'riverdist'*. δ_1 represents the distance of the centroid of the chosen spatial characteristic of the catchment concerning the average value of the flow distance (i.e., the catchment centroid). A value of $\delta_1 < 1$ indicates that the chosen characteristic is closer to the outlet (i.e., on the downstream side of the catchment), and a value greater than 1 indicates that the selected spatial characteristic is on the upstream side of the catchment. δ_2 represents the dispersion of the characteristic-weighted flow distances about their mean value with respect to the dispersion of the flow distances. A value of $\delta_2<1$ indicates that the chosen spatial characteristic is characterized by multimodal distribution, and a value greater than 1 indicates unimodal distribution. When a characteristic is uniformly distributed in the catchment, the value of δ_1 and δ_2 will be equal to one, indicating a spatially homogenous region. The following steps are followed to discretize the watershed into homogeneous sub-basins:

1) Calculation of sub-basins using a grid-based Digital Elevation Model (DEM). This step incorporates information about the topography of the catchment, which plays a crucial role in runoff generation.

2) Computation of the spatial moments (δ_1 and δ_2) of Curve Number (indicative property of land-use land cover and soil type) for the sub-basins.

3) Check whether the computed spatial moments are closer to 1 with a tolerance threshold of Δ (i.e., $1- \Delta \geq \delta_1, \delta_2 \leq 1+ \Delta$) in all sub-basins.

4) Delineate the sub-basins again into finer sub-basins where the spatial moments are beyond the tolerance threshold

5) Repeat steps 2 to step 4 until both the spatial moments at all the sub-basins are closer to 1 within the tolerance threshold. Terminate the delineation process if the catchment size is below 750 $km^{2,}$ which is the approximate spatial resolution of available gridded rainfall in this study.

The detailed flowchart describing the nested catchment discretization method using spatial variability indices is given in Figure 4.2.

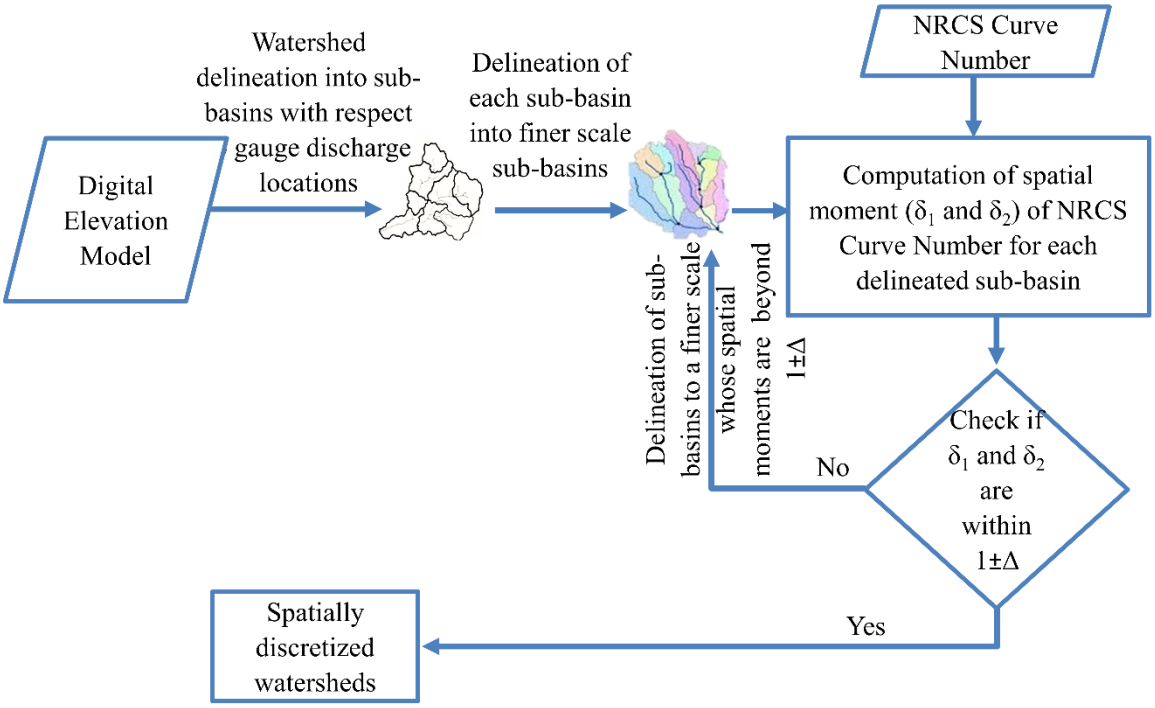

Figure 4.2 Flow chart of the proposed scheme of nested watershed discretization into homogenous sub-basins based on spatial moments of NRCS Curve Number.

The delineation of watersheds in this study is carried out by plugging R with GRASS GIS using the '*rgrass7*' package. The study watersheds were delineated for three tolerance thresholds, i.e., 0.1, 0.075 and 0.05, denoted as Discretization levels 1, 2 and 3, respectively. These tolerance thresholds represent ±10%, ±7.5% and ±5% deviation from the value 1. The plots of discretized watersheds for three discretization levels, along with the spatial distribution of curve number, are shown in Figure 4.3.

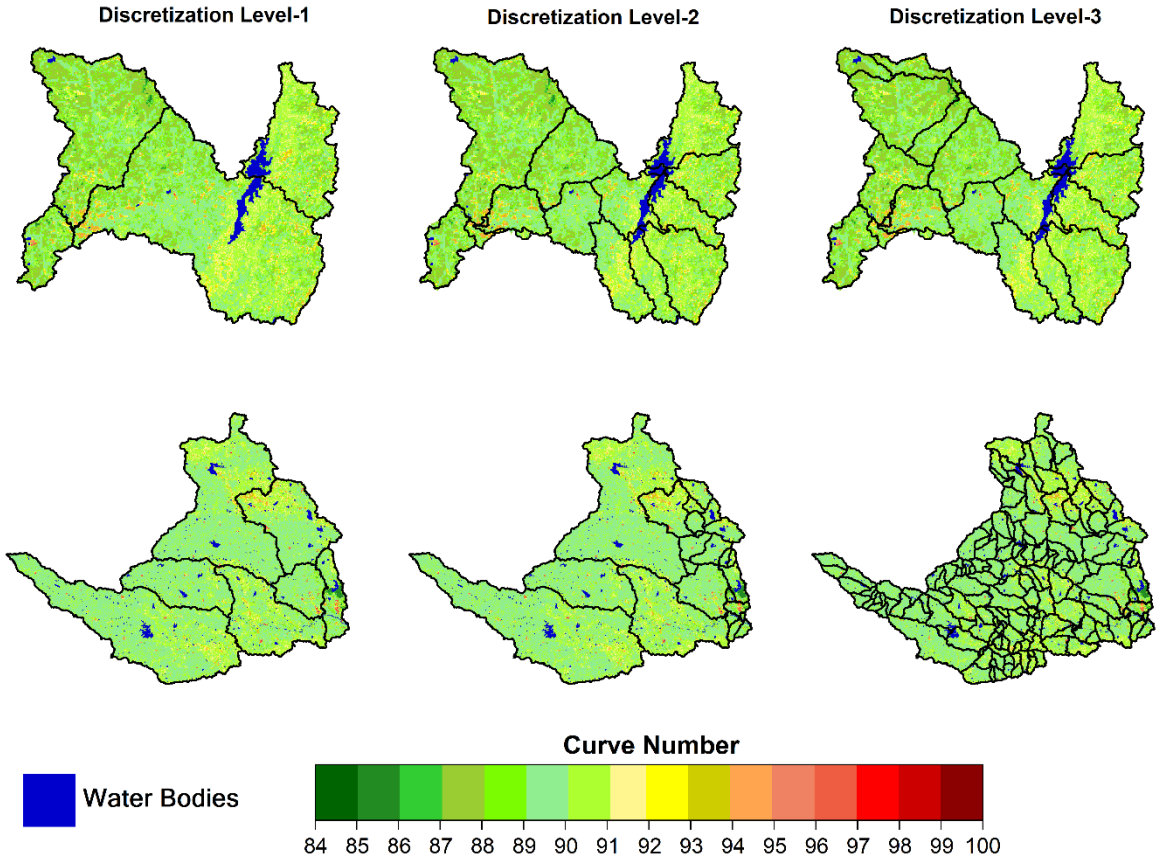

Figure 4.3 Jagdalpur (top) and Wardha (bottom) basins discretized for three different discretization levels. Discretization Level 1, 2 and 3 represents that the spatial moments of the Curve Number are deviatingdeviate within ±10%, ±7.5% and ±5%.

4.4 Model Setup

In this study, three versions (lumped, semi-distributed and semi-lumped) of modèle du Génie Rural à 4 paramètres Journalier (GR4J) were used to simulate the streamflow in the two catchments. A brief description of the employed GR4J model structures is presented in the subsequent sections.

4.4.1 GR4J lumped model

GR4J is a lumped conceptual rainfall-runoff model developed by Perrin et al. (2003), simulates the streamflow at a daily time step (Pushpalatha et al., 2012). The structure of the GR4J model is similar to the other soil moisture accounting based conceptual models. However, it was developed following an empirical approach to reduce the number of parameters that should be calibrated.

The Net Evapotranspiration (E_n) and Net Precipitation (P_n) are calculated using input evapotranspiration (ET) and precipitation (P) time series using Eq. 4.3

$$if\ P > ET;\ \ P_n = P - ET\ and\ ET_n = 0 \quad else \quad P_n = 0\ and\ ET_n = ET - P \qquad (4.3)$$

The net precipitation P_n is partitioned into two parts; P_s and $P_n - P_s$, where the former is allocated to production storage and the latter to surface runoff. The water in the production store is drained away as percolation (P_{perc}) and evapotranspiration. The summation of surface runoff and the percolated water is considered as total runoff (P_r), which is routed by partitioning into two parts at the routing store. The first partition comprising 90% of P_r is routed using Unit Hydrograph 1 (UH 1) and a non-linear routing store. The remaining 10 % of P_r is routed using Unit Hydrograph 2 (UH 2) alone. The groundwater exchange is accounted by applying a dynamic water loss/gain function to both the routed partitions. GR4J is a parsimonious model consisting of four parameters, one for the production function (X1 - Capacity of production store) and three for the routing function (X2 - Groundwater Exchange Coefficient, X3 - Capacity of routing store, X4 - Time base of unit hydrograph).

4.4.2 Semi-distributed GR4J

In the semi-distributed version of GR4J, the streamflow at each hydrological unit (sub-basins in this case) is generated using GR4J by forcing the mean areal precipitation, temperature and evapotranspiration of the corresponding sub-basin. Muskingum routing, a channel routing method, is employed to route the generated streamflow to the catchment outlet using a channel routing method where the observed streamflow data is available. The Muskingum routing method consists of two parameters K and x that address the travel time of the hydrograph through the routing reach and the dimensionless weighting factor, respectively. Sequentially, the Muskingum's K is a function of wave celerity (C) and the reach length (L). The reach length from the outlet of each sub-basin to the catchment outlet is computed using '*riverdist*' package in R by taking the stream network generated in GRASS GIS as input. To reduce the parameter dimensionality, the Celerity is assumed to be constant across all sub-basins. In the semi-distributed model structure, the model parameters describing the runoff generation through GR4J (4 parameters) and routing of generated runoff through the Muskingum method (2 parameters) are different for each sub-basin. The semi-distributed model setup for three discretization levels is denoted as SD-1, SD-2 and SD-3, respectively, in the preceding sections.

4.4.3 Semi-lumped GR4J

In a semi-lumped model, the parameters addressing the soil moisture routine are identical across all the sub-basins. However, it is important to note that the routing parameters vary for each sub-basin. The input meteorological forcing (precipitation, temperature and evapotranspiration) differs for each sub-basin, identical to the semi-distributed model structure. The schematic representation of the GR4J model structure, along with the semi-distributed and semi-lumped model versions of GR4J, is presented in Figure 4.4. The semi-lumped model setup for three discretization levels is denoted as SL-1, SL-2 and SL-3, respectively, in the preceding sections. The list of parameters to be calibrated in the lumped, semi-lumped and semi-distributed models, along with their ranges, are tabulated in Table 4.1.

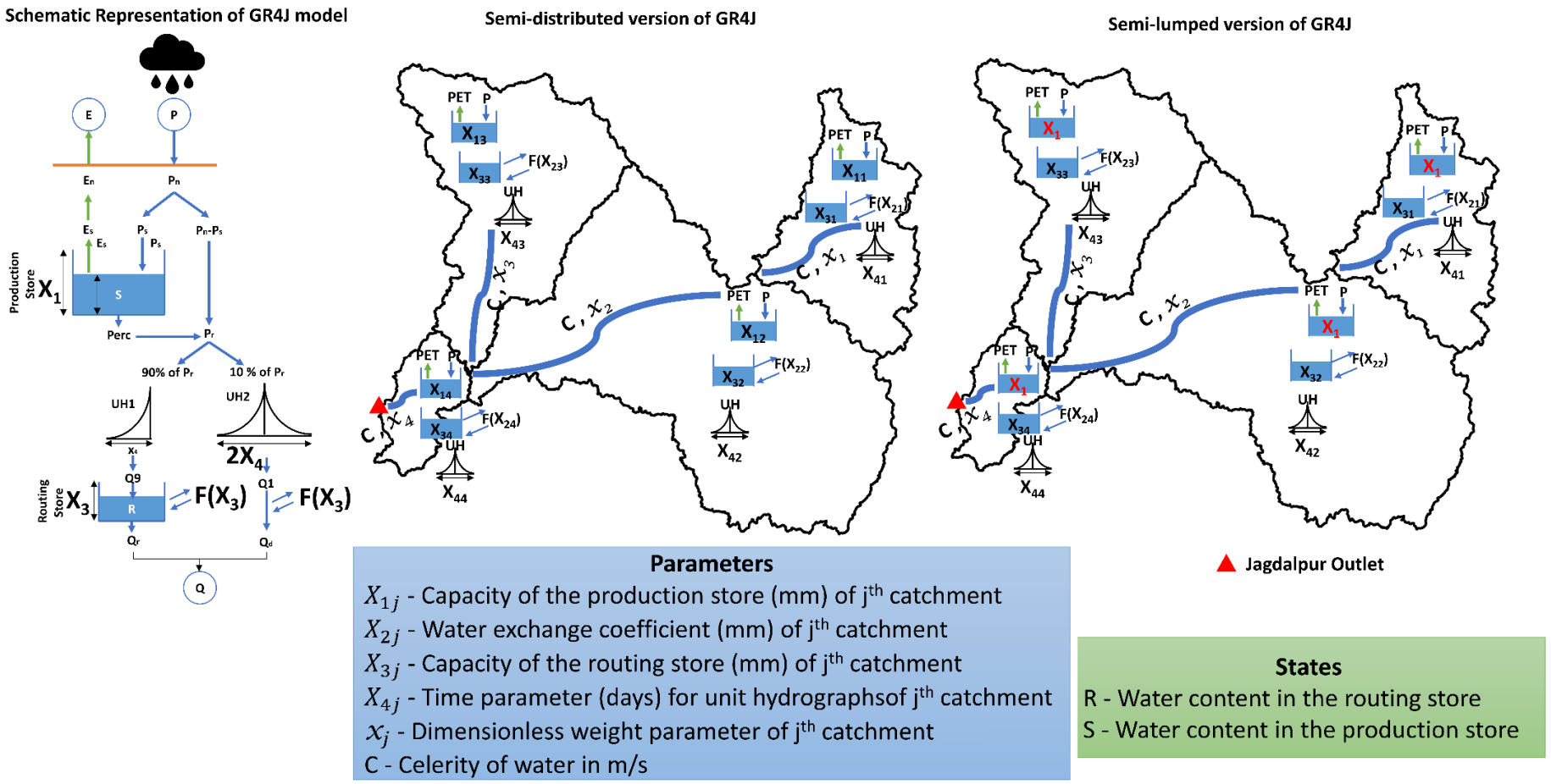

Figure 4.4 Schematic representation of GR4J lumped model structure (left) along with the semi-distributed (middle) and semi-lumped (right) versions of GR4J, the state variables and parameters of the model.

Table 4.1 List of parameters of the lumped, semi-distributed and semi-lumped versions of the GR4J model, along with their description and ranges.

		Parameter	Description	Range
Semi-distributed/Semi-lumped GR4J	GR4J	X1	production store capacity [mm]	1 to 1500
		X2	Inter-catchment exchange coefficient [mm/day]	-10 to 5
		X3	routing store capacity [mm]	1 to 500
		X4	unit hydrograph time constant [days]	0.5 to 4
	Muskingum routing	C	Average wave celerity in the stream network in m/s	0 to 5
		x	Dimensionless weighting factor	0 to 0.5

4.4.4 Model Calibration and Validation

In this study, the meteorological forcing data is partitioned into two parts, which are used for calibration (1970-1980 for Jagdalpur and 1970-1984 for Wardha) and validation (1981-1993 for Jagdalpur and 1985-2013 for Wardha). The period from 1967-1969 is used as a warmup period for both basins. Along with a continuous model, an event-based model was set up in this study to evaluate the performance of the model in capturing the observed flood events. Event-based hydrological models are often preferred over continuous hydrologic models for operational flood

forecasting as they exclusively simulate flood events. However, there is a need to estimate the Initial Hydrologic conditions or Antecedent Wetness Conditions before a flood event to accurately simulate flood events (Tramblay et al., 2010). The initial hydrologic conditions of a catchment in this study are estimated by extracting the states obtained from continuous rainfall-runoff modelling. A total of 20 flood events were obtained for both Jagdalpur and Wardha sub-basins. Ten events were used to calibrate model parameters, and the model's performance was validated using the remaining 10 events. The schematic representation of the event-based GR4J model setup is presented in Figure 4.5. The parameters of both continuous and event-based models are automatically calibrated using Genetic Algorithm with Nash-Sutcliffe Efficiency (NSE) (Nash and Sutcliffe, 1970) as an objective function. A population size of 100 has been used, and the maximum number of generations in the genetic algorithm was fixed as 2000, i.e., 2 lakhs simulations were run to calibrate the model. The lumped, semi-distributed and semi-lumped models of GR4J are set up in the R Environment using the 'airGR' package (Coron et al., 2017), and the automatic calibration is performed using the 'GA' package (Scrucca, 2013).

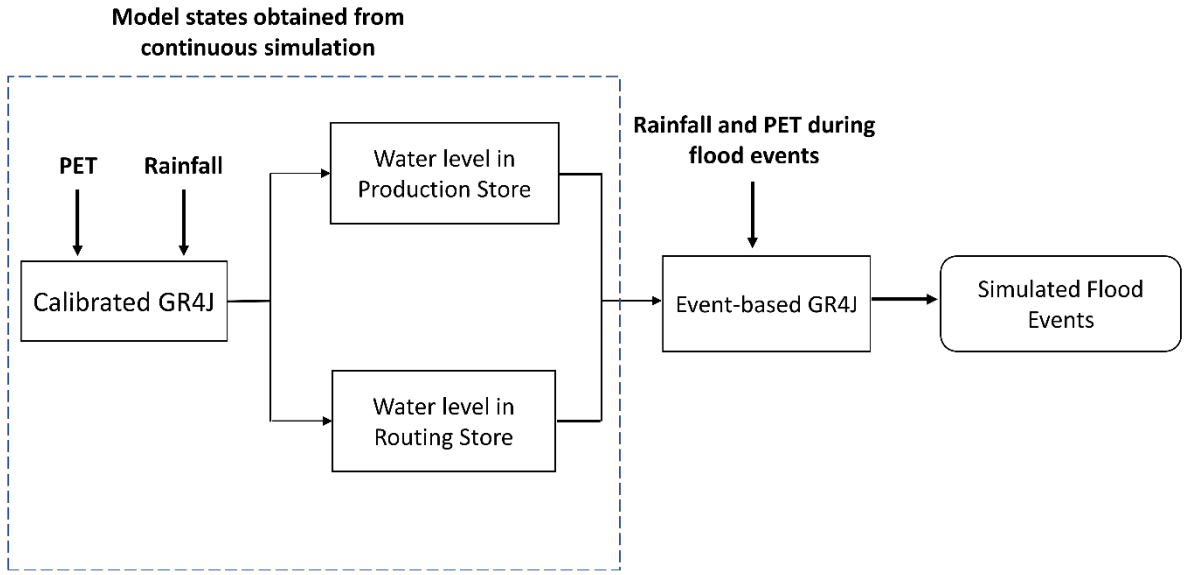

Figure 4.5 Schematic representation of Event-based GR4J model setup.

4.5 Performance Evaluation

The performance of continuous streamflow simulation generated from the calibrated GR4J model setup at varying spatial scales should be evaluated by a set of hydrologically relevant signature measures instead of classical single regression-based lumped measures (Manikanta and Vema, 2022). Hence, a diagnostic performance evaluation strategy suggested by Yilmaz et al. (2008) is

chosen to evaluate the model performance in simulating the streamflow. For this purpose, multiple model performance evaluation measures were employed to analyse the match between distinct features of the observed hydrograph, such as overall water balance, peak flows and low flows. Five performance evaluation criteria were chosen for this purpose, namely, Percentage Bias (PBIAS) (Ahmadi et al., 2014), Fourth root Mean Quadruple Error (R4MS4E) (Baratti et al., 2003), NSE of logarithmic transformed flows (logNSE) (Oudin et al., 2006), Skill Score (SS) (Raju et al., 2017) and Mean Absolute Error (MAE) (Jung et al., 2017). PBIAS is an indicative measure of over/underestimation, and a PBIAS value of zero suggests a good overall water balance. The performance of model simulations in matching the low and high flows is evaluated using logNSE and R4MS4E, respectively. The logarithmically transformed streamflow values used in logNSE emphasize low flows, whereas the errors raised to the power of four in R4MS4E allocate more weightage to the peak flows. Skill Score is a frequency domain-based metric that measures the match between observed and simulated histograms, i.e., Probability Density Functions (PDFs). This study also uses MAE to give an unbiased error estimate by allocating equal weightage to all the error magnitudes. The expressions for calculating R4MS4E, logNSE, PBIAS, SS and MAE are given in Eq. 4.4 to 4.8.

$$R4MS4E = \sqrt[4]{\frac{\sum_{t=1}^{T}(S_t - O_t)^4}{T}} \tag{4.4}$$

$$logNSE = 1 - \frac{\sum_{t=1}^{T}\left(\ln(S_t) - \ln(O_t)\right)^2}{\sum_{t=1}^{T}\left(\ln(O)_t - \overline{\ln(O)}\right)^2} \tag{4.5}$$

$$PBIAS = \frac{\sum_{t=1}^{T} S_t - O_t}{\sum_{t=1}^{T} O_t} \times 100 \tag{4.6}$$

$$SS = \frac{\sum_{i=1}^{nb} \min(f_{Si}, f_{Oi})}{T} \tag{4.7}$$

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|S_t - O_t| \tag{4.8}$$

where, S_t, O_t represents the simulated and observed streamflow values at time t, respectively, T represents the number of data points, f_{Si}, f_{Oi} represent the frequencies of simulated and observed streamflow values at the i^{th} bin, and nb denotes the total number of bins. In this study, the frequency of observed and simulated data for every 5 m3/sec was computed to assess the match between observed and simulated streamflow frequency characteristics. Along with these five performance criteria, a Euclidean Distance based composite metric (ED) is also used to

simultaneously analyse the model's performance in achieving the five criteria. The equation of ED is given in Eq. 4.9

$$ED = \sqrt{\sum_{k=1}^{K}(M_k - M_k^*)^2} \quad (4.9)$$

where, M_k is the kth model evaluation metric and M_k^* is the ideal value of the kth model evaluation metric. The ideal R4MSE, PBIAS and MAE values are 0, whereas the SS and logNSE values are 1. To account for varying ranges of the chosen criteria, the linear sum normalization technique is used to rescale them for the computation of ED (Vafaei et al., 2016).

Further, the performance of the employed event-based models in simulating the flood events is assessed using four performance evaluation metrics, namely, Nash Sutcliffe Efficiency (NSE), Percentage Error in Peak Flow (PEPF), PBIAS and Percentage Error in Timing to Peak (PETP). The equations to compute the criteria mentioned above are given from Eq. 4.10 to Eq. 4.12, and the performance rating of the employed statistics is given in Table 4.2 (Katwal et al., 2021).

$$NSE = 1 - \frac{\sum_{t=1}^{T}(S_t - O_t)^2}{\sum_{t=1}^{T}(O_t - \bar{O})^2} \quad (4.10)$$

$$PEPF = \frac{P_{sim} - P_{obs}}{P_{obs}} \times 100 \quad (4.11)$$

$$PETP = \frac{T_{sim} - T_{obs}}{T_{obs}} \times 100 \quad (4.12)$$

Table 4.2 Performance ratings of various statistics for evaluation of event-based model

Performance evaluation measure	Range	Performance Rating
NSE	0.75–1.00	Very good
	0.65–0.75	Good
	0.50–0.65	satisfactory
PBIAS	< ±10 %	Very good
	±10 % to ±15 %	good
	±15 % to ±25 %	satisfactory

PEPF	<±15 %	Very good
	±15 % to ±30 %	good
	±30 % to ±40 %	satisfactory
PETP	< ±10 %	Very good
	±10 % to ±15 %	good
	±15 % to ±30 %	satisfactory

4.6 Results

4.6.1 Performance Assessment of continuous streamflow simulation

As described in Section 4, the GR4J model simulates the streamflow at Jagdalpur and Wardha basins by setting up the model at four spatial resolutions (lumped and three discretization levels). Semi-distributed (SD-1, SD-2 and SD-3) and semi-lumped (SL-1, SL-2 and SL-3) models are considered parameter calibration strategies for three discretization levels, respectively. The parameters of the models were optimized using Genetic Algorithm for maximizing NSE. Along with NSE, the performance of simulated streamflow was assessed using five criteria, as mentioned in Section 5. The computed performance evaluation criteria and NSE at Jagdalpur and Wardha during the calibration period, are presented in Table 4.3. The performance of all models in simulating the streamflow during the calibration period is measured using NSE and is found to be greater than 0.76 for all models, suggesting very good simulations. When compared to the performance of lumped models in terms of NSE (0.779 at Jagdalpur and 0.763 at Wardha), both semi-distributed and semi-lumped showcased significant improvement in terms of NSE (>0.85 at Jagdalpur and >0.84 at Wardha). However, the performance of semi-distributed models in terms of NSE is found to be the highest in both Jagdalpur and Wardha basins. Interestingly, the performance of SD-3 is found to be the best at Jagdalpur, and SD-1 is doing well at Wardha. Wardha is approximately 6 times larger than Jagdalpur, and due to the finer discretization level in SD-3, it generally is anticipated to address the rainfall and landscape heterogeneity, thereby

leading to improved simulations. However, the parameter dimensionality has been significantly increased due to the finer discretization level, posing a challenge to the optimizer in the calibration process. This might be the plausible reason behind the slight underperformance of SD-3 when compared to SD-1 in Wardha.

The R4MS4E value of the lumped model in Jagdalpur and Wardha is 344.87 and 1853.43 m^3/sec, respectively, whereas the R4MS4E values of other models were less than the lumped model by approximately 25% at Jagdalpur and 16% at Wardha. These results show that the discretization-based models perform well in capturing the peak flows. The logNSE values of Jagdalpur and Wardha for discretized models range from 0.823 to 0.908 and 0.786 to 0.873. The PBIAS values between ±10 % denote that the performance of all models in terms of overall water balance is very good. The performance of discretized models regarding Skill Score during the calibration period ranges from 0.685 to 0.856 for Jagdalpur and 0.628 to 0.847, respectively, showing a good match between observed and simulated frequencies. The MAE of discretized models ranges from 41.08 to 43.4 m3/sec at Jagdalpur and 136.5 to 166.4 m3/sec at Wardha basins, respectively, comparatively lower than the MAE of lumped models. The performance evaluation metrics were also computed for the validation period and are tabulated in Table 4.4. The validation results in terms of NSE suggest that the streamflow simulation obtained from calibrated models in the validation period is very good. The performance of the models in terms of R4MS4E, logNSE, Skill Score and MAE were similar to the calibration period. An increase in the PBIAS values at Wardha is observed in all the models, suggesting that the models are overestimating of streamflow. However, the PBIAS values are satisfactory, i.e., < ±25 % at most of the models except for SD-3 showing that ineffective calibration of SD-3 due to higher dimensionality.

Table 4.3 Performance Evaluation metrics computed for observed streamflow and stream simulated using different GR4J model setups at Jagdalpur and Wardha during the calibration period. The optimum values are given in bold**.**

Discretization	R4MS4E	logNSE	PBIAS	Skill Score	MAE	NSE
Jagdalpur						
Lumped	344.877	0.763	-3.126	0.681	52.377	0.779
SD-1	238.586	0.872	-1.405	0.702	41.79	0.873
SD-2	**232.925**	0.823	-1.832	0.804	43.403	0.875
SD-3	235.398	0.868	**-0.657**	0.692	41.359	**0.876**
SL-1	258.821	**0.908**	-3.676	**0.856**	**41.086**	0.85
SL-2	237.596	0.894	-0.195	0.685	41.624	0.87
SL-3	251.159	0.866	-0.703	0.785	42.643	0.859
Wardha						
Lumped	1853.437	0.81	5.813	0.73	209.838	0.763
SD-1	**1360.035**	0.871	3.676	0.822	**136.511**	**0.881**
SD-2	1393.276	**0.873**	5.629	0.832	162.898	0.855
SD-3	1545.398	0.802	9.815	0.629	166.41	0.841
SL-1	1489.82	0.786	0.657	0.628	149.75	0.853
SL-2	1531.362	0.848	**0.166**	0.734	150.84	0.851
SL-3	1388.005	0.809	2.271	**0.847**	145.745	0.867

Table 4.4 Performance Evaluation metrics computed for observed streamflow and stream simulated using different GR4J model setups at Jagdalpur and Wardha during the validation period. The optimum values are given in bold.

Discretization	R4MS4E	logNSE	PBIAS	Skill Score	MAE	NSE
Jagdalpur						
Lumped	279.725	0.84	-3.391	0.783	45.315	0.791
SD-1	224.916	**0.894**	4.438	0.883	**41.435**	**0.846**
SD-2	239.445	0.819	-2.079	0.797	45.883	0.82
SD-3	243.89	0.885	5.041	**0.871**	42.772	0.828
SL-1	247.964	0.893	-1.075	0.865	41.939	0.827
SL-2	252.526	0.878	**0.457**	0.676	44.175	0.81
SL-3	**222.667**	0.88	1.837	0.852	42.143	0.84
Wardha						
Lumped	2269.95	0.706	25.812	0.794	228.741	0.683
SD-1	1509.805	0.752	24.287	0.864	177.174	**0.821**
SD-2	1730.79	0.768	24.502	0.819	200.797	0.774
SD-3	**1368.304**	0.673	32.411	0.655	206.86	0.804
SL-1	1819.869	0.734	21.165	0.713	180.822	0.773
SL-2	1877.877	**0.785**	**19.403**	0.786	180.201	0.771
SL-3	1426.177	0.753	22.439	**0.869**	**175.033**	0.817

The ability of the model to achieve the five performance criteria (R4MS4E, logNSE, PBIAS, SS and MAE) simultaneously is assessed by computing the Euclidean Distance (ED). The Euclidean distance calculated for Jagdalpur and Wardha basins during calibration and validation periods is shown in Figure 4.6. The figure shows that the overall performance of discretized models (except SD-3) in terms of ED is comparatively better than their lumped counterparts. The better ED value of Jagdalpur in the calibration period and the deteriorated ED value in the validation period are plausibly due to the overfitting of the model. Subsequently, the ED value of SD-3 at the Wardha basin is very high compared to other models during both calibration and validation periods denoting the ineffective calibration of parameters. The SD-3 model is prone to under/overfitting where the underfitting is due to the increased parameter dimensionality, and the overfitting is plausibly due to equifinality. The performance of semi-lumped models in terms of ED is

comparatively lower in the validation periods displaying their effective parameter transferability to the validation period. The percentage increase in the performance of discretized models in terms of ED is computed. The mean percentage decrement in the ED value of semi-distributed models compared to lumped models at Jagdalpur and Wardha is approximately 59% and 20%, respectively, during the calibration period and 3% and 25% during the validation period. For semi-lumped models, the mean percentage decrement of ED at both stations is approximately 50% during calibration and 40% during validation. This shows that the performance of semi-lumped models in simultaneously matching different segments of the observed hydrograph is comparatively better than the simulation of semi-distributed models.

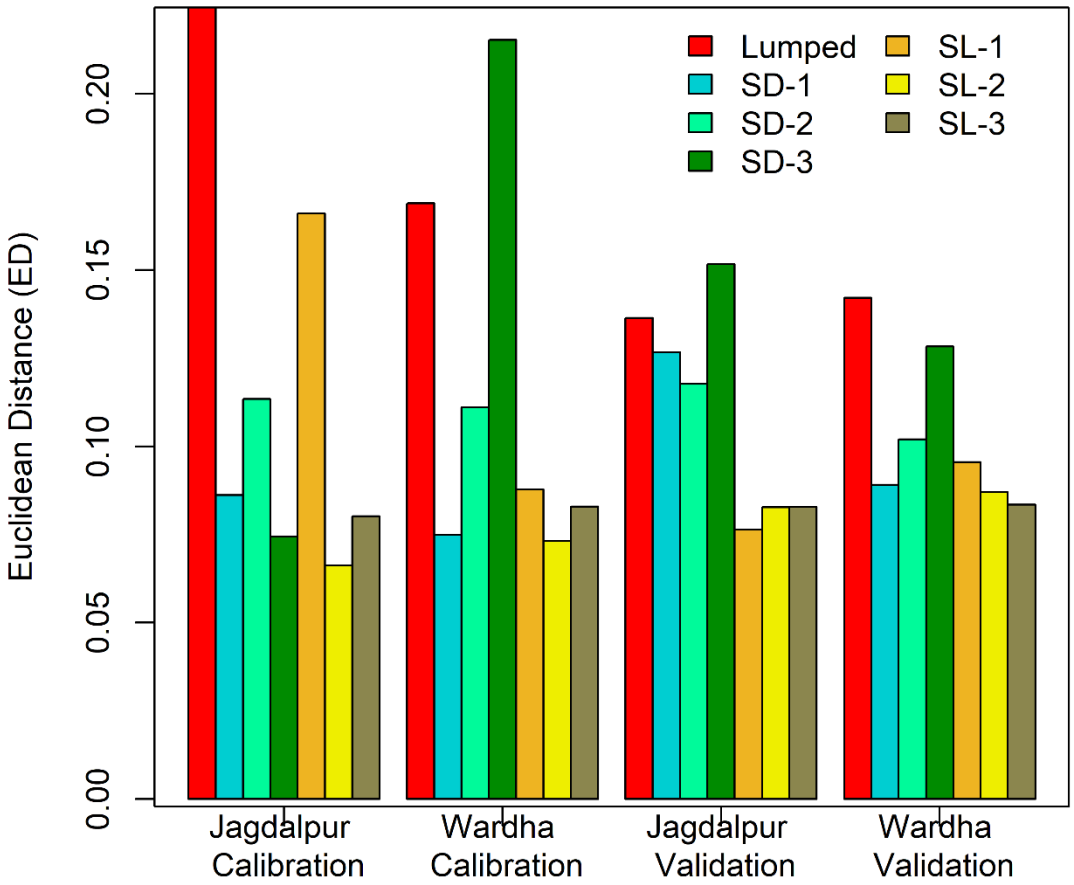

Figure 4.6 Euclidean distance computed for five performance criteria (R4MS4E, logNSE, PBIAS, SS and MAE) for all the models at Jagdalpur and Wardha basins during calibration and validation periods. An ED value closer to zero represents better model simulations.

The Percent Bias computed between the observed and simulated Flow Duration Curves (FDCs) in high, medium and low flow segments at Jagdalpur and Wardha during calibration and validation periods is presented in Figure 4.7. The bias in the high flows is below 15% at both stations during calibration and validation periods indicating good performance of models in predicting the peak flows. The performance model simulations in medium flows at Jagdalpur are found to be good (<15) whereas at Wardha, the bias in the simulations of semi-distributed models in the validation period exceeds 50%. The bias in the low flows at both stations during validation is higher when compared to other flow segments. Further, the bias of all the models in simulating the low flows at Wardha is >50% indicating an overestimation of low flows during both calibration and validation periods. The higher bias in the low flow is attributable to the model structure of GR4J, as some of the previous studies have reported its weakness in simulating the low flows (Demirel et al., 2015; Zeng et al., 2019).

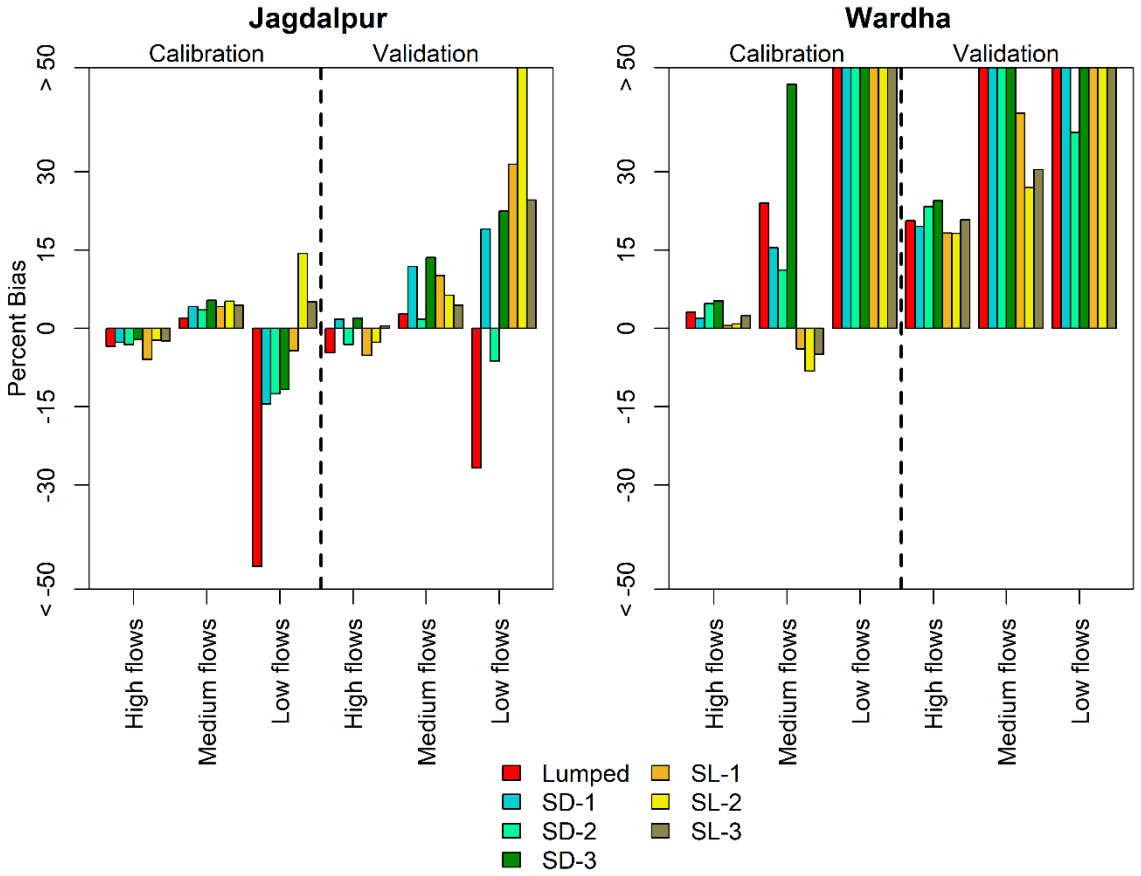

Figure 4.7 Percent Bias computed between different flow segments of observed and simulated Flow Duration Curves at Jagdalpur and Wardha basins during calibration and validation periods. A Percent Bias value closer to zero represents unbiased model simulations.

4.6.2 Performance assessment of event-based streamflow simulation

The performance of the lumped and discretization-based GR4J models to simulate historical flood events has been tested in this study. For assessing the performance of an event-based model, Berthet et al. (2009) suggested that an objective function, indicative error measures of the match between the magnitude and timing of peak error, and a visual comparison of the observed and simulated hydrographs are necessary. For this purpose, four performance evaluation measures, NSE (objective function), PEPF, PBIAS and PETP, were used the evaluate the model simulated flood response (Table 4.5). The median NSE value of lumped models during the calibration period is above 0.68 at both Jagdalpur and Wardha, indicating good performance. The performance of lumped models in the validation period regarding median NSE values at Jagdalpur (0.56) and Wardha (0.269) indicates their inability to account for spatial variability.

Further, the median NSE value of the flood events during the calibration period is greater than 0.84 (calibration) and 0.71 (validation) for all the SD and SL models at Jagdalpur, and the median NSE is greater than 0.79 (calibration) and 0.67 (validation) at Wardha indicating the very good performance of the models in capturing the flood peaks. The performance of semi-lumped models increases with an increase in the level of discretization at both stations during calibration and validation periods, with SL-3 performing the best. It is also important to note that the difference between median NSE values in calibration and validation periods of semi-lumped models is comparatively lower than the semi-distributed models indicating an efficient parameter transferability. The PEPF, PBIAS, and PETP values of the SD and SL models vary between -0.8 to 12.5%, -3.2 to 7.5% and -3.9 to 1.8%, respectively, during the calibration period at both stations. These values indicate very good performance of the discretization-based models in capturing the magnitude, timing and volume of flood during the calibration period. A decline in these performance statistics has been observed at Wardha, but most of the models are well within acceptable limits.

Table 4.5 Median values of performance evaluation metrics computed for the flood events simulated using different event-based GR4J model setups at Jagdalpur and Wardha during Calibration and Validation periods. The optimum values are given in bold.

Discretization	NSE	PEPF	PBIAS	PETP	NSE	PEPF	PBIAS	PETP
Jagdalpur								
	Calibration				**Validation**			
Lumped	0.7	-26.7	5.5	-3.9	0.568	-26.6	-3	6.3
SD-1	**0.91**	-3.3	0.4	-2.5	0.772	-13.2	-8.7	3.7
SD-2	0.85	-12.5	3.9	**-1.2**	0.743	-18.8	-4.8	2
SD-3	0.887	**-0.8**	1.9	1.8	0.731	-12.3	-2.9	5.1
SL-1	0.845	-7.8	-3.2	-2.9	0.714	-19.7	-4.9	1.3
SL-2	0.893	-9.2	-2.1	-3.9	0.748	-17.1	-6.5	7.5
SL-3	0.898	-7.9	**-0.1**	-2.5	**0.818**	**-7.7**	**-2.2**	**0.4**
Wardha								
	Calibration				**Validation**			
Lumped	0.687	**2.7**	**2.6**	-0.7	0.269	58.4	23.9	-3.6
SD-1	**0.843**	-8.2	3.8	-6.7	0.782	25.3	23.9	1.4
SD-2	0.821	-11.6	7.5	-0.9	0.701	31.5	27.8	**0**
SD-3	0.794	-12.6	6.8	-3.7	0.714	26.6	29.5	-0.7
SL-1	0.793	-9.9	4.6	**-0.6**	0.72	34.8	25.7	-0.7
SL-2	0.796	-12.3	7.2	-0.9	0.676	28.2	30.1	-1.4
SL-3	0.816	-10.3	6.1	-2.2	**0.821**	**20.1**	**22.7**	-2.1

A visual inspection of the observed and simulated flood hydrographs is necessary to gain more insights into the ability of the models to simulate the streamflow. The observed and simulated flood hydrographs at Wardha during calibration and validation are presented in Figures 4.8 and Figure 4.9, respectively; the flood hydrographs for Jagdalpur for calibration and validation periods are plotted in Figures 4.10 and 4.11, respectively. The spatial moment δ_1 of precipitation for each flood event is computed and presented along with the flood hydrographs to understand the role of rainfall spatial variability on the performance of the models. Figure 4.8 shows that the SD and SL models are more efficient in capturing the magnitude and timing of flood peaks compared to their lumped counterparts. The deterioration in the ability of lumped models to capture the timing of

flood peaks can be mainly attributed to rainfall spatial variability. The spatial variability of rainfall was measured using δ_1 is less than 0.8 during Events 5 and 8, where the lumped model failed to capture the timing of the peak. To further understand this, the spatial distribution of rainfall is plotted in Figure 4.12, along with the computed spatial moment δ_1 of rainfall for each day during the flood event. It can be observed that in Event Number 5, the δ_1 values before the flood peaks are below 0.8 with a magnitude greater than 60 mm, indicating that the rainfall is highly concentrated towards the downstream side (closer to the catchment outlet). This highly concentrated storm at the downstream side of the catchment leads to reduced lag time while routing the flood wave. The better performance of SD and SL models can be attributed to their ability to account for rainfall spatial variability.

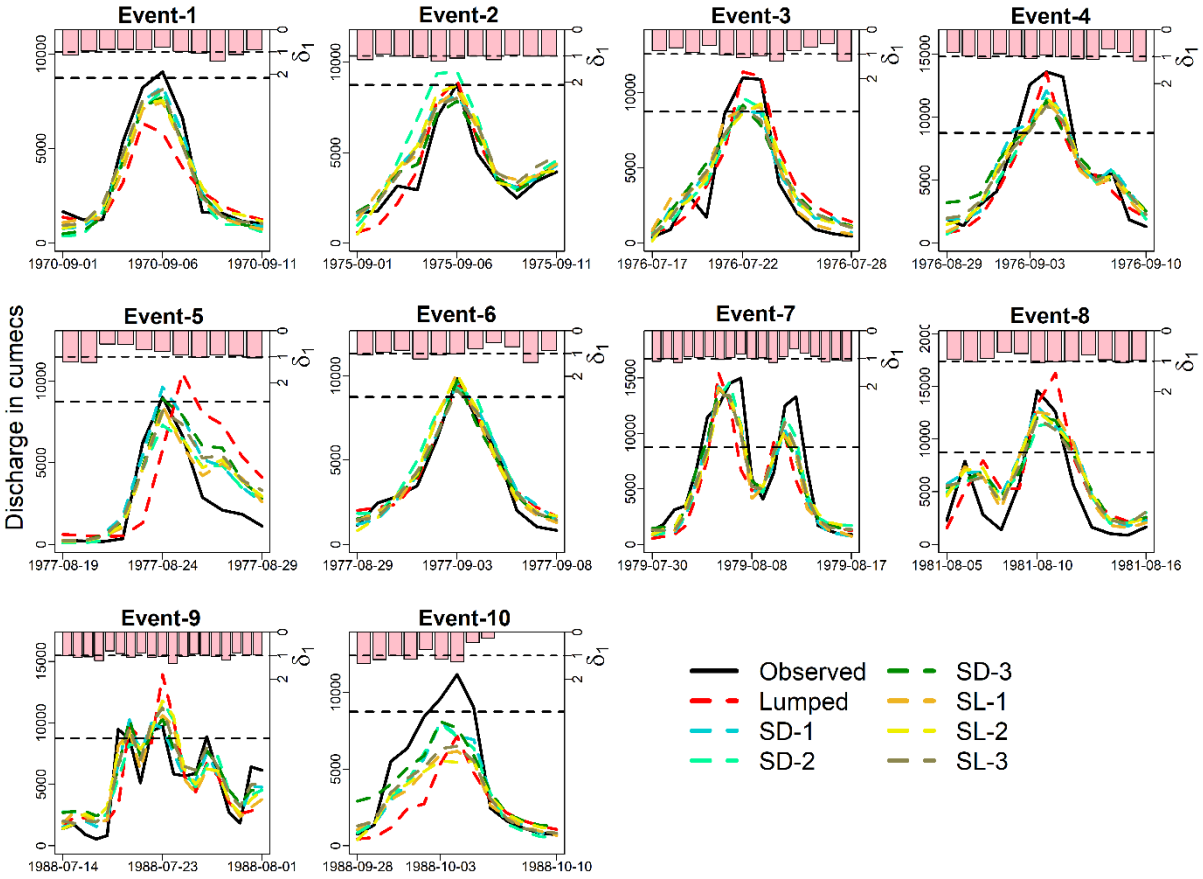

Figure 4.8 Observed and simulated flood hydrographs for selected flood events at Wardha during the calibration period, along with the computed spatial moment δ_1 of rainfall for each day during the flood events.

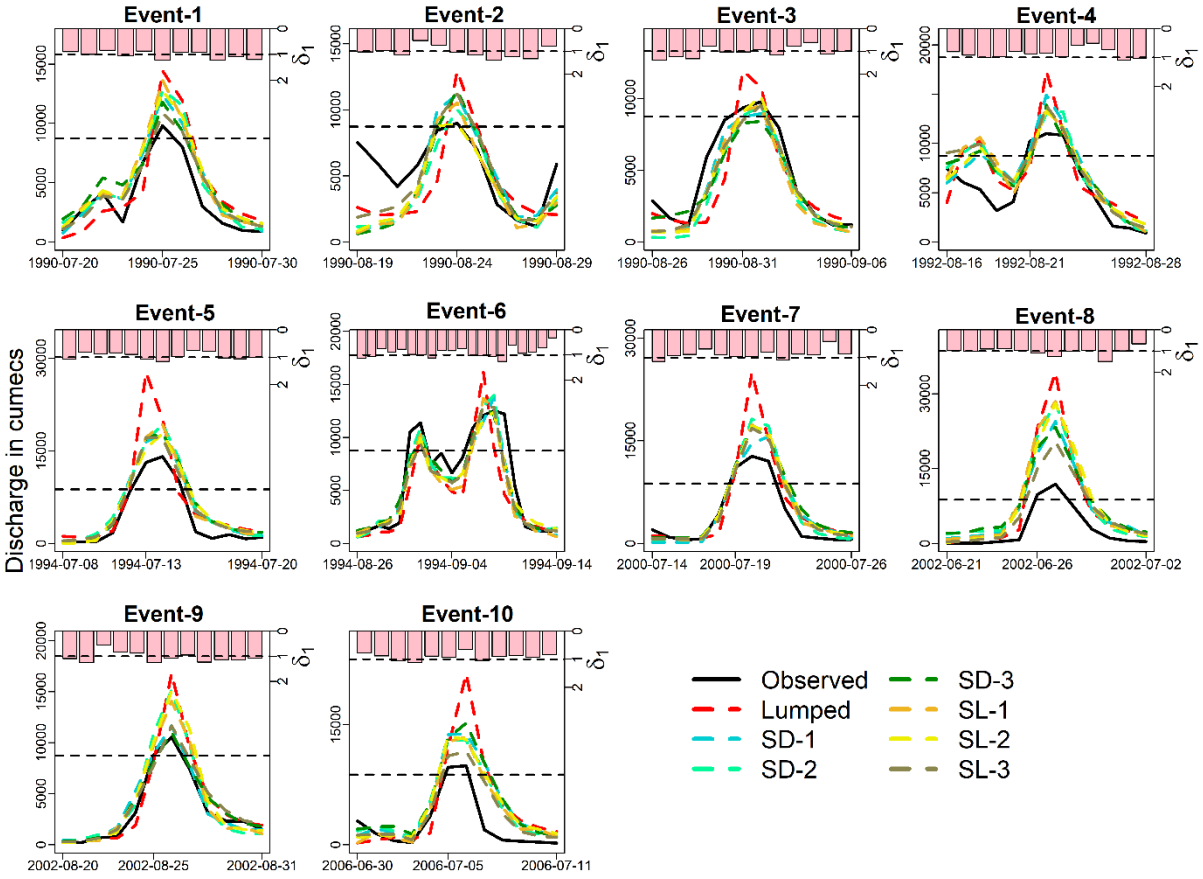

Figure 4.9 Observed and simulated flood hydrographs for selected flood events at Wardha during validation period along with the computed spatial moment δ_1 of rainfall for each day during the flood events. (SD represents semi-distributed and SL represented semi-lumped models; 1, 2 and 3 represents discretization level 1, 2 and 3)

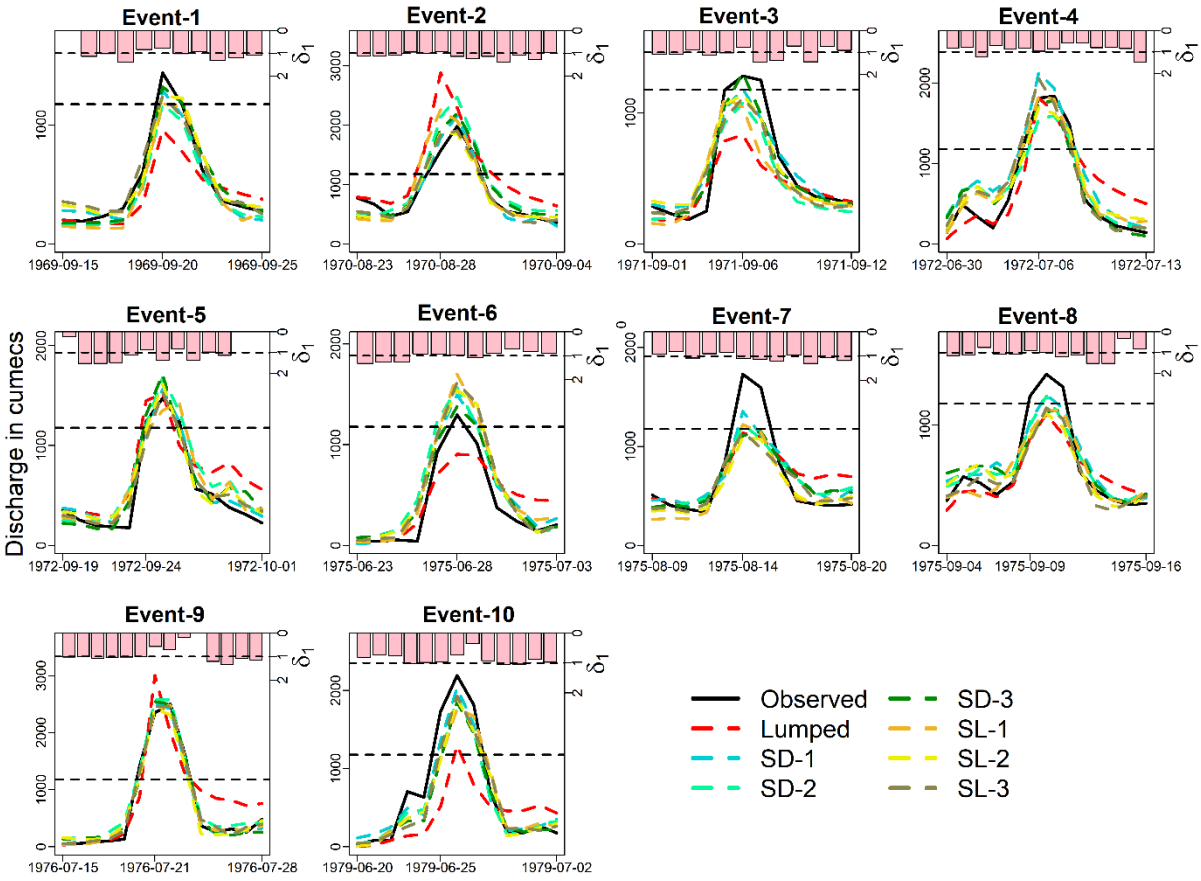

Figure 4.10 Observed and simulated flood hydrographs for selected flood events at Jagdalpur during calibration period along with the computed spatial moment δ_1 of rainfall for each day during the flood events. (SD represents semi-distributed and SL represented semi-lumped models; 1, 2 and 3 represents discretization level 1, 2 and 3)

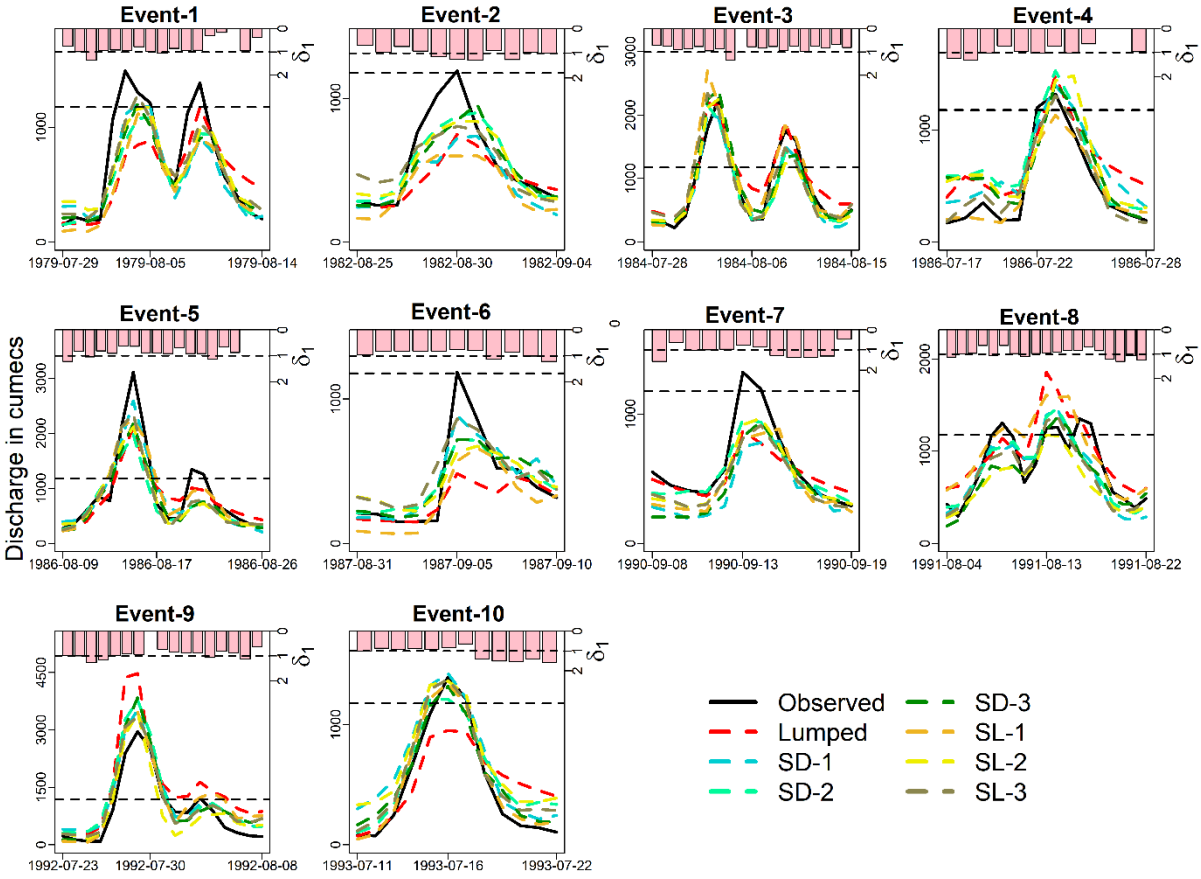

Figure 4.11 Observed and simulated flood hydrographs for selected flood events at Jagdalpur during validation period along with the computed spatial moment δ_1 of rainfall for each day during the flood events. (SD represents semi-distributed and SL represented semi-lumped models; 1, 2 and 3 represents discretization level 1, 2 and 3)

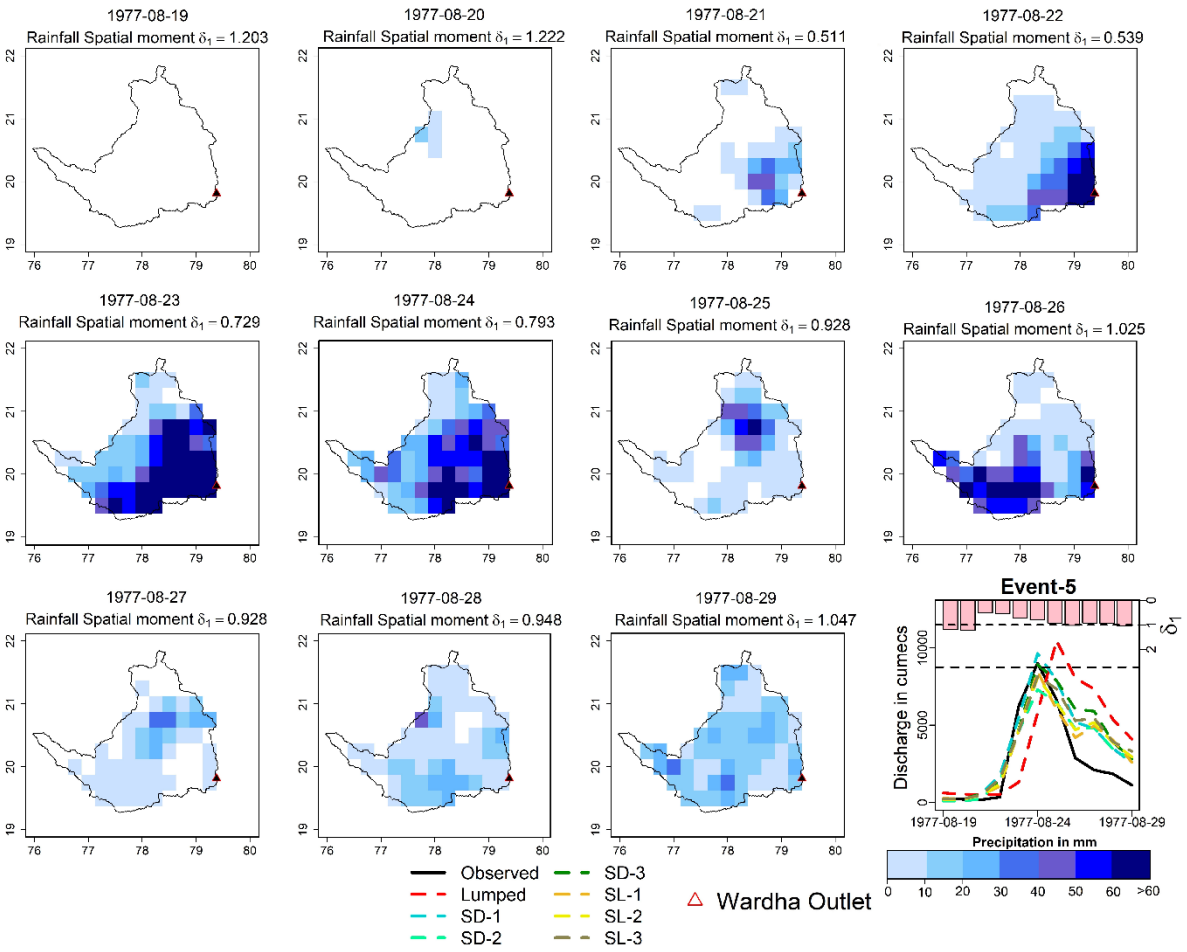

Figure 4.12 Spatial plots of rainfall for each day during Event-5 for Wardha during calibration. The observed and simulated hydrographs were plotted at the bottom right. (δ_1 represents the spatial moment computed for rainfall on each day.

4.7 Discussion

This study tested the influence of a nested landscape discretization scheme on the performance of a conceptual hydrological model GR4J in simulating the streamflow at two catchments. Lumped, semi-lumped and semi-distributed GR4J models were set up for various discretization levels. The performance of the simulated streamflow in matching the observed peak flows, low flows and water balance components is assessed by employing different performance evaluation criteria facilitating diagnostic evaluation. The results of the analysis show that the semi-distributed and semi-lumped models outperformed the lumped models in continuous and event-based simulations. The increment in the performance of the model set up at a finer discretization

level (level 3 in this study) is not significant as the model set up at discretization level 1. This analysis suggests that higher model resolution forced with finely resolved input data does not intrinsically enhance the model performance unless the input information is enhanced through finer spatial resolution datasets. The better performance of semi-distributed and semi-lumped models can be attributed to the landscape discretization based on the information about catchment physical characteristics (topography and CN) might help in the determination of model parameters. Another plausibility is that the parameters might have compensated the bias in the input meteorological forcing by calibration procedure, leading to better streamflow simulations (Brath et al., 2004). However, at model resolutions finer than the spatial resolution of the input meteorological forcing datasets, the error in representing spatial variability of meteorological data dominates the bias compensation in the input variables by the calibration procedure (Das et al., 2008). Moreover, it is also highly challenging to calibrate the model at finer resolutions due to increased degrees of freedom in the parameter sampling space and their underlying interactions.

Theoretically, SD-3 is expected to outperform the other models, whereas the increase in the model complexity (170 sub-basins) resulted in increase in the number of parameters to be calibrated. The optimising algorithm requires a lot of computational effort to reach a global optimal value when the dimension of parameter search space is high. Hence, it often results in local optimal values. Moreover, the uncertainty in the model prediction is increased in finer scale semi-distributed model (SD-3) due to the interaction between the resolution of input data and the conceptually lumped model structure of GR4J. Furthermore, the discrepancies between the performance of SD-3 between calibration and validation periods are also plausibly due to overfitting and uncertainties in the channel characteristics (Khakbaz et al., 2012). Reducing the parameter sample space by constraining the soil moisture accounting parameters in the semi-lumped model structure has addressed these issues to some extent (Jehn et al., 2019). Hence, the

performance drop in the semi-lumped models during the validation period is comparatively smaller than in the semi-distributed models.

The major advantage of using a coarser resolution model (SD-1 and SL-1) is that the calibration budget at Wardha, which is a large basin, is approximately 153% low when compared to models setup for discretization level-2 (SD-2 and SL-2) and 1636% lower than models setup for discretization level-3 (SD-3 and SL-3). However, a lumped model still gives acceptable results with a computational effort of approximately 40 % less than the SD-1 and SL-1 models. The computational advantage of the lumped model should be used efficiently by the seamless integration of rainfall spatial variability in its model structure (Zhou et al., 2021). Introducing the rainfall spatial variability as a parameter into the conceptual model to improve the model simulations should be investigated in future studies. Further, there is a need to test the proposed discretization scheme using multiple model structures and multiple calibration objective functions on large and diversified catchment sets to obtain more detailed insights.

4.8 Conclusions

This study explored the trade-offs between a nested discretization scheme and the streamflow simulated by a conceptual hydrological model. The catchments are discretized iteratively by assessing the spatial heterogeneity of the NRCS Curve Number using spatial moments computed from the catchment outlet. Lumped, semi-lumped and semi-distributed GR4J models were used to simulate the observed streamflow at Jagdalpur and Wardha basins in India. The results of the study suggest that the performance of semi-distributed and semi-lumped models was improved compared to lumped models in capturing the low flows, high flows and water balance components. The ability of models to capture the observed flood events through an event-based model setup is assessed, and similar results were obtained as the continuous simulation. The performance of semi-lumped models was found to be increasing with respect to the discretization level attributable to

its calibration strategy by constraining the soil moisture accounting parameters. This study concludes that a finer and more complex model leads to enhanced simulations when calibrated long enough. However, increased model resolution might not get translated as improved simulations or only provide marginally better performance at the cost of exponentially increased computational budgets. However, it is the modeller's responsibility to optimally choose between the model complexity and model performance at cheaper computational efforts.

Chapter 5

Accuracy evaluation of IHC estimation methods

5.1 Introduction

Flood is one of the most destructive natural disasters worldwide that can cause catastrophic impacts on day-to-day operations (Alfieri et al., 2017; Manohar Reddy and Ray, 2023). Real-time flood forecasting systems with sufficient lead time are considered as the most efficient and effective way to mitigate flood risk (Yatheendradas et al., 2008). Flood forecasting based on integrating meteorological and hydrological models has shown marked better improvement in recent times due to the advancements in computational power, remote sensing based observations, and an improved understanding of hydrological processes (Das et al., 2022). In the context of a flood modelling framework, event-based models are often considered as the sound alternatives to the continuous models as they are easier to calibrate and requires only data at the event scale (Tramblay et al., 2012). However, the efficiency of the both continuous and event-based hydrologic predictions, irrespective of the model structure, is highly influenced by optimal simulation of model states/initial hydrological conditions (IHC) (often considered as soil moisture) within the modelled watershed (Crow and Ryu, 2009; Loizu et al., 2018). Consequently, multitudes of modelling efforts have been reported so far on the accurate estimation of the magnitude of floods, particularly on minimizing the uncertainties associated with the estimation of initial hydrologic conditions (Alvarez-Garreton et al., 2014, 2015, Berthet et al., 2009, Brocca et al., 2008, 2009, 2010, Tramblay et al., 2010).

In general, a separate method is needed to estimate the initial states of the event-based models are separately calculated by establishing relationships between IHC and external predictors. For instance, the Antecedent Precipitation Index (API), computed by cumulating rainfall values of preceding days, is often used as a predictor in models based on the Soil Conservation Service Curve Number method (SCS–CN). Various predictors have been reported so far in the literature to accurately estimate the IHC for event-based models, such as Piezometric levels (Coustau et al., 2012), Baseflow (Franchini et al., 1996; Longobardi and Villani, 2003), Antecedent Discharge

Index (Tramblay et al., 2012), outputs from continuous models (Hegdahl et al., 2020; Huang et al., 2016; Yao et al., 2019) and in situ or remote-sensing based observables (Bahramian et al., 2021; Meng et al., 2017). Real-time measurements or climatology-based values can be used to update if the simulated model states reliably represent the observable model states (Berthet et al., 2009). Moreover, it is also worth noting that the relationship between the model states and the external predictors is determined by the model structure and uncertainty in the input data.

Data Assimilation (DA) techniques are widely employed to reduce the uncertainties associated with inputs, model states and output variables (McLaughlin, 2002). DA updates the model states by optimally combining the information from observations and model simulations, leading to improved estimates of initial states of a hydrological model (Sun et al., 2016). Observed datasets used to update the states of a hydrological model include streamflow (Seo et al., 2003), soil moisture (Luca Brocca et al., 2010), snow-covered area and snow water equivalent (Andreadis and Lettenmaier, 2006; Clark et al., 2006), and satellite observations of soil moisture and discharge (Andreadis et al., 2007). Discharge data is the most frequently assimilated variable since it contains the collective information of all other hydrological states (Clark et al., 2008). However, due to the difficulty in obtaining observations in real-time, many studies have used satellite-based soil moisture for data assimilation (Abbaszadeh et al., 2020; Alvarez-Garreton et al., 2015; Baguis and Roulin, 2017). In most DA studies, soil moisture observations are rescaled into the model space before assimilation, which influences the assimilation efficiency (Crow and Reichle, 2008; Crow and Van Loon, 2006). For instance, a poor rescaling method coupled with incorrect assumptions of the observational and model structural errors deteriorates the performance of DA (Tugrul Yilmaz and Crow, 2013).

A recent study by Nayak et al. (2021) showed that the efficiency of DA in reliable simulation of streamflow depends on the model structure. Various model structures are available based on runoff generation mechanisms such as physics-based, conceptual and data-driven, and based on the spatial resolution of models such as lumped, semi-distributed and distributed. However, in real-time operational applications, a trade-off between input data, model structural complexities and computation costs is necessary to issue reliable and timely flood warnings (Butts et al., 2004). Although physics-based fully distributed models account for the spatial heterogeneity of meteorological forcing and physical features within the basin leading to better simulations, they are data intensive and demand more computational resources (Young, 2002). On the contrary, the utilization of conceptual hydrologic models (CHMs) in operational streamflow forecasting has been amplified over the past few years due to their computational efficiency and simplicity,

proving their effectiveness in streamflow forecasting (Hapuarachchi et al., 2011). Many of the DA studies were reported on the lumped versions of conceptual hydrological models, and only a few were reported on the semi-distributed continuous model (Alvarez-Garreton et al., 2015, 2014; Berthet et al., 2009; Loizu et al., 2018; Pathiraja et al., 2012). Previous studies reported that accounting for spatial variability in CHMs enhances the model simulations (Das et al., 2008; De Lavenne et al., 2016). Hence, it is essential to understand the relevance of DA-estimated initial states in improving the performance of an event-based conceptual hydrological model.

In the scenarios where the initial states for event-based models are estimated through their continuous counterparts, the choice of calibration metrics plays a key role. This is due to the fact that the commonly used calibration objective functions have the tendency to emphasize particular flow segments of a hydrograph, and this bias will get translated into the model states simulated by the continuous models (Mizukami et al., 2019). Hence, this study is intended to assess the performance of a lumped and semi-distributed event-based conceptual model with initial conditions estimated using DA and their continuous counterparts corresponding continuous models calibrated using different calibration objective functions.

5.2 Study Area and Hydrometeorological Data

5.2.1 Study Area

The Jagdalpur catchment, a subbasin of the Godavari River Basin in India, is chosen as the study watershed in this study. The catchment of Jagdalpur was delineated from the flood forecasting stations at Jagdalpur, resulting in a catchment area of 7382 km^2. Figure 5.1 illustrates the location map of the study watershed, including its elevation and the stream network.

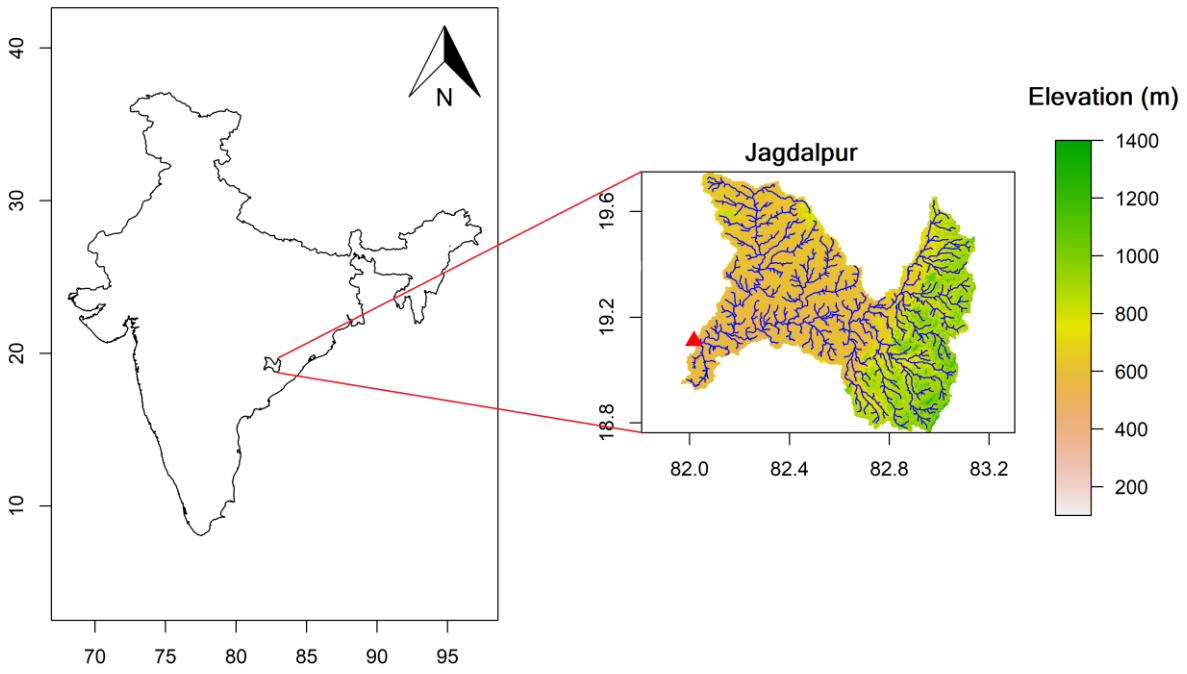

Figure 5. 1 Location map of Jagdalpur watershed with detailing of stream network and elevation.

5.2.2 Hydrometeorological Data

The daily streamflow data at Jagdalpur is available at the India-WRIS portal from 1965-2018. The observed flood events are separated from the streamflow data using the warning level obtained from the Central Water Commission appraisal report on Flood Forecasting and Warning Network Performance (CWC, 2018b). In this study, the daily gridded meteorological forcing datasets are obtained from India Meteorological Department (IMD) at a spatial resolution of 0.25^{o}x0.25^{o} (precipitation) and 1^{o}x1^{o} (minimum and maximum temperature) (Pai et al., 2014; Srivastava et al., 2009). For assimilation of soil moisture, root zone soil moisture is obtained from the Global Land Data Assimilation System (GLDAS), Catchment Land Surface Model, V2.0, where the data is available from 1948 to 2014.

5.3 Methodology

5.3.1 Hydrological model

In this study, two versions (lumped and semi-distributed) of modèle du Génie Rural à 4 paramètres Journalier (GR4J) were used to simulate the flood events at Jagdalpur. GR4J, originally developed by Perrin et al. (2003), is a soil moisture accounting-based daily lumped conceptual rainfall-runoff

model with four parameters to be calibrated. The semi-distributed version of GR4J utilizes the lumped model to simulate streamflow at each sub-basin, using mean areal precipitation, evapotranspiration, and temperature. The Muskingum routing method is then used to route the simulated streamflow from each sub-basin outlet to the catchment outlet. This requires two parameters, K and x, where K is determined by the reach length (L) and wave celerity (C), and x is a dimensionless weighting factor. The wave celerity is considered constant across all sub-basins to reduce parameter dimensionality. The study area is divided into four sub-basins, and a total of 21 parameters (four parameters for each sub-basin and one constant for wave celerity) must be optimized. The parameters to be optimized were tabulated in Table 5.1. Figure 5.2 provides a schematic representation of the lumped and semi-distributed versions of the GR4J model.

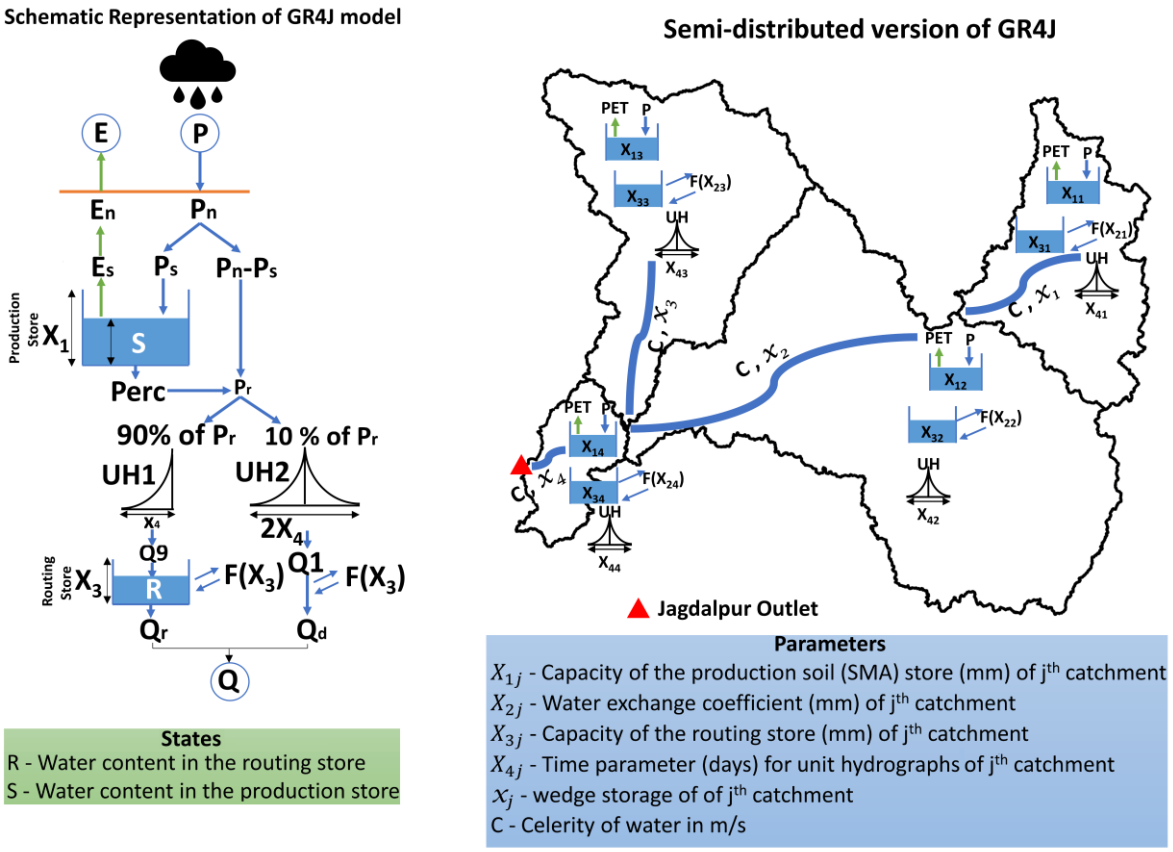

Figure 5. 2 Schematic representation of lumped GR4J model structure (left) along with the semi-distributed version of GR4J (right), the state variables and parameters of the model.

Table 5.1 Description and ranges of parameters to be calibrated in the lumped and semi-distributed versions of the GR4J model.

		Parameter	Description	Range
Semi-distributed GR4J	GR4J	X1	production reservoir capacity [mm]	1 to 1500
		X2	Inter-catchment exchange coefficient [mm/day]	-10 to 5
		X3	routing store capacity [mm]	1 to 500
		X4	unit hydrograph time constant [days]	0.5 to 4
	Muskingum routing	C	Average wave celerity in the stream network in m/s	0 to 5
		x	Dimensionless weighting factor	0 to 0.5

5.3.2 Implementation of Data Assimilation

5.3.2.1 Ensemble Kalman Filter (EnKF)

A background matrix (X^b) of dimensions $nstate \times nens$ is defined in Eq. (5.1), in which $nstate$ represents the number of state variables of the hydrological model and $nens$ denotes the number of ensemble members.

$$X^b = (x_1^b, \dots\dots, x_{ens}^b) \quad (5.1)$$

where, $x_1^b, \dots\dots, x_{ens}^b$ represents the model state vectors for each ensemble member in $nens$ before updating the states. The ensemble mean is the given by Eq. (5.2)

$$\bar{x}^b = \frac{1}{nens}\sum_{i=1}^{nens} x_i^b \quad (5.2)$$

The anomaly for the i^{th} ensemble member is computed as $x'^b_i = x_i^b - \bar{x}_i^b$, and the ensemble of anomalies (X'^b) is computed as given in Eq. (5.3)

$$X'^b = (x_1'^b, \dots\dots, x_{ens}'^b) \quad (5.3)$$

The ensemble anomalies are used to estimate the model error covariance using Eq. (5.4)

$$P^b = \frac{1}{nens-1} X'^b X^{b^T} \quad (5.4)$$

The equation to update the model states using the model error covariance is given in Eq. (5.5)

$$x_i^a = x_i^b + K(y_i - Hx_i^b) \quad (5.5)$$

where, $K = P^b H^T (HP^b H^T + R)^{-1}$

In Eq. (5.5), x_i^a represents the analysis of model states posterior to the update, H is an operator to covert model states to the observational space with dimensions $nobs \times nstate$, y_i represents the vector of observations with dimensions $nobs \times 1$ ($nobs$ denotes the number of observations), K represents the Kalman gain, and R is the observation error covariance matrix with dimensions $nobs \times nobs$. It should be noted that each ensemble member is updated individually using Eq. (5.5). In general, in the implementation of EnKF, each of the $nens$ ensemble members is updated using $nens$ vectors of observations. The $nobs$ values of y_i for each ensemble member are used to generate $nens$ vectors of observations by sampling from a distribution with observations as mean and R as a variance.

5.3.2.2 Bias correction of the GLDAS root zone soil moisture data

The model simulated soil moisture is commonly utilized to correct the bias in GLDAS root zone soil moisture before assimilation, given that both datasets do not accurately represent the true conditions. To address this, a straightforward mean-variance approach is employed, which demonstrates comparable performance to more intricate techniques. The simulated soil moisture is used to correct the raw values of GLDAS root zone soil moisture using Eq. (5.6).

$$SM_c = \overline{SM_s} + (SM_o - \overline{SM_o}) \frac{\sigma_s}{\sigma_o} \quad (5.6)$$

where SM_c denotes bias-corrected root zone soil moisture, $\overline{SM_s}$ and $\overline{SM_o}$ represents the average value of simulated soil moisture and GLDAS root zone soil moisture, respectively, and σ_s and σ_o are their respective standard deviations.

5.3.2.3. Forecast error

To address input uncertainty, meteorological datasets such as precipitation and evapotranspiration are stochastically perturbed. To achieve this, Clark et al. (2008) recommends

perturbing daily meteorological observations using multiplicative stochastic noise at each time step. A first-order autoregressive model is employed to ensure temporal correlation of time-variant forcings and physical consistency. To generate 100 ensemble meteorological forcing datasets, a temporal decorrelation length of one day and two days were used for rainfall and evapotranspiration, respectively, with a fractional error parameter set to 0.65. For a more detailed explanation of the perturbation method, please refer to Clark et al. (2008).

5.3.2.4. Observation error

In addition to ensuring an optimal ensemble spread to represent the model uncertainties, it also important to estimate the observation error for enhanced assimilation. In this study, a normally distributed random noise with zero mean and variance of ${\sigma_{obs}}^2$. The random noise is formulated as a function of observations, as given in Eq 5.7.

$${\sigma_{obs}}^2 = (\varepsilon_{obs}\, Obs)^2 \qquad (5.7)$$

Following the study by Piazzi et al. (2021), the error parameter ε_{obs} was set to 0.1. To prevent underestimation of error variances at lower values of observations, the minimum threshold for defining the error variance was assumed to be the 10th percentile value of observations (Obs_{10}). Following Thirel et al. (2010), the variance is evaluated proportionally to ${Obs_{10}}^2$ for observed values below Obs_{10}.

5.3.3 Model Calibration and Validation

In this study, the lumped and semi-distributed GR4J model is calibrated from 1970-1980 and validated from 1981-1993, with a warmup period from 1967-1969. The IHC/initial model states to initiate the event-based GR4J models are obtained from their corresponding calibrated continuous models. To calibrate the continuous models, four calibration objective functions were used, namely, Nash-Sutcliffe Efficiency (NSE) (Nash and Sutcliffe, 1970), NSE of logarithmic transformed flows (logNSE) (Oudin et al., 2006), Kling-Gupta Efficiency (KGE) (Gupta et al., 2009), and Fourth root Mean Quadruple Error (R4MS4E) (Baratti et al., 2003). The equations for the chosen calibration objective functions are given in Eq. (5.8) to Eq. (5.11)

$$NSE = 1 - \frac{\sum_{t=1}^{T}\left(Q_{s,t} - Q_{o,t}\right)^2}{\sum_{t=1}^{T}\left(Q_{o,t} - \overline{Q_{o,t}}\right)^2} \qquad (5.8)$$

$$logNSE = 1 - \frac{\sum_{t=1}^{T}\left(\ln(Q_{s,t}) - \ln(Q_{o,t})\right)^2}{\sum_{t=1}^{T}\left(\ln(Q_{o,t}) - \overline{\ln(Q_{o,t})}\right)^2} \quad (5.9)$$

$$KGE = 1 - \sqrt{(1-r)^2 + \left(1 - \frac{\sigma_s}{\sigma_o}\right)^2 + \left(1 - \frac{\mu_s}{\mu_o}\right)^2} \quad (5.10)$$

$$R4MS4E = \sqrt[4]{\frac{\sum_{t=1}^{T}\left(Q_{s,t} - Q_{o,t}\right)^4}{T}} \quad (5.11)$$

where, $Q_{s,t}$, $Q_{o,t}$ denotes simulated and observed discharge values at time t, respectively, and T denotes to total number of values, σ_s and σ_o σ_s and σ_o represents the standard deviations of simulated and observed values, respectively, and μ_s and μ_o represents the mean value of simulated and observed streamflow.

A total of 20 flood events were identified using the flood warning level, with 10 events used for calibration and the other 10 for validation. The parameters of the event-based models are automatically calibrated using Genetic Algorithm with NSE as an objective function. For both continuous and event-based models, the population size is set as 100 for a maximum of 2000 generations, i.e., 2 lakh simulations were performed in the calibration of models. A detailed flow chart demonstrating the methodology adopted in this study is given in Figure 5.3.

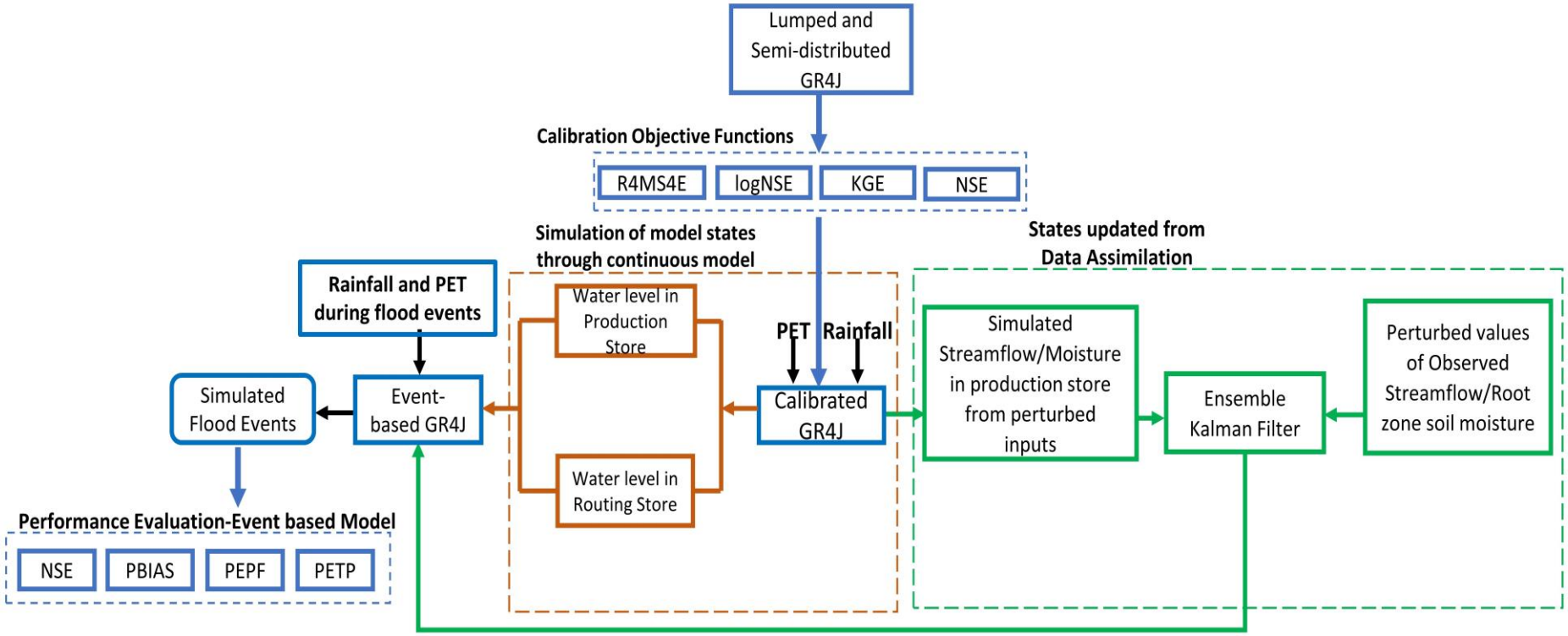

Figure 5.3 Flow chart demonstrating the detailed methodology adopted to evaluate the performance of event-based models under different schemes of estimating Initial Hydrologic Condition.

5.3.4 Performance Evaluation

A set of hydrologically relevant performance evaluation metrics are employed in lieu of a single traditional lumped measures to assess the accuracy of simulated hydrographs at all flow segments of the observed hydrograph (Manikanta and Vema, 2022; Yilmaz et al., 2008). Four evaluation metrics, namely Fourth root Mean Quadruple Error (R4MS4E), NSE of logarithmically transformed flows (logNSE), Percentage Bias (PBIAS), and Skill Score (SS), were employed to assess the agreement between observed and simulated hydrographs in terms of peak flow, low flow, water balance, and flow frequencies, respectively. The equations to compute the values of PBIAS and SS are given in Eq. (5.12) to Eq. (5.13).

$$PBIAS = \frac{\sum_{t=1}^{T} Q_{s,t} - Q_{o,t}}{\sum_{t=1}^{T} Q_{o,t}} \times 100 \quad (5.12)$$

$$SS = \frac{\sum_{i=1}^{nb} \min(f_{Q_{s,i}}, f_{Q_{o,i}})}{T} \quad (5.13)$$

where, $f_{Q_{s,i}}$, $f_{Q_{o,i}}$ are the frequency values of simulated and observed discharge values at the i^{th} bin, respectively, and nb represents the total number of bins. In this study, a bin size of 5 m^3/sec was used to compare the frequency distributions of observed and simulated streamflow.

The performance of the event-based models in simulating the observed flood hydrographs is evaluated using four metrics, namely, Percentage Error in Peak Flow (PEPF), Percentage Error in Timing to Peak (PETP), PBIAS and Nash Sutcliffe Efficiency (NSE) given from Eq. 5.14 to Eq. 5.15. In addition, the indicative performance ratings of the chosen metrics are shown in Table 5.2 (Katwal et al., 2021).

$$PEPF = \frac{P_{sim} - P_{obs}}{P_{obs}} \times 100 \quad (5.14)$$

$$PETP = \frac{T_{sim} - T_{obs}}{T_{obs}} \times 100 \quad (5.15)$$

where, P_{sim}, P_{obs} denotes simulated and observed flood peaks and T_{sim}, T_{obs} denotes the time taken to simulated and observed flood peaks.

Table 5.2 Indicative performance ratings of the statistics employed for evaluation of the event-based model

Performance evaluation measure	Range	Performance Rating
NSE	0.75–1.00	Very good
	0.65–0.75	good
	0.50–0.65	satisfactory
PBIAS	< ±10 %	Very good
	±10 % to ±15 %	good
	±15 % to ±25 %	satisfactory
PEPF	<±15 %	Very good
	±15 % to ±30 %	good
	±30 % to ±40 %	satisfactory
PETP	< ±10 %	Very good
	±10 % to ±15 %	good
	±15 % to ±30 %	satisfactory

5.4 Results

5.4.1 Performance evaluation of continuous streamflow simulation

The accuracy of the streamflow simulated by the continuous lumped and semi-distributed GR4J models is evaluated using a set of metrics as described in Section 3.4. The performance evaluation metrics for the calibration and validation periods are tabulated in Table 5.3 and Table 5.4, respectively. Table 5.3 shows that the performance of the R4MS4E calibrated lumped GR4J model is good at simulating the peak flow values. Higher values of logNSE and Skill Score of the lumped

GR4J model with streamflow assimilation indicate its better performance in terms of matching the observed flow frequencies and also in terms of simulating the low flows. The PBIAS of KGE calibrated lumped model is found to be low attributable to the bias factor ($\frac{\mu_s}{\mu_o}$) in the formulation of KGE. Further the PBIAS values of R4MS4E calibrated model shows its overestimation bias whereas other models are slightly underestimating the flows (< -10%). In the case of semi-distributed GR4J, a significant improvement is observed in both open-loop (without assimilation) and assimilated models in terms of the all the chosen evaluation criteria. The R4MS4E values in semi-distributed GR4J model were reduced on an average of 18.9% in all models when compared to their lumper counterparts. The PBIAS values of streamflow and soil moisture assimilated models along with NSE and KGE calibrated open-loop models were found to be less than ±5%, whereas logNSE an R4MS4E calibrated models were found to be underestimating and overestimating the flows, respectively. It can also be noticed the performance of KGE calibrated models in terms of logNSE and Skill Score is low (-0.07 and 0.45), showing its inability to capture the low flows.

In summary, the performance of the semi-distributed GR4J model with soil moisture assimilation is poor in the continuous streamflow simulation. The semi-distributed GR4J model with streamflow assimilation and model calibrated with NSE, logNSE and KGE performs well regarding chosen evaluation criteria in the calibration period. Similar performance can be observed in the validation period for both lumped and semi-distributed GR4J models (Table 5.4). This indicates better transferability of the calibrated parameters and better assimilation efficiency in the validation period.

Table 5.3 Performance Evaluation metrics computed for observed streamflow and ensemble mean of streamflow simulated using lumped and semi-distributed GR4J model setups with and without data assimilation during the calibration period.

	R4MS4E	**logNSE**	**Skill Score**	**PBIAS**
	Lumped GR4J			
Soil Moisture Assimilation	543.47	-0.75	0.65	-12.12
Streamflow Assimilation	407.5	**0.9**	**0.89**	-10.61
NSE calibrated	345.23	0.74	0.67	-3.08
KGE calibrated	360.22	0.66	0.63	**-0.7**
logNSE calibrated	402.33	0.87	0.79	-11.79
R4MS4E calibrated	**312.36**	0.7	0.78	40.22
	Semi-Distributed GR4J			
Soil Moisture Assimilation	426.07	0.06	0.64	5.15
Streamflow Assimilation	340.38	**0.92**	0.85	4.55
NSE calibrated	238.59	0.87	0.66	-1.41
KGE calibrated	291.23	-0.07	0.45	**-0.73**
logNSE calibrated	380.12	0.91	**0.86**	-22.58
R4MS4E calibrated	**251.94**	0.73	0.66	29.05

Table 5.4 Performance Evaluation metrics computed for observed streamflow and ensemble mean of streamflow simulated using lumped and semi-distributed GR4J model setups with and without data assimilation during the validation period.

	R4MS4E	**logNSE**	**Skill Score**	**PBIAS**
	Lumped GR4J			
Soil Moisture Assimilation	438.33	-0.86	0.69	-14.73
Streamflow Assimilation	337.84	**0.9**	**0.91**	-10.7
NSE calibrated	280	0.83	0.77	-3.34
KGE calibrated	325.5	0.78	0.73	**-0.94**
logNSE calibrated	363.18	0.85	0.86	-10.24
R4MS4E calibrated	**256.75**	0.6	0.66	45.7
	Semi-Distributed GR4J			
Soil Moisture Assimilation	341.77	0.36	0.69	17.03
Streamflow Assimilation	286.17	**0.92**	**0.87**	4.75
NSE calibrated	**224.92**	0.89	0.84	4.44
KGE calibrated	316.53	0.36	0.53	**2.67**
logNSE calibrated	368.96	0.89	0.75	-22.05
R4MS4E calibrated	253.09	0.59	0.53	35.62

5.4.2 Performance Evaluation of event-based models

In this study, the performance of the lumped and semi-distributed GR4J models in simulating the historical flood events using the IHC estimated from different methods (through data assimilation and continuous models) has been tested. Berthet et al. (2009) suggested that, for evaluating the predictive performance of event-based models, careful selection of evaluation criteria is essential which emphasize on match between the magnitude, volume and timing observed and simulated flood hydrographs. In addition to quantitative measures, a visual inspection on the match between observed and simulated flood hydrographs is necessary. Accordingly, four quantitative performance assessment metrics were selected namely, NSE (objective function), Percentage Error in Peak Flow (PEPF), PBIAS and Percentage Error in Timing to Peak (PETP) to evaluate the simulated flood response.

5.4.2.1 Performance of event-based models using IHC obtained through DA

The above statistics are computed for both lumped and semi-distributed event-based models, with IHC estimated by assimilating soil moisture and streamflow (Table 5.5). Table 5.5 calculates the statistics for each flood event using the observed flow and ensemble mean of simulations. The median value of the statistics computed for the events used for calibration and validation periods separately are given in Table 5.5. The median NSE value of the event-based lumped model based on soil moisture-based data assimilation (SM-DA) and streamflow-based data assimilation (Q-DA) is higher than 0.5 in the calibration period, indicating satisfactory performance. However, the performance of continuous simulations from the lumped model based on SM-DA and Q-DA in simulating the flood hydrographs during the calibration period is found to be low (median NSE < 0.4). In the validation period, the performance of all the lumped models based on DA is not satisfactory in terms of median NSE value (< 0.5). Additionally, the median NSE values in the validation period suggest that the temporal transferability of the calibrated parameters is poor, leading to poor simulations.

In summary, the performance of both continuous and event-based lumped models based on Q-DA is found to be satisfactory, with lower PEPF, PBIAS, and PETP indicating satisfactory performance compared to other models. The simulated flood hydrographs obtained from the event-based and continuous lumped models with SM-DA and Q-DA during the calibration period are plotted for calibration and validation periods in Figure 5.4 and Figure 5.5, respectively, along with the observed values. The figure shows that the ensemble spread of the simulations from the continuous models is higher than that of their event-based counterparts. This indicates that updating the model states continuously by assimilation preserves the spread induced through

perturbation. In contrast, despite the ensemble of IHCs used in event-based models, the spread in the event-based simulations is lower than that of their continuous counterparts. Furthermore, the ensemble spread of SM-DA-based simulations is higher than that of Q-DA-based simulations.

Table 5.5 Median values of performance evaluation metrics computed for the flood events simulated using lumped and semi-distributed event-based GR4J model setups with data assimilation at Jagdalpur during calibration and validation periods.

Assimilation Variable	**NSE**	**PEPF**	**PBIAS**	**PETP**	**NSE**	**PEPF**	**PBIAS**	**PETP**
Lumped								
	Calibration				**Validation**			
SM-EVT	0.56	-29.37	-12.5	-19.46	0.28	-48.58	0	-34.94
Q-EVT	**0.6**	**-22.64**	-6.25	**-16**	**0.39**	**-37.91**	**0**	**-29.96**
SM-CONT	-0.04	-59.45	-16.67	-46.8	-0.95	-77.94	-15.48	-67.82
Q-CONT	0.39	-47.16	**0**	-18.13	0.46	-44.63	0	-14.93
Discretized								
	Calibration				**Validation**			
SM-EVT	**0.82**	**-4.45**	**0**	**-6.25**	0.44	-39.93	0	-23.54
Q-EVT	0.64	-14.7	0	-14.13	**0.57**	-26.33	**0**	-27.16
SM-CONT	0.41	-46.92	0	-19.54	0.3	-50.83	0	-25.56
Q-CONT	0.42	-14.3	14.29	12.04	0.38	**-16.66**	15.48	**9.47**

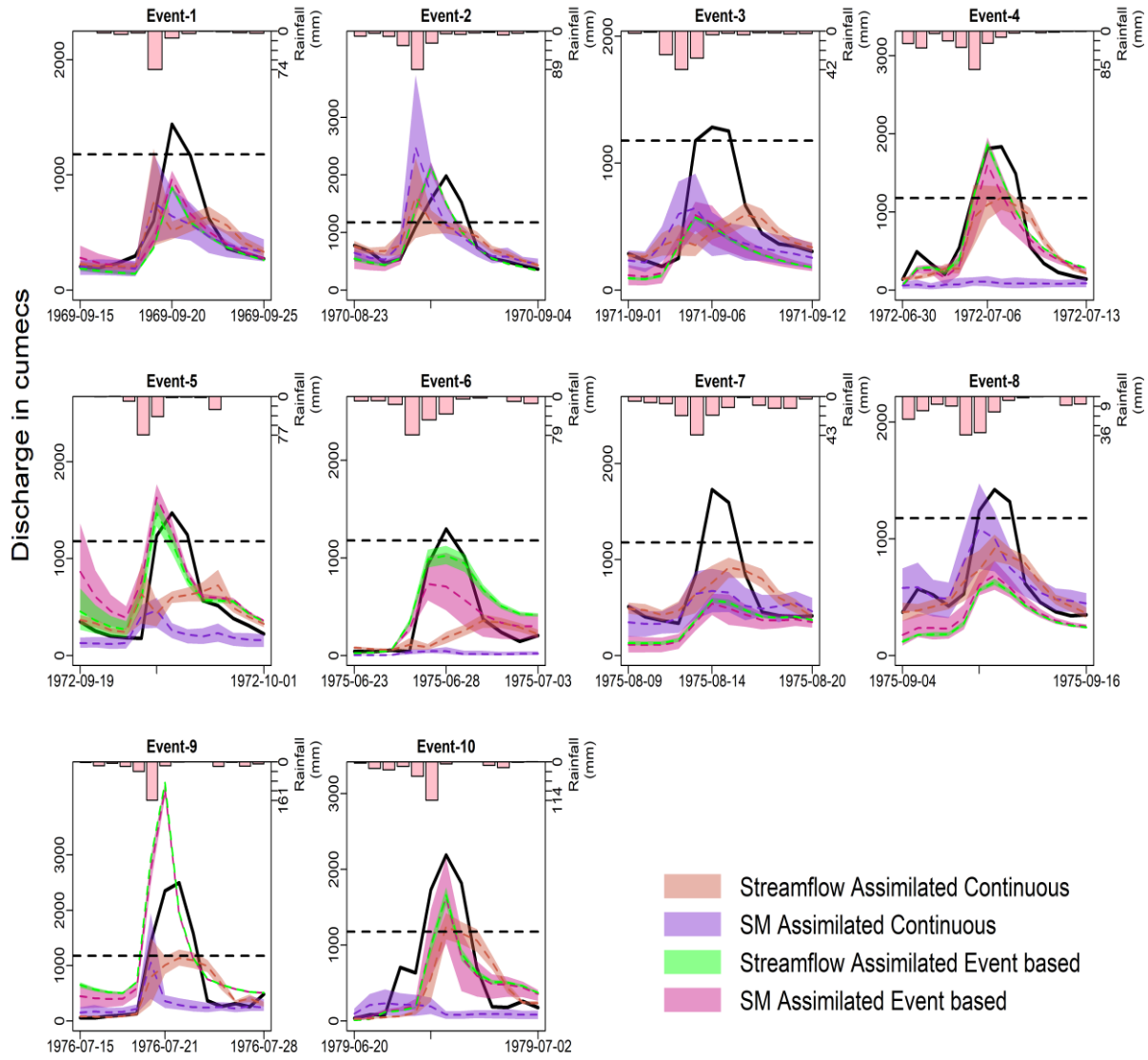

Figure 5. 4 Simulated flood hydrographs by the continuous and event-based lumped model based on Soil Moisture (SM) and streamflow data assimilation. Observed flood hydrographs were also plotted for the selected flood events at Jagdalpur during calibration period.

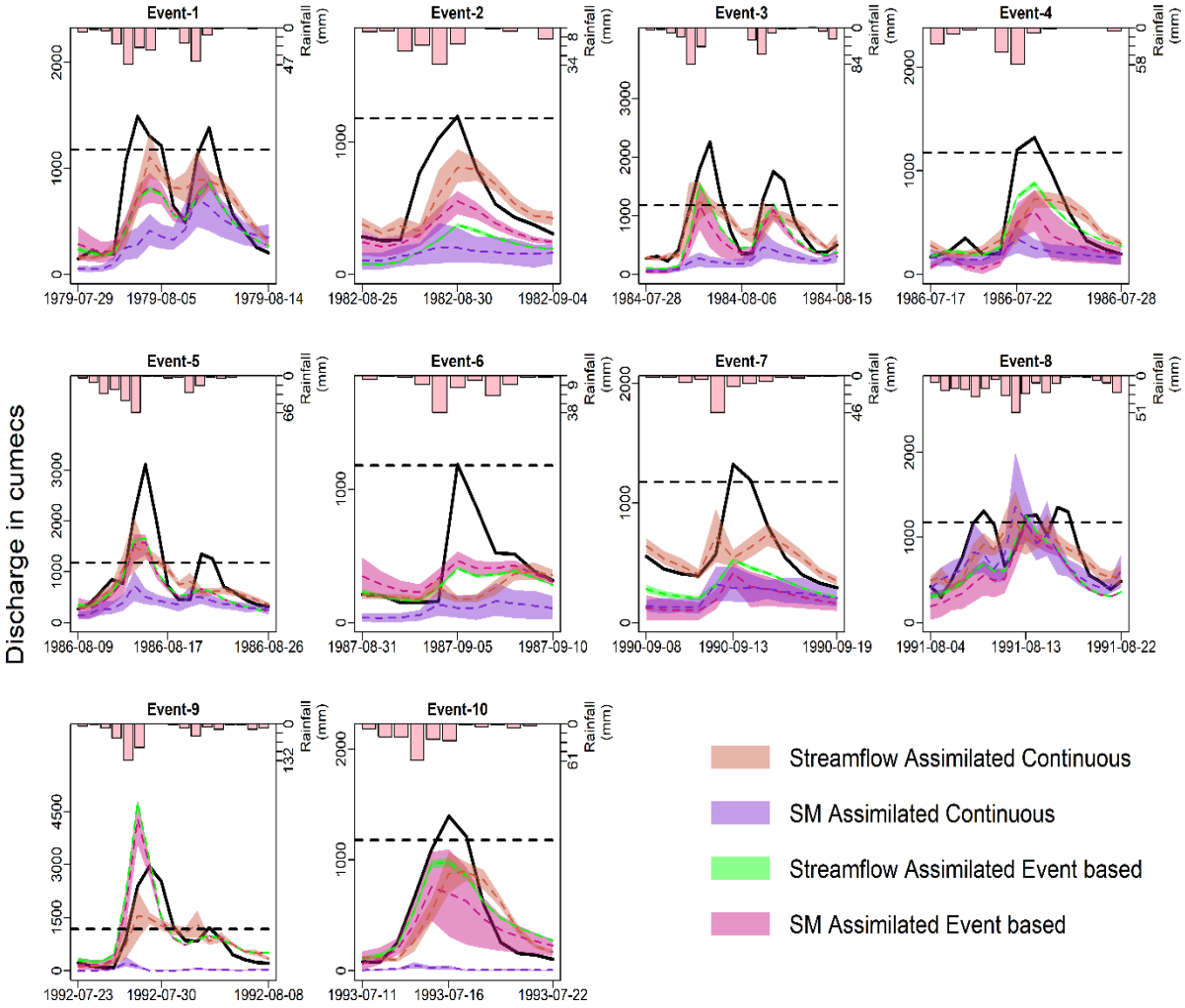

Figure 5. 5 Simulated flood hydrographs by the continuous and event-based lumped model based on Soil Moisture (SM) and streamflow data assimilation. Observed flood hydrographs were also plotted for the selected flood events at Jagdalpur during validation period.

A significant improvement has been observed in the predictive ability of both continuous and event-based semi-distributed based GR4J based on SM-DA and Q-DA than their lumped counterparts in terms of NSE, PEPF, PETP and PBIAS. During the calibration period, the event-based semi-distributed model with SM-DA showed the highest performance, with a median NSE value of 0.82. This highlights the advantage of incorporating spatial variability in the model. However, similar to lumped models, the temporal transferability of the semi-distributed models in the validation period was poor. Nevertheless, the event-based semi-distributed model with Q-DA demonstrated satisfactory performance with median NSE values of 0.64 and 0.57 in calibration and validation periods, respectively. Compared to lumped models, the semi-distributed models exhibited better PETP values, indicating a more accurate capture of the timing to peak.

Additionally, the PEPF and PBIAS values of the semi-distributed models were relatively lower than those of the lumped models. The flood hydrographs generated from the semi-distributed models during the calibration period and validation period were presented in Figure 5.6 and Figure 5.7, respectively. The ensemble spread of the semi-distributed model simulations was lower than that of the lumped models, plausibly due to accounting for spatial heterogeneity. The calibration process may have also compensated for biases in the soil moisture, leading to improved performance of SM-DA in the semi-distributed models.

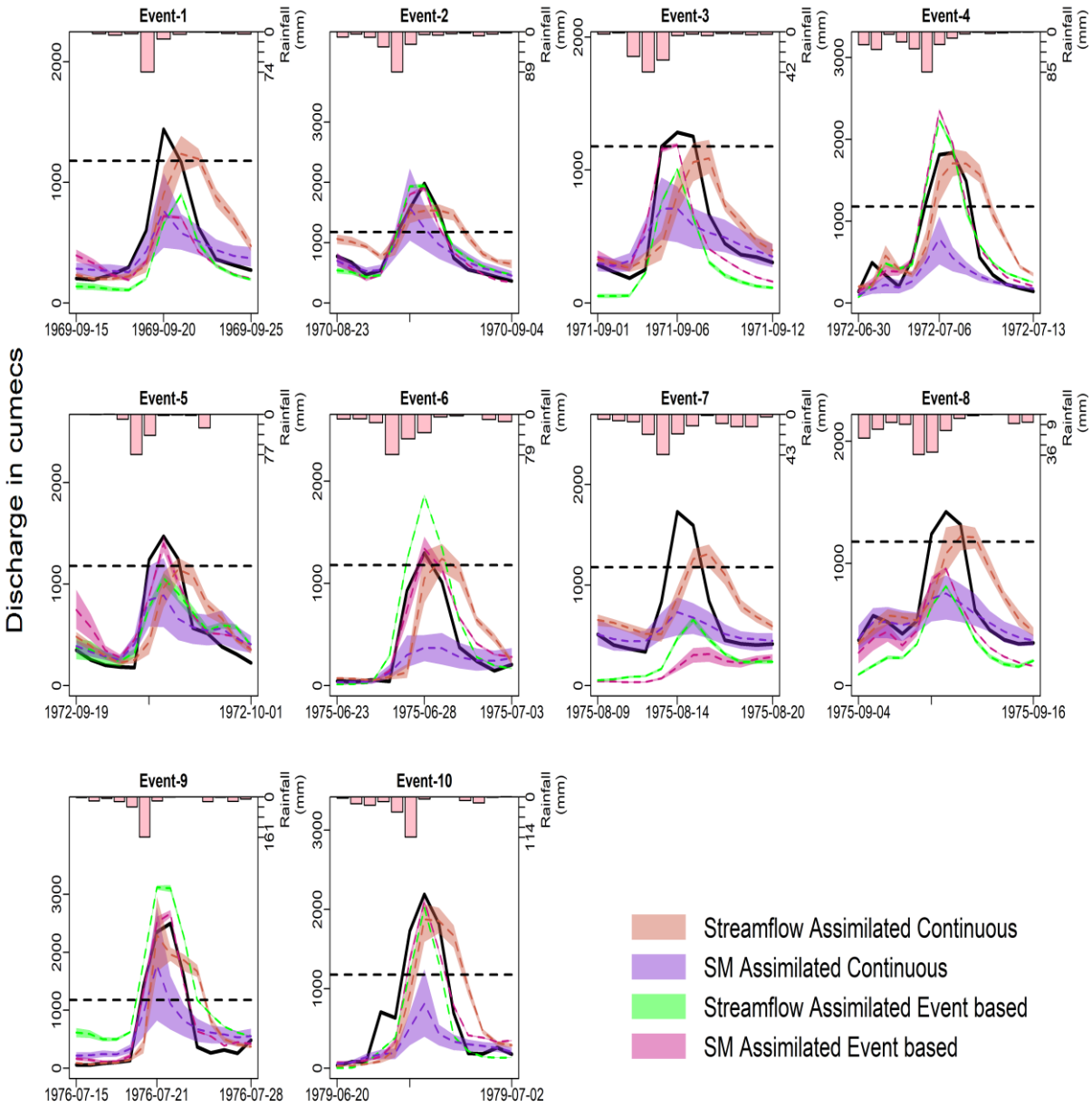

Figure 5.6 Simulated flood hydrographs by the continuous and event-based semi-distributed model based on Soil Moisture (SM) and streamflow data assimilation. Observed flood hydrographs were also plotted for the selected flood events at Jagdalpur during calibration.

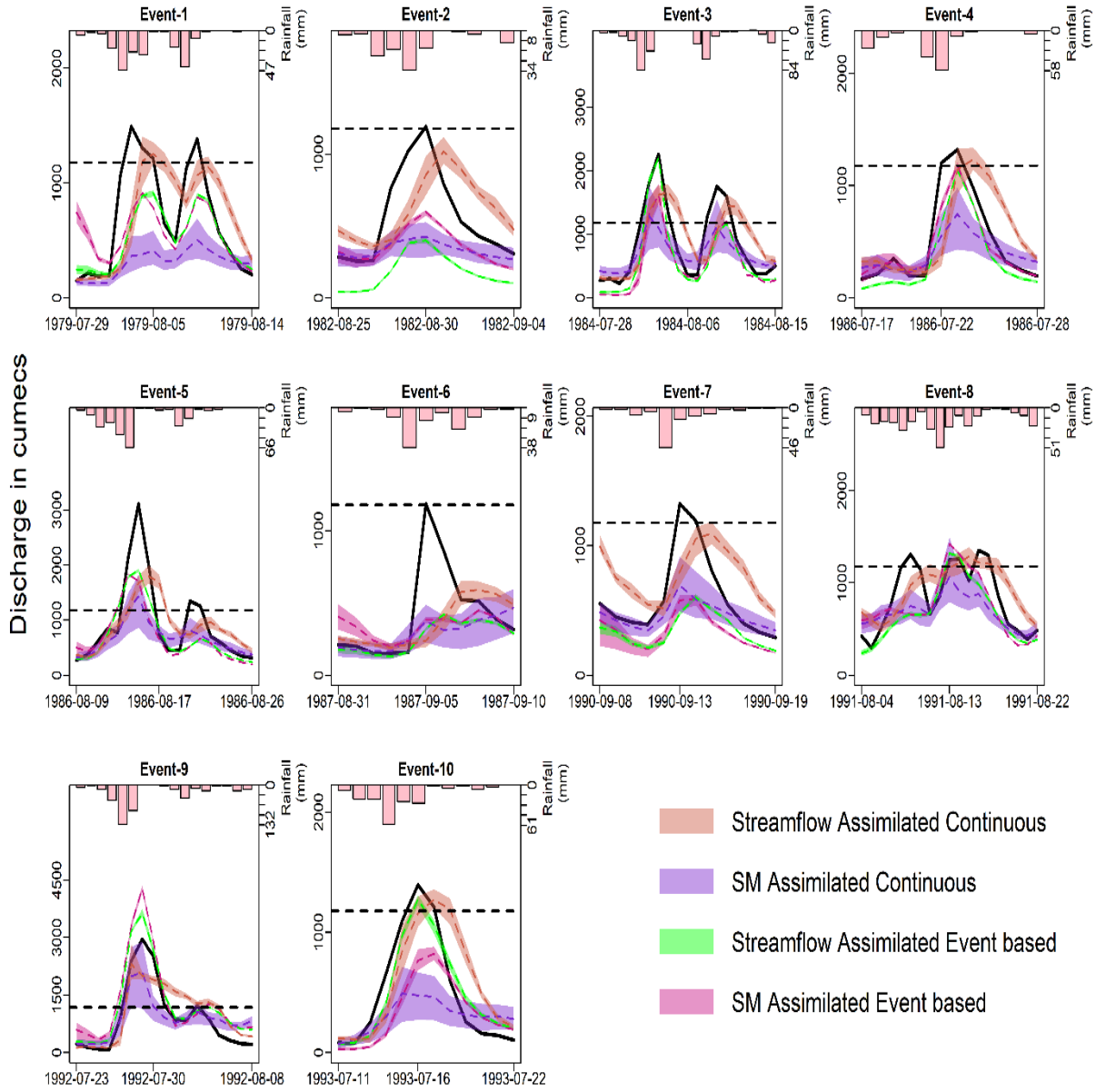

Figure 5.7 Simulated flood hydrographs by the continuous and event-based semi-distributed model based on Soil Moisture (SM) and streamflow data assimilation. Observed flood hydrographs were also plotted for the selected flood events at Jagdalpur during validation.

5.4.3 Performance of event-based models using IHC obtained through continuous models calibrated with different objective functions

The continuous lumped and semi-distributed models were calibrated using NSE, logNSE, KGE, and R4MS4E. The resulting model states were utilized as initial hydrological conditions (IHCs) for their respective event-based models. The performance assessment statistics of these event-based models are provided in Table 5.6. The median NSE values of the event-based lumped models calibrated using NSE, KGE, and R4MS4E in the calibration period were good (> 0.65), except for the logNSE calibrated model. The PEPF (< 30%) and PETP values indicate that the performance of all the lumped models in capturing the magnitude and timing of flood peaks was satisfactory during the calibration period. The KGE-calibrated lumped event-based model performed best in capturing the flood peak in both calibration and validation periods. However, a decline in the performance of all the models in terms of the chosen evaluation statistics was observed during the validation period. Nevertheless, the performance of all the models during the validation period regarding NSE values was satisfactory (>0.5). The flood hydrographs observed and simulated during the calibration period and validation period are presented in Figure 5.8 and Figure 5.9, respectively. The figure shows that all the simulated flood hydrographs are approximately parallel, and their disposition highly depends on the IHC. This similarity in performance of all the event-based lumped models coupled with continuous models calibrated using different metrics is due to their reliance on the same IHC.

Table 5.6 Median values of performance evaluation metrics computed for the flood events simulated using lumped and semi-distributed event-based GR4J model setups calibrated using multiple objective functions at Jagdalpur during calibration and validation periods.

Calibration Objective Function	**NSE**	**PEPF**	**PBIAS**	**PETP**	**NSE**	**PEPF**	**PBIAS**	**PETP**
Lumped								
	Calibration				**Validation**			
KGE	0.69	**-25.65**	**0**	9.34	0.55	**-24.72**	**0**	5.33
R4MS4E	0.67	-25.96	0	**-2.61**	0.55	-25.58	0	**-0.35**
LogNSE	0.49	-30.63	0	-12.57	0.51	-30.1	0	-14.15
NSE	**0.7**	-26.57	0	5.58	**0.57**	-26.43	0	-2.86
Discretized								
	Calibration				**Validation**			
KGE	0.85	2.99	0	4.23	0.65	**-12.88**	**0**	**-2.27**
R4MS4E	0.79	-13.67	0	3.86	0.68	-23.32	0	-10.03
LogNSE	0.77	-11.36	0	2.57	0.56	-24.83	0	-8.42
NSE	**0.91**	**-3.25**	**0**	**0.44**	**0.77**	-13.21	0	-8.73

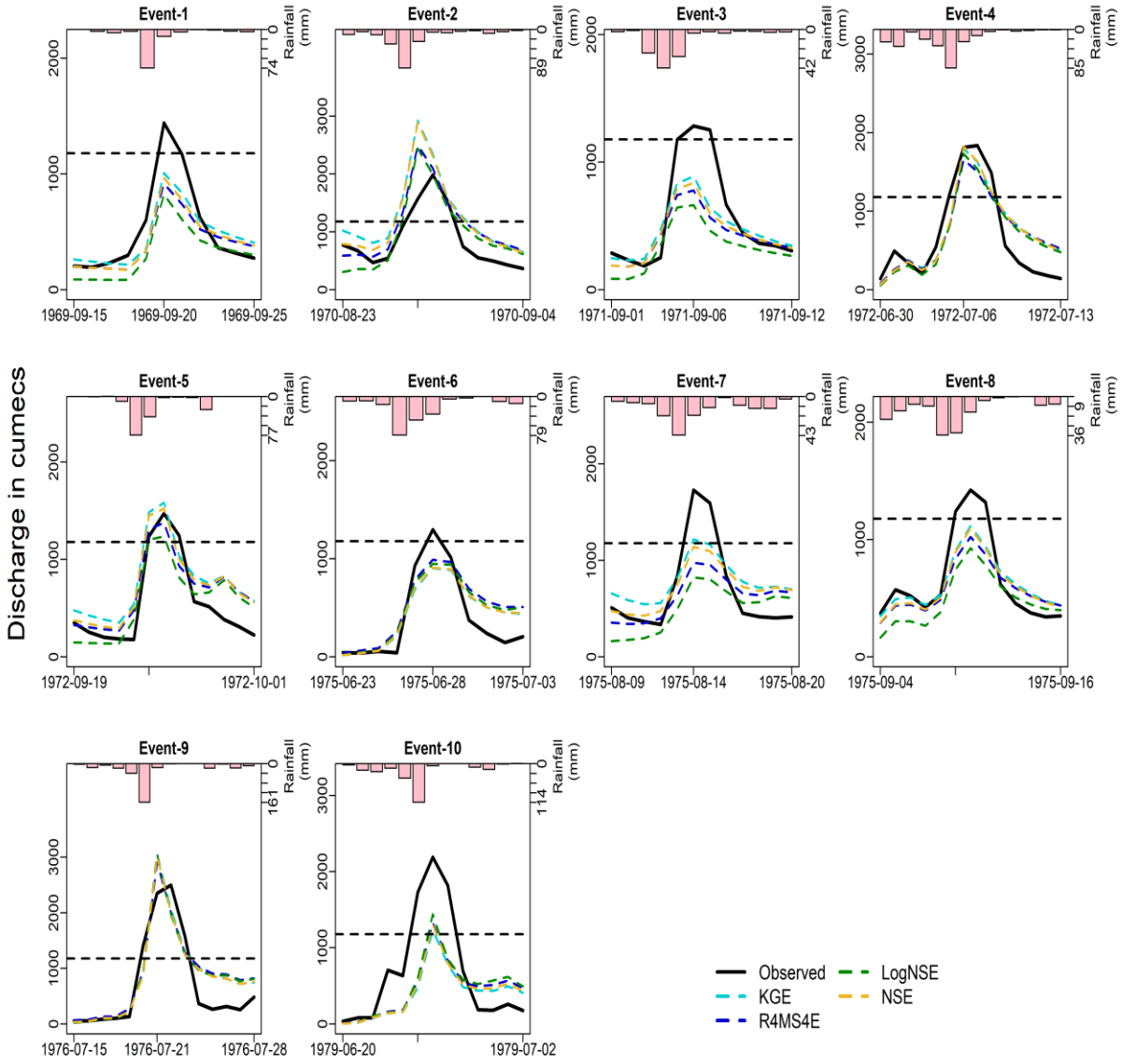

Figure 5.8 Observed and simulated flood hydrographs from the event-based lumped model at Jagdalpur for selected flood events during the calibration period.

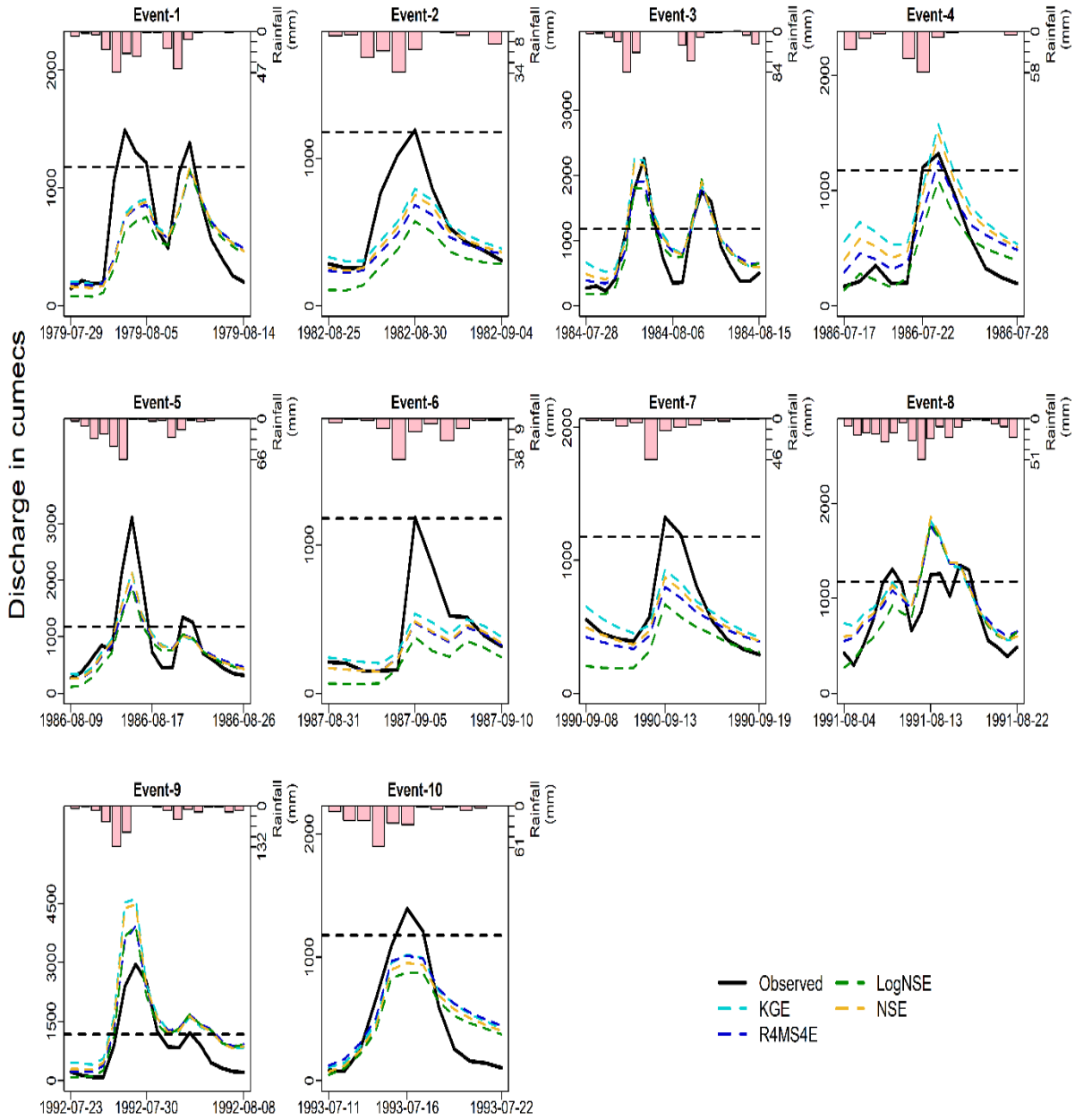

Figure 5.9 Observed and simulated flood hydrographs from the event-based lumped model at Jagdalpur for selected flood events during the validation period.

In the case of semi-distributed models, the simulated flood hydrographs accurately represent the observed flood hydrographs. During the calibration period, the median NSE value of all semi-distributed models exceeds 0.77, indicating good performance. Moreover, the semi-distributed model based on KGE, R4MS4E, and NSE-calibrated continuous models demonstrated good performance in the validation period (NSE > 0.65). The PEPF values are below 30% in both calibration and validation periods, indicating that the models can capture the observed flood magnitudes effectively. The observed and simulated hydrographs for the semi-distributed event-based model coupled with multiple calibrated continuous models are shown for both calibration and validation periods in Figure 5.10 and Figure 5.11, respectively. Unlike the lumped models, the simulated hydrographs for the semi-distributed models are not parallel because of the unique parametrization of each sub-basin and channel routing. Moreover, the simulated hydrographs are capable of capturing the observed peaks during most of the events in the calibration period. However, in the validation period, a slight underestimation bias is present in all semi-distributed models.

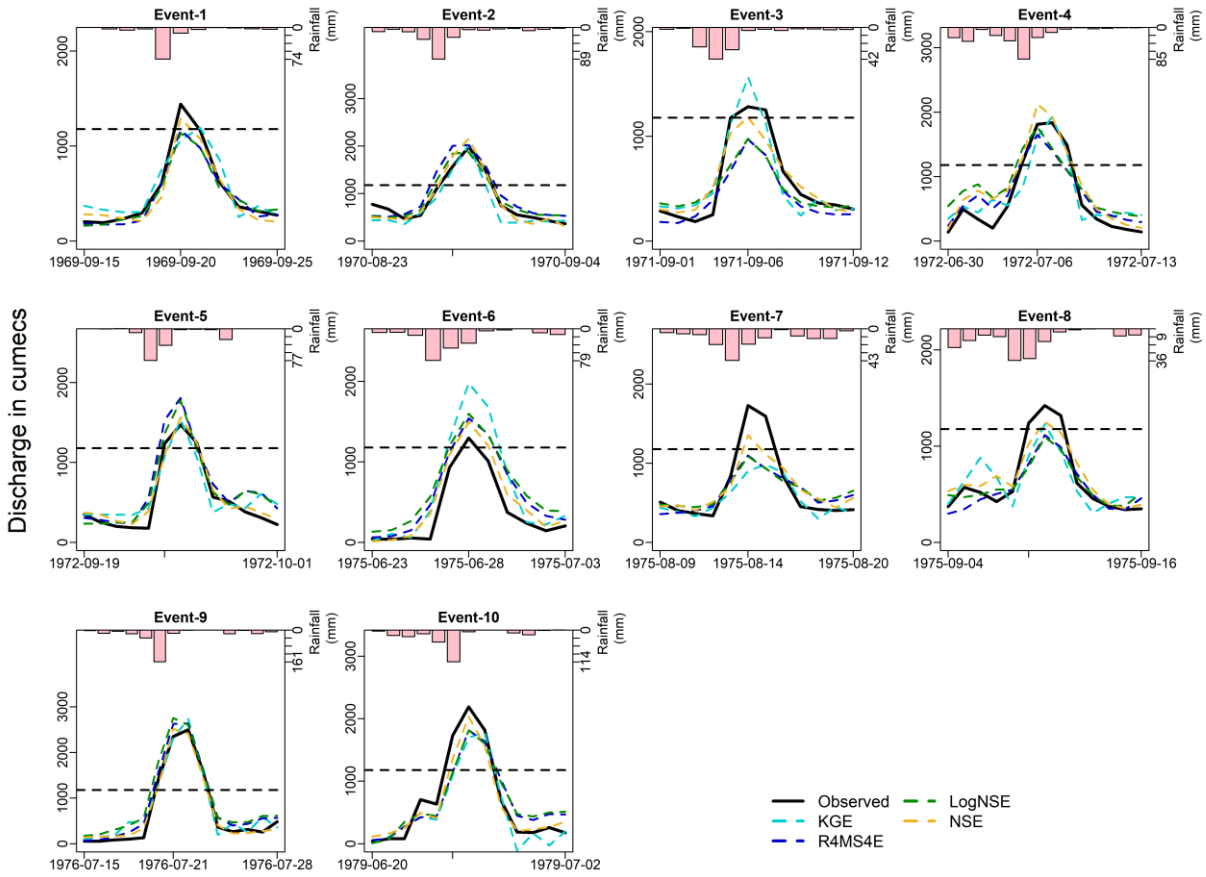

Figure 5.10 Observed and simulated flood hydrographs from the event-based semi-distributed model at Jagdalpur for selected flood events during the calibration period.

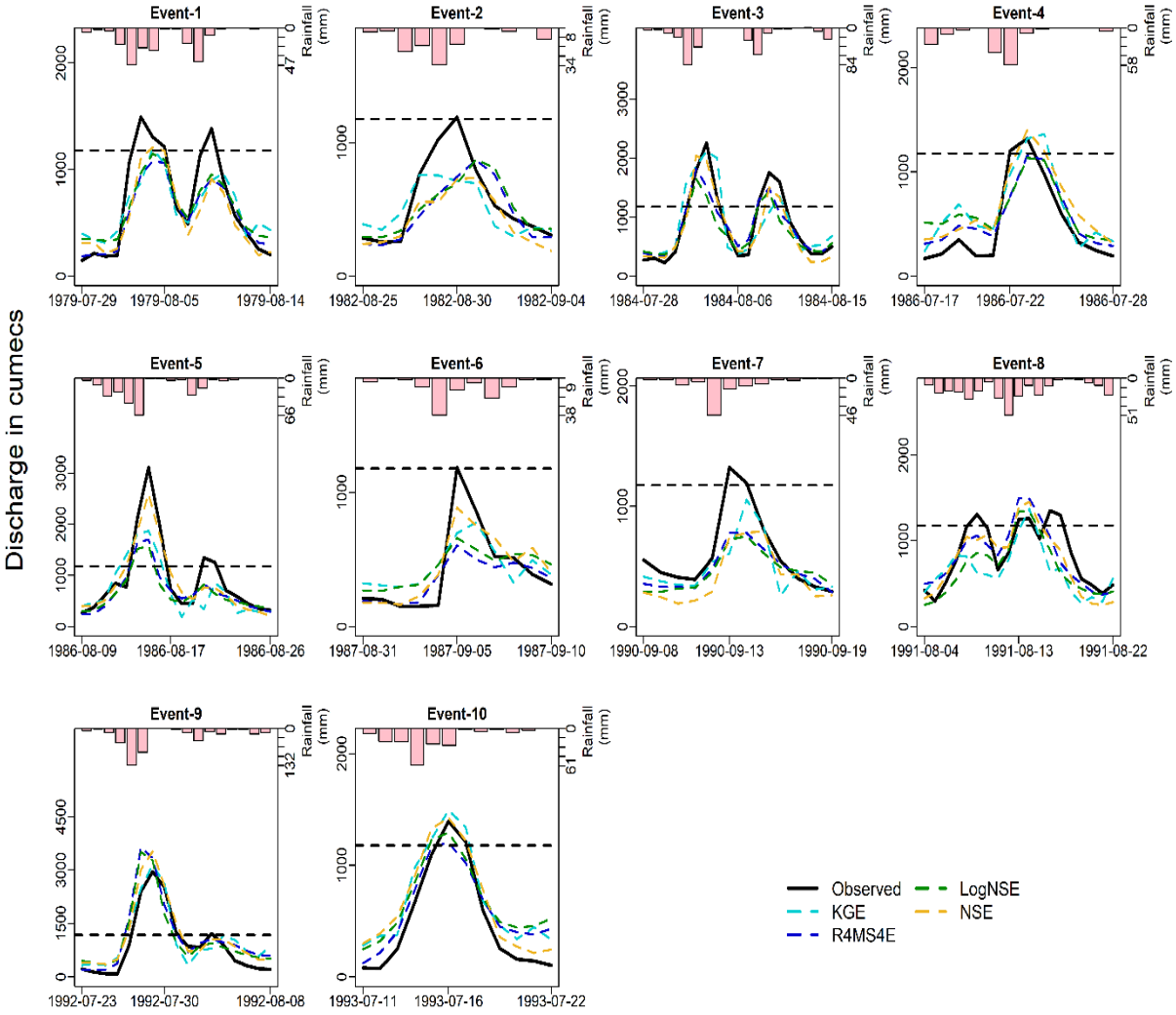

Figure 5.11 Observed and simulated flood hydrographs from the event-based semi-distributed model at Jagdalpur for selected flood events during the validation period.

5.5 Summary and Conclusions

This study evaluates the performance of event-based models based on the Initial Hydrologic Conditions (IHC)/model states using different methods. The current study on event-based models led to the following insights:

i. Streamflow Assimilation (Q-DA) in both lumped and semi-distributed models led to improved simulations at the catchment outlet during both calibration and validation periods. In contrast, assimilating soil moisture (SM-DA) did not enhance continuous streamflow simulations in the lumped model, but showed a slight improvement in the semi-distributed model.

ii. Semi-distributed model resulted in enhanced streamflow simulation in both SM-DA and Q-DA when compared to their lumped counterparts. This result is in line with the findings of Alvarez-Garetton et al. (2015).

iii. Compared to simulations produced by continuous models with DA, the ensemble spread of simulations generated from the event-based model using IHCs obtained through DA is lower. However, simulations based on SM-DA have a higher ensemble spread than those based on Q-DA.

iv. In event-based models with initial conditions generated through assimilation, SM-DA based semi-distributed model performed exceptionally well in the calibration period exploiting the advantage of the model spatial resolution.

v. The event-based semi-distributed GR4J initialised by the IHCs extracted from the corresponding NSE calibrated continuous model is found to be outperforming other models in terms of capturing the magnitude, timing and volume of observed flood events.

vi. In conclusion, the findings suggest the potential benefits of assimilating streamflow and employing a semi-distributed model structure. However, limitations such as overfitting of parameters and uncertainties in channel characteristics may affect model performance (Jehn et al., 2019; Khakbaz et al., 2012). Future research can explore the integration of the rainfall spatial variability in a lumped model to exploit its computational advantage (Zhou et al., 2021). Additionally, further investigation is needed to evaluate the efficiency of combined assimilation of soil moisture in IHC estimation and to assess the suitability of EnKF for assimilation in various hydrological models. Overall, this study provides valuable insights for

practical applications and calls for large sample investigations using multiple model structures (Gupta et al., 2014).

Chapter6

Generation of Ensemble Flood Forecasts

6.1 Introduction

Over the past few decades, numerous unexpected natural disasters have occurred, and floods, in particular, have been recognized as highly unpredictable events that can cause devastation to both human and animal lives and properties (Alfieri et al., 2017). Predicting the timing and frequency of severe hydrometeorological events that lead to flooding is challenging. To minimize human casualties and property damage, early warning systems and flood mitigation measures rely on accurate estimations of runoff volume and flood peaks. To effectively plan and manage water resources sustainably, a comprehensive understanding of the factors influencing surface runoff in catchment areas is essential.

Simulating rainfall and runoff to comprehend the hydrologic response of a region has long been established as a reliable method. Rainfall-runoff models are commonly used to predict floods, measure water levels under different conditions, and provide flood forecasts. Various factors, such as land use, slope, vegetation, and storm attributes, including rainfall duration, volume, and intensity, play significant roles in determining the amount of surface runoff. During emergencies caused by heavy rainfall, decision-makers heavily rely on accurate predictions. Precipitation data is crucial for flood forecasting, with two key factors influencing its effectiveness: accuracy and lead time. The lead time of flood forecasts can be enhanced by using precipitation forecasts obtained from numerical weather prediction (NWP) models. NWPs facilitate the collection of valuable flood-related information and enable the dissemination of early flood warnings. Forecasters are transitioning from deterministic to ensemble forecasts to address the uncertainties in meteorological and hydrological systems. A deterministic forecast can be transformed into a fully probabilistic ensemble by considering uncertain boundary conditions and incorporating data assimilation techniques. This shift in approach is particularly relevant for precipitation and runoff forecasting, introducing a novel way of conducting these predictions.

An interactive grand global ensemble system has been developed to account for the uncertainties in projections from multiple global models. The International Grand Global Ensemble (TIGGE) initiative collects and analyzes prediction data from major forecasting centres worldwide. By combining various sources of uncertainty, the TIGGE ensemble prediction generates a probability distribution that provides valuable insights for decision-making. When the ECMWF (European Centre for Medium-Range Weather Forecasts) and the NCEP (National Centers for Environmental Prediction) initially developed ensemble predictions in 1992, they were extensively used in inflow forecasts due to their reliance on a wide range of models. The application of meteorological predictions, particularly in flood forecasting, has proven beneficial. Ensemble predictions, which involve generating multiple forecasts using different physical parameterizations and initial conditions, have significantly improved meteorological forecasts. The TIGGE (The International Grand Global Ensemble) project serves another purpose: establishing an early warning system based on operational medium-range ensemble predictions from various numerical weather centres such as the ECMWF, UKMO, and NCEP.

NWP models' Quantitative Precipitation Forecasts (QPF) are crucial inputs for hydrological models when estimating streamflow. Uncertainty in QPF arises from inaccuracies in initial conditions, approximations of atmospheric processes, and the NWP model's overall predictive ability. Given that streamflow estimates heavily influence decision support for water infrastructure management, flood and drought warnings, and reservoir operations, there is a growing interest in probabilistic forecasts that can estimate the likelihood of future weather events. Post-processing techniques based on statistical models have been developed to improve the validity of QPFs, whether deterministic or ensemble, for estimating streamflow. These techniques leverage the relationship between observations and NWP forecasts, estimate model parameters using historical data, and generate post-processed ensemble forecasts for the future.

With over 5,500 major dams, India requires accurate inflow projections for effective flood damage reduction. The Central Water Commission (CWC) currently provides inflow projections for over 150 locations in India, but expanding this coverage to include all major and minor dams and key cities is necessary. Improving the flood forecasting system in India is essential to provide accurate predictions at critical locations and sites. Specifically, it is also important to generate and verify ensemble flood forecasts over the Godavari River basin as it is often prone to floods that can cause devastating damages (Garg and Mishra, 2019; Rakhecha and Singh, 2017). The increasing frequency and severity of extreme precipitation events, including cloud bursts, emphasize the need

for ensemble predictions in flood forecasting. This study aims to develop an atmospheric-hydrologic flood forecasting model over the Wardha basin (a sub-basin of Godavari River Basin) by forcing quantitative precipitation forecasts obtained from ECMWF into calibrated hydrological models for short and medium range lead times.

6.2 Results

The raw and post-processed (using Quantile Regression Forests (QRF)) ensemble precipitation forecasts were forced into the calibrated lumped and semi-distributed GR4J model to generate ensemble streamflow forecasts. Short and medium-range ensemble flood forecasts were developed for seven historical flood events at Bamni (Wardha), where the first four events fall under the calibration period of the post-processor, and the following three events fall under the validation period of the post-processor. The IHC of the event-based models are initialized using the outputs from their respective continuous models.

The generated ensemble flood forecasts were plotted against the observed discharge values for the events during calibration and validation periods of the post-processor in Figures 6.1 and 6.2, respectively. The figures show a discrepancy between the observed and simulated flood peaks obtained from the raw forecasts in both the calibration and validation periods. Additionally, the ensemble spread appears to be smaller. The flood forecasts generated by the QRF in the post-processor exhibit good performance during the calibration period, accurately capturing the magnitude and timing of the observed flood peaks. However, this improved performance does not extend to the validation period, as the QRF-based flood forecasts underestimate the observed flood peaks. During the validation period, there is still a mismatch between the peaks of the raw flood forecasts and the observed flood peaks. However, regarding the magnitude of the flood peak, the raw flood forecasts perform reasonably well compared to the post-processed forecasts. Regarding the hydrological model, semi-distributed models outperform lumped models in both the calibration and validation periods.

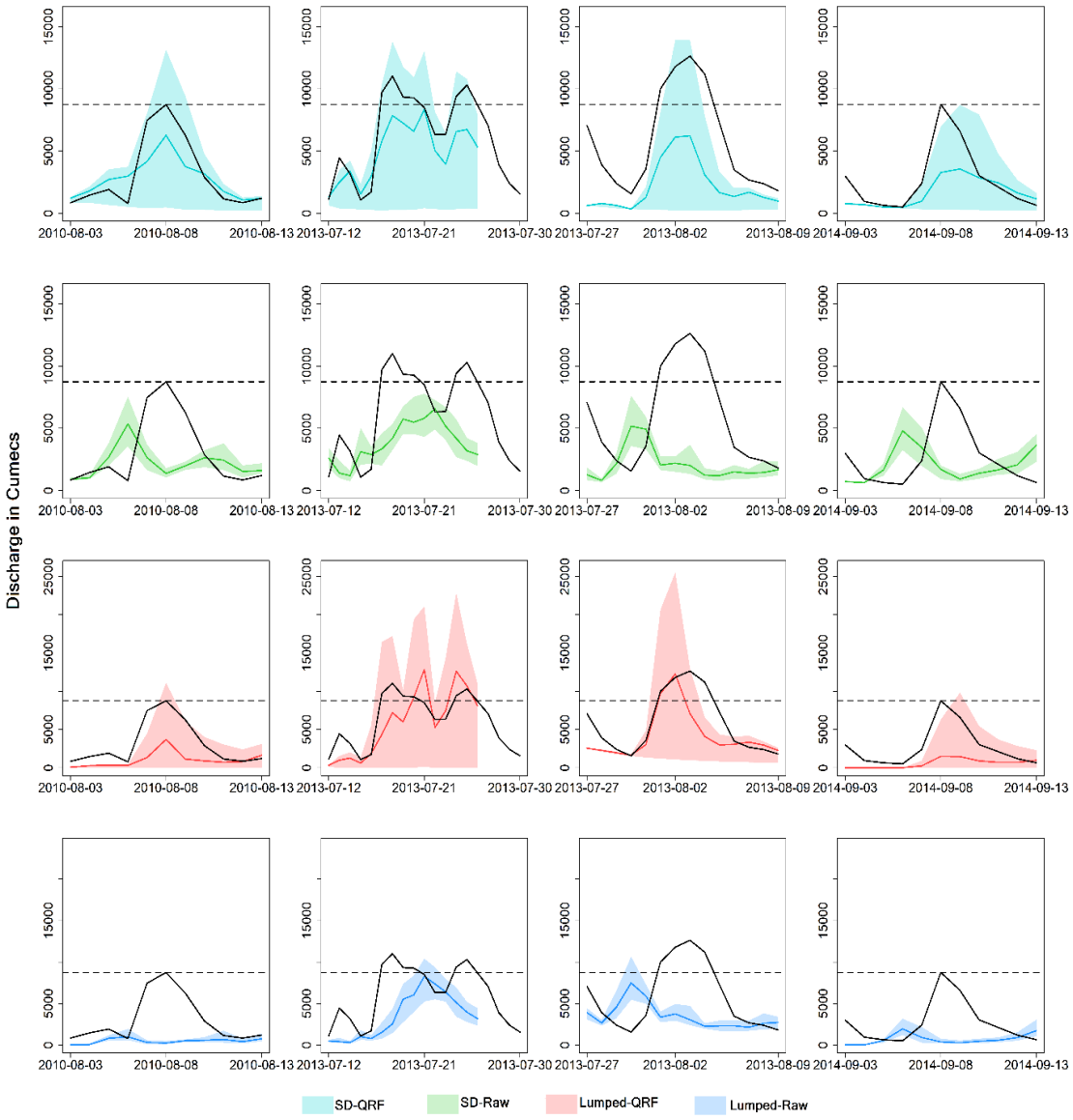

Figure 6.1 Ensemble flood forecasts generated by forcing raw and post-processed ensemble precipitation forecasts into the lumped and semi-distributed GR4J model along observed discharge (black).

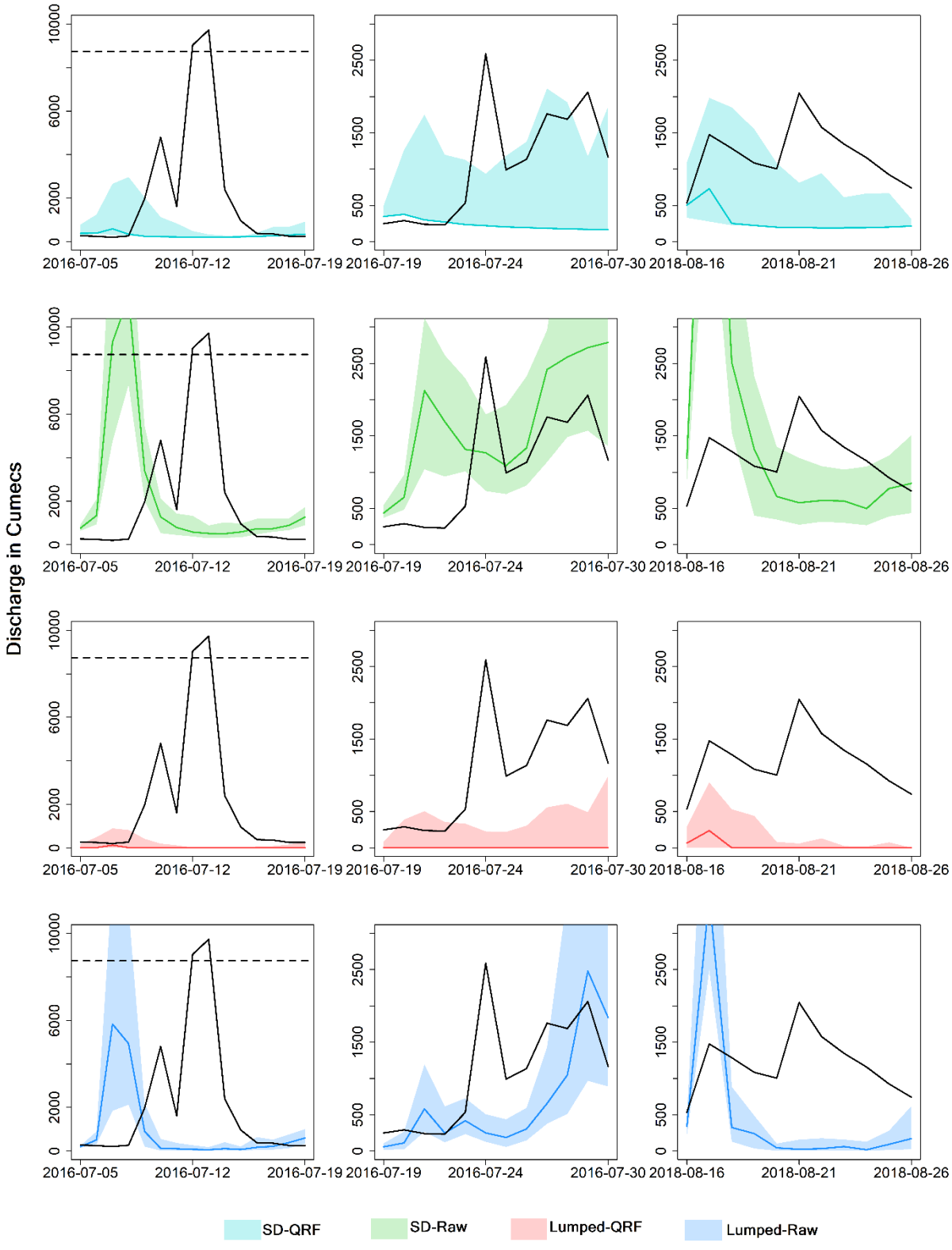

Figure 6.2 Ensemble flood forecasts generated by forcing raw and post-processed ensemble precipitation forecasts into the lumped and semi-distributed GR4J model along observed discharge (black).

Moreover, the performance of the EFF is evaluated using the Nash Sutcliffe Efficiency (NSE) and mean Continuous Ranked Probability Score (CRPS) as performance evaluation measures. Figures 6.3 and 6.4 illustrate the mean CRPS and maximum NSE of the ensemble members plotted against lead times during the calibration and validation periods of the post-processor, respectively. Based on Figure 6.3, it is evident that the EFF performs satisfactorily in terms of both NSE and mean CRPS during the calibration period. The EFF, based on the QRF+semi-distributed model, exhibits good performance with an NSE>0.5 and mean CRPS<1700 m^3/s at all lead times. However, the performance during the validation period is unsatisfactory, with an NSE<0.5. Additionally, the performance of the generated EFF is found to be satisfactory at shorter lead times, specifically 1 to 3 days, with Percentage Error in Peak Flow (PEPF), Percentage Error in Volume (PEV), and Percentage Error in Timing to Peak (PETP) < 30% in most cases during the calibration period of the post-processor. However, the performance of the post-processed forecasts is poor during the validation period, and as the lead time increases, there is a discrepancy in predicting the timing of flood peaks. This result is in line with the results of Patel and Yadav (2022) for the Sabarmati River basin. The results suggest that the performance of semi-distributed model-based EFF is found to be better in comparison to its lumped counterparts.

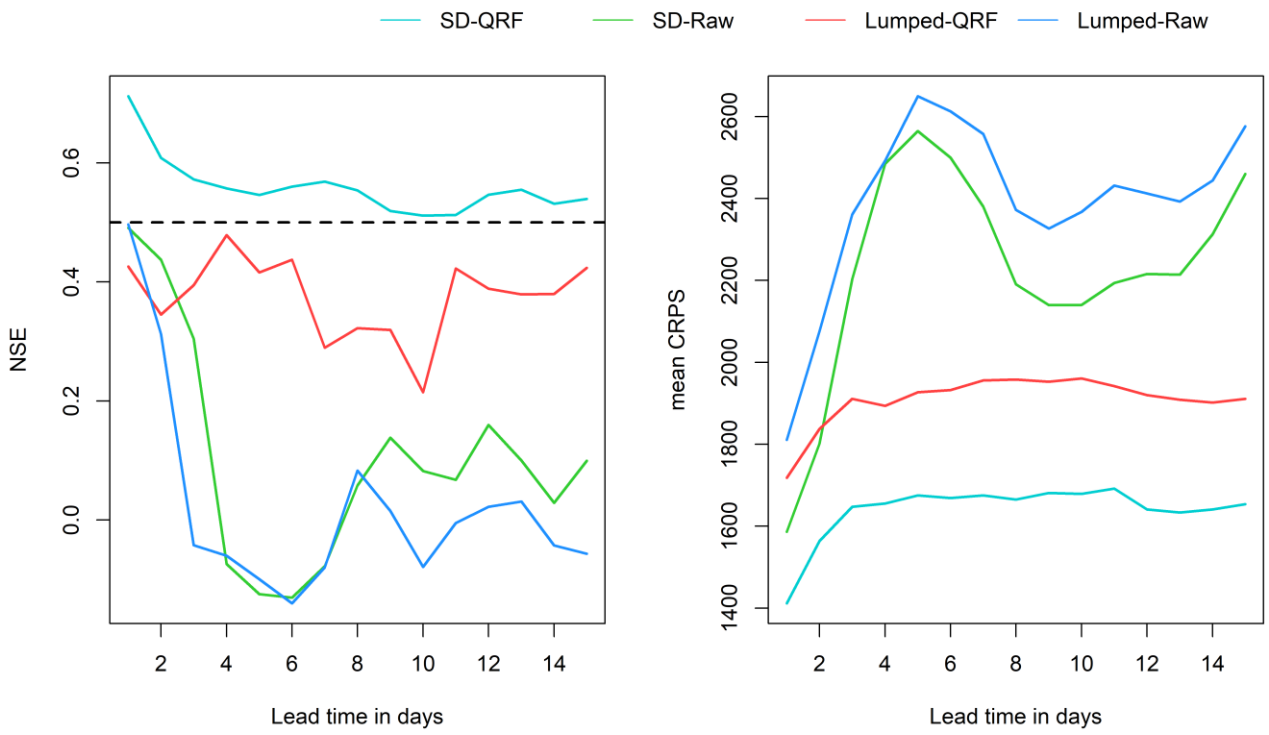

Figure 6.3 NSE and mean CRPS values of ensemble flood forecasts with varying lead times during the calibration period of the post-processor.

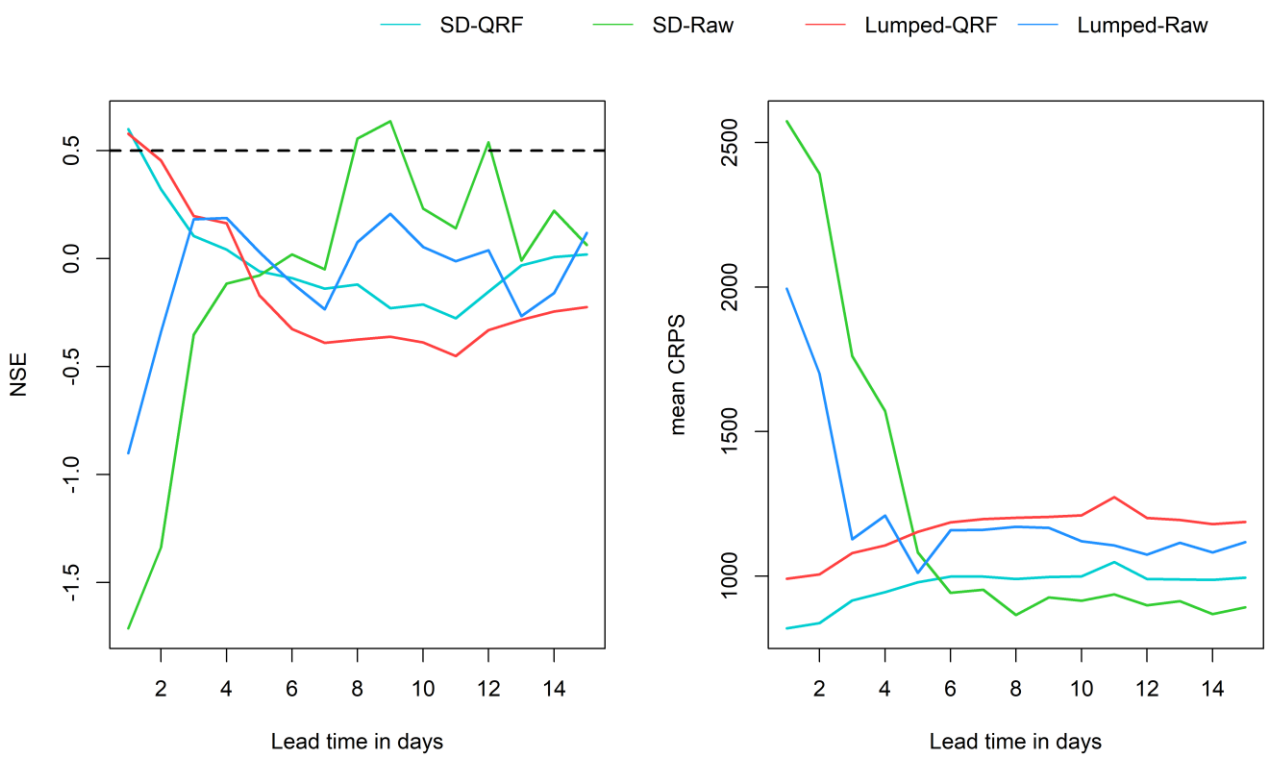

Figure 6.4 NSE and mean CRPS values of ensemble flood forecasts with varying lead times during the validation period of the post-processor.

6.3 Conclusions

The present study utilized raw and post-processed ensemble precipitation forecasts, incorporating Quantile Regression Forests (QRF), to generate ensemble streamflow forecasts using the lumped and semi-distributed GR4J models. The following conclusions were drawn from the study:

- The ensemble spread of raw forecasts was smaller, but post-processing techniques successfully addressed this issue by increasing the spread of the Ensemble Flood Forecasts (EFF).
- During the calibration period, the post-processed flood forecasts, utilizing QRF, exhibited improved performance by accurately capturing both the magnitude and timing of observed flood peaks. However, this improved performance did not carry over to the validation period, as the QRF-based forecasts underestimated the observed flood peaks.
- Despite a mismatch between the peaks of the raw flood forecasts and the observed flood peaks during the validation period, the raw forecasts performed reasonably well in terms of the magnitude of the flood peak when compared to the post-processed forecasts.

- The semi-distributed models outperformed the lumped models in both the calibration and validation periods of the hydrological model, demonstrating the superiority of the semi-distributed approach.
- The EFF based on the QRF+semi-distributed model exhibited satisfactory performance during the calibration period, with NSE>0.5 and mean CRPS<1700 m3/s at all lead times. However, the performance during the validation period was unsatisfactory, with NSE<0.5, indicating the need for further improvements.
- The generated EFF showed satisfactory performance at shorter lead times (1 to 3 days) during the calibration period, as indicated by Percentage Error in Peak Flow (PEPF), Percentage Error in Volume (PEV), and Percentage Error in Timing to Peak (PETP) < 30% in most cases.
- However, the post-processed forecasts performed poorly during the validation period, and as the lead time increased, there was a discrepancy in predicting the timing of flood peaks, suggesting a limitation in long-term forecast accuracy.

Chapter 7

Summary

7.1 Summary

This thesis aims to address key research gaps in ensemble flood forecasting for the Godavari River Basin (GRB) in India. The study focuses on developing an ensemble flood forecasting framework for the GRB by identifying the best ensemble weather forecast products and post-processing methods, analyzing the impact of model resolution on flood peak simulation using event-based semi-distributed models, and evaluating the accuracy of assimilating observed initial hydrologic states/conditions and soil moisture data into event-based rainfall-runoff models. The research aims to improve flood predictions and enhance flood risk management strategies by incorporating hydrologic uncertainties, determining optimal model resolution, and integrating real-time observational data. The following paragraphs give a summary and conclusions of the study presented in the thesis.

- The findings of this study reveal that both NCEP and ECMWF raw forecasts exhibit poor skill in capturing observed extreme precipitation events across all lead times, and statistical post-processing methods are unable to alleviate this issue. This highlights the need for further development of the underlying physics of Numerical Weather Prediction (NWP) models to improve the accuracy of forecasting extreme precipitation events. The correlation between ensemble mean and observed precipitation decreases with increasing lead time, while the Root Mean Squared Error (RME) does not depend on lead time. The ensemble mean of post-processed NCEP and MME forecasts using Quantile Regression Forest (QRF) outperform other forecasts in terms of correlation coefficient and RME in all subbasins and at all lead times. The post-processed forecasts also exhibit an improved ensemble spread-error relationship compared to the raw forecasts. QRF outperforms Quantile Mapping (QM) in preserving the ensemble spread-error relation. Rank histograms indicate that both NCEP and ECMWF raw forecasts are under-dispersive and biased in all subbasins, but post-processing mitigates the bias issue. Reliability diagrams show that raw NCEP and ECMWF forecasts tend to be overconfident, while post-processed forecasts perform well at a 1-day lead time but exhibit declining reliability with increasing lead

times. The Area Under the Curve (AUC) values, indicating forecast usefulness, are found to be greater than 0.75 at all lead times and subbasins. However, the discrimination ability of the forecasts, as measured by AUC, decreases with lead time, indicating a higher false alarm rate. Comparatively, the performance of raw MME forecasts is better than raw NCEP and ECMWF forecasts, but the QRF post-processed NCEP and MME forecasts perform equally well. Considering computational costs, a 20-member QRF post-processed NCEP forecast ensemble is recommended for hydrologic forecasting applications in the study area. Overall, the performance of NCEP and MME forecasts surpasses that of ECMWF forecasts, and QRF post-processed forecasts outperform both QM post-processed and raw forecasts. The satisfactory performance of QRF post-processed NCEP and MME forecasts is observed in Lower Godavari, Middle Godavari, Indravati, Manjira, and Weinganga subbasins for both deterministic and probabilistic measures.

- This thesis investigated the trade-offs associated with a nested discretization scheme and the simulated streamflow by a conceptual hydrological model. Comparative analysis with continuous lumped models revealed that both semi-distributed and semi-lumped approaches exhibited substantial improvements in terms of Nash-Sutcliffe Efficiency (NSE), with values surpassing 0.85 at Jagdalpur and exceeding 0.84 at Wardha, while the lumped models achieved NSE values of 0.779 at Jagdalpur and 0.763 at Wardha. However, it was observed that semi-distributed models at finer discretization levels could encounter challenges during optimization, as they may become trapped in local optima due to a higher number of parameters requiring optimization. Consequently, the calibrated SD-3 parameters demonstrated slightly inferior performance during the validation period, with NSE values of 0.78 and 0.71 for SD1 and SD3, respectively, at Wardha. The semi-distributed and semi-lumped models showcased enhanced capabilities in capturing low flows, high flows, and water balance components compared to lumped models. In the case of semi-lumped models, performance increased with higher levels of discretization, as evidenced by NSE values of 0.72 and 0.82 for SL1 and SL3, respectively, during the validation period at Wardha. However, the performance gains associated with finer discretization levels (specifically level 3 in this study) were not significantly superior to those achieved by models at discretization level 1. These findings indicate that elevating model resolution with finely resolved input data does not inherently enhance model performance unless the input information is improved through the utilization of finer spatial resolution datasets.

- The accuracy of different methods to estimate the Initial Hydrologic Conditions of the event-based hydrological models was assessed. The application of Streamflow Assimilation (Q-DA) in both lumped and semi-distributed models yielded improved simulations at the catchment outlet during both calibration and validation periods. However, assimilating soil moisture (SM-DA) did not significantly enhance continuous streamflow simulations in the lumped model, whereas a slight improvement was observed in the semi-distributed model. Notably, the semi-distributed model demonstrated superior streamflow simulation performance in both SM-DA and Q-DA compared to the lumped models. When compared to continuous models with data assimilation (DA), the event-based model utilizing Initial Hydrological Conditions (IHCs) obtained through DA exhibited a lower ensemble spread in simulations. However, simulations based on SM-DA exhibited a higher ensemble spread compared to those based on Q-DA. In event-based models employing assimilated initial conditions, the semi-distributed model with SM-DA excelled during the calibration period, capitalizing on the advantages of its spatial resolution. Furthermore, the event-based semi-distributed GR4J model initialized by the IHCs extracted from the corresponding Nash-Sutcliffe Efficiency (NSE) calibrated continuous model outperformed other models in accurately capturing the magnitude, timing, and volume of observed flood events.
- Finally, Ensemble Flood Forecasts were generated by forcing quantitative precipitation forecasts obtained from NWPs into a calibrated hydrological model. The ensemble spread of raw forecasts was initially narrower, but post-processing techniques effectively addressed this issue by increasing the spread of the Ensemble Flood Forecasts (EFF). During the calibration period, the post-processed flood forecasts utilizing the Quantile Random Forest (QRF) method displayed improved performance by accurately capturing both the magnitude and timing of observed flood peaks. However, this improved performance did not extend to the validation period, as the QRF-based forecasts underestimated the observed flood peaks. Despite a discrepancy in peak timing between the raw flood forecasts and the observed peaks during validation, the raw forecasts performed reasonably well in terms of peak magnitude when compared to the post-processed forecasts. In terms of the hydrological model, the semi-distributed approach outperformed the lumped models in both the calibration and validation periods, showcasing its superiority. The EFF based on the QRF+semi-distributed model exhibited satisfactory performance during the calibration period, with Nash-Sutcliffe Efficiency (NSE) greater

than 0.5 and mean Continuous Ranked Probability Score (CRPS) less than 1700 m3/s at all lead times. However, the performance during the validation period fell below satisfactory levels, with NSE less than 0.5, indicating the need for further improvements. The generated EFF demonstrated satisfactory performance at shorter lead times (1 to 3 days) during calibration, with Percentage Error in Peak Flow (PEPF), Percentage Error in Volume (PEV), and Percentage Error in Timing to Peak (PETP) consistently below 30% in most cases. However, the post-processed forecasts performed poorly during the validation period, and as the lead time increased, there was a discrepancy in predicting the timing of flood peaks, highlighting a limitation in long-term forecast accuracy.

7.2 Future Scope

- The present work is mainly based on conceptual hydrologic models. However, utilization of process-based fully distributed models such as VIC, SWAT and HEC-HMS might give more reliable simulations as they address the spatial heterogeneity of both catchment characteristics and rainfall as well.
- The work was carried out with data of available spatial and temporal resolution. Finer resolution data would have given better results. Particularly, hourly meteorological forcing is essential in flood modelling.
- Root zone soil moisture from a single source (reanalysis product) has been used in this study. Assessing the suitability of different satellite-based soil moisture products is necessary for detailed understanding of the assimilation efficiency of models.
- The computational advantage of lumped models should be exploited by integrating the rainfall spatial variability to increase the model efficiency during anomalistic rainfall events.
- The influence of increasing model resolution on the performance of the model in capturing the observed flood events through assimilation of streamflow/soil moisture should be investigated.
- The trade-offs between model structure, spatial resolution and assimilation variable in assimilation efficiency is reserved for future research.

References

Abbaszadeh, P., Gavahi, K., Moradkhani, H., 2020. Multivariate remotely sensed and in-situ data assimilation for enhancing community WRF-Hydro model forecasting. Adv. Water Resour. 145, 103721. https://doi.org/10.1016/j.advwatres.2020.103721

Acero Triana, J.S., Chu, M.L., Guzman, J.A., Moriasi, D.N., Steiner, J.L., 2019. Beyond model metrics: The perils of calibrating hydrologic models. J. Hydrol. 578, 124032. https://doi.org/10.1016/j.jhydrol.2019.124032

Acharya, N., Chattopadhyay, S., Mohanty, U.C., Dash, S.K., Sahoo, L.N., 2013. On the bias correction of general circulation model output for Indian summer monsoon. Meteorol. Appl. 20, 349–356. https://doi.org/10.1002/met.1294

Ahmadi, M., Arabi, M., Ascough, J.C., Fontane, D.G., Engel, B.A., 2014. Toward improved calibration of watershed models: Multisite multiobjective measures of information. Environ. Model. Softw. 59, 135–145. https://doi.org/https://doi.org/10.1016/j.envsoft.2014.05.012

Alfieri, L., Bisselink, B., Dottori, F., Naumann, G., de Roo, A., Salamon, P., Wyser, K., Feyen, L., 2017. Global projections of river flood risk in a warmer world. Earth's Futur. 5, 171–182. https://doi.org/10.1002/2016EF000485

Alvarez-Garreton, C., Ryu, D., Western, A.W., Crow, W.T., Robertson, D.E., 2014. The impacts of assimilating satellite soil moisture into a rainfall-runoff model in a semi-arid catchment. J. Hydrol. 519, 2763–2774. https://doi.org/10.1016/j.jhydrol.2014.07.041

Alvarez-Garreton, C., Ryu, D., Western, A.W., Su, C.H., Crow, W.T., Robertson, D.E., Leahy, C., 2015. Improving operational flood ensemble prediction by the assimilation of satellite soil moisture: Comparison between lumped and semi-distributed schemes. Hydrol. Earth Syst. Sci. 19, 1659–1676. https://doi.org/10.5194/hess-19-1659-2015

Anderson, S.R., Csima, G., Moore, R.J., Mittermaier, M., Cole, S.J., 2019. Towards operational joint river flow and precipitation ensemble verification: considerations and strategies given limited ensemble records. J. Hydrol. 577, 123966. https://doi.org/10.1016/j.jhydrol.2019.123966

Andreadis, K.M., Clark, E.A., Lettenmaier, D.P., Alsdorf, D.E., 2007. Prospects for river discharge and depth estimation through assimilation of swath-altimetry into a raster-based hydrodynamics model. Geophys. Res. Lett. 34.

Andreadis, K.M., Lettenmaier, D.P., 2006. Assimilating remotely sensed snow observations into a macroscale hydrology model. Adv. Water Resour. 29, 872–886.

Bafitlhile, T.M., Li, Z., 2019. Applicability of ε-Support Vector Machine and artificial neural network for flood forecasting in humid, semi-humid and semi-arid basins in China. Water (Switzerland) 11. https://doi.org/10.3390/w11010085

Baguis, P., Roulin, E., 2017. Soil moisture data assimilation in a hydrological model: A case study in Belgium using large-scale satellite data. Remote Sens. 9, 1–26. https://doi.org/10.3390/rs9080820

Bahramian, K., Nathan, R., Western, A.W., Ryu, D., 2021. Towards an ensemble-based short-term flood forecasting using an event-based flood model- incorporating catchment-average estimates of soil moisture. J. Hydrol. 593, 125828. https://doi.org/10.1016/j.jhydrol.2020.125828

Baratti, R., Cannas, B., Fanni, A., Pintus, M., Sechi, G.M., Toreno, N., 2003. River flow forecast for reservoir management through neural networks. Neurocomputing 55, 421–437. https://doi.org/10.1016/S0925-2312(03)00387-4

Baxter, M.A., Lackmann, G.M., Mahoney, K.M., Workoff, T.E., Hamill, T.M., 2014. Verification of Quantitative Precipitation Reforecasts over the Southeastern United States. Weather Forecast. 29, 1199–1207. https://doi.org/10.1175/WAF-D-14-00055.1

Beck, H.E., Pan, M., Miralles, D.G., Reichle, R.H., Dorigo, W.A., Hahn, S., Sheffield, J., Karthikeyan, L., Balsamo, G., Parinussa, R.M., van Dijk, A.I.J.M., Du, J., Kimball, J.S., Vergopolan, N., Wood, E.F., 2021. Evaluation of 18 satellite- And model-based soil moisture products using in situ measurements from 826 sensors. Hydrol. Earth Syst. Sci. 25, 17–40. https://doi.org/10.5194/hess-25-17-2021

Bellier, J., Bontron, G., Zin, I., 2017. Using Meteorological Analogues for Reordering Postprocessed Precipitation Ensembles in Hydrological Forecasting. Water Resour. Res. 53, 10085–10107. https://doi.org/10.1002/2017WR021245

Ben Bouallègue, Z., Theis, S.E., 2014. Spatial techniques applied to precipitation ensemble forecasts: From verification results to probabilistic products. Meteorol. Appl. 21, 922–929. https://doi.org/10.1002/met.1435

Berthet, L., Andréassian, V., Perrin, C., Javelle, P., 2009. How crucial is it to account for the antecedent moisture conditions in flood forecasting? Comparison of event-based and continuous approaches on 178 catchments. Hydrol. Earth Syst. Sci. 13, 819–831. https://doi.org/10.5194/hess-13-819-2009

Beven, K., 2006. Searching for the Holy Grail of scientific hydrology: Q t=(S, R, Δt) A as closure. Hydrol. Earth Syst. Sci. 10, 609–618.

Bhuiyan, M.A.E., Nikolopoulos, E.I., Anagnostou, E.N., Quintana-Seguí, P., Barella-Ortiz, A.,

2018. A nonparametric statistical technique for combining global precipitation datasets: Development and hydrological evaluation over the Iberian Peninsula. Hydrol. Earth Syst. Sci. 22, 1371–1389. https://doi.org/10.5194/hess-22-1371-2018

Bischiniotis, K., van den Hurk, B., Zsoter, E., Coughlan de Perez, E., Grillakis, M., Aerts, J.C.J.H., 2019. Evaluation of a global ensemble flood prediction system in Peru. Hydrol. Sci. J. 64, 1171–1189. https://doi.org/10.1080/02626667.2019.1617868

Booij, M.J., 2005. Impact of climate change on river flooding assessed with different spatial model resolutions. J. Hydrol. 303, 176–198. https://doi.org/10.1016/j.jhydrol.2004.07.013

Bournas, A., Baltas, E., 2021. Increasing the Efficiency of the Sacramento Model on Event Basis in a Mountainous River Basin. Environ. Process. 8, 943–958. https://doi.org/10.1007/s40710-021-00504-4

Bourqui, M., Loumagne, C., Chahinian, N., Plantier, M., 2006. Accounting for spatial variability: A way to improve lumped modelling approaches? An assessment on 3300 chimera catchments. IAHS-AISH Publ. 300–310.

Brath, A., Montanari, A., Toth, E., 2004. Analysis of the effects of different scenarios of historical data availability on the calibration of a spatially-distributed hydrological model. J. Hydrol. 291, 232–253. https://doi.org/10.1016/j.jhydrol.2003.12.044

Brocca, L., Melone, F., Moramarco, T., 2008. On the estimation of antecedent wetness conditions in rainfall-runoff modelling. Hydrol. Process. 22, 629–642. https://doi.org/10.1002/hyp.6629

Brocca, L., Melone, F., Moramarco, T., Singh, V.P., 2009. Assimilation of Observed Soil Moisture Data in Storm Rainfall-Runoff Modeling. J. Hydrol. Eng. 14, 153–165. https://doi.org/10.1061/(asce)1084-0699(2009)14:2(153)

Brocca, Luca, Melone, F., Moramarco, T., Wagner, W., Naeimi, V., Bartalis, Z., Hasenauer, S., 2010. Improving runoff prediction through the assimilation of the ASCAT soil moisture product. Hydrol. Earth Syst. Sci. 14, 1881–1893.

Brocca, L., Melone, F., Moramarco, T., Wagner, W., Naeimi, V., Bartalis, Z., Hasenauer, S., 2010. Improving runoff prediction through the assimilation of the ASCAT soil moisture product. Hydrol. Earth Syst. Sci. 14, 1881–1893. https://doi.org/10.5194/hess-14-1881-2010

Brown, J.D., Seo, D.-J., Du, J., 2012. Verification of precipitation forecasts from NCEP's short-range ensemble forecast (SREF) system with reference to ensemble streamflow prediction using lumped hydrologic models. J. Hydrometeorol. 13, 808–836. https://doi.org/10.1175/JHM-D-11-036.1

Brown, J.D., Wu, L., He, M., Regonda, S., Lee, H., Seo, D.J., 2014. Verification of temperature, precipitation, and streamflow forecasts from the NOAA/NWS Hydrologic Ensemble Forecast Service (HEFS): 1. Experimental design and forcing verification. J. Hydrol. 519, 2869–2889. https://doi.org/10.1016/j.jhydrol.2014.05.028

Butts, M.B., Payne, J.T., Kristensen, M., Madsen, H., 2004. An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. J. Hydrol. 298, 242–266. https://doi.org/10.1016/j.jhydrol.2004.03.042

Caldeira, T.L., Mello, C.R., Beskow, S., Timm, L.C., Viola, M.R., 2019. LASH hydrological model: An analysis focused on spatial discretization. Catena 173, 183–193. https://doi.org/10.1016/j.catena.2018.10.009

Casati, B., Wilson, L., Nurmi, P., Ghelli, A., 2008. Forecast Verification : Current Status and Future Directions. Meteorol. Appl. 15, 3–18. https://doi.org/10.1002/met.52

Chakraborty, P., Sarkar, A., Bhatla, R., Singh, R., 2021. Assessing the skill of NCMRWF global ensemble prediction system in predicting Indian summer monsoon during 2018. Atmos. Res. 248, 105255. https://doi.org/10.1016/j.atmosres.2020.105255

Christensen, O.B., Gaertner, M.A., Prego, J.A., Polcher, J., 2001. Internal variability of regional climate models. Clim. Dyn. 17, 875–887. https://doi.org/10.1007/s003820100154

Clark, M.P., Rupp, D.E., Woods, R.A., Zheng, X., Ibbitt, R.P., Slater, A.G., Schmidt, J., Uddstrom, M.J., 2008. Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model. Adv. Water Resour. 31, 1309–1324. https://doi.org/10.1016/j.advwatres.2008.06.005

Clark, M.P., Slater, A.G., Barrett, A.P., Hay, L.E., McCabe, G.J., Rajagopalan, B., Leavesley, G.H., 2006. Assimilation of snow covered area information into hydrologic and land-surface models. Adv. Water Resour. 29, 1209–1221.

Cloke, H. L., Pappenberger, F., 2009. Ensemble flood forecasting: A review. J. Hydrol. 375, 613–626. https://doi.org/10.1016/j.jhydrol.2009.06.005

Cloke, H L, Pappenberger, F., 2009. Ensemble flood forecasting: A review. J. Hydrol. 375, 613–626.

Cloke, H.L., Schaake, J.C., 2018. Handbook of Hydrometeorological Ensemble Forecasting, Handbook of Hydrometeorological Ensemble Forecasting. https://doi.org/10.1007/978-3-642-40457-3

Coron, L., Thirel, G., Delaigue, O., Perrin, C., Andréassian, V., 2017. The suite of lumped GR hydrological models in an R package. Environ. Model. Softw. 94, 166–171. https://doi.org/10.1016/j.envsoft.2017.05.002

Coustau, M., Bouvier, C., Borrell-Estupina, V., Jourde, H., 2012. Flood modelling with a distributed event-based parsimonious rainfall-runoff model: Case of the karstic Lez river catchment. Nat. Hazards Earth Syst. Sci. 12, 1119–1133. https://doi.org/10.5194/nhess-12-1119-2012

Crow, W.T., Reichle, R.H., 2008. Comparison of adaptive filtering techniques for land surface data assimilation. Water Resour. Res. 44. https://doi.org/10.1029/2008WR006883

Crow, W.T., Ryu, D., 2009. A new data assimilation approach for improving runoff prediction using remotely-sensed soil moisture retrievals. Hydrol. Earth Syst. Sci. 13, 1–16. https://doi.org/10.5194/hess-13-1-2009

Crow, W.T., Van Loon, E., 2006. Impact of incorrect model error assumptions on the sequential assimilation of remotely sensed surface soil moisture. J. Hydrometeorol. 7, 421–432. https://doi.org/10.1175/JHM499.1

CWC, 2018a. Flood Forecasting Appraisal Report Godavari River Basin-2018.

CWC, 2018b. Flood Forecasting and Warning Network Performance: Appraisal Report, Central Water Commission, New Delhi.

CWC, NRSC, 2014. WATERSHED ATLAS.

Das, J., Manikanta, V., Nikhil Teja, K., Umamahesh, N. V., 2022. Two decades of ensemble flood forecasting: a state-of-the-art on past developments, present applications and future opportunities. Hydrol. Sci. J. 67, 477–493. https://doi.org/10.1080/02626667.2021.2023157

Das, T., Bárdossy, A., Zehe, E., He, Y., 2008. Comparison of conceptual model performance using different representations of spatial variability. J. Hydrol. 356, 106–118. https://doi.org/10.1016/j.jhydrol.2008.04.008

Day, G.N., 1985. Extended streamflow forecasting using NWSRFS. J. Water Resour. Plan. Manag. 111, 157–170.

De Lavenne, A., Thirel, G., Andréassian, V., Perrin, C., Ramos, M.H., 2016. Spatial variability of the parameters of a semi-distributed hydrological model. IAHS-AISH Proc. Reports 373, 87–94. https://doi.org/10.5194/piahs-373-87-2016

Demirel, M.C., Booij, M.J., Hoekstra, A.Y., 2015. The skill of seasonal ensemble low-flow forecasts in the Moselle River for three different hydrological models. Hydrol. Earth Syst. Sci. 19, 275–291. https://doi.org/10.5194/hess-19-275-2015

Douinot, A., Roux, H., Garambois, P.A., Larnier, K., Labat, D., Dartus, D., 2016. Accounting for rainfall systematic spatial variability in flash flood forecasting. J. Hydrol. 541, 359–370. https://doi.org/10.1016/j.jhydrol.2015.08.024

Dube, A., Ashrit, R., Singh, H., Arora, K., Iyengar, G., Rajagopal, E.N., 2017. Evaluating the

performance of two global ensemble forecasting systems in predicting rainfall over India during the southwest monsoons. Meteorol. Appl. 24, 230–238. https://doi.org/10.1002/met.1621

Durai, V.R., Bhardwaj, R., 2014. Forecasting quantitative rainfall over India using multi-model ensemble technique. Meteorol Atmos Phys 126, 31–48. https://doi.org/10.1007/s00703-014-0334-4

Ekmekcioğlu, Ö., Koc, K., Özger, M., Işık, Z., 2022. Exploring the additional value of class imbalance distributions on interpretable flash flood susceptibility prediction in the Black Warrior River basin, Alabama, United States. J. Hydrol. 610. https://doi.org/10.1016/j.jhydrol.2022.127877

Emerton, R.E., Stephens, E.M., Pappenberger, F., Pagano, T.C., Weerts, A.H., Wood, A.W., Salamon, P., Brown, J.D., Hjerdt, N., Donnelly, C., Baugh, C.A., Cloke, H.L., 2016. Continental and global scale flood forecasting systems. Wiley Interdiscip. Rev. Water 3, 391–418. https://doi.org/10.1002/wat2.1137

Emmanuel, I., Andrieu, H., Leblois, E., Janey, N., Payrastre, O., 2015. Influence of rainfall spatial variability on rainfall-runoff modelling: Benefit of a simulation approach? J. Hydrol. 531, 337–348. https://doi.org/10.1016/j.jhydrol.2015.04.058

Epstein, E.S., 1969. Stochastic dynamic prediction. Tellus 21, 739–759.

Estacio, A.B.S., Costa, A.C., Souza Filho, F.A., Rocha, R.V., 2021. Uncertainty analysis in parameter regionalization for streamflow prediction in ungauged semi-arid catchments. Hydrol. Sci. J. 66, 1132–1150. https://doi.org/10.1080/02626667.2021.1913281

Euser, T., Hrachowitz, M., Winsemius, H.C., Savenije, H.H.G., 2015. The effect of forcing and landscape distribution on performance and consistency of model structures. Hydrol. Process. 29, 3727–3743.

Flügel, W., 1995. Delineating hydrological response units by geographical information system analyses for regional hydrological modelling using PRMS/MMS in the drainage basin of the River Bröl, Germany. Hydrol. Process. 9, 423–436.

Franchini, M., Wendling, J., Obled, C., Todini, E., 1996. Physical interpretation and sensitivity analysis of the TOPMODEL. J. Hydrol. 175, 293–338. https://doi.org/10.1016/S0022-1694(96)80015-1

Garambois, P.A., Larnier, K., Roux, H., Labat, D., Dartus, D., 2014. Analysis of flash flood-triggering rainfall for a process-oriented hydrological model. Atmos. Res. 137, 14–24. https://doi.org/10.1016/j.atmosres.2013.09.016

Garg, S., Mishra, V., 2019. Role of Extreme Precipitation and Initial Hydrologic Conditions on

Floods in Godavari River Basin, India. Water Resour. Res. 55, 9191–9210. https://doi.org/10.1029/2019WR025863

Ghimire, U., Agarwal, A., Shrestha, N.K., Daggupati, P., Srinivasan, G., Than, H.H., 2020. Applicability of lumped hydrological models in a data-constrained river basin of Asia. J. Hydrol. Eng. 25, 5020018.

Gill, P.G., Buchanan, P., 2014. An ensemble based turbulence forecasting system. Meteorol. Appl. 21, 12–19. https://doi.org/10.1002/met.1373

Gomez, M., Sharma, S., Reed, S., Mejia, A., 2019. Skill of ensemble flood inundation forecasts at short- to medium-range timescales. J. Hydrol. 568, 207–220. https://doi.org/10.1016/j.jhydrol.2018.10.063

González, V.I., Carkovic, A.B., Lobo, G.P., Flanagan, D.C., Bonilla, C.A., 2016. Spatial discretization of large watersheds and its influence on the estimation of hillslope sediment yield. Hydrol. Process. 30, 30–39.

Gudmundsson, L., Bremnes, J.B., Haugen, J.E., Engen-Skaugen, T., 2012. Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations – A comparison of methods. Hydrol. Earth Syst. Sci. 16, 3383–3390. https://doi.org/10.5194/hess-16-3383-2012

Gupta, H. V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J. Hydrol. 377, 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Gupta, H. V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M., Andréassian, V., 2014. Large-sample hydrology: A need to balance depth with breadth. Hydrol. Earth Syst. Sci. 18, 463–477. https://doi.org/10.5194/hess-18-463-2014

Haghnegahdar, A., Tolson, B.A., Craig, J.R., Paya, K.T., 2015. Assessing the performance of a semi-distributed hydrological model under various watershed discretization schemes. Hydrol. Process. 29, 4018–4031.

Hamill, T.M., 2012. Verification of TIGGE multimodel and ECMWF reforecast-calibrated probabilistic precipitation forecasts over the contiguous United States. Mon. Weather Rev. 140, 2232–2252. https://doi.org/10.1175/MWR-D-11-00220.1

Hamill, T.M., 2001. Interpretation of rank histograms for verifying ensemble forecasts. Mon. Weather Rev. 129, 550–560. https://doi.org/10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2

Hamill, T.M., Whitaker, J.S., Wei, X., 2004. Ensemble Reforecasting: Improving Medium-Range Forecast Skill Using Retrospective Forecasts. Mon. Weather Rev. 132, 1434–1447.

https://doi.org/10.1175/1520-0493(2004)132<1434:ERIMFS>2.0.CO;2

Han, J.-C., Huang, G.-H., Zhang, H., Li, Z., Li, Y.-P., 2014. Effects of watershed subdivision level on semi-distributed hydrological simulations: case study of the SLURP model applied to the Xiangxi River watershed, China. Hydrol. Sci. J. 59, 108–125.

Hapuarachchi, H.A.P., Wang, Q.J., Pagano, T.C., 2011. A review of advances in flash flood forecasting. Hydrol. Process. 25, 2771–2784. https://doi.org/10.1002/hyp.8040

Hegdahl, T.J., Engeland, K., Müller, M., Sillmann, J., 2020. An event-based approach to explore selected present and future atmospheric river–induced floods in Western Norway. J. Hydrometeorol. 21, 2003–2021. https://doi.org/10.1175/JHM-D-19-0071.1

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., Kanae, S., 2013. Global flood risk under climate change. Nat. Clim. Chang. 3, 816–821. https://doi.org/10.1038/nclimate1911

Huang, P., Li, Z., Chen, J., Li, Q., Yao, C., 2016. Event-based hydrological modeling for detecting dominant hydrological process and suitable model strategy for semi-arid catchments. J. Hydrol. 542, 292–303. https://doi.org/10.1016/j.jhydrol.2016.09.001

Jaafar, H.H., Ahmad, F.A., Beyrouthy, N. El, 2019. GCN250 , new global gridded curve numbers for hydrologic modeling and design. Sci. Data 1–9. https://doi.org/10.1038/s41597-019-0155-x

Jain, Sharad Kumar, Mani, P., Jain, Sanjay K., Prakash, P., Singh, V.P., Tullos, D., Kumar, S., Agarwal, S.P., Dimri, A.P., 2018. A Brief review of flood forecasting techniques and their applications. Int. J. River Basin Manag. 16, 329–344. https://doi.org/10.1080/15715124.2017.1411920

Jehn, F.U., Chamorro, A., Houska, T., Breuer, L., 2019. Trade-offs between parameter constraints and model realism: a case study. Sci. Rep. 9, 1–12. https://doi.org/10.1038/s41598-019-46963-6

Jung, D., Choi, Y.H., Kim, J.H., 2017. Multiobjective automatic parameter calibration of a hydrological model. Water (Switzerland) 9. https://doi.org/10.3390/w9030187

Kalantari, Z., Lyon, S.W., Jansson, P.E., Stolte, J., French, H.K., Folkeson, L., Sassner, M., 2015. Modeller subjectivity and calibration impacts on hydrological model applications: An event-based comparison for a road-adjacent catchment in south-east Norway. Sci. Total Environ. 502, 315–329. https://doi.org/10.1016/j.scitotenv.2014.09.030

Katwal, R., Li, J., Zhang, T., Hu, C., Rafique, M.A., Zheng, Y., 2021. Event-based and continous flood modeling in Zijinguan watershed, Northern China. Nat. Hazards 108, 733–753. https://doi.org/10.1007/s11069-021-04703-y

Khakbaz, B., Imam, B., Hsu, K., Sorooshian, S., 2012. From lumped to distributed via semi-distributed: Calibration strategies for semi-distributed hydrologic models. J. Hydrol. 418–419, 61–77. https://doi.org/10.1016/j.jhydrol.2009.02.021

Kim, S., Sadeghi, H., Limon, R.A., Saharia, M., Seo, D.J., Philpott, A., Bell, F., Brown, J., He, M., 2018. Assessing the skill of medium-range ensemble precipitation and streamflow forecasts from the Hydrologic Ensemble Forecast Service (HEFS) for the upper Trinity River basin in North Texas. J. Hydrometeorol. 19, 1467–1483. https://doi.org/10.1175/JHM-D-18-0027.1

Kouwen, N., Soulis, E.D., Pietroniro, A., Donald, J., Harrington, R.A., 1993. Grouped response units for distributed hydrologic modeling. J. Water Resour. Plan. Manag. 119, 289–305.

Krysanova, V., Müller-Wohlfeil, D.-I., Becker, A., 1998. Development and test of a spatially distributed hydrological/water quality model for mesoscale watersheds. Ecol. Modell. 106, 261–289.

Kulkarni, B.D., Nandargi, S., Mulye, S.S., 2010. Analysis of severe rainstorm characteristics of the Godavari basin in Peninsular India. J. Hydrometeorol. 11, 542–552. https://doi.org/10.1175/2009JHM1153.1

Kumar, R., Samaniego, L., Attinger, S., 2010. The effects of spatial discretization and model parameterization on the prediction of extreme runoff characteristics. J. Hydrol. 392, 54–69. https://doi.org/10.1016/j.jhydrol.2010.07.047

Leavesley, G.H., Lichty, R.W., Troutman, B.M., Saindon, L.G., 1983. Precipitation-runoff modeling system: User's manual. Water-resources Investig. Rep. 83, 207.

Leith, C.E., 1974. Theoretical skill of Monte Carlo forecasts. Mon. Weather Rev. 102, 409–418.

Leonardo, N.M., Colle, B.A., 2017. Verification of multimodel ensemble forecasts of North Atlantic tropical cyclones. Weather Forecast. 32, 2083–2101. https://doi.org/10.1175/WAF-D-17-0058.1

Liu, Y., Duan, Q., Zhao, L., Ye, A., Tao, Y., Miao, C., Mu, X., Schaake, J.C., 2012. Evaluating the predictive skill of post-processed NCEP GFS ensemble precipitation forecasts in China's Huai river basin. Hydrol. Process. 27, 57–74. https://doi.org/10.1002/hyp.9496

Lobligeois, F., Perrin, C., Tabary, P., Loumagne, C., 2014. When does higher spatial resolution rainfall information improve streamflow simulation ? An evaluation using 3620 flood events 575–594. https://doi.org/10.5194/hess-18-575-2014

Loizu, J., Massari, C., Álvarez-Mozos, J., Tarpanelli, A., Brocca, L., Casalí, J., 2018. On the assimilation set-up of ASCAT soil moisture data for improving streamflow catchment simulation. Adv. Water Resour. 111, 86–104.

https://doi.org/10.1016/j.advwatres.2017.10.034

Longobardi, a, Villani, P., 2003. On the relationship between runoff coefficient and catchment initial conditions. Model. Simul. Soc. Aust. New Zeal. 2, 1–6.

Lorenz, E.N., 1963. Deterministic nonperiodic flow. J. Atmos. Sci. 20, 130–141.

Madadgar, S., Moradkhani, H., Garen, D., 2014. Towards improved post-processing of hydrologic forecast ensembles. Hydrol. Process. 28, 104–122. https://doi.org/10.1002/hyp.9562

Madsen, H., 2000. Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. J. Hydrol. 235, 276–288. https://doi.org/10.1016/S0022-1694(00)00279-1

Mahdi El Khalki, E., Tramblay, Y., Massari, C., Brocca, L., Simonneaux, V., Gascoin, S., El Mehdi Saidi, M., 2020. Challenges in flood modeling over data-scarce regions: How to exploit globally available soil moisture products to estimate antecedent soil wetness conditions in Morocco. Nat. Hazards Earth Syst. Sci. 20, 2591–2607. https://doi.org/10.5194/nhess-20-2591-2020

Manguerra, H.B., Engel, B.A., 1998. HYDROLOGIC PARAMETERIZATION OF WATERSHEDS FOR RUNOFF PREDICTION USING SWAT 1. JAWRA J. Am. Water Resour. Assoc. 34, 1149–1162.

Manikanta, V., Vema, V.K., 2022. Formulation of Wavelet Based Multi-Scale Multi-Objective Performance Evaluation (WMMPE) Metric for Improved Calibration of Hydrological Models. Water Resour. Res. 58. https://doi.org/10.1029/2020WR029355

Manohar Reddy, V., Ray, L.K., 2023. Development of Machine Learning Based Flood Prediction Model for Godavari River Basin BT - River Dynamics and Flood Hazards: Studies on Risk and Mitigation, in: Pandey, M., Azamathulla, H., Pu, J.H. (Eds.), . Springer Nature Singapore, Singapore, pp. 363–383. https://doi.org/10.1007/978-981-19-7100-6_20

Markstrom, S.L., Niswonger, R.G., Regan, R.S., Prudic, D.E., Barlow, P.M., 2008. GSFLOW-Coupled Ground-water and Surface-water FLOW model based on the integration of the Precipitation-Runoff Modeling System (PRMS) and the Modular Ground-Water Flow Model (MODFLOW-2005). US Geol. Surv. Tech. methods 6, 240.

McLaughlin, D., 2002. An integrated approach to hydrologic data assimilation: Interpolation, smoothing, and filtering. Adv. Water Resour. 25, 1275–1286. https://doi.org/10.1016/S0309-1708(02)00055-6

McMillan, H.K., Hreinsson, E.O., Clark, M.P., Singh, S.K., Zammit, C., Uddstrom, M.J., 2013. Operational hydrological data assimilation with the recursive ensemble Kalman filter. Hydrol. Earth Syst. Sci. 17, 21–38. https://doi.org/10.5194/hess-17-21-2013

Medina, H., Tian, D., Marin, F.R., Chirico, G.B., 2019. Comparing GEFS, ECMWF, and postprocessing methods for ensemble precipitation forecasts over Brazil. J. Hydrometeorol. 20, 773–790. https://doi.org/10.1175/JHM-D-18-0125.1

Meinshausen, N., 2017. Package 'quantregForest.' Quantile Regres. For. Packag e version 1.3–7).

Meinshausen, N., 2006. Quantile Regression Forests. J. Mach. Learn. Res. 7, 983–989.

Melsen, L., Teuling, A., Torfs, P., Zappa, M., Mizukami, N., Clark, M., Uijlenhoet, R., 2016. Representation of spatial and temporal variability in large-domain hydrological models: Case study for a mesoscale pre-Alpine basin. Hydrol. Earth Syst. Sci. 20, 2207–2226. https://doi.org/10.5194/hess-20-2207-2016

Mendoza, P.A., McPhee, J., Vargas, X., 2012. Uncertainty in flood forecasting: A distributed modeling approach in a sparse data catchment. Water Resour. Res. 48.

Meng, S., Xie, X., Liang, S., 2017. Assimilation of soil moisture and streamflow observations to improve flood forecasting with considering runoff routing lags. J. Hydrol. 550, 568–579. https://doi.org/10.1016/j.jhydrol.2017.05.024

Merz, R., Parajka, J., Blöschl, G., 2009. Scale effects in conceptual hydrological modeling. Water Resour. Res. 45, 1–15. https://doi.org/10.1029/2009WR007872

Mizukami, N., Rakovec, O., Newman, A.J., Clark, M.P., Wood, A.W., Gupta, H. V, Kumar, R., 2019. On the choice of calibration metrics for "high-flow" estimation using hydrologic models.

Molnar, D.K., Julien, P.Y., 2000. Grid-size effects on surface runoff modeling. J. Hydrol. Eng. 5, 8–16.

Moore, B.J., Mahoney, K.M., Sukovich, E.M., Cifelli, R., Hamill, T.M., 2015. Climatology and Environmental Characteristics of Extreme Precipitation Events in the Southeastern United States. Mon. Weather Rev. 143, 718–741. https://doi.org/10.1175/MWR-D-14-00065.1

Mullen, S.L., Buizza, R., 2002. The impact of horizontal resolution and ensemble size on probabilistic forecasts of precipitation by the ECMWF ensemble prediction system. Weather Forecast. 17, 173–191. https://doi.org/10.1175/1520-0434(2002)017<0173:TIOHRA>2.0.CO;2

Murphy, A.H., Winkler, R.L., 1987. A General Framework for Forecast Verification. Mon. Weather Rev. 115, 1330–1338. https://doi.org/10.1175/1520-0493(1987)115<1330:AGFFFV>2.0.CO;2

Nanditha, J.S., Mishra, V., 2021. On the need of ensemble flood forecast in India. Water Secur. 12, 100086. https://doi.org/10.1016/j.wasec.2021.100086

Nash, J.E., Sutcliffe, J. V, 1970. River flow forecasting through conceptual models part I — A discussion of principles. J. Hydrol. 10, 282–290. https://doi.org/http://dx.doi.org/10.1016/0022-1694(70)90255-6

National Disaster Management Plan, 2019. A publication of the National Disaster Management Authority, Government of India. November 2019, New Delhi.

Nayak, A.K., Biswal, B., Sudheer, K.P., 2021. Role of hydrological model structure in the assimilation of soil moisture for streamflow prediction. J. Hydrol. 598, 126465. https://doi.org/10.1016/j.jhydrol.2021.126465

Novak, D.R., Bailey, C., Brill, K.F., Burke, P., Hogsett, W.A., Rausch, R., Schichtel, M., 2014. Precipitation and Temperature Forecast Performance at the Weather Prediction Center. Weather Forecast. 29, 489–504. https://doi.org/10.1175/WAF-D-13-00066.1

Ogarekpe, N.M., Nnaji, C.C., Antigha, R.E.-E., 2022. A preliminary case for modification of the SCS-CN hydrologic model for runoff prediction in Imo River sub-basin. Arab. J. Geosci. 15. https://doi.org/10.1007/s12517-022-09995-3

Oudin, L., Andréassian, V., Mathevet, T., Perrin, C., Michel, C., 2006. Dynamic averaging of rainfall-runoff model simulations from complementary model parameterizations. Water Resour. Res. 42, 1–10. https://doi.org/10.1029/2005WR004636

Pai, D.S., Sridhar, L., Rajeevan, M., Sreejith, O.P., Satbhai, N.S., Mukhopadyay, B., 2014. Development of a new high spatial resolution (0.25° × 0.25°) Long Period (1901-2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region. Mausam 65, 1–18.

Palmer, T.N., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G.J., Steinheimer, M., Weisheimer, A., 2009. Stochastic parametrization and model uncertainty.

Park, Y.-Y., Buizza, R., Leutbecher, M., 2008. TIGGE: Preliminary results on comparing and combining ensembles. Q. J. R. Meteorol. Soc. 134, 2029–2050. https://doi.org/https://doi.org/10.1002/qj.334

Pathiraja, S., Westra, S., Sharma, A., 2012. Why continuous simulation? the role of antecedent moisture in design flood estimation. Water Resour. Res. 48, 1–15. https://doi.org/10.1029/2011WR010997

Perrin, C., Michel, C., Andréassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. J. Hydrol. 279, 275–289.

Piazzi, G., Thirel, G., Perrin, C., Delaigue, O., 2021. Sequential Data Assimilation for Streamflow Forecasting: Assessing the Sensitivity to Uncertainties and Updated Variables of a Conceptual Hydrological Model at Basin Scale. Water Resour. Res. 57.

https://doi.org/10.1029/2020WR028390

Pilz, T., Francke, T., Bronstert, A., 2017. LumpR 2.0.0: An R package facilitating landscape discretisation for hillslope-based hydrological models. Geosci. Model Dev. 10, 3001–3023. https://doi.org/10.5194/gmd-10-3001-2017

Pushpalatha, R., Perrin, C., Moine, N. Le, Andréassian, V., 2012. A review of efficiency criteria suitable for evaluating low-flow simulations. J. Hydrol. 420–421, 171–182. https://doi.org/10.1016/j.jhydrol.2011.11.055

Raftery, A.E., Gneiting, T., Balabdaoui, F., Polakowski, M., 2005. Using Bayesian model averaging to calibrate forecast ensembles. Mon. Weather Rev. 133, 1155–1174.

Rahimi, L., Deidda, C., De Michele, C., 2021. Origin and variability of statistical dependencies between peak, volume, and duration of rainfall-driven flood events. Sci. Rep. 11. https://doi.org/10.1038/s41598-021-84664-1

Raju, K.S., Sonali, P., Kumar, D.N., 2017. Ranking of CMIP5-based global climate models for India using compromise programming. Theor. Appl. Climatol. 128, 563–574. https://doi.org/10.1007/s00704-015-1721-6

Rakhecha, P., Singh, V., 2017. Enveloping Curves for the Highest Floods of River basins in India. Int. J. Hydrol. 1, 79–84. https://doi.org/10.15406/ijh.2017.01.00015

Ran, Q., Fu, W., Liu, Y., Li, T., Shi, K., Sivakumar, B., 2018. Evaluation of Quantitative Precipitation Predictions by ECMWF, CMA, and UKMO for Flood Forecasting: Application to Two Basins in China. Nat. Hazards Rev. 19. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000282

Ratri, D.N., Whan, K., Schmeits, M., 2019. A comparative verification of raw and bias-corrected ECMWF seasonal ensemble precipitation reforecasts in Java (Indonesia). J. Appl. Meteorol. Climatol. 58, 1709–1723. https://doi.org/10.1175/JAMC-D-18-0210.1

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.-J., DMIP Participants, and, 2004. Overall distributed model intercomparison project results. J. Hydrol. 298, 27–60. https://doi.org/10.1016/j.jhydrol.2004.03.031

Scrucca, L., 2013. GA: A package for genetic algorithms in R. J. Stat. Softw. 53, 1–37. https://doi.org/10.18637/jss.v053.i04

Seo, D.-J., Koren, V., Cajina, N., 2003. Real-time variational assimilation of hydrologic and hydrometeorological data into operational hydrologic forecasting. J. Hydrometeorol. 4, 627–641.

Sharma, S., Siddique, R., Balderas, N., Fuentes, J.D., Reed, S., Ahnert, P., Shedd, R., Astifan, B., Cabrera, R., Laing, A., Klein, M., Mejia, A., 2017. Eastern U.S. verification of

ensemble precipitation forecasts. Weather Forecast. 32, 117–139. https://doi.org/10.1175/WAF-D-16-0094.1

Sharma, V.C., Regonda, S.K., 2021. Multi-spatial resolution rainfall-runoff modelling—a case study of sabari river basin, india. Water (Switzerland) 13. https://doi.org/10.3390/w13091224

Siddique, Ridwan, Mejia, A., Brown, J., Reed, S., Ahnert, P., 2015. Verification of precipitation forecasts from two numerical weather prediction models in the Middle Atlantic Region of the USA: A precursory analysis to hydrologic forecasting. J. Hydrol. 529, 1390–1406. https://doi.org/10.1016/j.jhydrol.2015.08.042

Siddique, R, Mejia, A., Brown, J., Reed, S., Ahnert, P., 2015. Verification of precipitation forecasts from two numerical weather prediction models in the Middle Atlantic Region of the USA: A precursory analysis to hydrologic forecasting. J. Hydrol. 529, 1390–1406. https://doi.org/10.1016/j.jhydrol.2015.08.042

Singh, O., Kumar, M., 2017. Flood occurrences, damages, and management challenges in India: a geographical perspective. Arab. J. Geosci. 10. https://doi.org/10.1007/s12517-017-2895-2

Siqueira, V.A., Fan, F.M., Paiva, R.C.D. de, Ramos, M.H., Collischonn, W., 2020. Potential skill of continental-scale, medium-range ensemble streamflow forecasts for flood prediction in South America. J. Hydrol. 590, 125430. https://doi.org/10.1016/j.jhydrol.2020.125430

Srivastava, A.K., Rajeevan, M., Kshirsagar, S.R., 2009. Development of a high resolution daily gridded temperature data set (1969-2005) for the Indian region. Atmos. Sci. Lett. n/a-n/a. https://doi.org/10.1002/asl.232

Sulis, M., Paniconi, C., Camporese, M., 2011. Impact of grid resolution on the integrated and distributed response of a coupled surface–subsurface hydrological model for the des Anglais catchment, Quebec. Hydrol. Process. 25, 1853–1865.

Sun, L., Seidou, O., Nistor, I., Liu, K., 2016. Review of the Kalman-type hydrological data assimilation. Hydrol. Sci. J. 61, 2348–2366. https://doi.org/10.1080/02626667.2015.1127376

Taillardat, M., Mestre, O., Zamo, M., Naveau, P., 2016. Calibrated ensemble forecasts using quantile regression forests and ensemble model output statistics. Mon. Weather Rev. 144, 2375–2393. https://doi.org/10.1175/MWR-D-15-0260.1

Teja, K.N., Umamahesh, N. V, 2020. Application of Ensemble Techniques for Flood Forecasting in India. Roorkee Water Conclave.

Thiemig, V., Bisselink, B., Pappenberger, F., Thielen, J., 2015. A pan-African medium-range ensemble flood forecast system. Hydrol. Earth Syst. Sci. 19, 3365–3385.

Thirel, G., Martin, E., Mahfouf, J.-F., Massart, S., Ricci, S., Regimbeau, F., Habets, F., 2010. A past discharge assimilation system for ensemble streamflow forecasts over France - Part 2: Impact on the ensemble streamflow forecasts. Hydrol. Earth Syst. Sci. 14, 1639–1653. https://doi.org/10.5194/hess-14-1639-2010

Tramblay, Y., Bouaicha, R., Brocca, L., Dorigo, W., Bouvier, C., Camici, S., Servat, E., 2012. Estimation of antecedent wetness conditions for flood modelling in northern Morocco. Hydrol. Earth Syst. Sci. 16, 4375–4386. https://doi.org/10.5194/hess-16-4375-2012

Tramblay, Y., Bouvier, C., Martin, C., Didon-Lescot, J.F., Todorovik, D., Domergue, J.M., 2010. Assessment of initial soil moisture conditions for event-based rainfall-runoff modelling. J. Hydrol. 387, 176–187. https://doi.org/10.1016/j.jhydrol.2010.04.006

Tran, Q.Q., De Niel, J., Willems, P., 2018. Spatially Distributed Conceptual Hydrological Model Building: A Generic Top-Down Approach Starting From Lumped Models. Water Resour. Res. 54, 8064–8085. https://doi.org/10.1029/2018WR023566

Tugrul Yilmaz, M., Crow, W.T., 2013. The optimality of potential rescaling approaches in land data assimilation. J. Hydrometeorol. 14, 650–660. https://doi.org/10.1175/JHM-D-12-052.1

Vafaei, N., Ribeiro, R.A., Camarinha-Matos, L.M., 2016. Normalization techniques for multi-criteria decision making: analytical hierarchy process case study, in: Doctoral Conference on Computing, Electrical and Industrial Systems. Springer, pp. 261–269.

Viney, N.R., Croke, B.F.W., Breuer, L., Bormann, H., Bronstert, A., Frede, H., Gräff, T., Hubrechts, L., Huisman, J.A., Jakeman, A.J., 2005. Ensemble modelling of the hydrological impacts of land use change, in: International Congress on Modelling and Simulation: Advances and Applications for Management and Decision Making, MODSIM05. pp. 2967–2973.

Vu Van Nghi, Ha Bui Nguyen Lam, Tai Pham Anh, Can Thu Van, 2020. Development and Application of a Distributed Conceptual Hydrological Model to Simulate Runoff in the Be River Basin and the Water Transfer Capacity to the Saigon River Basin, Vietnam. J. Environ. Sci. Eng. A 9, 12. https://doi.org/10.17265/2162-5298/2020.01.001

Wanders, N., Karssenberg, D., De Roo, A., De Jong, S.M., Bierkens, M.F.P., 2014. The suitability of remotely sensed soil moisture for improving operational flood forecasting. Hydrol. Earth Syst. Sci. 18, 2343–2357. https://doi.org/10.5194/hess-18-2343-2014

Wang, H., Sankarasubramanian, A., Ranjithan, R.S., 2013. Integration of climate and weather information for improving 15-day-ahead accumulated precipitation forecasts. J. Hydrometeorol. 14, 186–202. https://doi.org/10.1175/JHM-D-11-0128.1

Wang, Q.J., Robertson, D.E., Chiew, F.H.S., 2009. A Bayesian joint probability modeling

approach for seasonal forecasting of streamflows at multiple sites. Water Resour. Res. 45.

Whitaker, J.S., Louche, A.F., 1998. The relationship between ensemble spread and ensemble mean skill. Mon. Weather Rev. 126, 3292–3302. https://doi.org/10.1175/1520-0493(1998)126<3292:TRBESA>2.0.CO;2

Wilks, D.S., 2011. Statistical Methods in the Atmospheric Sciences, 3rd ed. Elsevier, San Diego.

Wilks, D.S., 2006. Comparison of ensemble-MOS methods in the Lorenz '96 setting. Meteorol. Appl. 13, 243–256. https://doi.org/10.1017/S1350482706002192

Wood, E.F., Sivapalan, M., Beven, K., Band, L., 1988. Effects of spatial variability and scale with implications to hydrologic modeling. J. Hydrol. 102, 29–47. https://doi.org/10.1016/0022-1694(88)90090-X

Wu, W., Emerton, R., Duan, Q., Wood, A.W., Wetterhall, F., Robertson, D.E., 2020a. Ensemble flood forecasting: Current status and future opportunities. WIREs Water 7, 1–32. https://doi.org/10.1002/wat2.1432

Wu, W., Emerton, R., Duan, Q., Wood, A.W., Wetterhall, F., Robertson, D.E., 2020b. Ensemble flood forecasting: Current status and future opportunities. Wiley Interdiscip. Rev. Water e1432. https://doi.org/10.1002/wat2.1432

Yao, C., Ye, J., He, Z., Bastola, S., Zhang, K., Li, Z., 2019. Evaluation of flood prediction capability of the distributed Grid-Xinanjiang model driven by weather research and forecasting precipitation. J. Flood Risk Manag. 12. https://doi.org/10.1111/jfr3.12544

Yatheendradas, S., Wagener, T., Gupta, H., Unkrich, C., Goodrich, D., Schaffner, M., Stewart, A., 2008. Understanding uncertainty in distributed flash flood forecasting for semiarid regions. Water Resour. Res. 44. https://doi.org/10.1029/2007WR005940

Yilmaz, K.K., Gupta, H. V, Wagener, T., 2008. A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. Water Resour. Res. 44. https://doi.org/https://doi.org/10.1029/2007WR006716

Young, P.C., 2002. Advances in real–time flood forecasting. Philos. Trans. R. Soc. London. Ser. A Math. Phys. Eng. Sci. 360, 1433–1450. https://doi.org/10.1098/rsta.2002.1008

Zamo, M., Naveau, P., 2018. Estimation of the Continuous Ranked Probability Score with Limited Information and Applications to Ensemble Weather Forecasts. Math. Geosci. 50, 209–234. https://doi.org/10.1007/s11004-017-9709-7

Zeng, L., Xiong, L., Liu, D., Chen, J., Kim, J.S., 2019. Improving parameter transferability of GR4J model under changing environments considering nonstationarity. Water (Switzerland) 11. https://doi.org/10.3390/w11102029

Zhou, Y., Liang, Z., Li, B., Huang, Y., Wang, K., Hu, Y., 2021. Seamless Integration of Rainfall

Spatial Variability and a Conceptual Hydrological Model. Sustainability 13, 3588. https://doi.org/10.3390/su13063588

Zhu, Y., Toth, Z., 2008. 2.2 Ensemble Based Probabilistic Forecast Verification.

Zhuo, L., Han, D., 2016. Could operational hydrological models be made compatible with satellite soil moisture observations? Hydrol. Process. 30, 1637–1648. https://doi.org/10.1002/hyp.10804

Zoccatelli, D., Borga, M., Viglione, A., Chirico, G.B., Blöschl, G., 2011. Spatial moments of catchment rainfall: Rainfall spatial organisation, basin morphology, and flood response. Hydrol. Earth Syst. Sci. 15, 3767–3783. https://doi.org/10.5194/hess-15-3767-2011

Research papers resulting from the thesis

1. Das, J., **Manikanta, V.**, Nikhil Teja, K., & Umamahesh, N. V. (2022). Two decades of ensemble flood forecasting: a state-of-the-art on past developments, present applications and future opportunities. *Hydrological Sciences Journal*, 67(3), 477-493. (Published)
2. **Manikanta, V.**, Nikhil Teja, K., Das, J., & Umamahesh, N. V. (2023). On the verification of ensemble precipitation forecasts over the Godavari River basin. *Journal of Hydrology*, 616, 128794. https://doi.org/10.1016/j.jhydrol.2022.128794 (Published)
3. **Manikanta, V.** & Umamahesh, N. V. (2023). Influence of Nested Landscape Discretization on an Event-based Conceptual Hydrological Model in simulating Flood Hydrographs. (Submitted to *Computational Geosciences*, Manuscript ID COMG-D-23-00071)
4. **Manikanta, V.** & Umamahesh, N. V. (2023). Performance assessment of methods to estimate initial hydrologic conditions for event-based rainfall-runoff modelling. (Accepted in *Journal of Water and Climate Change*)

Acknowledgements

I would like to express my deepest gratitude to my supervisor, **Prof. N.V. Umamahesh,** for his invaluable guidance and support throughout the journey of completing this thesis. His expertise, encouragement, and unwavering commitment to academic excellence have been instrumental in shaping my research and academic growth. I am grateful for his mentorship, patience, and insightful feedback, which have contributed immensely to the quality and

rigour of this work. I am also grateful for the academic freedom given by him to choose and pursue my research. It has been an honour to be his PhD student.

I would like to extend my heartfelt gratitude to three individuals who have been instrumental during my PhD journey and have played a significant role in the completion of my thesis. First and foremost, I am immensely grateful to **Dr Titas Ganguly** for his invaluable mentorship and unwavering support throughout my research. His expertise inculcated in me the necessary research capabilities, and his constant encouragement and belief in my abilities have been a source of motivation. I would also like to express my deepest appreciation to **Dr Jew Das** for his unwavering support, motivation, and relentless push towards excellence. His passion for research, dedication, and commitment to academic rigour have inspired me to work diligently and consistently. I am grateful to Dr Jew Das for teaching me the art of writing manuscripts and helping me refine my research skills. His constructive feedback and guidance have played a crucial role in shaping my thesis. Furthermore, I would like to acknowledge **Dr Vamsi Krishna Vema** for his invaluable contribution to my understanding of hydrological modelling. His expertise, patience, and willingness to address my doubts and queries have been immensely helpful in navigating the complexities of hydrological modelling, a key aspect of my thesis.

I am immensely grateful to Prof. K. Venkata Reddy, Dr Vamsi Krishna Vema, and Prof. Debasish Dutta for their unwavering encouragement and invaluable guidance during the progress review in each semester of my PhD. I would also like to extend my thanks to all the faculty members of the Water and Environmental Division for their continuous support and assistance.

I would like to extend a special thanks to **Dr Manish Pandey** and **Dr Sanjit Biswas** for their support and encouragement throughout my PhD journey. I am also grateful to them for fostering a healthy balance between academic pursuits and my passion for cricket, allowing me to find joy and relaxation outside of my studies. Their support in both my academic and personal pursuits is deeply appreciated.

My special thanks to **Shri. V.N. Kameswara Rao** for his profound influence on my academic journey, despite not having the chance to interact with him directly. As his "Ekalavya shishya," I have been deeply inspired by his teaching approach and his insightful views on research.

I would like to express my deepest thanks to my parents, **Mr Venkateswarlu** and **Mrs Anjali Devi**. Although I lost my father seven years ago, his unwavering support and encouragement to pursue higher studies have always remained with me. I strongly believe that he would have been the happiest person to know that I am on the verge of receiving a doctoral degree. I am grateful for the opportunity to fulfil his wish and make him proud. On the other hand, my mother has been my pillar of strength throughout this entire journey. Her constant love, support, and sacrifices have been invaluable, and I cannot imagine having reached this point without her unwavering belief in me. Her guidance and encouragement have been my driving force, and I am truly indebted to her for everything she has done for me. I am blessed to have such wonderful parents who have shaped me into the person I am today, and I am forever grateful for their love and support. In addition, my sister and brother-in-law have always helped me throughout this journey.

I would like to express a special thanks to **Mr Malla Mani Kanta** for his instrumental role in my academic journey. When no one else believed in my capabilities, and I was uncertain about pursuing my PhD career, he recommended me for a JRF position at IIT Roorkee. This opportunity turned out to be the turning point in my career, providing me with the platform and support to embark on my PhD journey. The skills and knowledge I acquired at IIT Roorkee have been invaluable in my research, and without this opportunity, my PhD journey would not have been possible. I am deeply grateful to my friend Mani for his unwavering belief in me, his recommendation that opened doors for me, and the profound impact he has had on my academic and professional growth.

I am immensely grateful to my friends Sai Kumar, Nikhil, Rudra, Srikanth, Manohar, Aravind, Satish, Sudhir, Hari, Sai Teja Jollu, and Soumya for their invaluable companionship, engaging research discussions, memorable coffee breaks, enjoyable cricket matches, captivating cinema outings, and all the delightful moments of fun and celebration. Your presence and friendship have truly made my time at NIT Warangal incredibly beautiful.

Saving the best for last, I would like to express my heartfelt gratitude to my wife, **Shami**. Her unwavering support and understanding have been the pillars that have held me up throughout this challenging journey. I am deeply grateful for her patience, as I know that I have spent countless hours immersed in my research, leaving little time for our family.

Despite the sacrifices she made and the limited time we had together, she has always been there, offering her unwavering support and encouragement. I am truly blessed to have such a strong and supportive partner by my side. I am deeply thankful to my sons, Parth and Charith, for being my stress-busters during my PhD. Their playful presence and infectious laughter provided me solace and motivation, reminding me of the importance of family amidst the academic rigours.

_____

Velpuri Manikanta

Date: