

## Modelling hydrological responses under climate change using machine learning algorithms – semi-arid river basin of peninsular India

G. Sireesha Naidu, M. Pratik and S. Rehana\*

Lab for Spatial Informatics, International Institute of Information Technology, Hyderabad 500032, India

\*Corresponding author. E-mail: rehana.s@iiit.ac.in

### Abstract

Catchment scale conceptual hydrological models apply calibration parameters entirely based on observed historical data in the climate change impact assessment. The study used the most advanced machine learning algorithms based on Ensemble Regression and Random Forest models to develop dynamically calibrated factors which can form as a basis for the analysis of hydrological responses under climate change. The Random Forest algorithm was identified as a robust method to model the calibration factors with limited data for training and testing with precipitation, evapotranspiration and uncalibrated runoff based on various performance measures. The developed model was further used to study the runoff response under climate change variability of precipitation and temperatures. A statistical downscaling model based on K-means clustering, Classification and Regression Trees and Support Vector Regression was used to develop the precipitation and temperature projections based on MIROC GCM outputs with the RCP 4.5 scenario. The proposed modelling framework has been demonstrated on a semi-arid river basin of peninsular India, Krishna River Basin (KRB). The basin outlet runoff was predicted to decrease (13.26%) for future scenarios under climate change due to an increase in temperature (0.6 °C), compared to a precipitation increase (13.12%), resulting in an overall reduction in water availability over KRB.

**Key words:** Budyko hypothesis, climate change impacts, general circulation models, hydrological modelling, statistical downscaling

### Highlights

- Catchment parameter estimation dynamically with precipitation, evapotranspiration and runoff using machine learning algorithms.
- Statistical downscaling model to relate climate and hydrological variables using machine learning algorithms.
- Assessment of hydrological responses under climate change of a semi-arid river basin of India.

### INTRODUCTION

Hydrological models are promising tools to study various water resources engineering problems such as flood prediction and design, drought assessment, water quantity and quality assessment and hydrological responses under climate variability (Sood & Smakhtin 2015). The most important classification of hydrological models is empirical (data-driven), conceptual and physically based

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

models (Aghakouchak & Habib 2010). Empirical models have evolved as data-driven using various regression (Zhang *et al.* 2020), machine learning algorithms such as Artificial Neural Network (ANN) (Hu *et al.* 2018), least square-Support Vector Regression (SVR) (Bharti *et al.* 2017), etc. Such empirical models are primarily based on observations without considering the various hydrological processes involved (Jaiswal *et al.* 2020), whereas the physically-based models work on the principles of conservation of mass, momentum and energy to consider various hydrological processes along with field measured parameters and other basin characteristics (Aghakouchak & Habib 2010). The conceptual models are based on the simplified mathematical conceptualization of a hydrological system by considering various components at catchment scales and use model parameters in simulating the discharges (Jaiswal *et al.* 2020). Physically-based models are complex and require fine spatio-temporal data, whereas purely data-driven empirical models are limited completely to observations. Compared to physical and empirical based hydrological models, the conceptual models have gained much interest in the research community due to their capability to include various hydrological processes (e.g. runoff, evapotranspiration, surface storage, etc.) and ease of implementation with available data (Ammann *et al.* 2020). In this context, various conceptual models have evolved in the past, such as HYMOD (Parra *et al.* 2018), HBV model (Vormoor *et al.* 2018), etc. Model parameters or calibration factors are most prominent in conceptual hydrological models affecting the accuracy of streamflow estimates (Arnold *et al.* 2012). In this context, auto-calibration approaches for hydrological models have progressed in recent years, for example, the Soil Water Assessment Tool – Calibration Uncertainty Programs (SWAT-CUP) (Ha *et al.* 2017), and the Shuffled Complex Evolution Algorithm (SCE-UA) which carries the limitation of multiple optimal parameters and corresponding accuracies (Pohlert *et al.* 2007). Calibration of hydrological model parameters using remotely sensed variables has also gained interest in recent years, based on satellite evapotranspiration data (Immerzeel & Droogers 2008; Rientjes *et al.* 2013). To this end, various catchment processes have been conceptualized in hydrological modelling by including model parameters whose values can be determined through calibration (Mianabadi *et al.* 2019). In this context, calibration factors were implemented for various processes such as closure of water balance (Jarsjö *et al.* 2008), radiation (Choudhury 1999) and vegetation dynamics (Donohue *et al.* 2007). Conceptual models developed by including such calibration factors on various processes have been used to study the hydrological responses under climate, land use and water changes as a function of various hydroclimatic variables and observed historical streamflows (e.g. Asokan *et al.* 2010).

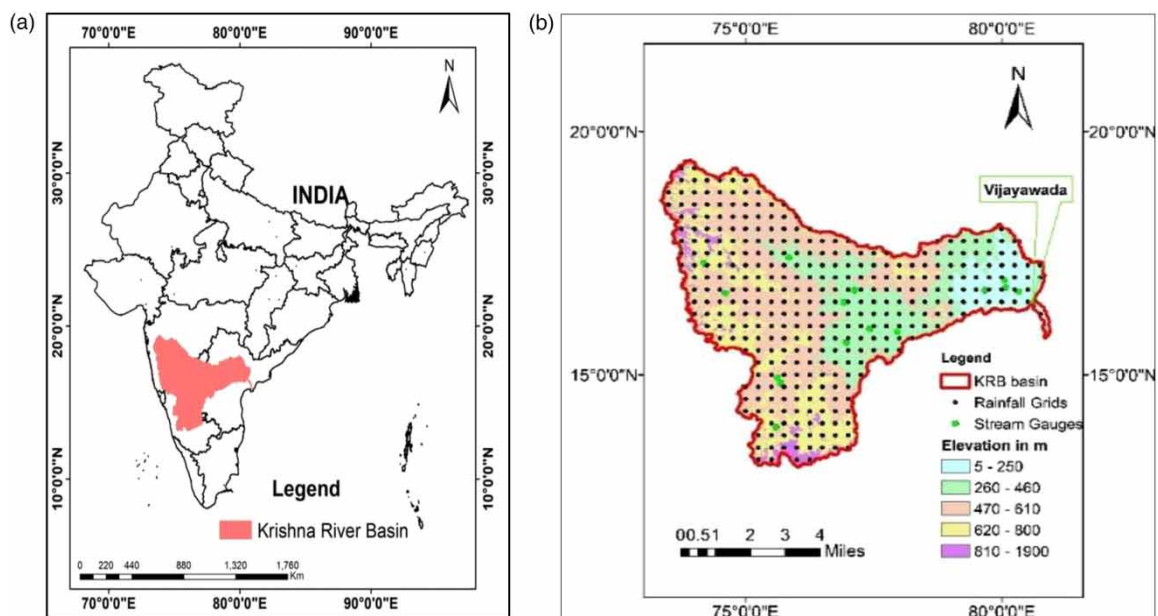
Hydrological variables such as precipitation, evapotranspiration and streamflows have shown pronounced changes under climate signals and therefore the model parameters also should account for such changes (Srinivasa Raju & Nagesh Kumar 2018). Most of the conventional hydrological models apply such calibration parameters which were established based on observed historical streamflow data and other hydrological variables in the hydrological climate change impact assessment (e.g. Meenu *et al.* 2013). Such models integrate hydrological models and statistical downscaling models to study the hydrological responses based on the historically established model parameters (Hengade *et al.* 2018; Saraf & Regulwar 2018). Inclusion of time-invariant model parameters which are independent of the hydro-climatic variables and estimated solely based on the observed data are limited to capture the temporal variability of water-energy balance variables under climate change (Rehana *et al.* 2020a). The present study proposes to establish a relationship between the calibration factors and various hydroclimatic variables and uncalibrated runoff. Such an established functional form of calibration factors can be easily implementable for future scenarios under projected hydroclimatic variables. In this context, developing a calibration approach relating precipitation, evapotranspiration and uncalibrated streamflows can improve the model performance and understanding of hydrological processes (López *et al.* 2017). Therefore, estimating model parameters dynamically accounting for the expected hydrological variability under climate change can be promising. Furthermore, such approaches can be further extended with climate change projections to study the hydrological response of river basins.

The study proposed a modelling framework by relating the conceptual hydrological model parameter with various hydrological variables of precipitation, evapotranspiration and simulated runoff. Such established relationships will ease the estimation of model calibration factors for future scenarios with climate change projections obtained from statistical downscaling models. The major challenges in implementing such a framework are limited data and poor understanding of complex relationships between the calibration factors and various hydrological process-based variables. To overcome such limitations, the present study proposed an approach to model the parameters of conceptual models using various hydrological variables by adopting machine learning algorithms. The study emphasized the use of advanced machine learning algorithms for developing such complex relationships between model parameters and hydrological variables, which can form as a basis for studying the hydrological responses under climate variability. The proposed modelling framework has been demonstrated on a semi-arid river basin, Krishna River Basin (KRB), India.

## MATERIALS AND METHODS

### Basin description

Krishna River Basin (KRB) is the fifth largest river system in India and second largest river basin of peninsular India with a total catchment area of 258,948 km<sup>2</sup> (Figure 1). The KRB, 73°17'–81°9'E and 13°10'–19°22'N, has heavy precipitation over the Western Ghats with decreasing precipitation towards the upper and lower parts of the basin. The KRB originates in the Western Ghats of India and travels for about 1,400 km across the four states of Karnataka, Maharashtra, Andhra Pradesh and Telangana before joining the Bay of Bengal at Hamsaladevi in Andhra Pradesh near the east coast of India. The major tributaries of the KRB include Ghataprabha, Malaprabha, Bhima, Tungabhadra and Musi. About 44% of the KRB lies in Karnataka, with about 26, 15 and 15% of the basin in Maharashtra, Telangana and Andhra Pradesh respectively. The annual average precipitation over the basin is about 784 mm, with 90% of the precipitation occurring during the south-west monsoon from June to October (<http://india-wris.nrsc.gov.in/wrpinfo/?title = Krishna>). Most of the basin

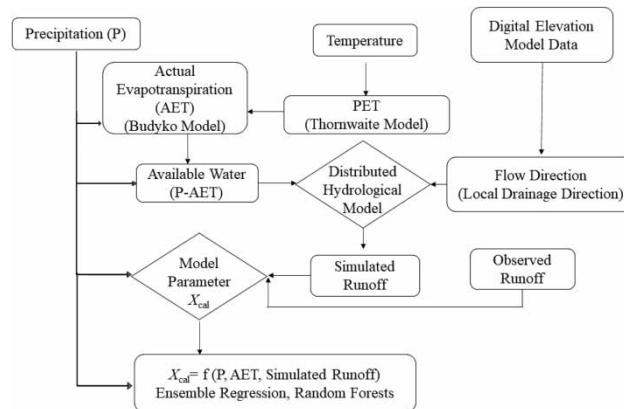


**Figure 1** | Map of the Krishna River Basin (KRB). (a) Location of the catchment in India; (b) various hydro-meteorological stations and elevation map superimposed on the basin.

has a prevailing semi-arid climate and the states covering the basin (Maharashtra, Karnataka and Telangana) are majorly drought-prone (<http://india-wris.nrsc.gov.in>). The major water use of the KRB is for irrigation with about 61.9 billion cubic meters per year (BCM/year) along with domestic and industrial water uses as 1.6 and 3.2 BCM/year (Rooijen *et al.* 2009).

### Distributed hydrological model: PCRaster

In this study, hydrological modelling of the KRB was performed using the distributed hydrological model, PCRaster-water flow module using POLFLOW code (Wit 2001; Jarsjö *et al.* 2008) ([http://pcraster.geo.uu.nl/pcraster/4.2.0/documentation/pcraster\\_manual/sphinx](http://pcraster.geo.uu.nl/pcraster/4.2.0/documentation/pcraster_manual/sphinx)). The KRB was delineated and a local drainage direction (ldd) map was created using the topographical data (Figure 2). The basin was discretized with a rectangular grid (1 × 1 km) and for each grid various hydrological processes can be provided as input. The Digital Elevation Map (DEM) data used was at a resolution of 30 arc-seconds (approximately 1 km) from the Global 30 Arc-Second Elevation (GTOPO30) data set provided by the US Geological Survey. Spatially distributed data of daily precipitation (P) and temperature (T) for the KRB were from the India Meteorological Department (IMD). The gridded daily precipitation data at a spatial resolution of 0.25 × 0.25° (Pai *et al.* 2014) and temperature at a resolution of 1 × 1° (Srivastava *et al.* 2009) were cropped to the basin for a common time period from 1951 to 2014. The temperature, which was available at 1 × 1° resolution, was interpolated to 0.25 × 0.25° using the inverse distance weighting method. The discharge data was obtained from Krishna & Godavari Basin Organization (KGBO), Central Water Commission (CWC), Hyderabad, Government of India ([www.kgbo-cwc.ap.nic.in](http://www.kgbo-cwc.ap.nic.in)) for about 25 discharge locations for the period 1966–2015 (Figure 1(b)).



**Figure 2** | Overview of the modelling approach to estimate catchment-scale calibration factors using machine learning algorithms.

The basic hydrological processes involved in the partitioning of precipitation (P) into evapotranspiration (AET), groundwater (GW), runoff (Q) and change of storage ( $\Delta S$ ) at catchment scale can be written as follows:

$$P + Q_{in} + GW_{in} - GW_{out} - Q_{out} - AET - \Delta S = 0 \quad (1)$$

For sufficiently long time-scales, the groundwater inflow ( $GW_{in}$ ) and outflow ( $GW_{out}$ ) volumes along with the change in storage volumes  $\Delta S$  can be neglected. Such basic assumptions ultimately represent a water balance equation at river basin scale with P, AET and R as the major hydrological variables which can be related to estimating the precipitation surplus or Available Water (AW) and

can be expressed as follows:

$$AW = P - AET \quad (2)$$

It can be noted that the study employed a conceptual catchment scale hydrological model on an annual time scale, neglecting the storage component of water in terms of groundwater in the water-balance equation. For long-term soil water storage changes the groundwater recharge influenced by geological features and human interaction water withdrawals are considered to be negligible (Wu *et al.* 2017). However, it can be noted that the storage component of water cannot be neglected at any temporal scale less than annual and the resulting phenomenon and the hydrological responses may vary due to the inclusion of groundwater storages being taken into account (Gunkel & Lange 2017).

If the basic hydrologic unit of a distributed hydrological model is considered as grid cell,  $i$ , the annual total precipitation ( $P_i$ ) and AET ( $AET_i$ ), the precipitation surplus ( $AW_{cal,i}$ ) can be estimated as follows:

$$AW_{cal,i} = P_i - AET_i \quad (3)$$

where  $P_i$  and  $AET_{clim,i}$  are at annual scale in mm/year. Due to the lack of direct and long-term AET data, most of the studies depend on either pan evaporation or empirical estimations of PET (Han *et al.* 2014). One of the widely applied methods to estimate AET flux in hydrology is estimating PET from the energy available for vaporization first, and then applying a limiting factor to account for the water availability (Anabalón & Sharma 2017). Such ET models based on P and PET works with readily available and modelled operational meteorological variables such as precipitation and temperature (Budyko 1974). In this context, Budyko's (1974) hypothesis has gained much interest in hydrology, which can relate to the most important hydrological variables of P, PET, AET along with model parameter ' $\omega$ ' to represent climate variability, basin characteristics, vegetation and terrain, etc. Several studies used parametric formulations of the Budyko framework to estimate AET globally (e.g. Bai *et al.* 2020) and Indian case studies (e.g. Singh & Kumar 2015; Goyal & Khan 2017). A bottom-up probabilistic Budyko framework has been proposed by Singh & Kumar (2015) to estimate the vulnerability of available water of India under climate change. In another study by Sinha *et al.* (2018), the Budyko framework was used to study the anthropogenic stress and climatic variance under warming shifts across 55 catchments in peninsular India.

Such a basic parametric formulation of Budyko framework can be written as follows:

$$\frac{AET}{P} = 1 + \frac{PET}{P} - \left(1 + \left(\frac{PET}{P}\right)^\omega\right)^{(1/\omega)} \quad (4)$$

Several studies have modified the above formulation and the most widely applied one is based on one by Zhang *et al.* (2004) as a calibration-free formulation for the Budyko equation relating to P, PET and AET as follows:

$$AET_{climate} = \left[ P \left(1 - \exp\left(\frac{-PET}{P}\right)\right) PET \tanh\left(\frac{P}{PET}\right) \right]^{0.5} \quad (5)$$

where PET can be estimated based on empirical formulations such as Penman (1950), Thornthwaite (1948) and Hargreaves & Allen (2003), etc. The present study considered the Thornthwaite model as one of the simplest models to estimate PET, which includes average air temperature and geographical location of the region of interest as input variables. The most dependable gridded observed meteorological data available for India is temperature among the other variables (e.g. wind speed, humidity,

radiation, etc.) required to implement detailed PET estimates such as the Penman model. Therefore, the present study adopted the Thornthwaite model as the temperature-based PET model for the Indian context. The AET can be estimated on a monthly time scale with PET estimates. The AET estimated using Equation (5) is region-specific and ideally developed for given climatic and basin characteristics. Therefore, model parameters have to be introduced in Equation (5) to obtain runoff which is in agreement with that observed. As the AET is the most complicated component of the hydrological cycle, defined by various factors of climatic, vegetative, soil moisture and amount of available water, implementation of calibration factors on AET will be more promising. We have introduced the calibration factor on AET as the basic conceptual hydrological model considered was based on the closure of water balance involving the prominent hydrological variables of precipitation (P), AET and runoff (R),  $R = P - AET$ . A calibration factor ( $X_{Cal}$ ) has been introduced on AET in the calibration of the observed runoff ( $R_{obs}$ ) using the water balance equation,  $R_{obs} = (P - X_{Cal}AET)$ . This way, a model parameter has been introduced on AET by accounting for the closure of water balance at the catchment outlet. Therefore, the AW with inclusion of calibration factor ( $X_{Cal}$ ) can be written as follows:

$$AW_{actual,i} = P_i - X_{Cal}AET_{clim,i} \quad (6)$$

The discharge at the basin outlet ( $R_{Cal,outlet}$ ) can be estimated by accumulating the flow at grid cell,  $i$ , ( $RAW_{cal,i}$ ) and from all upstream grid cells ( $RAW_{cal,\Omega}$ ) following the flow direction of the river and corresponding to the area of each grid cell ( $A_{Cell}$ ) as follows:

$$R_{Cal,outlet} = (AW_{cal,i} + AW_{cal,\Omega}) * A_{Cell} \quad (7)$$

where  $R_{Cal,outlet}$  is the uncalibrated total runoff from the basin outlet at steady state and it is generally not consistent with observed runoff,  $R_{Obs,outlet}$ :

$$R_{Obs,outlet} = (AW_{actual,i} + AW_{actual,\Omega}) * A_{Cell} \quad (8)$$

A comparison of observed and uncalibrated runoff estimations can be made to estimate the model parameter ( $X_{Cal}$ ) based on Jarsjö *et al.* (2008) as follows:

$$\frac{R_{Cal,outlet}}{R_{Obs,outlet}} = \frac{P - AET_{clim}}{P - (X_{Cal} * AET_{clim})} \quad (9)$$

$$X_{Cal} = \frac{R_{Obs,outlet}}{R_{Cal,outlet}} + \left(1 - \frac{R_{Obs,outlet}}{R_{Cal,outlet}}\right) \frac{\sum P}{\sum AET_{clim}} \quad (10)$$

where  $R_{Obs,outlet}$  and  $R_{Cal,outlet}$  are the long-term annual average observed and simulated runoff at the basin outlet in m<sup>3</sup>/s respectively and  $\sum P$  and  $\sum AET_{clim}$  are the long-term accumulated annual average observed precipitation and AET (Equation (5)) over the basin in mm/year. Such a basin averaged single calibration factor has been widely applied to study hydrological responses of soil, land cover and water use by Asokan *et al.* (2010) and Jarsjö *et al.* (2008). However, to study the hydrological responses under climate change, use of single time-invariant calibration factors may be limited to capture the variability of precipitation and AET. A dynamic model parameter which can capture the observed variability of precipitation and ET has been proposed, which can be further used to study the hydrological responses under climate change. Such a dynamic calibration factor can vary as a function of hydrological variables (P, AET and uncalibrated runoff) for each time step. To model the complex relation between the model parameters and hydrological variables of P, AET, and R,

the present study proposed two machine learning algorithms, Ensemble Regression (ER) and Random Forest (RF).

### Machine learning algorithms for modelling the calibration factors

The annual scale calibration factors ( $X_{Cal}$ ) were estimated and considered as predictand variable and uncalibrated runoff (P-AET) resulting from the PCRaster model, along with P and AET which were considered as predictor variables in training and testing of machine learning models. The present study used two machine learning models, Ensemble Regression (ER) and Random Forest (RF) to predict the model parameters.

#### Ensemble regression (ER) model

Ensemble Regression (ER) methods are machine learning paradigms in which multiple methods which are often referred to as weak learners are trained to solve a problem and are combined to obtain better results (Friedman 2001). The ER model works on the hypothesis that a diverse set of models can make better predictions in comparison with an individual model. The ER models have gained interest in hydrological model assessments in recent years (Sajedi-Hosseini *et al.* 2018).

The input training dataset of 'N' points  $\{X, Y\} = \{(x_i, y_i)\}_{i=1}^N$ , where  $x_i$  is the set of predictors and  $y_i$  as the observed predictand value at the  $i$ th timestep, will be considered. Initially, all predictors are given equal weighting coefficients (equal importance –  $\alpha_i = \frac{1}{N}$ ) and an initial model ( $F_0(x)$ ) will be developed to predict values of the form  $y = F_0(x)$ .

At every iteration,  $m$ , the residuals ( $r_{im}$ ) will be calculated between the observed ( $y_i$ ) and modelled predictand value ( $F_{m-1}(x_i)$ ):

$$r_{im} = - \left[ \frac{d \left( \sum_{i=1}^N [y_i - F_{m-1}(x_i)]^2 \right)}{d(F_{m-1}(x_i))} \right] \text{ for } i = 1..N \quad (11)$$

A base-learner ( $h_m$ ) will be fitted to these residuals using a loss function 'L' in the direction of the steepest gradient, i.e. weight,  $\alpha_i$  of point 'i' is increased corresponding to a higher value of residual using the training set  $\{(x_i, r_{im})\}_{i=1}^N$ . The model is then sequentially updated as follows:

$$F_m(x) = F_{m-1}(x) + \underset{h}{\operatorname{argmin}} \sum_{i=1}^n L(y_i, F_{m-1}(x) + h_m(x_i, r_{im})) \quad (12)$$

where the  $\underset{h}{\operatorname{argmin}}$  refers to minimization of base-learner error ( $h_m$ ). The least-squares loss function  $L(y, F(x)) = \frac{1}{2} (y - F(x))^2$  is used to update the residuals.

#### Random forest (RF) model

In recent years, data mining algorithms such as Random Forest (RF) have been widely applied in various studies related to water resources engineering such as hydrological (Moore *et al.* 1991), soil moisture forecasting (Prasad *et al.* 2018) and water quality assessment (Bui *et al.* 2020). RFs are an ensemble of decision trees, which split the data space similar to the way humans and each node represents a predictor variable (Breiman 2001). The data splitting is generally done based on inequality condition, Gini index, which is a measure of the performance of the split and measures how diverse the data is until it reaches a terminal node (Breiman 2017). The RF tree size is determined based on

the number of nodes, which serve as a basis to minimize the variance of each split. RFs can perform regression as well as classification using various decision trees based on bootstrap aggregation (bagging), which is a random data re-sampling methodology with replacement. The bootstrapping process was used to sample  $S$  data points with  $N$  variables by replacement. A decision tree was trained on each of these random samples. Bagging is the process of making overall predictions from average predictions of various bootstrapped samples trained on individual decision trees, making averaging predictions from each decision tree. The steps involved in designing the RF models are as follows:

1. Draw a bootstrap samples  $S^*$  of size  $N$  from the training data
2. An RF tree  $T_b$  will be developed to the bootstrapped data recursively for each terminal node of the tree.
  - Selecting  $S$  variables at random from  $N$  variables
  - Choosing the best split-point among the  $S$
  - Split the node into two daughter nodes
  - Continuing the above steps until the size of the minimum node  $n_{min}$  is reached.
3. The ensemble of trees can be estimated as  $\{T_b\}_1^B$
4. Prediction at a new point,  $x$ , can be made using the regression formulation as

$$\hat{f}_{rf}^B(x) = \frac{1}{B} \sum_{b=1}^B T_b(x) \quad (13)$$

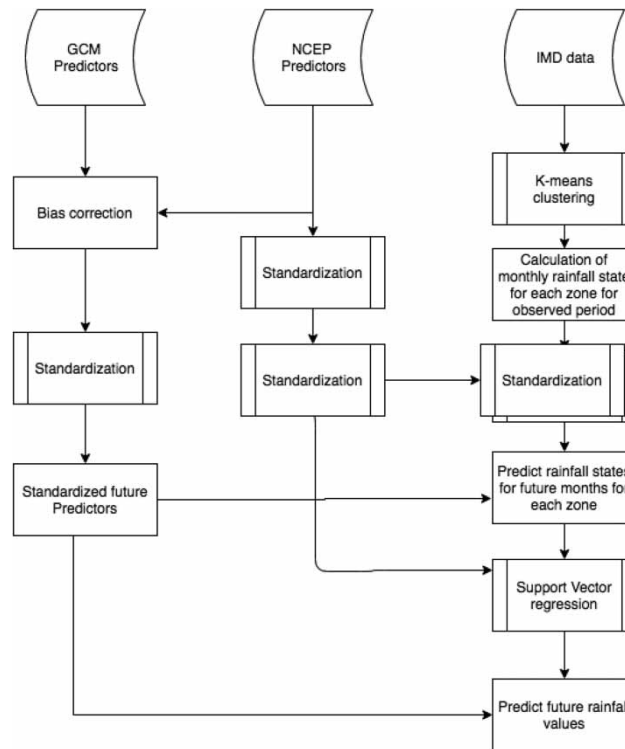
The RF model has been identified as a robust model due to its ability to capture the complex relationships between features and labels efficiently, accounting for the uncorrelated decision trees in the prediction.

The trained and tested machine learning algorithms will be further used to predict the calibration factors which can be used to estimate the calibrated runoff at the catchment scale. To quantify the performance of models and to understand the dependence between calibration factors and hydrological variables, Root Mean Square Error (RMSE), Absolute Percentage Bias (APB), Mean Absolute Error (MAE), Nash–Sutcliffe Efficiency (NSE) (Nash & Sutcliffe 1970), Kling-Gupta Efficiency (KGE) (Knoben *et al.* 2019) and Likelihood were selected as performance measures.

### Climate change projection model: statistical downscaling models (SDM)

Statistical downscaling models are the state-of-the-art climate change projections prediction models relating to large-scale climate variables with surface hydrological variables (e.g. precipitation) using statistical methods (Wilby & Dawson 2013). The present study employed a multisite statistical downscaling model to predict the climate change projections of precipitation and temperature (Figure 3). The basic formulation of the downscaling model includes data preprocessing to remove systematic bias in the modelled and actual climate observations (bias correction), data reduction method (principal component analysis, PCA), predictand variable states estimation (K-means clustering), fitting algorithm to relate predictors and predictand states (CART), and transfer function (Support Vector Regression). National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) climate data from 1948 to the present at a resolution of  $2.5 \times 2.5^\circ$  were considered as large-scale observed climate variables. The MIROC-ESM derived from the Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology, Japan, was considered as the GCM output. Spatial resolution mismatch between NCEP and GCM data was resolved by applying an inverse distance weighting interpolation. The GCM predictor data undergoes bias correction based on quantile mapping method by comparing the cumulative distribution functions (CDFs) of NCEP and GCM predictor data for the





**Figure 3** | Overview of the statistical downscaling model to predict projections of precipitation and temperatures over KRB.

historical and future scenarios to remove the systematic bias associated with the climate outputs (Li *et al.* 2010). Both bias-corrected GCM and reanalysis climate data undergoes standardization, which involves subtraction of mean and division by standard deviation estimated with all the data points of the time series. After data processing, statistically significant climate variables (predictors) were considered to predict the precipitation over various grid points of the basin. The predictor variables which have shown significant correlation coefficients with precipitation were identified as surface air temperature, wind speed, humidity, etc., which were also found as potential predictors in other studies over India (Rehana & Mujumdar 2012; Salvi *et al.* 2013). NCEP/NCAR reanalysis data for surface air temperature, mean sea level pressure, specific humidity at 500 mb pressure level, zonal and meridional wind velocity at the surface level is extracted for the latitudes in the range of 12.5–20°N and the longitudes in the range of 72.5–82.5°E, surrounding the entire Krishna Basin.

### K-mean clustering

K-means clustering is an unsupervised machine learning algorithm that partitions  $n$  observations into  $K$  clusters for which each observation belongs to the cluster with the nearest mean. In this study, we use the K-means algorithm to achieve cross-correlation among the rain stations and group the months having similar rainfall. This technique reads the observed rainfall values for all grids in the basin in a month, clusters them, and provides a single representative value that is referred to as the state of rainfall for that particular month. Consider  $X = x_1, x_2, x_3, \dots, x_n$  as the set of data points and  $V = \mu_1, \mu_2, \mu_3$  as the set of cluster means. For each data point, the nearest mean was identified and the point is assigned to the corresponding cluster. The objective of k-means algorithm is to iteratively minimize the squared error function given as follows:

$$J(V) = \sum_{i=1}^c \sum_{j=1}^{c_i} (\|x_i - v_j\|)^2 \quad (14)$$

where  $\|x_i - v_j\|$  represents the Euclidian distance between  $x_i$  and  $v_j$ ,  $C$  is the number of clusters and  $c_j$  is the number of points in the cluster.

### Classification and Regression Trees (CART)

Classification and Regression Trees is a decision tree learning technique used in prediction modelling. A decision tree is a machine learning model in which categorization of the data is carried out with each feature partitioning the data based on conditions or range of the features and works when the data has a finite set of values. CART helps in categorizing the rainfall into states by building a statistical relation between the continuous principal components extracted from predictor data and the rainfall states estimated using K-means clustering. The established relationship is assumed to be intact for the future predictors which are then taken as input for CART model and for which the future rainfall states are estimated. An advantage of CART over linear classification models is that they can capture non-parametric and nonlinear relationships as well as yield simple models. Cross-validation is carried out in order to ensure there is no risk of overfitting the data. For every iteration, calculate the Gini's impurity/diversity Index (E) (Loh 2011) of the data using the formula given in the equation as follows:

$$E = \sum_{i \neq j} P(w_i)P(w_j) = 1 - \sum_{i \neq j} P(w_i)^2 \quad (15)$$

where  $P(w_j)$  represents the probability of data being in  $j$ th class for each value of the attribute. An attribute splitting will be selected which has minimum impurity index (E). It can be considered as 0 when all the patterns at node have the same class label. Continue until impurity (E) is less than a certain threshold ( $\eta$ ) or on reaching the maximum number of iterations where  $E < \eta$ .

### Support vector regression (SVR)

Individual regression models are built on separating the predictor and observed data based on the weather state category into individual datasets. The SVR model tries to fit the error within a certain threshold ( $\epsilon$ ), identifying a single separating hyperplane which maximizes the margin rather than solely minimizing the error, which helps to find the best model. Linear SVR was used as the regression technique to predict the precipitation and temperature without overfitting. The training data of  $\{(x_1, y_1), \dots, (x_n, y_n)\}$  with  $n$  patterns and function  $f(x)$  will be identified with the consideration of the deviation from the observed target variables  $y_i$  (Lima *et al.* 2012). The input variables,  $X$ , will be mapped into a higher dimensional feature space using a nonlinear mapping function  $\varphi$ :

$$f(x; w) = \langle W, \varphi(x) \rangle + b \quad (16)$$

where  $\langle, \rangle$  denotes the inner product, and  $W$  and  $b$  are the regression coefficients, which can be estimated by minimizing the error between  $f(x)$  and the observed values of  $y$ . SVR uses the  $\epsilon$ -insensitive error to measure the error between  $f(x)$  and the observed values of  $y$ , where  $\epsilon$  is the hyper-parameter:

$$|f(x; w) - y|_{\epsilon} = \begin{cases} 0, & \text{if } |f(x; w) - y| < \epsilon \\ |f(x; w) - y| - \epsilon, & \text{otherwise,} \end{cases} \quad (17)$$

Using the training data of  $(x_i, y_i)$  the values of  $w$  and  $b$  are estimated by minimizing the objective function:

$$F = \frac{C}{N} \sum_{i=1}^n |f(x_i, w) - y_i|_{\in} + \frac{1}{2} \|w\|^2 \quad (18)$$

where  $C$  and  $\in$  are the hyper-parameters. The minimization of the objective function,  $F$ , uses the Lagrange multiplier method, and the final regression equation with kernel function  $K(X, X')$  can be in the form:

$$f(X) = \sum_i K(X, X_i) + b \quad (19)$$

## RESULTS AND DISCUSSION

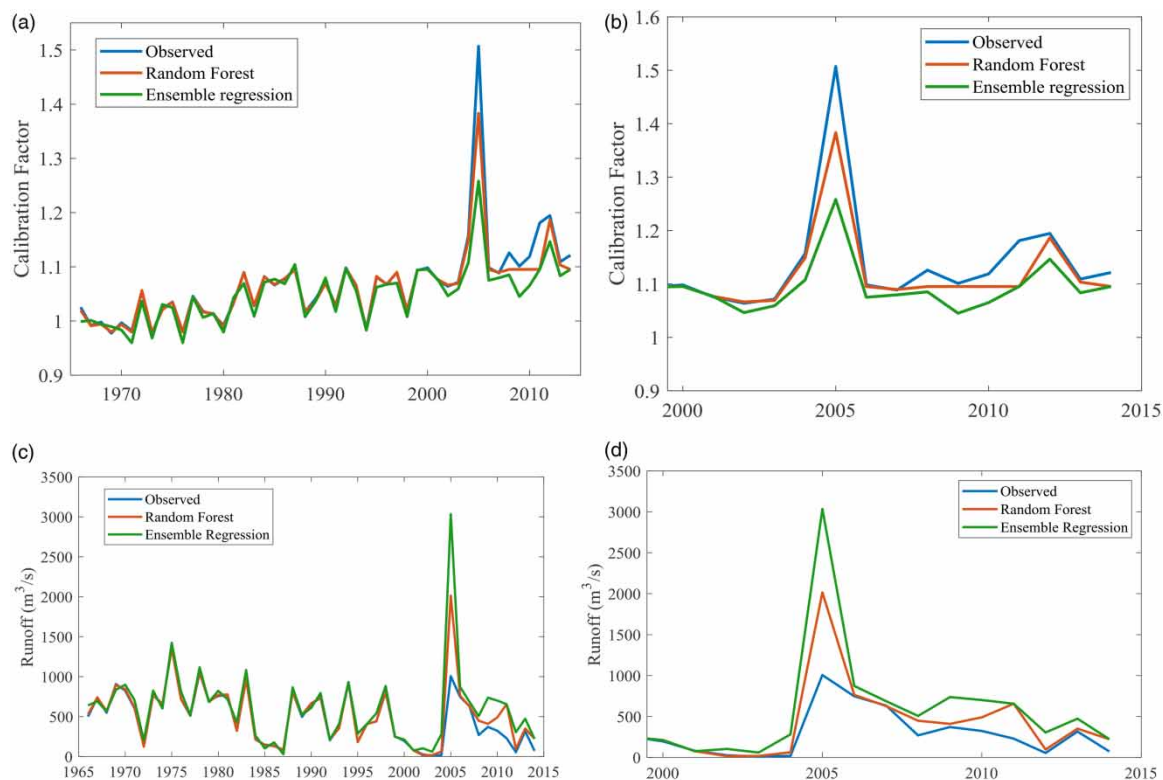
The study applied a PCRaster distributed hydrological model for 348 grids encompassing the KRB at  $0.25 \times 0.25^\circ$  resolution at the annual scale. The original water balance equation of (P-AET) was estimated for each grid from 1965 to 2014 to estimate the precipitation surplus. The precipitation surplus will be converted to runoff by using the area of each grid cell (Equation (7)). The basin outlet uncalibrated runoff will be estimated using the accuthresholdflux function of PCRaster using a local drainage network created based on DEM data. The observed and uncalibrated runoff along with precipitation and AET will be used to estimate the basin averaged model parameter,  $X_{\text{cal}}$  on an annual scale. The model parameters estimated based on the closure of water balance by comparing the observed and uncalibrated runoff will be considered as the predictand. The hydrological variables of P, AET and uncalibrated runoff ( $R_{\text{cal}}$ ) are considered as the predictor variables in modelling of calibration factors using machine learning algorithms.

The calibration factors obtained based on the closure of water balance (Equation (10)) were considered as observed ones and we have used P, AET (Equation (5)) and uncalibrated runoff resulting from PCRaster (Equation (8)) as independent variables in the modelling of dynamic calibration factors using ER and RF models. The ER and RF models were trained with observed calibration factors obtained from Equation (10) as a dependent variable and P, AET and uncalibrated runoff obtained from PCRaster as independent variables for the 1966–2000 data. The trained ER and RF models have been tested with the independent data of P, AET and uncalibrated runoff for the period 2000–2014. The trained and tested ER and RF models were compared in terms of various performance measures, as given in Table 1. The error between the modelled and predicted values of calibration factors were found to be less for the RF model compared to ER. The RMSE and MAE values for the RF (ER) model was 0.04 (0.02) and 0.14 (0.09), respectively. Furthermore, the efficiency

**Table 1** | Comparison of performance measures for RF and ER models in predicting the model parameters

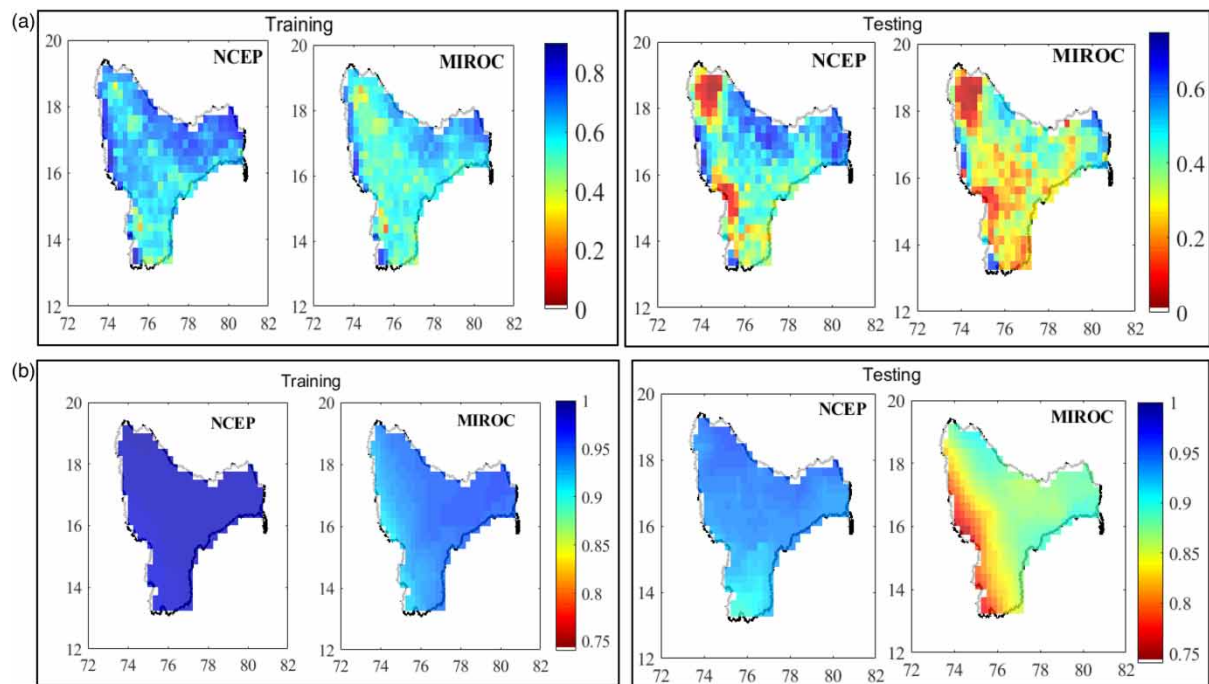
Performance metric	Random Forest	Ensemble Regression
RMSE	0.041	0.142
APB	1.96	8.09
MAE	0.02	0.09
NSE	0.84	0.41
KGE	0.85	0.17
Likelihood	0.46	0.0001

performance measures such as NSE and KGE have also shown that the RF model is robust to predict the calibration factors compared to the ER model with limited data points in training (34) and testing (15). Also, the RF model has been identified as a better fit compared to ER in terms of similarity measure (likelihood) and bias (APB), as given in Table 1 and Figure 4.



**Figure 4** | Comparison of observed, Ensemble Regression and Random Forest modelled calibration factors obtained for: (a) training and (b) testing periods, simulated runoff for (c) training and (d) testing periods.

It can be noted that the calibration factors in Figure 4(a) and 4(b) are in the range of 0.9–1.5, which represents the multiplying factor to be applied on the AET. These are the range of values which have to be applied on the estimated AET based on Equation (5) to implement Equation (6) in the estimation of actual water availability in terms of runoff. The resulting runoff obtained after application of the calibration factor on AET will represent the calibrated runoff which is comparable with the observed runoff on the catchment scale. Figure 4(c) and 4(d) show the observed and simulated runoff values with ER and RF models for the training and testing periods. Due to the robustness of the RF model in modelling the calibration factors, the present study used the RF model to predict the calibration factors for the current and future scenarios. The trained and tested RF model has been used with projections of precipitation and temperature to estimate the projected runoff response over the basin. Figure 5 shows the comparison of NSE coefficients estimated between observed and NCEP and GCM simulations for the training period of 1951–1989 and the testing period of 1990–2005 of the statistical downscaling model for precipitation and temperature. The basin averaged NSE values estimated between observed and NCEP data for precipitation for the training and testing period were 0.65 and 0.44 respectively, whereas the NSE values estimated between observed and MIROC simulations of precipitation for the training and testing period were 0.57 and 0.32 respectively. Temperature predictions were more convincing compared to precipitation due to less variability in data. The basin averaged NSE values estimated between observed and NCEP data for



**Figure 5** | Comparison of Nash–Sutcliffe coefficients estimated between observed and modelled with NCEP and GCM simulations for (a) precipitation and (b) temperature for the training and validation period.

temperature for the training and testing period were estimated as 0.98 and 0.93 respectively, whereas the NSE values estimated between observed and MIROC simulations of temperature for the training and testing period were 0.94 and 0.86 respectively. The observed basin averaged annual precipitation for the training and validation was estimated as 763.06 and 812.04 mm respectively, whereas the NCEP (MIROC) simulations for training and testing period were obtained as 663.68 (650.57) and 611.83 (648.26) mm, respectively. The observed mean annual temperature for the training (testing) period was noted as 26.26 (26.54) °C, with NCEP simulated as 26.25 (26.42) °C and MIROC simulated as 26.27 (26.20) °C. The future projections of precipitation and temperatures for the period of 2021–2080 were estimated with MIROC GCM under the Representative Concentration Pathway (RCP) 4.5 representing atmospheric radiation at  $4.5 \text{ Wm}^{-2}$  at the end of 2100. Both precipitation and temperatures were predicted to increase over the KRB for the period of 2021–2080. The annual average precipitation increase was predicted to be about 13.12% with a temperature increase of 0.6 °C for the period 2061–2080 compared to the observed period of 1951–1989 with MIROC GCM outputs. Overall, the precipitation and temperature are predicted to increase over the KRB, affecting the water-energy balance variables (Rehana *et al.* 2020a). Such increase of precipitation and temperature projections based on GCM outputs was found to be comparable with the earlier research findings over the KRB based on Regional Circulation Model (RCM) based Coordinated Regional Downscaling Experiment (CORDEX) projections (e.g. Rehana *et al.* 2020a, 2020b). Their study noted an approximate increase of 2.19% in precipitation and an approximate increase of 1.29 °C in temperature for the period 2021–2040 and 2041–2060 respectively compared to the observed period of 1966–2003 with the RCP 4.5 scenario with various CORDEX model outputs.

The projections of precipitation and temperature have been used as input to the hydrological modelling framework as shown in Figure 2 to estimate the runoff for the current and future scenarios. The annual AET for observed period of 1951–1989 and 1990–2005 was estimated as 340.22 and 495.96 mm respectively. The estimated AET with MIROC GCM projections of precipitation and temperature for the periods 2020–2040, 2041–2060 and 2061–2080 are 580.63, 613.75 and 648.49 mm respectively (Table 2). Overall, the AET has been predicted to increase under an increase

**Table 2** | Spatial average annual precipitation, temperature, PET, AET and runoff for current period (1951–1989 (training), 1990–2005 (testing)) and future period (2021–2040, 2041–2060, 2061–2080) for KRB

Hydrological variable	Observed/simulated	Current		Future		
		Training	Validation	2021–2040	2041–2060	2061–2080
Average annual precipitation (mm)	Observed	763.06	812.04	–	–	–
	NCEP	663.68	611.83	–	–	–
	MIROC	650.57	648.26	646.13	695.67	735.91
Annual temperature (°C)	Observed	26.26	26.54	–	–	–
	NCEP	26.25	26.42	–	–	–
	MIROC	26.27	26.20	26.95	26.79	26.85
Annual PET (mm)	Observed	1,772.31	1,831.13	–	–	–
	MIROC	1,717.49	1,697.60	1,909.87	1,859.31	1,868.61
Annual AET (mm)	Observed	340.22	495.96	–	–	–
	MIROC	575.09	592.16	580.63	613.75	648.49
Annual runoff at basin outlet (m <sup>3</sup> /s)	Observed	586.56	362.63	–	–	–
	MIROC	687.60	575.68	498.83	500.39	527.79

of temperature and the consequent increase of PET over the basin. The decrease of basin outlet runoff for the periods 2021–2040, 2041–2060 and 2061–2080 compared to the observed period of 1951–1989 was noted as 87.73 (15%), 86.17 (14.69%), 58.77 (10.09%) m<sup>3</sup>/s respectively over the KRB (Table 2). The increase of AET over the basin was found to be more pronounced compared to the increase of precipitation over the KRB using GCM climate change projections.

## CONCLUSIONS

The present study proposed a modelling framework to study the hydrological response under climate change over the KRB. A conceptual hydrological model was adopted to study the runoff response under variability of P and AET. A time-variant calibration approach was developed to estimate the model parameters relating to precipitation, evapotranspiration and uncalibrated runoff using two machine learning algorithms, Ensemble Regression and Random Forest model. The Random Forest algorithm was identified as a robust method to model the calibration factors with limited data available for training and testing with precipitation, evapotranspiration and uncalibrated runoff. The developed model was further used to study the runoff response under climate change variability of precipitation and temperature. A statistical downscaling model based on K-means clustering, CART and SVR was developed to predict the future scenarios of precipitation and temperature. Both precipitation and temperature were predicted to increase by 13.12% and 0.6 °C respectively for the period 2061–2080 compared to the observed period of 1951–1989 with MIROC GCM outputs. The increase of temperature has predicted an increase of PET for the future scenarios. It can be noted that the AET estimates are defined by P and PET changes. The increase of temperature has resulted in an increase of PET and consequent increase of AET over KRB for the future scenarios. It can be noted that the projected increase of AET is more pronounced under an increase of PET compared to the increase of P, resulting in a decrease of runoff for the future scenarios over the KRB. The projected decrease of basin outlet runoff for the periods 2021–2040, 2041–2060 and 2061–2080 compared to the observed period 1951–1989 was noted as 15, 14.69 and 10.09% respectively. Overall, the basin outlet runoff was predicted to decrease (13.26%) for future scenarios under climate change due to an increase in temperature (0.6 °C) compared to the precipitation increase (13.12%) over the KRB. As the conceptual model applied is based on the basic hydrological variables of P and AET, the resulting runoff outlets were predicted to decrease under the projected increase of AET. Overall, severe water shortage was predicted under climate change due to the projected increase of AET compared to the precipitation over KRB.

The proposed modelling framework has considered one hydrological model with one particular GCM simulation under a particular concentration pathway scenario for demonstration purposes. It should be noted that hydrological response can be expected to vary with various hydrological models, calibration parameters, statistical downscaling models, various concentration path scenarios, and various machine learning algorithms. A detailed study may be evident under the possible range of uncertainties arising at each stage of the modelling framework, which can be a potential research problem.

The present study considered a calibration-free formulation of Budyko's hypothesis to estimate the annual AET on the catchment scale. The study intended to demonstrate how a hydrological calibration factor can be modelled based on various hydro-climatological variables and the same can be utilized for the climate change impact assessment. For this purpose, the catchment scale single calibration factor based conceptual hydrological model was adopted. However, the proposed methodology can be implemented with any such standard hydrological model, accounting for the number of calibration factors involved and the suitable data-driven model to be adopted. Furthermore, the present modelling framework can be extended by introducing multiple calibration factors accounting for the land use, soil, etc. (Jarsjö *et al.* 2008) in the hydrological modelling, which is a potential area of research. The main emphasis of the study was to develop a modelling framework to study the hydrological responses under climate change using a conceptual hydrological model with an integration of a calibration framework based on machine learning algorithms. The proposed modelling framework can be extendable for other river basins with sufficient data availability and validation by applying various machine learning algorithms. With the use of most readily available and operational hydrometeorological variables such as precipitation, temperature and runoff, the proposed model can be promising to study the catchment scale hydrological responses under climate change. However, validation of the modeled results is most critical in the implementation for other catchment studies. Specifically, the use of calibration-free Budyko formulation has to be validated with the water-balance or most dependable and state-of-the-art satellite based AET estimates. Furthermore, the study limited for annual water balances by neglecting the storage changes which can be further extended with the inclusion of appropriate storage components to study the water balances other than on annual scale. Due to the limitation over the availability of various meteorological variables in the estimation of PET, the present study adopted the Thornthwaite model, which can be extendable to the standard Penman model. Overall, the proposed methodology of inclusion of time-variant calibration factors in the conceptual hydrological model accounting for various hydrological variables is more reliable for catchment-scale water management under climate change.

---

## ACKNOWLEDGEMENTS

The research work presented in the paper was funded by Science and Engineering Research Board (SERB), Department of Science and Technology, Government of India through Start-up Grant for Young Scientists (YSS) Project no. YSS/2015/002111 to Dr S. Rehana. We sincerely acknowledge the India Meteorological Department (IMD) for providing the gridded observations of precipitation and temperatures and details of the data are given in the website at <https://iridl.ldeo.columbia.edu/SOURCES/.IMD/index.html?Set-Language=en>. The authors sincerely thank the editors and anonymous reviewers for reviewing the manuscript and providing insightful comments.

---

## DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

## REFERENCES

- Aghakouchak, A. & Habib, E. 2010 Application of a conceptual hydrologic model in teaching hydrologic processes. *International Journal of Engineering Education* **26** (4(S1)), 963–973. Retrieved from <https://escholarship.org/uc/item/3sv066q5>.
- Ammann, L., Doppler, T., Stamm, C., Reichert, P. & Fenicia, F. 2020 Characterizing fast herbicide transport in a small agricultural catchment with conceptual models. *Journal of Hydrology* **586**, 124812. <https://doi.org/10.1016/j.jhydrol.2020.124812>.
- Anabalón, A. & Sharma, A. 2017 On the divergence of potential and actual evapotranspiration trends: an assessment across alternate global datasets. *Earth's Future* **5** (9), 905–917. <https://doi.org/10.1002/2016EF000499>.
- Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., Santhi, C., Harmel, R. D., van Griensven, A., Van Liew, M. W., Kannan, N. & Jha, M. K. 2012 SWAT: model use, calibration, and validation. *Transactions of the ASABE* **55** (4), 1491–1508. <https://doi.org/10.13031/2013.42256>.
- Asokan, S. M., Jarsjö, J. & Destouni, G. 2010 Vapor flux by evapotranspiration: effects of changes in climate, land use, and water use. *Journal of Geophysical Research: Atmospheres* **115** (D24). <https://doi.org/10.1029/2010JD014417>.
- Bai, P., Liu, X., Zhang, D. & Liu, C. 2020 Estimation of the Budyko model parameter for small basins in China. *Hydrological Processes* **34** (1), 125–138. <https://doi.org/10.1002/hyp.13577>.
- Bharti, B., Pandey, A., Tripathi, S. K. & Kumar, D. 2017 Modelling of runoff and sediment yield using ANN, LS-SVR, REPTree and M5 models. *Hydrology Research* **48** (6), 1489–1507. <https://doi.org/10.2166/nh.2017.153>.
- Breiman, L. 2001 Random forests. *Machine Learning* **45** (1), 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Breiman, L. 2017 *Classification and Regression Trees*. Routledge, New York. <https://doi.org/10.1201/9781315139470>.
- Budyko, M. 1974 *Climate and Life*. Academic Press, New York.
- Bui, D. T., Khosravi, K., Tiefenbacher, J., Nguyen, H. & Kazakis, N. 2020 Improving prediction of water quality indices using novel hybrid machine-learning algorithms. *Science of The Total Environment* **721**, 137612. <https://doi.org/10.1016/j.scitotenv.2020.137612>.
- Choudhury, B. J. 1999 Evaluation of an empirical equation for annual evaporation using field observations and results from a biophysical model. *Journal of Hydrology* **216** (1), 99–110. [https://doi.org/10.1016/S0022-1694\(98\)00293-5](https://doi.org/10.1016/S0022-1694(98)00293-5).
- Donohue, R. J., Roderick, M. L. & McVicar, T. R. 2007 On the importance of including vegetation dynamics in Budyko's hydrological model. *Hydrology and Earth System Sciences* **11**, 983–995. <https://doi.org/10.5194/hess-11-983-2007>.
- Friedman, J. H. 2001 Greedy function approximation: a gradient boosting machine. *Annals of Statistics* **29** (5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>.
- Goyal, M. K. & Khan, M. 2017 Assessment of spatially explicit annual water-balance model for Sutlej River Basin in eastern Himalayas and Tungabhadra River Basin in peninsular India. *Hydrology Research* **48** (2), 542–558. <https://doi.org/10.2166/nh.2016.053>.
- Gunkel, A. & Lange, J. 2017 Water scarcity, data scarcity and the Budyko curve – an application in the Lower Jordan River Basin. *Journal of Hydrology: Regional Studies* **12**, 136–149. <https://doi.org/10.1016/j.ejrh.2017.04.004>.
- Ha, L. T., Bastiaanssen, W. G. M., van Griensven, A., van Dijk, A. I. J. M. & Senay, G. B. 2017 SWAT-CUP for calibration of spatially distributed hydrological processes and ecosystem services in a Vietnamese river basin using remote sensing. *Hydrology and Earth System Sciences Discussions* 1–35. <https://doi.org/10.5194/hess-2017-251>.
- Han, S., Tian, F. & Hu, H. 2014 Positive or negative correlation between actual and potential evaporation? Evaluating using a nonlinear complementary relationship model. *Water Resources Research* **50** (2), 1322–1336. <https://doi.org/10.1002/2013WR014151>.
- Hargreaves, G. H. & Allen, R. G. 2003 History and evaluation of Hargreaves evapotranspiration equation. *Journal of Irrigation and Drainage Engineering* **129** (1), 53–63. [https://doi.org/10.1061/\(ASCE\)0733-9437\(2003\)129:1\(53\)](https://doi.org/10.1061/(ASCE)0733-9437(2003)129:1(53)).
- Hengade, N., Eldho, T. I. & Ghosh, S. 2018 Climate change impact assessment of a river basin using CMIP5 climate models and the VIC hydrological model. *Hydrological Sciences Journal* **63** (4), 596–614. <https://doi.org/10.1080/02626667.2018.1441531>.
- Hu, C., Wu, Q., Li, H., Jian, S., Li, N. & Lou, Z. 2018 Deep learning with a long short-term memory networks approach for rainfall-runoff simulation. *Water* **10** (11), 1543. <https://doi.org/10.3390/w10111543>.
- Immerzeel, W. W. & Droogers, P. 2008 Calibration of a distributed hydrological model based on satellite evapotranspiration. *Journal of Hydrology* **349** (3), 411–424. <https://doi.org/10.1016/j.jhydrol.2007.11.017>.
- Jaiswal, R. K., Ali, S. & Bharti, B. 2020 Comparative evaluation of conceptual and physical rainfall-runoff models. *Applied Water Science* **10** (1), 48. <https://doi.org/10.1007/s13201-019-1122-6>.
- Jarsjö, J., Shibuo, Y. & Destouni, G. 2008 Spatial distribution of unmonitored inland water discharges to the sea. *Journal of Hydrology* **348** (1–2), 59–72. <https://doi.org/10.1016/j.jhydrol.2007.09.052>.
- Knoben, W. J. M., Freer, J. E. & Woods, R. A. 2019 Technical note: inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrology and Earth System Sciences* **23** (10), 4323–4331. <https://doi.org/10.5194/hess-23-4323-2019>.
- Li, H., Sheffield, J. & Wood, E. F. 2010 Bias correction of monthly precipitation and temperature fields from intergovernmental panel on climate change AR4 models using equidistant quantile matching. *Journal of Geophysical Research: Atmospheres* **115** (D10). <https://doi.org/10.1029/2009JD012882>.



- Lima, A. R., Cannon, A. J. & Hsieh, W. W. 2012 **Downscaling temperature and precipitation using support vector regression with evolutionary strategy**. In: *The 2012 International Joint Conference on Neural Networks (IJCNN)*, Brisbane, Australia. pp. 1–8. <https://doi.org/10.1109/IJCNN.2012.6252383>.
- Loh, W.-Y. 2011 **Classification and regression trees**. *WIREs Data Mining and Knowledge Discovery* **1** (1), 14–23. <https://doi.org/10.1002/widm.8>.
- López, L., Sutanudjaja, P., Schellekens, E. H., Sterk, J., & Bierkens, G. & P, M. F. 2017 **Calibration of a large-scale hydrological model using satellite-based soil moisture and evapotranspiration products**. *Hydrology and Earth System Sciences* **21** (6), 3125–3144. <https://doi.org/10.5194/hess-21-3125-2017>.
- Meenu, R., Rehana, S. & Mujumdar, P. P. 2013 **Assessment of hydrologic impacts of climate change in Tunga–Bhadra river basin, India with HEC-HMS and SDSM**. *Hydrological Processes* **27** (11), 1572–1589. <https://doi.org/10.1002/hyp.9220>.
- Mianabadi, A., Coenders-Gerrits, M., Shirazi, P., Ghahraman, B. & Alizadeh, A. 2019 **A global Budyko model to partition evaporation into interception and transpiration**. *Hydrology and Earth System Sciences* **23**, 1–29. <https://doi.org/10.5194/hess-23-4983-2019>.
- Moore, I. D., Grayson, R. B. & Ladson, A. R. 1991 **Digital terrain modelling: a review of hydrological, geomorphological, and biological applications**. *Hydrological Processes* **5** (1), 3–30. <https://doi.org/10.1002/hyp.3360050103>.
- Nash, J. E. & Sutcliffe, J. V. 1970 **River flow forecasting through conceptual models part I – a discussion of principles**. *Journal of Hydrology* **10** (3), 282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6).
- Pai, D. S., Sridhar, L., Rajeevan, M. C., Sreejith, O. P., Satbhai, N. S. & Mukhopadhyay, B. 2014 **Development of a new high spatial resolution (0.25° × 0.25°) long period (1901–2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region**. *Mausam* **65** (1), 1–18.
- Parra, V., Fuentes-Aguilera, P. & Muñoz, E. 2018 **Identifying advantages and drawbacks of two hydrological models based on a sensitivity analysis: a study in two Chilean watersheds**. *Hydrological Sciences Journal* **63** (12), 1831–1843. <https://doi.org/10.1080/02626667.2018.1538593>.
- Penman, H. L. 1950 **The dependence of transpiration on weather and soil conditions**. *Journal of Soil Science* **1** (1), 74–89. <https://doi.org/10.1111/j.1365-2389.1950.tb00720.x>.
- Pohlert, T., Breuer, L., Huisman, J. A. & Frede, H.-G. 2007 **Assessing the model performance of an integrated hydrological and biogeochemical model for discharge and nitrate load predictions**. *Hydrology and Earth System Sciences* **11** (2), 997–1011. <https://doi.org/10.5194/hess-11-997-2007>.
- Prasad, R., Deo, R. C., Li, Y. & Maraseni, T. 2018 **Soil moisture forecasting by a hybrid machine learning technique: ELM integrated with ensemble empirical mode decomposition**. *Geoderma* **330**, 136–161. <https://doi.org/10.1016/j.geoderma.2018.05.035>.
- Rehana, S. & Mujumdar, P. P. 2012 **Climate change induced risk in water quality control problems**. *Journal of Hydrology* **444–445**, 63–77. <https://doi.org/10.1016/j.jhydrol.2012.03.042>.
- Rehana, S., Sireesha Naidu, G., Monish, N. T. & Sowjanya, U. 2020a **Modeling hydro-climatic changes of evapotranspiration over a semi-arid river basin of India**. *Journal of Water and Climate Change*. jwc2020173. <https://doi.org/10.2166/wcc.2020.173>.
- Rehana, S., Naidu, G. S. & Monish, N. T. 2020b **Spatiotemporal variations of precipitation and temperatures under CORDEX climate change projections: a case study of Krishna River Basin, India**. In: *Contemporary Environmental Issues and Challenges in Era of Climate Change* (Singh, P., Singh, R. P. & Srivastava, V. eds). Springer, Singapore, pp. 157–170. [https://doi.org/10.1007/978-981-32-9595-7\\_8](https://doi.org/10.1007/978-981-32-9595-7_8).
- Rientjes, T. H. M., Muthuwatta, L. P., Bos, M. G., Booij, M. J. & Bhatti, H. A. 2013 **Multi-variable calibration of a semi-distributed hydrological model using streamflow data and satellite-based evapotranspiration**. *Journal of Hydrology* **505**, 276–290. <https://doi.org/10.1016/j.jhydrol.2013.10.006>.
- Rooijen, D. J. V., Turrall, H. & Biggs, T. W. 2009 **Urban and industrial water use in the Krishna Basin, India**. *Irrigation and Drainage* **58** (4), 406–428. <https://doi.org/10.1002/ird.439>.
- Sajedi-Hosseini, F., Malekian, A., Choubin, B., Rahmati, O., Cipullo, S., Coulon, F. & Pradhan, B. 2018 **A novel machine learning-based approach for the risk assessment of nitrate groundwater contamination**. *Science of the Total Environment* **644**, 954–962. <https://doi.org/10.1016/j.scitotenv.2018.07.054>.
- Salvi, K., Kannan, S. & Ghosh, S. 2013 **High-resolution multisite daily rainfall projections in India with statistical downscaling for climate change impacts assessment**. *Journal of Geophysical Research: Atmospheres* **118** (9), 3557–3578. <https://doi.org/10.1002/jgrd.50280>.
- Saraf, V. R. & Regulwar, D. G. 2018 **Impact of climate change on runoff generation in the Upper Godavari River Basin, India**. *Journal of Hazardous, Toxic, and Radioactive Waste* **22** (4), 04018021. [https://doi.org/10.1061/\(ASCE\)HZ.2153-5515.0000416](https://doi.org/10.1061/(ASCE)HZ.2153-5515.0000416).
- Singh, R. & Kumar, R. 2015 **Vulnerability of water availability in India due to climate change: a bottom-up probabilistic Budyko analysis**. *Geophysical Research Letters* **42** (22), 9799–9807. <https://doi.org/10.1002/2015GL066363>.
- Sinha, J., Sharma, A., Khan, M. & Goyal, M. K. 2018 **Assessment of the impacts of climatic variability and anthropogenic stress on hydrologic resilience to warming shifts in Peninsular India**. *Scientific Reports* **8** (1), 13833. <https://doi.org/10.1038/s41598-018-32091-0>.
- Sood, A. & Smakhtin, V. 2015 **Global hydrological models: a review**. *Hydrological Sciences Journal* **60** (4), 549–565. <https://doi.org/10.1080/02626667.2014.950580>.

- Srivastava, A. K., Rajeevan, M. & Kshirsagar, S. R. 2009 Development of a high resolution daily gridded temperature data set (1969–2005) for the Indian region. *Atmospheric Science Letters* **10** (4), 249–254. <https://doi.org/10.1002/asl.232>.
- Srinivasa Raju, K. & Nagesh Kumar, D. 2018 Hydrological modeling. In: *Impact of Climate Change on Water Resources: With Modeling Techniques and Case Studies* (Srinivasa Raju, K. & Nagesh Kumar, D., eds). Springer, Singapore, pp. 137–167. [https://doi.org/10.1007/978-981-10-6110-3\\_5](https://doi.org/10.1007/978-981-10-6110-3_5).
- Thornthwaite, C. W. 1948 An approach toward a rational classification of climate. *Geographical Review* **38** (1), 55–94. <https://doi.org/10.2307/210739>.
- Vormoor, K., Heistermann, M., Bronstert, A. & Lawrence, D. 2018 Hydrological model parameter (in)stability – ‘crash testing’ the HBV model under contrasting flood seasonality conditions. *Hydrological Sciences Journal* **63** (7), 991–1007. <https://doi.org/10.1080/02626667.2018.1466056>.
- Wilby, R. L. & Dawson, C. W. 2013 The statistical downScaling model: insights from one decade of application. *International Journal of Climatology* **33** (7), 1707–1719. <https://doi.org/10.1002/joc.3544>.
- Wit, M. J. M. d. 2001 Nutrient fluxes at the river basin scale. I: the PolFlow model. *Hydrological Processes* **15** (5), 743–759. <https://doi.org/10.1002/hyp.175>.
- Wu, J., Miao, C., Wang, Y., Duan, Q. & Zhang, X. 2017 Contribution analysis of the long-term changes in seasonal runoff on the Loess Plateau, China, using eight Budyko-based methods. *Journal of Hydrology* **545**, 263–275. <https://doi.org/10.1016/j.jhydrol.2016.12.050>.
- Zhang, L., Hickel, K., Dawes, W. R., Chiew, F. H. S., Western, A. W. & Briggs, P. R. 2004 A rational function approach for estimating mean annual evapotranspiration. *Water Resources Research* **40** (2), W025021. <https://doi.org/10.1029/2003WR002710>.
- Zhang, J., Zhang, Y., Song, J., Cheng, L., Kumar Paul, P., Gan, R., Shi, X., Luo, Z. & Zhao, P. 2020 Large-scale baseflow index prediction using hydrological modelling, linear and multilevel regression approaches. *Journal of Hydrology* **585**, 124780. <https://doi.org/10.1016/j.jhydrol.2020.124780>.

First received 30 April 2020; accepted in revised form 17 August 2020. Available online 11 November 2020