Adaptive hybrid architecture for enhancement of the complex hydroclimatic system and assessment of freshwater security

Venkatesh Budamala and Amit Baburao Mahindrakar*
Department of Environmental and Water Resources Engineering, School of Civil Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India
*Corresponding author. E-mails: amahindrakar@vit.ac.in, amahindrakar@gmail.com

ABSTRACT
Future freshwater security relies on hydroclimatic (HC) shifts and regimes for sustainable development. The approximation of the HC system faces major uncertainties and complexities due to the incorporation of heavy datasets, characteristics, and constraints. The proposed study focused on the parallel computing of emulator modeling-based spatial optimization to enhance the HC systems with the perspective of future freshwater security in the Upper Chattahoochee River basin (UCR). Here, the framework compiles both physical and machine learning concepts with adaptive technology for the replication of real-world scenarios. Besides, it contains 2Emulator Model Fitting, Spatial Optimization, Parallel Computing, and Initial and Adaptive sampling to upgrade model efficiency. While UCR has inadequate groundwater and the assessment of freshwater security in UCR is more necessary for varying future climatic conditions. The results displayed that the proposed spatial optimization algorithm proved to be an effective and efficient approach in the approximation of HC models. The assessment of water security in UCR was showed in terms of scarcity and vulnerability indicators for median and low-level conditions, respectively. Moreover, this study provides the potential framework for the enhancement of physical model predictions with the incorporation of hybrid concepts for problem-solving technology which can provide significant information on HC issues.

Key words: adaptive machine learning, hybrid simulators, hydroclimatology, parallel computing, spatial optimization, water security

HIGHLIGHTS
• A comprehensive framework for the integration of the physical and machine learning concepts to enhance the hydroclimatic system.
• Adaptive emulator-based spatial optimization was introduced to control expensive simulations.
• Parallel computing was incorporated in the framework to restrict the spatial variability in large-scale watersheds.
• Assessed the future freshwater security based on the Blue/Green Water dynamics.
1. INTRODUCTION

Integrated Watershed Planning and Management (IWPM) delivers potential solutions in the basin behavior with long-term planning, promotes consistency and efficiency, optimizes the use of the water system, encourages and facilitates regional planning, provides flexible solutions, and enhances communication and community support (Freas et al. 2008; Black 2014; Wang et al. 2016b; Goharian et al. 2017; Budamala & Mahindrakar 2020a). The IWPM system relies on different characteristics and consequences, it ultimately reflects uncertainty and complexity during the replication of the real-world phenomena (Wang et al. 2016b). Besides, future scenarios are also a major concern that affect the identification of the system due to overwhelming variations in emission levels (Luo et al. 2019). Complexity unveils with an upsurge of the model requirements. Harmonizing both uncertainty and complexity can provide efficiency and feasibility in the model outcomes (Budamala & Mahindrakar 2020a). The level of water security relies on different interactions in the ecosystem functioning, societal needs, and hydroclimatic (HC) conditions. However, it focuses on the attention of hydrological systems by shifting the boundary between environment and society, whose dynamics are crucial for determining water security, human safety, and sustainable development (Giacomoni & Berglund 2015). Moreover, freshwater security is partitioned into Blue (BW) and Green Water (GW) based on the hydrological processes. BW flows either above or below the land surface and is stored in lakes, rivers, aquifers, and reservoirs, whereas GW indicates the portion of water that infiltrates the soil moisture and temporally stores on the top of the soil or vegetation and ultimately returns to the atmosphere via evapotranspiration. The consumption of these water components inclined by human activities is referred to as water footprint. Consequently, Blue Water Footprint (BW-footprint) is the consumptive use of surface water and groundwater bodies, while Green Water Footprint (GW-Footprint) is expended by agricultural activities.

Future water security of both BW and GW plays a vital role in IWPM which is a prerequisite to develop the HC models for the hefty uncertainty and complexity that may arise during model development (Luo et al. 2019). For instance, the global climate models generate the data over a period considering dynamic formulations for the representation of different concentrative pathways, which primes to uncertainty in the selection of scenarios and enhances the complexity to cascading from the global conditions to local conditions (Luo et al. 2017). Whereas the hydrological models face uncertainty and complexity during the depiction of water balance components for varying climate changes. However, hybridizing of both hydrological and climatic models results in colossal variability due to anthropogenic activities and it is essential to consider it during approximation or optimization (Luo et al. 2019).
Initially, the models may not have the optimal setup due to the incorporation of huge datasets, equations, parameters, and constraints. Later, it needs efficient calibration or optimization approaches to approximate the phenomena (Abbaspour et al. 2015). Different approaches are used to enhance the hydrological and climate models separately (Abbaspour et al. 2015; Luo et al. 2017, 2019; Osei et al. 2018), but hybridizing both the components with the help of a single framework is not addressed. The hybrid models can afford the characteristics of two or more systems and provide a detailed analysis of the structure (Zhang et al. 2012). For different water management applications, there are various models with advantages. Of all the models, the conceptual (or physical) and data-driven models had more advantages than the other models. At this point, the conceptual model represents a physical phenomenon of the system with the incorporation of heavy parameters and datasets. However, the data-driven models cannot follow physical phenomena, but it fits the data effectively according to the target. Hence, hybridizing both conceptual and data-driven models provides an efficient and effective system with accurate model outcomes (Osei et al. 2018).

The approximation of HC models has a stepwise strategy. Firstly, the climate model needs to be enhanced to that of the hydrological model. Here, the output of the climate model served as the input of the hydrological model. Different climate models are evolved based on their perspectives for diverse meteorological variables. But the climate models are generated based on the global scenarios and misleads to local conditions. Therefore, they need to downscale from global to local conditions for effective watershed analysis. On the other hand, the conceptual hydrological models provide outcomes by adopting the physical and meteorological characteristics of the watershed. Since it contains an enormous number of parameters, it is desirable for the effective optimization approach to analyze the system behavior. Hence, the present study used data-driven models for the optimization of HC models. With the help of the approximated HC model one can address the BW and GW Security for historical as well as future scenarios.

An effective optimization tool can offer ease in the processing of the system and also helps in analyzing the system behavior. This research focused on the adaptive optimization strategy for enhancing the predictions of HC model outcomes. Besides, the framework developed in the R programming platform supports the open source. Furthermore, the framework complied with bias correction and extraction of future climatic variables, optimization of the hydrological system, and finally the assessment of HC applications. The Parallel Computing of Emulator Modelling-based Spatial Optimization (PCESO) was introduced to optimize an expensive HC simulation. It follows an adaptive strategy to restrict or control the sample size and navigates the optimal search space (Wang et al. 2014; Budamala & Mahindrakar 2020a). However, the large-scale watersheds contain heavy spatial variability and it is essential to optimize different stations together. To support that, the optimization algorithm incorporated parallel computing to confine the variability by spatial optimization (Rouholahnejad et al. 2012; Budamala & Mahindrakar 2020a).

To validate this framework, a complex urban watershed of the Metropolitan North Georgia Water Planning District has been selected for future water accessibility applications. Here, Metro Water District (MWD) serves the reuse, conservation, and supply of water resources to the metropolitan region. Here, MWD has made efforts to develop the future needs for water supply, water conservation, wastewater, and watershed management actions. It focuses more on the interconnections between the water resource policies encountered in the planning and management of the watershed (Black 2014; Budamala & Mahindrakar 2020a). Hence, an efficient system needs to be designed for supporting watershed management and addressing the future perspectives of water availability in MWD. Sustainable water dynamics can provide useful information that helps to improve water security for future needs (Ahn & Kim 2018; Du et al. 2018). The present study focused to provide a user-friendly framework for HC modeling with an end product of problem-solving applications for present and future conditions. The major objectives are (a) extraction of future climate variables, (b) optimization of the complex hydrological system, and (c) assessment of future BW and GW accessibility for scarcity and vulnerability conditions.

2. STUDY AREA

The Upper Chattahoochee River (UCR) is one of the major river basins in the Metropolitan North Georgia Water Planning District and it covers 18% of the total MWD area (Figure 1). Besides, UCR has contributed 57% of its total area to MWD, and this area provides drinking and primary receiving water from treated wastewater effluent in the Atlanta Metro Region (Black 2014). This river basin flows from the Blue Ridge Mountains of the northeast to southwest with an elevation of 194–135 m from the mean sea level. Besides, UCR is serving nearly 29 cities and 7 counties including Cherokee, Cobb, DeKalb, Forsyth, Fulton, Gwinnett, and Hall. The entire Northern Fulton county is assimilated to UCR in which one-third of the portion is
comprised of the Atlanta Metro Region and newly integrated cities are also incorporated in UCR like Brookhaven and the City of Peachtree Corners in DeKalb and Gwinnett Counties, respectively. As new cities are developing, proper water resources management across the basin needs to be ensured. UCR deposits large quantities of the weathered and

Figure 1 | The location of the Upper Chattahoochee River basin in the Metropolitan North Georgia Water Planning District. *Note: WETN (Wetlands-Non-Forested), WETF (Wetlands-Forested), WATR (Water), URMD (Residential-Medium Density), URLD (Residential-Low Density), URHD (Residential-High Density), UIDU (Industrial), SWRN (Southwestern US (Arid) Range), RNGE (Range-Grasses), RNGB (Range-Brush), HAY (Hay), FRST (Forest-Mixed), FRSE (Forest-Evergreen), FRSD (Forest-Deciduous), AGRR (Agricultural Land-Row Crops).
unconsolidated rock debris in the aquifer spaces. These deposits are heavy in valleys and result in inadequate yield for users other than very low-density residential areas and thus surface water is the primary source of potable water for the UCR basin (Black 2014). Moreover, this basin contains free-flowing natural tributaries, and the availability of groundwater is inadequate due to geologic conditions, which block the potential yield for water supply.

The flow of UCR in MWD is operated by the Bufford Dam, which was built for capturing the Lake Lanier (LL). LL is a multi-purpose reservoir and it extends up to 2,800 km² which provides flood protection, power production, water supply, navigation, recreation, and fish and wildlife management. It is the largest reservoir in MWD and provides the majority of MWD’s water supply, either through direct withdrawals or downstream releases (Black 2014). Monthly average flows near to the Atlanta region range from 253 to 24 m³/s with a mean flow of 70 m³/s. While precipitation ranges from 1,524 to 1,326 mm in the northeastern part to southwestern part of the watershed, respectively. Land Use and Land Cover (LULC) is fully dominated by an effective impervious area in the UCR basin which occupied nearly 46% of the total watershed area (Figure 1). Also, the southwestern part of the basin and downstream of the Lake Lanier are considered to be a more densely developed part of the watershed, which encompasses the Metro Atlanta Region, Buckhead, and Decatur. In Figure 1, Station 7 is ungauged, hence we considered Stations 1–6 for calibration of the SWAT model. Table 1 displays obligatory datasets of the UCR basin in each stage to run this framework.

Table 1 | Information of datasets used for hydroclimatic modeling to assess for water accessibility

| Variable                                      | Period of study | Time step | Source                                      |
|-----------------------------------------------|-----------------|-----------|---------------------------------------------|
| Meteorological features – observed climate variables |                 |           |                                             |
| Precipitation                                 | 1991–2015       | Daily     | National Oceanic and Atmospheric Administration (NOAA) |
| Temperature                                   |                 |           |                                             |
| (Minimum and Maximum)                         |                 |           |                                             |
| Solar Radiation                               |                 |           | Climate Forecast System Reanalysis (CPSR)    |
| Wind                                          |                 |           |                                             |
| Relative Humidity                             |                 |           |                                             |
| Topographical features for hydrological modeling |                 |           |                                             |
| Digital Elevation Model (DEM)                 | 10 m            | 2006      | National Hydrography Dataset Plus (NHD-Plus) |
| Land Use Land Cover (LULC)                    | 30 m            | 2011      | National Land Cover Database (NLCD)         |
| Soil Classification                           | 1:250,000       | 1995      | National Cooperative Soil Survey and supersedes the State Soil Geographic (STATSGO) |

| S. No. | Station ID | Station Name                  | Period of study | Source                          |
|--------|------------|--------------------------------|-----------------|---------------------------------|
| 1      | 02331600   | Chattahoochee River Near Cornelia |                 | Georgia State Surface Water, USGS |
| 2      | 02333500   | Chestatee River Near           |                 |                                  |
| 3      | 02334430   | Dahlonega                      |                 |                                  |
| 4      | 02335700   | Chattahoochee River at Buford Dam |                 |                                  |
| 5      | 02336000   | Big Creek Near Alpharetta      |                 |                                  |
| 6      | 02336300   | Peachtree Creek at Atlanta     | 1991–2015       |                                  |
3. FRAMEWORK

A BW/GW security assesses through a stepwise framework of an adaptive hybrid approach (Figure 2). The hybrid concepts are developed here for coupling of the physical and data-driven models for hydroclimatology applications. The main components involved in the adaptive hybrid framework in this study are given as follows:

(a) extraction and bias correction of climate variables,
(b) development of the SWAT physical hydrological model,
(c) optimization of the uncalibrated hydrological model, and
(d) assessment of BW/GW security from the approximated HC model.

3.1. Climate model

In this study, the Canadian Centre for Climate Modelling and Analysis Regional Climate Model (CanRCM4) was adopted as a climate model for the prediction of both historical and future water security applications. CanRCM4 is an innovative approach of coordinated global and regional climate modeling by a parent Global Climate Model (GCM) CanESM2 (Ben Alaya et al. 2019). Generally, CanRCM4 was designed to downscale the climate projections and its predictions for regional conditions with the close association of the parent GCM. Also, it provides a novel approach in model development to offer the data where independent regional climate modeling centers’ data are not accessible. It enhanced the quality of simulations by incorporating the driving information of prognostic variables. Moreover, CanRCM4 adopted a spectral nudging procedure for downscaling large-scale driving data of CanESM2 (Scinocca et al. 2016). The RCM data of essential meteorological variables (i.e., Precipitation (Prec.), Maximum Temperature, and Minimum Temperature (Max. Temp and Min. Temp) are used for building the hydrological phenomena obtained from the NA-CORDEX portal. Historical scenario of 1950–2005 and Future scenarios (RCP4.5 and RCP8.5) of 2006–2100 with a resolution of NAM-22i (i.e., quarter degree of the latitude–longitude grid) daily time step are considered for the assessment of BW/GW security. Here, RCP4.5 is a scenario of long-term,
global emissions of greenhouse gases, short-lived species, and LULC stabilizes radiative forcing at 4.5 W/m² in the year 2100 without ever. Similarly, RCP8.5 refers to the concentration of carbon that delivers global warming at an average of 8.5 W/m² across the planet. The RCP8.5 pathway will deliver a temperature increase of about 4.3 °C by 2100, relative to pre-industrial temperatures.

3.2. Hydrological model

The Soil and Water Assessment Tool (SWAT) is a physical, semi-distributed, and conceptual hydrological model (Dile et al. 2016). It helps to identify the hydrological phenomena and it analyzes the impacts on the water quantity as well as water quality aspects based on the physical characteristics of the watershed (Gassman et al. 2007; Dile et al. 2016). Initially, SWAT delineates the watershed into sub-watersheds based on the elevation inputs. Then, it identifies the Hydrological Response Units (HRUs) at every sub-watershed for dominant land use, soil, and slope characteristics. The meteorological data provide input to the hydrological system on a daily and sub-daily time-series basis. Here, the SWAT provided weather generators option to fill the missing datasets. Finally, the model setup is based on the quantity and quality aspects of outcomes for expected periods. The SWAT model follows the water balance equation for simulating the hydrological phenomena. It helps to predict the hydrological phenomena for basin, sub-basin, and HRU levels (Athira et al. 2016; Luo et al. 2019; Budamala & Mahindrakar 2020c).

3.3. Optimization algorithm

The physical model may not obtain optimal results at the initial setup due to the involvement of a huge number of parameters, datasets, and structures (Zhang et al. 2009; Ahmadi et al. 2014; Budamala & Mahindrakar 2020a). Hence, it is necessary to validate the default model outputs with the observed data. If the default model is not satisfied, then the model needs to be calibrated for the approximation of outcomes. In this study, the effective optimization algorithm ‘PCESO’ was incorporated to enhance the conceptual SWAT hydrological model. PCESO employs adaptive pseudoregression using parallel spatial optimization to replicate the real-world simulation (Budamala & Mahindrakar 2020a). PCESO follows the principle of data-driven concept by the input–output response. Here, the algorithm fits with initial samples and checks the criteria; if it is not satisfying the criteria, it adds extra samples iteratively until it reaches the stopping criteria. The components of PCESO are Initial Sampling or Design (IS), Emulator Fitting, Adaptive Sampling (AS), Spatial Optimization, and Parallel computing (Figure 2) which are further explained below.

3.3.1. Sampling design

The sampling design of PCESO is used in two ways during optimization like IS and AS. Initially, the parameter sets of IS fit the emulator model based on the pseudoresponse. Later, if the emulator model does not meet the criteria, it will add a few more parameter sets from AS (Wang et al. 2014; Budamala & Mahindrakar 2020a). Here, the parameter sets are considered as input–output responses such as influential parameters (input response) and the objective function (output response). The influential parameters of the UCR basin screened out using Global Sensitivity Analysis (GSA) and Nash–Sutcliffe Efficiency (NSE) are considered as the objective function.

Table 2 displays the most influential parameters of the UCR basin in terms of BW and GW security based on Budamala & Mahindrakar (2020). Here, the Curve Number (CN2) showed the most influential parameter due to rapid changes in the LULC which stimulates the surface runoff (Zhang et al. 2016). Hence, the adjusting CN2 is a major part of hydrological model simulation for avoiding false assessments in the BW and GW security in terms of runoff and storage. TheESCO parameter represents the evapotranspiration compensation factor which focused on the moisture content that can be absorbed from the bottom of the layers. Similarly, SOL_AWC influences water security due to peculiar changes in the water capacity of the soil layers. The groundwater parameters represent the percolation, depth, and delay to reach from soil layers to the aquifer in the basin, while SURLAG denotes the time of concentration from the starting point to the river bed. However, the rapid changes in the basin can affect SURLAG which ultimately will be reflected in the model approximation. Finally, the snowfall parameters showed the influence in UCR for snowmelt and snowfall in distinct seasons. Hence, it is also playing a major role in UCR for water security even though it deliberates less.

In this study, Quasi Random Sampling (QRS) and Lola-Voronoi have selected for IS and AS, respectively. QRS has uniform sampling with a lower discrepancy sequence in effective response for space-filling (Wang et al. 2014; Gong et al. 2016; Budamala & Mahindrakar 2020a). QRS is following the Monte Carlo method for the integration and simulation of response surface without following any random behavior. It represents the replacement of the original simulation by following
surrogates or pseudo response to reduce clumping by correlating the sample points which enhances the uniformity in the state of the discrepancy system. Thus, it is also termed as low discrepancy. On the other hand, AS is considered the LOLA-Voronoi algorithm and it contains hybrid characterization. Where Voronoi measures a gradient in each available point and helps create a design for nonlinear regions. While LOLA (Local Linear Approximation) component is designed to compare an assessment between dynamic to smooth regions. Hence, the updated surface needs both Voronoi exploration and the LOLA strategy to show the sampling performance of the space surrounded by the reference surface. Therefore, the combination of both LOLA and Voronoi can provide accurate results for linear to nonlinear regions with handling lower- to higher-dimensional updated surface problems (Budamala & Mahindrakar 2020b).

### 3.3.2. Emulator model

The emulator model replicates the real-world simulation by adopting the black-box concepts (Forrester et al. 2008; Luo & Lu 2014; Wang et al. 2016a; Budamala & Mahindrakar 2020c). In this study, Extreme Learning Machines (ELM) adopted as an emulator model to fit and enhance the hydrological response of SWAT. Here, the emulator follows the relationship between the input–output response and provides the best likelihood. ELM is a novel algorithm to fit machine learning problems like classification, clustering, compression, feature learning, regression, and space approximation by adopting the concept of feed-forward neural networks. ELM is adaptive to a single layer or multilayers of hidden neurons, where hidden neurons need not be tuned. To minimize the training error, the minimal norm least square method is used in the original implementation of ELM. It has wide advantages when comparing with other models like the extremely fast learning speed, the default nonadjusting hidden nodes, and homogeneous architectures for different types of machine learning problems. Moreover, ELM follows empirical risk minimization theory which needs only a single iteration for fixing nodes during the learning process which prevents multiple iterations as well as local minimization or approximation (Song et al. 2018; Budamala & Mahindrakar 2020a).

### 3.3.3. Spatial optimization

After fitting the emulator model, the best likelihood parameter set needs to be obtained for the validation of observed data. Here, the UCR basin is a larger area in size and there is a lot of spatial variability over different sub-basins. So, the single station cannot able to provide an approximate result over the entire river basin. It is essential to optimize the multi-station

---

Table 2 | List of influential parameters selected to optimize the SWAT hydrological model

| S. No. | Parameter   | Description                                                                 | Method | Range Minimum | Range Maximum |
|-------|-------------|------------------------------------------------------------------------------|--------|---------------|---------------|
| 1     | CN2         | SCS runoff curve number                                                      | Relative | -0.4          | 0.4           |
| 2     | ESCO        | Evapotranspiration compensation factor                                       | Replace | 0             | 1             |
| 3     | SOL_AWC     | Available water capacity of the soil layer (mm H2O/mm soil)                  | Relative | -0.5          | 0.5           |
| 4     | GW_REVAP    | Groundwater ‘revap’ coefficient                                              | Replace | 0.02          | 0.2           |
| 5     | REVAPMN     | Threshold depth of water in the shallow aquifer for ‘revap’ to occur (mm)   | Absolute | 0             | 500           |
| 6     | GWQMN       | Threshold depth of water in the shallow aquifer required for return flow to occur (mm) | Absolute | 0             | 5,000         |
| 7     | GW_DELAY    | Groundwater delay time (days)                                               | Absolute | 0             | 50            |
| 8     | ALPHA_BF    | Baseflow alpha-factor (1/day)                                               | Replace | 0             | 1             |
| 9     | RCHRG_DP    | Deep aquifer percolation fraction                                            | Absolute | 0             | 1             |
| 10    | CH_K2       | Effective hydraulic conductivity in main channel alluvium (mm/h)              | Absolute | 0.01          | 50            |
| 11    | SFTMP       | Snowfall temperature                                                        | Replace | 0             | 5             |
| 12    | SMTMP       | Snow melt base temperature                                                   | Replace | 0             | 5             |
| 13    | SMFMX       | Maximum melt rate for snow during the year (occurs on summer solstice)       | Replace | 0             | 10            |
| 14    | SMFMN       | Minimum melt rate for snow during the year (occurs on the winter solstice)   | Replace | 0             | 10            |
| 15    | TIMP        | Snow pack temperature lag factor                                             | Replace | 0             | 1             |
| 16    | SURLAG      | Surface runoff lag time                                                      | Replace | 0             | 24            |
for this type of basins or watersheds. Hence, spatial optimization is the best alternative to optimize the basins to restrict the variability over the stations. Spatial optimization may lead to a heavy computational burden and complexity due to multi-station optimization. To overcome this problem, the PCESO algorithm incorporated parallel computing of spatial optimization. Here, the algorithm segregates the stations based on the dependency test and fits the emulator model for each station with the help of parallel computation. In UCR, the first independent stations fit the emulator model, and later, it fits the dependent stations based on the updated response from independent stations.

3.4. Validation measures

The validation measures can help to judge the model predictions. The present study followed different validation measures for each component:

- **Validation of Climate Model Predictions:** CanRCM4 obtained from dynamic downscaling of parent GCM CanESM2 (Scinocca et al. 2016). Additionally, the raw data bias-corrected through Cannon’s MBCn algorithm against the gridMET or Daymet gridded observational datasets (Scinocca et al. 2016). This data validated through performance indicators of NSE and R2. Furthermore, the bias-corrected climate data formatted into SWAT weather files.

- **Validation of SWAT hydrological model predictions:** According to Moriasi et al. (2015), NSE above 0.75 is considered the best model prediction for streamflow optimization. Therefore, NSE > 0.75 is selected as the threshold limit to evaluate the model predictions. If the SWAT model does not meet the threshold limit and it is forwarded to the PCESO algorithm for the enhancement of SWAT model predictions by the emulator concept. Furthermore, the predictions compared with different performance metrics (i.e., Coefficient of Determination (R2), Percentage of Bias (PBIAS)) for identification of model behavior.

- **Validation of PCESO Optimization:** PCESO has two validation sections like the accuracy assessment of emulator model and the effectiveness of optimization predictions. Here, the optimization predictions validated similar to threshold limit followed in the SWAT model predictions (i.e., NSE > 0.75). While the validation of the emulator model needs to go through convergence criteria. Here, the convergence criteria are also called stopping criteria. PCESO followed the AS strategy to restrict the sampling space. Here, the PCESO measures the emulator model accuracy assessment in every iteration by using stopping criteria. By the presence of stopping criteria, it helps to restrict the sample size for exploration and updates for the next iteration. Here, convergence criteria will be stopped on two conditions: (a) minimum feasible value of cross-validation and (b) total number of samples reached (1,600). Furthermore, the detailed characterization of each component is displayed in Table 3.

| Table 3 | Explanation of methods incorporated in the PCESO algorithm |
|---------|------------------------------------------------------------|
| Components | Subcomponents | Methods | Remarks |
| Generation of parameter sets | Influential parameters | Identified 16 streamflow parameters Nash–Sutcliffe Efficiency (NSE) | Represented in Table 2 |
| | Objective function | | NSE = 1 − \[ \frac{\sum_{i=1}^{n} \left( Y_{\text{OBS}}^i - Y_{\text{SIM}}^i \right)^2}{\sum_{i=1}^{n} \left( Y_{\text{OBS}}^i - Y_{\text{MEAN}}^i \right)^2} \], where Y_{\text{OBS}} is streamflow observed values, Y_{\text{SIM}} is simulated or predicted values, Y_{\text{MEAN}} is the mean of observed values, n is the number of data points (or months) |
| Sampling | Initial | Quasi Random Design (QRD) | Initial Sample size was not more than 20 times of parameter dimensions (i.e., 20 \[ \times 16 = 320 \]) (Wang et al. 2014; Budamala & Mahindrakar 2021). Therefore, 300 initial samples were selected for building the model. |
| | Adaptive | LOLA-Voronoi | The adaptive sample size should not more than 10% of the total samples. Here, Total samples = 1,600 considered for this study (i.e., parameter dimensions \[ \times 100 \]) Therefore, 100 extra samples for every iteration. |
| Emulator | Extreme Learning Machines (ELM) | Accuracy of the emulator model assessed through 5-fold cross-validation by the minimal value of RMSE. |
| Convergence criteria | NSE > 0.75 | Here, convergence criteria will be stopped on two conditions like (a) Target (NSE > 0.75), (b) Total number of samples reached (1,600). |
3.5. Assessment of BW and GW security

BW and GW security provides information on freshwater resource availability in a basin. Hence, it is essential to show the BW and GW security for freshwater resources accessibility of future needs. In the present work, BW and GW security analyzed using the scarcity and vulnerability indicators. The water scarcity index is termed as the ratio of water consumption to the median (or) 50th percentile of water provision. While the water vulnerability index is the proportion of water consumption to the 30th percentile of water provision (Rodrigues et al. 2014). The vulnerability index contributes to analyzing low flow conditions or drought periods. Furthermore, the BW and GW security based on the above indicators is stated in Equations (1)–(6).

\[ \text{GW - Scarcity}_{(x,t)} = \frac{\text{GW - Footprint}_{(x,t)}}{\text{Median of GW - Provision}_{(x,t)}} \times 100\% \]  
(1)

\[ \text{GW - Vulnerability}_{(x,t)} = \frac{\text{GW - Footprint}_{(x,t)}}{\text{30th percentile of GW - Provision}_{(x,t)}} \times 100\% \]  
(2)

where GW-Footprint represents the potential sum of maximum transpiration and soil water evaporation, and GW-Provision refers to the soil water content at the beginning of the time period. Here, \( x \) refers to the position of sub-basin and \( t \) refers to the time period.

\[ \text{BW - Scarcity}_{(x,t)} = \frac{\text{BW - Footprint}_{(x,t)}}{\text{Median of GW - Provision}_{(x,t)}} \times 100\% \]  
(3)

\[ \text{BW - Vulnerability}_{(x,t)} = \frac{\text{BW - Footprint}_{(x,t)}}{\text{30th percentile of GW - Provision}_{(x,t)}} \times 100\% \]  
(4)

While BW-Footprint provides information on effective water consumption and BW-Provision signifies the consumptive use of water resources. Here, BW-Provision developed based on the Environmental Flow Indicators (EFRs) using the Smakhtin methodology (Smakhtin & Eriyagama 2008; Equation (5)). This methodology explains the EFRs for ecosystem consideration of both baseflow and quick flow in Equation (6) (Smakhtin & Eriyagama 2008).

\[ \text{BW - Provision}_{(x,t)} = Q_{(x,t)} - \text{EFR}_{(x,t)} \]  
(5)

\[ \text{EFR}_{(x,t)} = \text{LFR}_{(x,t)} + \text{HFR}_{(x,t)} \]  
(6)

Here, \( Q_{(x,t)} \) refers to discharge at the position of sub-basin with a particular time period. LFR indicates the low flow rate at \( Q_{90\%} \) and HFR is a high flow rate at \( Q_{10\%} \) for extreme conditions, which is derived from a specific percentage of flow duration curves (FDCs). When HFR indicates more than Average Discharge (\( Q_{\text{mean}} \)) and it is termed as zero.

These indicators describe supply and demand-driven variables of water vulnerability which outline the boundary between vulnerable and secure (Rodrigues et al. 2014). Here, the hotspots that help identify the locations and period facing threats from source variability.

4. RESULTS

In this framework, the validation measures run in each section to show the effectiveness of hybrid model predictions (i.e., validation of bias-corrected climatic variables, validation of the SWAT hydrological model, and validation of emulator model fitting and validation of optimal model results) and, finally, the application of BW/GW security in the UCR basin.

4.1. Evaluation of bias-corrected climate data

The accuracy assessment of bias-corrected climate data evaluated through the performance indicators of NSE and R2. In Table 4, the bias-corrected data proved its efficiency by comparing with observed variables in every station by satisfying the threshold limit (i.e., NSE > 0.75 and R2 > 0.75). Moreover, the validation results of temperature data followed the trend better than the precipitation data in every station. Hence, the CanRCM4 was adopted for the identification of hydrological phenomena to evaluate water security.
4.2. Quality assessment of the optimization tool (PCESO)

The default SWAT physical model of the UCR basin without calibration could not meet the criteria (i.e., NSE >0.75), and it redirected to the calibration approach. Figure 3 compares the minimum feasible value for six different optimized stations over the AS of 300–1,000. The optimization target for all stations reached 1,000 samples and proven the least sampling efficiency while comparing to the total number of samples (i.e., 1,600). In Stations 3 and 4 maintained the same level, while the remaining stations observed a drop during the optimization of UCR. With an initial number of samples at 300, Station 6 recorded the highest value, and Station 4 performed the least value. Whereas the outstanding stations lied in between the range of 0.2–0.5, respectively. Stations 1 and 5 observed the gradual drop of up to 500 samples, whereas Stations 2 and 6 recorded the minimal value at a maximum number of samples 1,000. Where Station 4 observed a minimum feasible value without any variation due to its smaller in the sub-basin area and less impact on the mainstream. While Station 5 is located near Station 4, it observed similar performance at the optimal point, but it has less variation in the initial set-up due to imperviousness. However, Stations 1 and 2 have similar feasibility at an optimal point because it contained near and similar topographical characteristics. In upstream of Station 3 has a reservoir, hence, the model observed a similar trend throughout the development. Especially, Station 6 contained substantial variability due to rapid urbanization and narrow streams which enables the flash floods. Hence, this station contains more complexity at the initial step and it needs a powerful optimization tool to handle. Therefore, it is transparent from Figure 3, the PCESO effectively played a crucial role to control complex systems and providing actionable results. The computational time for PCESO consumed nearly 1.5 h and the entire framework ran nearly 2 h to achieve the water dynamics and accessibility for conditions of historical, near future, mid future, and far future.

4.3. Quality assessment of the optimal SWAT model

The flow signatures describe the relationship between the observed data to predicted data with a range of 1–100%. This plot is also called FDCs and it explains how much discrepancy between the predicted vs. observed values (Kundu et al. 2016). For the representation of immense data points, the FDC plots are better than the hydrograph plots in terms of high, medium, and

| Weather station | Precipitation (Prec.) | Maximum temperature (Max. Temp.) | Minimum temperature (Max. Temp.) |
|-----------------|----------------------|---------------------------------|---------------------------------|
|                 | NSE                  | R²                              | NSE                             | R²                              | NSE                             | R²                              |
| 1               | 0.76                 | 0.78                            | 0.78                            | 0.82                            | 0.79                            | 0.88                            |
| 2               | 0.75                 | 0.75                            | 0.82                            | 0.81                            | 0.88                            | 0.87                            |
| 3               | 0.79                 | 0.76                            | 0.86                            | 0.82                            | 0.9                             | 0.89                            |
| 4               | 0.81                 | 0.89                            | 0.79                            | 0.84                            | 0.79                            | 0.81                            |

Figure 3 | Minimum feasible value of six different stations for every adaptive sampling through the PCESO optimization tool.
low flows, because the FDC characterizes the whole data points into percentages (i.e., 0–100). While the hydrograph plot displays the entire period in a single graph which leads to heavy noise and complicates during interpretation. The FDC plot is subdivided into different signatures like high flows (i.e., from \(Q_0\) to \(Q_{20}\)), medium flows (i.e., from \(Q_{20}\) to \(Q_{75}\)), and low flows (i.e., \(Q_{75}\) to \(Q_{100}\)). Here, the FDC analyzed for three different conditions as Training, Testing, and Validation. Where the training and testing developed in the optimization tool to improve the system behavior. Besides, the validation dataset helps in finding the scenario with respect to the optimized SWAT model. In Figure 4(a) and 4(b), both training and testing modules for six different stations captured the observed data effectively. While the testing set has some inconsistency in Stations 3 and 6. This is due to Station 3 that contains the reservoir operations upstream and Station 6 contains the effective impervious area. The assessment of the scenarios is displayed in Figure 5, it was evident that the RCP4.5 scenarios for the near, mid, and far future following a similar trend compared with the Historical period. However, the RCP8.5 scenario showed the extremes conditions over the period. Hence, the RCP4.5 scenario can provide an effective hydrological analysis for UCR until the near-future period. Additionally, Table 5 shows the performance measures for the optimized hydrological model. Here, the performance indicators for six different stations satisfied the threshold limit and it directed to the assessment of freshwater security indicators via BW/GW.

![Figure 4](image-url)  
**Figure 4** | Quality assessment of streamflow predictions for six different stations through the FDC plot. (a) Training dataset and (b) Testing dataset.

![Figure 5](image-url)  
**Figure 5** | Comparison of streamflow predictions for historical and RCP scenarios of near future, mid future, and far future. Here, historical (1951–2005), near future (2006–2040), mid future (2041–2070), and far future (2071–2100).
4.4. Assessment of freshwater security

Scarcity and vulnerability are two major indicators to provide information on water security. These indicators enable the preparedness and information of available water resources with varying spatial and temporal timescales for future climatic conditions. The assessment of BW and GW offers the relation of freshwater availability, thereby facilitating the hotspots of water scarcity and vulnerability shows the basin flow regimes for critical conditions over the period. In UCR, the GW of the whole basin in different climatic conditions (i.e., historical, near future, mid future, and far future) is in a critical spot for both median and low-level conditions. This is due to low soil moisture content, rapid urbanization, heavy unconsolidated rock debris, and hilly regions at the UCR basin. Figure 6 represents GWS and GWV, which showed the entire UCR basin of GW security is in a critical condition.

Similarly, BW-Scarcity (BWS) and BW-Vulnerability (BWV) are defined by the consumptive use of water resources to the available BW-Provision at the median and low-level conditions, respectively. In UCR, no aquifer has contributed to water

| Station | Performance indicators | R² | PBIAS |
|---------|------------------------|----|-------|
| 1       | 0.87                   | 0.89 | −1.96 |
| 2       | 0.79                   | 0.8  | +2.56 |
| 3       | 0.81                   | 0.81 | +7.71 |
| 4       | 0.76                   | 0.80 | −8.98 |
| 5       | 0.77                   | 0.81 | −6.98 |
| 6       | 0.79                   | 0.76 | +10.58 |

Figure 6 | Assessment of Green Water security through scarcity and vulnerability indicators for different climatic scenarios. The green color represents critical zones and the blue color indicates a safe zone or green zone. While the percentage of critical zones mentioned in the left-hand corner of every image. Here, near future (2006–2040), mid future (2041–2070), and far future (2071–2100). *Note: the hotspots indicated that the water footprint is greater than the water provisional. Please refer to the online version of this paper to see this figure in color: https://doi.org/10.2166/hydro.2021.182.
withdrawals, hence the surface water is the main source of potable water (Black 2014). Therefore, the surface water is considered for the assessment of BW in this study. Overall, the hotspots laid in between 30 and 44% of total sub-basins in the UCR basin for four different climatic conditions with two different representative concentration pathways of the historical to future extreme conditions (Figure 7). Historical conditions are considered as the benchmark for comparing the number of hotspots to other climatic conditions. In both water security indicators of RCP4.5 fluctuated and it observed the peak at mid future with a maximum number of hotspots (Figure 7). While RCP8.5, the scarcity indicator dropped in the mid future and followed the same number of hotspots respective to the historical scenario. But, the vulnerability indicator of RCP8.5 gradually increased in the near future, and later, it observed a considerable decrease (Figure 7). Here, RCP8.5 showed more hotspots than RCP4.5 in the water security for UCR because RCP8.5 considers the maximum possibilities for extreme conditions. In Figure 7, the hotspots correlated heavily with the effective impervious area and less with forest land. Therefore, water resources will be scarce in future periods, and hence, necessary water policies should bring in the developed and developing impervious areas.

5. DISCUSSION AND CONCLUSIONS

The main target of this research is to provide an effective framework for application-oriented hydrological issues. This framework developed with less manual inspection and a more automated process in extraction, development, enhancement, and assessment of hydrological problems. The entire structure developed in the R programming platform, which can enable to handle all types of users. The automated code assists in bias correction, optimization, and assessment of the hydrological system in a single platform. The entire system optimized every section and automatically directed or redirected to another component. For optimizing the hydrological system, Parallel Computing of Emulator Modelling-based Optimization incorporated in the framework by considering the hybrid concepts of integrating the physical and data-driven models. This calibration approach effectively restricted the complexity and computational burden during enhancement, while the adaptive

---

**Figure 7** | Assessment of Blue Water security through scarcity and vulnerability indicators for different climatic scenarios. The red color represents critical zones and the blue color indicates a safe zone or green zone. While the percentage of critical zones mentioned in the left-hand corner of every image. Here, near future (2006–2040), mid future (2041–2070), and far future (2071–2100). *Note: the hotspots indicated that the water footprint is greater than the water provisional. Please refer to the online version of this paper to see this figure in color: https://doi.org/10.2166/hydro.2021.182.
strategy supports controlling expensive sampling in each stream gauge station. Parallel computing aided to optimize the hydrological system spatially with the consideration of independent and dependent stations separately.

In the application section, the freshwater security specified the hotspots for future water accessibility using the scarcity and vulnerability indicators in UCR. It is transparent that the GW of the entire watershed is in the hotspot zone over the future climatic conditions of both scarcity and vulnerability indicators due to the formation of heavy weathered rock debris and less soil moisture content in the basin. Hence, UCR has difficult conditions for yielding agricultural activities (Black 2014). The water security indicators of BW ranged between 30 and 40% of total sub-basins in UCR. The hotspots of BW directly correlated with an effective impervious area. Hence, the UCR basin can face a water shortage problem due to rapidly changing the impervious area. Noticeably, the RCP4.5 scenario following the present circumstances, and RCP8.5 showed extreme conditions. Therefore, RCP4.5 can provide an effective hydrological analysis for UCR until the near-future period ends. Consequently, the required water policy for surface water should bring in hotspot zones concerning the RCP4.5 scenario to control future water scarcity.

The water dynamics of UCR are explained through the traditional concepts and it approximated by the recent advanced technology to handle the complex HC models. This study developed an effective framework that assists academicians, planners, managers, policymakers, scientists, and stakeholders to identify the threats related to hydroclimatology. Moreover, this framework helpful for real-time analysis and can be focused on future research works.

ACKNOWLEDGEMENTS
The authors are grateful to the editor, the three anonymous reviewers for their valuable comments and suggestions. Our sincere thanks to Dr Parthiban L. for providing valuable suggestions and the Centre for Disaster Mitigation and Management (CDMM), VIT Vellore, for providing the lab resources.

CONFLICT OF INTEREST
No conflict of interest.

DATA AND CODE AVAILABILITY
The source code for this framework available in GitHub Repository: https://github.com/venky5194/R-codes. The meteorological variables obtained from the National Oceanic and Atmospheric Administration (NOAA) and Climate Forecast System Reanalysis (CFSR). While DEM acquired from the National Hydrography Dataset Plus (NHD-Plus), LULC from the National Land Cover Database (NLCD), and soil data from the National Cooperative Soil Survey and supersedes the State Soil Geographic (STATSGO). Finally, the streamflow data attained from the United States Geological Survey (USGS).

DATA AVAILABILITY STATEMENT
All relevant data are included in the paper or its Supplementary Information.

REFERENCES
Abbaspour, K. C., Rouholahnejad, E., Vaghefi, S., Srinivasan, R., Yang, H. & Kløve, B. 2015 A continental-scale hydrology and water quality model for Europe: calibration and uncertainty of a high-resolution large-scale SWAT model. J. Hydrol. 524, 733–752.
Ahmadi, M., Arabi, M., Ascough, J. C., Fontane, D. G. & Engel, B. A. 2014 Toward improved calibration of watershed models: multistage multiobjective measures of information. Environ. Modell. Softw. 59, 135–145.
Ahn, S. R. & Kim, S. J. 2018 Analysis of water balance by surface–groundwater interaction using the SWAT model for the Han River basin, South Korea. Paddy Water Environ. 16, 543–560.
Athira, P., Sudheer, K. P., Cibin, R. & Chaubey, I. 2016 Predictions in ungauged basins: an approach for regionalization of hydrological models considering the probability distribution of model parameters. Stochastic Environ. Res. Risk Assess. 30, 1131–1149.
Ben Alaya, M. A., Zwiers, F. & Zhang, X. 2019 Evaluation and comparison of CanRCM4 and CRCM5 to estimate probable maximum precipitation over North America. J. Hydrometeorol. 20, 2069–2089.
Black, V. 2014 Water Resources Management Plan for Metropolitan North Georgia Metro Water District 2014.
Budamala, V. & Mahindrakar, A. B. 2020a Approximation of Metro Water District Basin using parallel computing of emulator based spatial optimization (PCESO). Water Resour. Manage. 34, 121–137.
Budamala, V. & Mahindrakar, A. B. 2020b Flexible user interface for machine learning techniques to enhance the complex geospatial hydroclimatic models with future perspective. Geocarto Int., 1–20.
Budamala, V. & Mahindrakar, A. 2020c Integration of adaptive emulators and sensitivity analysis for enhancement of complex hydrological models. *Environ. Process.*

Budamala, V. & Mahindrakar, A. 2021 Enhance the prediction of complex hydrological models by pseudo-simulators. *Geocarto Int.* 36, 1027–1045.

Dile, Y. T., Daggupati, P., George, C., Srinivasan, R. & Arnold, J. 2016 Introducing a new open source GIS user interface for the SWAT model. *Environ. Modell. Softw.* 85, 129–138.

Du, L., Rajib, A. & Merwade, V. 2018 Large scale spatially explicit modeling of Blue and Green water dynamics in a temperate mid-latitude basin. *J. Hydrol.* 562, 84–102.

Forrester, A. I. J., Sóbester, A. & Keane, A. J. 2008 Engineering Design via Surrogate Modelling. *J. Hydrol.* 562, 84–102.

Freas, K., Bailey, B., Munévar, A. & Butler, S. 2008 Incorporating climate change in water planning. *J. Am. Water Work. Assoc.* 100, 92–99.

Gassman, P. W., Reyes, M. R., Green, C. H. & Arnold, J. G. 2007 The soil and water assessment tool: historical development, applications, and future research directions. *Am. Soc. Agric. Biol. Eng.* 50, 1211–1250.

Giacomoni, M. H. & Berglund, E. Z. 2015 Complex adaptive modeling framework for evaluating adaptive demand management for urban water resources sustainability. *J. Water Resour. Plan. Manage.* 141, 04013024.

Goharian, E., Burian, S. J., Lillywhite, J. & Hile, R. 2017 Vulnerability assessment to support integrated water resources management of metropolitan water supply systems. *J. Water Resour. Plann. Manage.* 143, 1–12.

Gong, W., Duan, Q. Y., Li, J. D., Wang, C., Di, Z. H., Ye, A. Z., Miao, C. Y. & Dai, Y. J. 2016 An intercomparison of sampling methods for uncertainty quantification of environmental dynamic models. *J. Environ. Inf.* 28, 11–24.

Kundu, D., van Ogtrop, F. F. & Vervoort, R. W. 2016 Identifying model consistency through stepwise calibration to capture streamflow variability. *Environ. Modell. Softw.* 84, 1–17.

Luo, J. & Lu, W. 2014 Comparison of surrogate models with different methods in groundwater remediation process. *J. Earth Syst. Sci.* 123, 1579–1589.

Luo, M., Meng, F., Liu, T., Duan, Y., Frankl, A., Kurban, A. & de Maeyer, P. 2017 Multi-model ensemble approaches to assessment of effects of local climate change on water resources of the Hotan River basin in Xinjiang, China. *Water (Switzerland)* 9.

Luo, M., Liu, T., Meng, F., Duan, Y., Bao, A., Xing, W., Feng, X., De Maeyer, P. & Frankl, A. 2019 Identifying climate change impacts on water resources in Xinjiang, China. *Sci. Total Environ.* 676, 613–626.

Moriasi, D. N., Gitau, M. W., Pai, N. & Daggupati, P. 2015 Hydrologic and water quality models: performance measures and evaluation criteria. *Trans. ASABE* 58, 1763–1785.

Osei, M. A., Amekudzi, L. K., Wemegah, D. D., Preko, K., Gyawu, E. S. & Obiri-Danso, K. 2018 Hydro-climatic modelling of an ungauged basin in Kumasi, Ghana. *Hydrol. Earth Syst. Sci.* 22, 3071–3090.

Rodrigues, D. B., Gupta, V. H. & Eduardo, M. M. 2014 A Blue/Green water-based accounting framework for assessment of water security. *Water Resour. Res.* 50, 7187–7205.

Rouholahnejad, E., Abbaspour, K. C., Vejdani, M., Srinivasan, R., Schulin, R. & Lehmann, A. 2012 A parallelization framework for calibration of hydrological models. *Environ. Modell. Softw.* 31, 28–36.

Scinocca, J. F., Kharin, V. V., Jiao, Y., Qian, M. W., Lazare, M., Solheim, L., Flato, G. M., Biner, S., Desgagne, M. & Dugas, B. 2016 Coordinated global and regional climate modeling. *J. Clim.* 29, 17–35.

Smakhtin, V. U. & Eriyagama, N. 2008 Developing a software package for global desktop assessment of environmental flows. *Environ. Modell. Softw.* 23, 1396–1406.

Song, J., Yang, Y., Wu, J., Wu, J., Sun, X. & Lin, J. 2018 Adaptive surrogate model based multiobjective optimization for coastal aquifer management. *J. Hydrol.* 562, 84–102.

Wang, C., Duan, Q., Gong, W., Ye, A., Di, Z. & Miao, C. 2014 An evaluation of adaptive surrogate modeling based optimization with two benchmark problems. *Environ. Modell. Softw.* 60, 167–179.

Wang, C., Duan, Q., Tong, C. H., Di, Z. & Gong, W. 2016a A GUI platform for uncertainty quantification of complex dynamical models. *Environ. Modell. Softw.* 76, 1–12.

Wang, G., Mang, S., Cai, H., Liu, S., Zhang, Z., Wang, L. & Innes, J. L. 2016b Integrated watershed management: evolution, development and emerging trends. *J. For. Res.* 27, 967–994.

Zhang, X., Srinivasan, R. & Bosch, D. 2009 Calibration and uncertainty analysis of the SWAT model using Genetic Algorithms and Bayesian Model Averaging. *J. Hydrol.* 374, 307–317.

Zhang, J., Chowdhury, S. & Messac, A. 2012 An adaptive hybrid surrogate model. *Struct. Multidiscip. Optim.* 46, 223–238.

Zhang, J., Li, Y., Huang, G., Chen, X. & Bao, A. 2016 Assessment of parameter uncertainty in hydrological model using a Markov-Chain-Monte-Carlo-based multilevel-factorial-analysis method. *J. Hydrol.* 538, 471–486.

First received 16 December 2020; accepted in revised form 7 June 2021. Available online 15 July 2021.