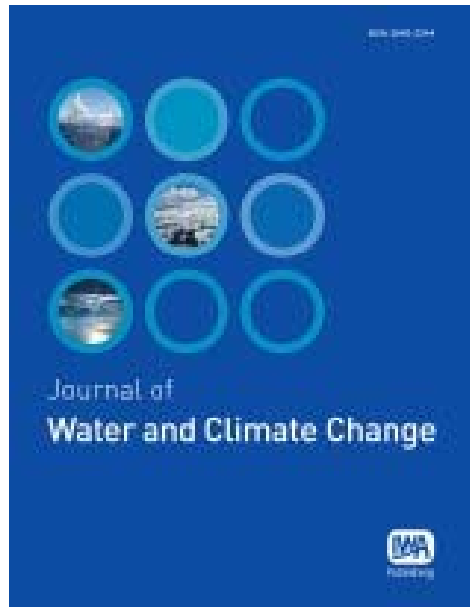


**Provided for non-commercial research and educational use only.
Not for reproduction or distribution or commercial use.**



This article was originally published by IWA Publishing. IWA Publishing recognizes the retention of the right by the author(s) to photocopy or make single electronic copies of the paper for their own personal use, including for their own classroom use, or the personal use of colleagues, provided the copies are not offered for sale and are not distributed in a systematic way outside of their employing institution.

Please note that you are not permitted to post the IWA Publishing PDF version of your paper on your own website or your institution's website or repository.

Please direct any queries regarding use or permissions to jwc@iwap.co.uk

Development of HydroClimatic Conceptual Streamflow (HCCS) model for tropical river basin

Parag P. Bhagwat and Rajib Maity

ABSTRACT

Combined processes of land-surface hydrology and hydroclimatology influence the response of a watershed to different hydroclimatic variables. In this paper, streamflow response of a watershed to hydrometeorological variables is investigated over a part of two tropical Indian rivers – Narmada and Mahanadi. The proposed HydroClimatic Conceptual Streamflow (HCCS) model is able to consider the time-varying basin characteristics and major hydrologic processes to model basin-scale streamflow using climate inputs at a daily scale. In addition, the proposed model is able to provide additional overall estimates of ground water recharge component and evapotranspiration component from the entire basin. Moreover, ability to consider the time-varying watershed characteristics and hydroclimatic inputs renders the proposed model usable for assessment of future streamflow variation. This application is also investigated for both the study basins. In general, the methodological approach of the proposed model can be applied to other tropical basins for daily streamflow modelling as well as future streamflow assessment.

Key words | climate, conceptual model, hydroclimatology, prediction, streamflow

Parag P. Bhagwat
Rajib Maity (corresponding author)
Department of Civil Engineering,
Indian Institute of Technology Kharagpur,
Kharagpur 721302,
West Bengal,
India
E-mail: rajib@civil.iitkgp.ernet.in;
rajibmaity@gmail.com

INTRODUCTION

In the context of a changing climate, identification of streamflow response to other hydroclimatological variables is a research challenge (Kumar & Maity 2008; Maity & Kashid 2011). Most of the existing approaches attempt to consider and conceptualize different hydrological processes. However, time varying watershed characteristics are not considered. As a consequence, consistency (decadal to climatic scale) of the model performance is affected under a changing climate and changing watershed characteristics that are not accounted for. Here lies the importance of this study since recent observation of climate change has added a new dimension towards this overall research direction in hydrologic modelling. The question examined was, is it possible to model watershed response as streamflow with the simultaneous consideration of changing climate and time varying watershed characteristics? As such, it was found that there is a need to develop a streamflow model having few parameters, which will be able to consider time varying watershed characteristics and climatic inputs, and provide

better or at least comparable performance to existing approaches.

Background and literature review

A brief review (with respect to the huge amount of literature available) of existing approaches reveals that the hydrological models for streamflow estimation can be grouped into three broad categories viz: (a) physically based models, (b) conceptual models, and (c) artificial intelligence (AI)-based models. In physically based models, existing knowledge of all possible hydrological processes is represented through a set of mathematical equations. Examples of some popular physically based models include System Hydrologique European (Abbott *et al.* 1986a,b), Better Assessment Science Integrating point and Non-point Sources (EPA 1998), Soil and Water Assessment Tool (SWAT) (Eckhardt & Arnold 2001; Grizzetti *et al.* 2003), and Community Land Model (Lawrence *et al.* 2011). However, an inability for accurate

mathematical representation of hydrological processes considering the huge spatial variability, computational demands and overparameterization effects, and non-availability of required comprehensive data sets make the results of physically based models poor in many cases (Piotrowski & Napiorkowski 2012; Zeng *et al.* 2012).

Conceptual models are also derived using (relatively simpler) mathematical equations, which mainly consider the major hydrological process of the complex hydrological cycle, which is difficult to understand fully. However, in the absence of complete knowledge, they may be represented in a simplified way by means of the system concept. A system is a set of connected parts that form the whole. By using the system concept, effort is directed to the construction of a model relating inputs and outputs rather than to the extremely difficult task of exact representation of the system details, which may not be significant from a practical point of view or may not be known. Nevertheless, knowledge of the physical system helps in developing a conceptually clear model. The major hydrological processes considered are mostly rainfall, evapotranspiration, infiltration, surface runoff, etc., which have significant importance towards the watershed output. As mentioned before, the physically based models require a large number of data and still produce poor results due to a variety of reasons. Hence, conceptual models are used as alternative models for watersheds. Some of the generally used conceptual models are the Stanford watershed model, which is an early finding in conceptual hydrological modelling by Crawford & Linsley (1966), the Institute Royal Meteorology Belgium (IRMB) model (Bultot & Dupriez 1976), HYDROLOG model which was later modified by Chiew & McMahon (1994) to the MODified HYDROLOG (MODHYDROLOG) model, Hydrologiska Byråns Vattenbalansavdelning (HBV) model which was developed at the Swedish Meteorological and Hydrological Institute (SMHI) in 1972 (Bergström 1980), and the Hydrologic Simulation Program-FORTRAN model which is a modified version of Stanford model (Johnson *et al.* 2003), and so on. These models are designed to approximate the general hydrological processes, which dictate the hydrological cycle. However, most of the conceptual models consider all hydrological processes to be lumped together, greatly simplify the hydrological processes involved, and do not consider time-varying

properties of the watershed. As a consequence, the outcome of such models sometimes becomes crude. Sometimes, even the major components, such as evapotranspiration, ground water recharge component and streamflow, are not separated from each other. Moreover, typical absence of short-term (i.e., fortnight to seasonal) and/or long-term (i.e., decadal to centennial) dynamic characteristics raise questions against its applicability to capture the dynamic, time-varying response of watershed to long-term climate change impact analysis studies, which are particularly essential in a changing climate.

AI-based models are much simpler with respect to the understanding of underlying physical processes. These models are developed based on interrelationships between input and output values of the concerned variables. In the last two decades, use of AI-based models has increased due to their lower dependence on physical understanding of the underlying process(es) by the users, and provide quick as well as reasonably impressive results compared with the other two types of models discussed before. Some of the most popular AI-based models, which are applied to hydrological modelling, are based on artificial neural network (Kumar *et al.* 2007), genetic programming (Maity & Kashid 2010), and support vector machines (SVMs). However, apart from the criticism of the 'black-box' nature, the performance of such models is often found to be excellent and difficult to replicate using other conceptual/physically based approaches. This said, the same criticism which was mentioned in the case of conceptual models earlier, exists for such modelling approaches as well. That is, due to their static nature, studies related to long-term climate change impact analysis may not be always possible with this modelling approach.

Motivation and objectives

Specifically, this study attempts to develop a basin-scale daily hydroclimatic streamflow model (named as HydroClimatic Conceptual Streamflow (HCCS) model), considering the time-varying watershed characteristics and major hydrologic processes to model basin-scale streamflow using climate inputs. The developed modelling approach is conceptual in nature. Its performance is investigated for two Indian tropical river basins. Performances of AI-based machine learning approaches are found to overshadow the

performance of other modelling approaches. Thus, the performance of the proposed model is compared with the performance of a highly popular AI-based machine-learning approach known as least square-support vector regression (LS-SVR) (Suykens *et al.* 2002). Performance of the proposed model is also compared with other popular conceptual modelling approaches that are effective at daily scale in addition to AI.

The ability to consider the hydroclimatic inputs is expected to make the proposed HCCS model usable also for future projected climate. Time-varying watershed characteristics are considered that renders the approach dynamic, which is found to be very useful for a changing climate condition and analysis of streamflow variation under future climate. The parameter that controls the time-varying watershed characteristics is conceptualized. This parameter may be projected to the future with different assumptions on change in watershed characteristics and may be used for future. Thus, the specific objectives of this paper are: (1) the development of the HCCS model considering the time varying watershed characteristics and using climatic inputs; (2) application of the developed HCCS model to two Indian river basins (Mahanadi and Narmada) and investigation of their performance; (3) comparison of the developed HCCS model with other popular conceptual modelling approaches as well as with AI-based the LS-SVR model; and (4) study of future streamflow variation under projected climate during 2026–2035, 2076–2085 using the proposed HCCS model.

METHODOLOGY

HCCS model

In the proposed HCCS model, the watershed is treated as a 'system' that receives rainfall, processes it, and generates streamflow as its 'response'. The 'response' depends on various factors, depending on the time-dependent as well as time-invariant characteristics of the watershed. For instance, topology, shape of the catchment are treated as time-invariant whereas wetness condition of watershed, rainfall over catchment, continuous loss due to evaporation, rate of ground water recharge, etc., are treated as time

dependent. The proposed methodology is motivated by SACramento Soil Moisture Accounting model (Burnash *et al.* 1973) and leaky bucket model (Huang *et al.* 1996) as far as an initial water balance equation is concerned. However, subsequent conceptualizations differ as the major focus of these models is on soil moisture simulation whereas the major focus of the proposed HCCS model is on streamflow prediction.

The HCCS model is conceptual in nature and able to predict the daily variation of streamflow while estimating the spatially averaged evapotranspiration loss and ground water recharge from the entire catchment. The model is also suitable for simulating the future streamflow variation using such future projected climate data over the basin. The model is described below.

The 'System Wetness Condition' (SWC) is a representation of the amount of water that is stored in the near-surface strata of the watershed as depression storage, soil water retention, reservoir storage, etc. SWC is time dependent and is denoted as $V(t)$, where t is the time subscript. $V(t)$ ranges between zero and maximum capacity of the system wetness, $V_{\max}$, which is the maximum water holding capacity of the watershed in the form of soil moisture at the top strata of the soil, depression storage, reservoirs, etc. Whereas $V(t)$ changes at a much faster rate depending on the rainfall and other weather conditions, $V_{\max}$ also varies over time but at a much slower pace. $V_{\max}$ depends on the combination of different types of land use land cover (LULC), such as, urbanization, forest cover, existence of reservoirs, etc.

Temporal change of this characteristic ($V_{\max}$) is slower than other influencing parameters, such as, rainfall, wetness condition ($V(t)$) and other weather conditions. In this modelling approach, $V_{\max}$ is also considered to vary over time, which makes it suitable for application over a longer time frame and more importantly in the changing climate. In the application of the proposed model, it is shown that this property changes over a multi-year scale. This is a very important aspect to be considered in hydroclimatic streamflow modelling to study the variation of streamflow under projected future climate owing to changed characteristics of watershed status.

SWC at time t , $V(t)$ is calculated based on the water balance in the watershed. The components of the water balance in the model are precipitation, evaporation, streamflow, and

loss due to ground water recharge. For a watershed, the change in SWC over time can be expressed as

$$\frac{dV(t)}{dt} = P(t) - E(t) - S(t) - G(t) \quad (1)$$

where $V(t)$ is the SWC at time t expressed in terms of depth of water per unit area of watershed, $P(t)$ is the depth of precipitation over the watershed, $E(t)$ is the total evapotranspiration loss from the catchment, which is the sum of evapotranspiration from land surface and direct evaporation loss from depression storages, reservoirs, etc., $S(t)$ represents the streamflow divided by the catchment area of watershed, and $G(t)$ is the loss due to deep percolation that joins ground water as ground water recharge component from the watershed.

As mentioned earlier, $V_{\max}$ is the maximum capacity of the system wetness. Hence, $V(t)/V_{\max}$ denotes dynamic condition of the system or the proportion of SWC, a value of 0 indicates a completely dry system and a value of 1 indicates a fully wet system. In this proposed model, it is assumed that this proportion determines system response and other components in the water balance equation are linked with this. Physically, this factor indicates the overall status of the watershed at time t that controls the loss of water from the system through various components. These are conceptualized as follows.

First, the streamflow is expected to be generated by the system depending on its wetness condition. The generated streamflow (per unit area of the watershed) is conceptualized to have a non-linear relationship with the SWC $V(t)$. If non-dimensionalized with respect to their respective maximum possible values which are denoted as $V_{\max}$ (defined earlier) and $S_{\max}$, the assumption is expressed as

$$\frac{S(t)}{S_{\max}} = a \left(\frac{V(t)}{V_{\max}} \right)^{b'} \quad (2)$$

whereas the conceptualization of $V_{\max}$ is explained earlier, the concept of $S_{\max}$ is not completely new. Rather it is analogous to estimated limiting value (ELV), which is defined as the largest magnitude possible for a hydrologic event at a given location, based on the best available hydrologic information (Chow *et al.* 1988). This assumption, expressed in Equation (2), will be checked for two watersheds considered

in this study later. Equation (2) can be rearranged as follows (for time t):

$$V(t) = B[S(t)]^b \quad (3)$$

where $B = V_{\max}/(aS_{\max})^b$ and $b = 1/b'$.

Another major component is the loss due to evapotranspiration, $E(t)$, which is estimated using the following relation:

$$E(t) = \frac{E_p(t) V(t)}{V_{\max}} \quad (4)$$

where $E_p(t)$ is the potential evapotranspiration and $V(t)$ is the SWC at time t . Substituting $V(t)$ in Equation (4) from Equation (3)

$$E(t) = E_p(t) \frac{B[S(t)]^b}{V_{\max}} \quad (5)$$

Potential evapotranspiration from climatic data can be estimated using any standard method, such as the Hargreaves method (Hargreaves *et al.* 1985; Hargreaves 1994) or the Penman–Monteith method (Monteith 1965). The Hargreaves method is used in this study and presented in Appendix A (available online at <http://www.iwaponline.com/jwc/005/015.pdf>).

The ground water recharge component $G(t)$ can be assumed to be non-linearly associated with SWC, $V(t)$. This is expressed as follows:

$$G(t) = \alpha[V(t)]^\beta \quad (6)$$

For simplicity, β is assumed to be 1, i.e., ground water recharge component is assumed to vary linearly with SWC. This assumption of linearity was also made by earlier researchers in the similar context of ground water recharge component in the leaky bucket model (Huang *et al.* 1996). However, it may be noted here that the development of the subsequent equations is still possible without this assumption. The aforementioned assumption reduces the burden of one extra parameter only. Using Equation (3) in Equation (6) with $\beta = 1$, $G(t)$ can be expressed as

$$G(t) = k[S(t)]^b \quad (7)$$

where $k = \alpha$, B . From Equation (1), expressing the left hand side with finite difference scheme, gives

$$\frac{V(t+1) - V(t)}{\Delta t} = P(t) - E(t) - S(t) - G(t) \quad (8)$$

Substituting the expressions of $V(t)$, $E(t)$ and $G(t)$ in term of $S(t)$ gives

$$\begin{aligned} \frac{B[S(t+1)]^b - B[S(t)]^b}{\Delta t} &= P(t) - E_p(t) \frac{B[S(t)]^b}{V_{\max}} - S(t) - k[S(t)]^b \\ &=> S(t+1) = \\ &\left[\{S(t)\}^b + \frac{\Delta t}{B} \left\{ P(t) - E_p(t) \frac{B[S(t)]^b}{V_{\max}} - S(t) - k[S(t)]^b \right\} \right]^{\frac{1}{b}} \quad (9) \end{aligned}$$

where B , b , k and $V_{\max}$ are the model parameters that characterize the basin considered in the study. All these parameters may change depending on the change in watershed condition, e.g., urbanization, deforestation, construction of dams and reservoirs, etc.

Discussion on different model parameters

The physical analogy of different parameters used in the model is discussed in this section.

Parameter k

The parameter k indicates the net contribution of catchment to ground water recharge. Theoretically, this value may be positive or negative. Spatially, surface water may contribute to ground water at some location whereas ground water may contribute to surface water at some other location within the catchment. Thus, the net contribution to ground water from entire catchment may be positive or negative. If the net contribution to ground water from entire catchment is positive, k will be positive and vice versa. This parameter is unitless.

Parameter $V_{\max}$

The parameter $V_{\max}$ indicates the overall water holding capacity of the watershed. Compared with a virgin river

basin, a basin with many dams and reservoirs is expected to have more water holding capacity. On the other hand, highly urbanized, deforested basins with increased paved surfaces may exhibit low water holding capacity. Thus, for a particular river basin this property is expected to change depending on the status of watershed development, such as, construction of dams and reservoirs, urbanization, deforestation, etc. It is expected to perceive such changes over a multi-year scale. Thus, a proper estimation of the projected status of this parameter should precede the estimation of the future streamflow. This interpretation of time varying characteristics differs, in general, from other standard conceptual models. A historical analysis is carried out for two study river basins in the next section to investigate the change of this parameter over the last couple of decades. This parameter has a unit of length (say m or mm).

Parameters b and B

The inverse of parameter b is the measure of degree of non-linearity between $S(t)/S_{\max}$ and $V(t)/V_{\max}$ as expressed in Equation (2). Parameter B is a function of $V_{\max}$ and $S_{\max} [= V_{\max}/(a S_{\max})^b]$. While making the analogy of $S_{\max}$ with ELV it should not be confused with the fact that the value of $S_{\max}$ does not change over time. Rather it is expected to be modified depending on the change in the characteristics of the upstream catchment. This fact is not unrealistic. For instance the chance and magnitude of flash flood increase with increased urbanization. Thus $S_{\max}$ may also be subjected to change depending on the characteristics of the upstream catchment which is associated with the change in $V_{\max}$. However, these changes are expected in long-term temporal scale (multi-year or decadal) depending on the rate of change in the watershed characteristics.

Parameter estimation, validation, and future projection

Model calibration is performed using the daily streamflow, rainfall, maximum temperature, minimum temperature, and average temperature data. Four parameters of conceptual rainfall-runoff model, i.e., $V_{\max}$, b , B and k , are estimated during the model calibration period. These parameters depend on the catchment characteristics, which influence the system response and may also be interrelated. Thus,

these parameters are to be simultaneously estimated during model calibration by minimizing the mean square error (MSE). The values of the parameters that yield minimum mean square error (MMSE) are used as estimated parameters. During both model development and testing periods, the model performance can be investigated through different statistics that measure the association between the observed and predicted daily streamflow values.

To investigate the slow change in any parameter, parameter estimation is carried out during a couple of years in a decade and the analysis is repeated for successive decades in the past. The estimated parameter values for different periods need to be investigated for possible change over time. This change might be modelled and if a trend is found that might be projected in future assuming different cases of change in watershed characteristics: (1) no further change (latest value is used for future); (2) same trend of change projected to future; and (3) change is slower approaching a constant (a non-linear equation may be fitted to observed parameter values in the past

and projected to future). However, some of these parameters may not show a liner/non-linear trend over the historical time. If no plausible trend information is obtained for some parameters, it is recommended to use either average value of that parameter over the historical period or the latest information of that parameter. This aspect is illustrated later with respect to the study basins considered for demonstration.

STUDY AREAS AND DATA USED

Tropical rainfed river basins in India are either east flowing or west flowing. The performance of the proposed model is investigated over the last couple of decades for two Indian River basins, Mahanadi (east flowing) and Narmada (west flowing). Upstream parts of both these basins are considered. In [Figure 1](#) (Mahanadi) and [Figure 2](#) (Narmada), study areas with locations of climate stations and streamflow gauging stations are shown. Study area considered for

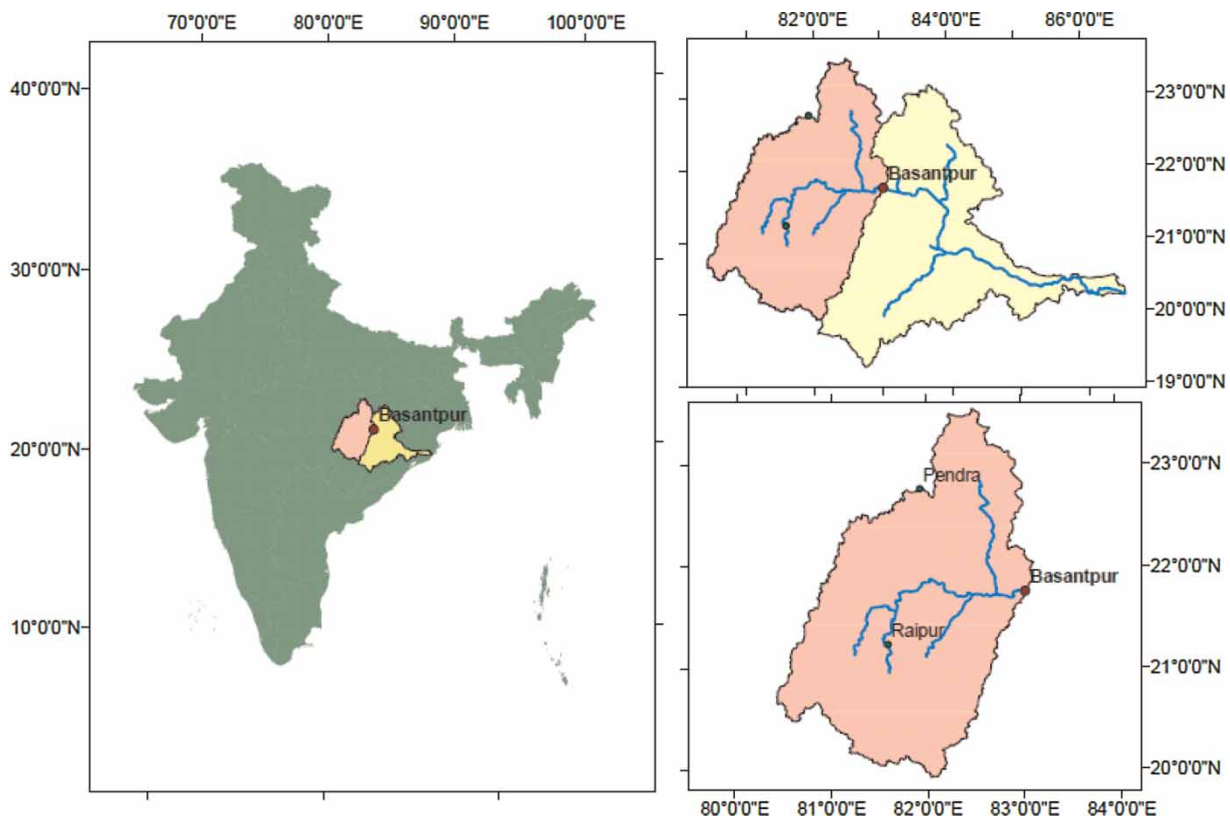


Figure 1 | Mahanadi River basin with the study area (up to Basantpur gauging station).

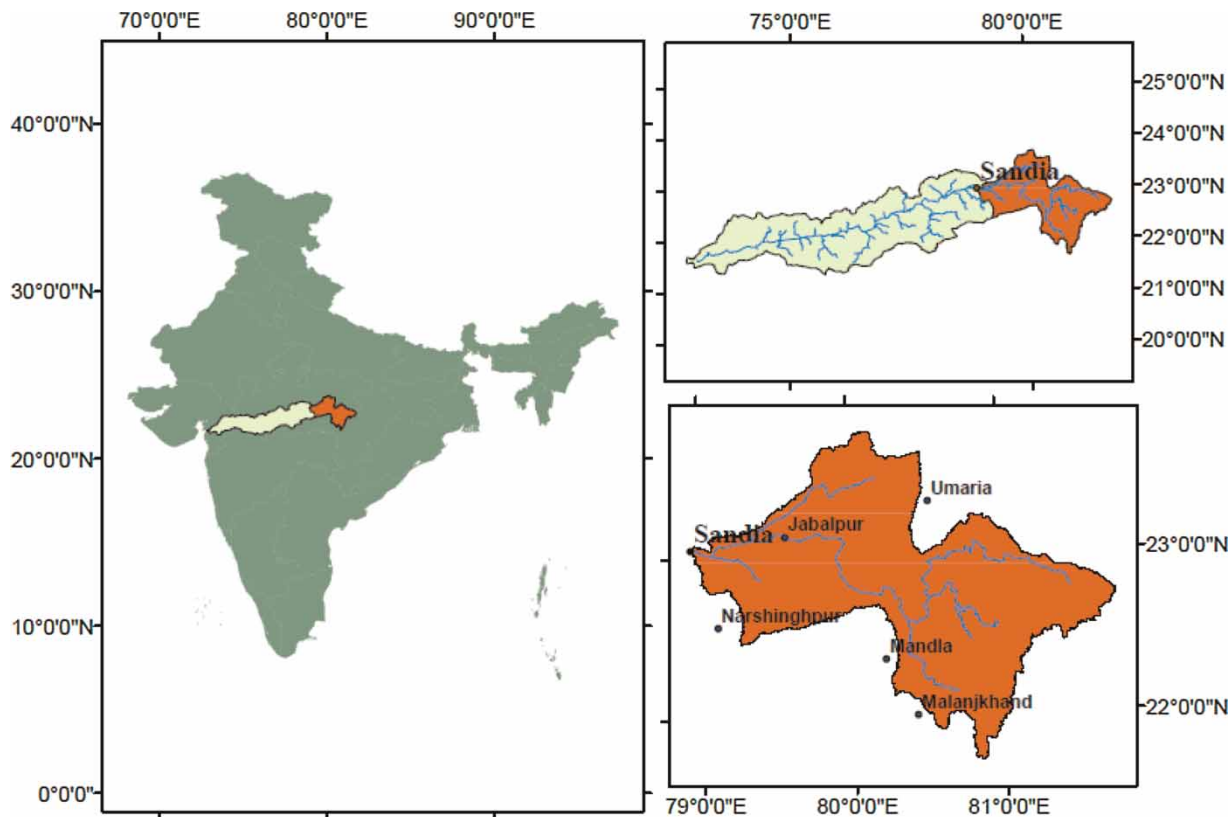


Figure 2 | Narmada River basin with the study area (up to Sandia gauging station).

Mahanadi River basin is 61,152 km² up to Bastantpur gauging station and that for Narmada River basin is 25,912 km² up to Sandia gauging station.

Daily streamflow at Bastantpur (Mahanadi River basin) is obtained from the office of Executive Engineer, Mahanadi Division, Central Water Commission (CWC), India. Daily streamflow data (Jan 1, 1973 to Dec 31, 2003) are used for this study. There is no major structure (apart from few medium and minor reservoirs) in the catchment of Mahanadi up to Bastantpur and there are no missing streamflow data for this basin. Streamflow data at Sandia (Narmada River basin), operated by the Water Resources Agency, are also obtained from CWC for the period June 1, 1978 to May 31, 2000. However, out of this entire duration of 23 years, streamflow data are missing for some non-contiguous periods (total 3 years 8 months). The missing data are from Dec 1, 1987 to May 31, 1988 (6 months); Jun 1, 1993 to May 31, 1994 (1 year); May 1–31, 1996 (1 month); Oct 1, 1996 to Nov 30, 1996 (2 months); Apr 1, 1997 to Aug 31, 1997 (5 months); Jun 1, 1998 to May 31, 1999 (1 year) and Dec 1, 1997 to May 31,

1998 (6 months). Since streamflow data are missing for these periods, other data corresponding to these periods are ignored from the analysis, i.e., these periods are excluded from analysis.

For Mahanadi basin, daily gridded (1° lat × 1° long) rainfall data are obtained from India Meteorological Department (IMD). There are no missing data for rainfall. Daily temperature records (maximum and minimum) from two stations (Raipur and Pendra Road) located in the catchment (see Figure 1) are obtained from the National Climatic Data Centre (NCDC) Climate Data Online (<http://www.ncdc.noaa.gov/cdo-web/>). As per the website information, these data are supplied by IMD to NCDC. Temperature data are missing for some (non-contiguous) days (approximately 250 days out of 31 years). These values are replaced by the long-term average of that date.

Daily observed rainfall and temperature data for the Narmada basin (also obtained from IMD) are recorded at five different stations within or around the basin. These stations are Jabalpur, Malanjkhanda, Mandla, Narshinghpur

and Umaria. Locations of these stations are shown in Figure 2. For some (non-contiguous) days (approximately 150 days out of 23 years), rainfall and temperature data are missing. These values are replaced by the average of other surrounding stations for that date.

RESULTS AND DISCUSSION

Model calibration and testing

Parameter estimation

Model calibration is performed using the daily streamflow, rainfall, maximum temperature, minimum temperature, and average temperature data. Four parameters of HCCS model, i.e., $V_{\max}$, b , B and k are estimated during model calibration (model development period). As mentioned above, an MMSE criterion is followed to estimate these parameters. An initial guess of the range of different parameters is necessary since the parameter may vary from watershed to watershed and also from one period to another period. However, the initial guess of range should be made in such a way that the value of estimated parameter should not lie on the border of the selected range. It is recommended to keep the initial guess as wide as possible at the cost of computation time. Due to the advent of fast computational facilities, computational time is not a major issue. Thus, the ranges of $V_{\max}$ is considered to be 50–1,000 with an interval of computation as 0.1. Similarly, ranges of b , B and k are considered to be 0.1–1, 10–200 and 0.0–0.5 with their interval of computation as 0.01, 1 and 0.001, respectively.

Three-fold model development and testing is adopted for the Mahanadi River basin. On the other hand, two-fold model development and testing is adopted for the Narmada River basin. This is based on the availability of data. For Mahanadi, model parameters are estimated during (a) Jan 1, 1973 to Dec 31, 1980, (b) Jan 1, 1984 to Dec 31, 1990 and (c) Jan 1, 1994 to Dec 31, 2000. Models developed over these periods are tested during (a) Jan 1, 1981 to Dec 31, 1983, (b) Jan 1, 1991 to Dec 31, 1993 and (c) Jan 1, 2001 to Dec 31, 2003, respectively. For the Narmada River basin, June 1978 to May 1987 is used as first development period and June 1990 to May 1997 is used as second

development period. Developed models are used for testing during June 1987 to May 1990 and June 1997 to May 2000. Estimated parameters for both the basins are shown in Table 1.

For both the river basins, it is noticed that the parameters vary over different periods of analysis. This issue is investigated with respect to LULC change in the context of future streamflow variation later. Before that, the model performances during development and testing periods are investigated to assess the potential of the developed model.

Model performances

Model performances are investigated for all the development periods and the corresponding testing periods for both the basins. Model performances are presented through (i) time series plot of observed and predicted streamflow to compare their correspondence, (ii) a scatter plot between observed and predicted streamflow to compare their association and (iii) statistical measures to quantify and assess the potential of prediction. The statistical measures include mean square error (MSE), correlation coefficient (CC), and Nash-Sutcliffe efficiency (NSE).

Model performance during the three-fold development and testing periods for the Mahanadi River is shown in Figure 3. A comparison plot between observed and predicted streamflow for the first model development period (Jan 01, 1973 to Dec 31, 1980) is shown in top panel of Figure 3(a). MSE, CC and NSE values for this period are found to be 1.133, 0.93 and 0.86, respectively (Table 2). Model parameters are estimated from this the period (Jan 01, 1973 to Dec 31, 1980) and the developed model is tested during Jan 01, 1981 to Dec 31, 1983 (First model testing period). The bottom panel of Figure 3(a) shows the comparison plots

Table 1 | Estimated Model Parameters during Different Development Periods

Basin Name	Calibration Period	Model Parameters			
		B	b	k	$V_{\max}$
Mahanadi	Jan 01, 1973 to Dec 31, 1980	22	0.63	0.113	214.0
	Jan 01, 1984 to Dec 31, 1990	98	0.28	0.043	244.4
	Jan 01, 1994 to Dec 31, 2000	48	0.54	0.080	270.3
Narmada	Jun 01, 1978 to May 31, 1987	26	0.57	0.093	315.0
	Jun 01, 1990 to May 31, 1997	52	0.44	0.015	316.3

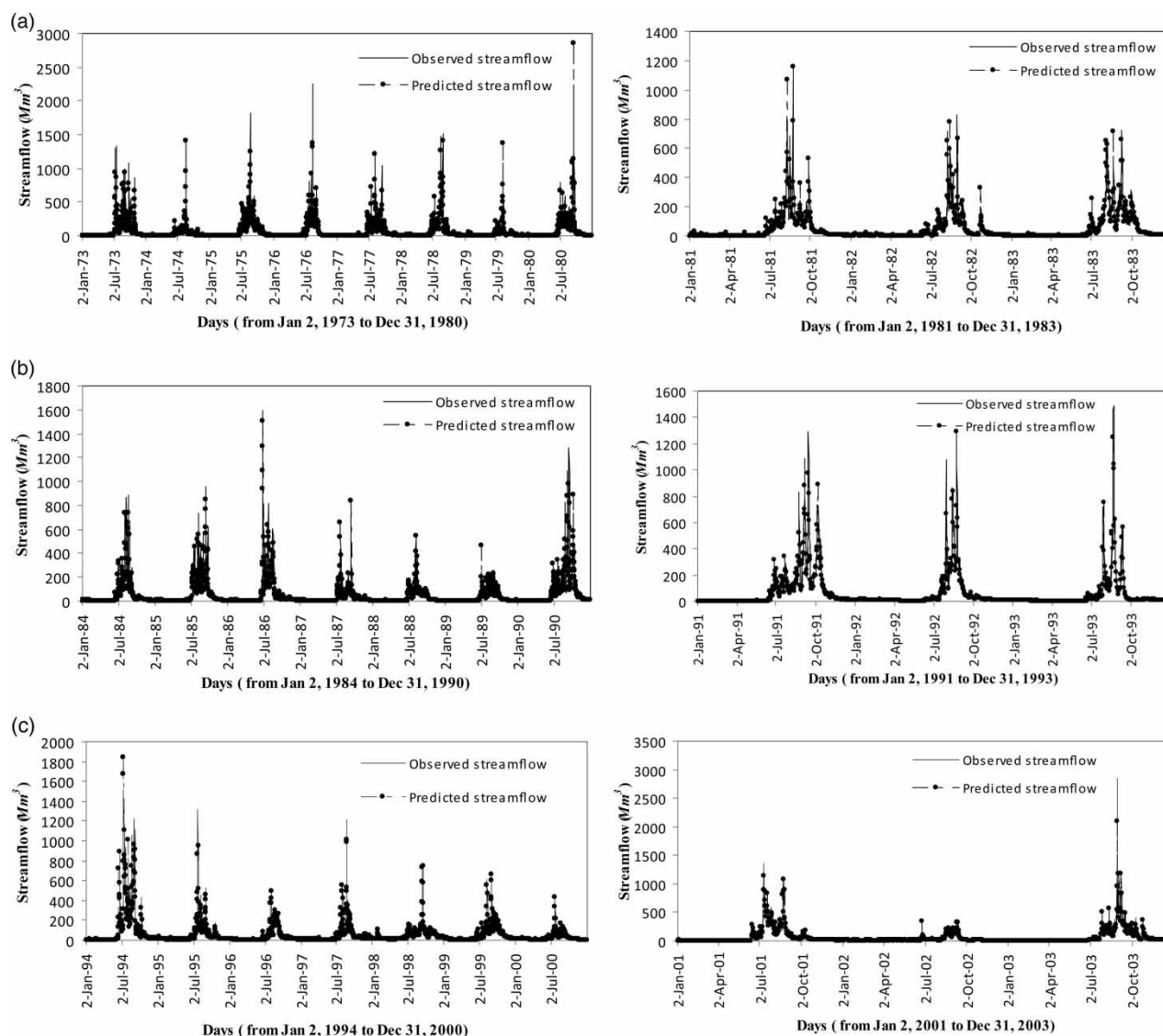


Figure 3 | (a) Observed and predicted streamflow for first development period (Jan 02, 1973 to Dec 31, 1980) and corresponding testing period (Jan 02, 1981 to Dec 31, 1983) for Mahanadi River Basin. (b) Observed and predicted streamflow for second development period (Jan 02, 1984 to Dec 31, 1990) and corresponding testing period (Jan 02, 1991 to Dec 31, 1993) for Mahanadi River Basin. (c) Observed and predicted streamflow for third development period (Jan 02, 1994 to Dec 31, 2000) and corresponding testing period (Jan 02, 2001 to Dec 31, 2003) for Mahanadi River Basin.

Table 2 | Model performance statistics for development and testing periods. Testing period values are shown in parentheses

Basin Name	Development Period (Testing Period)	Performance Statistics		
		MSE	CC	NSE
Mahanadi	Jan 01, 1973 to Dec 31, 1980 (Jan 01, 1981 to Dec 31, 1983)	1.133 (0.379)	0.93 (0.94)	0.86 (0.88)
	Jan 01, 1984 to Dec 31, 1990 (Jan 01, 1991 to Dec 31, 1993)	0.546 (0.601)	0.93 (0.95)	0.87 (0.90)
	Jan 01, 1994 to Dec 31, 2000 (Jan 01, 2001 to Dec 31, 2003)	0.551 (1.513)	0.94 (0.88)	0.88 (0.77)
Narmada	Jun 01, 1978 to May 31, 1987 (Jun 01, 1987 to May 31, 1990)	4.43 (2.77)	0.83 (0.86)	0.69 (0.72)
	Jun 01, 1990 to May 31, 1997 (Jun 01, 1997 to May 31, 2000)	4.66 (7.20)	0.90 (0.96)	0.81 (0.86)

between observed and predicted streamflows during this period. MSE, CC and NSE values for this period are found to be 0.379, 0.94 and 0.88, respectively (Table 2). For the second and third periods of analyses, plots between observed and predicted streamflows are shown in Figure 3(b) and 3(c), respectively, with the development period (upper) and the testing period (lower) shown in each.

Comparison plots between observed and predicted streamflow for Narmada River basin during Jul 01, 1978 to May 31, 1987 (First model development period) and corresponding testing period (Jun 02, 1987 to May 31, 1990) are shown in Figure 4(a), top and bottom panel, respectively. Similarly, the plots between observed and predicted streamflows during Jun 01, 1990 to May 31, 1997 (Second model development period) and corresponding testing period (Jun 01, 1997 to May 31, 2000) are shown in the top and bottom panel, respectively, of Figure 4(b). In addition to these plots, readers may refer to the scatter plots between observed and predicted streamflow values during all the

development and testing periods. These are shown in the supplementary document (Figures S-1–S-10, available online at <http://www.iwaponline.com/jwc/005/015.pdf>).

Visual inspection of these plots reveals that the performance of the proposed HCCS model is reasonably good for both Mahanadi and Narmada River basins. Performance in terms of statistical measures for all pairs of development and testing periods for both the study basins are shown previously in Table 2. The high value of NSE (range ~ 0.69–0.9) and CC (range ~ 0.83–0.96) between observed and predicted streamflow indicate the efficacy of the proposed model. In some cases, slightly better values of these statistics during the testing period as compared with the development period may be contradictory to the usual experience. However, this is not impossible. A close inspection of observed and predicted values for such cases during development and testing periods reveals the existence of some very high values in the development period. Such high values are not present during the testing period. Thus, the model is

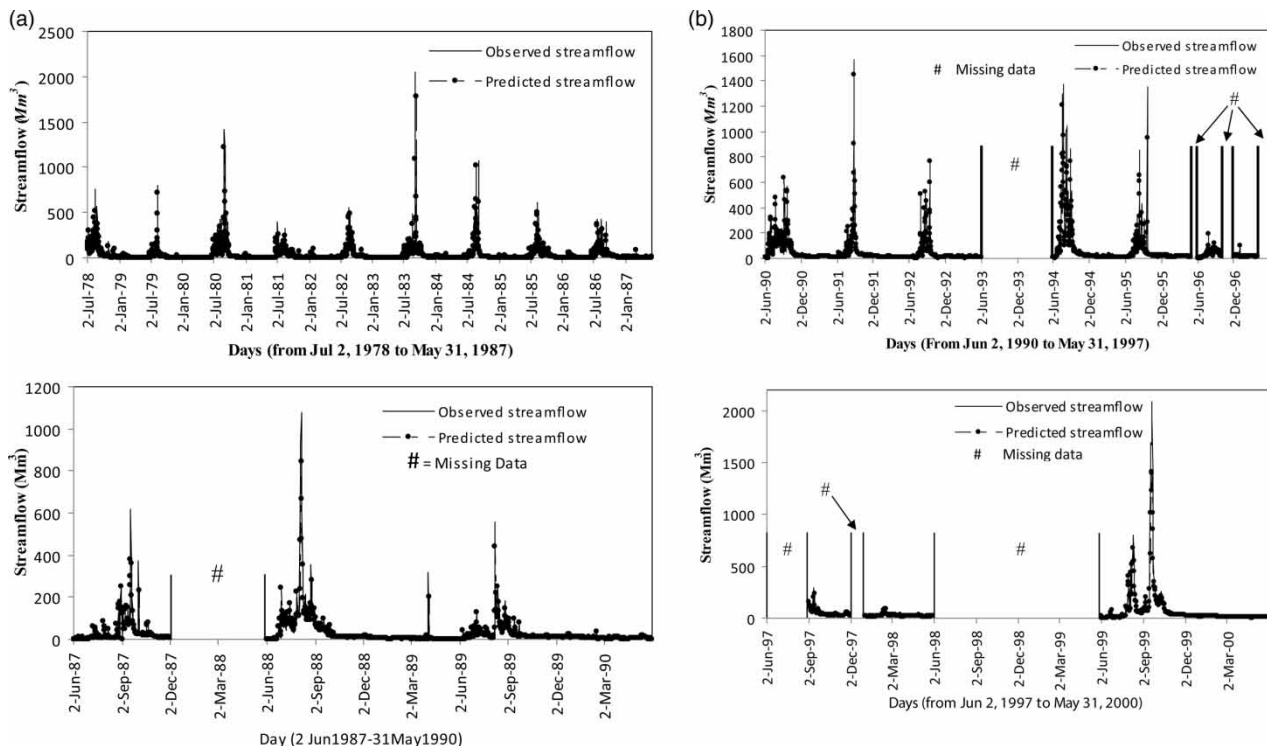


Figure 4 | (a) Observed and predicted streamflow for the first development period (Jul 02, 1978 to May 31, 1987) and corresponding testing period (Jun 02, 1987 to May 31, 1990) for Narmada River Basin. (b) Observed and predicted streamflow for the second development period (Jun 02, 1990 to May 31, 1997) and corresponding testing period (Jun 02, 1997 to May 31, 2000) for Narmada River basin.

able to predict with slightly better accuracy during the testing period.

While comparing the performance between two basins, it is noticed that the performance of Mahanadi River basin even better than that of the Narmada River. This is due to the fact that the available data length is more and discontinuity/missing data are less for Mahanadi basin. Thus, availability of longer data length leads to better performance of the proposed HCCS model.

Other outputs from HCCS model

As mentioned before, the ground water component and evaporation components can also be estimated as other outputs of the model. To note again, these estimates are spatially averaged magnitudes over the entire catchment. The daily series of these components and monthly estimates (derived from the daily estimates) are shown for Narmada River Basin for the period Jun 1987 to May 1990 (Figure 5). These estimates for Mahanadi River Basin for the period Jan 1981 to Dec 1983 are shown in Figure 6. Figures for the rest of periods are shown in

Figures S-11–S-13 (available online at <http://www.iwaponline.com/jwc/005/015.pdf>). Though these estimates are not compared with the observed values (due to non-availability), seasonality in those estimated is visible and matches with the nature of Indian hydroclimatology. However, some observations may appear contrary to the general experience. For instance, in 1987, the actual evapotranspiration value in October (14.15 mm) is more than that in June (4.06 mm) and July (6.31 mm) for Narmada. It is due to the notably different SWC ($V(t)$) for these months. As mentioned in the Methodology section, actual evapotranspiration $E(t)$, is conceptualized as a function of potential evapotranspiration ($E_p(t)$) and $V(t)$. For instance, $V(t)$ was found to be very low in July due to the sustained low rainfall. In June (just previous month), total observed rainfall (56.7 mm) was much lower (55%) than that is observed normally (mean = 122.3 mm). Thus, even if the total observed rainfall in July (299.0 mm) was near normal (mean = 278.3 mm), $E(t)$ was found to be low.

It may be further noted that the values of ground water component and evaporation components seems to

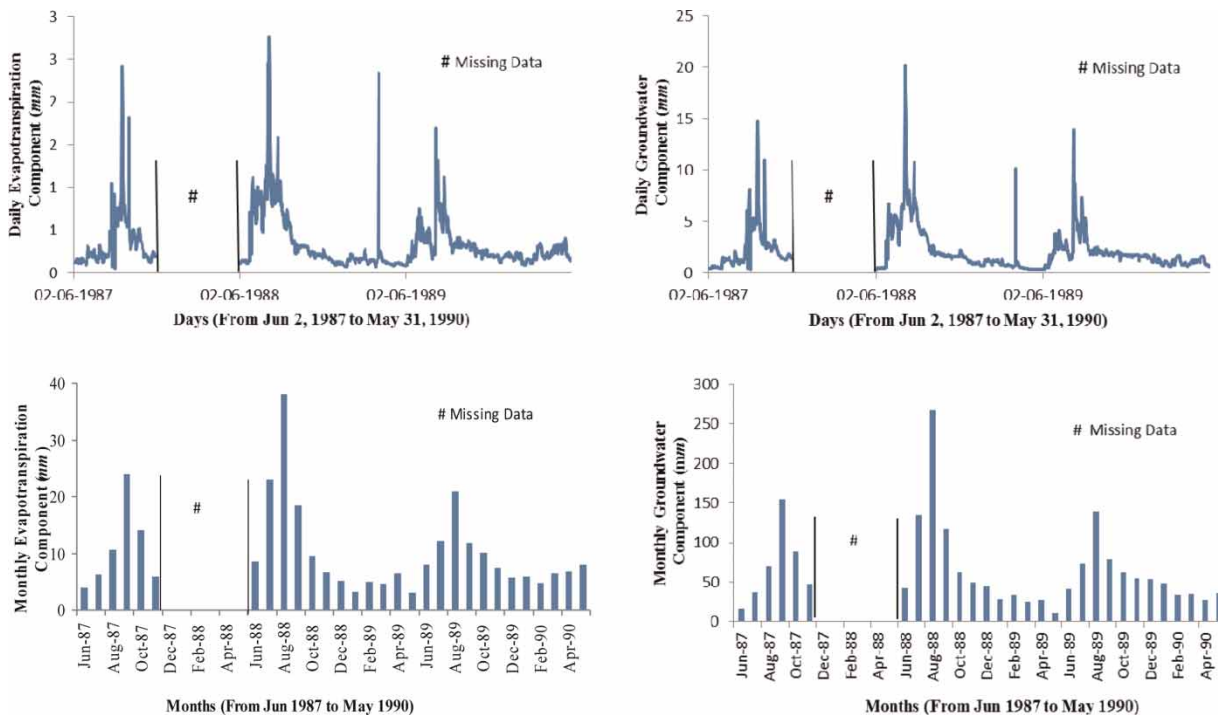


Figure 5 | Daily (top) and monthly (bottom) evapotranspiration (left) and ground water (right) component for Narmada River basin for the period Jun 1987 to May 1990.

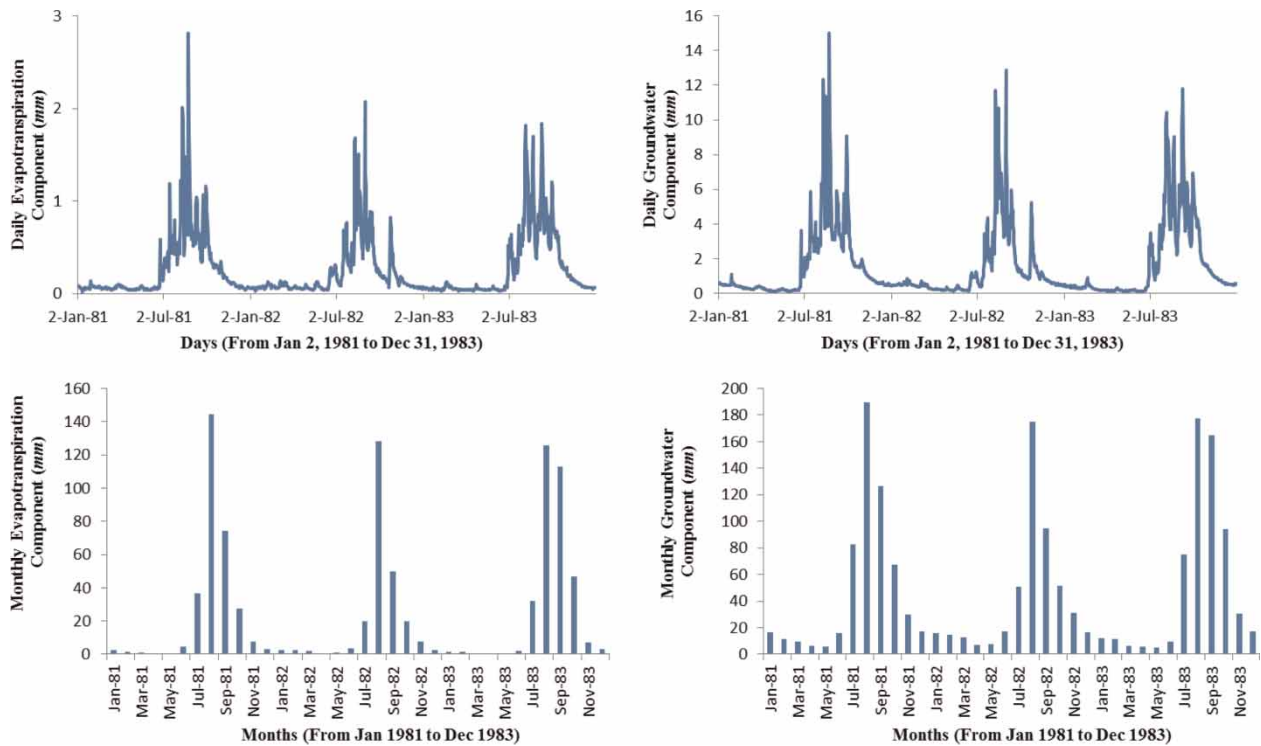


Figure 6 | Daily (left) and monthly (right) evapotranspiration and ground water (right) component for Mahanadi River basin for the period Jan 1981 to Dec 1983.

correspond well with each other. As reported in Equation (4), actual evapotranspiration is conceptualized as a function of potential evapotranspiration and SWC. Whereas ground water recharge is conceptualized as a non-linear function of SWC (Equation (6)). Thus, the correspondence between these two outputs is indeed due to the model structure. It was also mentioned earlier that the parameter β in Equation (6) is assumed to be 1, i.e., ground water recharge component is assumed to vary linearly with SWC. Thus, as can be noticed from Equation (6), non-linearity can be invoked at the cost one additional parameter. This may result in the non-linear association between these model outputs.

While comparing these estimates between two basins, it is observed that, in general, the evapotranspiration is less in Narmada than Mahanadi. This is due to a combined effect of low SWC (V_t) and meteorological conditions responsible for potential evapotranspiration. On the other hand, ground water recharge component varies over different time periods both for Narmada and for Mahanadi. SWC (V_t) is responsible for this as conceptualized in the methodology, resulting in time-varying recharge amount.

Relationship between $S(t)$ and $V(t)$

As explained in the Methodology, that the streamflow (per unit area of the watershed), $S(t)$ is conceptualized to have a non-linear relationship with the SWC, $V(t)$. In this section, model generated $V(t)$ and observed streamflow values ($S(t)$) are analyzed to investigate this issue. Different plots are prepared for Narmada and Mahanadi River basin. Percentage explained through a non-linear curve of form, $y = ax^b$ is computed, and displayed in those figures. Figure 7 shows such plots for Narmada basin during model development and testing periods. Similar plots for the Mahanadi River basin are shown in Figure 8. It can be observed that for most of the cases, the best fit non-linear curves can explain, on an average 93% (range 74–98%) of the association between observed $S(t)$ and model generated $V(t)$. For Mahanadi River basin, the percentage explained by the non-linear curve is even better than that observed for Narmada River. This is perhaps due to the existence of a reservoir (Bargi) that started operating in 1990 inducing the effect of human operation. However, a sort of generalization on the

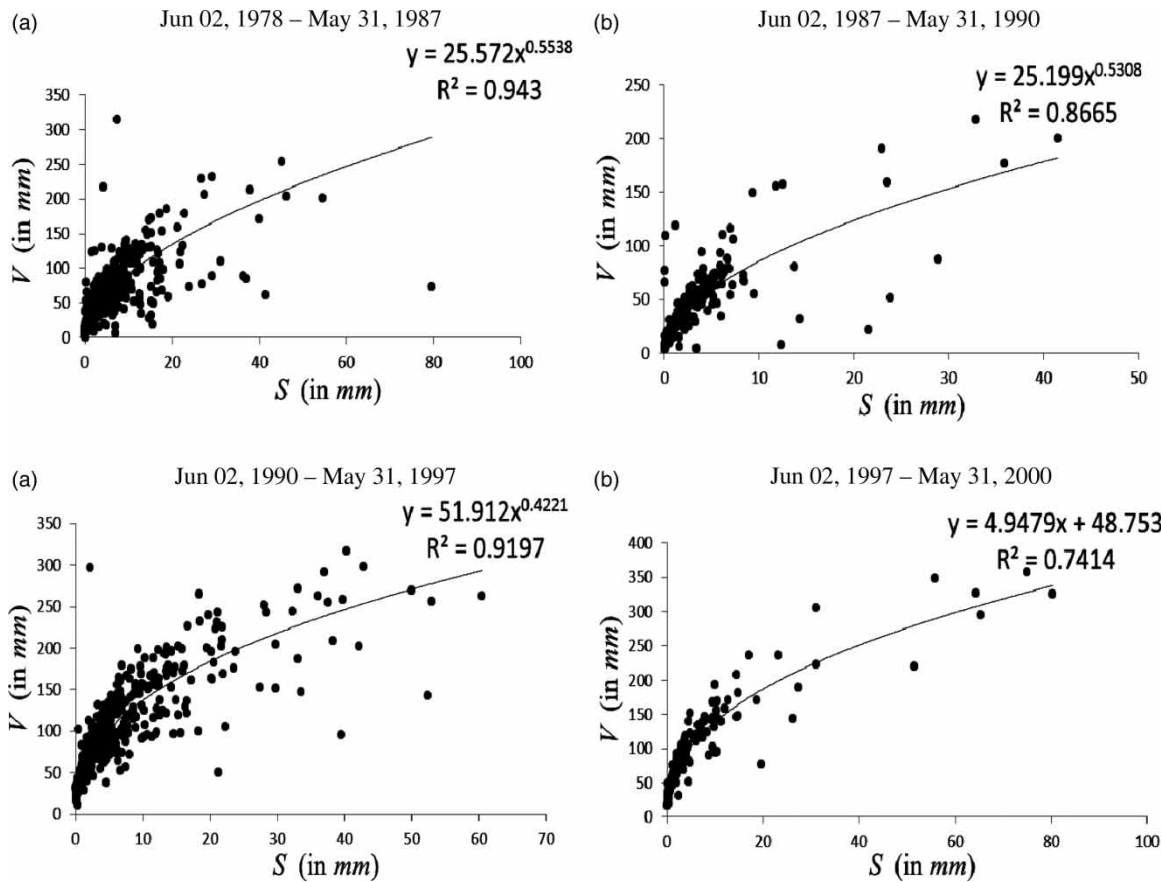


Figure 7 | Relationship between observed $S(t)$ and modeled $V(t)$ for Narmada River basin during different periods.

nature of this curve for both the basins can be made. Towards this, a linear equation between observed $S(t)$ and model generated $V(t)$ was also derived (not shown in figures). It is observed that the percentage of association captured by the linear equation drops to as low as 58% (range 58–77%). This confirms the assumption of a non-linear relationship between streamflow and SWC made in the Methodology.

The equations for the best-fit non-linear curves are shown in the respective plots. The coefficients are expected to be similar to the estimated parameters. By comparing the equations of the best-fit curves with the Equation (3) in the Methodology, the multiplying constants and the powers are the estimates of B and b , respectively. By comparing them with the values shown in Table 1, it is noticed that the estimated parameters and these values are fairly close to each other.

Comparison of performances of proposed HCCS model with other models

HCCS model performance is compared with other conceptual models that are effective at daily scale. These are Australian Water Balance Model (AWBM) (Boughton 1993; Boughton & Carroll 1993), Sacramento model (Burnash *et al.* 1973), SIMplified HYDrolog (SIMHYD) (Porter 1972; Porter & McMahon 1975, 1976), Soil Moisture Accounting and Routing (SMAR) model (O'Connell *et al.* 1970) and Tank Model (Sugawara 1967; Sugawara *et al.* 1974). Details of these models and its software package one can be found from Rainfall Runoff Library (RRL) (available at <http://www.toolkit.net.au/Tools/RRL>). The performances of these models are shown in Table 3a along with their parameter values for the study period (presented in the supplementary document in Tables

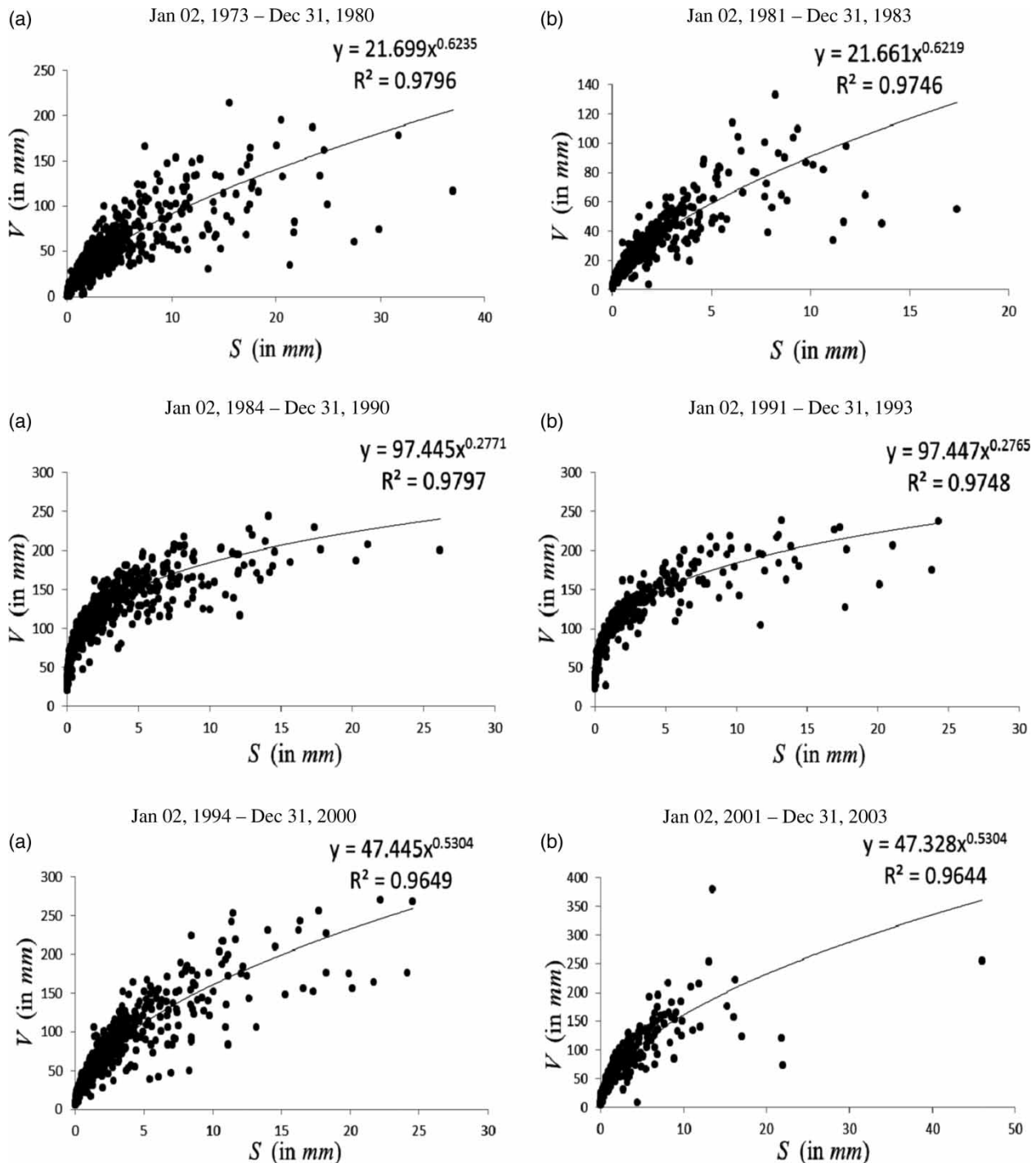


Figure 8 | Relationship between streamflow $S(t)$ and modeled $V(t)$ for Mahanadi River basin.

S1–S5, available online at <http://www.iwaponline.com/jwc/005/015.pdf>. It is noticed that the performances of all these models are inferior (if not remarkably inferior in some cases) to the proposed HCCS model.

It is worthwhile to mention here (based on the literature review) that the performance of AI-based machine learning approaches are found to overshadow the performance of other modelling approaches. Thus, the performance of the

Table 3a | Performances of other conceptual models during different model development and testing periods (within parentheses)

Basin Name	Development Period (Testing Period)	AWBM			Sacramento			SimHyd			SWAR			Tank		
		MSE	CC	NSE	MSE	CC	NSE	MSE	CC	NSE	MSE	CC	NSE	MSE	CC	NSE
Mahanadi	Jan 01, 1973 to Dec 31, 1980 (Jan 01, 1981 to Dec 31, 1983)	2.10 (0.88)	0.86 (0.86)	0.74 (0.71)	2.56 (1.13)	0.83 (0.85)	0.68 (0.63)	2.69 (0.99)	0.82 (0.83)	0.66 (0.68)	2.76 (1.77)	0.82 (0.71)	0.65 (0.66)	2.87 (1.44)	0.82 (0.79)	0.64 (0.53)
	Jan 01, 1984 to Dec 31, 1990 (Jan 01, 1991 to Dec 31, 1993)	1.08 (1.47)	0.86 (0.85)	0.73 (0.72)	1.11 (1.72)	0.85 (0.82)	0.73 (0.67)	1.20 (1.68)	0.84 (0.82)	0.71 (0.68)	1.38 (1.98)	0.82 (0.80)	0.66 (0.62)	1.91 (2.58)	0.81 (0.79)	0.53 (0.50)
	Jan 01, 1994 to Dec 31, 2000 (Jan 01, 2001 to Dec 31, 2003)	1.44 (2.60)	0.85 (0.79)	0.70 (0.63)	1.47 (3.64)	0.85 (0.75)	0.70 (0.47)	1.38 (2.90)	0.85 (0.77)	0.71 (0.58)	2.34 (4.44)	0.75 (0.65)	0.52 (0.36)	1.63 (3.30)	0.83 (0.74)	0.66 (0.52)
	Jun 01, 1978 to May 31, 1987 (Jun 01, 1987 to May 31, 1990)	6.04 (7.41)	0.76 (0.67)	0.57 (0.49)	6.85 (7.18)	0.73 (0.67)	0.52 (0.43)	6.04 (6.98)	0.76 (0.69)	0.58 (0.44)	8.02 (8.08)	0.66 (0.61)	0.44 (0.35)	5.87 (6.79)	0.77 (0.69)	0.59 (0.46)
Narmada	Jun 01, 1990 to May 31, 1997 (Jun 01, 1997 to May 31, 2000)	12.90 (40.6)	0.73 (0.82)	0.49 (0.23)	11.91 (39.4)	0.77 (0.87)	0.53 (0.25)	13.53 (40.2)	0.70 (0.72)	0.47 (0.24)	11.90 (33.9)	0.74 (0.86)	0.53 (0.37)	12.52 (33.87)	0.71 (0.70)	0.51 (0.35)

proposed model is also compared only with a highly popular AI-based machine learning approach known as LS-SVR (Suykens *et al.* 2002). The streamflow is expressed in the same way (divided by the catchment area) as in the HCCS model to facilitate the performance comparison. Rainfall, maximum temperature, minimum temperature, and lagged streamflow values are used as input data. Streamflow of the current day is considered as the output. It is established by Bray & Han (2004) that SVM-based approaches with normalized input data outperform those with non-normalized input data. Therefore, the data are normalized and finally the model outputs are back-transformed to their original range by denormalization. The normalization (also back-transformation) is done using (Samsudin *et al.* 2011),

$$y_i = 0.1 + \frac{S_i}{1.2 \max(S_i)} \quad (10)$$

where y_i are the normalized data for i th day, S_i is the observed value for i th day and $\max(S_i)$ is the maximum of all S_i (during the model development period).

The LS-SVR model has two parameters (regularization parameter, γ , kernel parameter, σ) to be determined. These parameters are interdependent, and their (near) optimal values are often obtained by a trial and error method. The grid search method is used to find the optimum parameters of LS-SVR. The CC is used for selecting the best performing model parameters. Different statistical measures are used to evaluate the performance of the model during development and testing period.

Both the study areas are explained in detail above. Training and testing data for Mahanadi River basin are from period Jan 1, 1973 to Dec 31, 1995 and Jan 1, 1996 to Dec 31, 2003, respectively, and for Narmada River basin, they are Jun 1, 1978 to May 31, 1995 and Jun 1, 1995 to May 31, 2000, respectively.

The performance of both study areas is assessed with the remaining testing data points. Model performance statistics are obtained between observed and modeled river flow values during training and testing period. The combination that yields comparable performance during training and testing period is selected, which ensures the optimum parameter values without the fear of overfitting. The values of γ and σ^2 for which the difference between the training and testing period performances is minimum, are identified. It

is found from that the best combination of γ and σ^2 is 1 and 0.36, respectively, for the Mahanadi River basin. Similarly, for the Narmada River basin the best combination of γ and σ^2 is 25 and 0.36, respectively.

Model performance during the training and testing period is assessed in terms of aforementioned statistical measures and shown in Table 3b. Comparison figures between observed and predicted streamflows are shown in the supplementary document (Figures S-14–S-17, available online at <http://www.iwaponline.com/jwc/005/015.pdf>) for both the river basins. Fair correspondence between observed and predicted river flow values is found. In low and medium range of flow values, the correspondence is found to be better. Different statistical measures like CC, MSE, and NSE are used to determine the performance of the developed LS-SVR model. Values of MSE, CC and NSE for Mahanadi River basin are 1.333, 0.89 and 0.79, respectively, during the training period, and are 1.011, 0.85 and 0.73, respectively, during the testing period (Table 3b). Similarly, for the Narmada River basin, these values are found to be 4.545, 0.86 and 0.75 (training), and 9.901, 0.86 and 0.70 (testing), respectively.

These performance measures are compared with the performance of the proposed HCCS model. Average NSE for the proposed HCCS model for Mahanadi river is found to be 0.87 whereas the same for LS-SVR (even using a longer data length) is 0.79. Similarly, the CC and MSE are also found to be better in case of HCCS model than that of LS-SVR for both Mahanadi and Narmada rivers. Only in the case of first development period (Jun 01, 1978 to May 31, 1987) and testing (Jun 01, 1987 to May 31, 1990) for Narmada river, performance of HCCS model was found to be less than that of LS-SVR while considering CC and NSE values. Again, MSE is better for HCCS model than LS-SVR for this set of development and testing periods.

Table 3b | Performances of LS-SVR during training period and testing period (within parentheses)

River Basin	MSE Training (Testing)*	CC Training (Testing)*	NSE Training (Testing)*
Mahanadi	1.333 (1.011)	0.89 (0.85)	0.79 (0.73)
Narmada	4.545 (9.901)	0.86 (0.86)	0.75 (0.70)

* For Mahanadi: Training period – Jan 1, 1973 to Dec 31, 1995; Testing period – Jan 1, 1996 to Dec 31, 2003.

* For Narmada: Training period – Jun 1, 1978 to May 31, 1995; Testing period – Jun 1, 1995 to May 31, 2000.

Thus, in brief, while comparing the performance of the proposed HCCS model (Table 2) with other conceptual models (Table 3a) and LS-SVR (Table 3b), overall it is found that the performance of the proposed HCCS model is better.

FUTURE STREAMFLOW VARIATION

It was mentioned earlier that in the proposed approach, time-varying watershed characteristics are considered that render the approach dynamic. Moreover, ability to consider the hydroclimatic inputs (rainfall, maximum temperature, minimum temperature, and average temperature) is useful for studying the change in streamflow under projected future climate. Thus, the proposed HCCS model is used for future climate study. The same watersheds are considered for this purpose to compare the current streamflow variation with the future streamflow variations over different months in the year.

Temporal change in maximum system wetness capacity ($V_{\max}$) and other parameters

Temporal change of $V_{\max}$ is a very important aspect to be considered in hydroclimatic modelling to study the variation of streamflow under a changing climate and also the changing characteristics of watershed. Maximum system wetness capacity ($V_{\max}$) and its variation over time are the unique characteristics of a particular watershed. As mentioned before, this concept is based on the physical properties and their change due to a combination of human activities (urbanization, deforestation, construction of reservoirs, etc.) and the effect of climatic change over the basin. To investigate this aspect with respect to the observed changes in the basin for a physical explanation, LANDSAT data are obtained from Earth Science Data Interface (ESDI) at the Global Land Cover Facility (available at <http://glcfapp.glcf.umd.edu:8080/esdi/>). Though good quality images are not available before 2000 at a continuous interval, analysis is carried out with the images that are available during some of the years over different study periods. Having this limitation on data availability, changes in LULC over the study basins are investigated. Three major categories, namely vegetation (forest,

agriculture, grassland, etc.), water body and settlement, are considered to be able to make a physical analogy with change in $V_{\max}$. Results are shown in Table 4 for both the basins. A little perturbation of total area in different years is due to some unclassified pixels, which can be ignored. Further, the conversion of a particular class to another class is also shown in Tables 5a and 5b (for Mahanadi), and Tables 5c and 5d (for Narmada).

Each cell of Tables 5a–d indicates the magnitude change from LULC type shown in row name to the type shown in column heading for that cell. Changes in different LULC classes are clearly noticed and a possible link with the change in the parameter $V_{\max}$ is indicated. $V_{\max}$ is supposed to increase (decrease) with the increase (decrease) in area under the water body. On the other hand, $V_{\max}$ is supposed to decrease due to conversion of land from vegetation or water body to settlement. However, if the land under vegetation converted into a water body (due to inundation of

Table 4 | Change in LULC over for the study basins

Basin name	Description	Area (km ²) in the year of		
		1972*	1990*	2000*
Mahanadi	Vegetation	52,505	52,665	45,957
	Water Body	1,473	1,463	2,758
	Settlement	7,026	7,023	12,435
Narmada	Vegetation	24,038	22,978	21,172
	Water Body	525	709	774
	Settlement	1,812	2,689	4,430

*Image dates for the study area in Mahanadi basin.

1972: Image dates – Dec 15, 1972 (1 block); Dec 16, 1972 (3 blocks); Dec 17, 1972 (1 block); Nov 21, 1975; Jan 08, 1977; Feb 27, 1973; .

1990: Image dates – Nov 10, 1990 (3 blocks); Nov 17, 1990 (3 blocks).

2000: Image dates – Nov 20, 2000 (1 block); Dec 15, 2000 (2 blocks); Dec 09, 2001 (2 blocks); Nov 11, 1999 (1 block);.

*Image dates for the study area in Narmada basin.

1972: Image dates – Dec 16, 1972; Dec 17, 1972; Nov 30, 1972; Feb 02, 1973 (1 block each).

1990: Image dates – Nov 21, 1989 (1 block); Nov 17, 1990 (2 blocks) .

2000: Image dates –Dec 29, 2000; Nov 20, 2000; Dec 09, 2001 (1 block each).

Table 5a | LULC change matrix in Mahanadi basin (up to Basantpur) between 1972 and 1990. Each cell indicates the magnitude (percentage) change from LULC type shown in row name to the type shown in column heading for that cell

Area (km ²) in the year of 1972	Area (km ²) in the year of 1990		
	Vegetation	Water Body	Settlement
Vegetation	47,902 (91.26%)	814 (1.55%)	3,774 (7.19%)
Water Body	460 (31.23%)	449 (30.48%)	564 (38.29%)
Settlement	4,153 (59.11%)	200 (2.85%)	2,673 (38.04%)

Table 5b | Land use and land cover change matrix in Mahanadi basin (up to Basantpur) between 1990 and 2000. Cell description as per Table 5a

Area (km ²) in the year of 1990	Area (km ²) in the year of 2000		
	Vegetation	Water Body	Settlement
Vegetation	41,277 (78.38%)	1,938 (3.68%)	9,450 (17.94%)
Water Body	518 (35.41%)	673 (46.00%)	272 (18.59%)
Settlement	4,163 (59.28%)	146 (2.08%)	2,714 (38.64%)

Table 5c | Land use and land cover change matrix in Narmada basin (up to Sandia) between 1972 and 1990. Cell description as per Table 5a

Area (km ²) in the year of 1972	Area (km ²) in the year of 1990		
	Vegetation	Water Body	Settlement
Vegetation	21,219 (88.31%)	557 (2.32%)	2,251 (9.37%)
Water Body	437 (83.24%)	73 (13.90%)	15 (2.86%)
Settlement	1,312 (72.45%)	79 (4.36%)	420 (23.19%)

Table 5d | Land use and land cover change matrix in Narmada basin (up to Sandia) between 1990 and 2000. Cell description as per Table 5a

Area (km ²) in the year of 1990	Area (km ²) in the year of 2000		
	Vegetation	Water Body	Settlement
Vegetation	19,161 (83.41%)	485 (2.11%)	3,327 (14.48%)
Water Body	331 (46.69%)	279 (39.35%)	99 (13.96%)
Settlement	1,675 (62.31%)	10 (0.37%)	1,003 (37.31%)

vegetation behind the newly constructed reservoir), $V_{\max}$ is expected to increase. For Mahanadi, significant increase in the water body is observed from 1990 to 2000 (almost 89%). There is a decrease in vegetation by 6,700 km² out of which almost 2,000 km² is converted to water body. These changes cause increase in $V_{\max}$. A gradual increase (77% over a 30-year period) in settlement was also found that might cause a decrease in $V_{\max}$. A combination of all these effects results in an overall increase in $V_{\max}$.

On the other hand, change in water body at Narmada basin from 1972 to 1990/2000 is due to construction of the Bargi Dam in 1989–90. As a result, $V_{\max}$ is supposed to increase. However, at the same time, significant increase in settlement (144% over 30 years) and decrease in vegetation cause $V_{\max}$ to decrease. As a combined outcome, $V_{\max}$ remains almost same (Table 1). Thus, $V_{\max}$ reflects a combined response of all types of LULC changes.

For other parameters (B , k and b), we have not found any plausible trend, though fluctuations are observed. A slight decreasing trend in k can be attributed to the decrease in vegetation which contributes maximum percentage of the total basin area. Further, increase in settlement also supports the decreasing trend of k . However, fluctuations are also observed in k as in other parameters (B and b) without any noticeable trend. Thus, the latest values of these parameters are used for the future. Use of latest value for some parameters may have an effect on the uncertainty in the streamflow estimate for the future time. However, maximum uncertainty may occur at a daily scale, that is, the day-to-day variation of streamflow, which is not the goal of this part of the analysis. For future time, we have shown a month-wise variation of streamflow, averaged over a decade (2026–2035 and 2076–2085). This may also not be free from the uncertainty; however, the effect might be less crucial than that at a daily scale.

As stated above, we were not able to provide a theoretical argument better than this due to much less information on LULC during the analysis period and lack of time synchronization between analysis periods and date of the satellite image. Secondly (not an argument though), even if it was possible to develop a theoretical relationship with improved LULC information, that may not be useful in the context of future assessment for which future information on LULC is required, which might be very tricky, if not impossible since LULC map is not available for future time. Thus, a trend line approach is adopted to project $V_{\max}$ into the future, which is discussed in the subsequent sections.

Optimum data length to estimate $V_{\max}$

First, the optimum length of data needed to get a more or less stable estimate of $V_{\max}$ is investigated. This is carried out by estimating the value of the parameter over a variable length of data (3–20 years). Estimated values of $V_{\max}$ are plotted against the data length. Figure 9 shows the variation of $V_{\max}$ values with respect to the data length for Narmada (top panel) and Mahanadi (bottom panel) basin, respectively. $V_{\max}$ is estimated deterministically, not probabilistically. Since $V_{\max}$ is not estimated probabilistically, it is not possible to show whether the values are statistically different or not. Rather the trend of the change in $V_{\max}$, shown in Figure 9,

can be tested for statistical significance. The trend being non-linear, is tested at logarithmic scale and null hypothesis of not having any trend (zero slope) is comfortably rejected with very low p-values ($\sim 10^{-6}$ – 10^{-7}). Construction of the Bargi dam was completed in 1989–90 that leads to 13/14 year data period and beyond in Figure 9. This is shown in this figure, which reflects a very small jump in $V_{\max}$ value from 15 to 17 years of data length. It is necessary to note that the gross storage capacity (GSC) of Bargi 392 McM. Let us take a typical value of $V_{\max} = 200$ mm. In volumetric unit (multiplied with catchment area), this leads to 5,182 McM. This means the construction of Bargi dam adds only a small part to $V_{\max}$ (7.6%). As indicated in the discussion on model parameters above, the concept of $V_{\max}$ reflects the combined effect of the entire catchment, in which reservoirs are only a part of it. For instance, urbanization and deforestation lead to a decrease in $V_{\max}$. Thus, $V_{\max}$ reflects a combined response of all such changes, which was also explained above with respect to the LULC change.

It can be seen from Figure 9 that the estimate of $V_{\max}$ gets more or less stable as the data length increases for both the river basins. For Narmada, the estimate remains more or less the same for data length greater than approximately 7 years whereas the same for Mahanadi River basin is observed for a data length approximately 5 years. To investigate the temporal change in $V_{\max}$, considered data length should not be too short to obtain a wrong (noisy) estimate and also, should not be too lengthy to miss the meaningful temporal trend/pattern, if any. Considering this, a data length of successive 5 to 10-year periods can be considered to assess the change of $V_{\max}$ over time.

Temporal change in $V_{\max}$

Estimates of $V_{\max}$ for successive 5-year periods (with one or two periods of different lengths) are obtained for both the basins. Results for Narmada River basin is shown in Figure 10. Similar analysis is performed for Mahanadi River basin and shown in Figure 11. In general, it is observed from these figures that these estimates show an overall gradual increase in $V_{\max}$. A smooth variation of $V_{\max}$ over time is observed in case of Narmada River basin whereas high fluctuations are noticed for the Mahanadi River basin.

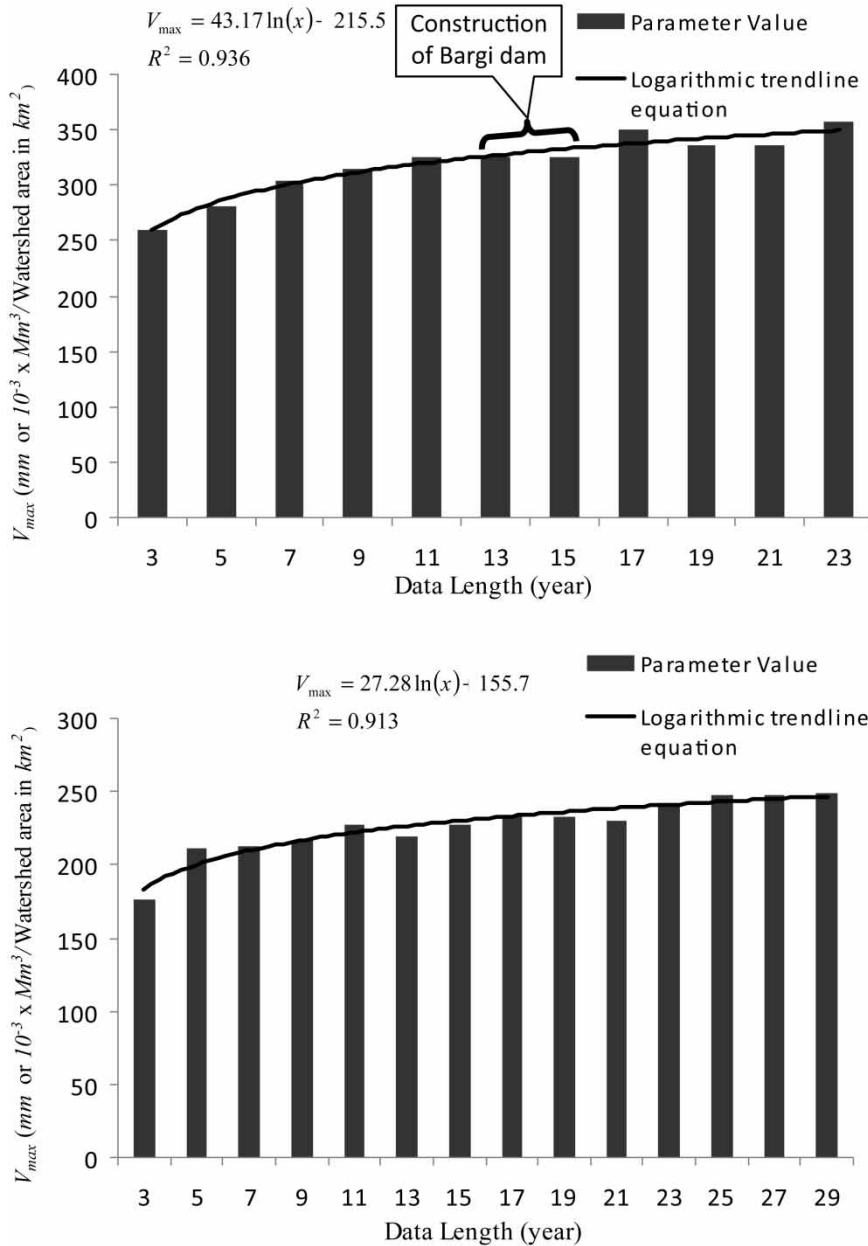


Figure 9 | Variation of $V_{\max}$ values with respect to the data length for Narmada River (top) and Mahanadi River (bottom).

Linear as well as logarithmic trend lines for the historical change in the estimates of $V_{\max}$ are also shown in Figures 10 and 11. For Narmada, linear trend line is described by the following equation:

$$V_{\max} = 26.3X + 233.7 \quad (11)$$

and the equation for best-fit logarithmic trend line is found to be

$$V_{\max} = 59.5 \ln(X) + 254.1 \quad (12)$$

where X is expressed as

$$X = \frac{(\text{Mid year} - 1^{\text{st}} \text{ Mid year})}{5} + 1 \quad (13)$$

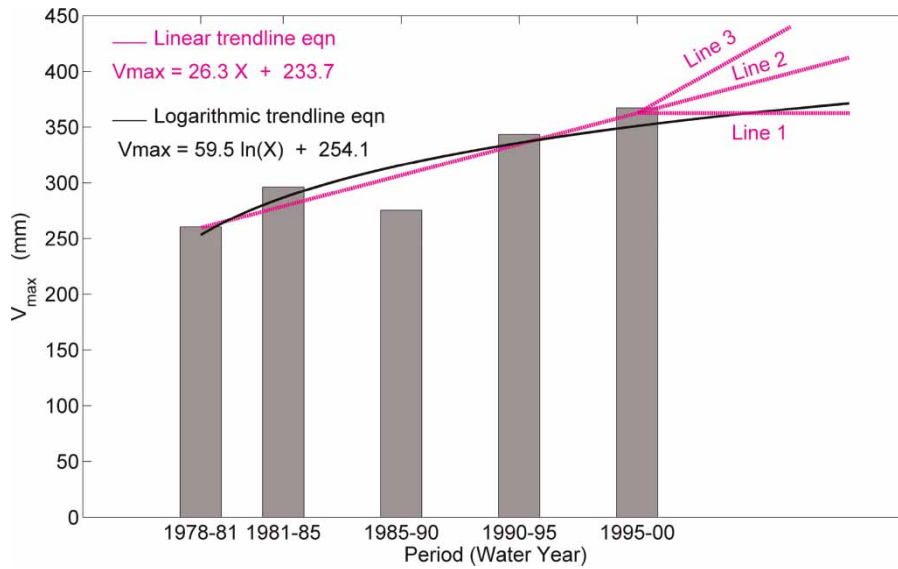


Figure 10 | Change of $V_{\max}$ over time for Narmada River fitted with linear and logarithmic trend lines. Here $X = (\text{Mid year} - 1^{\text{st}} \text{Mid year})/5 + 1$ and 1st Mid year for this case is 1979.5. Refer text for the explanation for three projected lines.

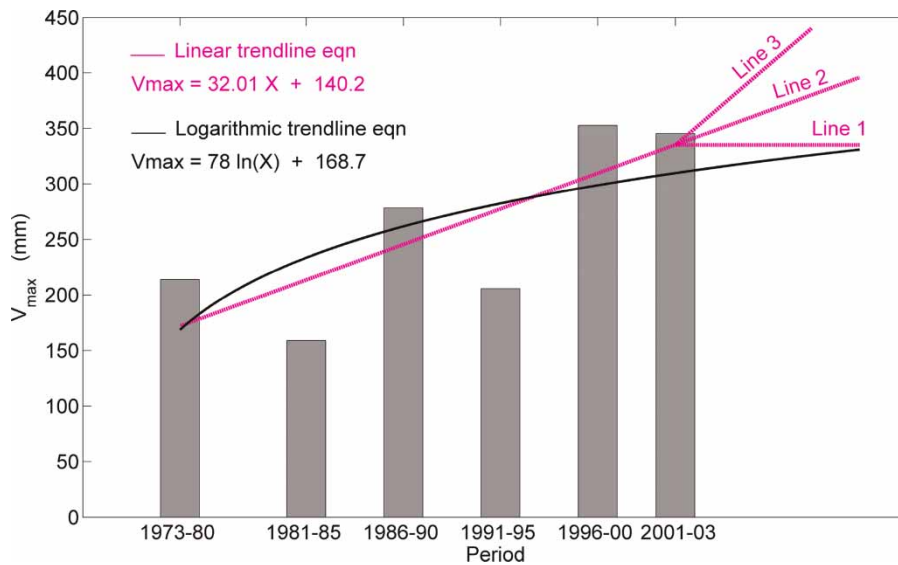


Figure 11 | Change of $V_{\max}$ over time for Mahanadi River fitted with linear and logarithmic trend lines. Here $X = (\text{Mid year} - 1^{\text{st}} \text{Mid year})/5 + 1$ and 1st Mid year for this case is 1976.5. Refer text for the explanation for three projected lines.

‘Mid year’ in Equation (13) is the middle of the period for which an estimate of $V_{\max}$ is required. For instance, 1978 is the ‘Mid year’ for the period 1976–1980 and 2030.5 is the ‘Mid year’ for the period 2026–2035. ‘1st Mid year’ for Narmada (as seen from Figure 10) is 1979.5.

Similarly, for Mahanadi River, linear trend line is described by

$$V_{\max} = 32.01 X + 140.2 \quad (14)$$

and the equation for best-fit logarithmic trend line is found to be

$$V_{\max} = 78 \ln(X) + 168.7 \quad (15)$$

where X is expressed similarly as shown in Equation (13) and '1st Mid year' for Mahanadi (as seen from Figure 11) is 1976.5.

For the linear trend line, the changes in the future may follow three possible paths. These are: (i) remain constant (line 1 – untouched henceforth); (ii) increase with the same linear trend observed in the past (line 2 – business-as-usual scenario); and (iii) increase at a faster rate than that observed in the past (line 3 – intense water management/utility activity in terms of construction of reservoirs). These are also shown in Figures 10 and 11. It might be very difficult to quantify the growth rate of the trend line without linking it to the policy makers' decision for the basin. For instance, higher demand of water may lead to construction of new reservoir, which will increase the water storage capacity for the basin. Thus, it is very difficult to quantify this rate. However, if an overall trend of change is found for the entire catchment that can be projected in future with specific assumptions. Here we demonstrate the future projection by adopting a logarithmic trend line that makes an assumption of a continuous increase over time but at a slower rate as time passes by. Please note that a logarithmic growth rate (and no change also) is used in the analysis; continuous rate of increase is not adopted. The logarithmic growth rate is estimated from the historical trend as explained before. On the other hand, a logarithmic trend line indicates a continuous increase over time but at a slower rate in the future.

Future climate data – PRECIS data

Daily rainfall and temperature data (maximum and minimum) for future climate of both river basins are obtained from the Indian Institute of Tropical Meteorology (IITM), Pune. For this study, averaged data over two future periods (2026–35 and 2076–85) are used to analyze the future streamflow scenario for both study basins.

These data were generated using the widely popular Hadley Centre's high resolution Regional Climate Model

(RCM), known as PRECIS (Providing REgional Climates for Impact Studies) (Jones *et al.* 2004). PRECIS simulations corresponding to the IPCC-SRES A1B emission scenario are carried out for a continuous period of 1961–2098. The base line period considered is from 1961 to 1990. As is well known, RCMs dynamically downscale global model simulations to superimpose the regional details of specific regions of interest. Analysis is carried out at the Indian Institute of Tropical Meteorology (IITM), Pune, to develop the high-resolution climate change scenarios for impact assessment studies (Krishna Kumar *et al.* 2011). While carrying out a rigorous analysis, they found that the model shows reasonable skill in simulating the monsoon climate over India.

Specific to the study basins, the correspondence is checked between observed and modelled (PRECIS) rainfall, maximum and minimum temperature at monthly scale over the study basins during the historical period of analysis (Narmada basin: July 1978 to May 2000; and Mahanadi basin: January 1973 to December 2003). Correlation coefficient (r) and index of agreement ($d1$) (Willmott *et al.* 2011) are computed to assess the correspondence. These statistics are expressed as

$$r = \frac{\sqrt{\sum_{i=1}^n (P_i - \bar{P})(O_i - \bar{O})}}{\sqrt{\sum_{i=1}^n (P_i - \bar{P})^2} \sqrt{\sum_{i=1}^n (O_i - \bar{O})^2}} \quad (16)$$

$$d1 = 1 - \frac{\sum_{i=1}^n |P_i - O_i|}{\sum_{i=1}^n (|P_i - \bar{O}| + |O_i - \bar{O}|)} \quad (17)$$

r varies from -1 to $+1$; -1 indicates negative association, 0 indicates no association and $+1$ indicates positive association. $d1$ varies from 0 to 1 ; 0 indicates complete disagreement and 1 indicates perfect agreement. Results are shown in Table 6. It is noticed that the observed and PRECIS data for all the variables correspond very well to each other. Among the three variables, minimum temperature is found to be the best in terms of both CC and degree of agreement. Rainfall is found to have a higher degree of agreement as compared with maximum temperature in both the basins.

Table 6 | Correspondence between observed and modeled (PRECIS) data over the study basins during the historical period of analysis (Narmada basin – July 1978 to May 2000 and Mahanadi basin – January 1973 to December 2003)

Basin Name	Variable	Statistics	
		<i>r</i>	<i>d1</i>
Mahanadi	Rainfall	0.74	0.69
	Max. temperature	0.79	0.63
	Min. temperature	0.93	0.83
Narmada	Rainfall	0.77	0.72
	Max. temperature	0.82	0.64
	Min. temperature	0.94	0.78

Streamflow scenario under future climate

Possible future streamflow variation using two possible cases, namely ‘no further change’ and ‘logarithmic growth’, for both the basins are investigated and analyses are carried out at daily scale. However, daily results are converted to monthly as it is more meaningful than daily variation for future periods (after 20 or 70 years) and then change in $V_{\max}$ for Narmada and Mahanadi River basin is analysed and, based on the ‘best fit line’, the $V_{\max}$ values are projected for both basins. The projected $V_{\max}$ values are used to model future streamflow generation. Other parameters like b , B and k are kept constant, as observed in the last decade training period. Results are reported separately in the following sections.

Narmada River basin

In ‘no further change’ condition, the same $V_{\max}$ values are considered, i.e., $V_{\max}$ for Narmada River basin is 316.3 mm for both future periods (2026–35 and 2076–85). Another possible condition is ‘logarithmic growth’ where $V_{\max}$ varies in a logarithmic scale. In this case, the $V_{\max}$ values for the Narmada River basin are 397.39 mm for 2026–35 and 435.62 mm for 2076–85. These are computed using Equation (12). Monthly variations of streamflow for the Narmada River basin are shown in Figure 12 for ‘no further change’ and ‘logarithmic growth’ conditions, respectively.

Average month-wise variations of streamflow values are shown for the past two decades (1978–1989 and 1990–2000) as well as future two decades (2026–2035 and 2076–2085) for comparison. Though there is not much difference for

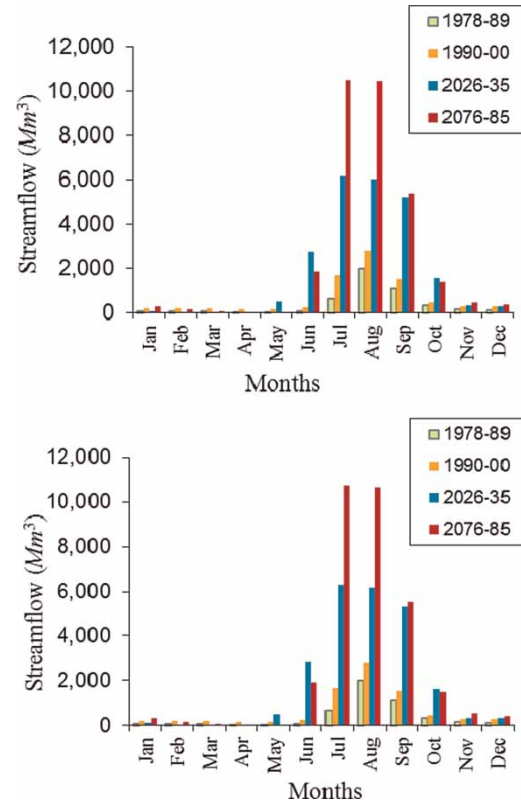


Figure 12 | Possible future monthly streamflow variation using ‘no further change’ (top) and ‘logarithmic growth’ (bottom) condition for Narmada River.

both the possible conditions, i.e., for ‘no further change’ and for ‘logarithmic growth’ conditions, it is found that streamflow magnitudes are going to increase in comparison to the past observations, particularly for monsoon periods. Also for the post-monsoon period, the streamflow is going to increase marginally.

Mahanadi River basin

For the Mahanadi River basin similar analysis is performed. Two possible future conditions (‘no further change’ and ‘logarithmic growth’) are analyzed. The results are shown in Figure 13 for ‘no further change’ and ‘logarithmic growth’ conditions. In ‘no further change’ condition, the same $V_{\max}$ values (270.3 mm) are considered for both future periods (2026–35 and 2076–85). For ‘logarithmic growth’ conditions, $V_{\max}$ values are computed using Equation (15). $V_{\max}$ values are obtained as 360.53 mm for 2026–35 and 408.71 mm for 2076–85. Figure 13 shows the comparison between four periods – past two decades

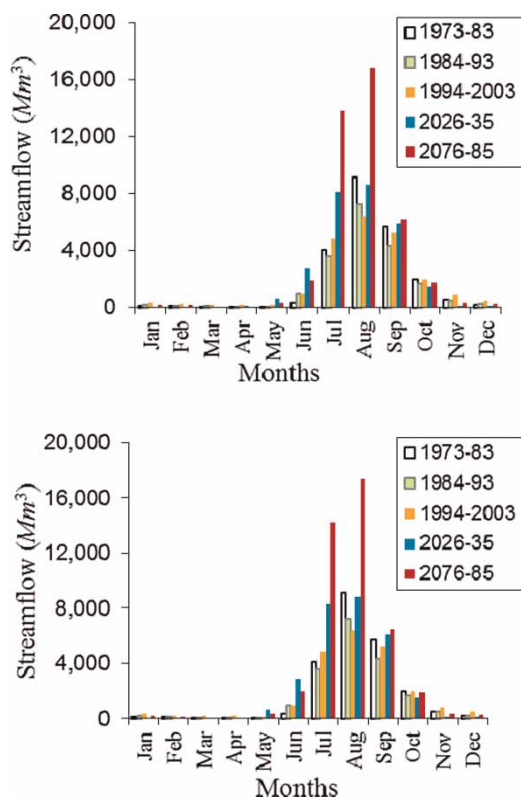


Figure 13 | Possible future monthly streamflow variation using 'no further change' (top) and 'logarithmic growth' (bottom) condition for Mahanadi River.

(1978–1989 and 1990–2000) as well as future two decades (2026–2035 and 2076–2085). For Mahanadi River basin also, streamflow magnitudes are found to increase for future time periods. It is further observed that these increased magnitudes are high particularly for the months of June to August.

These observations indicate that in the near future, monsoon is going to be more concentrated over fewer months. In general, it is found that streamflow is going to increase for the monsoon period for both the study areas. Thus, specific measures may be required in order to control streamflow in future. Implication of the temporal redistribution of streamflow magnitude lies in modification of water management, hydraulic design practices, etc. Since the monsoon streamflow increases and non-monsoon streamflow decreases, higher volume of storage requirement might be necessary in future in order to meet the required demand. Design practices for different hydraulic structures should be suitably revised in order to avoid natural hazards due to extreme streamflow during monsoon months.

Generalization for other tropical river basins

The methodology, being general and having some potential aspects, can be applied to any other tropical watersheds. It is recommended to have a good representative rainfall, temperature and streamflow time series for the study basin. Model parameters are to be obtained from this historical information to capture the watershed characteristics. Watershed characteristics can be more accurately represented through the parameters if the streamflow series consists of all possible flow ranges depicting all possible conditions that are expected to be available in a reasonably long period of data set. A long period of information will also ensure the representation of the gradual change in parameters for the study basin over the successive period of 5–10 years. Thus, a minimum of 30 years of data is recommended. Once the historical trends of different parameters are obtained, these are to be projected to the future time of interest. Since, different watershed might have gone through different process of LULC change, each watershed will have a unique signature in the trend of parameter values over the historical period. Thus, the trend has to be uniquely determined for individual watersheds. These are some of the important aspects in generalizing the approach, presented in this paper, for other tropical basin.

SUMMARY AND CONCLUSIONS

Recent observations on the impacts of climate change motivate to develop a basin-scale streamflow model having few parameters, which will be able to consider time varying watershed characteristics and climatic inputs, and provide better or at least comparable performance to that of AI-based approaches. In this paper, HCCS is developed considering the time varying watershed characteristics and using climatic inputs. Performance of the proposed HCCS model is investigated over last couple of decades for two Indian river basins, Mahanadi and Narmada. Ability to consider the time-varying watershed characteristics and hydroclimatic inputs renders the proposed model usable for assessment of future streamflow variation. The HCCS model is also applied to study future streamflow variation for both the basins.

The specific conclusions from this study are as follows.

The proposed HCCS model is able to consider the time-varying watershed characteristics and major hydrologic processes to model basin-scale streamflow using climate inputs. Performance of the proposed HCCS model is found to be impressive for both the study basins considered in the study. While comparing the performance of the proposed HCCS model with the performance of other popular conceptual models and with LS-SVR (one of the popular AI-based machine learning approaches), it is found that the overall performance of HCCS model is better in general and remarkably better in some cases.

In addition, the proposed model is also able to provide additional overall (spatially averaged) estimates of ground water recharge component and evapotranspiration component from the entire catchment. Though these estimates are not compared with the observed values (due to non-availability), seasonality in these estimated is visible which matches with the reality for Indian hydroclimatology.

The proposed HCCS model is also suitable for future streamflow modelling with projected climate data. The proposed HCCS model is used to model future streamflow variation utilizing the projected climate data (PRECIS data) during two future periods from 2026–2035 and 2076–2085 using two possible cases related to watershed characteristic changes, namely, ‘no further change’ and ‘logarithmic growth’. Compared with historical observation, it is observed that the streamflow magnitudes are going to increase during early monsoon months and marginally increase or remain almost same during the late monsoon and non-monsoon months.

The methodology, being general and having some potential aspects, can be applied to any other tropical watersheds. However, the methodology needs a good representative rainfall, temperature and streamflow time series, consisting of all possible flow ranges, for estimation of parameters during calibration.

ACKNOWLEDGMENT

The study is partially supported by a project sponsored by Space Application Centre (SAC), Indian Space Research Organisation (ISRO), Ahmedabad.

REFERENCES

- Abbott, M. B., Bathurst, J. C., Cunge, J. A., O'Connell, P. E. & Rasmussen, J. 1986a [An introduction to the European Hydrological System – Système Hydrologique Européen, SHE – 1: history and philosophy of a physically based, distributed modelling system](#). *Journal of Hydrology* **87**, 45–59.
- Abbott, M. B., Bathurst, J. C., Cunge, J. A., O'Connell, P. E. & Rasmussen, J. 1986b [An introduction to the European Hydrological System – Système Hydrologique Européen, SHE – 2: structure of a physically based, distributed modelling system](#). *Journal of Hydrology* **87**, 61–77.
- Bergström, S. 1980 Operational applications of conceptual hydrological models developed for representative and experimental basins. *Proceedings of the Helsinki Symposium, IAHS-AISH* **130**, 421–426.
- Boughton, W. C. 1993 A hydrograph-based model for estimating the water yield of ungauged catchments. In: *Proceedings of the 1993 Hydrology and Water Resources Conference*, Institution of Engineers, Australia, National Conference Publication no. 93/14, pp. 317–324.
- Boughton, W. C. & Carroll, D. G. 1993 A simple combined water balance/flood hydrograph model. In: *Proceedings of the 1993 Hydrology and Water Resources Conference*, Institution of Engineers, Australia, National Conference Publication no. 93/14, pp. 299–304.
- Bray, M. & Han, D. 2004 Identification of support vector machines for runoff modelling. *Journal of Hydroinformatics* **6**, 265–280.
- Bultot, F. & Dupriez, G. L. 1976 [Conceptual hydrological model for an average-sized catchment area, I. concepts and relationships](#). *Journal of Hydrology* **29**, 251–272.
- Burnash, R. J. C., Ferral, R. L. & McGuire, R. A. 1973 *A Generalized Streamflow Simulation System – Conceptual Modelling for Digital Computers*. Report, U. S. Department of Commerce, National Weather Service, Silver Springs, MD, and State of California, Dept of Water Resources, Sacramento, California.
- Chiew, F. H. & McMahon, T. A. 1994 [Application of the daily rainfall-runoff model MODHYDROLOG to 28 Australian catchments](#). *Journal of Hydrology* **153**, 383–416.
- Chow, V. T., Maidment, D. R. & Mays, L. W. 1988 *Applied Hydrology*. McGraw Hill Higher Education International, New York.
- Crawford, N. H. & Linsley, R. S. 1966 *Digital Simulation in Hydrology: The Stanford Watershed Model IV*. Technical Report No. 39, Department of Civil Engineering, Stanford University, Palo Alto, California.
- Eckhardt, K. & Arnold, J. G. 2001 [Automatic calibration of a distributed catchment model](#). *Journal of Hydrology* **251**, 103–109.
- EPA 1998 *Better Assessment Science Integrating Point and Nonpoint Sources: BASINS Version 2.0. User's Manual*. EPA-823-B-98-006, US Environmental Protection Agency, Washington, DC.
- Grizzetti, B., Bouraoui, F., Granlund, K., Rekolainen, S. & Bidoglio, G. 2003 [Modelling diffuse emission and retention of](#)

- nutrients in the Vantaanjoki watershed (Finland) using the SWAT model. *Ecological Modelling* **169**, 25–38.
- Hargreaves, G. H. 1994 Defining and using reference evapotranspiration. *Journal of Irrigation and Drainage Engineering, ASCE* **120**, 1132–1139.
- Hargreaves, G. L., Hargreaves, G. H. & Riley, J. P. 1985 Irrigation water requirements for Senegal River Basin. *Journal of Irrigation and Drainage Engineering, ASCE* **111**, 265–275.
- Huang, J., van den Dool, H. & Georgakakos, K. P. 1996 Analysis of model-calculated soil moisture over the United States (1931–1993) and application to long range temperature forecasts. *Journal of Climate* **9**, 1350–1362.
- Johnson, M. S., Coon, W. F., Mehta, V. K., Steenhuis, T. S., Brooks, E. S. & Boll, J. 2003 Application of two hydrologic models with different runoff mechanisms to a hillslope dominated watershed in the northeastern US: a comparison of HSPF and SMR. *Journal of Hydrology* **284**, 57–76.
- Jones, R., Noguer, M., Hassell, D., Hudson, D., Wilson, S., Jenkins, G. & Mitchell, J. 2004 *Generating High Resolution Climate Change Scenarios using PRECIS, PRECIS Handbook*, Hadley Centre for Climate Prediction and Research, UK.
- Krishna Kumar, K., Patwardhan, S. K., Kulkarni, A., Kamala, K., Koteswara Rao, K. & Jones, R. 2011 Simulated projections for summer monsoon climate over India by a high resolution regional climate model (PRECIS). *Current Science* **101**, 312–326.
- Kumar, D. N. & Maity, R. 2008 Bayesian dynamic modeling for nonstationary hydroclimatic time series forecasting along with uncertainty quantification. *Hydrological Processes* **22** (17), 3488–3499.
- Kumar, D. N., Reddy, M. J. & Maity, R. 2007 Regional rainfall forecasting using large scale climate teleconnections and artificial intelligence techniques. *Journal of Intelligent Systems* **16** (4), 307–322.
- Lawrence, D., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J., Zeng, X., Yang, Z. L., Levis, S., Skaguchi, K., Bonan, G. B. & Slater, A. G. 2011 Parameterization improvements and functional and structural advances in Version 4 of the community land model. *Journal of Advances in Modelling Earth Systems* **3** (M03001), 27.
- Maity, R. & Kashid, S. S. 2010 Short-term basin-scale streamflow forecasting using large-scale coupled atmospheric oceanic circulation and local outgoing longwave radiation. *Journal of Hydrometeorology* **11** (2), 370–387.
- Maity, R. & Kashid, S. S. 2011 Importance analysis of local and global climate inputs for basin-scale streamflow prediction. *Water Resources Research* **47**, W11504.
- Monteith, J. L. 1965 Evaporation and the environment. In: The state and movement of water in living organisms. XIXth Symposium Society for Experimental Biology, Cambridge University Press, Swansea, Wales, pp. 205–234.
- O'Connell, P. E., Nash, J. E. & Farrell, J. P. 1970 River flow forecasting through conceptual models. Part 2. The Brosna catchment at Ferbane. *Journal of Hydrology* **10**, 317–329.
- Piotrowski, A. P. & Napiorkowski, J. J. 2012 Product-Units neural networks for catchment runoff forecasting. *Advances in Water Resources* **49**, 97–113.
- Porter, J. W. 1972 *The Synthesis of Continuous Streamflow from Climatic Data by Modelling With a Digital Computer*. PhD Thesis, Monash University, Australia.
- Porter, J. W. & McMahon, T. A. 1975 Application of a catchment model in south eastern Australia. *Journal of Hydrology* **24**, 121–134.
- Porter, J. W. & McMahon, T. A. 1976 The Monash model: User manual for daily program HYDROLOG. Research Report no. 2/76, Department of Civil Engineering, Monash University, Australia, p. 41.
- Samsudin, R., Saad, P. & Shabri, A. 2011 River flow time series using least squares support vector machines. *Hydrology and Earth System Sciences* **15**, 1835–1852.
- Sugawara, M. 1967 The flood forecasting by a series storage type model. International Symposium on floods and their computation, Leningrad, USSR, pp. 1–6.
- Sugawara, M., Ozaki, E., Watanabe, I. & Katsuyama, Y. 1974 Tank model and its application to Bird Creek, Wollombi Brook, Bikin River, Kitsu River, Sanga River and Nam Mune. Research Note, National Research Center for Disaster Prevention, Kyoto, Japan, 11, pp. 1–64.
- Suykens, J. A. K., Van Gestel, T., De Brabanter, J., De Moor, B. & Vandewalle, J. 2002 *Least Squares Support Vector Machines*. World Scientific, Singapore.
- Willmott, C. J., Robeson, S. M. & Matsuura, K. 2011 Short Communication, A refined index of model performance. *International Journal of Climatology* **32**, 2088–2094.
- Zeng, X., Kiviat, K. L., Sakaguchi, K. & Mahmoud, A. M. A. 2012 A toy model for monthly river flow forecasting. *Journal of Hydrology* **452–453**, 226–231.