



INDIAN NATIONAL COMMITTEE ON SURFACE WATER (INCSW- CWC)

UID	23/68/2012-R&D/427-437
Type (State whether final or draft report)	FINAL REPORT
Name of R&D Scheme	"ENSEMBLE MODELLING OF RAINFALL RUNOFF TRANSFORMATION PROCESS"
Name of PI & Co-PI	PI - Prof. RAJENDRA SINGH Co-PI - Prof. CHANDRANATH CHATTERJEE & Dr. ASHOK MISHRA
Institute Address	AGRICULTURAL AND FOOD ENGINEERING DEPARTMENT, INDIAN INSTITUTE OF TECHNOLOGY, KHARAGPUR – 721 302
Circulation(State whether Open for public or not)	PUBLIC
Month & Year of Report Submission	FEBRUARY, 2018

©INCSW Sectt.

Central Water Commission
Room No. 2, Wing 4, 1st Floor

West Block I, R.K.Puram, New Delhi-110066

E-Mail: incsw-cwc@nic.in

1. Title of the scheme: **Ensemble Modelling of Rainfall Runoff Transformation process**

2. Name and addresses of the PI and other investigators

Principal Investigator	Singh Rajendra Professor, AgFE Department, IIT Kharagpur
Co-Investigator	Chatterjee Chandranath Professor, AgFE Department, IIT Kharagpur
Co-Investigator	Mishra Ashok Associate Professor, AgFE Department, IIT Kharagpur

3. Name and address of the Institute

Indian Institute of Technology Kharagpur, Kharagpur

4. Financial details (Sanctioned cost; amount released; expenditure; unspent balance (if any) and return of unspent balance)

Please refer Annexure I

5. Utilisation Certificate

Please refer Annexure II

6. Statement of equipment purchased under the scheme

Please refer Annexure III

7. Original objectives and methodology as in the sanctioned proposal.

- i) **To develop multi-model ensembles for studying the rainfall-runoff transformation process.**
- ii) **To analyse the performance of the developed multi-model ensembles.**
- iii) **To determine the optimal size and optimal combination of models for the ensemble modelling of the rainfall-runoff transformation process.**
- iv) **To evaluate the reliability of the simulated outputs obtained through multi-model ensembles.**

8. Any changes in the objectives during the operation of the scheme.

Uncertainty assessment was performed rather than reliability assessment of simulated outputs in the last objective.

9. All data collected and used in the analysis with sources of data.

Please refer Annexure IV

10. Methodology actually followed. (observations, analysis, results and inferences)

Please refer Annexure V for detailed methodology, analysis, results & discussion

11. Conclusions/ Recommendations

Please refer Annexure VI for conclusions

12. How do the conclusions/recommendations compare with current thinking

Ensemble hydrological modelling is attempted for the first time in India, aimed at improved simulation and prediction of river discharges. Ensemble of hydrological models reduces uncertainty and provides better results in terms of real valued scores as compared to a single hydrological model. The application of ensembles towards uncertainty reduction and findings are in sync with the current trend in the hydrological modelling fraternity.

13. Field tests conducted: **No**

14. Software generated, if any: **Nil**

15. Possibilities of any patents/copyrights. If so, then action taken in this regard: **No. The following publications, however, have emerged from this study:**

Journal

Kumar, A., Singh, R., Jena, P.P., Chatterjee, C. and Mishra, A. (2015). Identification of the best multi-model combination for simulating river discharge. Journal of Hydrology, 525, 313–325.

Conference

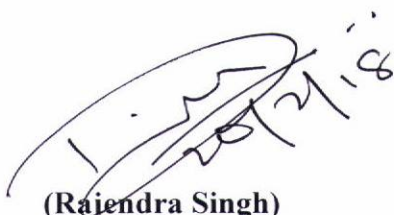
- (a) Kumar, A., Singh, R., Chatterjee, C., and Mishra, A. (2014). “Improving accuracy in modeling river discharge using Ensemble”, Proc. Of National Conference on Emerging**

Technology Trends in Agricultural Engineering, Excel India Publishers, New Delhi held at NERIST, Itanagar, India 7-9 November, 2014, pp. 437-445.

- (b) Kumar, A., Singh, R. and Chatterjee, C. (2016). Uncertainty Comparison in Simulating River Discharge by Hydrological models and its Ensemble, International Conference on Climate Change, Water, Agriculture and Food Security (ICCCWAFS), ICRISAT, Hyderabad, India, November 2-3, 2016.
- (c) Kumar, A., Singh, R., Mishra, A. and Chatterjee, C. (2015). Assessment of Climate Change Impact on River Discharge using Reduced Uncertainty Ensemble Modeling Framework. American Geophysical Union, Fall Meeting 2015, abstract #GC23C-1155.

16. Suggestions for further work

- a. Multi-model ensembles may be developed based on water balance.
- b. Impact assessment of climate change may be attempted using multi-model ensemble.
- c. Contribution to overall prediction uncertainty due to hydrological models and climate models may be assessed.



(Rajendra Singh)

Principal Investigator

Rajendra Singh
Brahmaputra Chair Professor for Water Resources
Professor
Agricultural & Food Engineering Department
Indian Institute of Technology
Kharagpur - 721302, India



21/2/18

Head of the Institute,
IIT Kharagpur

Dean
Sponsored Research &
Industrial Consultancy
I.I.T. Kharagpur-721302

Prof. Dr. Ramesh Chandra
Indian Institute of Technology
Department of Agricultural & Food Engineering
Professor
Brahmapur Chair Professor for Water Resources
Rajendra Singh

Annexure I

**SPONSORED RESEARCH AND INDUSTRIAL CONSULTANCY
INDIAN INSTITUTE OF TECHNOLOGY, KHARAGPUR**

STATEMENT OF RECEIPTS AND PAYMENTS OF RESEARCH SCHEME/PROJECT FOR THE
PERIOD 01/04/2014 TO 31/03/2015

NAME OF THE PROJECT **"Ensemble Modeling of Rainfall-Runoff transformation Process (ERT)"**

SPONSORING AGENCY Ministry of Water Resources

DEPARTMENT Agricultural & Food Engineering

INVESTIGATOR-IN-CHARGE Prof. Rajendra Singh

Amount in ₹

RECEIPTS	Amount Sanctioned	Amount Received	Amount Spent
By, Salary / Fellowship	672000	0	0
By, Travel Expenditure	155000	0	0
By, Infrastructure/Equipment	988000	0	0
By, Experimental charges	300000	0	0
Sub Total	2115000	0	0
By, Contingency @ 5 %	105750	0	27998
Total	105750	0	27998
By, Institute Overhead	423000	0	0
GRAND TOTAL	2643750	0	27998
Closing Balance as on date			-7402

CERTIFIED THAT THE ABOVE MONEY HAVE BEEN UTILIZED FOR THE PURPOSE
FOR WHICH THE SAME WAS SANCTIONED



Principal Investigator



Deputy Registrar (SRIC)

Sumit Kumar Biswas
Deputy Registrar
Sponsored Research & Industrial Consultancy
I.I.T. Kharagpur-721302

**SPONSORED RESEARCH AND INDUSTRIAL CONSULTANCY
INDIAN INSTITUTE OF TECHNOLOGY, KHARAGPUR**

STATEMENT OF RECEIPTS AND PAYMENTS OF RESEARCH SCHEME/PROJECT FOR THE
PERIOD **01/04/2015 TO 31/03/2016**

NAME OF THE PROJECT **"Ensemble Modeling of Rainfall-Runoff transformation Process (ERT)"**

SPONSORING AGENCY Ministry of Water Resources

DEPARTMENT Agricultural & Food Engineering


INVESTIGATOR-IN-CHARGE Prof. Rajendra Singh

Amount in ₹

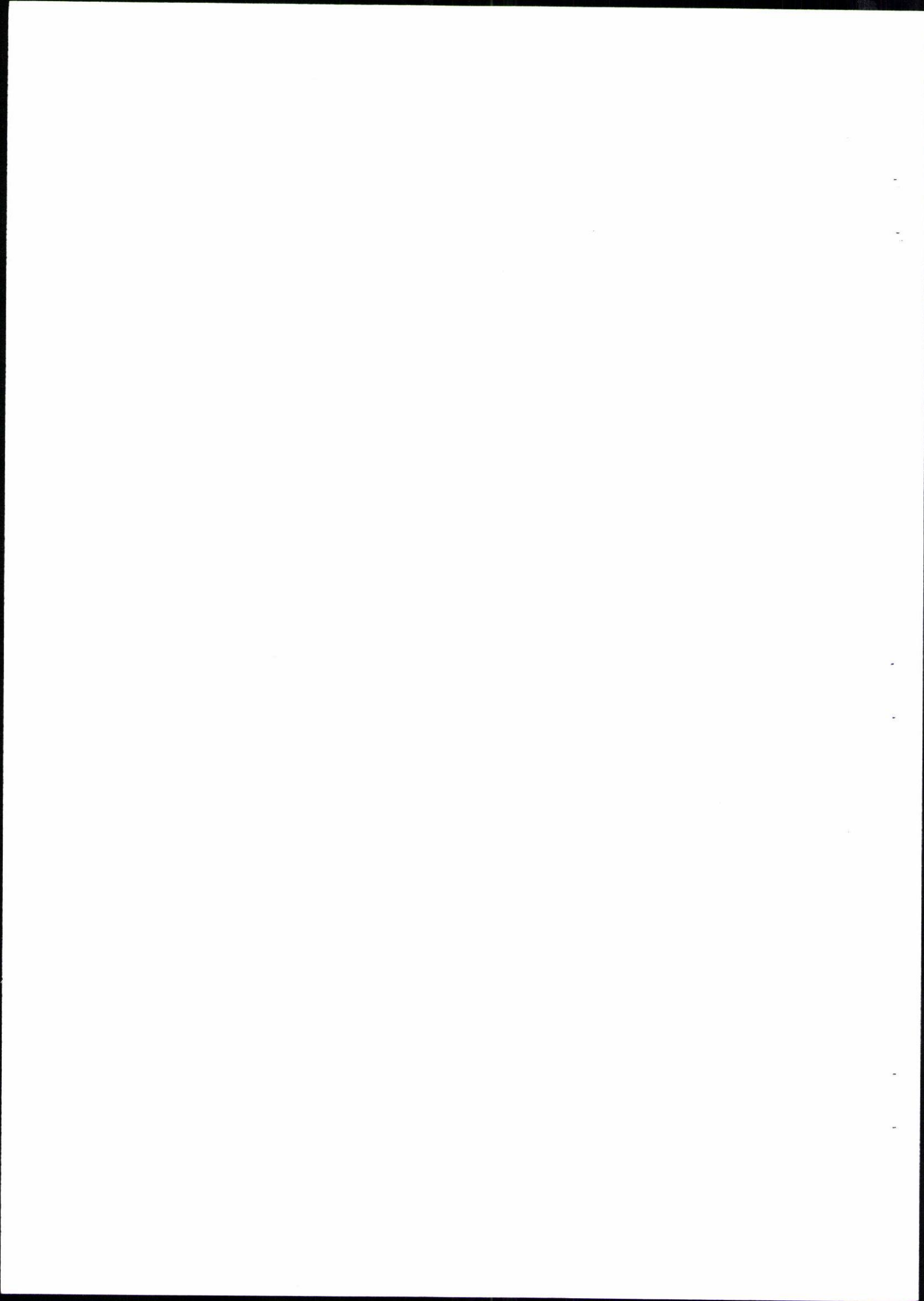
RECEIPTS	Amount Sanctioned	Amount Received	Amount Spent
By, Salary / Fellowship	672000	236000	4500
By, Travel Expenditure	155000	87675	0
By, Infrastructure/Equipment	988000	0	0
By, Experimental charges	300000	100000	17668
Sub Total	2115000	423675	22168
By, Contingency @ 5 %	105750	80000	6934
Total	105750	80000	6934
By, Institute Overhead	423000	0	0
GRAND TOTAL	2643750	503675	29102
Closing Balance as on date			467171

CERTIFIED THAT THE ABOVE MONEY HAVE BEEN UTILIZED FOR THE PURPOSE
FOR WHICH THE SAME WAS SANCTIONED


Principal Investigator


Deputy Registrar (SRIC)

Sumit Kumar Biswas
Deputy Registrar
Sponsored Research & Industrial Consultancy
I.I.T. Kharagpur-721302



**SPONSORED RESEARCH AND INDUSTRIAL CONSULTANCY
INDIAN INSTITUTE OF TECHNOLOGY, KHARAGPUR**

STATEMENT OF RECEIPTS AND PAYMENTS OF RESEARCH SCHEME/PROJECT FOR THE
PERIOD **01/04/2016 TO 31/12/2016**

NAME OF THE PROJECT **"Ensemble Modeling of Rainfall-Runoff transformation Process (ERT)"**

SPONSORING AGENCY Ministry of Water Resources

DEPARTMENT Agricultural & Food Engineering

INVESTIGATOR-IN-CHARGE Prof. Rajendra Singh

Amount in ₹

RECEIPTS	Amount Sanctioned	Amount Received	Amount Spent
By, Salary / Fellowship	672000	0	0
By, Travel Expenditure	155000	0	125028
By, Infrastructure/Equipment	988000	0	230233
By, Experimental charges	300000	0	179740
Sub Total	2115000	0	535001
By, Contingency @ 5 %	105750	0	24652
Total	105750	0	24652
By, Institute Overhead	423000	0	0
GRAND TOTAL	2643750	0	559653
Closing Balance as on date			-92482

CERTIFIED THAT THE ABOVE MONEY HAVE BEEN UTILIZED FOR THE PURPOSE
FOR WHICH THE SAME WAS SANCTIONED



Principal Investigator



Deputy Registrar (SRIC)

Sumit Kumar Biswas
Deputy Registrar
Sponsored Research & Industrial Consultancy
I.I.T. Kharagpur-721302

Annexure II

Utilization Certificate for the financial years 2014-15

Title of the Project/Scheme: **"Ensemble Modeling of Rainfall-Runoff transformation process (ERT)"**

Name of the Institution: **Indian Institute of Technology, Kharagpur- 721 302**

Principal Investigator: **Prof. Rajendra Singh**

Ministry of Water Resources letter reference sanctioning the project Head of account as given in the original sanction letter: MoWR Administrative Approval No. 23/68/2012-R&D/427-437,
Dated: 06/03/2012.

Financial Year to which UC pertains: **2014-15**

(Hereinafter referred to as the **UC Financial Year**)

	Reference (MoWR letter and date)	Amount in ₹
Amount brought forward from the previous financial year	23/68/2012-R&D/218-230 Date: 10/02/2016	20,596.00
Amount received during the UC financial year	N/A	0.00
Total amount that was available for expenditure during the UC financial year	Nil	20,596.00
Actual expenditure incurred during the UC financial year	Nil	27,998.00
Balance amount available at the end of the UC financial year	Nil	(-) 7,402.00

Unspent balance is to be ~~refunded~~ OR carried forward to the next financial year

Certified that:

1. The information given above is correct
2. The balance of Rs. NA remaining unutilized at the end of the year has been refunded to MoWR vide DD No.....Dated.....drawn on (bank).....

The negative balance of **Rs. 7,402.00** remaining unutilized at the end of the year is carried forward for ~~utilization~~/adjustment during the next year i.e. 2015-16.



Signature of Principal Investigator

Date: 24/5/17



Signature of Registrar/
Accounts Officer

Date:

Sumit Kumar Biswas
Deputy Registrar
Sponsored Research & Industrial Cell
I.I.T. Kharagpur-721 302

Signature of Head
of the Institute

Date:

25.5.17
Dean
Sponsored Research &
Industrial Consultancy
I.I.T. Kharagpur

Utilization Certificate for the financial years 2015-16

Title of the Project/Scheme: **"Ensemble Modeling of Rainfall-Runoff transformation process (ERT)"**

Name of the Institution: **Indian Institute of Technology, Kharagpur- 721 302**

Principal Investigator: **Prof. Rajendra Singh**

Ministry of Water Resources letter reference sanctioning the project Head of account as given in the original sanction letter: MoWR Administrative Approval No. 23/68/2012-R&D/427-437,

Dated: 06/03/2012.

Financial Year to which UC pertains: **2015-16**

(Hereinafter referred to as the **UC Financial Year**)

	Reference (MoWR letter and date)	Amount in ₹
Amount brought forward from the previous financial year	Nil	(-) 7,402.00
Amount received during the UC financial year	23/68/2012-R&D/218-230 Date: 10/02/2016	5,03,675.00
Total amount that was available for expenditure during the UC financial year	Nil	4,96,273.00
Actual expenditure incurred during the UC financial year	Nil	29,102.00
Balance amount available at the end of the UC financial year	Nil	4,67,171.00

Unspent balance is to be ~~refunded~~ **OR** carried forward to the next financial year

Certified that:

1. The information given above is correct
2. The balance of Rs. **4,67,171.00** remaining unutilized at the end of the year has been refunded to MoWR vide DD No.....Dated.....drawn on (bank).....

The balance of **Rs. 4,67,171.00** remaining unutilized at the end of the year is carried forward for utilization during the next year i.e. 2016-17.

Signature of Principal Investigator

Date: 24/5/17

Signature of Registrar/
Accounts Officer

Date:

Sumit Kumar Biswas
Deputy Registrar
Sponsored Research & Industrial Consultancy
I.I.T., Kharagpur- 721302

Signature of Head
of the Institute

Date:

25.5.17
Dean
Sponsored Research &
Industrial Consultancy
I.I.T., Kharagpur

Utilization Certificate for the financial years 2016-17 (till 31/12/2016 date of termination)

Title of the Project/Scheme: **"Ensemble Modeling of Rainfall-Runoff transformation process (ERT)"**

Name of the Institution: **Indian Institute of Technology, Kharagpur- 721 302**

Principal Investigator: **Prof. Rajendra Singh**

Ministry of Water Resources letter reference sanctioning the project Head of account as given in the original sanction letter: MoWR Administrative Approval No. 23/68/2012-R&D/427-437,

Dated: 06/03/2012.

Financial Year to which UC pertains: **2016-17**

(Hereinafter referred to as the **UC Financial Year**)

	Reference (MoWR letter and date)	Amount in ₹
Amount brought forward from the previous financial year	Nil	4,67,171.00
Amount received during the UC financial year	Nil	0.00
Total amount that was available for expenditure during the UC financial year	Nil	4,67,171.00
Actual expenditure incurred during the UC financial year	Nil	5,59,653.00
Balance amount available at the end of the UC financial year	Nil	(-) 92,482.00

Unspent balance is to be ~~refunded~~ **OR** carried forward to the next financial year

Certified that:

1. The information given above is correct
2. The balance of Rs. **Nil** remaining unutilized at the end of the year has been refunded to MoWR vide DD No.....Dated.....drawn on (bank).....

The negative balance of **(-) 92,482.00** remaining unutilized at the end of the year is carried forward for utilization/adjustment during the ~~next year~~/next release of grants.

Signature of Principal Investigator

Date: 24/5/17

Signature of Registrar/
Accounts Officer

Date:

Sunil Kumar Biswas
Deputy Registrar
Sponsored Research & Industrial Cell
I.I.T., Kharagpur- 721 302

Signature of Head
of the Institute

Date:

Date
Sponsored Research &
Industrial Consultancy
I.I.T., Kharagpur

Annexure-III

Statement of Equipment Purchased under the Scheme

Assets Acquired Wholly or Substantially out of Government grants Registrar maintained by grantee institution

Name of Sanctioning Authority: Ministry of Water Resources, Govt. of India

Serial No	1
Name of Grantee Institution	Indian Institute of Technology, Kharagpur
No. and date of sanction	MoWR Administrative Approval No. 23/68/2012-R&D/427-437 dated 06 March, 2012
Amount of the Sanctioned grant	Rs. 3,38,000
Brief purpose of the grant	To procure Intel Processor Based Server, Laptop (Notebook), and MIKE SHE Software Upgrade
Whether any condition regarding the right of ownership of Government in the property or other assets acquired out of the grant was incorporated in the grant-in-aid sanction.	As per administrative approval.
Particulars of assets actually credited or acquired	(i) Laptop (ii) MIKE SHE software Upgrade (iii) Server
Value of the Assets as on (Date)	INR 4,25,300
Purpose for which utilised at Present	The Server and the laptop are being used to carry out rainfall-runoff simulations. Laptop is also being used while collecting data from different agencies. The MIKE SHE software is being used for hydrological modelling of the selected catchments.
Encumbered or not	No
Reasons if encumbered	Not Applicable
Disposed of or not	No
Reasons and authority, if any, for disposal	Not Applicable
Amount realised on disposal	Not Applicable
Remarks	The equipment are being fully utilised

Annexure -IV

Data collected and used in the analysis along with source of data

Data	Sources
Digital Elevation Model	SRTM, CARTOSAT-1
Landuse	Landsat ETM+, RESOURCESAT-2 (LISS –III)
Rainfall	IMD, Hydrometry Bhubaneswar
Maximum temperature	IMD, Hydrometry Bhubaneswar
Minimum temperature	IMD, Hydrometry Bhubaneswar
River discharge	CWC Bhubaneswar
Soil map	National Bureau of Soil Survey and Land Use Planning (NBSS & LUP), Kolkata

Annexure -V

Study Area, Methodology, Results & Discussion

This section includes a brief description of the study area, different hydrological models and ensemble methods along with theoretical background for model /ensemble evaluation. Methodology behind uncertainty analysis is also discussed in this section. Furthermore, results are presented in latter part followed by associated discussion.

1 Study Area

Two catchments of the middle reaches of Mahanadi River Basin namely, Kesinga and Salebhata, located in Odisha, India, and having concurrent rainfall-runoff data are chosen as the study area (Fig. 1). Kesinga lies between 19°16'10" and 20°44'42" N latitudes, and 82°02'50" and 83°24'09" E longitudes, and covers an area of 12371 km², whereas, Salebhata lies between 20°40'12" and 21°25'08" N latitudes, and 82°33'24" and 83°34'11" E longitudes, and covers an area of 4515 km². The main river of the study area is Mahanadi which has a drainage area of 141589 km² including 75136 km², 65580 km², 635 km² and 238 km² in Chattisgarh, Odisha, Maharastra and Jharkhand states of India, respectively. The total length of the river from origin to its outfall into Bay of Bengal is 851 km, of which 357 km lies in Chhattisgarh and 494 km in Odisha. The middle reaches of the Mahanadi river extends from Hirakud dam to Munduli weir having a total length of 358.37 km. The rivers joining the main Mahanadi River are Ong, Bagh, Suktel and Tel. Tel river originates in and passes through Kesinga, whereas Ong river originates in and passes through Salebhata. About 50% of the area of middle reaches of the basin is hilly varying in elevation from 300 m to 1321 m above mean sea level, and remaining area lies between 0 m and 300 m on either side of the Mahanadi River. The red and yellow soils cover a major part of the area. Paddy is the main crop grown on the cultivable land. The average annual rainfall of the basin is 1360 mm. The Normal rainfall season is of five months, i.e., from June to October, though the maximum precipitation is usually observed during July – Mid-September (CWC, 1997). The coldest and hottest months in the basin are December and May, respectively. The entire basin area is divided into smaller catchments as per the location of the discharge gauging sites. However, due to data availability constraints, only two catchments, i.e., Kesinga and Salebhata, are chosen for this study.

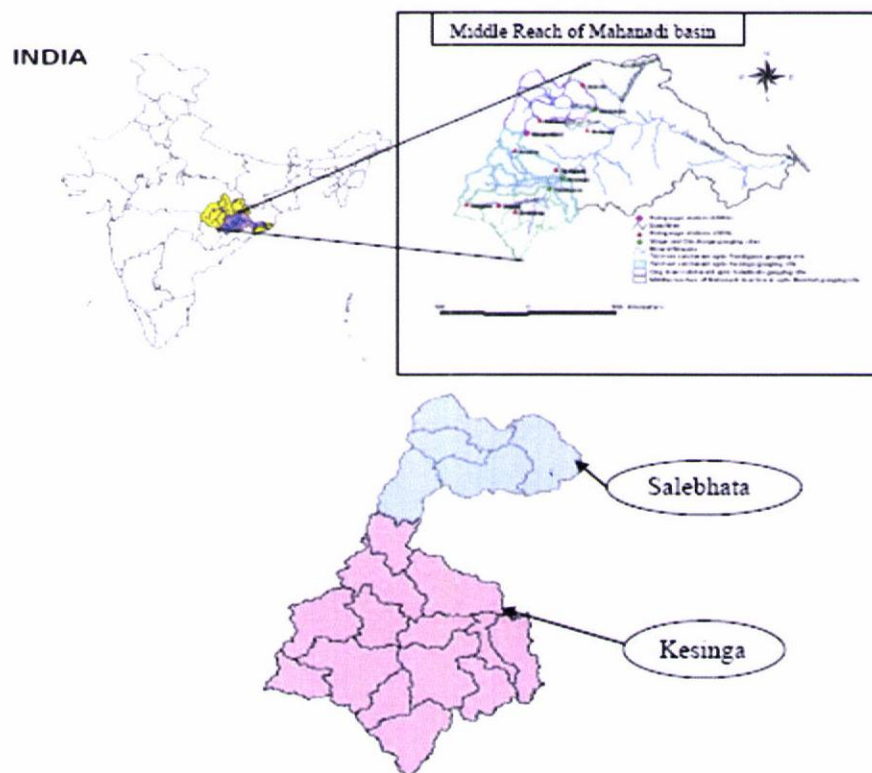


Fig. 1: Location of the study area

2 Theory and Methodology

2.1 Hydrological Models

As the primary objective of the study is to test the efficacy of the concept of combining various rainfall-runoff model outputs, the readily available models were used. This selection of models is neither considered to be optimum nor even considered to be entirely representative of the range of models present in the literature. However, based on data availability, model requirements and expertise, eight hydrological models, varying from lumped conceptual to distributed physically based were chosen in this study. The models include two distributed physically based models, i.e., MIKE SHE (Abbott et. al. 1986a, 1986b; Refsgaard and Storm, 1995) and SWAT (Arnold et al., 1998), and six lumped conceptual models, i.e., HEC-HMS (Feldman, 2000; USACE, 2000), TANK (Sugawara, 1979), AWBM (Boughton, 2004), SIMHYD (Chiew et al., 2002), SACRAMENTO (Burnash et al., 1973) and SMAR (O'Connell et al., 1970). The chosen models represent a broad cross-section of complexity ranging from fully distributed, physically-based models to lumped, conceptual models in terms of their spatial resolution and number of model parameters. Process parameterization of these models is different; hence they behave

differently, which is a necessary criterion for ensemble creation. A brief description of these models is given in the following sub-sections.

2.1.1 MIKE SHE

MIKE SHE is a comprehensive deterministic, distributed and physically based modelling system, capable of describing the entire land phase of the hydrological cycle in a given watershed (Abbott et al., 1986a, b; Refsgaard and Storm, 1995). The model area is discretized by two analogous horizontal-grid square networks for surface and groundwater flow components. These are linked by vertical column of nodes at each grid representing the unsaturated zone. MIKE SHE modelling system is designed with a modular structure. Its core module is MIKE SHE Water Movement Module (MIKE SHE WM). Other MIKE SHE modules are built around this core module. Each component solves a corresponding equation as follows: 3-D Boussinesq Equation for saturated groundwater flow, 1-D Richards' Equation for unsaturated flow, 2-D diffusive wave approximation of the Saint Venant equations for overland flow, and 1-D diffusive wave approximation of the Saint Venant equations for river flow. MIKE SHE uses the Kristensen and Jensen (1975) method for calculating actual evapotranspiration based on potential evaporation, leaf area index, root depth for each vegetation type, and a set of empirical parameters. Fig. 2 presents the schematic diagram of MIKE SHE model structure.

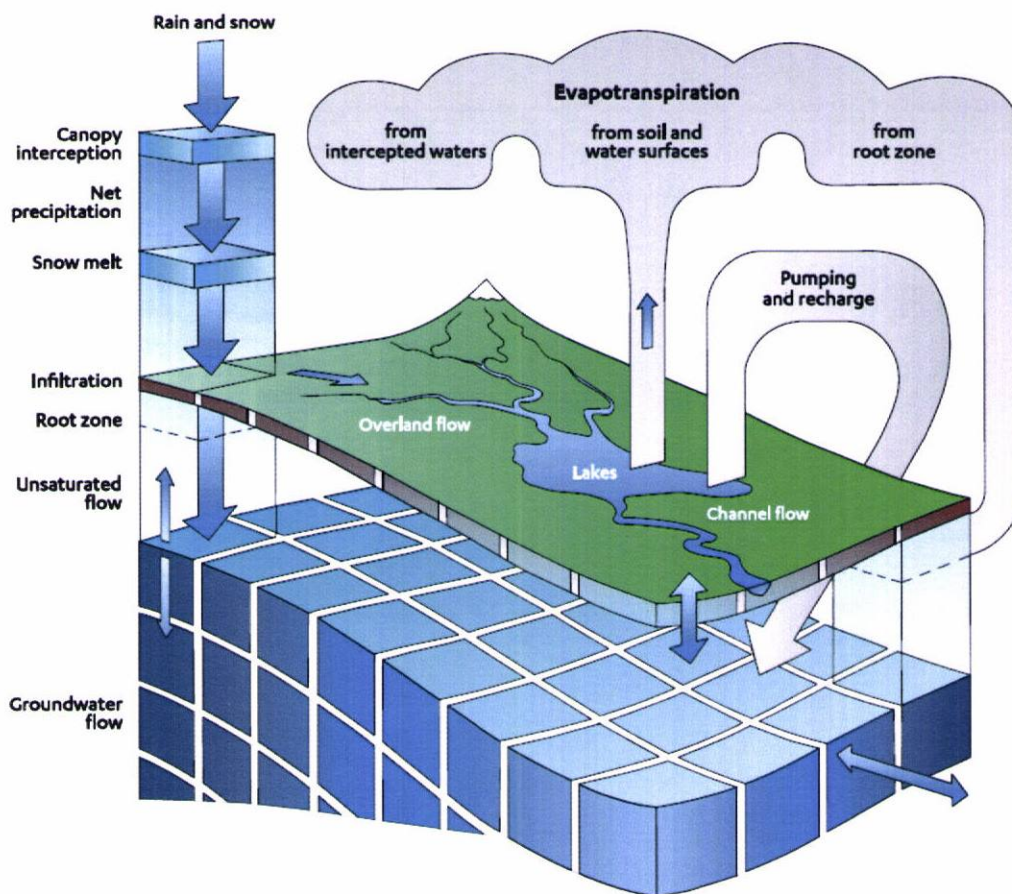


Fig. 2: Schematic diagram of MIKE SHE model (DHI, 2007)

2.1.2 SWAT

Soil and Water Assessment Tool (SWAT) (Arnold et al., 1998) is a physically based, continuous simulation watershed model which is capable of simulating surface runoff, sediment yield and nutrient losses from small, medium and large watersheds. The physical processes associated with water and sediment movement, crop growth, and nutrient cycling are directly modeled by SWAT using input data. The model simulates the hydrological cycle based on the water balance equation using precipitation, surface runoff, evapotranspiration, infiltration and return flow. The model uses SCS curve number technique and Green-Ampt model to compute surface runoff volume and infiltration, respectively. Potential evaporation may be computed using Hargreaves, Priestley-Taylor or Penman-Monteith methods. For return flow estimation, SWAT partitions groundwater into two aquifer systems – a shallow, unconfined aquifer which contributes return flow to streams within the watershed and

a deep, confined aquifer which contributes return flow to streams outside the watershed.

2.1.3 HEC-HMS

Hydrologic Modelling System (HEC-HMS) (Feldman, 2000; USACE, 2000) was designed to simulate the precipitation-runoff processes of dendritic watershed systems, which is applicable in a wide range of geographic areas for solving the widest possible range of problems. The physical representation of a watershed is accomplished with a basin model in which hydrologic elements such as subbasin, reach, junction, reservoir, diversion, source, and sink are connected in a dendritic network to simulate runoff processes. Various methods are available to simulate infiltration losses and to transform excess precipitation into surface runoff, and for representing base flow contributions to sub-basin outflow. A variety of hydrologic routing methods such as Muskingum method, straddle stagger method, the modified Puls methods etc. are available to model a reach. Channels with trapezoidal, rectangular, triangular, or circular cross sections can be modeled with the kinematic wave or Muskingum-Cunge method. Channels with overbank areas can be modeled with the Muskingum-Cunge method using 8-point cross section.

2.1.4 TANK

TANK model (Sugawara, 1979) is composed of four tanks laid vertically in series. Precipitation is put into the top tank, and evaporation is subtracted sequentially from the top tank downwards. As each tank is emptied the evaporation shortfall is taken from the next tank down until all tanks are empty. The outputs from the side outlets are the calculated runoffs. The output from the top tank is considered as surface runoff, output from the second tank as intermediate runoff, from the third tank as sub-base runoff and output from the fourth tank as base flow. Despite this simple schematisation, the assessment of the tank model is not so simple. The assessment of the model is strongly influenced by the content of each of the stores. Under the same rainfall and different storage volumes the runoff generated is significantly different. The tank model is applied to analyse daily discharge from daily precipitation and evaporation inputs. The concept of initial loss of precipitation is not necessary,

because its effect is included in the non-linear structure of the tank model. Fig. 3 presents the schematic diagram of TANK model structure.

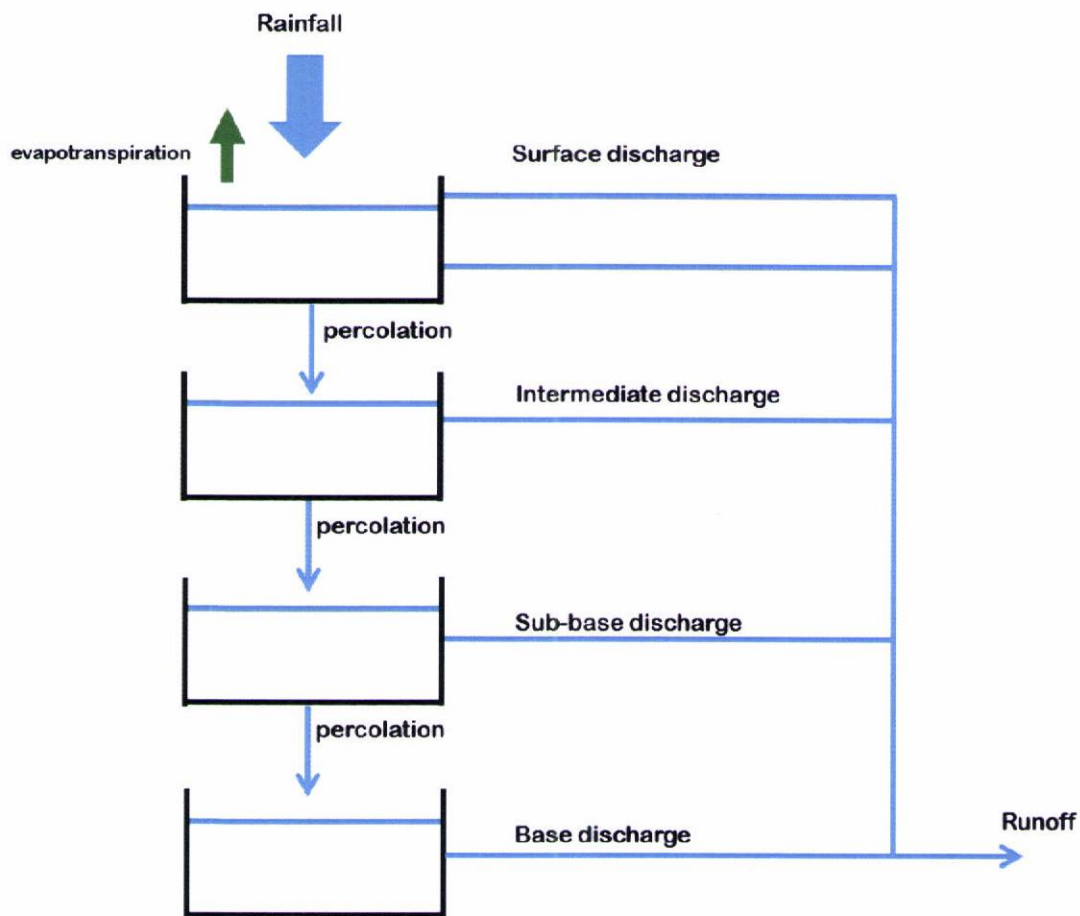


Fig. 3: Structure of TANK rainfall-runoff model (Clanor et al., 2015)

2.1.5 AWBM

AWBM (Boughton, 2004) is a catchment water balance model that relates runoff to rainfall with daily or hourly data, and calculates losses from rainfall for flood hydrograph modelling. The model uses three surface stores to simulate partial areas of runoff. The water balance of each surface store is calculated independently of the others. At each time step, rainfall is added to each of the three surface moisture stores and evapotranspiration is subtracted from each store. If the value of moisture in the store becomes negative, it is reset to zero, as the evapotranspiration demand is superior to the available moisture. If the value of moisture in the store exceeds the capacity of the store, the moisture in excess of the capacity becomes runoff and the store is reset to the capacity. When runoff occurs from any store, part of the runoff

becomes recharge of the base flow store and the remainder goes as surface runoff. The surface runoff can be routed through a store, if required, to simulate the delay of surface runoff reaching the outlet of a medium to large catchment. Fig. 4 presents the schematic diagram of AWBM model structure.

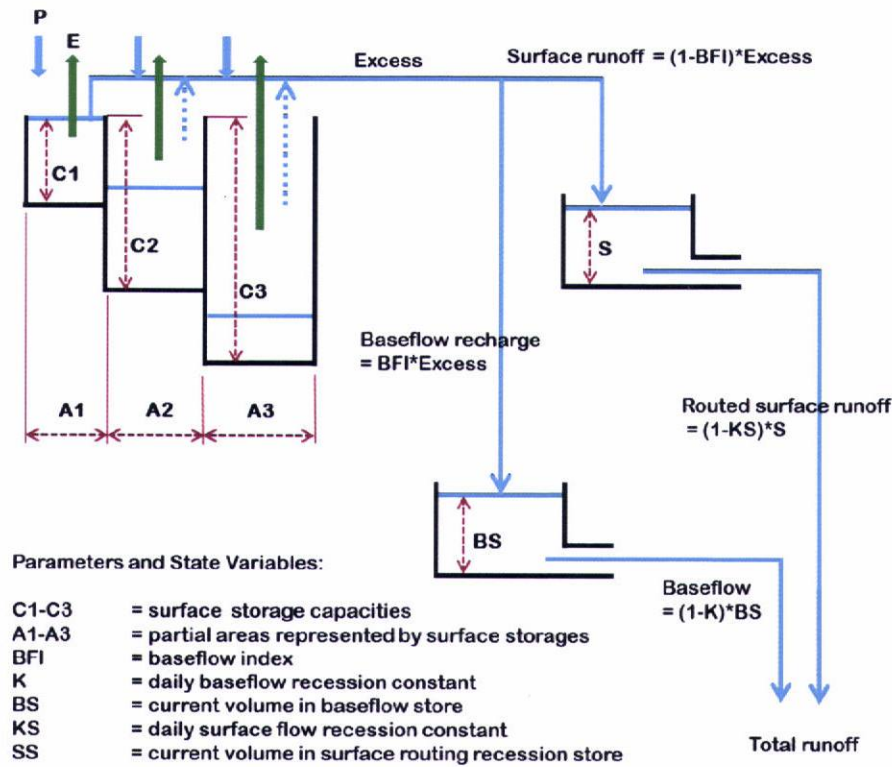


Fig. 4: Structure of AWBM rainfall-runoff model (Clanor et al., 2015)

2.1.6 SIMHYD

SIMHYD (Chiew et al., 2002) is a daily conceptual rainfall-runoff model that estimates daily streamflow from daily rainfall and areal evapotranspiration data. In SIMHYD, daily rainfall first fills the interception store, which is emptied by evaporation. The excess rainfall is then subjected to an infiltration function that determines the infiltration capacity. The excess rainfall that exceeds the infiltration capacity becomes infiltration excess runoff. Moisture that infiltrates is subjected to a soil moisture function that diverts the water to the stream (interflow), groundwater store (recharge) and soil moisture store. The equation used to simulate interflow, therefore, attempts to mimic both the interflow and saturation excess runoff processes (with the soil wetness used to reflect parts of the catchment that are saturated from which saturation excess runoff can occur). Groundwater recharge is then estimated as

a linear function of the soil wetness. The remaining moisture flows into the soil moisture store. Evapotranspiration from the soil moisture store is estimated as a linear function of the soil wetness, but cannot exceed the atmospherically controlled rate of areal potential evapotranspiration. The soil moisture store has a finite capacity and overflows into the groundwater store. Base flow from the groundwater store is simulated as a linear recession from the store. Finally, model estimates runoff generated as the sum of three sources, i.e., infiltration excess runoff, interflow (and saturation excess runoff) and base flow. Fig. 5 presents the schematic diagram of SIMHYD model structure.

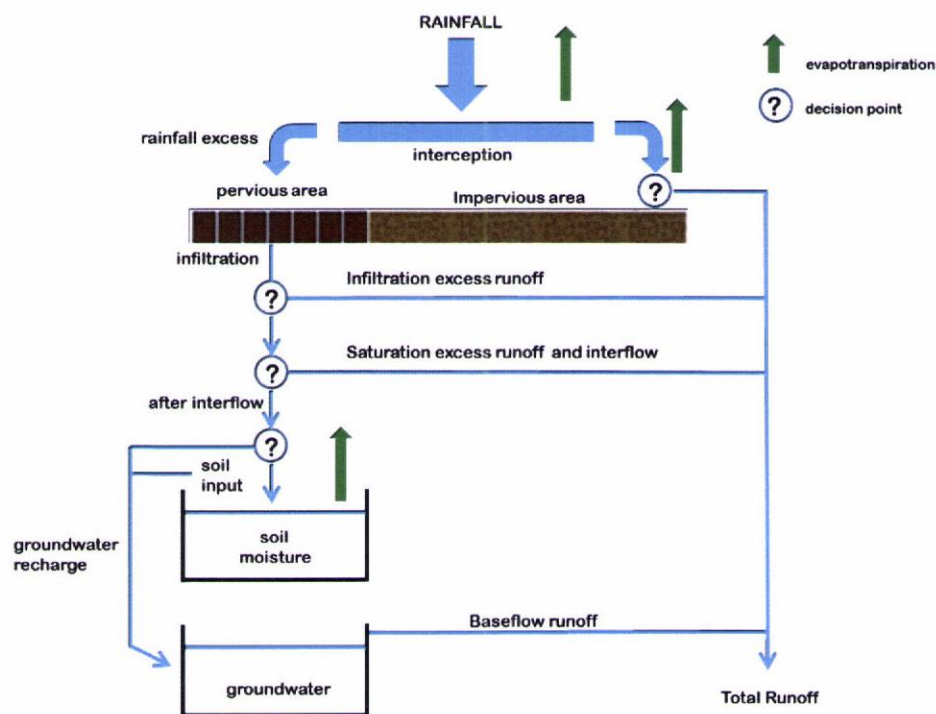


Fig. 5: Structure of SIMHYD rainfall-runoff model (Clanor et al., 2015)

2.1.7 SACRAMENTO

SACRAMENTO (Burnash et al., 1973) model is a continuous rainfall-runoff model used to generate daily streamflow from rainfall and evaporation records. The Model uses soil moisture accounting to simulate the water balance within the catchment. Soil moisture storage is increased by rainfall and reduced by evaporation and by flow of water out of the storage. There are five stores in the SACRAMENTO Model: Upper zone tension water, Upper zone free water, Lower zone tension water, Lower zone primary free water and Lower zone supplementary free water. The tension water

stores represent the volume of water that is held in the soil matrix by surface tension. Water can only be removed from tension stores by evapotranspiration. In the free water stores, water can move through the soil vertically to other stores, or laterally as interflow (upper zone) or as baseflow (lower zone). Stream flow generated with this model is made up of three flow components: surface runoff, interflow and base flow. The generation of these components depends on the amount of water in each store relative to the stores capacity, and the rate at which water moves into and out of these stores. Fig. 6 presents the schematic diagram of SACRAMENTO model structure.

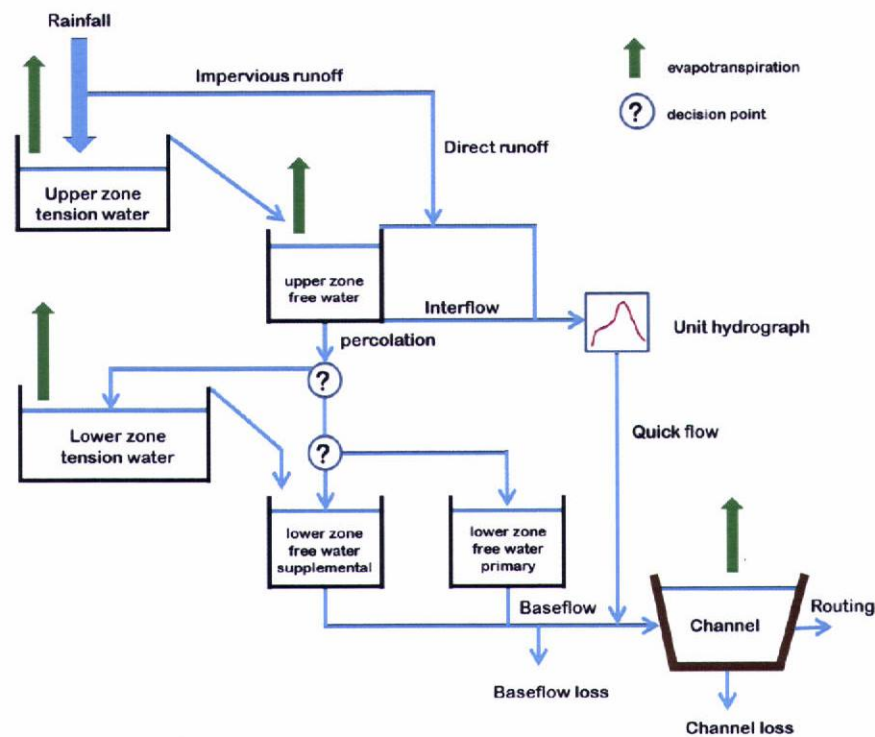


Fig. 6: Structure of SACRAMENTO rainfall-runoff model (Clanor et al., 2015)

2.1.8 SMAR

Soil Moisture and Accounting Model (SMAR) (O'Connell et al., 1970) is a lumped conceptual rainfall runoff water balance model with soil moisture as a central theme. The model consists of two components in sequence, a water balance component and a routing component. The model simulates stream flow at the catchment outlet by using time series of rainfall and pan evaporation. The water balance component divides the soil column into horizontal layers, which contain a prescribed amount of water (usually 25 mm) at their field capacities. Evaporation from soil layers is treated in a

way that reduces the soil moisture storage in an exponential manner from a given potential evapotranspiration demand. The routing component transforms the surface runoff generated from the water balance component to the catchment outlet by a gamma function model form (Nash, 1960), a parametric solution of the differential routing equation in a single input single output system. The generated groundwater runoff is routed through a single linear reservoir and provides the groundwater contribution to the stream at the catchment outlet. The surface runoff generated from the landscape is routed (attenuation and lag) to the catchment outlet using the linear cascade model of Nash (1960). Fig. 7 presents the schematic diagram of SMAR model structure.

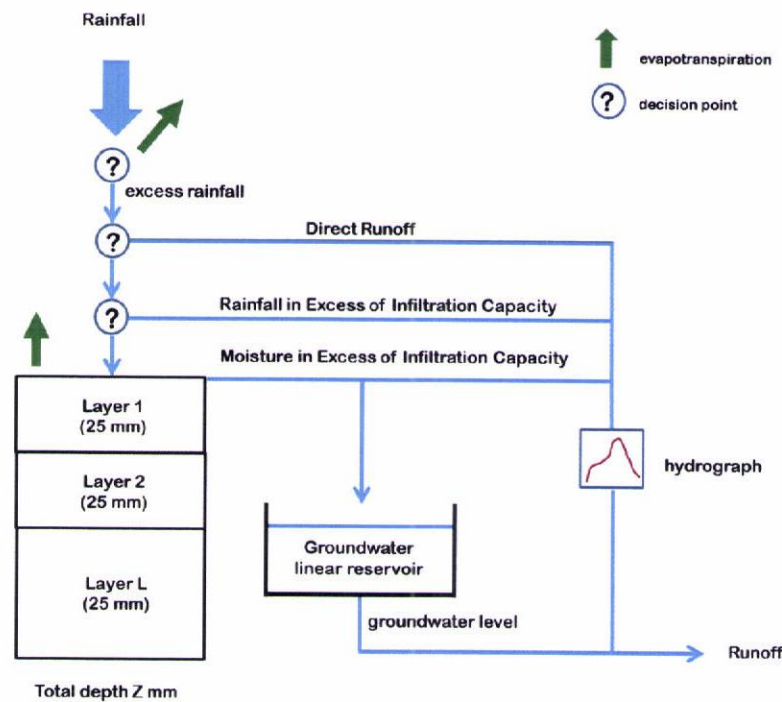


Fig. 7: Structure of SMAR rainfall-runoff model (Clanor et al., 2015)

2.2 Model Setup and Ensemble Preparation

2.2.1 Data Collection

Boundary and topography maps of the study catchments were derived from Shuttle RADAR Topography Mission (SRTM) image of 90 m resolution which was downloaded from Consultative Group on International Agricultural Research-Consortium for Spatial Information (CGIAR-CSI) website. Land use/land cover map

was derived from LANDSAT -7 (ETM+) images, downloaded from Global Land Cover Facility (GLCF) website. Soil map was collected from the National Bureau of Soil Survey and Land Use Planning (NBSS & LUP), Kolkata and corresponding soil properties were estimated using Rosetta 1.0 software (Marcel et al., 2001). Daily rainfall and temperature data of eight and six meteorological stations, respectively, for Kesinga and Salebhata, were collected from India Meteorological Department (IMD), Bhubaneswar for the period 2004 to 2009. Figs. 8 and 9 present the rainfall station data of Kesinga and Salebhata, respectively. Thiessen polygon method was used to obtain the average daily rainfall of the catchment. Figs. 10 & 11 present the location of raingauge stations and Thiessen polygons for Kesinga and Salebhata, respectively.

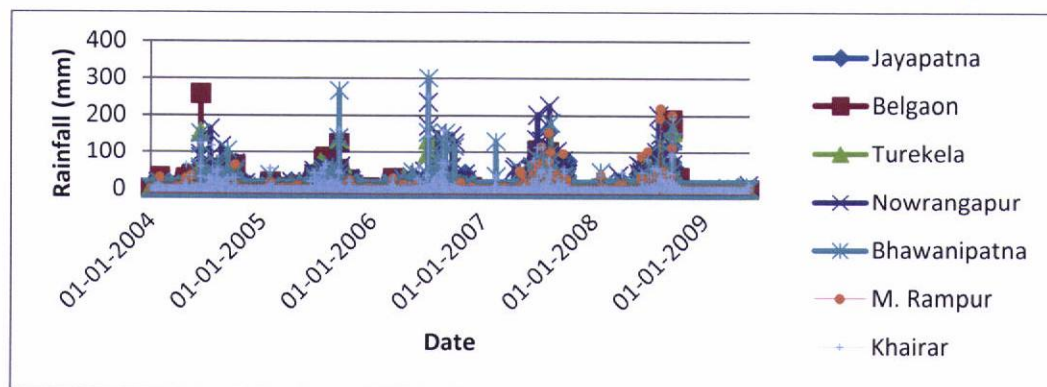


Fig. 8: Rainfall of different raingauges in Kesinga

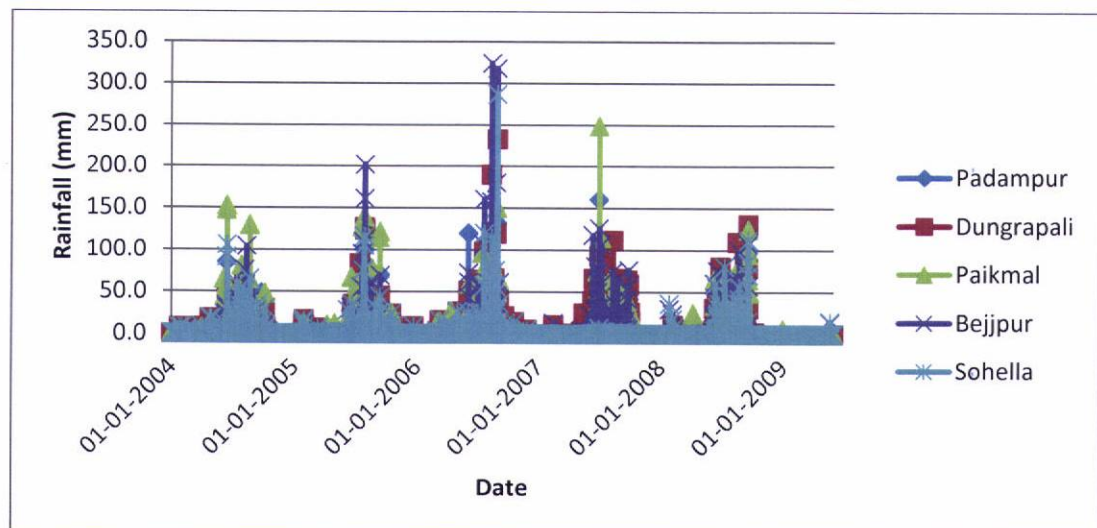


Fig. 9: Rainfall of different raingauges in Salebhata

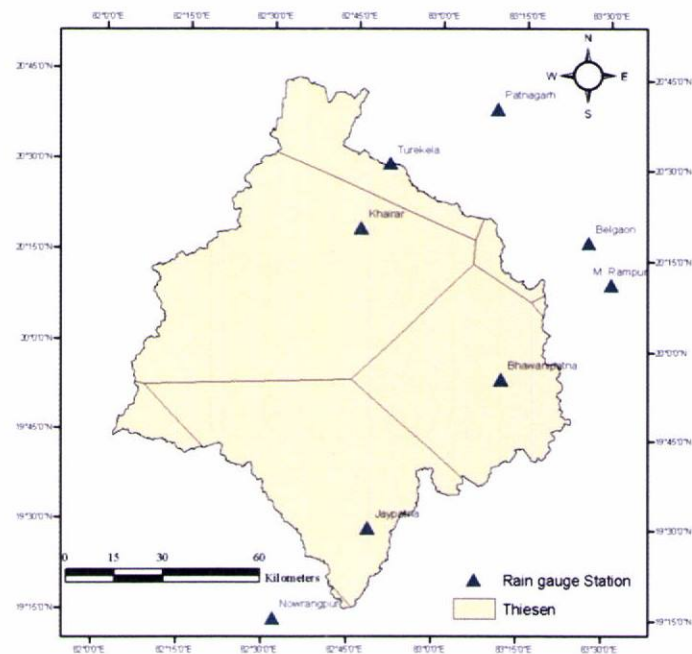


Fig. 10: Location of rain gauge stations, and Thiessen polygon for Kesinga

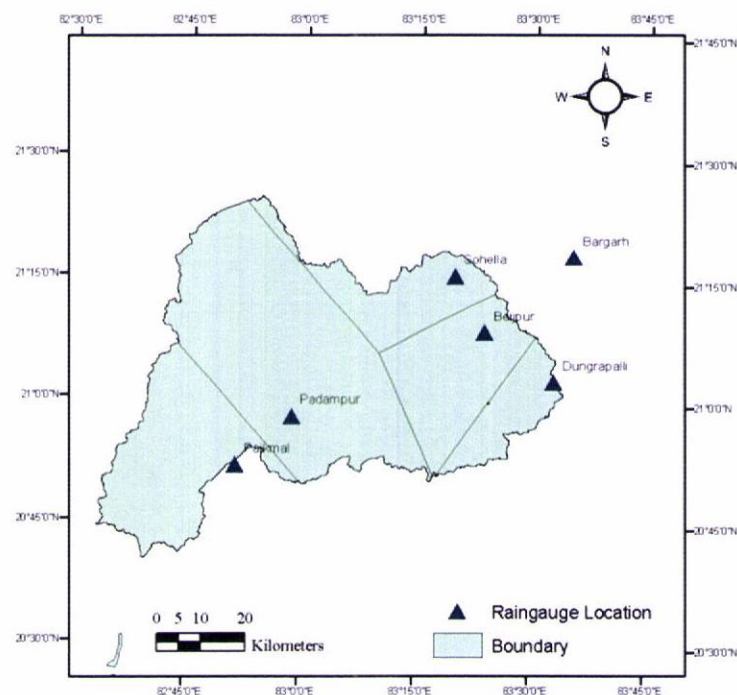


Fig. 11: Location of rain gauge stations, and Thiessen polygon for Salebhata

The average annual rainfall over the catchments is found to be 1475 mm and 1284 mm for Kesinga and Salebhata, respectively. The discharge data at the outlet of the catchments was obtained from Central Water Commission (CWC), Bhubaneswar for

the period 2004 to 2009. Figs. 12 and 13 present the average rainfall of Kesinga and Salebhata, respectively, along with observed discharge at the outlet of the respective catchments. The average daily discharge at the outlet of the catchments is found to be $313 \text{ m}^3/\text{s}$ and $72 \text{ m}^3/\text{s}$ for Kesinga and Salebhata, respectively. The entire dataset was split into two parts: June 2004-May 2007, which was used for calibration, and June 2007- May 2009, which was used for validation of different hydrological models. Mean and standard deviation of discharge at the outlet of Kesinga are found to be 303 and $909 \text{ m}^3/\text{s}$, respectively during the calibration period, and 358 and $1069 \text{ m}^3/\text{s}$, during the validation period. Similarly, mean and standard deviation of daily discharge at the outlet of Salebhata are found to be 72 and $275 \text{ m}^3/\text{s}$, respectively, during the calibration period, and 59 and $220 \text{ m}^3/\text{s}$, during the validation period. The difference in average of observed discharges, during calibration and validation periods, is analyzed using Student's t-test and found to be statistically insignificant at 95% level of significance. The pre- and post-monsoon groundwater table depths at 32 and 34 different locations in Kesinga and Salebhata, respectively, were collected from hydrometry division, Bhubaneswar. Reference evapotranspiration was determined by the Hargreaves equation using DSS_ET software (Bandyopadhyay et al., 2012).

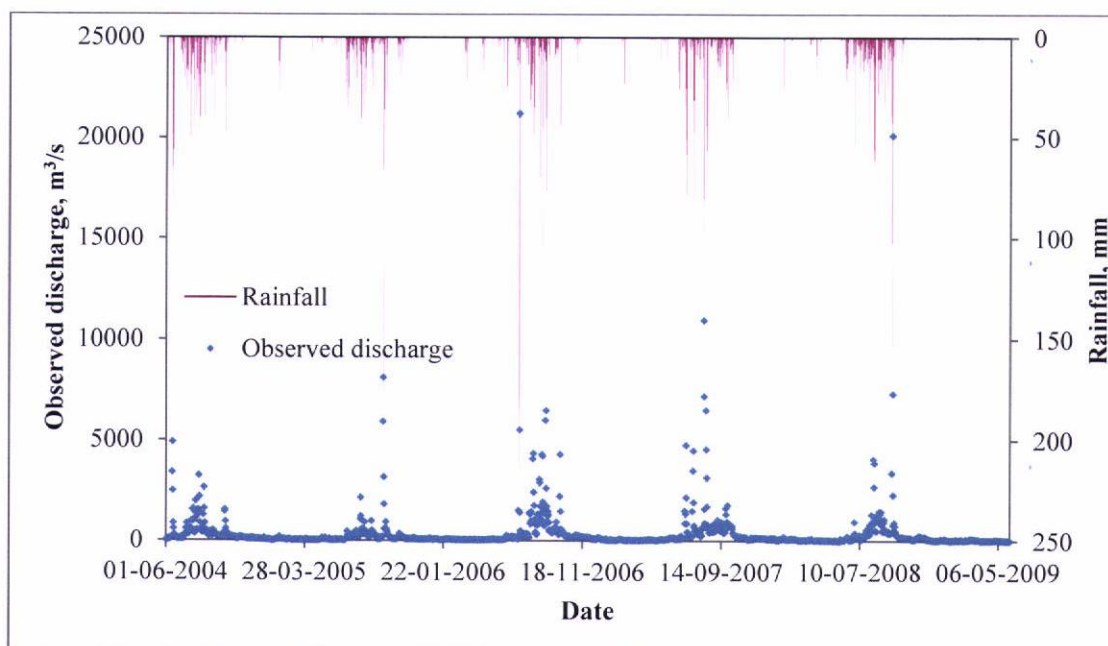


Fig. 12: Average rainfall over the catchment and observed discharge at the outlet of Kesinga

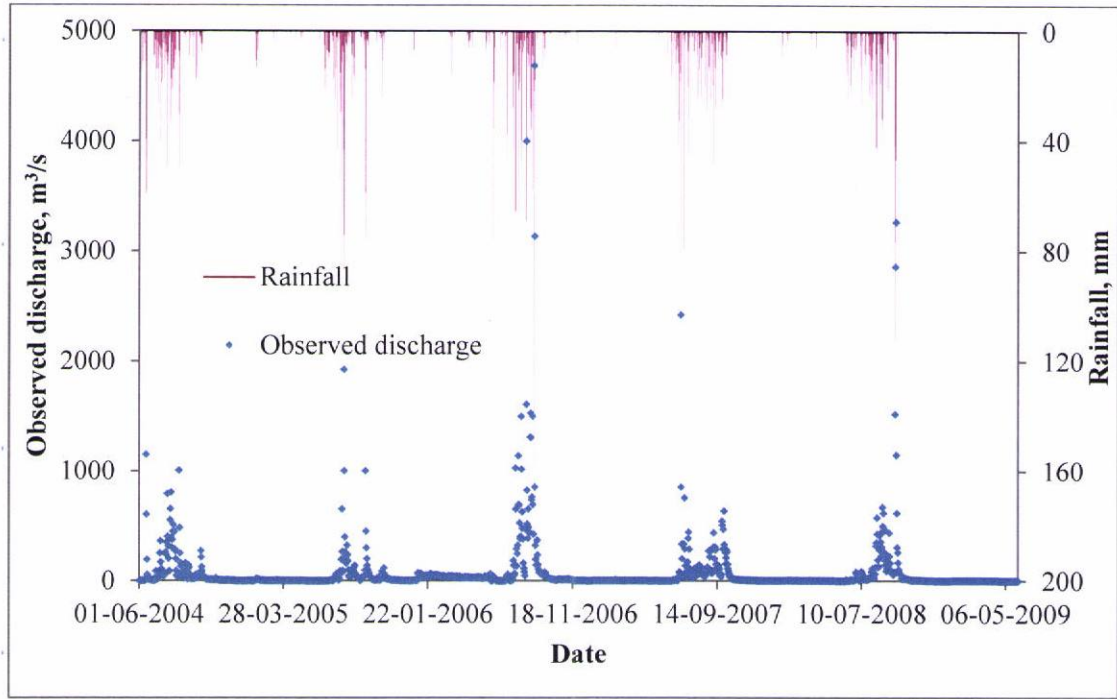


Fig. 13: Average rainfall over the catchment and observed discharge at the outlet of Salebhata

2.2.2 Model Setup, Calibration and Validation

Model setup was developed for all eight models using the collected data followed by their calibration for the period of 1/6/2004 to 31/5/2007. Manual calibration was performed by trial and error method of parameter adjustments. Several simulations were performed for getting a calibrated model. After each adjustment of parameters, the simulated and the observed results were compared to find out the improvement in predictions. The performance of the model for simulating surface runoff was evaluated using Nash Sutcliffe efficiency (NSE) and percent bias (PBIAS) as performance index as per Eq. 1 and 2. All the calibrated models were validated for the period of 1/6/2007 to 31/5/2009 by keeping the parameters of model unaltered.

$$PBIAS = \frac{\sum_{i=1}^n (Q_{sim,i} - Q_{obs,i})}{\sum_{i=1}^n Q_{obs,i}} \times 100 \quad (1)$$

$$R = \frac{\sum_{i=1}^n (Q_{obs,i} Q_{sim,i} - n \bar{Q}_{obs} \bar{Q}_{sim})}{\sqrt{\left(\sum_{i=1}^n Q_{obs,i}^2 - n \bar{Q}_{obs}^2 \right) \left(\sum_{i=1}^n Q_{sim,i}^2 - n \bar{Q}_{sim}^2 \right)}} \quad (2)$$

where, $Q_{sim,i}$ = simulated discharge on i^{th} day, $Q_{obs,i}$ = observed discharge on i^{th} day, \bar{Q}_{sim} = average simulated discharge, \bar{Q}_{obs} = average observed discharge, and n = total number of days under consideration.

2.2.3 Multi-Model Ensemble Methods

There are several methods available for developing multi-model ensembles. Here, we have compared the performance of eight methods for creating the multi-model ensembles and for selecting the most suitable method for the study area. The selected methods include simple methods like mean, median and trimmed mean and more complex methods like weighted mean based on calibration performance (two variants), linear programming, simple model average and multi-model super ensemble. A brief description of these methods is given in the following sub-sections.

2.2.3.1 Mean

The mean is the simplest method of combining the outputs of different individual models. Given the estimated discharges of M rainfall-runoff models, a combined estimate of the discharge, Q , at the i^{th} time period, using mean, is given by

$$Q_{ens,i} = \frac{\sum_{j=1}^M Q_{sim,i,j}}{M} \quad (3)$$

where, $Q_{ens,i}$ = ensemble discharge at i^{th} time step, $Q_{sim,i,j}$ = simulated discharge by the j^{th} model at i^{th} time step.

In this method, equal emphasis (i.e. weight) is assigned to the outputs of all the considered models. The average can produce forecasts that are better than that of the individual models and its accuracy depends mainly on the number of models involved, and on the actual forecasting ability of the specific models included in the simple average (Makridakis and Winkler, 1983; Velázquez et al., 2011).

2.2.3.2 Median

The method estimates the median of discharges simulated by different models at different time steps (Viney et al., 2009) and is given by

$$Q_{ens,i} = Median(Q_{sim,i,j}) \quad (4)$$

2.2.3.3 Trimmed Mean

Ensemble using this method is estimated by excluding the result of two extreme models and taking the average of remaining models at every time step (Viney et al., 2009).

2.2.3.4 Weighted Mean Method (WAM_K=1, WAM_K=1.5)

Ensemble discharge using this method is calculated as:

$$Q_{ens} = PA \quad (5)$$

where, P = simulated discharge matrix, and A = weight matrix of the member models.

The weights of different models are estimated using Nash Sutcliffe efficiency of member models during the calibration period as (Viney et al., 2009):

$$a(j,k) = \frac{1}{\sum_{x=1}^M \frac{(1 - NSE_j)^k}{(1 - NSE_x)^k}} \quad (6)$$

where, NSE_x = Nash Sutcliffe efficiency of x^{th} model during calibration, NSE_j = Nash Sutcliffe efficiency of j^{th} model during calibration, $a(j,k)$ = weight of j^{th} model and k = a positive exponent which provides higher weight to the model having higher efficiency in the ensemble model.

Two variants of this method exist in literature, with $k=1$ and $k=1.5$.

NSE is given by:

$$NSE = 1 - \frac{\sum_{i=1}^n (Q_{obs,i} - Q_{sim,i})^2}{\sum_{i=1}^n (Q_{obs,i} - \bar{Q}_{obs})^2} \quad (7)$$

NSE ranges between $-\infty$ and 1, and it is often considered that the NSE higher than 0.5 characterizes a good simulation of the discharges (Moriassi et al., 2007).

2.2.3.5 Linear Programming (LP)

Ensemble discharge using this method is also calculated from Eq. 5.

For obtaining the weight matrix of the member models, sum of deviations between the simulated ensemble discharge and observed discharge is minimized. This is expressed mathematically as:

$$\text{Minimize: } Z = \sum_{i=1}^n U_i + V_i \quad (8)$$

$$\text{Subject to } \sum_{j=1}^M Q_{sim,i,j} W_j + U_i + V_i = Q_{obs,i} \quad \text{for } i=1 \text{ to } n. \quad (9)$$

$$\sum_{j=1}^M W_j = 1 \quad (10)$$

where, U_i = negative deviation variable on i^{th} day of simulation, V_i = positive deviation variable on i^{th} day of simulation, and W_j = weight of j^{th} model.

2.2.3.6 Simple Model Average (SMA)

This method uses the logic of bias reduction with respect to a member model (Georgakakos et al., 2004), and is considered as the benchmark in development of sophisticated methods of multi-model ensemble. The ensemble discharge is estimated as:

$$Q_{ens,i} = \bar{Q}_{obs} + \sum_{j=1}^M \frac{[Q_{sim,i,j} - \bar{Q}_{sim,j}]}{M} \quad (11)$$

where, $\bar{Q}_{sim,j}$ = average simulated discharge by j^{th} model.

2.2.3.7 Multi-Model Super Ensemble (MMSE)

This method uses the logic of bias reduction with respect to individual member models along with variance reduction in simulation/prediction (Krishnamurti et al., 2000). The ensemble discharge is estimated as:

$$Q_{ens,i} = \bar{Q}_{obs} + \sum_{j=1}^M a_j [Q_{sim,i,j} - \bar{Q}_{sim,j}] \quad (12)$$

where, a_j = the weight of j^{th} model in the weight matrix A . The weight matrix is estimated by unconstrained least square technique, which is estimated as:

$$A = (P^T P)^{-1} P^T Q \quad (13)$$

where, Q = observed discharge matrix and P^T = transpose of matrix P .

2.2.4 Evaluation of Multi-Model Ensemble Methods

Ensembles of the eight models were created using eight different ensemble methods and their performance was evaluated to identify the most suitable method for the selected catchment. For this purpose, RMSE and R , two commonly used statistical criteria in the field of hydrology, were used (Georgakakos et al., 2004; Ajami et al., 2006). The lower the RMSE and higher the R , better is the performance of the ensemble with respect to these scalar criteria. RMSE is defined as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Q_{sim,i} - \bar{Q}_{obs})^2} \quad (14)$$

2.3 Ensemble Selection

2.3.1 Development of Multi-Model Ensembles

Selected models belong to two classes: lumped conceptual and distributed physically based. Hence, at least one model from each class was considered while developing the multi-model ensembles. Table 1 presents 189 different possible combinations of the selected models. In this table, different models, viz., MIKE SHE, SWAT, HEC-HMS, TANK, AWBM, SIMHYD, SACREMANTO and SMAR are denoted by numbers 1, 2, 3, 4, 5, 6, 7 and 8, respectively. The combination of models is denoted by concatenation of their corresponding numbers in the last column of Table 1. For example, combination '124' denotes the combination of models '1', '2', and '4', i.e., MIKE SHE, SWAT and TANK.

Table 1: Possible ensembles with model combinations

Ensemble size	No. of physically based model	No. of conceptual model	Model combination [*]
2	1	1	13, 14, 15, 16, 17, 18, 23, 24, 25, 26, 27, 28
3	1	2	134, 135, 136, 137, 138, 145, 146, 147, 148, 156, 157, 158, 167, 168, 178, 234, 235, 236, 237, 238, 245, 246, 247, 248, 256, 257, 258, 267, 268, 278
3	2	1	123, 124, 125, 126, 127, 128
4	1	3	1345, 1346, 1347, 1348, 1356, 1357, 1358, 1367, 1368, 1378, 1456, 1457, 1458, 1467, 1468, 1478, 1567, 1568, 1578, 1678, 2345, 2346, 2347, 2348, 2356, 2357, 2358, 2367, 2368, 2378, 2456, 2457, 2458, 2467, 2468, 2478, 2567, 2568, 2578, 2678
4	2	2	1234, 1235, 1236, 1237, 1238, 1245, 1246, 1247, 1248, 1256, 1257, 1258, 1267, 1268, 1278
5	1	4	13456, 13457, 13458, 13467, 13468, 13478, 13567, 13568, 13578, 13678, 14567, 14568, 14578, 14678, 15678, 23456, 23457, 23458, 23467, 23468, 23478, 23567, 23568, 23578, 23678, 24567, 24568, 24578, 24678, 25678
5	2	3	12345, 12346, 12347, 12348, 12356, 12357, 12358, 12367, 12368, 12378, 12456, 12457, 12458, 12467, 12468, 12478, 12567, 12568, 12578, 12678
6	1	5	134567, 134568, 134578, 134678, 135678, 145678, 234567, 234568, 234578, 234678, 235678, 245678
6	2	4	123456, 123457, 123458, 123467, 123468, 123478, 123567, 123568, 123578, 123678, 124567, 124568, 124578, 124678, 125678
7	1	6	1345678, 2345678
7	2	5	1234567, 1234568, 1234578, 1234678, 1235678, 1245678
8	2	6	12345678

^{*} *MIKE SHE, SWAT, HEC-HMS, TANK, AWBM, SIMHYD, SACREMANO and SMAR are denoted by numbers 1, 2, 3, 4, 5, 6, 7 and 8, respectively.*

2.3.2 Contingency Table

Contingency table is a complete representation of joint probability distribution of simulated and observed discharges. This table can be constructed by grouping the simulations and observations of the event in N bins in ascending order. However, ten equi-probable bins of probability interval 0.1 are usually taken in practice (Hamill, 1997; Sahai et al., 2008). The lower probability limit for first bin is zero, while the upper limit for the last bin is one. The threshold discharge for each bin is calculated

by frequency analysis of the observed discharge using the Weibull plotting position method. The probability of an event Q not exceeding a given threshold Q_t is given by

$$P(Q \leq Q_t) = 1 - \frac{x}{X+1} \quad (15)$$

where, Q_t = threshold discharge or maximum discharge limit for different bins, x = rank of Q_t in considered discharge, and X = total number of observations considered for the frequency analysis.

Hit rate (HR), representing the proportion of the simulated as well as observed events, and false alarm rate (FAR), representing the proportion of simulated but not observed events, are calculated for each probability bin. Computation of HR and FAR includes the summation of observed occurrences (O), i.e., number of observed values (discharge) which is simulated by the model, and non-occurrences (NO), i.e., number of observed values (discharge) which is not simulated by the model for each probability bin. The HR and FAR for k^{th} bin is defined as:

$$HR_k = \frac{\sum_{l=k}^N O_l}{\sum_{l=1}^N O_l} \quad (16)$$

$$FAR_k = \frac{\sum_{l=k}^N NO_l}{\sum_{l=1}^N NO_l} \quad (17)$$

where, N = total number of bins, usually taken as ten.

2.3.3 Categorical Evaluation of Ensembles

The relative operating characteristic (ROC) curve is a representation of skill of a prediction/simulation system in which the hit rate and the false-alarm rate are compared. The ROC curve was first introduced into the meteorological literature by Mason (Mason, 1982), although it has a longer history of use in disciplines of psychology and medicine. Multi categories contingency table is constructed for a set of simulated values that can range from 0%, representing less than the lowest possible

discharge, to 100%, indicating higher than the highest possible discharge. A point in the ROC curve is defined by the FAR on the x-axis and HR on the y-axis. Fig. 14 presents a typical ROC curve. The ROC curve of a perfect ensemble connects O(0,0), A(0,1) and B(1,1), and that of a no-skill ensemble connects O(0,0) and B(1,1) along diagonal (Wilks, 1995). The upper left corner of the ROC-diagram represents a perfect forecast system where there are no false alarms and only hits (HR=1 and FAR=0). There are number of indices for summarizing the performance of a simulation/forecast (Mason, 1982). However, the area under the ROC curve is the most commonly used index. ROC area corresponding to more skillful simulation/forecast is higher compared to less skillful simulation/forecast. The area under the ROC curve of all the ensembles are computed numerically using trapezoidal rule. The ROC area corresponding to a perfect system is one (joining points O, A and B) whereas, that of no skill system is less than 0.5 (below the diagonal). The details of ROC curve, and its construction and interpretation, may be referred from Wilks (1995).

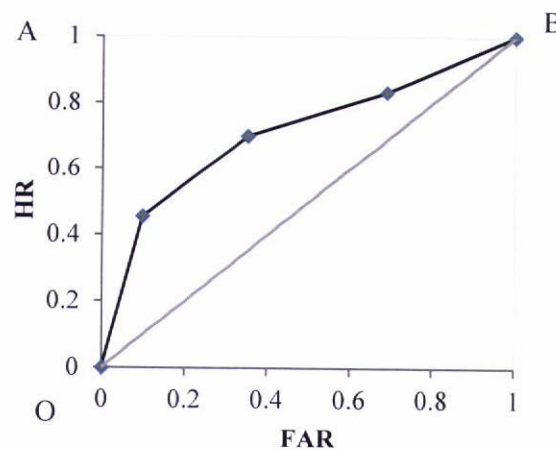


Fig. 14: Relative Operating Characteristic (ROC) curve of a typical ensemble

2.3.4 Temporal Evaluation of Ensemble

Two of the most basic quantities used to verify an ensemble system are ensemble dispersion, or spread, and ensemble skill of the control. Large ensemble spread corresponds to relatively uncatchable situation whereas small spread corresponds to relatively highly catchable situation. Spread is defined as the distance function of n^{th} order between ensemble and simulated quantities, whereas skill is defined as the distance function of n^{th} order between ensemble and observed quantity (Buizza and

Palmer, 1998). Some of the studies related to atmospheric sciences reveal that distance function of 2^{nd} order may be used to define ensemble spread and skill (Barker, 1991). Hence, daily skill can be defined as the root mean square distance (RMSD) between daily discharge simulated by ensemble and observed daily discharge, i.e., the absolute difference between daily discharge simulated by ensemble and observed daily discharge, whereas spread as the RMSD between daily discharge simulated by ensemble and the discharge simulated by different models (Barker, 1991). Spread-skill relationships can be characterized by number of cases, in which ensemble skill is less than spread, i.e., ensemble is providing advantage over spread (Buizza and Palmer, 1998). We call this number of cases as number of skillful days (N_d) due to consideration of daily time step in this study. This index of spread-skill relationship is used as temporal evaluation criteria for ensembles.

2.3.5 Selection of the Best Ensemble

Developed ensembles may be evaluated using scatter plot of ROC area and N_d for combined categorical and temporal evaluation. The upper right corner of scatter plot corresponds to both maximum number of skillful days and maximum ROC area. Hence, the upper right ensemble in scatter plot is the best performing ensemble from both categorical and temporal evaluation point of view. However, the best ensemble must perform consistently better for the validation period as well. Hence, the ensemble whose performance is good in both calibration and validation periods may be considered as good ensemble. This is not possible using scatter plot only. Hence, an analytical procedure is developed in which both the number of skillful days and ROC area is normalized separately for all ensembles such that both become unitless. ROC area and N_d are normalized as follows:

$$ROC_{Norm} = \frac{ROC - ROC_{mean}}{ROC_{SD}} \quad (18)$$

$$N_{d,Norm} = \frac{N_d - N_{d,mean}}{N_{d,SD}} \quad (19)$$

where, ROC_{Norm} = normalized ROC area of individual ensemble, ROC = original ROC area of individual ensemble, ROC_{mean} = mean ROC area of all ensembles, ROC_{SD} = standard deviation of ROC area of all ensembles, $N_{d,Norm}$ = normalized N_d of

individual ensemble, N_d = original N_d of individual ensemble, $N_{d,mean}$ = mean N_d of all ensembles and $N_{d,SD}$ = standard deviation of N_d of all ensembles.

In order to identify the ensemble, which performs better in both the calibration and validation periods, a *SCORE* (S) is introduced as the sum of normalized ROC area and normalized number of skillful days. This *SCORE* is an indicator of distance from origin to the ensemble in scatter plot. An ensemble is supposed to perform the best if it follows two criteria: *SCORE* has to be higher, and 'location' has to be in the upper right corner on scatter plot. The developed ensembles were ranked according to their performance using developed score for both calibration and validation periods separately, and the ensemble performing consistently during calibration and validation, i.e., having minimum deviation in performance from calibration to validation, was chosen as the best.

2.4 Uncertainty Analysis of Ensembles and Models

Quantile regression technique, a stochastic approach, was used to assess the uncertainty resulting from all sources collectively (Koenker and Bassett, 1978). This methodology has been widely adopted to estimate the uncertainty of deterministic forecasts (Weerts et al., 2011). This methodology was used to quantify the uncertainty in the discharge estimated using different ensembles and models. The observed daily discharge may be expressed mathematically in terms of ensemble discharge and residual as:

$$Q(t) = \hat{Q}(t) + e(t) \quad (20)$$

where $Q(t)$ = observed daily discharge, $\hat{Q}(t)$ = ensemble discharge, and $e(t)$ = residual.

The method assumes a functional relationship between residuals and estimates in Gaussian domain, i.e., normalized quantile discharge (NQD) and normalized quantile residual (NQR). Linear relation between NQD and NQR is supported by existing literature (Koenker and Hallock, 2001; Weerts et al., 2011). Hence, NQR may be expressed as:

$$NQR = a \times NQD + b \quad (21)$$

Different quantile regression lines may be obtained by minimizing the absolute bias by assigning different weights to positive and negative residuals in Gaussian domain. Absolute bias may be considered for this purpose as objective function which is expressed mathematically as:

$$\text{Min} \sum \rho_{\tau} [|NQR - (a \times NQD + b)|] \quad (22)$$

where, a = slope, b = intercept, and ρ_{τ} = quantile regression function which pushes the regression line to desired location.

To estimate the discharge corresponding to a given confidence limit, the ensemble discharge is transformed to Gaussian domain as NQD first, and then error in Gaussian domain, NQR is estimated using the regression line indicated by Eq. 21. The estimated error, NQR is transformed back to original domain using pre-estimated mean and standard deviation of residual. Finally, the estimated residual is added to daily ensemble discharge to obtain the discharge which includes uncertainty.

Five regression lines were used to analyze the uncertainty in ensemble discharge: two for maximum and minimum limits of 90% confidence band, two for 50% confidence band, and one for median. The slope and intercept of these lines were estimated by Eq. 22 using the calibration period data. Furthermore, to verify the correctness of error models, the models were applied on the ensemble discharge during both calibration and validation periods.

2.5 Uncertainty Comparison of Ensembles and Models

To evaluate the discharge simulated by different models and ensembles, an index, Uncertainty, which is the area between upper and lower limit of uncertainty band was proposed to be a measure of uncertainty. It was assumed that lower the Uncertainty, thinner will be the uncertainty band, resulting in lower level of uncertainty in the estimate of river discharge. Similarly, higher the Uncertainty, wider will be the band, resulting in higher level of uncertainty. UC_Area was estimated numerically using following equation:

$$UC_{Area} = \frac{(Q_1 + Q_n)}{2} \times \sum_{i=2}^{n-1} Q_i \quad (23)$$

where, Q_1 , Q_n and Q_i is the simulated daily discharge of first day, last day and i^{th} day of considered period.

3 Results and Discussion

3.1 Model Calibration

Select models are calibrated for the period of 1/6/2004 to 31/5/2007. The performance of the model for simulating surface runoff was evaluated using Nash Sutcliffe efficiency (NSE) and bias as performance index (Table 2). The values of Nash Sutcliffe efficiency in Table 2 indicate that runoff is better simulated by conceptual models than physically based models for both catchments. Time series of model predictions for calibration period is presented in Fig. 15 for Kesinga and in Fig. 16 for Salebhata.

Table 2: Bias and NSE of models during calibration period

Models	Kesinga		Salebhata	
	Calibration		Calibration	
	Bias (%)	NSE	Bias (%)	NSE
MIKE SHE	-22.62	0.72	0.26	0.78
SWAT	-24.38	0.76	29.75	0.71
HEC HMS	45.49	0.42	20.02	0.67
AWBM	-28.04	0.80	-17.27	0.90
SACRAMENTO	-11.77	0.89	-10.23	0.90
SIMHYD	-16.04	0.89	6.32	0.86
SMAR	-40.80	0.76	-45.73	0.83
TANK	-2.71	0.88	0.33	0.82

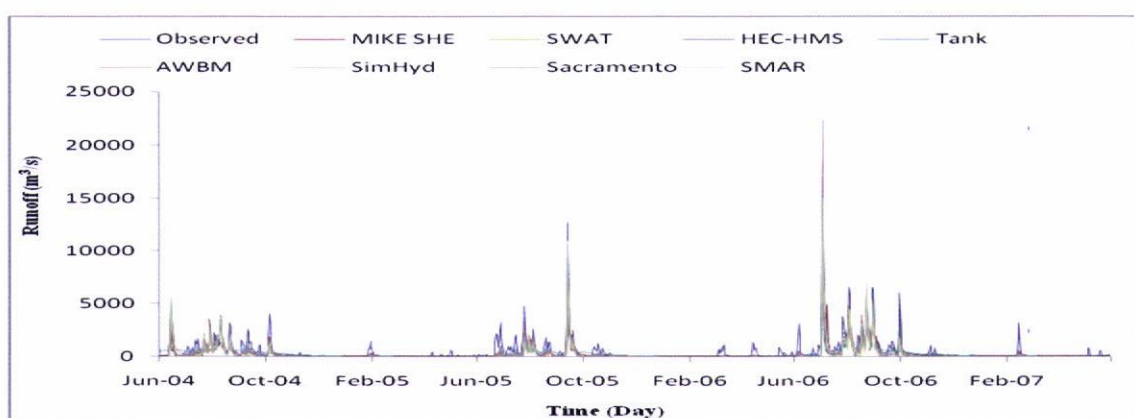


Fig. 15: Observed and simulated streamflow during calibration for Kesinga

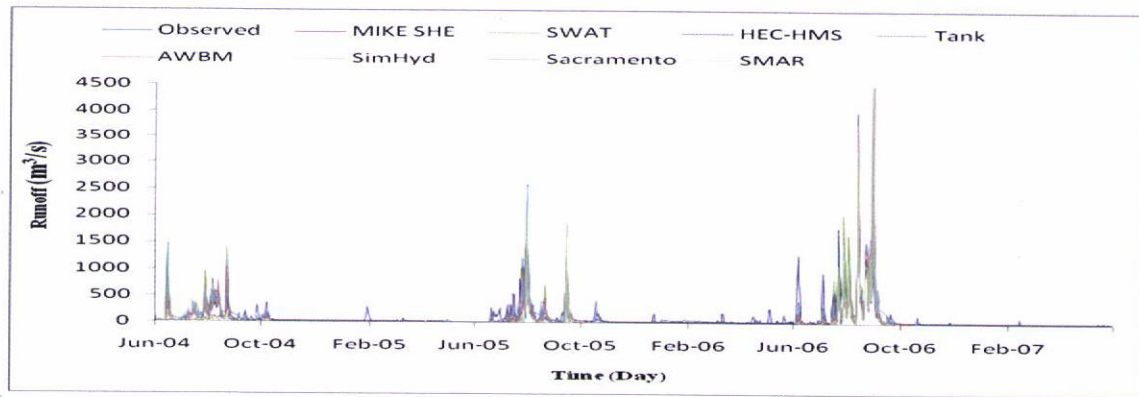


Fig. 16: Observed and simulated streamflow during calibration for Salebhata

3.2 Model Validation

All the calibrated models were simulated for the period 1/6/2007 to 31/5/2009 by keeping the parameters of model unaltered. Figs. 17 and 18 present the simulated runoff during validation period for Kesing and Salebhata, respectively. Table 3 presents NSE and bias of different models during the validation period. The obtained values of NSE indicate satisfactory performance of models.

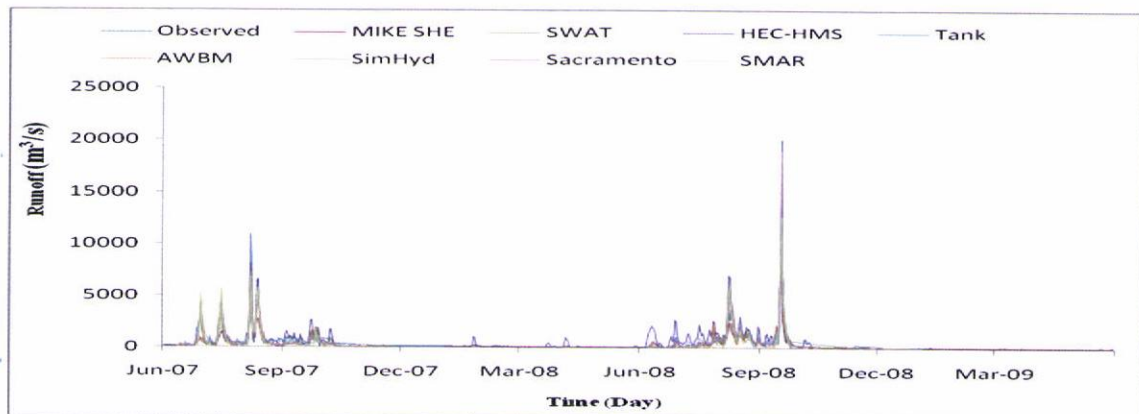


Fig. 17: Observed and simulated streamflow during validation for Kesing.

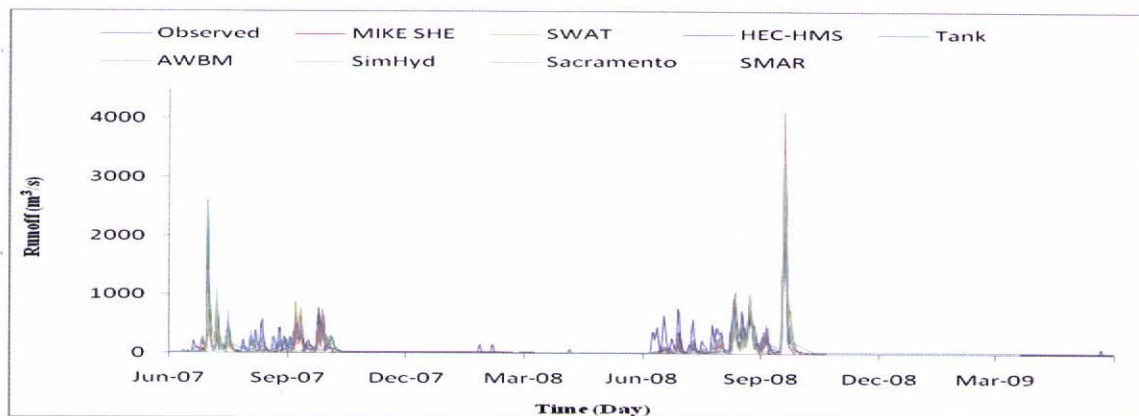


Fig. 18: Observed and simulated streamflow during validation for Salebhata

Table 3: Bias and NSE of models during validation period

Models	Kesinga		Salebhata	
	Validation		Validation	
	Bias (%)	NSE	Bias (%)	NSE
MIKE SHE	-34.15	0.64	8.60	0.78
SWAT	-19.08	0.77	32.29	0.83
HEC HMS	13.80	0.70	25.05	0.61
AWBM	-26.41	0.77	-20.52	0.87
SACRAMENTO	-12.98	0.84	-12.69	0.82
SIMHYD	-15.58	0.83	7.22	0.82
SMAR	-42.20	0.76	-46.47	0.75
TANK	-10.76	0.83	0.41	0.82

Scatter plot for the two performance statistics (NSE and bias) for each of the models is shown in the Fig. 19 for Kesinga and in Fig. 20 for Salebhata. Statistically the best models are those with efficiencies approaching 1.0 and biases near 0%.

From Fig. 19 it clear that seven of the eight models are under-predicting (negative bias) with absolute bias as high as -28.04% for AWBM, with only HEC-HMS resulting in over-prediction (45.49 %) for calibration period for Kesinga. Among all the models, TANK is giving optimum results with the bias just near 0% followed by SACRAMENTO, and both of them have high NSE efficiency too compared to other models. When predictions in validation period are assessed, the relative positions of the models remain largely unchanged. Two of the models, i.e., HEC-HMS and SWAT have increased efficiencies in the validation periods. NSE values are varying from 0.42 to 0.89.

Similar analysis was done for Salebhata as presented in Fig. 20 which reveals that during calibration period most of the models are over-predicting (positive bias) except for SACRAMENTO, SMAR and AWBM, with over-prediction as high as 29.75% for SWAT and under-prediction as high as -45.73% for SMAR. Among the models, TANK and MIKE SHE are giving bias near 0%, thus giving optimum prediction. When predictions in validation period are assessed, the relative positions of the models remain largely unchanged; TANK and MIKE SHE are showing little change in results for both calibration and validation while for HEC-HMS and SWAT

validation efficiency is better than calibration. NSE values are varying from 0.61 to 0.90.

It is observed that in both cases lumped conceptual models have higher efficiency (high NSE value) than the physically based models. All the models are performing better in case of Salebhata than in Kesinga because of better correlation between rainfall-runoff data in Salebhata.

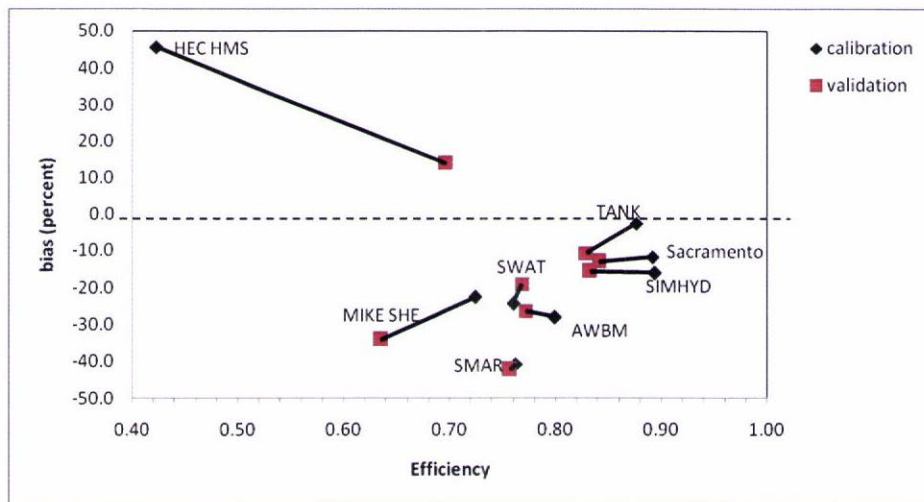


Fig. 19: Bias and NSE of eight models for both calibration and validation for Kesinga

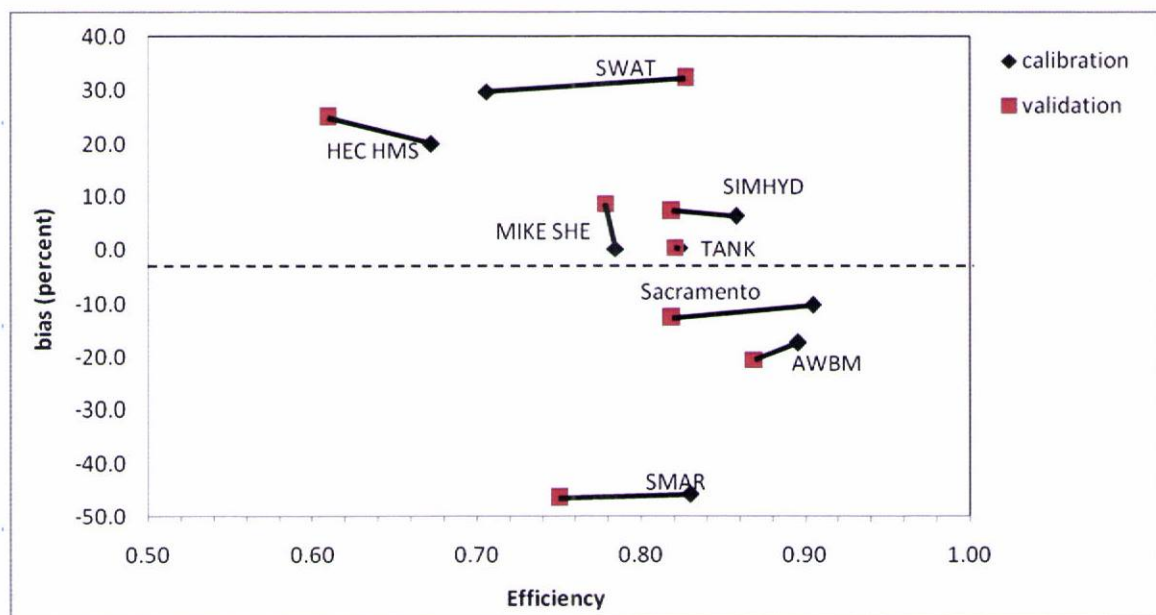


Fig. 20: Bias and NSE of eight models for both calibration and validation for Salebhata

3.3 Identification of the Best Multi-Model Ensemble Method

Multi-model ensemble of all eight models were developed using eight ensemble methods listed in section 2.2.3 considering all models at a time for both the catchments. Eq. 2 and 14 were employed to estimate correlation coefficient R and root mean square error (RMSE) for all the eight methods using the ensemble simulated and observed discharges. Fig. 21 presents the scatter plot of RMSE and R for different ensemble methods in case of Kesinga. Each method is denoted by two points in Fig. 21: one corresponding to calibration and the other validation. RMSE for different methods varies from $262 \text{ m}^3/\text{s}$ to $328 \text{ m}^3/\text{s}$ during the calibration period and from $412 \text{ m}^3/\text{s}$ to $500 \text{ m}^3/\text{s}$ during the validation period. The correlation coefficient, R , varies from 0.93 to 0.96 during the calibration period and from 0.91 to 0.93 during the validation period. Fig. 21 shows that MMSE method results in the lowest RMSE and the highest correlation coefficient during the calibration period, however its performance deteriorates during the validation period. The Weighted average method (WAM_K_1.5) method, on the other hand, performs satisfactorily during both calibration and validation periods. Hence, WAM_K_1.5 was chosen as the best performing ensemble method for Kesinga.

Similarly, Fig. 22 presents the scatter plot of RMSE and R of different ensembles resulting from different ensemble methods during calibration and validation periods in case of Salebhata catchment. RMSE for different methods vary from $75 \text{ m}^3/\text{s}$ to $90 \text{ m}^3/\text{s}$ during the calibration period and from $72 \text{ m}^3/\text{s}$ to $77 \text{ m}^3/\text{s}$ during the validation period. The correlation coefficient varies from 0.94 to 0.96 during the calibration period and 0.94 to 0.95 during the validation period. Fig. 22 shows that MMSE method results in lowest RMSE and highest correlation during the calibration period, whereas, the RMSE using this method is very high during the validation period. However, linear programming method (LP), on the other hand, performs satisfactorily during both calibration and validation periods. Hence, LP was chosen as the best performing ensemble method for Salebhata.

The results show that though the multi-model super ensemble (MMSE), recommended earlier as the superior method (Ajami et al., 2006), performed well during the calibration period, its performance deteriorated during the validation

period. This shows that an ensemble method must be checked for its performance during both calibration and validation before its application.

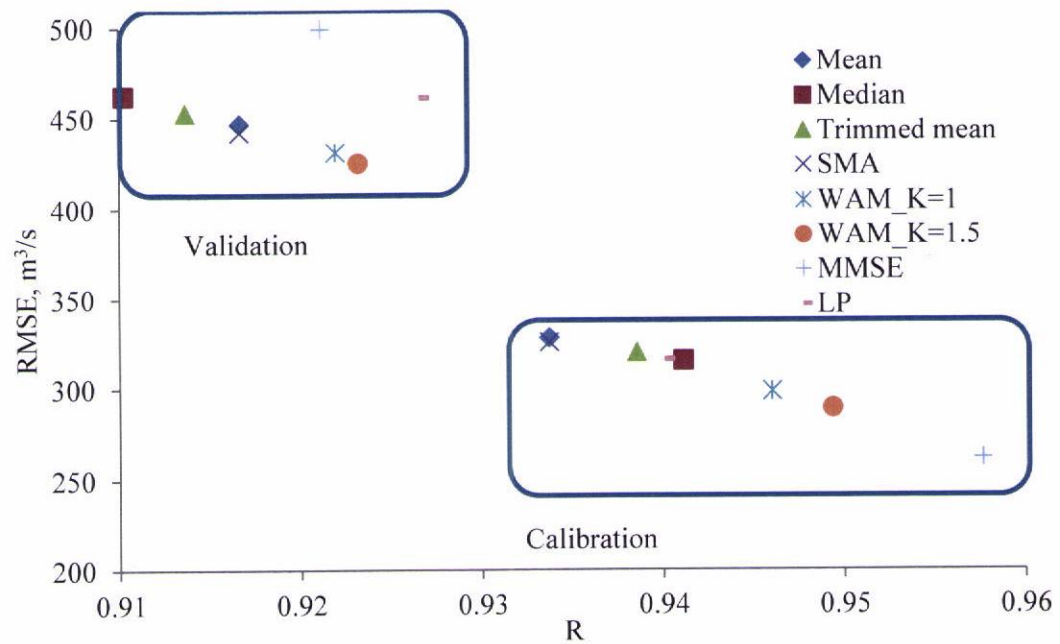


Fig. 21: Scatter plot of root mean square error (RMSE) and correlation coefficient R for different ensemble methods in case of Kesinga

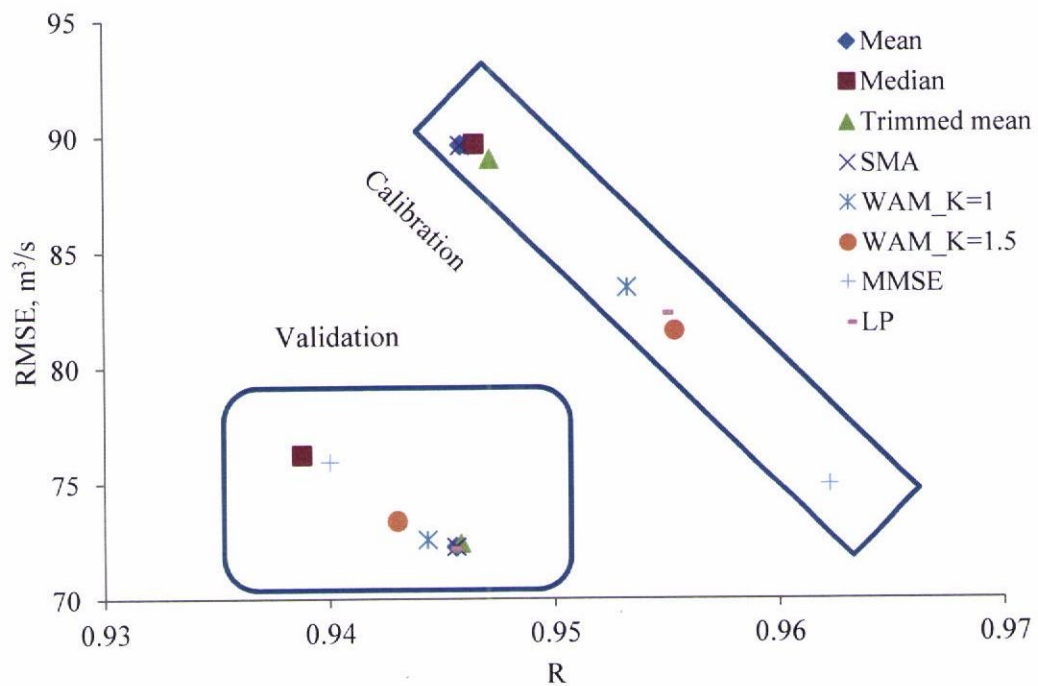


Fig. 22: Scatter plot of root mean square error (RMSE) and correlation coefficient R for different ensemble methods in case of Salebhata

3.4 Identification of the Best Ensemble

As described in section 2.3.1, 189 possible ensembles were developed using the weighted average method (WAM_K_1.5) and the linear programming (LP) method, for Kesinga and Salebhata, respectively. Weights of different members in an ensemble were estimated using the data of calibration period only. The estimated weights were then used for estimation of corresponding ensemble discharge during validation period. In order to construct multi-category contingency table, frequency analysis of the observed discharge during 1st Jun 2004 to 31st May 2009 was performed using the Weibull plotting position method for both catchments. Figs. 23 and 24 present the frequency distribution of the observed discharge, which was used to identify discharges corresponding to different levels of probability for Kesinga and Salebhata, respectively. These discharge levels were subsequently used to define the upper and lower limits of ten equi-probable bins in case of both catchments. Tables 4 and 5 present the upper and lower discharge, i.e., threshold discharge, corresponding to different bins for Kesinga and Salebhata, respectively.

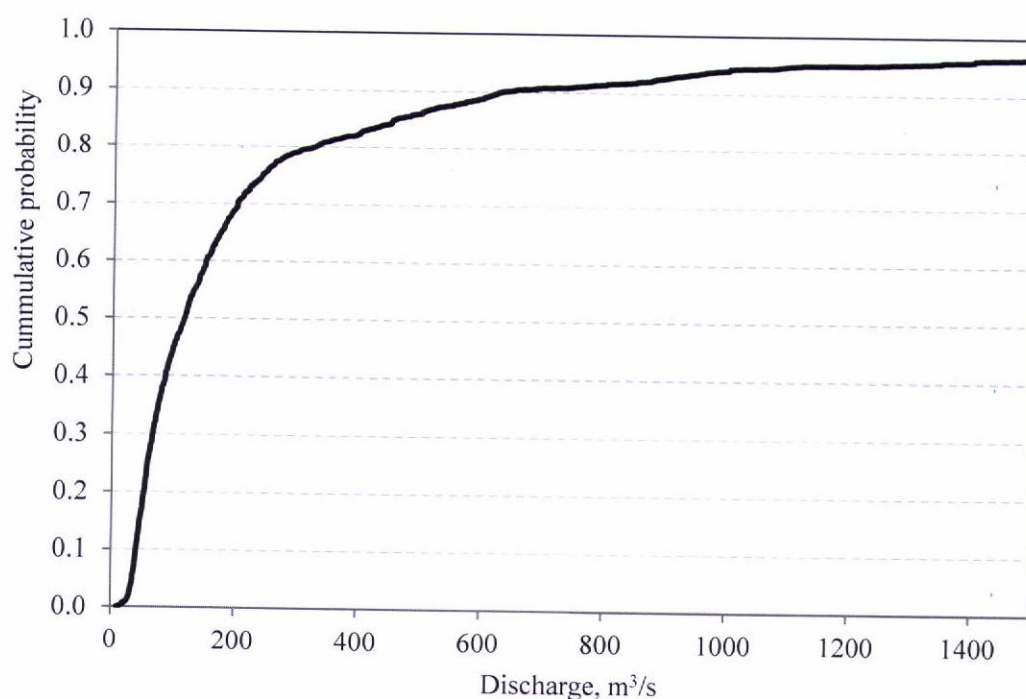


Fig. 23: Frequency distribution of observed discharge for Kesinga

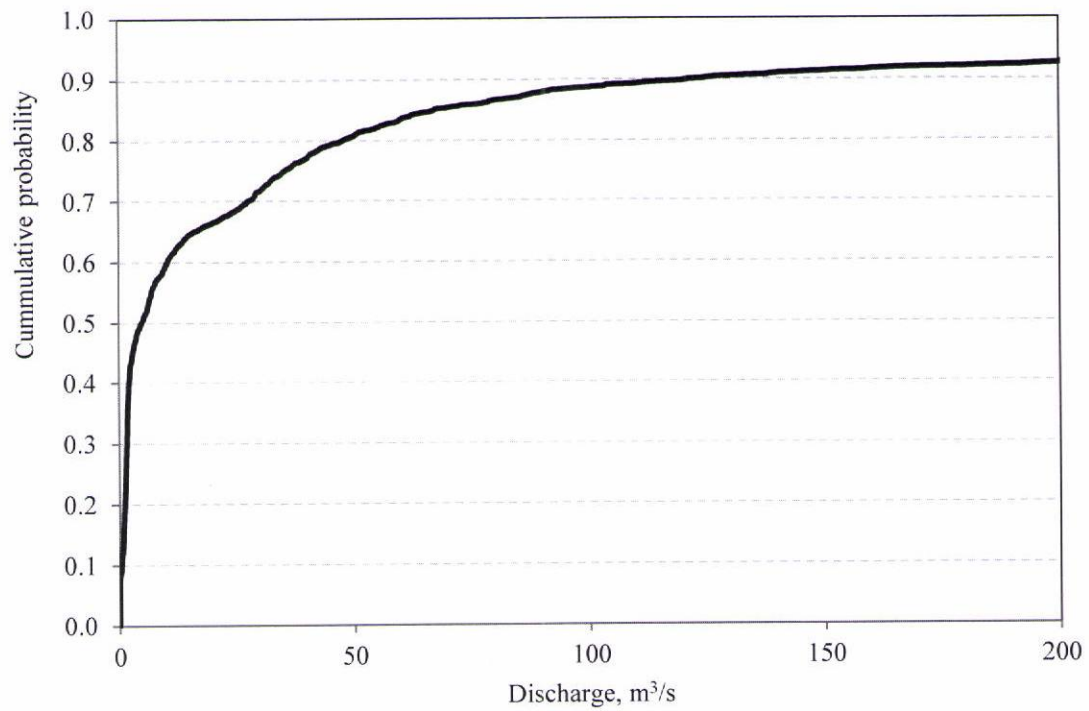


Fig. 24: Frequency distribution of observed discharge for Salebhata

Table 4: Thershold discharge for different bins for Kesinga

Bin	Probability range		Discharge range	
	Lower	Upper	Lower (m ³ /s)	Upper (m ³ /s)
1	0.0	0.1	0.00	40.32
2	0.1	0.2	40.32	52.68
3	0.2	0.3	52.68	65.85
4	0.3	0.4	65.85	85.52
5	0.4	0.5	85.52	115.96
6	0.5	0.6	115.96	150.14
7	0.6	0.7	150.14	200.00
8	0.7	0.8	200.00	325.77
9	0.8	0.9	325.77	630.18
10	0.9	1.0	630.18	inf

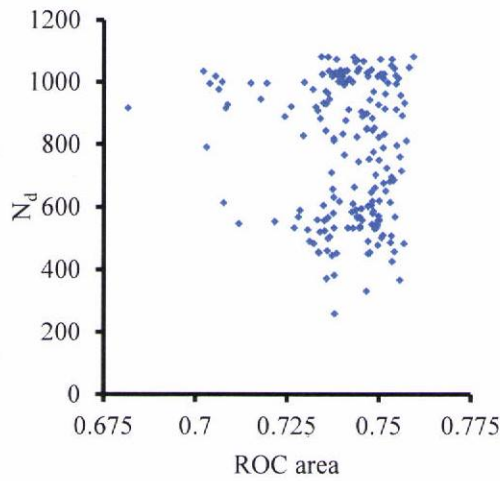
Table 5: Threshold discharge for different bins for Salebhata

Bin	Probability range		Discharge range	
	Lower	Upper	Lower (m ³ /s)	Upper (m ³ /s)
1	0.0	0.1	0	0.37
2	0.1	0.2	0.37	1.15
3	0.2	0.3	1.15	1.62
4	0.3	0.4	1.62	2.12
5	0.4	0.5	2.12	4.85
6	0.5	0.6	4.85	10.44
7	0.6	0.7	10.44	27.67
8	0.7	0.8	27.68	47.96
9	0.8	0.9	47.96	123.79
10	0.9	1.0	123.79	inf

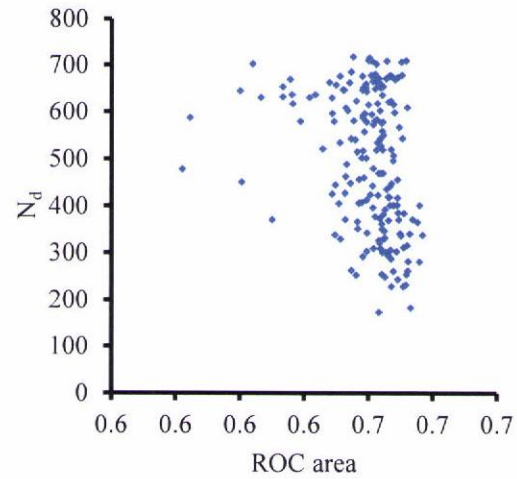
3.4.1 Categorical and Temporal Performance

Once the bins were defined, these were used to develop the multi category contingency table for all 189 ensembles for both calibration and validation periods for both the catchments. This was followed by development of Relative Operating Characteristic (ROC) curve and estimation of area under the curve as discussed in section 2.3.3. Number of skillful days, N_d was also estimated for all the ensembles, as discussed in section 2.3.4. Figs. 25a and 25b present the scatter plot of ROC Area and N_d for various ensembles (represented by dots) during the calibration and validation periods, respectively, in case of Kesinga catchment and Figs 25c and 25d, respectively, in case of Salebhata catchment. A total of 189 dots in these plots represent different ensembles. The values of ROC area and N_d were further used to estimate the normalized ROC area, and normalized N_d as discussed in section 2.3.5. The mean and standard deviation used for the normalization of ROC area are 0.741 and 0.0125, respectively, during the calibration period, and 0.65 and 0.015 during the validation period in case of Kesinga. The mean and standard deviation used for the normalization of N_d were 769.4 and 219.16, respectively during the calibration period, and 496.84 and 146.16 during the validation period in case of Kesinga. Similarly, in case of Salebhata, mean and standard deviation used for the normalization of ROC were 0.754 and 0.025, respectively, during the calibration period, and 0.69 and 0.049 during the validation period. The mean and standard deviation used for the normalization of N_d were 841.17 and 170.64, respectively, during the calibration

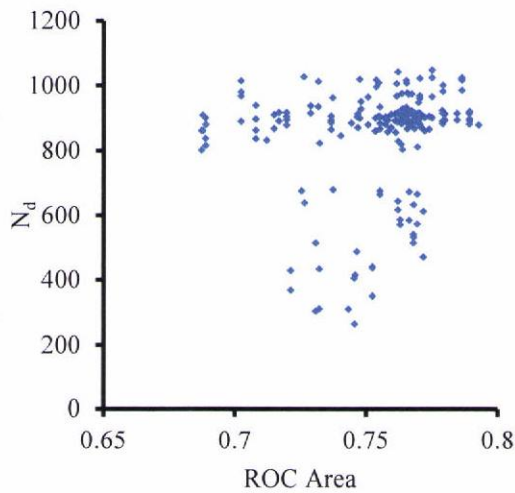
period, and 610.78 and 126.23 during the validation period in case of Salebhata. Figs. 26a and 26b present the scatter plot of normalized ROC area (Norm_ROC) and normalized N_d (Norm_ N_d) for various ensembles during the calibration and validation periods for Kesinga catchment, and Figs. 26c and 26d for Salebhata catchment. The ensembles in the upper right corner of the scatter plot represent the ensembles that have superior performance with respect to both categorical and temporal performances, i.e., higher ROC Area and higher number of skillful days.



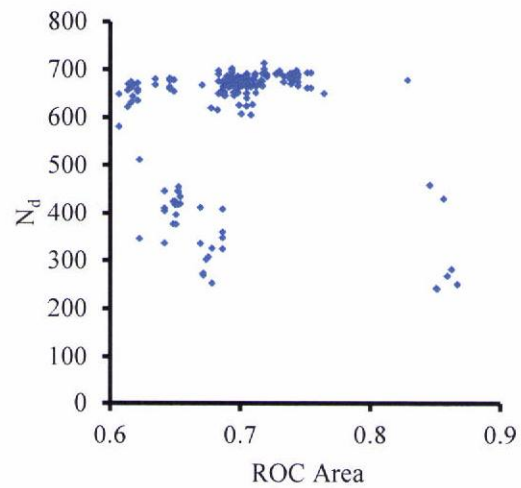
(a) Kesinga calibration



(b) Kesinga validation

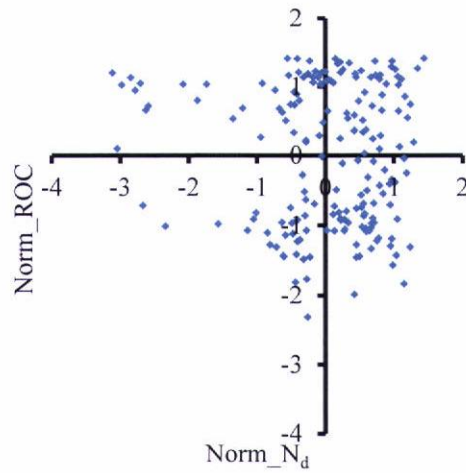


(c) Salebhata calibration

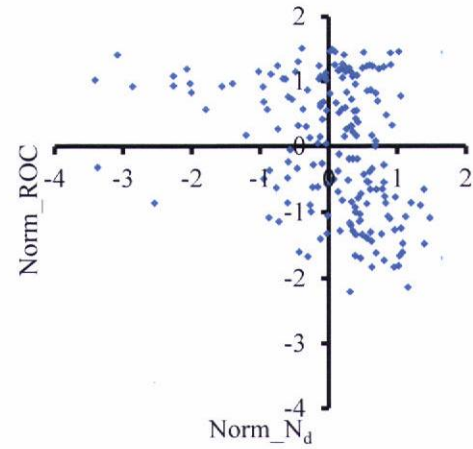


(d) Salebhata validation

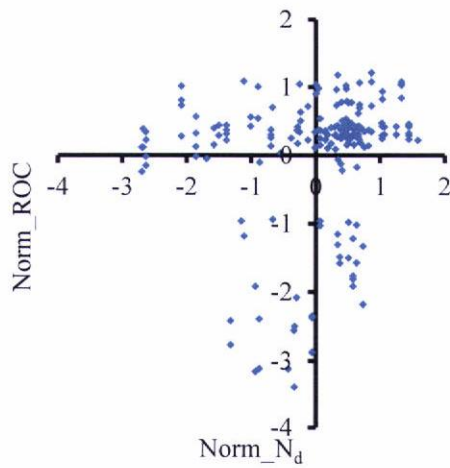
Fig. 25: Scatter plot of number of skillful days (N_d) and area under Relative operating characteristic (ROC) curve



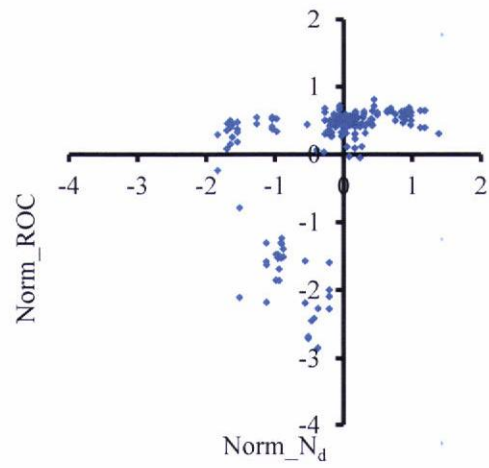
(a) Kesinga calibration



(b) Kesinga validation



(c) Salebhata calibration



(d) Salebhata validation

Fig. 26: Scatter plot of normalized number of skillful days ($\text{Norm_}N_d$) and normalized area under relative operating characteristic curve ($\text{Norm_}ROC$)

3.4.2 Ensemble selection

Normalized ROC area and N_d were used to develop SCORE as discussed in section 2.3.5, which was further used to rank the performance of different ensembles during both calibration and validation periods separately. Table 6 presents the performance of ten top performing ensembles of Kesinga during the calibration period, and their corresponding performance during the validation period. The ensembles were ranked according to their SCORE, i.e., the ensemble corresponding to highest SCORE was

Table 6: Performance of top ten ensembles during calibration period, and their corresponding performance during validation period for Kesinga

Ensemble	Calibration						Validation					
	ROC area	N _d	Norm ROC area	Norm N _d	SCORE [*]	Rank [#]	ROC area	N _d	Norm. ROC area	Norm N _d	SCORE [^]	Rank ^{\$}
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
245678	0.759	1081	1.439	1.422	2.860	1	0.653	619	0.256	0.836	1.092	36
247	0.758	1047	1.344	1.267	2.611	2	0.651	680	0.086	1.253	1.340	29
24678	0.754	1072	0.972	1.381	2.353	3	0.651	710	0.091	1.458	1.549	13
24567	0.754	1044	1.020	1.253	2.273	4	0.652	581	0.188	0.576	0.763	53
2467	0.753	1053	0.962	1.294	2.256	5	0.652	682	0.195	1.267	1.462	22
23457	0.755	1013	1.110	1.111	2.221	6	0.653	665	0.246	1.151	1.396	26
23478	0.755	1021	1.056	1.148	2.204	7	0.654	673	0.293	1.205	1.498	16
2457	0.750	1072	0.723	1.381	2.103	8	0.655	637	0.383	0.959	1.342	28
2347	0.755	995	1.066	1.029	2.096	9	0.647	648	-0.146	1.034	0.888	45
1245	0.756	958	1.162	0.860	2.022	10	0.641	609	-0.592	0.767	0.175	86

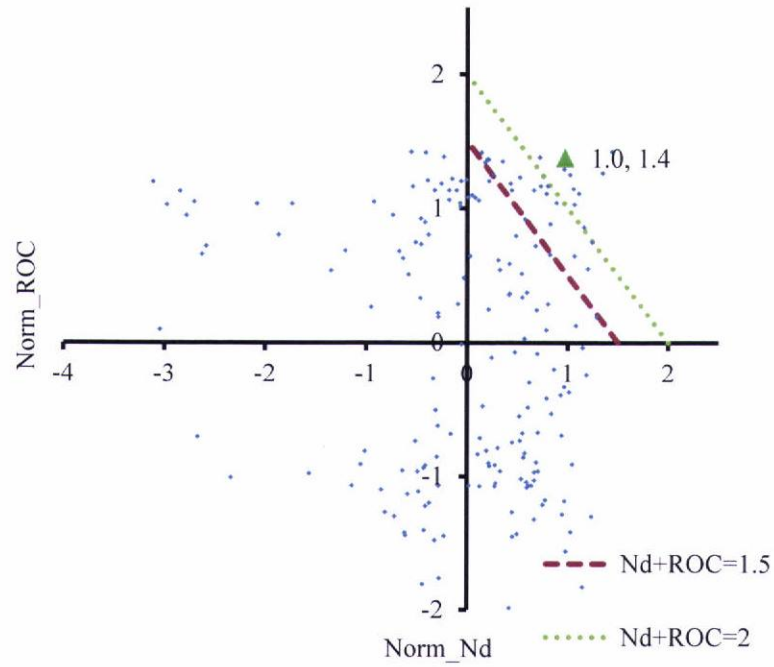
^{*} Sum of column 4 and 5 [#] Rank of ensembles during the calibration

[^] Sum of column 10 and 11 ^{\$} Rank of ensembles during the validation

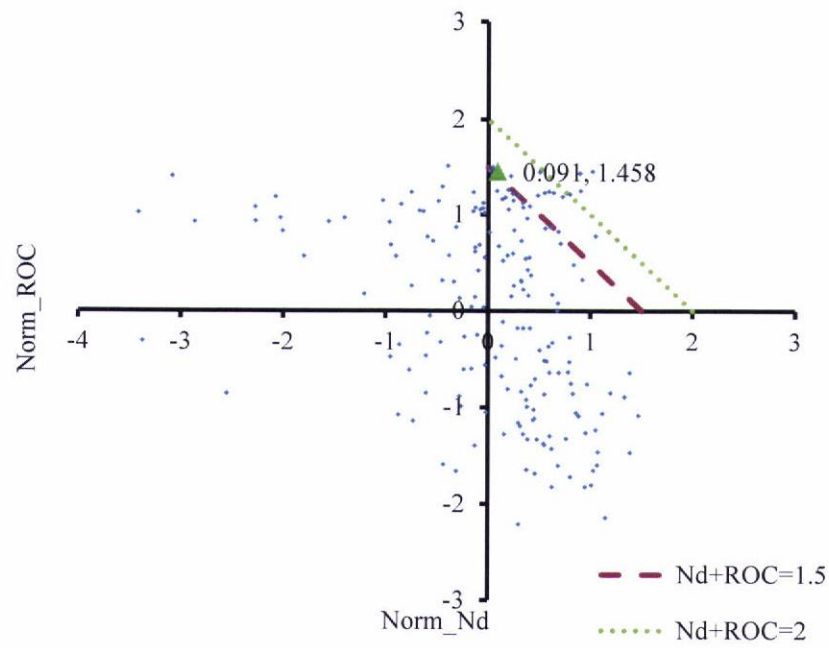
given as first rank and that of lowest SCORE was given as the last. Ranks of various top performing ensembles during calibration period were analyzed for minimum deviation from calibration to validation period. As evident from Table 6, the top performing ensemble during the calibration period, i.e., '245678' performs poorly during the validation period and obtains 36th rank. The same is true for the second best ensemble during the calibration period, i.e., '247' which goes to rank 29th during the validation period. The ensemble that performs consistently better during both calibration and validation is '24678', with ranks of 3 and 13 during the calibration and validation periods, respectively. The location of ensemble '24678' in scatter plot is also shown using a triangle in Figs. 27a and 27b in case of calibration and validation periods, respectively. As expected, this ensemble lies in the upper-right corner during the calibration (Fig. 27a). However, during validation there is a slight deviation from the upper right corner (Fig. 27b), though the ensemble still lies in between 1.5-2.0 band of SCORE. The location of the selected ensembles in Figs. 28a and 28b, depicting high SCORE values, re-establishes their superior performance. Hence, ensemble '24678' i.e., the ensemble of SWAT, TANK, SIMHYD, SACRAMENTO and SMAR, developed using weighted average method based on calibration performance, was selected as the best ensemble for Kesinga.

Similarly, Table 7 presents the performance of ten top performing ensembles of Salebhata during the calibration period, and their performance during the validation period. As evident from Table 7, the top performing ensemble during the calibration period, i.e., '24678' performs poorly during the validation period. The same is true for the second, third, fourth and fifth best ensemble during the calibration period. The ensemble that performs consistently better during both calibration and validation is '23467', with ranks of 6 and 10 during the calibration and validation periods, respectively. The location of ensemble '23467' is also shown in Figs. 28a and 28b using a triangle in case of calibration and validation, respectively. As expected, this ensemble lies in the upper-right corner relative to other ensembles in both the plots, re-establishing its superior performance. Hence, the ensemble '23467', i.e., the ensemble of SWAT, HEC-HMS, TANK, SIMHYD and SACRAMENTO, developed using linear programming method, was selected as the best ensemble for Salebhata. The results show that the best ensemble of both catchments i.e., '24678' for Kesinga and '23467' for Salebhata, have five ensemble members. This finding is in agreement

with the earlier findings that four to five models are necessary in an ensemble for obtaining the skillful results (Buizza and Palmer, 1998; Georgakakos et al., 2004; Ajami et al., 2006). The results also show that the best ensemble of Kesinga includes the best performing model for Kesinga, i.e., TANK (section 3.2), whereas, the best ensemble of Salebhata doesn't include the best performing model for Salebhata, i.e., MIKE SHE (section 3.2) based on calibration and validation performance. These findings are in agreement with earlier findings of Viney et al. (2009), i.e., the best ensemble may not necessarily contain the best individual model.



(a) Calibration



(b) Validation

Fig. 27: Relative position of the best performing ensemble '24678' for Kesiing in scatter plot of normalized values of number of skillful days (Norm_Nd) and area under Relative operating characteristic curve (Norm_ROC)

Table 7: Performance of top ten ensembles during calibration period, and their corresponding performance during validation period for Salebhata

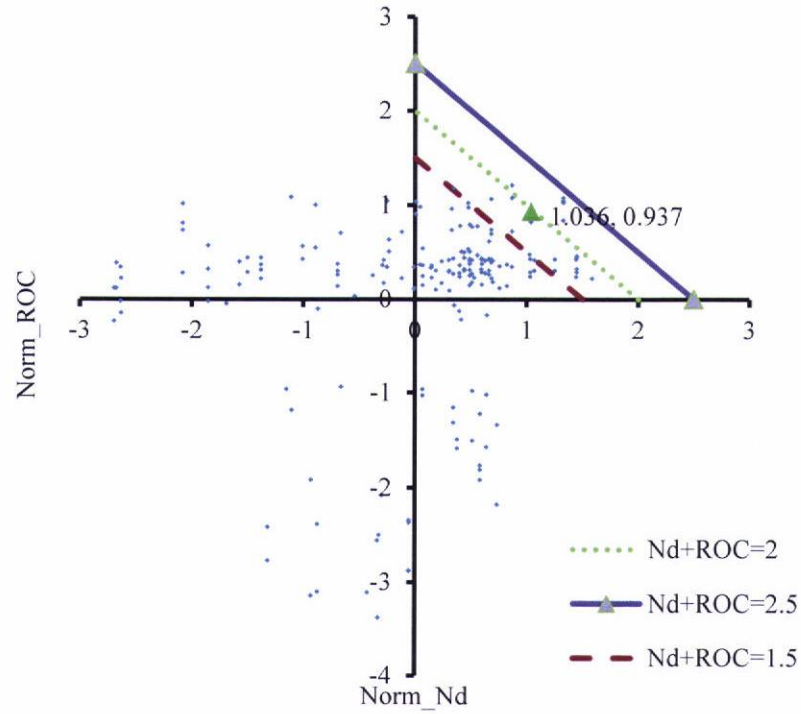
Ensemble	Calibration						Validation					
	ROC area	N _d	Norm ROC area	Norm N _d	SCORE [*]	Rank [#]	ROC area	N _d	Norm. ROC area	Norm N _d	SCORE [^]	Rank ^{\$}
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
24678	0.786	1025	1.327	1.077	2.404	1	0.693	694	-0.060	0.659	0.600	76
23478	0.786	1020	1.327	1.048	2.375	2	0.693	702	-0.060	0.723	0.663	62
234678	0.786	1017	1.327	1.030	2.357	3	0.693	696	-0.060	0.675	0.616	74
2478	0.786	985	1.327	0.843	2.170	4	0.693	700	-0.060	0.707	0.647	66
2468	0.775	1048	0.865	1.212	2.077	5	0.730	697	0.689	0.683	1.372	27
23467	0.779	1001	1.036	0.937	1.973	6	0.744	690	0.984	0.628	1.611	10
23468	0.775	1025	0.865	1.077	1.943	7	0.730	693	0.689	0.651	1.341	29
1234578	0.789	920	1.446	0.462	1.908	8	0.693	684	-0.069	0.580	0.511	87
12345678	0.789	917	1.446	0.444	1.890	9	0.693	685	-0.069	0.588	0.519	86
134578	0.789	914	1.443	0.427	1.870	10	0.689	686	-0.143	0.596	0.453	92

* Sum of column 4 and 5

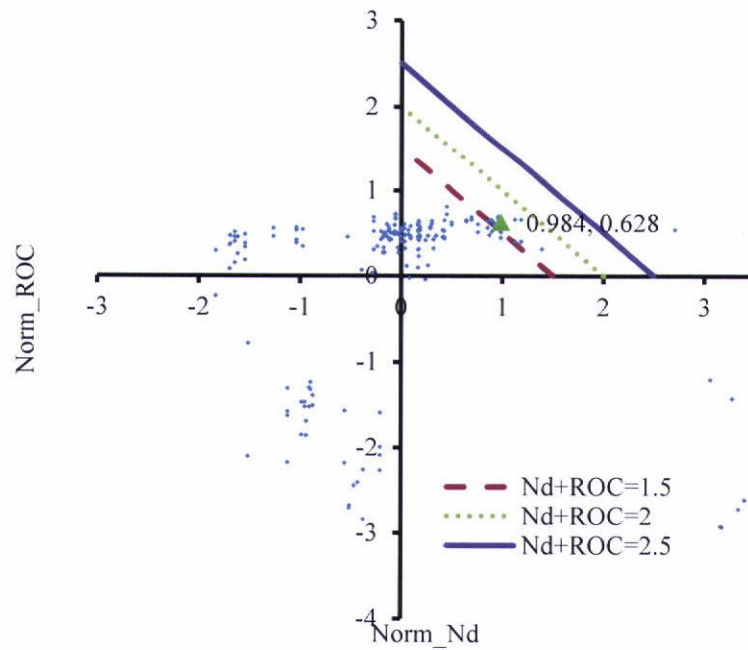
[#] Rank of ensembles during the calibration

[^] Sum of column 10 and 11

[§] Rank of ensembles during the validation



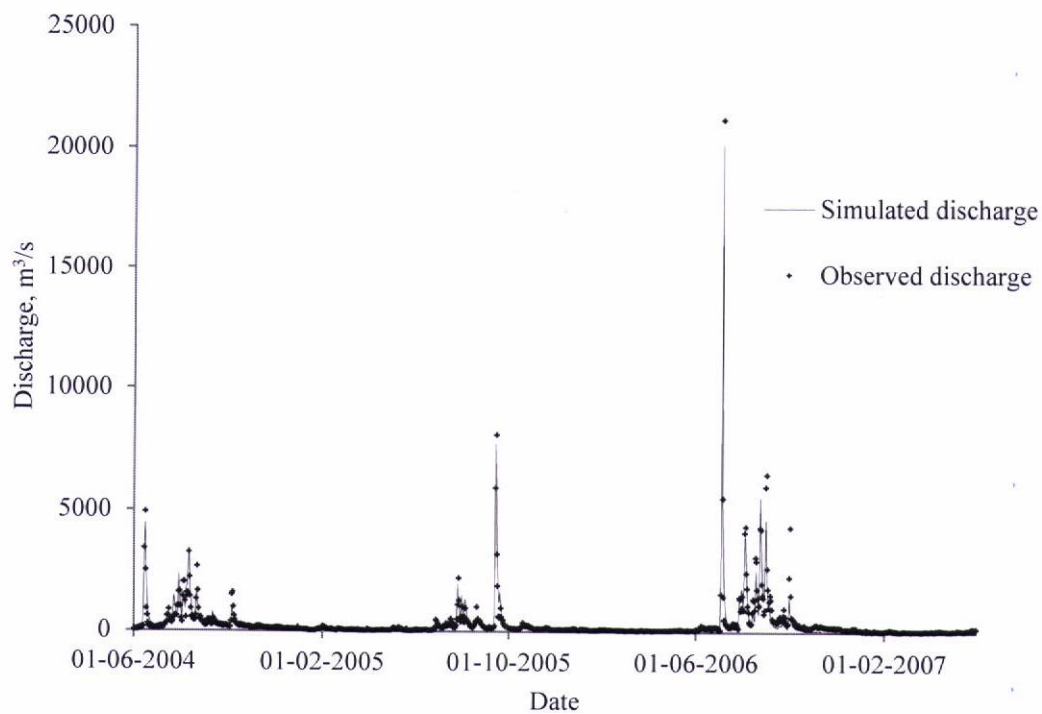
(a) Calibration



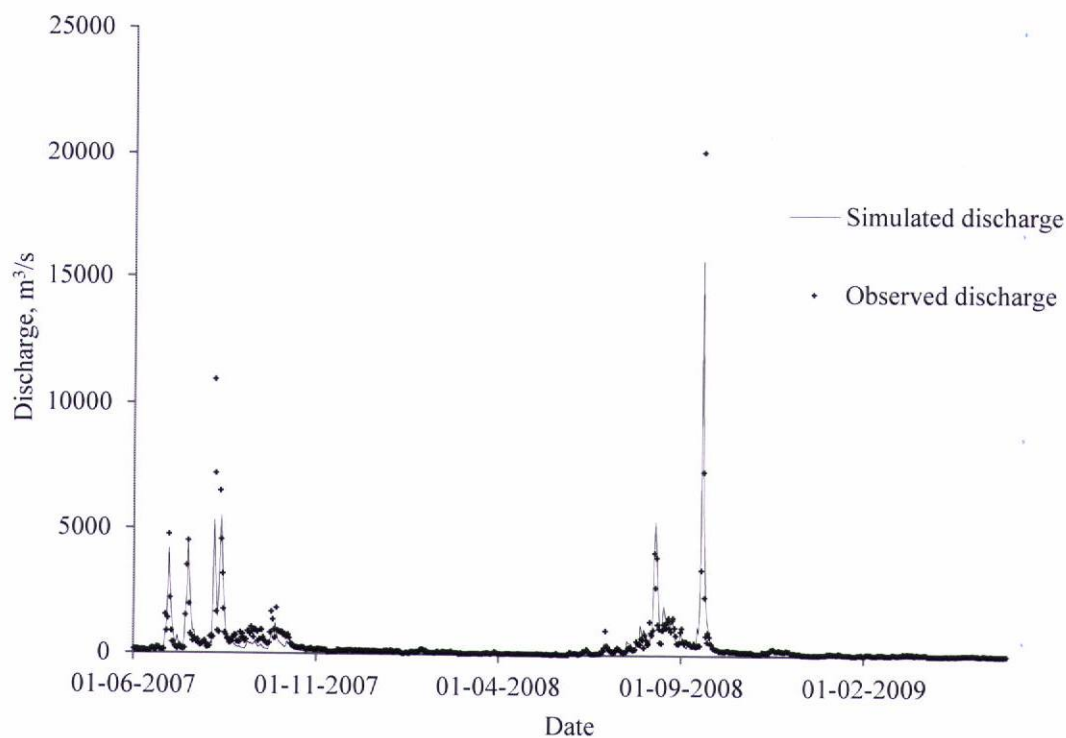
(b) Validation

Fig. 28: Relative position of the best performing ensemble '23467' for Salebhata in scatter plot of normalized values of number of skillful days (Norm_Nd) and area under Relative operating characteristic curve (Norm_ROC)

Figs. 29a and 29b present the temporal variation of observed and simulated ensemble discharges (using the best ensemble '24678') for Kesinga during the calibration and validation periods, respectively. Similarly, Figs. 30a and 30b present the temporal variation of observed and simulated ensemble discharges (using the best ensemble '23467') for Salebhata. It is evident that the selected ensembles are able to capture the observed discharges well. To reinforce this observation, the observed and simulated ensemble discharges for the calibration and validation periods are also plotted on 1:1 line in Figs. 31a and 31b for Kesinga, and Figs. 32a and 32b for Salebhata. The scatter plots show that during both calibration and validation periods, data points are spread equally on either side of the 1:1 line with minor deviation from 1:1 line.

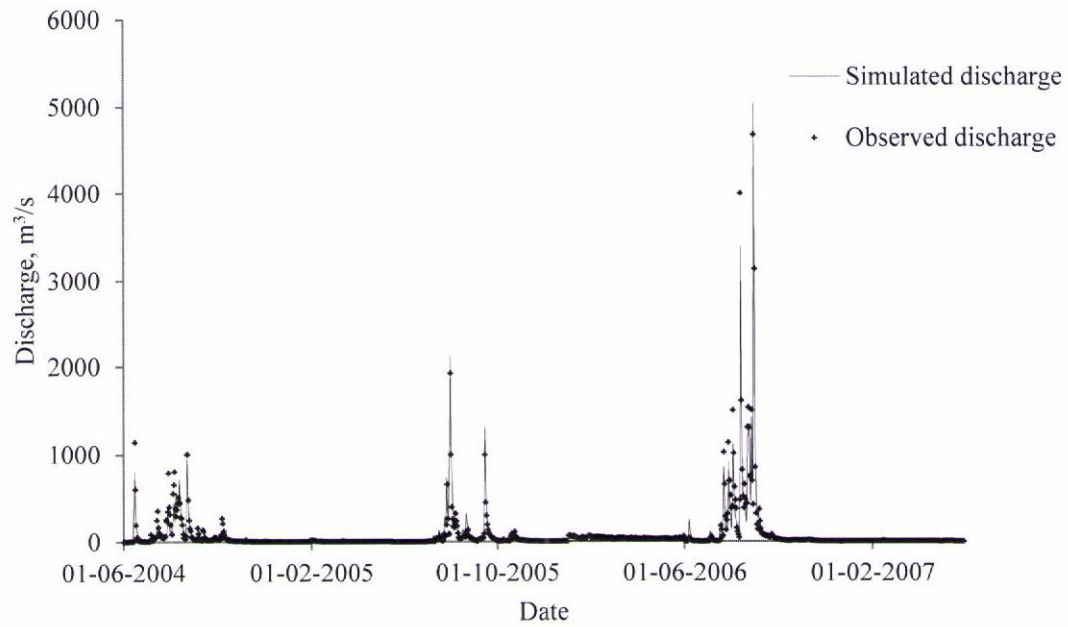


(a) calibration

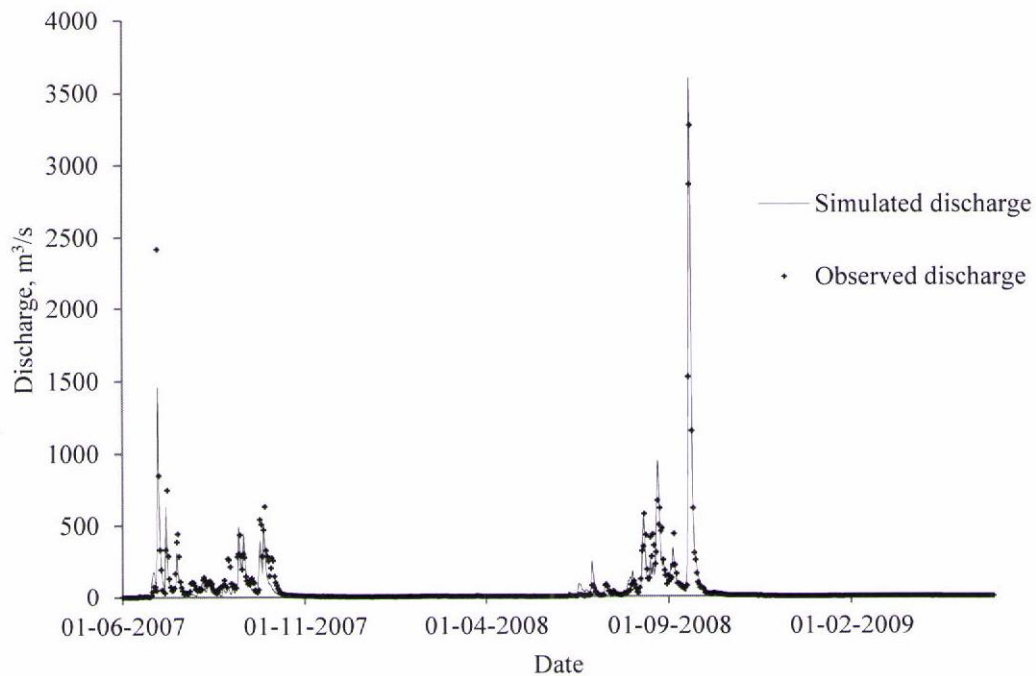


(b) Validation

Fig. 29: Observed and simulated ensemble discharges, using the best ensemble '24678' for Kesinga, during (a) calibration and (b) validation



(a) Calibration



(b) Validation

Fig. 30: Observed and simulated ensemble discharges, using the best ensemble '23467' for Salebhata, during (a) calibration and (b) validation

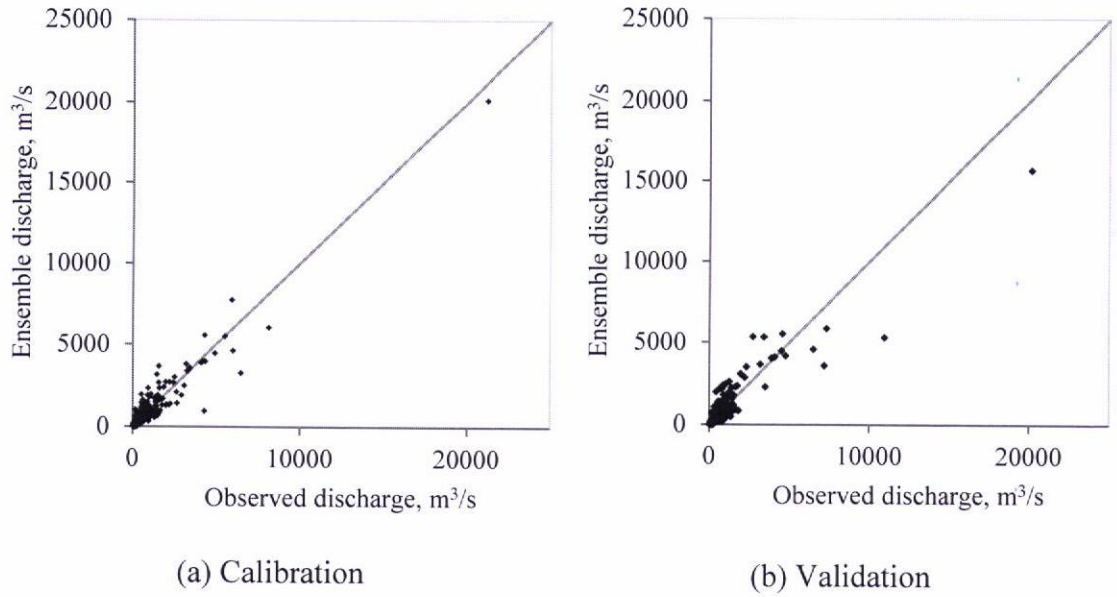


Fig. 31: Scatter plot of observed and simulated ensemble discharges for Kesinga during (a) calibration and (b) validation

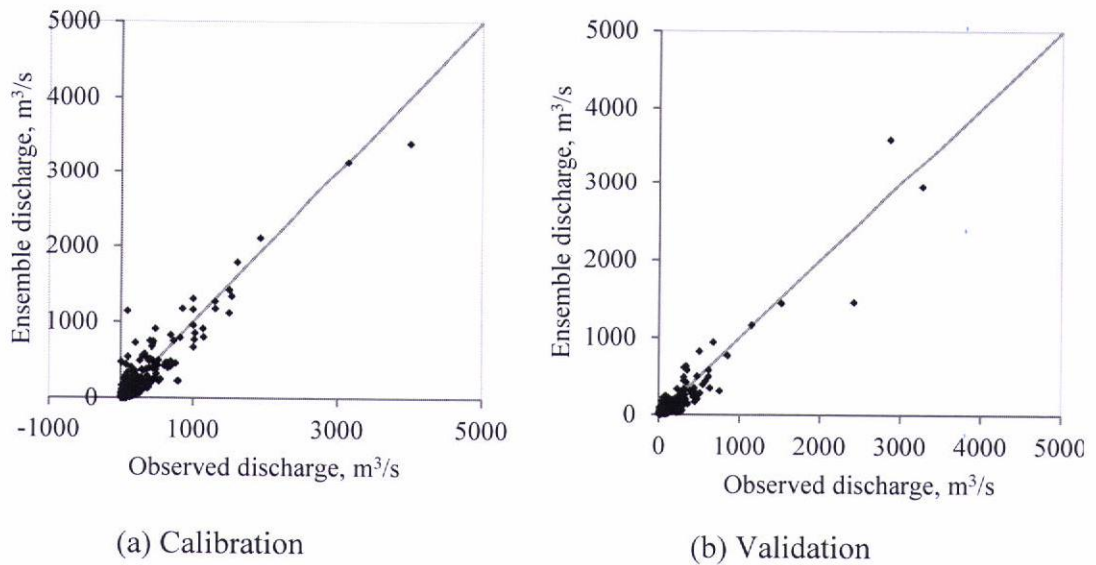
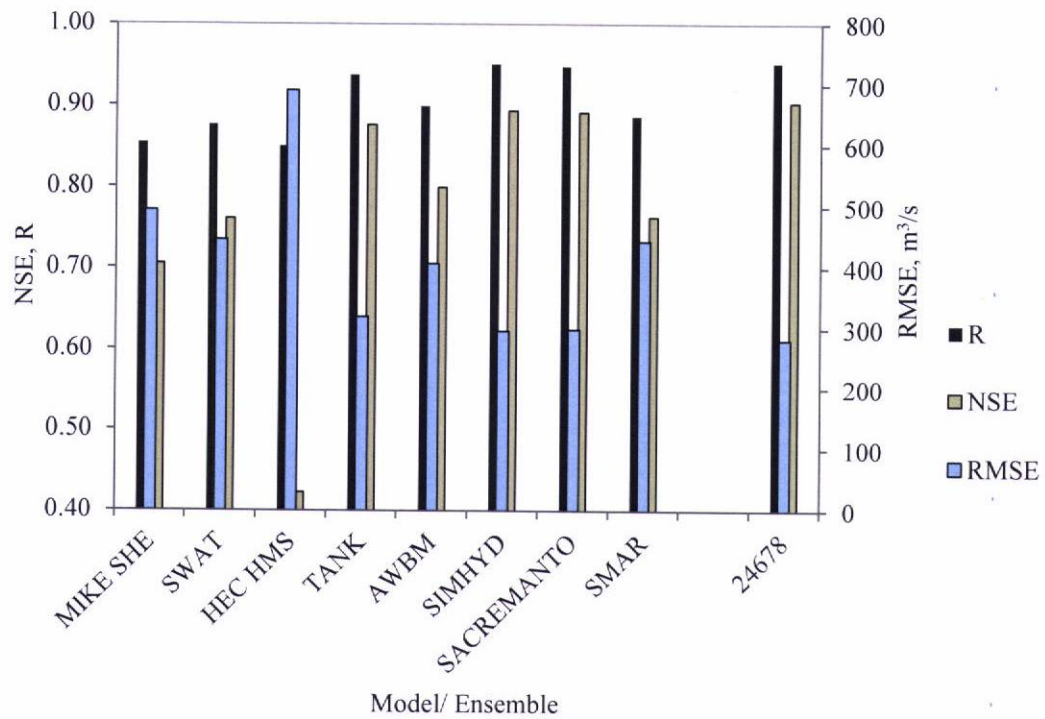


Fig. 32: Scatter plot of observed and simulated ensemble discharges for Salebhata during (a) calibration and (b) validation

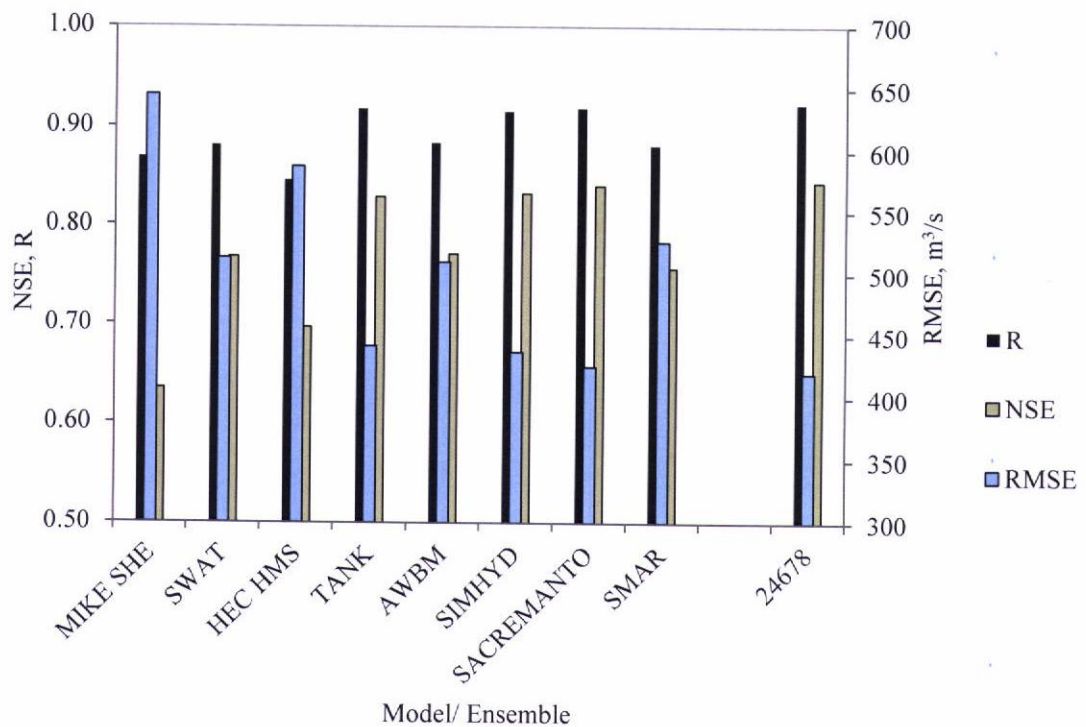
3.4.3 Comparison of the Best Ensemble with Individual Models

Performance of the selected ensembles, i.e., '24678' for Kesinga and '23467' for Salebhata, were analyzed statistically for calibration and validation periods using Nash Sutcliffe Efficiency (NSE), RMSE and R. Figs. 33a and 33b present the

comparison of the selected ensemble performance for Kesinga with that of individual models during calibration and validation, respectively. Similarly, Figs 34a and 34b present the comparison of the selected ensemble for Salebhata with that of individual models during calibration and validation. NSE of the selected ensemble is found to be 0.9 during the calibration period and 0.85 during the validation period (NSE of individual models ranges from 0.42-0.89 during calibration and 0.64-0.84 during validation) in case of Kesinga. Similarly, NSE is 0.92 during the calibration period and 0.9 during the validation period (NSE of models ranges from 0.67-0.90 during calibration and 0.60-0.87 during validation) in case of Salebhata. RMSE of ensemble '24678' is 282 m³/s during calibration period and 420 m³/s during validation period (RMSE ranges from 296-691 m³/s during calibration and 426-645 m³/s during validation) in case of Kesinga. Similarly, RMSE is 79.6 m³/s during calibration period and 70.8 m³/s during validation period (RMSE ranges from 85-157 m³/s during calibration and 79.6-137 m³/s during validation) in case of Salebhata. The figures also show that correlation coefficient of the selected ensemble is higher than that of individual models for both catchments. This shows that NSE and R of the selected ensemble is higher and RMSE is lower than any of the individual models for both catchments. The statistical values, thus, show that the selected ensembles perform better than individual models.

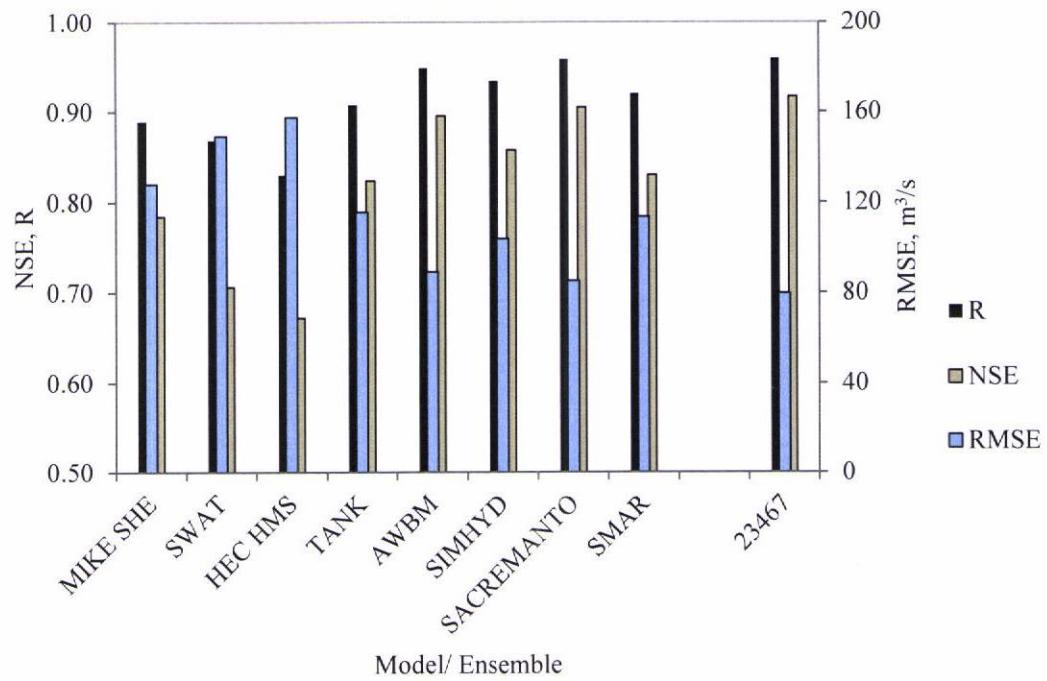


(a) Calibration

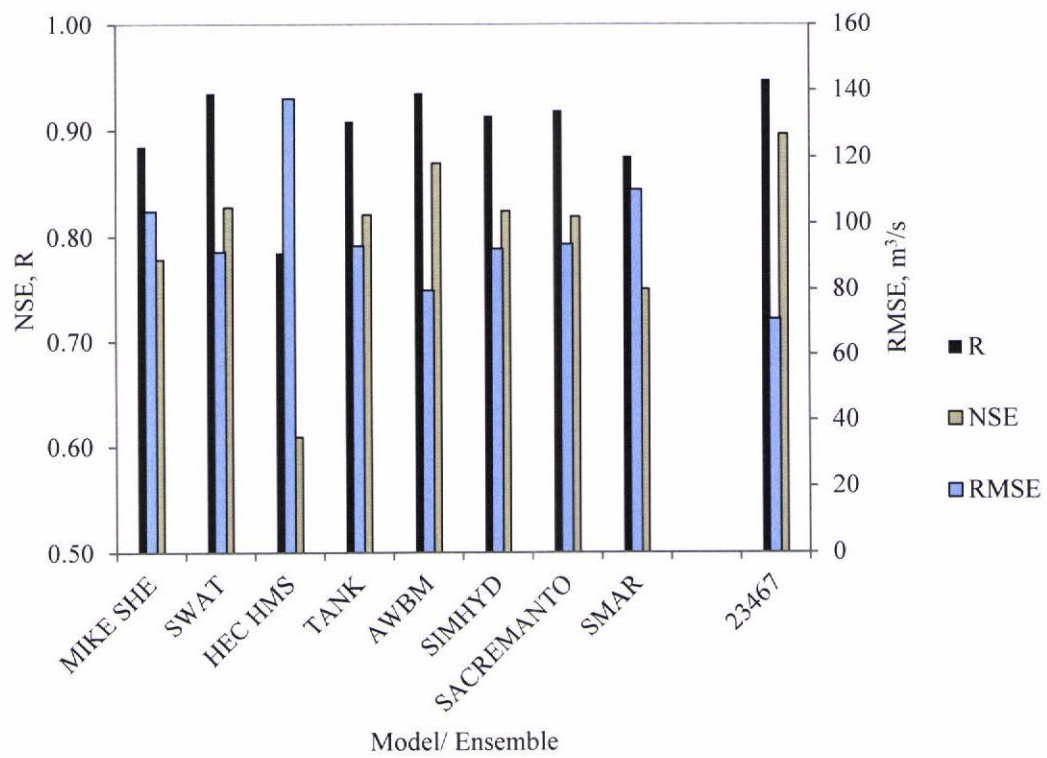


(b) Validation

Fig. 33: R, NSE and RMSE of different individual models and the best ensemble for Kesinga



(a) Calibration



(b) Validation

Fig. 34: R, NSE and RMSE of different individual models and the best ensemble for Salebhata

3.5 Uncertainty Analysis

In this section uncertainty in the discharge simulated by the best ensemble, different models and different ensembles are estimated and analyzed for both catchments. This is followed by assessment of advantages gained in terms of uncertainty reduction as a result of using ensemble over individual models.

3.5.1 Uncertainty Assessment of the Best Ensemble

Uncertainty in river discharge estimated by the selected ensembles for both catchments were analyzed using quantile regression technique, as discussed in section 2.4. Mean and standard deviation of ensemble discharge for the calibration period were estimated as 267.3 m³/s and 890 m³/s, respectively, for Kesinga, and 68.7 m³/s and 274.2 m³/s for Salebhata. The mean and standard deviation for residual are 45.6 m³/s and 278.6 m³/s, respectively, for Kesinga, and 3.3 m³/s and 79.6 m³/s, respectively, for Salebhata. Figs. 35 and 36 present the scattered plot of normalized quantile residual (NQR) and normalized quantile discharge (NQD) along with five regression lines for Kesinga and Salebhata, respectively. In these plots, two lines correspond to upper and lower limits of 90% confidence interval (CI), two corresponding to upper and lower limits of 50% CI, and one corresponding to median. Table 8 presents the slope and intercept of these regression lines, which provide the relationship between residual and discharge in Gaussian domain for Kesinga and Salebhata.

Table 8: Slope and intercept of error line for Kesinga and Salebhata

S.N.	CI	Limit	Kesinga		Salebhata	
			Slope	Intercept	Slope	Intercept
1	Median	-	-0.42	-0.07	-0.25	-0.11
2	90%	Lower	-2.05	-0.67	-2.90	-0.80
3	90%	Upper	1.75	0.91	1.68	0.93
4	50%	Lower	-1.09	-0.33	-1.05	-0.32
5	50%	Upper	0.19	0.22	0.59	0.24

Figs. 37a–37c present the confidence interval (CI) of daily discharge for the selected ensemble of Kesinga for three monsoon seasons during the calibration period, i.e., 2004, 2005 and 2006, respectively, whereas, Figs. 38a and 38b present the CI of daily discharge during the validation period for two monsoon seasons, i.e., 2007 and 2008, respectively. Similarly, Figs. 39a–39c and Figs 40a–40b present the corresponding CI of daily discharge for the selected ensemble of Salebhata during the calibration and

validation periods. It is evident from Figs. 37–40 that most of the observed data fall within 50% or 90% CI bands which indicate the accuracy of the developed error model. However, in case of validation period (Figs. 38 and 40), a few data points fall outside the defined bands.

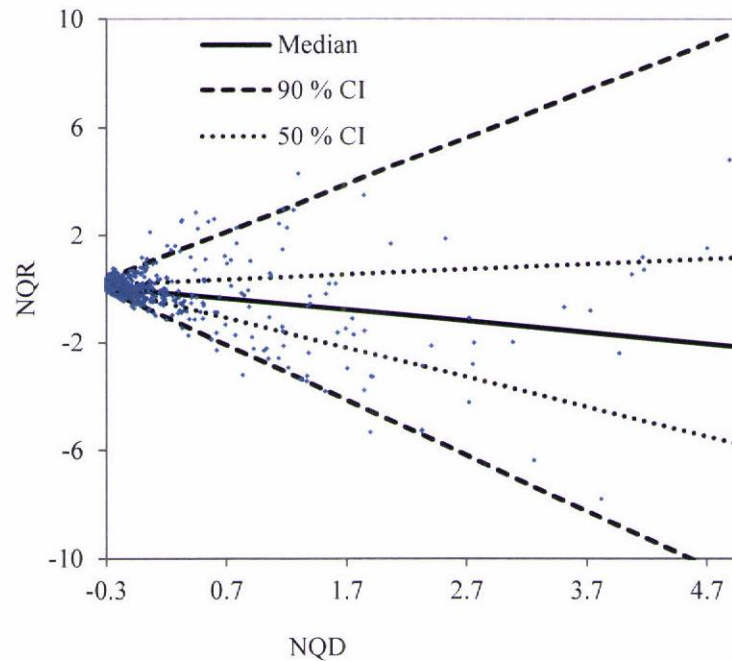


Fig. 35: Error model of ensemble '24678' in normalized domain for Kesinga

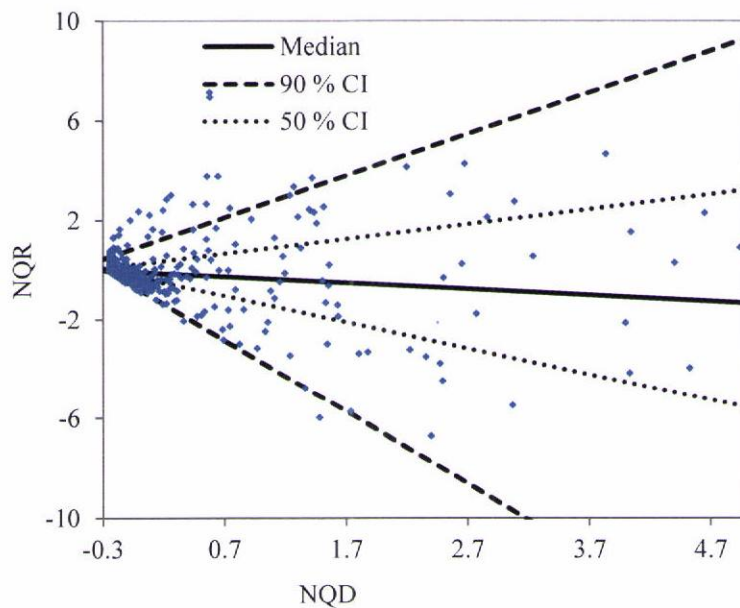
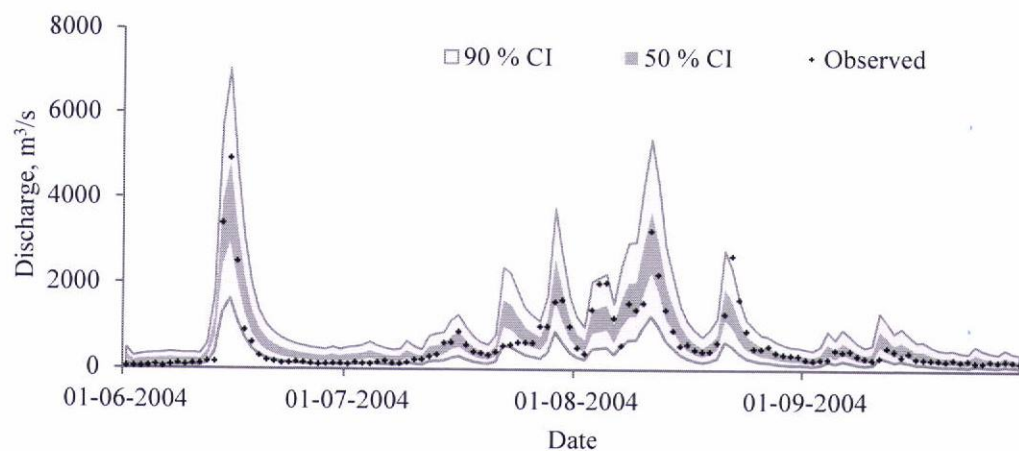
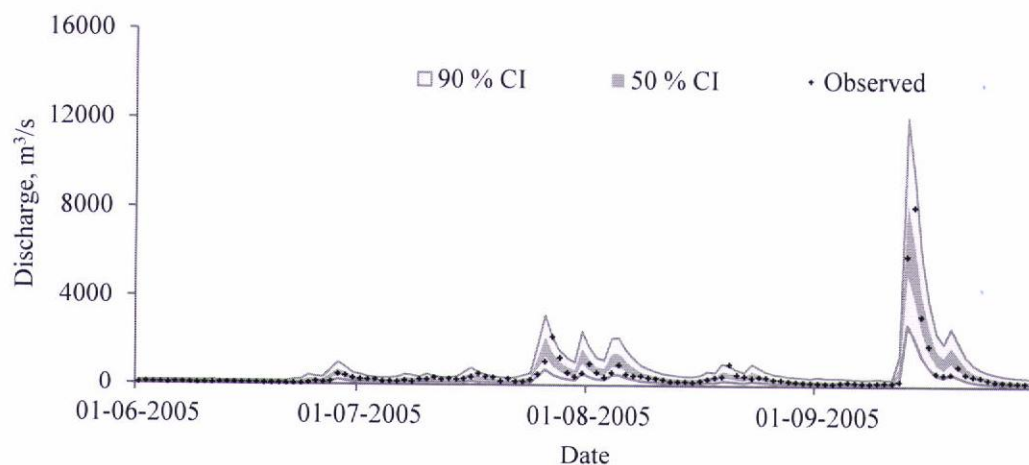


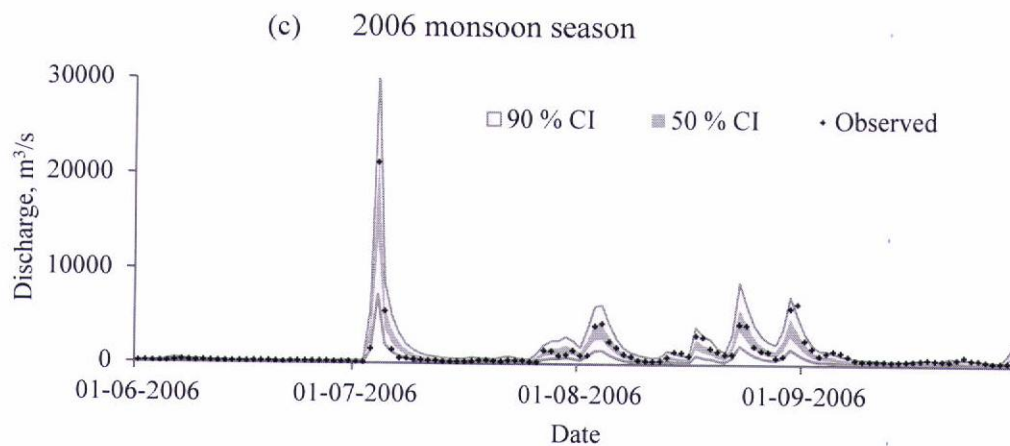
Fig. 36: Error model of ensemble '23467' in normalized domain for Salebhata



(a) 2004 monsoon season

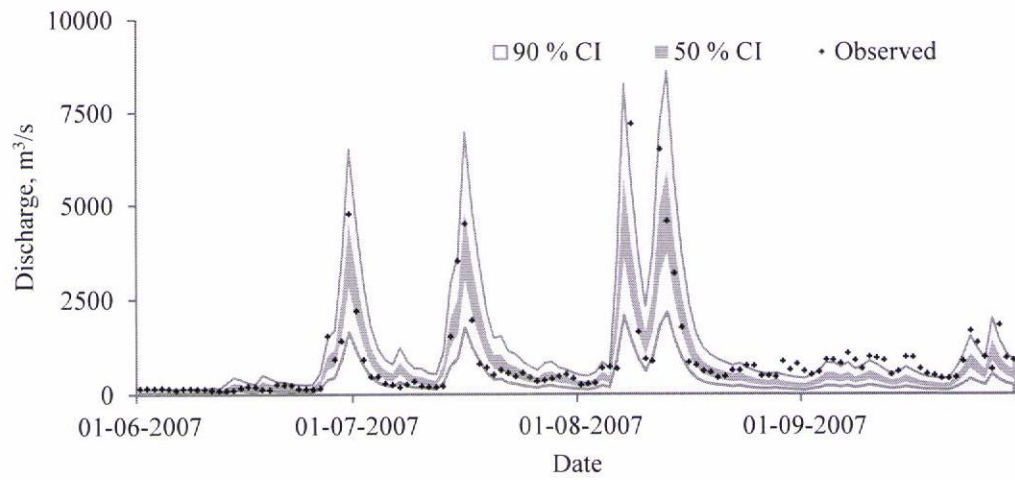


(b) 2005 monsoon season

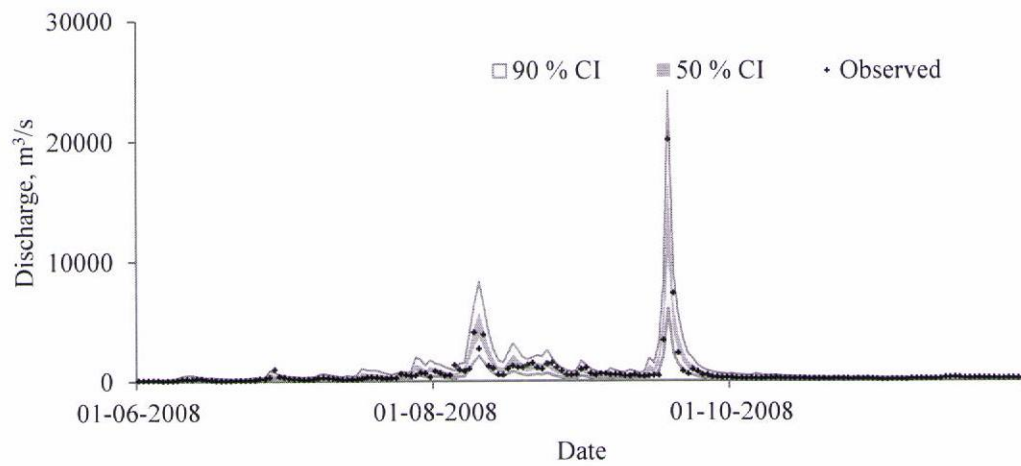


(c) 2006 monsoon season

Fig. 37: Confidence interval of discharge for Kesinga during calibration period

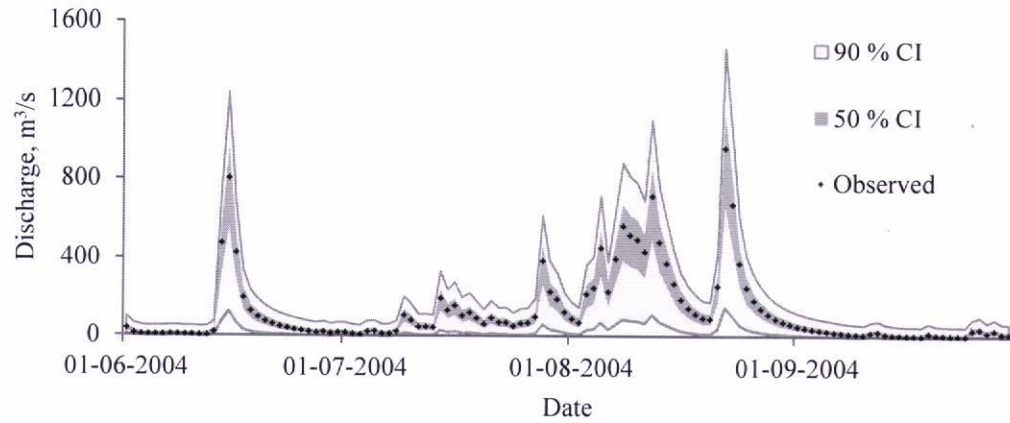


(a) 2007 monsoon season

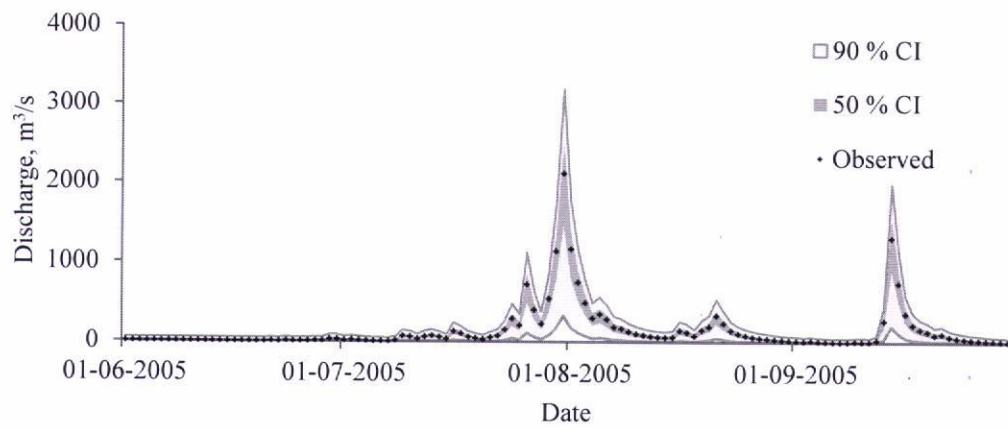


(b) 2008 monsoon season

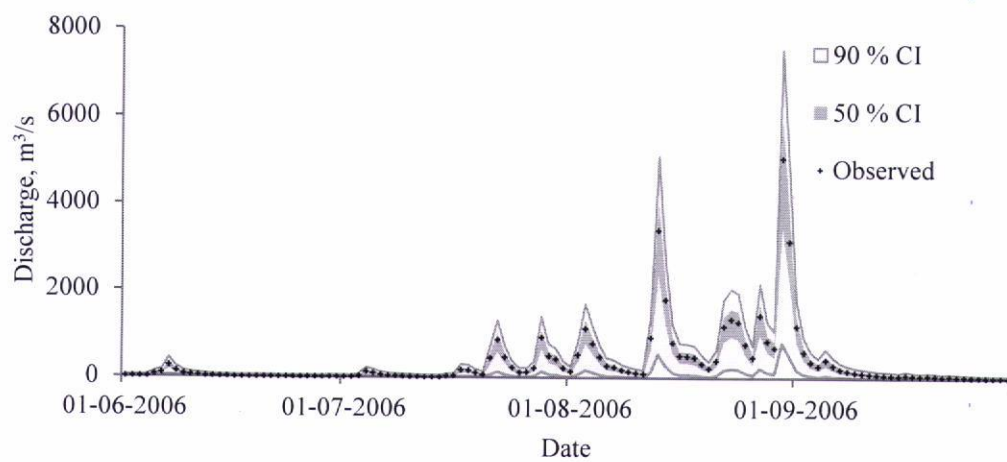
Fig. 38: Confidence interval of discharge for Kesinga during validation period



(a) 2004 monsoon season

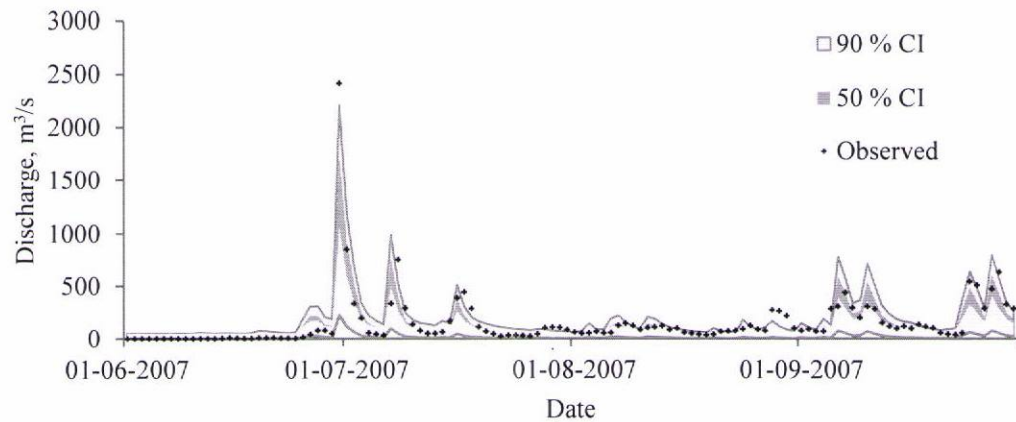


(b) 2005 monsoon season

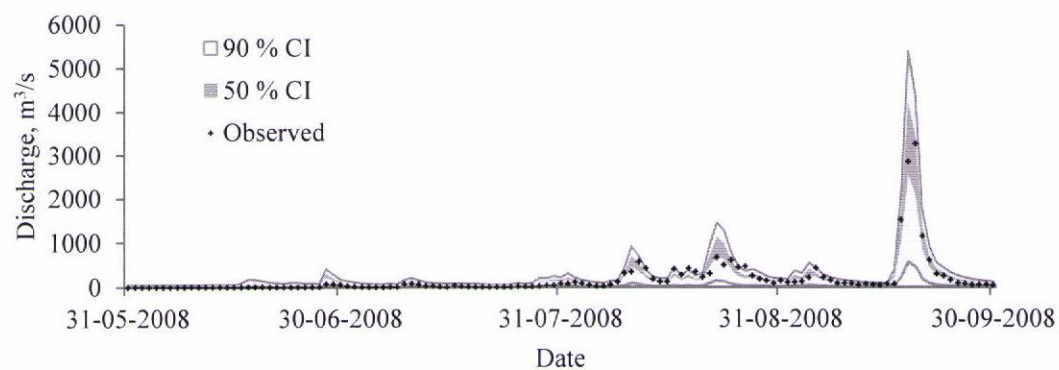


(c) 2006 monsoon season

Fig. 39: Confidence interval of discharge for Salebhata during calibration period



(a) 2007 monsoon season



(b) 2008 monsoon season

Fig. 40: Confidence interval of discharge for Salebhata during validation period

3.5.2 Uncertainty Assessment of the Hydrological Models in Simulating River Discharge

To estimate the uncertainty in river discharge simulated by different hydrological models, error models were developed for each of them for both catchments using quantile regression technique. Figs. 41 and 42 present the error model of different hydrological models for Kesinga and Salebhata, respectively. The dots in these figures present the daily simulated discharge and residual in normalized domain, whereas, red and green lines in figures 41 and 42 correspond to 90% and 50% confidence interval, respectively. Two lines in both the color category indicate the upper (upper line) and lower limits (lower line) of respective confidence band. Tables 9 and 10 present the intercept and slope of these lines for Kesinga and Salebhata, respectively.

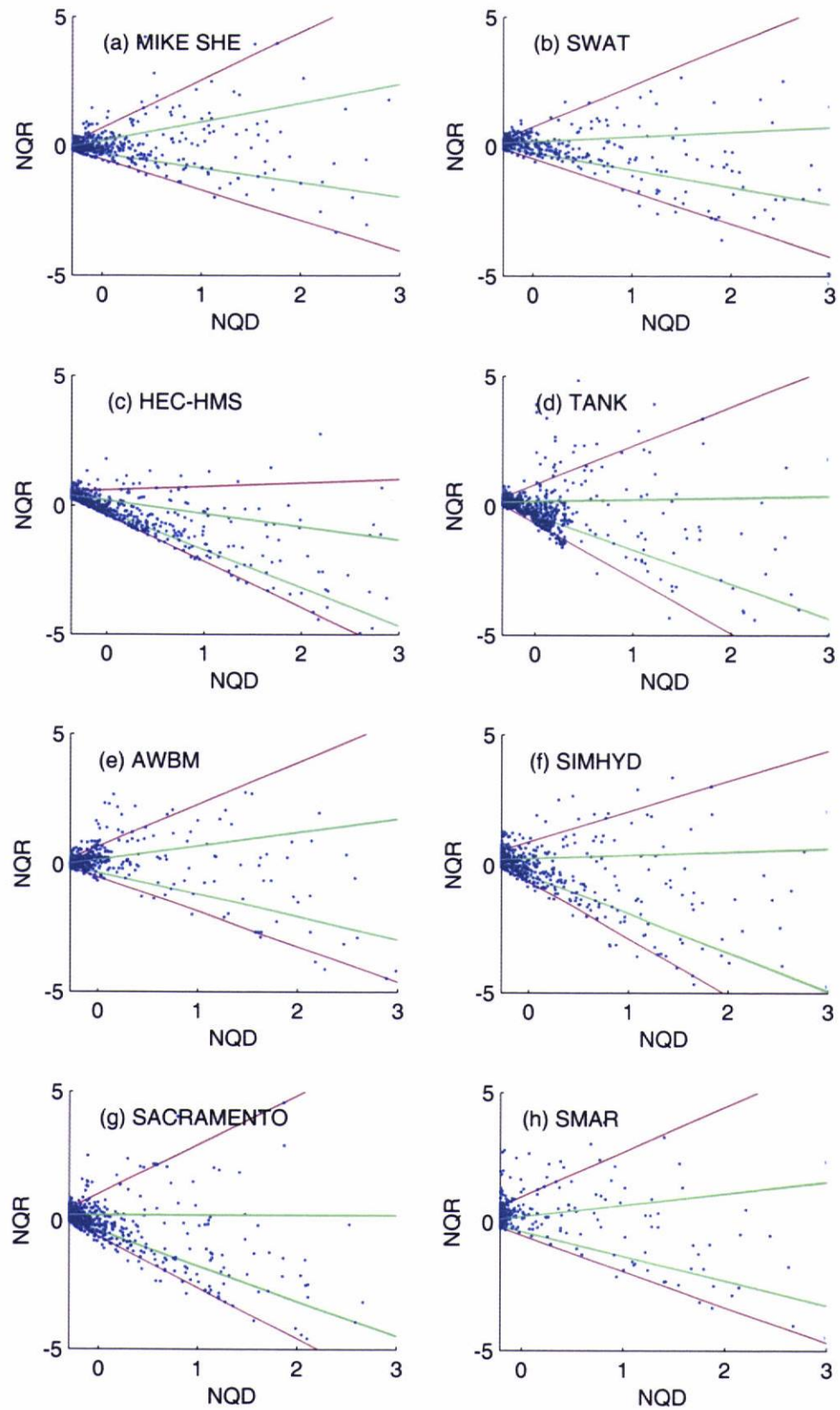


Fig. 41: Error models of different hydrological models for 90% (red) and 50% (green) CI in case of Kesinga

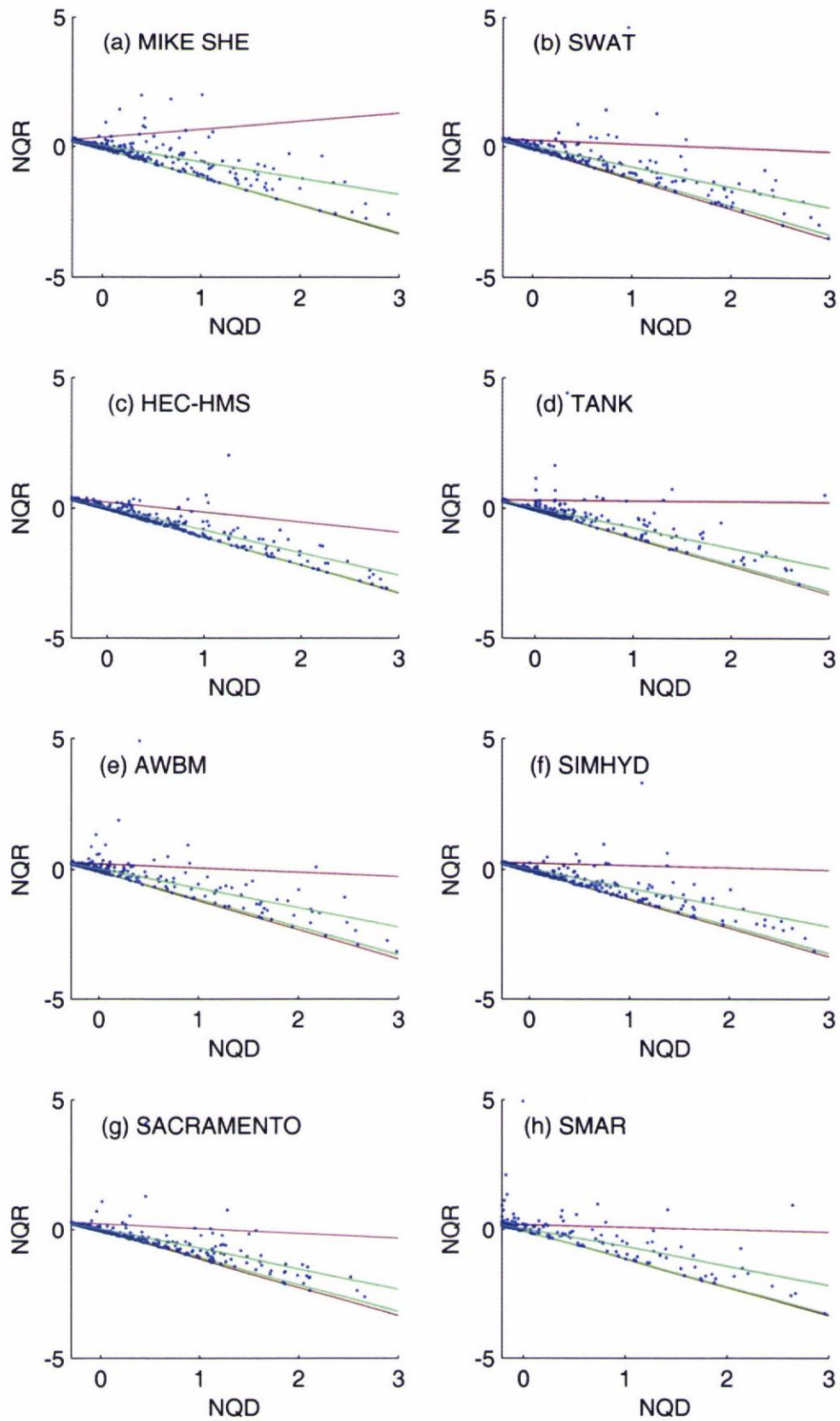


Fig. 42: Error models of different hydrological models for 90% (red) and 50% (green) CI in case of Salebhata

Table 9: Intercept and slope of error models for different hydrological models in case of Kesinga

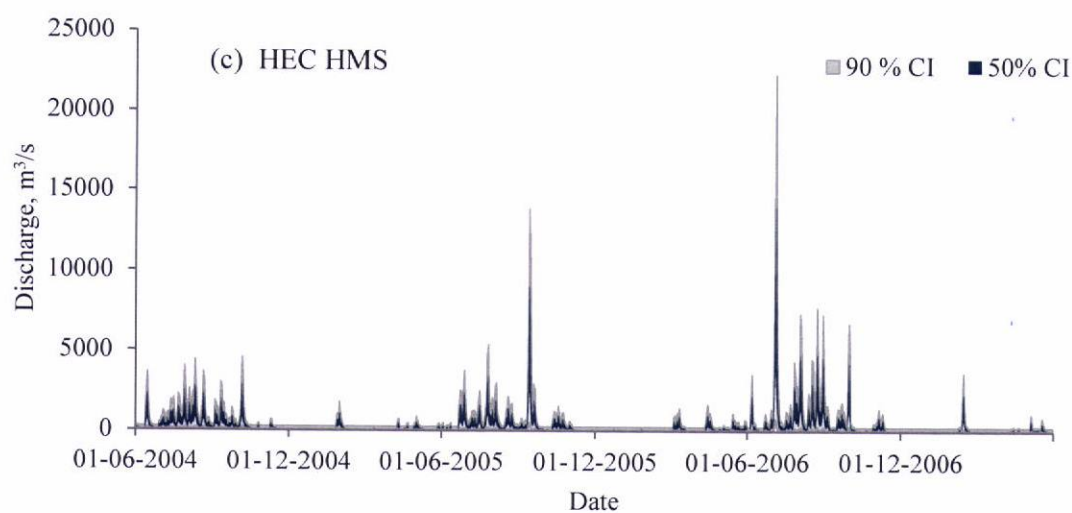
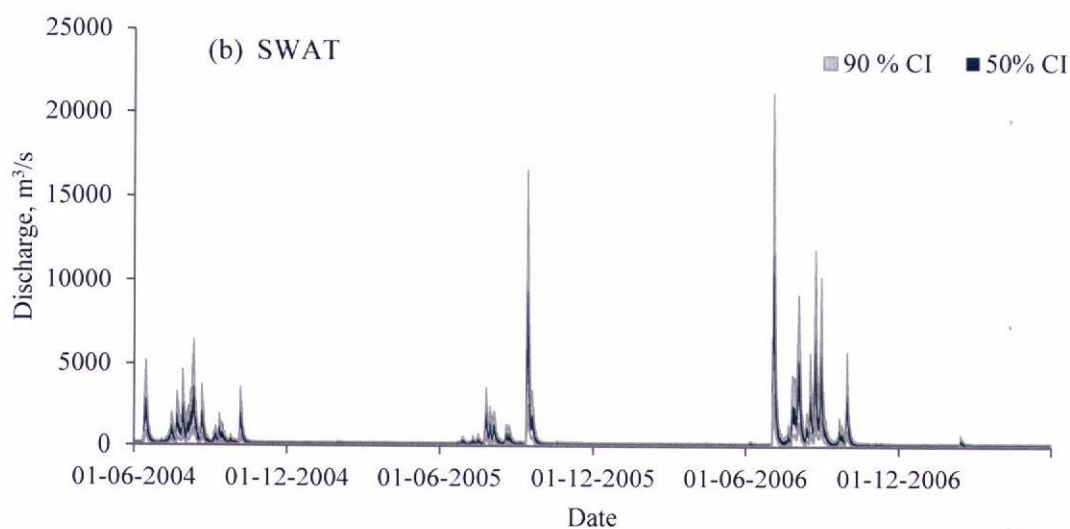
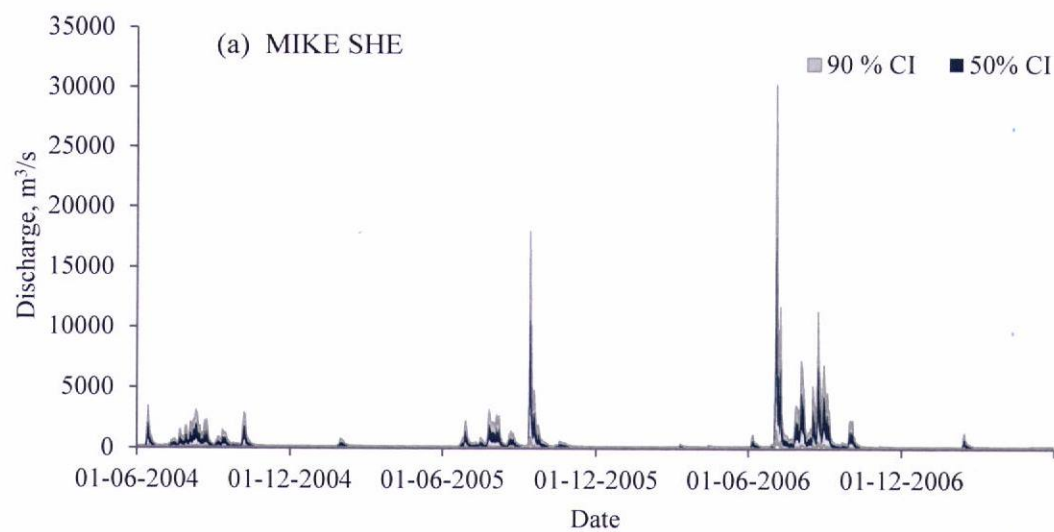
Model	Limit	50% CI		90% CI	
		Intercept	Slope	Intercept	Slope
MIKE SHE	Upper limit	0.20	0.73	0.70	1.85
	Lower limit	-0.30	-0.55	-0.52	-1.18
SWAT	Upper limit	0.19	0.18	0.76	1.58
	Lower limit	-0.25	-0.66	-0.47	-1.26
HEC HMS	Upper limit	0.20	-0.51	0.59	0.14
	Lower limit	-0.25	-1.47	-0.39	-1.78
TANK	Upper limit	0.19	0.07	0.83	1.49
	Lower limit	-0.37	-1.33	-0.67	-2.13
AWBM	Upper limit	0.19	0.52	0.68	1.61
	Lower limit	-0.33	-0.88	-0.51	-1.37
SIMHYD	Upper limit	0.27	0.13	0.90	1.16
	Lower limit	-0.42	-1.51	-0.67	-2.22
SACRAMENTO	Upper limit	0.23	0.00	1.05	1.90
	Lower limit	-0.37	-1.37	-0.61	-1.99
SMAR	Upper limit	0.22	0.45	0.98	1.73
	Lower limit	-0.37	-0.95	-0.51	-1.40

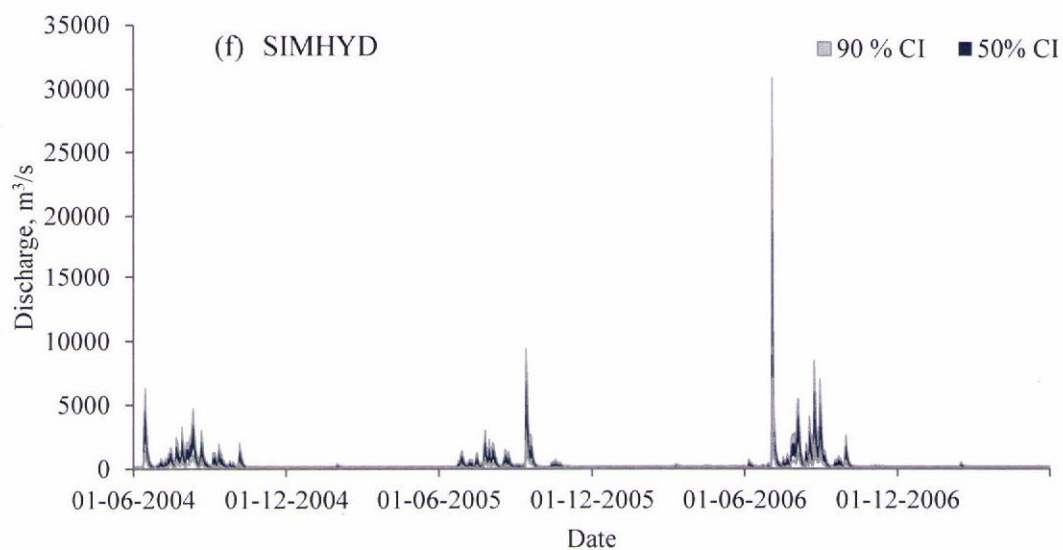
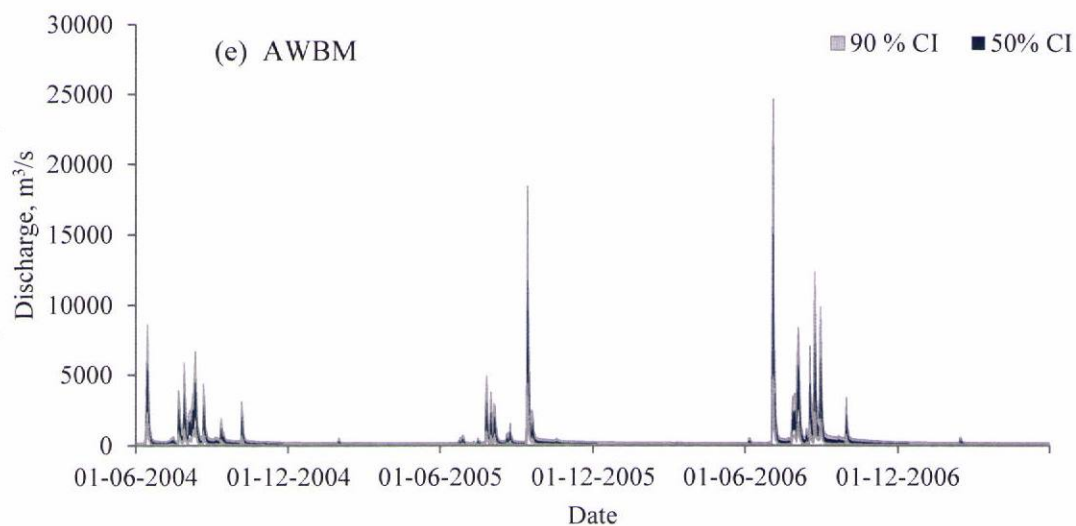
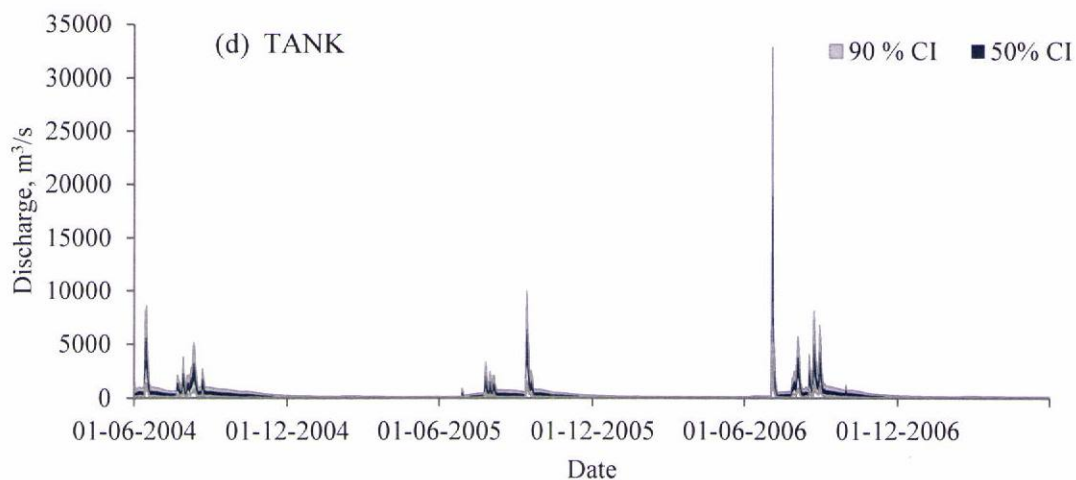
The developed error models were used to estimate the uncertainty in river discharges simulated by different models during calibration and validation periods. Fig. 43 presents the 50% and 90% confidence band of river discharge simulated by different models in case of Kesinga during calibration period. The gray and blue band represents the 90% and 50% confidence band of discharge, respectively. In order to clarify Fig. 43, a part of this figure (uncertainty band of SACRAMENTO model i.e., 43g) is enlarged in Fig. 44. This figure shows that most of the observed data points are within the uncertainty band. Similarly, Fig. 45 presents the uncertainty bands of river discharge simulated by different models in case of Salebhata during calibration period. Area under the uncertainty bands at 50% CI and 90% CI was considered as a measure of uncertainty in discharge simulated by different models and ensembles. The estimated area was used to compare the uncertainty level of selected ensembles

with that of individual hydrological models. Uncertainty bands were also developed during validation periods for both catchments to estimate the area under the band, and subsequently to compare the uncertainty level of selected ensembles with that of individual hydrological models.

Table 10: Intercept and slope of error models for different hydrological models in case of Salebhata

Model	Limit	50% CI		90% CI	
		Intercept	Slope	Intercept	Slope
MIKE SHE	Upper limit	0.03	-0.62	0.36	0.30
	Lower limit	-0.11	-1.06	-0.12	-1.07
SWAT	Upper limit	0.02	-0.78	0.26	-0.15
	Lower limit	-0.09	-1.09	-0.10	-1.14
HEC HMS	Upper limit	0.02	-0.87	0.23	-0.38
	Lower limit	-0.06	-1.06	-0.06	-1.07
TANK	Upper limit	0.02	-0.77	0.31	-0.02
	Lower limit	-0.08	-1.05	-0.09	-1.08
AWBM	Upper limit	0.02	-0.74	0.22	-0.16
	Lower limit	-0.08	-1.07	-0.10	-1.12
SIMHYD	Upper limit	0.02	-0.74	0.24	-0.09
	Lower limit	-0.07	-1.06	-0.08	-1.09
SACRAMENTO	Upper limit	0.01	-0.78	0.23	-0.20
	Lower limit	-0.07	-1.04	-0.08	-1.09
SMAR	Upper limit	0.02	-0.73	0.20	-0.10
	Lower limit	-0.09	-1.07	-0.09	-1.09





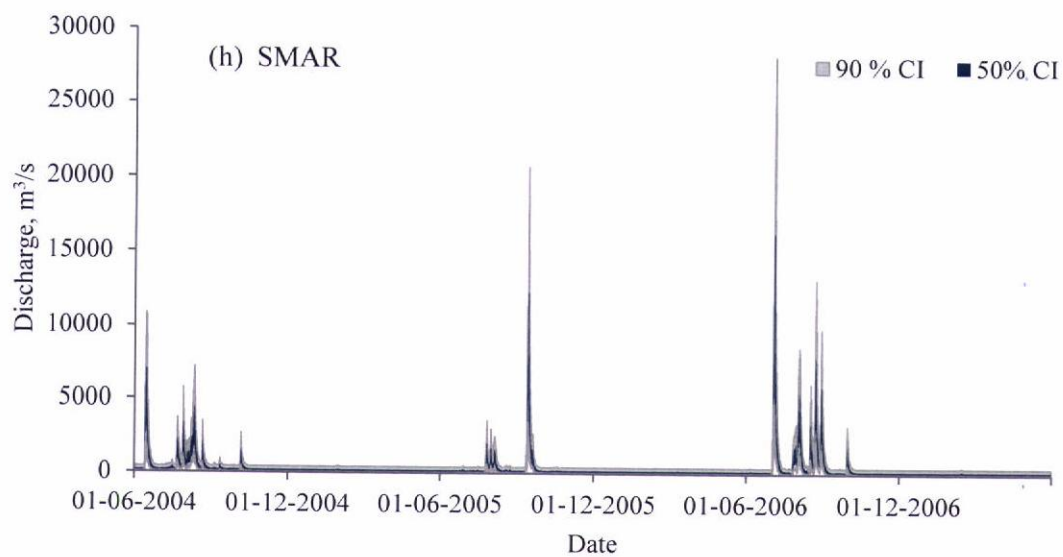
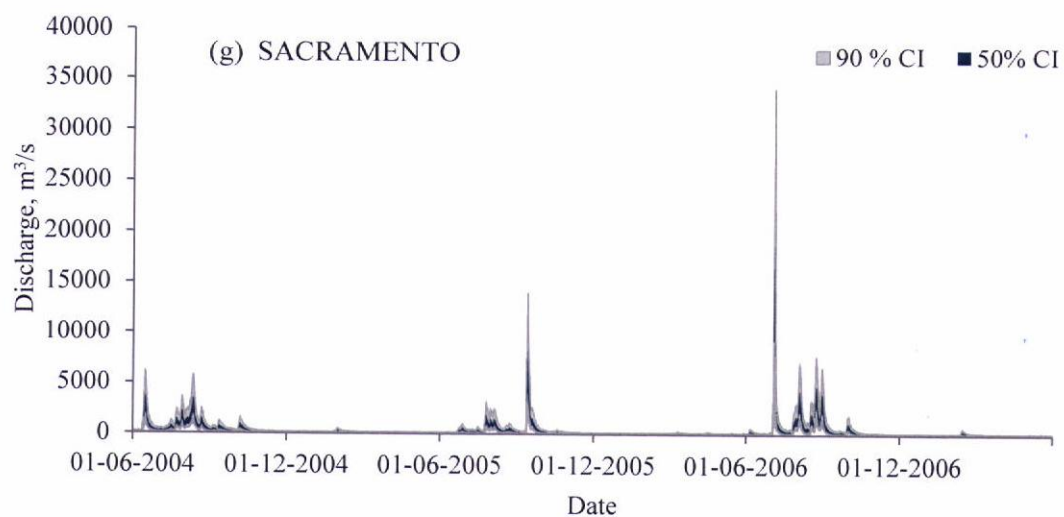


Fig. 43: 90% (grey) and 50% (blue) uncertainty band of discharge simulated by different models in case of Kesinga during calibration period

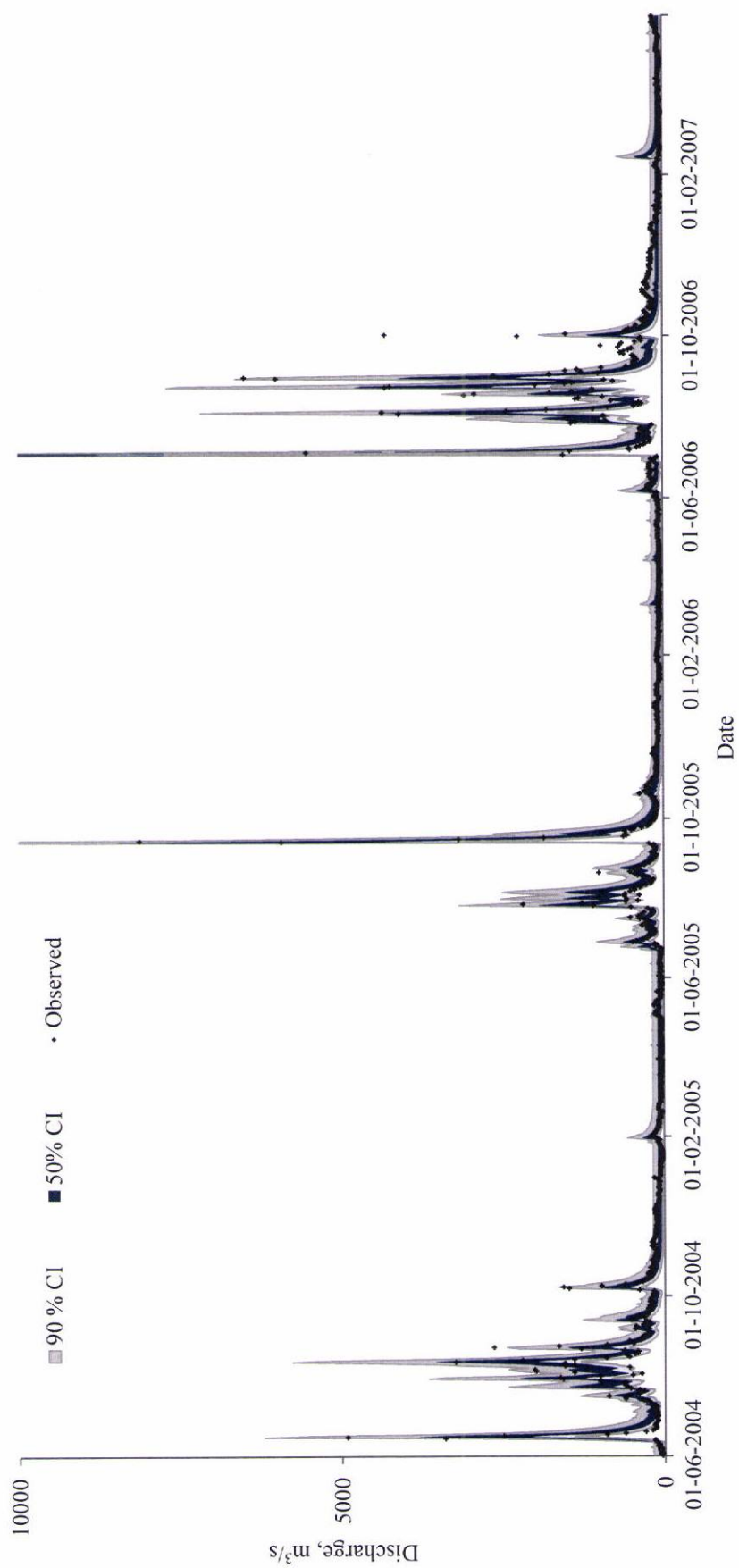
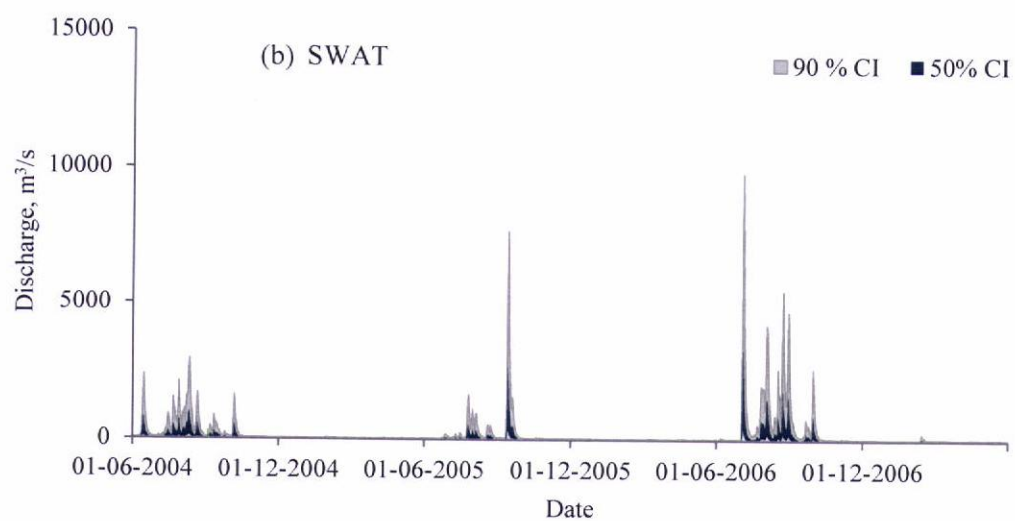
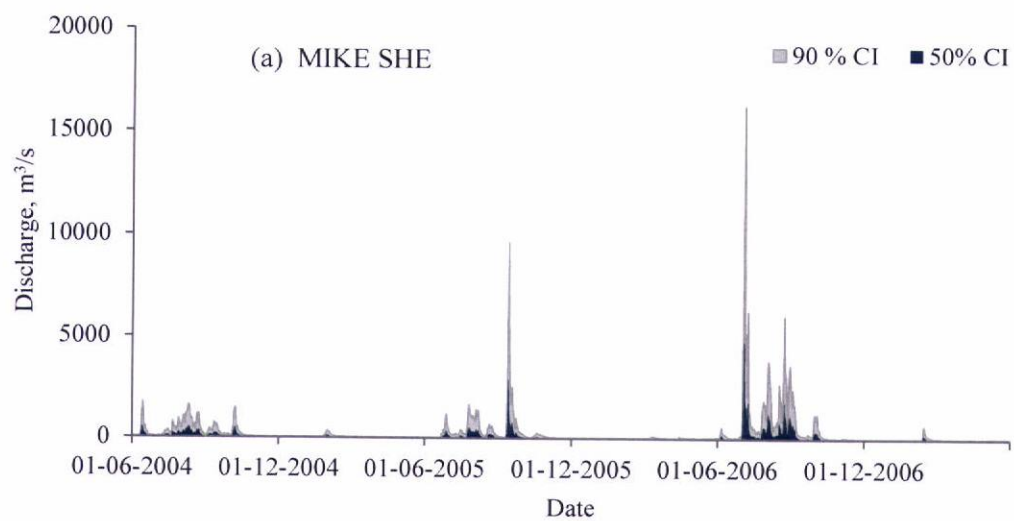
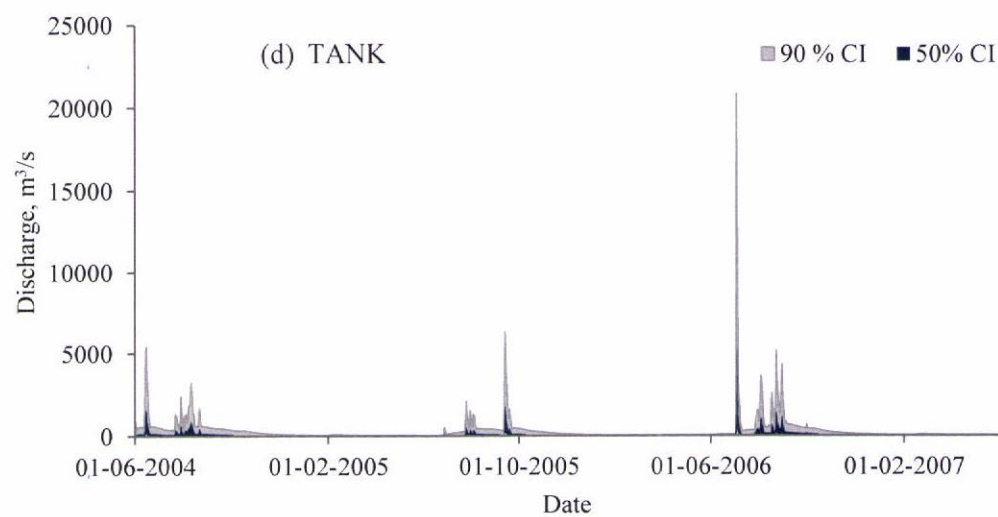
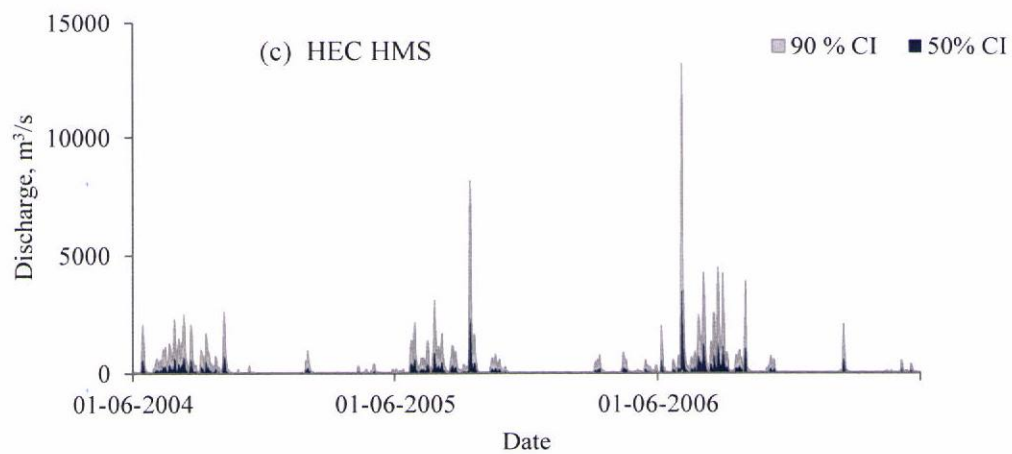
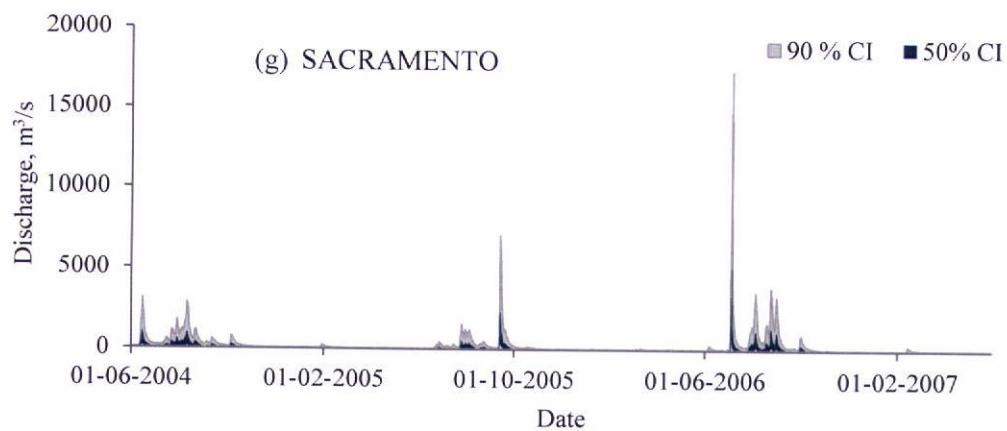
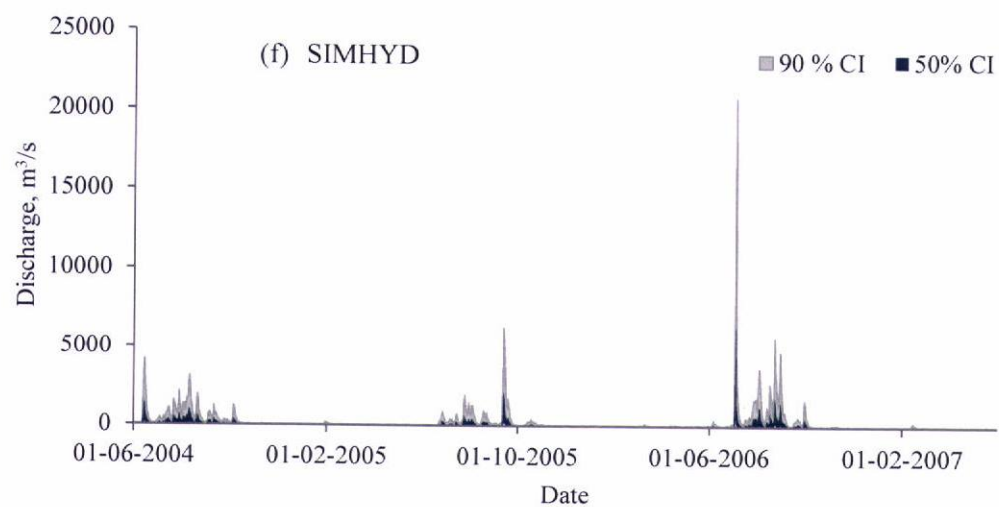
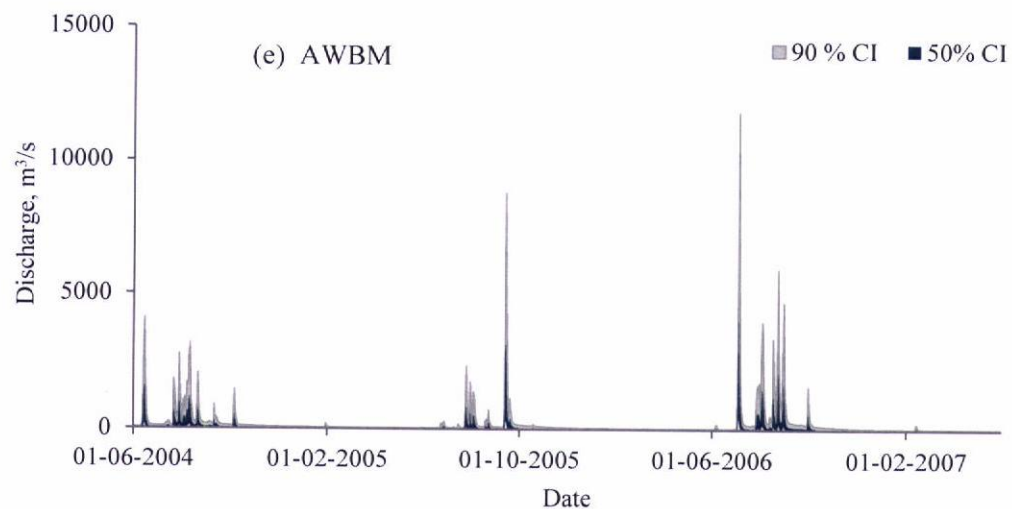


Fig. 44: A zoomed view of uncertainty band of discharge for SACRAMENTO model in case of Kesinga during calibration period







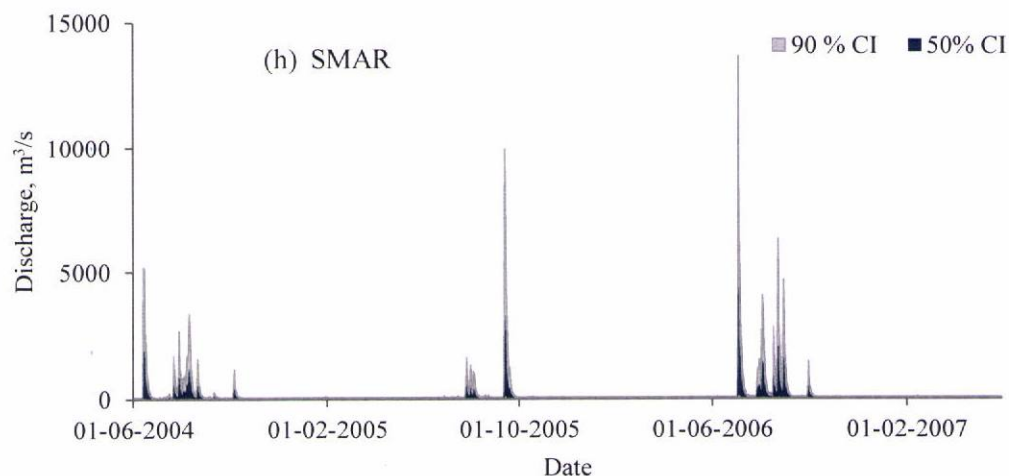
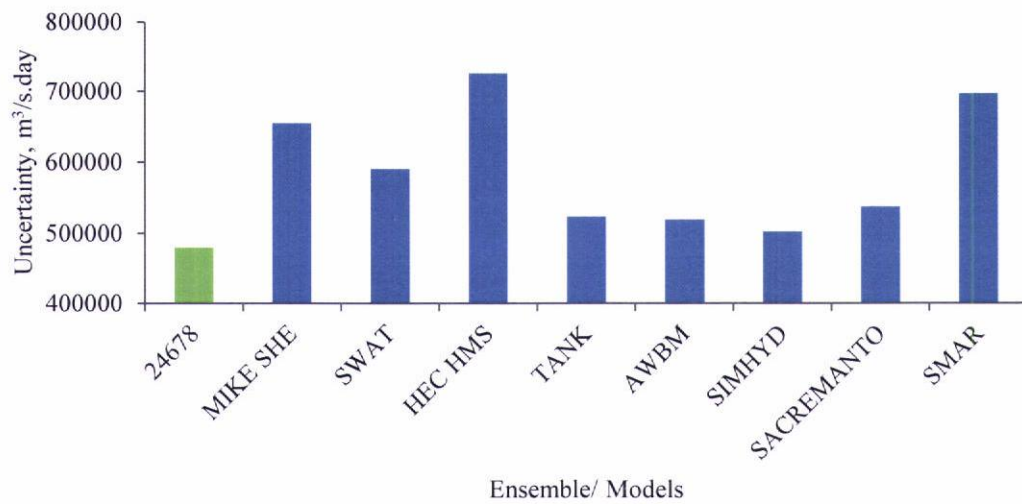


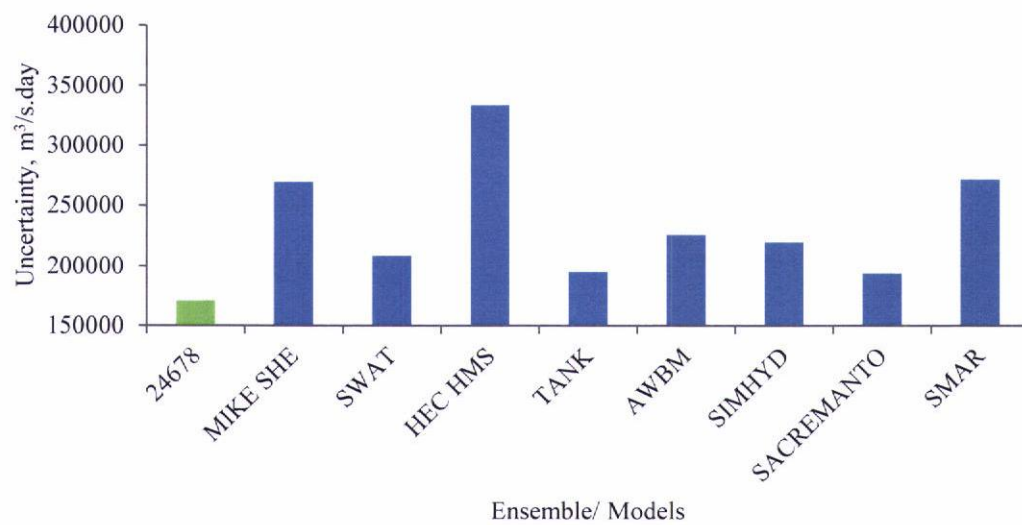
Fig. 45: 90% (grey) and 50% (blue) uncertainty band of discharge simulated by different models in case of Salebhata during calibration period

3.5.3 Uncertainty Comparison of the Best Ensemble with Individual Models

In order to assess the advantage gained using ensemble rather than individual models, uncertainty level (referred as Uncertainty), i.e., area under the uncertainty band for different uncertainty levels, of the selected ensemble during calibration and validation periods was compared with that of individual models for both catchments. Figs. 46a–46b and Figs. 47a–47b present the uncertainty level of the best ensemble and the individual hydrological models in simulating river discharge during calibration and validation periods for Kesinga at 90% and 50% CI, respectively. Similarly, Figs. 48a–48b and Figs. 49a–49b present the uncertainty level of the best ensemble and the individual hydrological models in simulating river discharge during calibration and validation periods for Salebhata. It is evident from Figs 46 –49 that uncertainty level of the best ensembles is lesser than any of the hydrological models considered in case of both catchments at both 50% and 90% confidence levels. These results show that ensembles reduce uncertainty. These results are in agreement with earlier findings that multi-model ensembles reduce uncertainty at different stages (Singh and Sankarasubramanian, 2014).

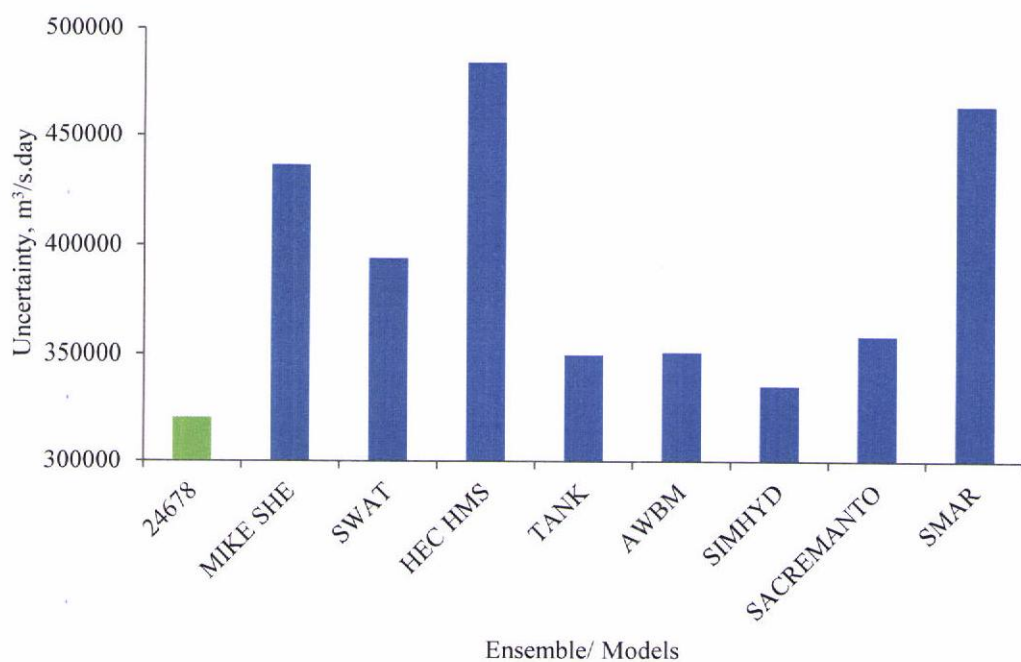


(a) 90% CI

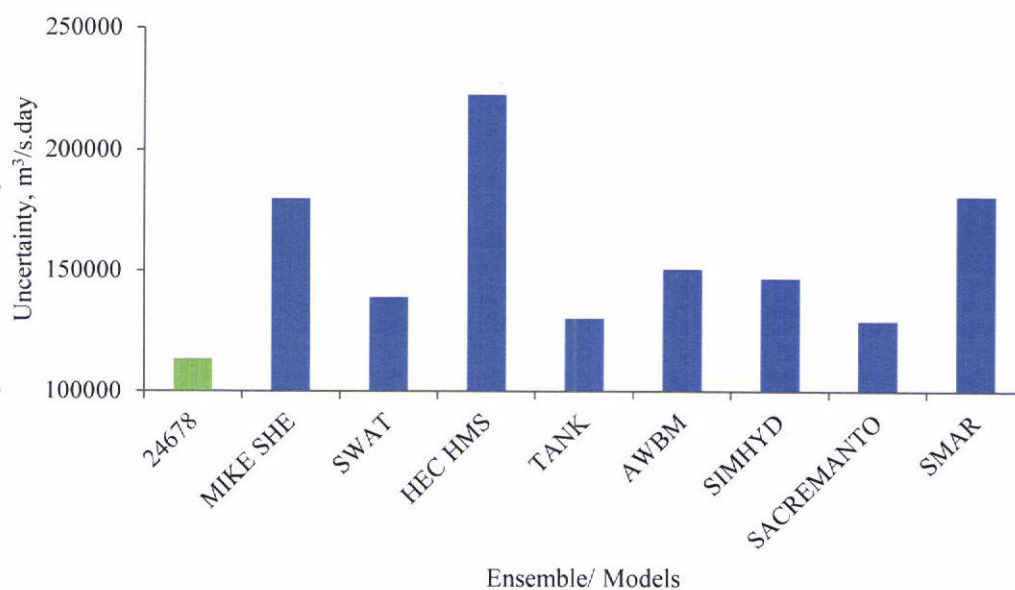


(b) 50% CI

Fig. 46: Uncertainty level of the best ensemble and individual models for Kesinga during calibration period

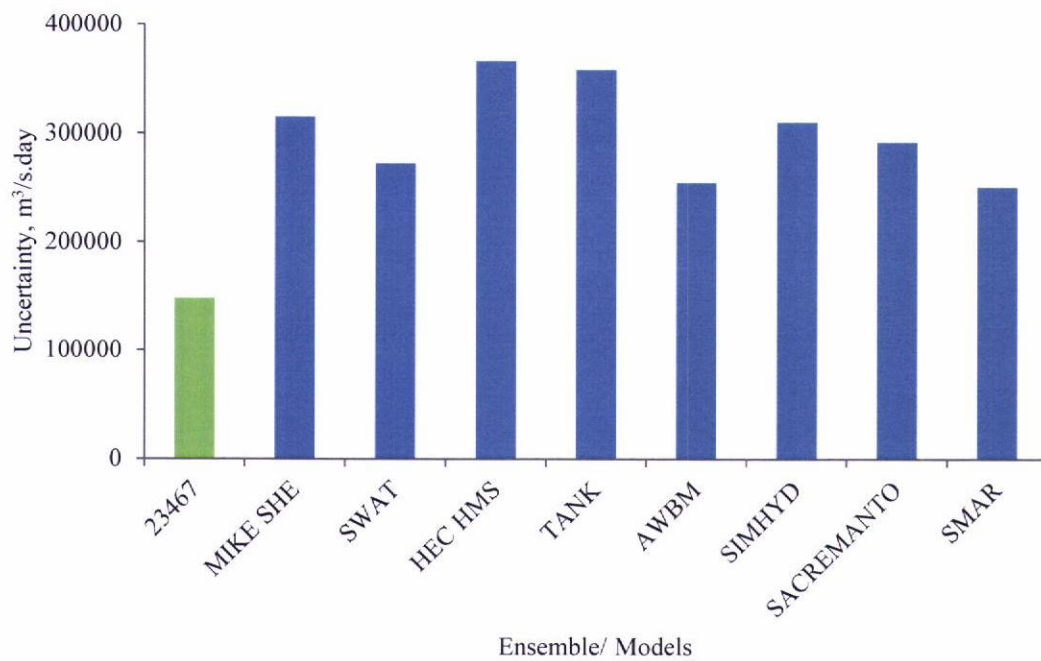


(a) 90% CI

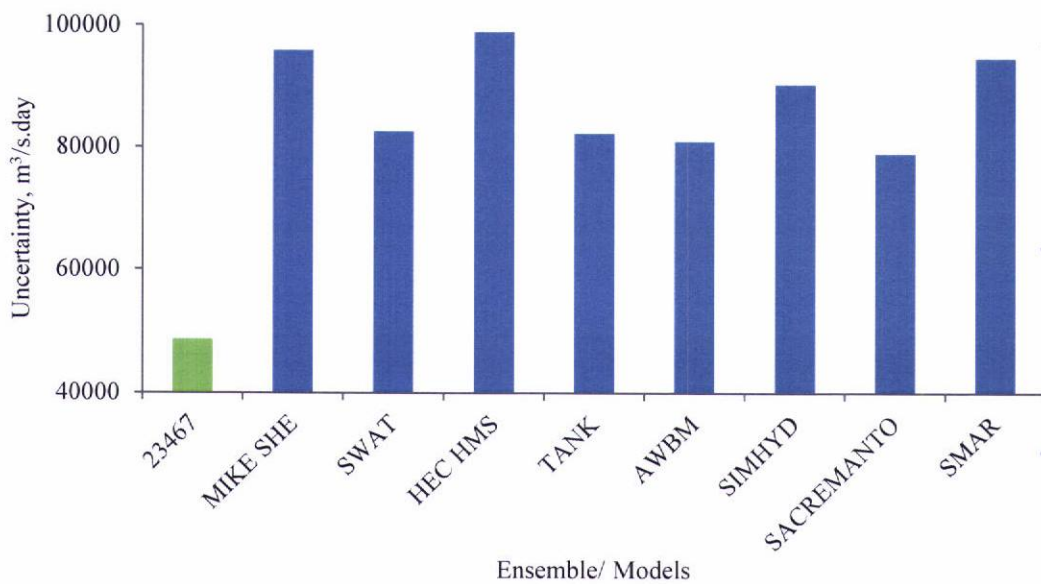


(b) 50% CI

Fig. 47: Uncertainty level of the best ensemble and individual models for Kesinga during validation period

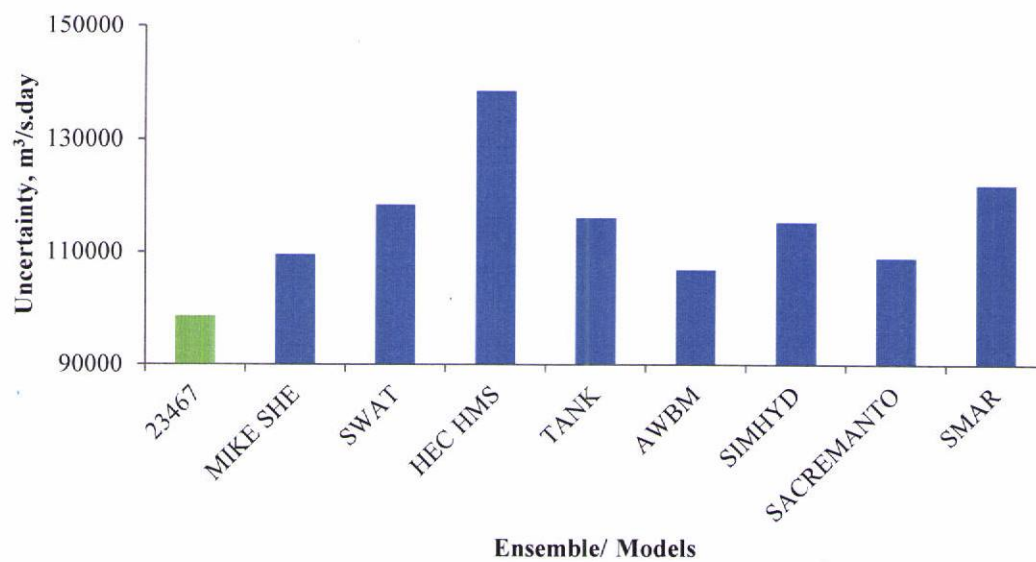


(a) 90% CI

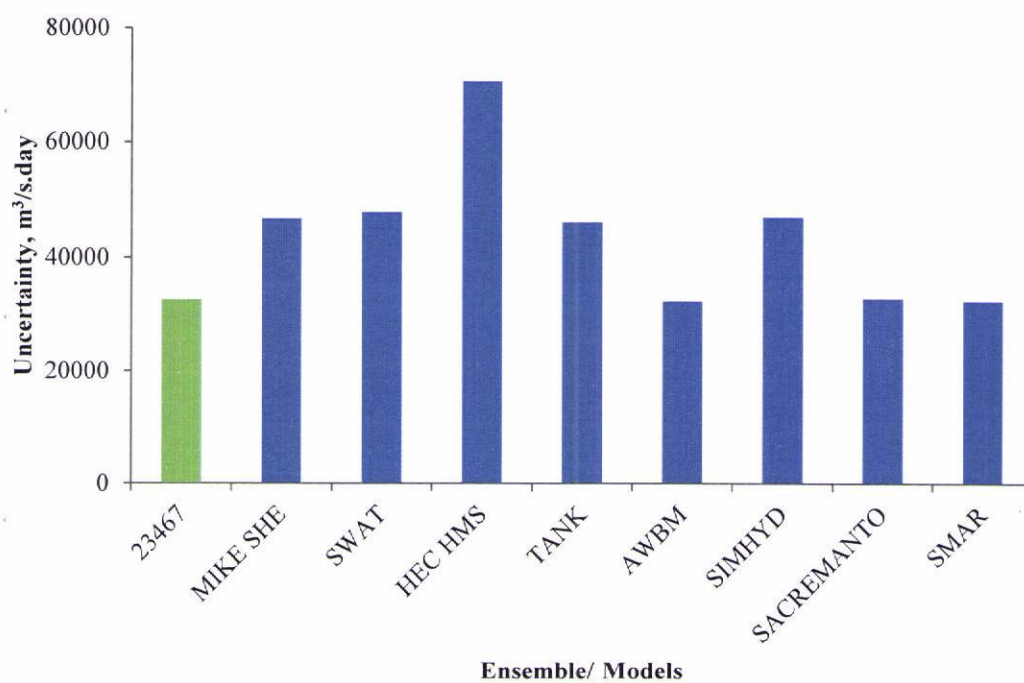


(b) 50% CI

Fig. 48: Uncertainty level of the best ensemble and individual models for Salebhata during calibration period



(a) 90% CI



(b) 50% CI

Fig. 49: Uncertainty level of the best ensemble and individual models for Salebhata during validation period

Annexure-VI

CONCLUSIONS/RECOMMENDATIONS

The following specific conclusions are drawn from the study:

1. Weighted average based on calibration performance (WAM_k_1.5) and linear programming methods are the best ensemble methods for Kesinga and Salebhata catchment, respectively.
2. The ensemble of SWAT, TANK, SIMHYD, SACRAMENTO and SMAR is the best ensemble for Kesinga, whereas, the ensemble of SWAT, HEC-HMS, TANK, SIMHYD and SACRAMENTO is the best ensemble for Salebhata.
3. The selected ensembles outperform individual models in simulating river discharge for both Kesinga and Salebhata.
4. The uncertainty level of the selected ensemble is lower than that of individual models.

References

- Abbott, M. B., Bathurst, J. C., Cunge, J. A., O'connell, P. E. and Rasmussen, J. (1986a). An introduction to European Hydrological System- Systeme Hydrologique Europeen, SHE, 1: history and philosophy of physically-based, distributed modelling system. *Journal of Hydrology*, 87, 45–59.
- Abbott, M. B., Bathurst, J. C., Cunge, J. A., O'connell, P. E. and Rasmussen, J. (1986b). An introduction to European Hydrological system-Systeme Hydrologique Europeen, SHE, 2: structure of a physically-based, distributed modelling system. *Journal of Hydrology*, 87, 61–77.
- Ajami, N. K., Duan, Q., Gao, X. and Sorooshian, S., (2006). Multimodel combination techniques for analysis of hydrological simulations: Application to distributed model intercomparison project results. *Journal of Hydrometeorology*, 7, 755–768.
- Arnold, J. G., Srinivasan, R., Muttiah, R. S. and Williams, J. R. (1998). Large area hydrologic modeling and assessment Part I: Model development. *Journal of the American Water Resources Association*, 34, 73–89.
- Bandyopadhyay, A., Bhadra, A., Swarnkar, R.K., Raghuwanshi, N.S. and Singh, R. (2012). Estimation of reference evapotranspiration using user-friendly decision support system: DSS_ET. *Agricultural and Forest Meteorology*, 154–155, 19–29.
- Barker, T. W. (1991). The relationship between spread and forecast error in extended-range forecasts. *Journal of Climate*, 4, 733–742.
- Boughton, W. (2004). The Australian water balance model. *Environmental Modelling & Software*, 19, 943–956.
- Buizza, R. and Palmer, T.N. (1998). Impact of ensemble size on ensemble prediction. *Monthly Weather Review*, 126, 2503–2518.
- Burnash, R. J., Ferral, R. L. and McGuire, R. A. (1973). A generalized streamflow simulation system – conceptual modelling for digital computers. U.S. Dept. of Commerce, National Weather Service, Silver Springs, Md. And State of California, *Dept. of Water Resources, Sacramento, Calif.*
- Chiew F.H.S., Peel M.C. and Western A.W. (2002). Application and testing of the simple rainfall-runoff model SIMHYD. In Singh VP, Frevert DK (eds) *Mathematical Models of Small Watershed Hydrology and Applications*, *Water Resources Publication, Littleton, Colorado*, 335–367.
- Clanor, M. D. M., Bondad, R. G. M., Escobar, E. C., Caburnay, J. M., Ventura, J.R. S., Dorado, A. A., Lu, M. M. D. and Glorioso, A. U. (2015). Rainfall-runoff modeling of the Molawin watershed of the Makiling forest reserve using five lumped conceptual models. *Philippine e-Journal for Applied Research and Development*, 5, 37–56.

- CWC. (1997). Flood estimation report for Mahanadi Sub-zone 3(d) (Revised). *Directorate of Central Water Commission, New Delhi*.
- DHI (2007). MIKE SHE user manual. *Danish Hydraulic Institute, Denmark*, pp 396.
- Feldman, A.D. (2000). Hydrologic Modeling System HEC-HMS, Technical Reference Manual. *U.S. Army Corps of Engineers, Hydrologic Engineering Center, HEC, Davis, CA, USA*.
- Georgakakos, K.P., Seo, D., Gupta, H., Schaake, J. and Butts, M.B. (2004). Towards the characterization of streamflow simulation uncertainty through multimodel ensembles. *Journal of Hydrology*, 298, 222–241.
- Hamill, T. M. (1997). Reliability diagrams for multicategory probabilistic forecasts. *Weather Forecast*, 12, 736–741.
- Koenker, R. and Basset, G. (1978). Regression quantiles. *Econometrica*, 46: 33-50.
- Koenker, R. and Hallock, K.F. (2001). Quantile regression. *Journal of Economic Perspectives*, 15(4), 143–156.
- Krishnamurti, T. N., Kishtawal, C.M., Zhang, Z., Larow, T., Bachiochi, D. and Williford, E. (2000). Multimodel ensemble forecasts for weather and seasonal climate. *Journal of Climate*, 13, 4196–4216.
- Kristensen, K. J. and Jensen, S. E. (1975). A model for estimating actual evapotranspiration from potential transpiration. *Nordic Hydrology*, 6, 170–188.
- Makridakis, S. and Winkler, R. L. (1983). Averages of forecasts: Some empirical results. *Management Science*, 29 (9), 987–996.
- Marcel G. S., Feike J. L. and Martinus Th. V. G. (2001). ROSETTA: a computer program for estimating soil hydraulic parameters with hierarchical pedo transfer functions. *Journal of Hydrology*, 251, 163–176.
- Mason, S. J. (1982). A model for assessment of weather forecast. *Australian Meteorological Magazine*, 30, 291–303.
- Moriasi, D.N., Arnold, J.G., Van Liew, W.V., Bingner, R.L., Harmel, R.D. and Veith, T.L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE*, 50 (3), 885–900.
- Nash, J. E. (1960). A unit hydrograph study, with particular reference to British catchments. *Proceedings of the Institution of Civil Engineers*, 17, 249–282.
- O'Connell, P.E., Nash, J.E. and Farrell, J.P. (1970). River flow forecasting through conceptual models. Part 2. The Brosna catchment at Ferbane. *Journal of Hydrology*, 10, 317–329.
- Refsgaard, J.C. and B. Storm (1995), MIKE SHE. In: V.P. Singh (ed.), *Computer Models of Watershed Hydrology*, *Water Resources Publications, Highland Ranch, Colorado, USA*, 809–846.

- Sahai, A. K., Chattopadhyay, R. and Goswami, B. N. (2008). A SST based large multi model ensemble forecasting system for Indian summer monsoon rainfall. *Geophysical Research Letters*, 35, L19705, doi: 10.1029/2008GL035461, 2008.
- Singh, H. and Sankarasubramanian, A. (2014). Systematic uncertainty reduction strategies for developing streamflow forecasts utilizing multiple climate models and hydrologic models. *Water Resources Research*, 50, 1288–1307.
- Sugawara, M. (1979). Automatic calibration of the tank model. *Hydrological Sciences Bulletin*, 24, 375–388.
- U. S. Army Corps of Engineers (USACE) (2000). Hydrologic Modeling System: Technical Reference Manual. *U.S. Army Corps of Engineers, Hydrologic Engineering Center, Davis, CA*.
- Velázquez, J. A., Anctil, F., Ramos, M. H. and Perrin, C. (2011). Can a multi-model approach improve hydrological ensemble forecasting? A study on 29 French catchments using 16 hydrological model structures. *Advances in Geosciences*, 29, 33–42.
- Viney, N. R., Bormann, H., Breuer, L., Bronstert, A., Croke, B. F. W., Frede, H., ... Willems, P. (2009). Advances in Water Resources Assessing the impact of land use change on hydrology by ensemble modelling (LUCHEM) II: Ensemble combinations and predictions. *Advances in Water Resources*, 32, 147–158.
- Weerts, A. H. Winsemius, H.C. and Verkade, J. S. (2011). Estimation of predictive hydrological uncertainty using quantile regression: examples from the national flood forecasting system (England and Wales). *Hydrology and Earth System Sciences*, 15, 255–265.
- Wilks, D. S. (1995). Statistical methods in the Atmospheric Sciences: An introduction. *Academic Press*, 449 pp.

