A robust skill verification of hindcast decadal data on hydroclimatic systems.

File Number : PDF/2023/000206

Submitted By : Dr. Venkatesh Budamala
[SERB Qualified Unique Identification Document: SQUID-1994-VB-8935]

Submission Date : 01-Aug-2023

PROPOSAL DETAILS

(PDF/2023/000206)

Principal Investigator	Mentor & Host Institution
Dr. Venkatesh Budamala	**Rajarshi DasBhowmik**
venky5194@gmail.com	rajarshidb@iisc.ac.in
Posdoctoral Researcher(Interdisciplinary Centre for Water Research)	AssistantProfessor(ICWaR)
Contact No : +919700586813	**Indian Institute of Science** Cv raman rd, bengaluru, Bangalore urban district, Karnataka-560012
Date of Birth : 05-Jan-1994	**Contact No.** : +919874819773
Name of Father/Spouse : B Muneendra Babu	**Registrar Email** : registrar@iisc.ac.in
	No. of PHD Scholars : 4
	No. Post-Doctoral Fellow : 2

Details of Post Doctorate

Ph.D. (HYDROCLIMATOLOGY)[Degree Awarded on : 27-May-2022]

ADAPTIVE FRAMEWORK TO ENHANCE THE HYDRO-CLIMATIC SYSTEM BY INTEGRATION OF PHYSICAL AND MACHINE LEARNING CONCEPTS

Research Supervisor/Guide & Institution :

Prof. Amit B. Mahindrakar

Vellore Institute of Technology, Vellore

Brief details of Thesis work :

Although climate change is understood as a gradual variation at large spatial and long temporal scales, real impacts of climate change are experienced on short-space time scales, such as extreme rainfall and floods. Some effects are also shared on large-space time scales, including droughts, hurricanes, heat waves, cold waves, excessive snowfall, and winds. Global climate change is understood and modelled using global climate models (GCMs). The outputs of these models regarding hydrological variables are only available on coarse or large spatial and time scales. Still, finer spatial and temporal resolutions are needed to assess the hydro-environmental impacts of climate change reliably. On the other hand, hydrological models derive the hydroclimatic variables for present and future scenarios based on the climate and topographical data. Further, the physical hydrological models show the typical methods of a basin with refined assessment. But, the physical models may not obtain the optimal results at the initial setup due to the incorporation of heavy datasets, and it is necessary to optimize the physical model. This study proposed a hybrid framework for downscaling climate variables from coarser to finer scale and optimising the hydrological model through an adaptive machine learning algorithm (Adaptive Emulator Modelling based Genetic Optimization (AEMGO)). Here, the AEMGO algorithm follows the adaptive strategy with adaptive sampling, emulator fit, adaptive tuning, convergence criteria and spatial optimization components. For validation of the framework, four
diverse application-oriented issues in four different study areas of India and the USA were selected, i. e., 1) assessment of water processes, 2) assessment of water security, 3) temporal downscaling of precipitation from daily to hour and, 4) temporal downscaling of streamflow for flash flood ungauged watersheds. This framework delivers a comprehensive methodology for developing the hydroclimatic system and analyzing the application-oriented issues.

Technical Details :

Research Area : Earth & Atmospheric Sciences (Earth & Atmospheric Sciences)

Project Summary :

Streamflow forecasting with multiple lead times assists several sectors, such as irrigation water allocations, flood mitigation efforts, water-use policies, and reservoir release, with short-term planning and operational decisions. Similarly, long-term simulation of streamflow, where general circulation model (GCM) projections/ simulations are considered as forcing, is useful to mitigate the impact of long-term climate change caused by anthropogenic activities. Although it is likely that a common hydrologic model is considered for streamflow forecasting and simulation, the forcing of the hydrologic model and the final objective of the modelling eventually separate the overall framework of forecasting and simulations. Former studies have reported that streamflow can be skillfully predicted in advance of a few days when numerical weather model output is forced in hydrologic models. However, the forecasting skills decrease as the lead time increases due to the increasing uncertainty in numerical weather models. To overcome the challenge, former studies have proposed ensemble streamflow prediction where past climate stages are considered as likely inputs for month-ahead streamflow predictions. Further, in the recent past, probabilistic streamflow prediction frameworks have been developed by several former studies as part of sub-seasonal to seasonal (S2S) initiatives. In S2S forecasting, GCM forecasts initialized by the observed Sea Surface Temperature (or persisted SST conditions) are considered input to hydrologic models. Apart from the forecasting initiatives, prior research has attempted long-term simulations of hydrologic fluxes where input and scenario uncertainties are noted as major contributors to the overall performance. Unlike streamflow forecasting, where observed to predicted correspondence is of interest, long-term simulations/projections estimate the change in the mean state of a hydrologic system. Between these two modelling paradigms, the theoretical concepts of decadal hydrologic forecasting existed that would bridge the gap between medium-range forecasting's and long-term simulations. While the inputs of decadal hydrologic forecasting are expected to arrive from climate models, the forecast skill should be verified by the observed to predicted correspondence at an appropriate spatio-temporal scale, assisting the policymakers and water managers in taking near and mid-term planning decisions. The current study adopts an integrated framework of organizing decadal hindcasts, pre-processing hindcasts, hydrologic model calibration and validation to issue decadal streamflow predictions for different major river basins in India. The study's overall objective is to suggest a single verification effort that will serve as a platform for a broader range of model validation and prediction verification operations related to initialized decadal prediction experiments for the Indian subcontinent context.

Objectives :

- To identify the appropriate climate and teleconnection drivers from large scale oceanic and atmospheric circulations.
- To identify the observations and optimal modelling approaches which are required to establish the robust relationship between hydroclimatic dynamics and decadal predictions.
- To assess the skill of hydro-meteorological variables of the river basin system for the decadal timescale by using adaptive AI techniques.
- To detect the impacts of decadal predictions of extreme scenarios on regional systems.

Keywords :

Climate Change Adaptation, Seasonal to Decadal Forecasts, Dry and Wet Extremes, Artificial Intelligence, Machine Learning, and Hybrid Models.

Expected Output and Outcome of the proposal :

Decadal climate hindcasts bridge the gap between extended forecasts and long-term simulation; hence, hindcasts may assist in mitigating the impact of near-term climate change. The current study attempted to understand (i) do the initialization years impact the prediction skill, (ii) at what lead-time decadal streamflow prediction (DSP) can yield the observed statistical attributes. The study applies an integrated framework of general circulation model, bias correction and statistical downscaling, hydrological modelling, and verification metrics over major Indian river basins. The current study also has important implications for near and mid-term planning decisions such as reservoir supply augmentation, additional water source inclusion for integrated water resource management, and a shift in irrigation practices from flood to micro-irrigation. In short, it suggests that issuing skillful DSPs may assist to yield near-term changes in statistical attributes resulting in a combination of natural climate variability and near-term climate change.

Reference Details :

S.No	Reference Details
1	**Dr. Kasiviswanathan KS** **Faculty at Water Resources Development and Management,** **Joint Faculty at Mehta Family School of Data Sciences and Artificial Intelligence** **Indian Institute of Technology, Roorkee.** **[+919952199196]** k.kasiviswanathan@wr.iitr.ac.in
2	**Dr. Amit B. Mahindrakar** **Faculty at School of Civil Engineering,** **Vellore Institute of Technology, Vellore.** **[+919360402842]** amahindrakar@vit.ac.in

PROPOSED WORK METHODOLOGY AND RESEARCH PLAN

1. Introduction of the proposal:

Global Climate Models (GCM) decadal runs (also known as decadal hindcasts or hindcasts) aim to capture the inherent low-frequency climate variability that emerges in conjunction with climate change by utilizing information on the starting state of a climate system, in addition to considering the changes resulting from the observed atmospheric composition (Mehrotra et al., 2014). Notably, decadal hindcast science is relatively new and is regarded as experimental compared to long-term climate change projections/simulations. Decadal hindcasts were introduced in Coupled Model Intercomparison Project Assessment Report Five (CMIP5) and then extended to CMIP6, where GCMs are initialized by the observed SST and are run for 10/30 years. In their seminal work, Hawkins and Sutton (2009) noted that the initial state of the atmospheric-oceanic system and the model structure are the two primary contributors to simulation uncertainty, which yields near-term climate change estimates. Hence, decadal hindcasts are often regarded as the best possible way in which a GCM could 'forecast' climate variables with a lead time of a decade. Nonetheless, it is essential to investigate the accuracy of decadal predictions and attain a comprehensive understanding of the events and processes that enable predictability. Several former studies attempted to evaluate the skill of decadal hindcasts by considering a series of model runs initialized in consecutive years (e.g., Meehl et al., 2009; Smith et al., 2019). However, to compare and validate the accuracy of information across different CMIP experiments, a certain degree of standardization (related to verification metrics, hindcast period, ensemble size, spatiotemporal smoothing, geographic representation, etc.) of hindcast outputs is necessary. Hence, the present proposal delivers a robust skill for the major river basins of India like Godavari, Mahanadi and Cauvery. Here, Godavari River basin selected to show the skill of decadal forecasts for perennial river system, and Mahanadi selected to provide for the non-perennial river system. On the other hand, Cauvery is chosen for identifying the peculiar and unprecedented extreme events in terms of both drought, flood events.

2. Workplan to achieve the proposed objectives:

The robust skill verification of decadal framework is a statistical-dynamic approach on hydro-climatic system that follows three major steps are illustrated in Figure 1 and explained below:

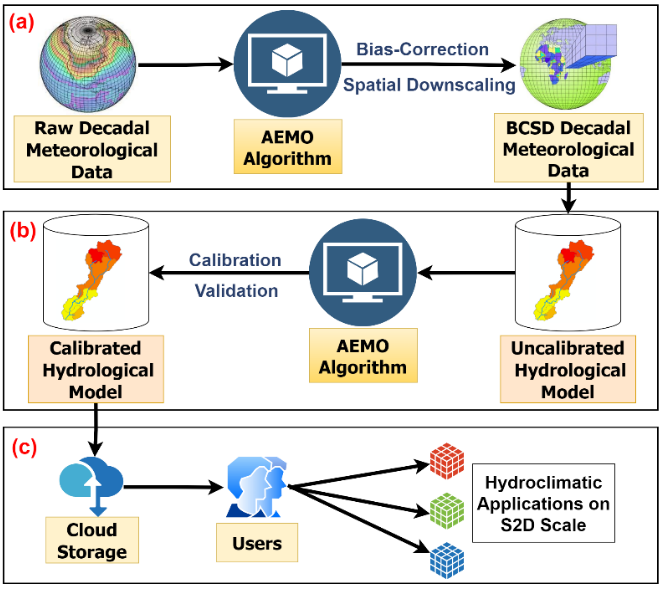

Figure 1. Framework of decadal skill verification on hydro-climatic system.

1.1 Data Preprocessing and Feature Engineering of decadal climate data

The current step would process decadal hindcast meteorological outputs before forcing them to the calibrated hydrologic model. GCM simulations and projections may exhibit substantial bias, which can be as strong as the prediction signal (Das Bhowmik et al., 2017). Additionally, a spatial resolution mismatch exists between the hindcasts and observed meteorological data. Hence, the current study performs a bias correction and statistical downscaling (BCSD) as a pre-processing measure to model forcing. As well as it performs feature engineering for identifying the optimal parameter sets for BCSD approach, to make system for reliable and effective. In this step, MIROC6 can be consider as decadal climate data to verify the skill for seasonal to decadal term. Notably, MIROC6 was collaboratively developed by a Japanese modelling community using a novel ocean-atmosphere coupled model (Tatebe & Watanabe, 2018). Major updates from MIROC5 to MIROC6 include a finer atmospheric vertical resolution with a higher model top and shallow convective parameterization incorporation. Furthermore, compared to MIROC5, MIROC6 has exhibited improvements in the overall reproducibility of mean climate and internal climate variability on intra-seasonal to decadal timescales.

1.2 Development of the optimal hydrological model

The study develops a lumped to semi-distributed physically based hydrological models for major river basins (like Godavari, Mahanadi and Cauvery) in India. In particular, the hydrological model predicts the streamflow at multiple gauge locations in the watershed (includes both gauged and ungauged stations). The physically based hydrological model which can assess the multiple applications of water quantity and quality aspects of a watershed. To note, these models have an enormous number of parameters to replicate the hydrological system. However, due to its immense parameterization, it may not pass the criteria with the default setup (Budamala & Mahindrakar, 2020). Hence, it is necessary to calibrate the model to obtain the optimal setup. For optimization of complex models, the emulator concept can be introduced in the proposal which elaborated in following section.

1.3 Assessment of different hydroclimatic applications

Finally, the results from approximated model stores in cloud, to assist in access for the multiple users. Once the model results are transferred to cloud storage, it will enable various applications in terms of both water quantity and quality aspects. A major part of the proposal will be delivered in this step, by verifying the skill with hydroclimatic variables and delivering the forecasting confidence in the application section. Firstly, it verifies the skill on decadal climate forecasts and later it verifies in the skill on the hydrological models which is induced by decadal data. Ultimately, the users can access the information with the enhanced skill in hydroclimatic system for intra-seasonal to decadal level.

1.4 Structure of Adaptive Emulator Modelling based Optimization (AEMO) Algorithm

AEMO concept is one of the essential steps in this proposal, to address two different problems with a single algorithm (i.e., BCSD of decadal data and Calibration of Hydrological Models). This AEMO strategy helps to optimize the hydro-climatic system. Here, the AEMO involves several steps including adaptive sampling, adaptive modelling, convergence criteria and global optimization. First, the parameters to be optimized are chosen. Sensitivity analysis is often used to screen out the insensitive parameters and identify the parameters that exert the most influence on model performance for further optimization (Budamala & Mahindrakar, 2020). Then an emulator model based on adaptive sampling is constructed to represent the input-output response surface of the simulation model. Finally, an optimization search is conducted on the emulator model.

3. Translational potential and relevance of the proposed work to India

- The proposed work has done on very few studies in Indian context. This would give a more weightage on the applying on Indian study areas for different environmental, hydrological and geophysical models for near to mid-term predictions. Based on Indian context, this proposal offers an innovative approach to forecast for the decadal streamflow for every year. Also, the approach adaptively combines the favourable characteristics of characteristically different emulator modelling methods (a hybrid emulator) would be able to address a broad range of applications (that demand function estimation).
- The major relevance of the proposed work in Indian condition to evaluate the large basins with reliable assessment of hydrological parameters. It helps to analyse from plot to basin level assessment of hydrological behaviour for intra-seasonal to decadal period. Further this framework can be able to provide potential applications of disaster modelling (i.e., flood, drought and unprecedented scenarios etc.,). Finally, it will assist on any sort of hydro-climatic applications in policy developments from Seasonal to Decadal level.

Scope of machine learning: To introduce different/ novel machine learning algorithms as the emulator model, and check for their relative performance in the decadal framework.

Scope of hydrologic sciences: Benchmarking of lumped to semi-distributed physics-based hydrologic models across various Indian basins conditioned on climate factors for decadal scale.

4. Dataset for developing the framework:

Available Datasets		Datasets to acquire
Name	**Period**	**Name**
IMD Meteorological Data (Precipitation, Minimum and Maximum Temperature)	1951 to 2022	Soil Maps
MIROC-6 Decadal data	1979 to 2022	
WRIS Stage and Discharge data	1981 to 2020	Land Use and Land Cover maps of multiple years
Elevation maps	-	

5. Time Schedule of activities giving milestones:

Tasks	1st Year						2nd Year					
	2	4	6	8	10	12	2	4	6	8	10	12
Literature Review												
Data procurement and analysis												
DSP-Emulator setup												
Hydrologic models setup												
Hybrid Algorithm Development												
Computation												
Validation												
Application and Analysis												
Reports												

6. References:

Budamala, V., & Mahindrakar, A. B. (2020). Approximation of Metro Water District Basin Using Parallel Computing of Emulator Based Spatial Optimization (PCESO). *Water Resources Management*, *34*(1). https://doi.org/10.1007/s11269-019-02424-3

Das Bhowmik, R., Sharma, A., & Sankarasubramanian, A. (2017). Reducing Model Structural Uncertainty in Climate Model Projections—A Rank-Based Model Combination Approach. *Journal of Climate*, *30*(24), 10139–10154. https://doi.org/10.1175/JCLI-D-17-0225.1

Meehl, G. A., Goddard, L., Murphy, J., Stouffer, R. J., Boer, G., Danabasoglu, G., Dixon, K., Giorgetta, M. A., Greene, A. M., Hawkins, E. D., Hegerl, G., Karoly, D., Keenlyside, N., Kimoto, M., Kirtman, B., Navarra, A., Pulwarty, R., Smith, D., Stammer, D., & Stockdale, T. (2009). Decadal prediction: Can it be skillful? *Bulletin of the American Meteorological Society*, *90*(10), 1467–1485. https://doi.org/10.1175/2009BAMS2778.1

Mehrotra, R., Sharma, A., Bari, M., Tuteja, N., & Amirthanathan, G. (2014). An assessment of CMIP5 multi-model decadal hindcasts over Australia from a hydrological viewpoint. *Journal of Hydrology*, *519*(PD), 2932–2951. https://doi.org/10.1016/J.JHYDROL.2014.07.053

Smith, D. M., Eade, R., Scaife, A. A., Caron, L. P., Danabasoglu, G., DelSole, T. M., Delworth, T., Doblas-Reyes, F. J., Dunstone, N. J., Hermanson, L., Kharin, V., Kimoto, M., Merryfield, W. J., Mochizuki, T., Müller, W. A., Pohlmann, H., Yeager, S., & Yang, X. (2019). Robust skill of decadal climate predictions. *Npj Climate and Atmospheric Science*, *2*(1). https://doi.org/10.1038/S41612-019-0071-Y

Tatebe, H., & Watanabe, M. (2018). *MIROC MIROC6 model output prepared for CMIP6 CMIP historical*. https://doi.org/https://doi.org/10.22033/ESGF/CMIP6.5603

BIO-DATA

Dr. Venkatesh Budamala

Postdoctoral Researcher
Indian Institute of Science, Bangalore

1. **Name and full correspondence address:**
   Dr. Venkatesh Budamala,
   ICWAR, Indian Institute of Science, CV Raman Road, Bangalore, Karnataka-560012, India.

2. **Email(s) and contact number(s)**
   Primary Email: venky5194@gmail.com,
   Secondary Email: bvenkatesh@iisc.ac.in
   Mobile: +91-9700586813

3. **Institution**
   Indian Institute of Science, Bangalore

4. **Date of Birth**
   5th January 1994

5. **Gender (M/F/T)**
   Male

6. **Category Gen/SC/ST/OBC**
   General

7. **Whether differently abled (Yes/No)**
   No

8. **Academic Qualification (Undergraduate onwards)**

	Degree	**Year**	**Subject**	**University/ Institution**	**% of marks**
1.	B.Tech	2015	Civil Engineering	JNTUA Anthapuramu	79.4
2.	M.Tech (By Research)	2019	Water Resources Engineering	Vellore Institute of Technology, Vellore	8.75/10
3.	PhD	2022	Water Resources Engineering	Vellore Institute of Technology, Vellore	9.5/10

9. Ph.D thesis title, Guide's Name, Institute/Organization/University, Year of Award.

Title: Adaptive framework to enhance the hydro-climatic system by integration of physical and machine learning concepts
Guide Name: Prof. Amit B. Mahindrakar
Institute: Vellore Institute of Technology, Vellore
Year of Award: 2022.

10. Work experience (in chronological order).

S.No.	Positions held	Name of the Institute	From	To	Pay Scale (Per Month)
1.	Teaching cum Research Assistant	Vellore Institute of Technology, Vellore	03-10-2016	15-05-2019	4000
2.	Teaching cum Research Assistant	Vellore Institute of Technology, Vellore	01-02-2020	25-02-2022	20000
3.	Senior Research Fellow	Indian Institute of Technology, Roorkee	01-04-2022	31-05-2022	35000
4.	Research Associate	Indian Institute of Technology, Roorkee	08-06-2022	07-09-2022	45000
5.	Postdoctoral Researcher	Indian Institute of Science, Bangalore	08-09-2022	Present	47000

11. Professional Recognition/ Award/ Prize/ Certificate, Fellowship received by the applicant.

S.No	Name of Award	Awarding Agency	Year
1.	Institute of Eminence (IOE) Fellowship	Indian Institute of Science, Bangalore	2022
2.	Raman Research Award	Vellore Institute of Technology, Vellore	2020,2021
3.	Highest Grade in Training Programme	GIAN, IIT Madras	2016
4.	Best Innovative Idea	Sri Padmavathi University	2015

12. **Publications** ***(List of papers published in SCI Journals, in year wise descending order).***

S.No.	Author(s)	Title	Name of Journal	Volume	Page	Year
1.	**Venkatesh Budamala** and Amit Baburao Mahindrakar	Flexible user interface for machine learning techniques to enhance the complex geospatial hydro-climatic models with future perspective	Geocarto International	37 (12)	3496-3488	2022
2.	**Venkatesh Budamala** and Amit Baburao Mahindrakar	Adaptive hybrid architecture for enhancement of the complex hydroclimatic system and assessment of freshwater security	Journal of Hydroinformatics	23 (5)	950-965	2021
3.	**Venkatesh Budamala** and Amit Baburao Mahindrakar	Enhance the prediction of complex hydrological models by pseudo-simulators	Geocarto International	36 (9)	1027-1043	2021
4.	**Venkatesh Budamala** and Amit Baburao Mahindrakar	Integration of adaptive emulators and sensitivity analysis for enhancement of complex hydrological models	Environmental Processes	7 (4)	1235-1253	2020
5.	**Venkatesh Budamala** and Amit Baburao Mahindrakar	Approximation of Metro Water District Basin using parallel computing of emulator based spatial optimization (PCESO)	Water Resources Management	34	121-137	2020

13. **Detail of patents**

NIL

14. **Books/Reports/Chapters/General articles etc.**

S. No	**Title**	**Author's Name**	**Publisher**	**Year of Publication**
1.	Temporal Downscaling of Daily to Minute Interval Precipitation by Emulator Modeling-Based Genetic Optimization	**Venkatesh Budamala**, Abhinav Wadhwa, Amit B. Mahindrakar, B. Srimuruganandam	CRC Press, Taylor and Francis	2022

19PHD0281

Sl. No.: 134577

VIT

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

The Board of Management of the
Vellore Institute of Technology (VIT)
hereby confers on

B. VENKATESH

the degree of

DOCTOR OF PHILOSOPHY

in recognition of the research work entitled

ADAPTIVE FRAMEWORK TO ENHANCE THE HYDRO-CLIMATIC SYSTEM BY INTEGRATION OF PHYSICAL AND MACHINE LEARNING CONCEPTS

for having fulfilled the prescribed requirements.

The degree has been awarded in compliance with the
"University Grants Commission, Regulations 2009"

given this day the 27 May 2022

Given under the seal of this university

R B Kodali

Vice - Chancellor

Chancellor

Dated: **28-Jul-2022**

Vellore - 632 014, Tamil Nadu, India

ಭಾರತೀಯ ವಿಜ್ಞಾನ ಸಂಸ್ಥೆ
ಬೆಂಗಳೂರು 560012, (ಭಾರತ)
भारतीय विज्ञान संस्थान
बैंगलोर 560 012, (भारत)
Indian Institute of Science
Bangalore 560 012, (India)
Telefax: 080-23600757, Ph: 22932440/444
Email: registrar@iisc.ac.in

Regr(ICWaR)/SERB NPDF Proposal-59/2023 25th July 2023

Endorsement Certificate from the Mentor & Host Institute

This is to certify that:

I. The applicant, Dr. Venkatesh B, will assume full responsibility for implementing the project.
II. The fellowship will start from the date on which the fellow joins University/Institute where he/she implements the fellowship. The mentor willsend the joining report to the SERB. SERB will release the funds on receipt of the joining report.
III. The applicant, if selected as SERB-N PDF, will be governed by the rules and regulations of the University/ Institute and will be under administrative control of the University/ Institute for the duration of the Fellowship.
IV. The grant-in-aid by the Science & Engineering Research Board (SERB) will be used to meet the expenditure on the project and for the period for which theproject has been sanctioned as indicated in the sanction letter/ order.
V. No administrative or other liability will be attached to the Science & Engineering Research Board (SERB) at the end of the Fellowship.
VI. The University/ Institute will provide basic infrastructure and other required facilities to the fellow for undertaking the research objectives.
VII. The University/ Institute will take into its books all assets received under this sanction and its disposal would be at the discretion of Science & Engineering Research Board (SERB).
VIII. University/ Institute assume to undertake the financial and other management responsibilities of the project.
IX. The University/ Institute shall settle the financial accounts to the SERB as per the prescribed guidelines within three months from the date of termination of the Fellowship.

Signature of the Mentor:

Name & Designation: Dr. Rajarshi Das Bhowmik, Assistant Professor, Interdisciplinary Centre for Water Research (ICWaR), Indian Institute of Science

Capt Sridhar Warrier (Retd) (Jul 25, 2023 15:07 GMT+5.5)

Dated: 25.07.2023
Seal of the Institution

Indian Institute of Science
Bangalore-560 012

Signature of the Registrar of University

ಕ್ಯಾಪ್ಟನ್ ಶ್ರೀಧರ್ ವಾರಿಯರ್ (ನಿವೃತ್ತ)/कप्तान श्रीधर वारियर (सेवानिवृत्त)
Capt. Sridhar Warrier (Retd.)
ಕುಲಸಚಿವರು/कुलसचिव/REGISTRAR
ಭಾರತೀಯ ವಿಜ್ಞಾನ ಸಂಸ್ಥೆ/भारतीय विज्ञान संस्थान/INDIAN INSTITUTE OF SCIENCE
ಬೆಂಗಳೂರು/बेंगलूरु/BENGALURU - 560 012

Undertaking by the Fellow

I, **Venkatesh B**, Son/~~Daughter/Wife~~ of Shri. **B Muneendra Babu**, resident of **India** have been awarded SERB N-PDF . I accept the award and undertake that:

1. I shall abide by the rules and regulations of SERB during the entire tenure of the fellowship.
2. I shall also abide by the rules, discipline of the institution where I will be implementing my fellowship
3. I shall devote full time to research work during the tenure of the fellowship
4. I shall prepare the progress report at the end of each year and communicate the same to SERB through the mentor
5. I shall send two copies of the consolidated progress report at the end of the fellowship period.
6. I further state that I shall have no claim whatsoever for regular/permanent absorption on expiry of the fellowship.

Date: 01 August, 2023

Signature

Brief CV of the Mentor

1. Name : Dr. Rajarshi Das Bhowmik

2. Affiliation : Assistant Professor, Department of Civil Engineering, Indian Institute of Science Bangalore.

3. Date of Birth : 24 Feb 1986

3. E-mail : rajarshidb@iisc.ac.in

4. Phone : 9874819773, 08022933224

5. Gender : Male

6. Academic Qualification (Undergraduate Onwards)

S.No.	Degree	Year	Subject	University/Institution	GPA/% of marks
1.	Ph.D.	2016	Civil Engineering	NCSU (USA)	3.8/4
2.	M.Tech	2012	Civil Engineering	IIT Kanpur	9.7/10
3.	B.E.	2009	Civil Engineering	IIEST Shibpur	72%

7. Ph. D. thesis title, Guide's Name, Institute/Organization/University, Year of Award

"Reducing model and downscaling uncertainties in CMIP5". Guide: Prof. Sankar Arumugam, North Carolina State University, 2016.

8. Work Experience

S.No.	Positions held	Name of the Institute	From	To	Pay Scale
1	Assistant Professor	IISc	2020	Present	Level 12, Stage 1
2	DST INSPIRE Faculty	IISc	2018	2020	Gross pay of INR 1,27,000
3	Postdoctoral Fellow	Hong Kong UST	2017	2018	NA
4	Research Assistant	NCSU	2012	2017	NA

9. Professional Recognition/ Award/ Prize/ Certificate, Fellowship received by the applicant

S.No.	Name of Award	Awarding Agency	Year
1	DST INSPIRE Award	DST	2018
2	Academic Excellence Award	IIT Kanpur	2012
3	International Travel Award	SERB, DST	2019

10. Mentoring of Students

- **PhD Students-3**
  - Sai Vikas Kona
  - Poornima Chandraleka
  - Shairik Sengupta
- **Post-doctoral Researchers-2**
  - Dr. Venkatesh Budamala
  - Dr. Tabasum Rasool

11. Publications

S.No.	Author(s)	Title	Name of Journal	Volume	Page	Year
1.	Basu, B., **Bhowmik, R. D.,** & Sankarasubramanian, A.	Changing Seasonality of Annual Maximum Floods over the Conterminous US: Potential Drivers and Regional Synthesis	Journal of Hydrologic Engineering	-	-	2022
2.	Joseph, R., **Bhowmik, R. D.,** & Mujumdar, PP. (2022**)**	Reconstruction of Urban Rainfall Measurements to Estimate the Spatio-Temporal Variability of Extreme Rainfall	Journal of Water	-	-	2022
3.	**Bhowmik, R. D.,** NG, T. L., & Wang, J. P.	Understanding the impact of observation data uncertainty on probabilistic streamflow forecasts using a dynamic hierarchical model	Water Resources Research	*56*(4)	NA	2020
4.	**Bhowmik, R. D**., & Sankarasubramanian, A	A Performance-based Multimodel Combination Approach to Reduce Uncertainty in Seasonal Temperature Change Projections	International Journal of Climatology	41	NA	2020
5.	**Bhowmik, R. D.,** Seo, S., Das, P., & Sankarasubramanian, A.	Synthesis of irrigation water supply use in the United States: spatio-temporal patterns	Journal of Water Resources Planning & Management	146	NA	2020
6.	**Bhowmik, R. D.,** Suchetana, B., & Li, M.	Shower effect of a rainfall onset on the heat accumulated during a preceding dry spell	Scientific Reports	9	2011	2019
7.	**Bhowmik, R. D.,** & Sankarasubramanian, A.	Limitations of Univariate Statistical Downscaling to	International Journal of Climatology	1-18	NA	2019

		Preserve Cross-Correlation between monthly precipitation and temperature				
8.	Seo, S., **Bhowmik, R. D.,** Sankarasubramanian, A., Mahinthakumar, G., & Kumar, M.	The role of cross-correlation between precipitation and temperature on basin-scale simulation of hydrologic variables.	Journal of Hydrology	570	304-314	2019
9.	**Bhowmik, R. D.,** Seo, S., & Sahoo, S.	Streamflow Simulation Using Bayesian Regression with Multivariate Linear Spline to Estimate Future Changes	Journal of Water	10(7)	875	2018
10.	**Bhowmik, R. D.,** Sharma, A., & Sankarasubramanian, A.	Reducing Model structural uncertainty in climate model projections—a rank-based model combination approach	Journal of Climate	30(24)	10139-10154	2017
11.	Sankarasubramanian, A., Sabo, J. L., Larson, K. L., Seo, S. B., Sinha, T., **Bhowmik, R D.,** … & Kominoski, J.	Synthesis of public water supply use in the United States: Spatio-temporal patterns and socio-economic controls	Earth's future	5(7)	771-788	2017
12.	**Bhowmik, R. D.,** Sankarasubramanian, A., Sinha, T., Patskoski, J., Mahinthakumar, G., & Kunkel, K. E.	Multivariate downscaling approach preserving cross-correlations across climate variables for projecting hydrologic fluxes	Journal of Hydrometeorology	18(8)	2187-2205	2017

12. Book Chapter:

1.	**Bhowmik, R. D.,** & Roy, T. (2022)	Challenges and Solution Pathways in Water Use Through the Lens of COVID-19	Global Pandemic and Human Security: Technology and Development Perspective	-	211–222	2022

Approximation of Metro Water District Basin Using Parallel Computing of Emulator Based Spatial Optimization (PCESO)

Venkatesh Budamala & Amit Baburao Mahindrakar

Water Resources Management
An International Journal - Published for the European Water Resources Association (EWRA)

ISSN 0920-4741
Volume 34
Number 1

Water Resour Manage (2020)
34:121-137
DOI 10.1007/s11269-019-02424-3

Springer

Water Resources Management (2020) 34:121–137
https://doi.org/10.1007/s11269-019-02424-3

Check for updates

Approximation of Metro Water District Basin Using Parallel Computing of Emulator Based Spatial Optimization (PCESO)

Venkatesh Budamala[1,2] · Amit Baburao Mahindrakar[1]

Received: 16 April 2019 / Accepted: 12 November 2019 /
Published online: 14 December 2019

Abstract
Metro Water District (MWD) is an agency that administers water distribution in a large geographic region. It targets for existing conditions with future projections of water resources for conservation, supply, and usage. Hence, it is required to show proper water resources management for MWD. Where the river basin profiles are projected to provide the water resources management with potential issues for MWD. Here, Upper Chattahoochee River (UCR) basin of the Metropolitan North Georgia Water Planning District (MNGWPD) selected for the study area. UCR is one of the largest river basins in the MNGWPD and it provides drinking and primary receiving water for nearly 3.5 million people of Atlanta Metro Region. In this study, Parallel Computing of Emulator based Spatial Optimization (PCESO) framework developed for spatial optimization of large complex watersheds. The proposed framework optimizes the hydrological model by parallel computing, emulator fit, sampling design, and spatial optimization. The results showed that 1) the computational time required for spatial optimization was significantly reduced by 50%, 2) goodness-of-fit reached its threshold limit in all stations inclusive in reservoir containing stations, 3) the water balance components and the optimized parameter values with sensitivity index provided the physical phenomena of the study area and showed the approximate hydrological processes in MWD. Further, this proposed work incorporates into future climate data can provide an accurate hydrological analysis with water allocation issues like water use, demand, conservation, and supply for MWD and it helps to identify the water-related disasters floods and droughts.

Keywords Spatial optimization · Hydrological process · Multi-site watershed calibration · Emulators · Parallel computing

✉ Amit Baburao Mahindrakar
amahindrakar@vit.ac.in

Extended author information available on the last page of the article

1 Introduction

Metro water district (MWD) serves to provide and protect the water resources with healthy and hygienic conditions for future needs and it employs for water conservation, water supply, wastewater management, and watershed management of urban municipalities. In MWD, watershed management is playing a key role to focus on sustainable development of resources and process which incorporates watershed functions that affect ecology and environment. As new cities have been created, additional levels of coordination should be implemented to ensure proper watershed management across each basin. Hence, it is necessary to develop an accurate watershed model which represents the real-world phenomena (Black 2017). Many of the watershed systems are highly complex and it may not understand in detail, but abstraction is essential to study or control some shades of behavior (Pecci et al. 2019). A watershed model provides valuable information on all associated parameters present in a catchment and replicates the real-world system. Due to enormous number of parameters, the spatial variability can be more in large-scale watershed models (Femeena et al. 2018). Moreover, the optimization of a single site in large-scale watersheds cannot explain an actual phenomenon and it displays the spatial variability.

Spatial optimization of a watershed can provide efficient and effective approach for dealing the spatial variability of multi-sites (Cibin and Chaubey 2015). During the spatial optimization of large-scale watersheds, most of the stations are correlated with each other because of dependency. Hence, it is essential to calibrate the independent stations first and follows by dependent stations. Optimization of each and individual stations separately can consume more time, where the parallelization process can minimize the computational burden (Rouholahnejad et al. 2012). Even in parallelization, the computational burden depends on the optimization algorithm. Furthermore, the optimization algorithm should be satisfying the following conditions like 1) it should mimic the original simulation model, 2) it should minimize the complexity of fitting and, 3) it should restrict the computational burden. One approach to approximate and replicate the expensive real-world simulation model is Emulator Modelling based Optimization (EMO) (Budamala and Mahindrakar 2019). It is extensively used for two purposes like 1) statistical fitting for prediction of model output, 2) for simplifying the original simulation model. Moreover, emulators are cheap models to represent the input-output response through real world simulator (Queipo et al. 2005; Razavi et al. 2012; Wang et al. 2014).

EMO developed for different optimization problems in the field of aerodynamics, metallurgy, geology, economics, electronics and water resources. Queipo et al. 2005; Forrester and Keane 2009 explained the selection criteria for different sampling techniques, emulator models, and convergence criteria. Zhang et al. 2009 implemented emulator model for optimizing the hydrological model and showed the correlation between folds and performance. For various water resources problems, Razavi et al. 2012 reviewed the articles for sustainable usage of emulators with justified solutions. Wang et al. 2014 explained effect of the initial design and sequential sampling with sensitivity indices. While Haftka et al. 2016 used the parallel computing for global optimization of emulator models and provided the parallelization concepts. Rehbach et al. 2018 compared the emulator models with parallelization process for electrostatic precipitators. The above studies concluded that the parallelization with EMO able to improve the model accuracy and reduce the computational time significantly. The main objective of this paper is to present a new approach for spatial optimization of the watershed model using parallel computing with EMO for effective watershed management.

In this research, the Parallel Computing of Emulator based Spatial Optimization (PCESO) algorithm is developed in MATLAB Environment to provide effective and efficient spatial watershed model which replicates the real-world system with less computational burden. PCESO algorithm consists of spatial optimization, parallel computing, initial design, fitting of emulator, and sequential sampling. This algorithm can divide the dependent and independent stations separately and runs the parallel computing based on the sequence. While each station fits the emulator model with initial design and checks the criteria if it will not meet the criteria and again it adds few more samples until it reaches stopping criteria. Moreover. this novel algorithm is applied to the Soil and Water Assessment Tool (SWAT) watershed model for MWD basin to show capable of watershed management. The objectives of study are 1) To develop an efficient spatial optimization framework for large-scale metro watershed model, 2) To show the efficacy of PCESO algorithm for spatial variability and computational burden, 3) Analysis of the long term hydrological components for MWD basin.

2 Materials and Methods

2.1 Study Area

Upper Chattahoochee River (UCR) basin located in Georgia state, USA, its headwaters in the Blue Ridge Mountains and flowing towards the southwest to the Chattahoochee river with Peachtree Creek (Fig. 1a). The UCR is one of the largest river basins covering 18% of the Metropolitan North Georgia Water Planning District (Metro Water District) with an area of 4291.87 km^2. The main tributaries of UCR consisting Chestatee River, Wahoo Creek, Suwanee Creek, Big Creek, Sope Creek, Rottenwood Creek and Peachtree Creek with 2 dams, Buford Dam and Morgan Falls Dam. Buford Dam constructed for Lanier Lake, which controlled the river flow with a drainage area of 150 km^2. Lake Lanier is a multipurpose reservoir which serves as flood protection, power production, water supply, navigation, recreation, fish and wildlife management. The southern stretch towards the Lanier Lake is primarily a suburban area that covers Forsyth, Gwinnett, and North Fulton. Perimeter center and Cobb Galleria are the areas that are highly developed and employed. UCR considered as low-land catchment due to its elevation ranges from mean sea level of 135.3 m to 19.4 m (Fig. 1c). Land Use Land Cover LULC) of UCR is dominated by impervious area of 46% and it is equal to sum of agriculture and forest cover percentage (Fig. 1d). The imperviousness of the watershed has a direct impact on the hydrological process such as increase in the amount and rate of peak storm flow that affects the strength and reliability of stream, quality of water, aquatic and biotic community. The impervious and pervious surface gives an entirely different response on the watershed and thus it is important to clearly understand the hydrological process for the study of watershed management. Groundwater availability is limited due to geologic conditions, which restrict the potential yield for water supply. UCR is spread over various types of soils which mainly are, Cecil-Madison-Pacolet, Madison-Davidson-Pacolet, Riverview-Chewacla-Cartecay and the "urban" soils in the North Fulton County (Fig. 1e). As the name suggests (Chattahoochee is a Muskogee word which means rock-marked, where Chato- rock and huchi- marked), the river basin is covered with rolling hills and isolated mountains consisting of crystalline rocks in the north extending up to the Fall Line. These rocks are deposited with weathered and unconsolidated rock debris making space for the aquifers that cover the crystalline rocks in the north. These superimposed deposits are thicker

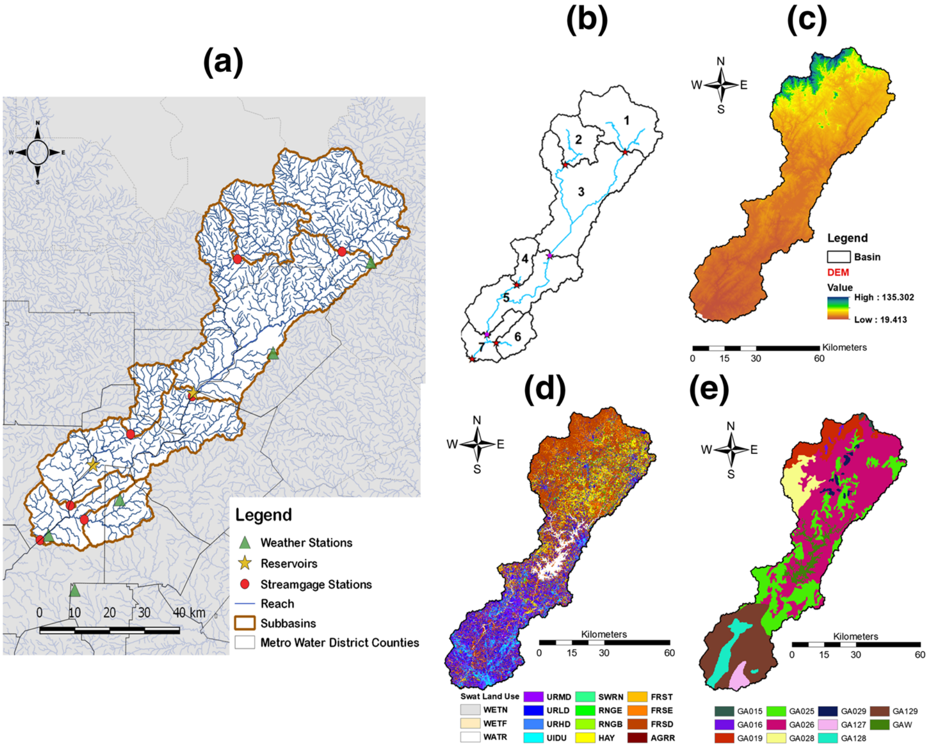

Fig. 1 **a** Map of Upper Chattahoochee River Basin with Metro Water District in Georgia state, USA. **b** Delineated sub-basins with streamgage stations (**c**) Digital Elevation Map (DEM) (**d**) Land Use and Land Cover (LULC) map and, (**e**) Soil classification map. *Note: 1, 2, 4 and 6 are considered as independent stations and 3, 5 and 7 are represented as dependent stations. Where, station 7 does not available any dataset. So, we considered station 7 as ungauged station. Hence, stations 1–6 considered for calibration

in the valleys but do not provide enough yield other than satisfying the very low-density residential which makes surface water as the primary source of potable water in the MWD (Black 2017). Based on 59 years of USGS record, the annual flow of UCR ranges from a minimum of 14.725 m^3 / sec to a maximum of 178.113 m^3 / sec, and a mean flow of 67.395 m^3 / sec. The average annual rainfall remained between 1346.2 mm to 1524 mm per year from the southwestern to the northeastern part respectively. Table 1 gives the entire data used for this study with data sources.

2.2 Steps for Integration of SWAT Conceptual Model and PCESO Algorithm

1) Select the SWAT hydrological model and apply the physical characteristics of the watershed. Execute the SWAT model with meteorological data for prediction of hydrological components.
2) Evaluate the SWAT model with respective of observed data. If the model does not meet its criteria, it should go through calibration or optimization using PCESO algorithm otherwise it will reach to approximate output for analyzing real-world phenomena.
3) During the calibration process, identify the influential parameters and objective function for watershed significance. Initially, PCESO algorithm segregates according to

dependency test. This algorithm optimizes the independent stations through Emulator Modelling based Optimization (EMO) by parallel processing and later, it follows by dependent stations.

4) In EMO, generate the initial parameter sets for respective design space for each station. Develop the emulator fitting using designed parameter sets as input. Once fitting each station according to dependency, the algorithm verifies through convergence criteria. If criteria do not approach at a satisfactory level, the design space adds few more samples until the model achieves the stopping criteria.

Figure 2 shows the framework of hybrid modelling for SWAT conceptual hydrological model and PCESO calibration algorithm for approximating watershed management. Step 1 shows the development of a conceptual hydrological model with necessary datasets to predict the hydrological components. Step 2 explains how conceptual model predicts the model output and moreover it may not get accurate results at initial setup of conceptual model because it involves huge datasets, equations, and parameters to capture the particular model output. So, the conceptual model may go through the calibration process. Step 3–4 demonstrates the calibration procedure using PCESO algorithm. This algorithm employed for parallelization concept for Emulator Modelling based Optimization (EMO). EMO links the input-output response with a real-world simulator for replicate the real-world phenomena and it employs on two broad ways like one-shot approach and sequential approach. During One-Shot Approach of EMO, the model fits with whole surface at single stretch but there is no point to calculate the entire surface and it is enough to carry out necessary sets. In Sequential Approach of EMO, the model fits with initial parameter sets and checks the convergence criteria, if the model does not reach the criteria and it should add few more samples using sequential sampling until model reaches the target or stopping criteria. In this framework, we adopted Sequential Approach methodology for calibration of SWAT conceptual hydrological model. Here, the conceptual model (SWAT) provides theoretical and physical knowledge of real-world phenomena and the data-driven models (PCESO) improve the efficiency of the conceptual model. Moreover, the integration of both conceptual and data-driven models shows an accurate theoretical background.

2.3 SWAT Hydrological Model Setup

SWAT model considered a conceptual model for the prediction of hydrological components. SWAT is a physical-based quasi-distributed model, it contains numerous parameters to capture the hydrological phenomena and it is useful to evaluate the assets of water quantity and quality from catchment to continental scales (Neitsch et al. 2009; Zhang et al. 2009). As size of watershed increases, there can be more spatial variability in hydrological phenomena. So, relying on single-site calibration may not provide proper variations throughout the watershed and it is necessary to optimize the model spatially. SWAT model primarily delineated the basin into sub-basins, and it generated the stream networks and monitoring points for whole basin. Further, the sub-basins divided into Hydrological Response Units (HRU's) based on land use, soil and slope characteristics. Meteorological datasets are provided to basin and later SWAT executables are used to simulate the output for time period of 1999–2015. Hence, the model setup into three components for boosting the SWAT-like warm-up period (1999–2000), calibration period (2001–2010) and validation period (2011–2015).

Table 1 List of hydroclimatic variables and topographic data used in developing the SWAT model for UCR

Data Type	Summary			
Variable	Time step	Period of Study	Number of stations	Source
Streamflow	Monthly	2001–2015	6	Georgia Water Data , USGD, USA
Rainfall	Daily	1999–2015	5	National Oceanic and Atmospheric Administration (NOAA), USA
Temperature	Daily	1999–2015	5	
Wind speed	Daily	1999–2015	5	Climate Forecast System Reanalysis (CFSR)
Humidity	Daily	1999–2015	5	
Solar radiation	Daily	1999–2015	5	
Topographic Data	Resolution	Period of acquisition	Source	
Digital elevation model (DEM)	30 m × 30 m	2006	National Hydrography Dataset Plus (NHD-Plus)	
Land-use map	30 m × 30 m	2011	National Land Cover Database (NLCD)	
Soil map	1:250,000	1995	National Cooperative Soil Survey and supersedes the State Soil Geographic (*STATSGO*) dataset	
Sl. No.	Streamflow Station Name		Station ID	
1.	Chattahoochee River Near Cornelia		02331600	
2.	Chestatee River Near Dahlonega		02333500	
3.	Chattahoochee River at Buford Dam		02334430	
4.	Big Creek Near Alpharetta		02335700	
5.	Chattahoochee River at Atlanta		02336000	
6.	Peachtree Creek at Atlanta		02336300	

2.4 PCESO Algorithm Setup

The PCESO algorithm employs on Emulator Modelling based Optimization (EMO) with parallel computing. It consists of major stages like parallelization, initial design, emulator fitting, convergence criteria and sequential sampling (Fig. 2b). The parallel computing can minimize the computational burden and it shows an effective path to reach the solution (Haftka et al. 2016). In PCESO, the algorithm segregates the stations into dependency and fits an emulator model for each station by parallel distribution. Here, the stations of 1,2,4, and 6 are optimized primarily due to independent for the hydrological process, while station 3 is dependent to 1 and 2; station 5 is dependent to 3 and 4 (Fig. 1b). During the parallel computing, first independent stations are optimized and followed by dependent stations to cover entire watershed space and encounter spatial variability. Further, the components of PCESO are explained below:

2.4.1 Initial Design

Initial design has two portions like initial sampling (like adjustable parameters) and objective function. In this study, a Quasi Random sampling (QRS) has selected for uniform initial sampling with lower discrepancy sequence of effective design filling. Quasi-random is a

Monte Carlo method used for simulation and integration as it is a very simple, direct and easy to use (Razavi et al. 2012). Quasi-Monte Carlo prefers quasi-random sequences than random or surrogate-random methods as it does not reproduce the random behavior. It proves to be a deterministic substitute to random or surrogate-random sequences as they are responsible to reduce clumping by correlating the sample points and enhance the uniformity that is stated in terms of discrepancy. Thus, Quasi-Random method is also termed as low-discrepancy. Due to its correlation property, they seem to be less adaptable than the other two sequences like random and non-random and thus Quasi-Random is preferred for integration for real-world simulation and optimization. The accuracy of integrating can be achieved by filtering a set of calculations. As mentioned, initial design comprised of adjustable parameters and objective function. Here, SWAT hydrological model contains a huge number of parameters, and calibration of whole parameters at a time is impossible, so it necessary to identify the influential parameters to model output. Therefore, sensitivity analysis of Sobol Total Effect identified 16 most influencing parameters which represent of basin, sub-basin and HRU level responses for the streamflow variable (Fig. 3).

In Fig. 3, the selected sensitive parameters are displayed based on the hydrological process. The most influencing parameter is the Curve Number (CN2) that depends on the LULC and soil type of the watershed. Increase in CN2 value indicates the increase of surface runoff but reduces base flow, also lesser CN2 indicates storage of water in the soil. Upper Chattahoochee watershed being dominated by impervious surface and it can face over-prediction in runoff which has to be reduced by adjusting the CN2 as to avoid more runoff as well as the storage of water-based on observed data. Larger values of ESCO indicates that more amount of moisture can be absorbed from the bottom layers of soil. CH_K2 depends on the river bed composition. Where Surface runoff time lag (SURLAG) denotes the time of concentration more than one day which means the entire surface runoff does not reach the river on a particular day it occurs. The decrease in SURLAG shows increases the storage of surface runoff in the catchment delaying the process. Soil Water Capacity (SOL_AWC) estimates the field capacity of every soil layer. As the soil water controls the groundwater percolation the infiltration of present soil layer is disturbed by the already available water capacity. The parameters related to the delay in the process are groundwater time delay (GW_DELAY) and baseflow recession constant (ALPHA_BF). ALPHA_BF represents the retention of GW_DELAY measures the suspension in recharging the shallow aquifers. If this delay is increased the recharge of aquifer is slowed down. Finally, the snow content is available in UCR basin and it is necessary to include snow parameters (SFTMP, SMTMP, SMFMX, SMFMN, and TIMP) due to lag and melt. Further, Nash Sutcliffe Efficiency (NSE) considered as objective function, which shows goodness-of-fit to deal for noise and representation of observed and predicted data. According to Wang et al. 2014, the initial design size should be within 20 times of parameter dimensions. Therefore, the volume of initial samples should not be greater than 20 times of parameter dimensions (i.e., $20 * 16 = 320$), and hence, we selected 300 initial samples for model approximation.

2.4.2 Emulator model

Extreme learning machine (ELM) proposed as emulator model, it is a novel learning algorithm for single-hidden layer feedforward neural networks and composed by a simple three-layer structure (input layer, output layer, and hidden layer) and it is useful for regression, clustering, and classification. ELM has three distinct advantages, i.e., the extremely fast learning speed,

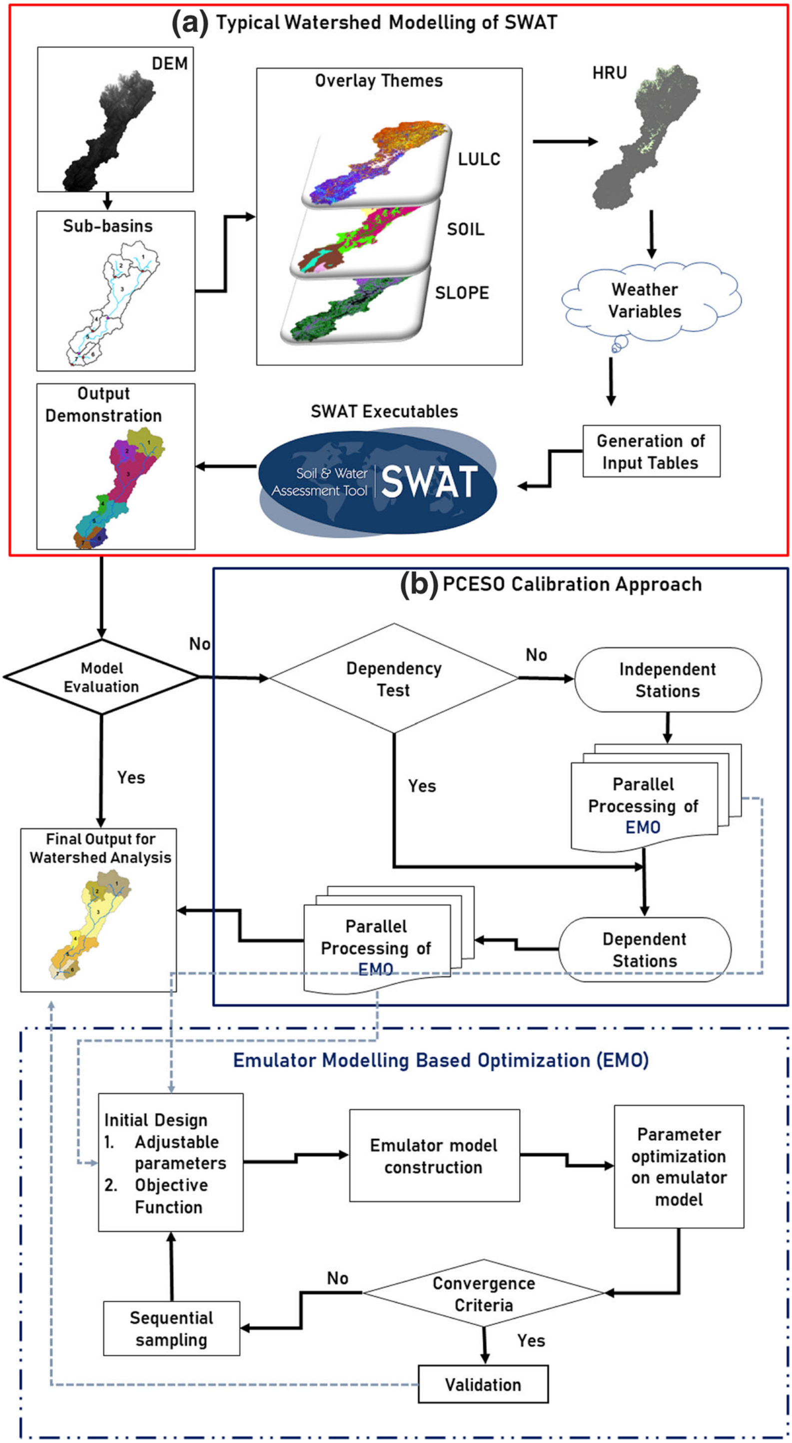

Fig. 2 The integration of SWAT-PCESO framework (**a**) Development of SWAT conceptual watershed model, (**b**) Calibration approach using PCESO algorithm

the non-adjusting hidden node parameters, and homogenous architectures for regression analysis or classification (Song et al. 2018). The minimal norm least square method is used in the original implementation of ELM to minimize the training error. ELM can use white box kernel mapping instead of black-box kernel used in support vector machines. Principal component analysis can be used as special feature where linear nodes are used in ELM. It has capability to deal classification and global approximation. Moreover, ELM is based on empirical risk minimization theory which needs only single iteration for learning process, and it prevents multiple iterations and local minimization.

2.4.3 Convergence criteria

Once fitting the emulator model with updated space, it should go through cross-validation for emulator model accuracy analysis. Here, Cross-Validation (CV) is a statistical method which is basically used for machine learning concepts. One of the most popular methods in CV is K-fold cross-validation, which enhances the holdout method. In CV, the function approximation fits with training data and it predicts the test set for the assessment of the model. The datasets are divided into K equal parts, the K-1 part can go for training and the remaining one part goes for testing. CV can justify both training and testing in available samples. Moreover, the method highly depends on endpoints at training and testing set, while it significantly varies with the division of folds. In this study, the root mean square error (RMSE) is used as a cross-validation score to test the performance of emulator models (Forrester and Keane 2009; Viana et al.

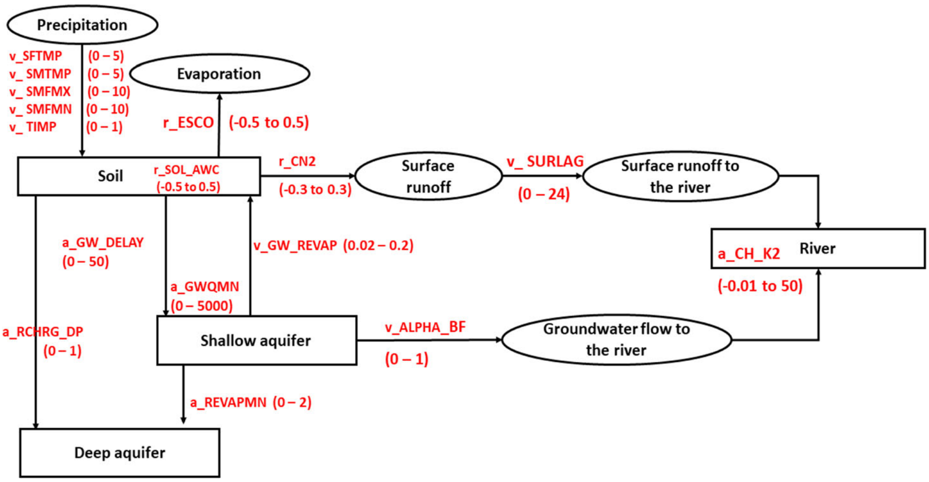

Fig. 3 Representation of influential streamflow parameters with ranges and method of UCR basin through hydrological process. *Note: Prefix of parameters represent method for selection of sample 'a' – absolute, 'r' – relative, and 'v' – replace. Where, **CN2** (SCS runoff curve number), **ESCO** (Soil evaporation compensation factor), **SOL_AWC** (Available water capacity of the soil layer (mm H2O/mm soil)), **GW_REVAP** (Groundwater "revap" coefficient), **REVAPMN** (Threshold depth of water in the shallow aquifer for "revap" to occur (mm)), **GWQMN** (Threshold depth of water in the shallow aquifer required for return flow to occur (mm)), **GW_DELAY** (Groundwater delay time (days)), **ALPHA_BF** (Baseflow alpha factor (1/days)), **RCHRG_DP** (Deep aquifer percolation fraction), **CH_K2** (Effective hydraulic conductivity in main channel alluvium (mm/h)), **SFTMP** (Snowfall temperature), SMTMP (Snow melt base temperature), **SMFMX** (Maximum melt rate for snow during the year (occurs on summer solstice)), **SMFMN** (Minimum melt rate for snow during the year (occurs on the winter solstice)), **TIMP** (Snow pack temperature lag factor), **SURLAG** (Surface runoff lag time)

2009; Wang et al. 2014). RMSE provides a global error measure over the entire design domain (Zhang et al. 2012). If CV score is not satisfactory then it adds few more samples and fits the model iteratively until the model reaches the stopping criteria. Here, the stopping criteria depend on the three classes like 1) When CV score target reaches, 2) When maximum number of sampling limit reaches and 3) when time limit reaches. In this study, we would not consider the time limit because to show computational efficient over other existing methods. Here, CV score target should be less than 5% of max flows for each station (Budamala and Mahindrakar 2019). The maximum number of samples is taken in this study is 1000.

2.4.4 Sequential sampling

Once convergence criteria are not satisfied, sequential sampling will be enabled to fill extra samples for the updated surface. Here, Lola-Voronoi sequential sampling is selected for additional samples. Lola-Voronoi is two different strategies for sample filling as ‘Lola’ and ‘Voronoi’. Voronoi estimates a gradient in each available point and creates a design focused on non-linear regions. It has been widely used in research areas related to electromagnetic compatibility, microwave system macro modeling, and exposure assessment. The idea behind LOLA (Local Linear Approximation) component is to compare the dynamic and smooth regions. The dynamic regions should be densely sampled compared to the smooth regions. Linearity is expressed by a linear fit which is estimated by building a linear approximation for each sample and then compared to the neighboring samples. Bad fit represents that the sample is in the dynamic region and needs exploitation. This method does not consider the size of the region: if the region is big as compared to the other sample area it integrates additional samples giving rise to new non-linear regions that are exploited in the subsequent iteration. The difference between the real output and the observed value in the neighborhood is termed as linearity measure. The linearity measure is provided with the Voronoi exploration that shows the sampling performance of the space surrounded by the reference sample which helps to rank the samples. Hence, the combination of LOLA and Voronoi can provide accurate results for non-linear regions with high dimensional problems. The sequential sampling of LOLA-Voronoi can provide extra 50 samples of every updated surface for enhancing the model output.

3 Results and Discussions

The present study focusses on two perspectives like 1) spatial optimization of large complex watersheds and 2) approximation of MWD basins for analyzing hydrological phenomena. For large complex watersheds contain high computational burden with substantial spatial variability. To encounter this problem, we proposed a framework called PCESO algorithm. To show the excellence of proposed framework with existing method, derived results are constructed like how emulators are performing during fitting of each station (i.e., Emulator Accuracy Analysis), after optimization of parameters in each station how well model predicting the model output (i.e., Validation Analysis), and how much computational burden involved during fitting the model (i.e., Parallel Computing Accuracy Analysis). Here, PCESO compared with Sequential Uncertainty Fitting (SUFI-2) algorithm. SUFI-2 is most popular method for optimizing SWAT model parameters and it encounters uncertainty in model output. So, we compared this framework with regularly using method over different perspectives. Finally, the results are enclosed with analysis of MWD basin for hydrological phenomena (i.e., Optimization Results with Sensitivity and Analysis of Long-Term Water Balance Components.)

3.1 Emulator accuracy analysis

The parallel computing for optimization of UCR based on independent and dependent stations as shown in Fig. 1b. PCESO algorithm used for optimizing the SWAT hydrological model with the help of stopping criteria. This optimization process initially checked with few samples, if it is not satisfying and again added few more samples until model reached its stopping criteria (300 initial samples took for this study and 50 extra samples added at each iteration). The stopping criteria of this study depended on two types like 1) once target reaches (i e., Cross-validation score should be <5% of max flows for each optimized station) or, 2) exceedance of maximum limit of samples (i e., 1000 samples). In Table 2, each station CV score with the target are displayed. Excluding 3 and 5 stations, remaining stations optimized the SWAT hydrological model within initial samples. While 3 and 5 stations took extra samples to optimize the model because the stations located at downstream of reservoirs. Hence, PSECO algorithm optimized UCR basin of SWAT model with a smaller number of samples and more accuracy.

3.2 Validation analysis

Once PSECO optimized the model, the next step is to check the performance of the predicted model output through metric and flow-wise. In this study, two major performance metrics are selected for validating the model namely Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe 1970), and Percentage of Bias (PBIAS) (Gupta et al. 2009; Yen et al. 2014). The NSE is dimensionless goodness of fit and differentiates the length and thickness of the observed and simulated hydrographs, where its optimum value is 1 (Moriasi et al. 2007). Percent bias tends to show whether the prediction errors are low or high. It provides information such as positive bias and negative bias indicating overestimation and underestimation, respectively. These measures were selected to compare and evaluate the model forecasts under different base-lines. For present research work, the 2011–2015 period considered for validating SWAT-PCESO model of UCR watershed. Figure 4a represents the performance measures of PCESO, SWAT-CUP and default SWAT model. According to Moriasi et al. 2007, calibration and validation with above 0.65 in NSE and $\leq\pm$ 15 in PBIAS is considered as good models. Here, PSECO achieved the threshold limit and satisfied the conditions in both cases. But, SUFI-2 not able to provide satisfactory results in most of the cases. While SWAT model without calibration failed in all stations and clearly showed it requires calibration. Positive PBIAS indicates the overprediction and negative PBIAS represents the under prediction. In Fig. 4a, SWAT default model without calibration showed over prediction due to high amount of imperviousness area present in UCR.

Flow Duration Curve's (FDC's) help to show how well the predicted flow captured the observed flow over different signatures. FDC plots contain flow signatures categorized into high (Q0 to Q20), medium (Q20 to Q75), low (Q75 to Q90), and very low (Q90 to Q100) flows (Fig. 4b). Here, PCESO followed and captured the observed data in all flow signatures while remaining methods are not followed. As mentioned in previous, SWAT default model without calibration had over predicted and that inference visually displayed in Fig. 4b. While, all stations captured peaks and base flow accurately in PCESO, but SUFI-2 are not captured accurately in peaks and

Table 2 Cross-validation scores of emulator fitting for each station

Stations											
1		2		3		4		5		6	
Target = 1.875		Target = 1.023		Target = 4.907		Target = 0.424		Target = 6.863		Target = 0.6186	
Samples	RMSE	Samples	RMSE	Samples	RMSE	Samples	RMSE	Samples	RMSE	Samples	RMSE
300	1.372	300	0.766	300	5.096	300	0.230	300	6.925	300	0.364
				350	5.436			350	6.535		
				400	3.650						

there is a shift in prediction. But, PCESO algorithm identified properly in stations 3 and 5 also, even though it contained reservoirs. Hence, PCESO provided the best results comparing with SUFI-2 and SWAT (without calibration) in optimizing UCR basin through both performance metric wise (NSE and PBIAS) and visual wise (by FDC (Flow signatures)).

3.3 Parallel Computing Accuracy Analysis

The main advantage of parallel computing is to save the computational burden. To show the excellence of parallel computing, this study compared to series computing. While SUFI-2 does not contain parallel computing, but it has parallel processing. So, the comparison of computational time was evaluated for Parallel Computing for EMO (PCESO), Series Computing for EMO (SC-EMO) and Series Computing of SUFI-2 (SC-EMO) using Core i5 3.2 GHz processor with 8 GB RAM. Series computing of both EMO and SUFI-2 acquired 280 and 370 min respectively. While PCESO consumed only half of the computational time of series with 160 min approximately. Therefore, Parallel Computing saves 50% of the computational time of Series Computing with more accuracy.

3.4 Optimization results with Sensitivity

Sensitivity analysis is performed in two different stages such as before optimization and after optimization. Before optimization can help to screen out the most influential parameters to calibrate the model. While after optimization can show robustness of model by varying parameter values. In Fig. 5, graph plotted for comparison of normalized optimal parameter value to sensitivity of each parameter after optimization. As observed in previous sections, the SWAT model flow predicted excessively due to presence of high urbanization. The reasons behind the over prediction is high surface runoff and low evapotranspiration. CN2 considered as most influential parameter during optimization because it influences the surface runoff. Abbaspour et al. 2015 recommended three parameters to control overprediction (i. e., decrease of CN2, increase of SOL_AWC and ESCO). To show brevity of PCESO algorithm prediction, we took parameter range of CN2 (−0.3 to 0.3), SOL_AWC (−0.5 to 0.5), and ESCO (−0.5 to 0.5). It clearly displayed that PCESO algorithm efficient in optimization by minimizing CN2 and integrating SOL_AWC, ESCO parameters (Fig. 5). Remaining parameters are also contributed for model performance, but it contained minimal effect. In stations 3 and 5, CN2 and ESCO played major role during optimization and rest could not show any effect in model output. While station 6 controlled by RCHRG_DP parameter, it represents the fraction to percolate from root zone to deep aquifer which affects the evapotranspiration. Hence, PCESO algorithm identified and followed the patterns to achieve an accurate SWAT model.

3.5 Analysis of Long-Term Water Balance Components

The long-term water balance components explain how well the hydrological process is performing in a watershed. Precipitation (PREC), Surface Runoff (SWR), Evapotranspiration (ET) and Water Yield (WYLD) are considered major water balance components to identify water resources management. Here, SWAT default model without calibration shown over prediction of surface runoff with long term water balance components of PREC is 1319.26 mm, SWR is 213.018 mm, ET is 658.25 mm and WYLD is 632.49 mm respectively. While the mean annual results of PCESO approximated model are PREC is 1319.26 mm, SWR is 96.42 mm, ET is 887.31 mm and WYLD is

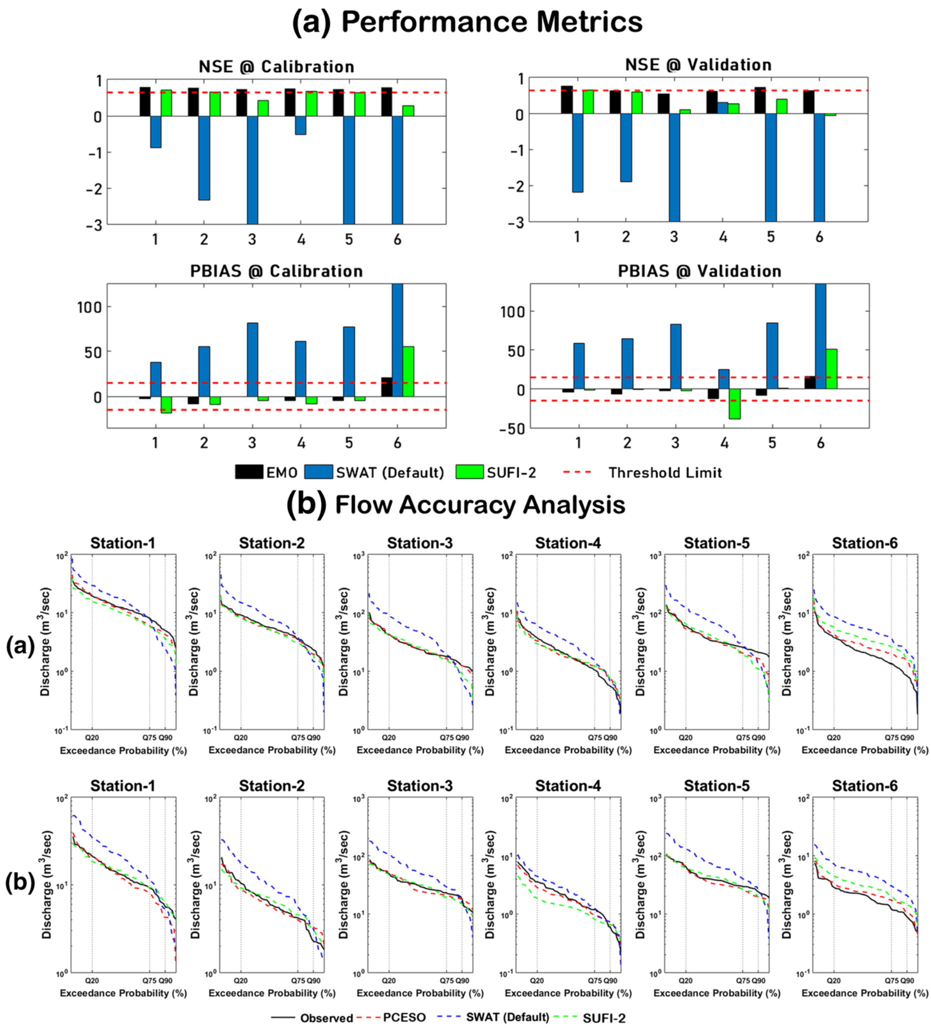

Fig. 4 Comparison of validation analysis for PCESO, SWAT-CUP and SWAT (default model). **a** Performance metrics of NSE and R2, (**b**) Flow Duration Curves

373.29 mm respectively. Hence, approximated watershed model of UCR contains high ET with moderate SWR and exceptional in WYLD of entire watershed which can helpful for sustainable water resources management to MWD.

4 Conclusions

The proposed PCESO framework effectively applied to Upper Chattahoochee River Basin to achieve an accurate watershed analysis with spatial optimization and less computational burden that will address water resources issues of Metro Water District Basin. This algorithm comprised of parallelization, initial sampling, emulator fitting, sequential sampling, and spatial optimization. The parallelization process carried out based on the independent and dependent stations for showing

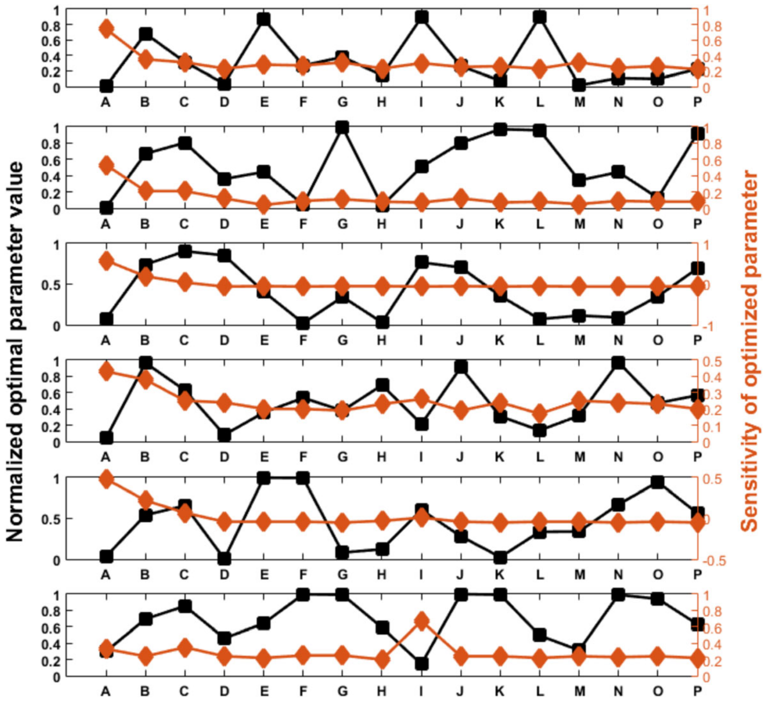

Fig. 5 Inter-comparison of normalized optimized parameter value to sensitivity index for 6 stations. *Note: Where x-axis represents the influential parameters, A: CN2, B: ESCO, C: SOL_AWC, D: GW_REVAP, E: REVAPMN, F: GWQMN, G: GW_DELAY, H: ALPHA_BF, I: RCHRG_DP, J: CH_K2, K: SFTMP, L: SMTMP, M: SMFMX, N: SMFMN, O: TIMP, P: SURLAG

spatial variability. ELM emulator fitted with the minimum discrepancy of QSR initial samples and Lola-Voronoi sequential sampling helped to add extra samples for reservoirs containing stations with less computational time nearly 50% of series computing. While validation results showed its best in performance metrics and flow signatures. Optimized parameter results with sensitive index showed the behavior of watershed, where CN2 (runoff parameter), ESCO (Evapotranspiration parameter) and SOL_AWC (soil water capacity parameter) are the major driving parameters in calibration of UCR and followed by RCHRG_DP (Groundwater parameter) showed the effect in evapotranspiration for Peachtree Creek station covered by Atlanta Metro region. Finally, the long-term water balance components showed the performance of PCESO algorithm related to hydrological processes, here 67% and 28% of precipitation consumed by evapotranspiration and water yield of UCR. Hence, this approximated model incorporates into future climate data can provide an accurate watershed analysis and water allocation issues for metro water regions with less computational burden. Finally, PCESO algorithm is not only limited to optimization of hydrological models but also applicable for different optimization problems.

Compliance with Ethical Standards

Conflict of Interest None.

References

Abbaspour KC, 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. https://doi.org/10.1016/j.jhydrol.2015.03.027

Black V (2017) Water resources management plan. Metropolitan North Georgia Water Planning District, Georgia, USA.

Budamala V, Mahindrakar AB (2019) Enhance the prediction of complex hydrological models by pseudo-simulators. Geocarto Int 0:1–17. https://doi.org/10.1080/10106049.2019.1629646

Cibin R, Chaubey I (2015) A computationally efficient approach for watershed scale spatial optimization. Environ Model Softw 66:1–11. https://doi.org/10.1016/j.envsoft.2014.12.014

Femeena PV, Sudheer KP, Cibin R, Chaubey I (2018) Spatial optimization of cropping pattern for sustainable food and biofuel production with minimal downstream pollution. J Environ Manag 212:198–209. https://doi.org/10.1016/j.jenvman.2018.01.060

Forrester AIJ, Keane AJ (2009) Recent advances in surrogate-based optimization. Prog Aerosp Sci 45:50–79. https://doi.org/10.1016/j.paerosci.2008.11.001

Gupta HV, Kling H, Yilmaz KK, Martinez GF (2009) Decomposition of the mean squared error and NSE performance criteria : implications for improving hydrological modelling. J Hydrol 377:80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Haftka RT, Villanueva D, Chaudhuri A (2016) Parallel surrogate-assisted global optimization with expensive functions – a survey. Struct Multidiscip Optim 54:3–13. https://doi.org/10.1007/s00158-016-1432-3

Moriasi DN, Arnold JG, Van Liew MW et al (2007) Model evaluation guidelines for Sytematic quantification of accuracy in watershed simulations. Trans ASABE 50:885–900

Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models part-1 - a discussion of principles. J Hydrol 10:282–290

Neitsch SL, Arnold JG, Kiniry JR, Williams JR (2009) Soil and Water Assessment Tool User's Manual: Version 2005. Texas Water Resources Institute, TR-192, College Station, Texas.

Pecci F, Abraham E, Stoianov I (2019) Model reduction and outer approximation for optimising the placement of control valves in complex water networks. J Water Resour Plan Manag 145:1–13. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001055

Queipo NV, Haftka RT, Shyy W, Goel T, Vaidyanathan R, Kevin Tucker P (2005) Surrogate-based analysis and optimization. Prog Aerosp Sci 41:1–28. https://doi.org/10.1016/j.paerosci.2005.02.001

Razavi S, Tolson BA, Burn DH (2012) Review of surrogate modeling in water resources. Water Resour Res 48: 1–32. https://doi.org/10.1029/2011WR011527

Rehbach F, Zaefferer M, Stork J, Bartz-Beielstein T (2018) Comparison of parallel surrogate-assisted optimization approaches. In: GECCO 2018 - Proceedings of the 2018 Genetic and Evolutionary Computation Conference. pp 1348–1355. https://doi.org/10.1145/3205455.3205587

Rouholahnejad E, Abbaspour KC, Vejdani M, Srinivasan R, Schulin R, Lehmann A (2012) A parallelization framework for calibration of hydrological models. Environ Model Softw 31:28–36. https://doi.org/10.1016/j.envsoft.2011.12.001

Song J, Yang Y, Wu J et al (2018) Adaptive surrogate model based multiobjective optimization for coastal aquifer managemen. J Hydrol 561:98-111t

Viana FAC, Haftka RT, Steffen V (2009) Multiple surrogates: how cross-validation errors can help us to obtain the best predictor. Struct Multidiscip Optim 39:439–457. https://doi.org/10.1007/s00158-008-0338-0

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 Model Softw 60:167–179. https://doi.org/10.1016/j.envsoft.2014.05.026

Yen H, Wang X, Fontane DG, Harmel RD, Arabi M (2014) Environmental Modelling & Software A framework for propagation of uncertainty contributed by parameterization , input data , model structure , and calibration / validation data in watershed modeling. Environ Model Softw 54:211–221. https://doi.org/10.1016/j.envsoft.2014.01.004

Zhang X, Srinivasan R, Van Liew M (2009) Approximating swat model using artificial neural network and support vector machine. J Am Water Resour Assoc 45:460–474. https://doi.org/10.1111/j.1752-1688.2009.00302.x

Zhang J, Chowdhury S, Messac A (2012) An adaptive hybrid surrogate model. Struct Multidiscip Optim 46:223–238. https://doi.org/10.1007/s00158-012-0764-x

Affiliations

Venkatesh Budamala[1,2] • **Amit Baburao Mahindrakar**[1]

[1] Department of Environmental and Water Resources Engineering, School of Civil Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India

[2] Centre for Disaster Mitigation and Management, Vellore Institute of Technology, Vellore, Tamil Nadu 632014, India

GEOCARTO INTERNATIONAL
2021, VOL. 36, NO. 9, 1027–1043
https://doi.org/10.1080/10106049.2019.1629646

Enhance the prediction of complex hydrological models by pseudo-simulators

Venkatesh Budamala[a,b] and Amit Baburao Mahindrakar[a]

[a]Department of Environmental and Water Resources Engineering, School of Civil Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India; [b]Centre for Disaster Mitigation and Management, Vellore Institute of Technology, Vellore, Tamil Nadu, India

ABSTRACT
Complex hydrological models demand significant computational cost for representing a hydrological system due to its greater number of parameters, the involvement of various datasets, selection of samples, objective functions and algorithms. Pseudo-models are the cheap simulators and it is the best alternative to represent the complex hydrological systems with an input–output response. In the present study, Soil Water Assessment Tool (SWAT) hydrological model is scrutinized by developing pseudo-modeling-based optimization (PMO) method for the agriculturally dominated watershed in India. Results conclude that Gaussian Process regression performed as the best pseudo-model with minimal discrepancy quasi-random sampling of 200 initial design and the prediction of streamflow showed as NSE-0.88 & 0.82 and R^2-0.9 & 0.84 in calibration and validation. The proposed model incorporating into future climate data can provide accurate water-related issues like water use, demand and quality with a less computational burden and moreover, it can identify the water-related disasters like floods and droughts.

ARTICLE HISTORY
Received 18 March 2019
Accepted 20 May 2019

KEYWORDS
Optimization; SWAT; watershed calibration; pseudo-models; adaptive sampling

1. Introduction

Complex conceptual hydrological models like Soil and Water Assessment Tool (SWAT) are widely used in recent years to understand and manage various activities that affect the watershed (Abbaspour et al. 2015). It integrates with Geographical Information System (GIS) to evaluate the quantity and quality of water for different resolutions and ecological conditions across the world (Gassman et al. 2007; Dile et al. 2016). SWAT contains an enormous number of parameters to capture the streamflow in a watershed. The parameters like process-based (curve number (CN2), plant compensation factor (EPCO), groundwater 'Revap' coefficient (GW_REVAP) etc.,) cannot measure directly and it must go through the adjustments between each parameter interval (like Calibration) (Zhang et al. 2009). While the adjustments of parameters for SWAT through manual calibration takes more time even minutes or hours to run one single simulation. Due to the massive computational cost included, it is imperative to progress the effectiveness of calibration for SWAT.

CONTACT Amit Baburao Mahindrakar amahindrakar@vit.ac.in; amahindrakar@gmail.com

Model calibration is a process to change the parameters and achieve the most favorable solution by minimizing the error between the predicted and observed data. The calibration of the watershed can be most expensive time-step in creating a precise SWAT hydrological model (Yang et al. 2008; Noori and Kalin 2016). Moreover, the calibration time (computational burden) depends on the quality of datasets (resolution), the area of the watershed, the model duration and the model calibration algorithms. Different approaches are developed to encounter the computational burden in SWAT-like Sequential Uncertainty Fitting (SUFI-2) (Abbaspour et al. 2007), Generalized Likelihood Uncertainty Estimator (GLUE) (Beven and Binley 1992), Parameter Solution (ParaSol) (Uniyal et al. 2015) and Markov Chain Monte Carlo (MCMC) (Kuczera and Parent 1998) and Particle Swarm Optimization (PSO) (Kennedy and Eberhart 1995) etc. However, it requires tens of thousands runs to get global optimal set and ultimately it leads to high computational CPU time. To deal with these problems, the pseudo-models are the best choice to optimize the complex system with effective simulation time as it mimics the original simulation model and shows the relationship between inputs (adjustable parameters) and outputs (objective function) (Razavi et al. 2012a).

Pseudos' are the cheap models and its principle is to simulate a pseudo-function between input parameters and model outputs (Forrester et al. 2008). For obtaining an accurate pseudo-model, the model runs with different parameter sets of input–output data based on the simulator with a defined parameter range. In traditional studies used one-shot approach (using a parameter sets at a single stretch) to get optimal set by the global optimization algorithm (Razavi et al. 2012a). For finding the optimum solution, only necessary sets can be enough instead of working with the whole surface. Where the adaptive sampling strategy can help to seize or restrict the model runs based on its optimization target. In adaptive pseudo-modelling-based optimization (APMO), the model runs with an initial design and checks the performance, if the performance is not meeting the criterion, then it can further continue with the additional samples iteratively until the model reaches its stopping criteria. Finally, the model simulations end up with the most favourable solution along with the optimal parameter set. Moreover, the adaptive sampling receives additional points which can enhance the accuracy of the pseudo-model with effective control of the high simulations (Wang et al. 2014).

Various studies show developments in the area of pseudo-based optimization and methods explaining the design of experiments (DOE), model selection and validation, sensitivity analysis. (Forrester and Keane 2009) explained about pseudo-modelling-based optimization including sensitivity analysis, adaptive sampling and multi-objective optimization in aerospace designs. (Razavi et al. 2012a) reviewed 48 articles of pseudo-modelling in water resources engineering field with the framework, approaches, limitations and applications. (Wang et al. 2014) used adaptive pseudo-modelling approach for calibration of the SAC-SMA hydrological model to obtain the parameter variations and performances in different pseudo-models. (Gong et al. 2015) developed the adaptive pseudo-model for large complex geophysical models to reduce computational time and increase the optimization effectiveness in multi-objective. While few types of research focussed to pseudo-based optimization using a one-shot approach in SWAT model and demonstrated with different sample sizes, crossvalidation (CV) folds and correlated parameters (Zhang et al. 2009). However, these studies provided valuable information of PMO for SWAT, appropriate methods in each stage of APMO, effects of varying initial design sizes, stopping criteria and control of samples is not explained. Due to these gaps, can lead to high computational burden during calibration. For restricting the computational burden, we concentrated on these gaps to show the effective modelling process.

The purpose of this article is to provide an effective calibration process for computer intensive hydrological models. The proposed approach can improve the performance with a drastic reduction of simulation time and sample size. Further, the following aspects are discussed in this article: (1) effects in model output with change of initial design; (2) appropriate pseudo-model with effective initial and adaptive sampling for SWAT hydrological model; (3) efficacy of adaptive strategy compared to the one-shot approach and SUFI-2 based on performance, number of samples, optimization and computational time; (4) the influence of adjustable parameters and its affect in prediction of flows. This study helps to model users for analysing different characteristics in the watershed with the help of more accurate and less computational burden model. This article is organized as follows; section 2 elaborates the study area and methods, section 3 discusses the results obtained from APMO approach and finally concluded with some remarks of the model in section 4.

2. Methods and materials

2.1. Study area

The Kagna watershed of Krishna river basin located in the state of Telangana, India (Figure 1) has selected as study area. This watershed covers a drainage area of 1854.05 km^2, with the mean altitude of 544.09 m, ranging from 423 m to 725 m. The major part of basin is covered with agricultural land accounting to 69.55% of the total area, 2.94% of the basin is covered by water bodies, 18.11% is covered by forest and 6.5% by fallow and scrub land (Figure 2). In Kagna watershed, built up area covers only 2.9% of the total area. The watershed has a tropical climate, where usually summers have much rainier than winters. The mean annual temperature of Kagna is 27.4 °C; while its

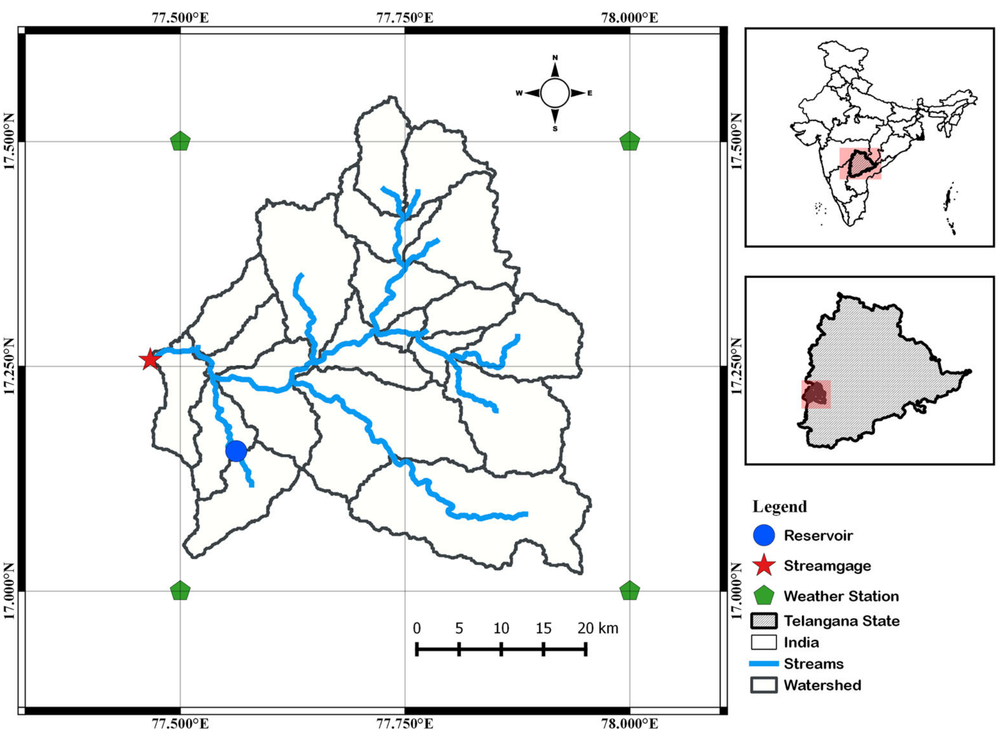

Figure 1. Location of the delineated watershed of Kagna, India.

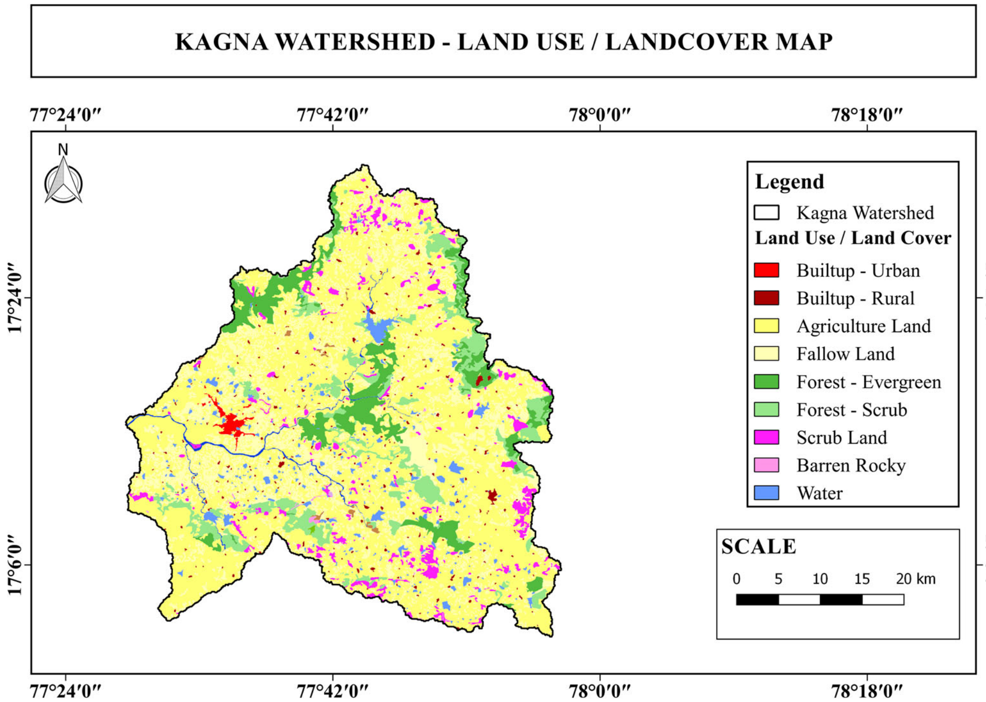

Figure 2. Land Use and Land Cover Classification of Kagna Watershed.

precipitation ranges within 800 mm–1200 mm with an average value of 1078 mm, though 78% of the rainfall occurs in summer due to influence of monsoon. The hydrological regime of Kagna watershed highly influenced by summer monsoon in between the periods of April to September. This summer monsoon is associated with heavy rainfall and humid climate and ultimately results in sudden increase in discharge. Moreover, agriculture relies on the yearly rain. This study area does not have large irrigation systems surrounding lakes and rivers. Aquifers, or supplies of underground water, are shallow. The summer monsoon fills wells and aquifers for the rest of the year.

2.2. Framework

The methodology of the proposed framework is described in Figure 3. The first step of the methodology is to provide model inputs to SWAT like topographical data (Digital Elevation Model (DEM), Land Use and Land Cover (LULC), soil map) and meteorological data (weather variables like precipitation, temperature, relative humidity, wind speed and solar radiation). Further, the model is developed based on given inputs. Here, SWAT is more prevalent due to physically based semi-distributed and continuous time hydrological model to evaluate the characteristics of the watershed by daily, monthly and yearly time step (Dile et al. 2016). It integrates with Geographical Information System (GIS) to evaluate spatial and temporal variations of hydrologic systems. The structure and the process of the SWAT model are further explained below.

2.2.1. Initialization

SWAT delineates the watershed into sub-watersheds based on the elevation data (DEM) or using predefined stream networks. In this study, the watershed delineated into 20

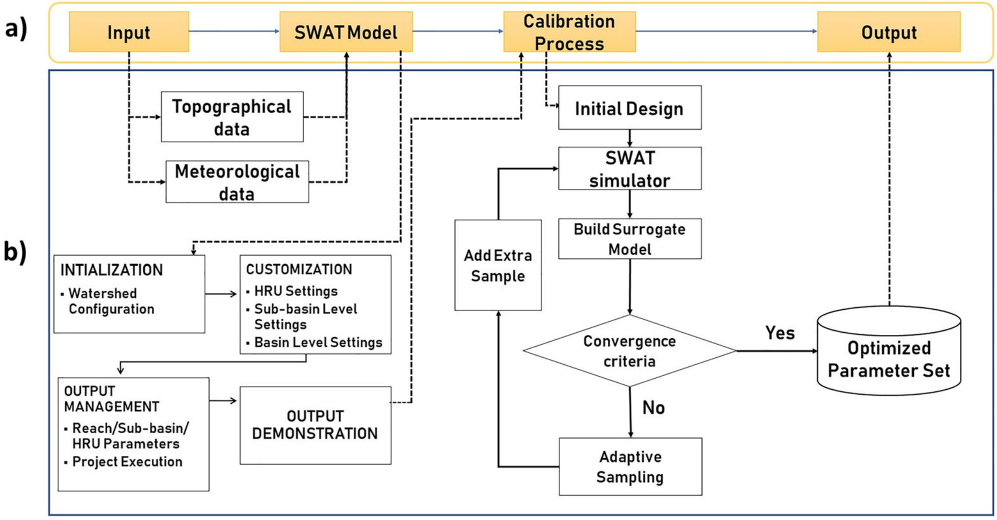

Figure 3. The schematic framework of adaptive pseudo-modelling-based optimization (APMO) for SWAT hydrological model. (a) the general process of calibration to achieve accurate model; (b) in detail representation of each process in every step to achieve an accurate model.

sub-basins with Jewangi outlet station of the catchment (Figure 1). DEM of 30 m resolution obtained from Shuttle Radar Topography Mission (SRTM).

2.2.2. Customization

Basin, sub-basin and hydrological response units (HRU's) level settings are adjusted based on the topographical features. Where, HRU contains the unique combination of land use, soil and slope. Without any adjustment in the model, the default system is considered to capture the actual behaviour. Then, weather data should provide as in daily or sub-daily time step to the SWAT model. Also, there is a possibility to include the weather generators for missing data. The meteorological data obtained from 'Global Weather Data For SWAT' (Fuka et al. 2013; Dile and Srinivasan 2014) of 1987 to 2014.

2.2.3. Output management and demonstration

Once the customization has done, the next step is to write the input tables. According to the given input data and adjustments, the model develops input tables. Finally, the model executes and simulates based on the input tables with specified time-period by daily, monthly or yearly. For evaluating the model performance, considered the available monthly dataset into calibration for 48 months and validation for 36 months in this study.

2.3. Calibration process

After developing the SWAT model, it may or may not obtain acceptable performance due to the handling of different datasets which are incorporated in generating the streamflow. If the model is not in acceptable limits, it must go through the calibration process. The general form of calibration (or optimization) is changing the parameters and achieving its target. A major concern in the process of calibration is expensive simulation time because of recursive runs of the computer artificial models, usually known as computationally intensive optimization. This can set the model for any given problem by objective function or constraint through expensive simulations. Different algorithms have been developed for calibration or optimization in the field of water resources to encounter computational burden (Yen et al. 2014; Zhang et al. 2016), among the most effective and popular approach is pseudo-modelling (Razavi et al. 2012a; Wang et al. 2014). It enables the approximation of the original simulation function based on the input–output response. The framework of PMO consists of input design set, objective function, regression and optimal parameter set. One-shot approach and adaptive strategy are two types of PMO. In the one-shot approach (or traditional approach), the model runs in a single stretch without any future updates. While adaptive pseudo-modelling can run with the initial design and checks the criteria if the criteria are not satisfied it goes to adaptive sampling iteratively until the model reaches its criteria. Different issues are faced while constructing the PMO such as the selection of appropriate methods, initial sample size, convergence criteria and the link between the original model, the pseudo-model and optimizer which are described in detail below.

2.3.1. Initial design

Construction of the initial design involves inputs (adjustable parameters) and outputs (objective function). Das et al. (2013) proposed 26 parameters to predict the streamflow in the SWAT model. These parameter dimensions are heavy, which increases the complexity and computational burden. To reduce this complexity, 10 most influential parameters are screened out for Kagna watershed using sensitivity analysis [Sobol-based variance decomposition (Saltelli et al. 2010)] (Table 1). Then, Nash–Sutcliffe efficiency (NSE) is selected as the objective function. Where NSE is a normalized statistic to determine the difference between the residual variance and observed data (Nash and Sutcliffe 1970). It shows how well the plot of observed versus predicted value fits the 1:1 line. NSE ranges between $-\infty$ and $+1$, wherenegative values indicate unacceptable performance and positive values shows an acceptable level of performance (Table 2). The NSE represents as follows:

$$\text{NSE} = 1 - \left[\frac{\sum_{I=1}^{n} \left(Y_i^{\text{OBS}} - Y_i^{\text{SIM}}\right)^2}{\sum_{I=1}^{n} \left(Y_i^{\text{OBS}} - Y_i^{\text{MEAN}}\right)^2} \right] \tag{1}$$

where, Y^{OBS} = streamflow observed values, Y^{SIM} = simulated or predicted values, Y^{MEAN} = mean of observed values, n = number of data points (or months).

Three different types of sampling methods are selected to build the pseudo-model such as random, quasi-random and Latin-hypercube sampling. Random sampling cannot follow any strategy, it picks up arbitrarily. While quasi-random follows the low-discrepancy sequence, it finds the exact location of sample points based on the sample size, construction and original method (Wang et al. 2014); and Latin-hypercube trails based on the probability density function of each parameter as well as each axis aligned with a hyperplane (Wang et al. 2014). Then, the selection of initial sample size follows a thumb rule of a maximum number of samples/25 $\geq$ initial sample size $\geq$ 20 times of parameter dimensions (Razavi et al. 2012a). Here, a maximum number of samples is 1000 and parameter dimensions is 10. Therefore, input design must lie between 40 (1000/25 = 40) to 200 (10 $\times$ 20 = 200) and further shows a wide gap between the minimum and maximum criteria (between 40 and 200). So, this study shows the effect on the model performance using a different initial sample size of 50, 100 and 200.

Table 1. Selected 10 most influential SWAT model parameters for tuning of a hydrological model using APMO.

				Range	
Sl. no.	Parameter	Description	Method	Min	Max
1	CN2	SCS curve number	Relative	−0.1	0.1
2	EPCO	Plant uptake compensation factor	Replace	0	1
3	SOL_AWC	Available water capacity of the soil layer (mm H_2O/mm soil)	Relative	−0.3	0.3
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	−750	750
6	GWQMN	Threshold depth of water in the shallow aquifer required for return flow to occur (mm)	Absolute	−1000	1000
7	GW_DELAY	Groundwater delay time (days)	Absolute	−30	60
8	ALPHA_BF	Baseflow alpha factor (1/days)	Replace	0	1
9	RCHRG_DP	Deep aquifer percolation fraction	Absolute	−0.05	0.05
10	CH_K2	Effective hydraulic conductivity in main channel alluvium (mm/h)	Absolute	0	50

Table 2. Classification of model performance criteria for model output (Moriasi et al. 2007; Yen et al. 2014).

Categories	NSE	R^2	PBIAS (%)
Very good	$0.75 < \text{NSE} \leq 1.00$	$0.75 < R^2 \leq 1.00$	$\text{PBIAS} \leq \pm 10$
Good	$0.65 < \text{NSE} \leq 0.75$	$0.65 < R^2 \leq 0.75$	$\pm 10 \leq \text{PBIAS} < \pm 15$
Satisfactory	$0.50 < \text{NSE} \leq 0.65$	$0.50 < R^2 \leq 0.65$	$\pm 15 \leq \text{PBIAS} < \pm 25$
Unsatisfactory	$\text{NSE} \leq 0.50$	$R^2 \leq 0.50$	$\text{PBIAS} \geq \pm 25$

2.3.2. SWAT simulator

SWAT simulator is used to determine the objective function based on the available parameter sets. Here, the initial and update parameters can go through the SWAT model and provides the output (streamflow). With reference to the observed data of streamflow, the objective function (NSE) is obtained.

2.3.3. Pseudo-models

Pseudo-model fits with available sample sets and provides the output. The main issue in the selection of an appropriate model function is to calibrate or optimize the parameters. Various models are evolved to analyse and replace the original simulation model, among which the effective models used for optimization of hydrological models are Gaussian Process (GP), Artificial Neural Networks (ANN) and Support Vector Machines (SVM) (Zhang et al. 2009; Razavi et al. 2012b; Wang et al. 2014; Gong et al. 2015) and further explained below. GP regression can easily adapt the additional learning which can enable the pseudo-model to train with the additional sample set. It contains several hyper-parameters to control process and uses the basis function with tuned parameters to mimic the original model. GP can also be used as a statistical interpreter as it can predict and compute the error in the predictor. Support Vector Machines are designed for both classification and regression problems. Sometimes, SVM depends on the concept of the basis function (Gaussian kernel) which is used in Kriging and RBF. Upon Kriging and RBF, SVM has an approximation using the ε-insensitive tube for formation of support vectors. The benefit in using ε-insensitive tube in SVM is that it can ignore the errors within a certain distance of true values during the model fitting and can directly control the sensitivities in noise. The foremost advantage of SVM is to formulate the two parameters, that is, the weight of regularization term and radius of ε-insensitive tube. ANN is becoming more popular for function approximation. The basic structure of ANN comprises of hidden neurons, inputs, outputs and the transfer function. According to the various problems, different methodologies are developed for ANN structures such as pruning or growing strategies, network geometry interpretation and Bayesian approaches. Apart from these methods, the trial and error method is useful for the prediction of a number of neurons for ANN-based pseudo-modelling studies. ANN's consists of inputs which are multiplied by weights and computed by the mathematical function to determine the activation function. While another function computes the output of the artificial neuron. Then, it combines the artificial neurons in order to process the information.

2.3.4. Convergence check

A convergence check is performed on every updated response surface. If the criteria are not meeting the satisfactory level, it can go through the adaptive sampling until unless the model reaches its stopping criteria. Here, the model stops with three criteria such as specified time, the maximum number of samples (1000) and reaches the target (minimum crossvalidation score). To show the performance of the individual model based on the time, we consider an infinite time limit as stopping criteria. In this study, we adopted the crossvalidation score as the convergence criteria and the Lola–Voronoi algorithm (Crombecq 2009) for adaptive sampling. Further, a detailed explanation of the crossvalidation and the Lola–Voronoi (adaptive sampling) is provided below.

2.3.4.1. Crossvalidation. Crossvalidation is a statistical method which is basically used for machine learning concepts. One of the most popular methods in CV is K-fold crossvalidation (KCV), which enhances the holdout method. In CV, the function approximation fits with training data and it predicts the test set for the assessment of model. The datasets are divided into K equal parts, the $K - 1$ part can go for training and the remaining one part goes for testing. CV can justify both training and testing in available samples. Moreover, the method highly depends on endpoints at training and testing set, while it significantly varies with the division of folds. In this study, the root mean square error (RMSE) is used as a crossvalidation score to test the performance of pseudo-models. RMSE provides a global error measure over the entire design domain (Zhang et al. 2012).

2.3.4.2. Lola–Voronoi. Lola–Voronoi is the powerful technique to provide additional samples to optimize the linear and the nonlinear regions. Generally, the adaptive sampling is divided into two types such as exploitation and exploration-based methods. Where exploitation-based methods aim at the sample regions of input space especially at steep ridges and nonlinear behaviour. While exploration-based methods are following sparse sampling and used in linear and nonlinear regions. The Lola–Voronoi method is the combination of both exploitation and exploration-based algorithm, this method is more flexible and robustness in linear and nonlinear regions for estimating a slope in each and every updating point. This algorithm helps in reducing the heavy computational burden and allows a desirable number of samples to obtain a certain model accuracy. The LOLA algorithm is developed based on fuzzy logic. LOLA helps to compare the surroundings of the other samples, while Voronoi identifies and adds samples in the specified region. This method potentially helps to determine the additional samples with the desired target output for both linear and nonlinear regions (Crombecq 2009; Van Der Herten et al. 2014).

3. Results and discussions

The results and discussions broadly divided into two phases like 'Phase–1 (for model selection) and Phase–2 (for model stability)'.

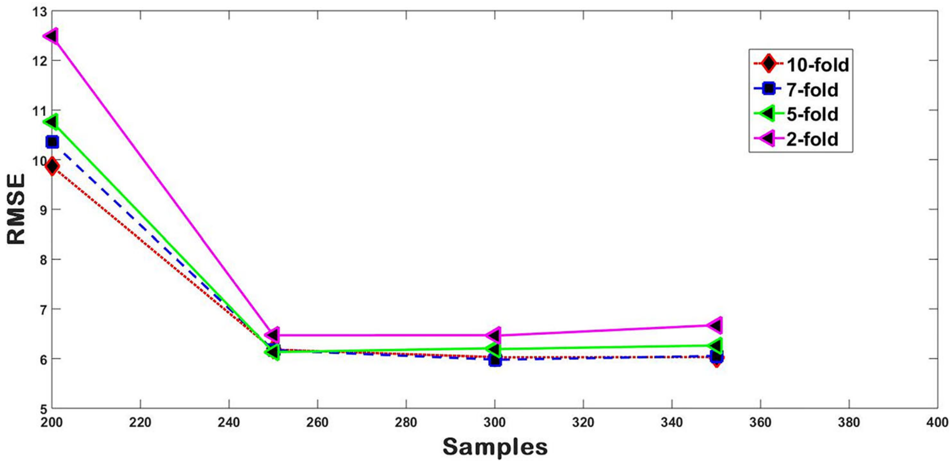

Figure 4. Comparison of different *K*-folds crossvalidation with respect to RMSE.

3.1. Phase–1

Different steps are involved in developing pseudo-modelling like initial design, regression, convergence criteria and adaptive sampling (if required). One of the biggest difficulties in the pseudo-modelling is setting initial design. Basically, the initial design comprises of inputs (adjustable parameters) and outputs (objective function), where the adjustable parameters are selected based on the Sobol sensitivity analysis. Here, screened out the 10 most influential parameters which are inducing the model output and considered as inputs to the pseudo-modelling with NSE as the objective function (output). For deciding on the volume of initial design, (Razavi et al. proposed a criterion between a maximum number of samples/25 $(1000/25 = 40)$ to 20 times of parameter dimensions $(20 \times 10 = 200)$). This study illustrated effects on the model output with varying different sampling methods (LHD, QSR and RND) of 50, 100 and 200 volume space. Once the model is setup for initial design, the next step is to fit the pseudo-models with specified samplings. After fitting the pseudo-model, it must check the model's accuracy. If the model performance is not satisfactory, it can go to 50 additional sampling (Lola–Voronoi). For analysing the pseudo-models, *K*-fold crossvalidation (KCV) chosen as convergence check because it can split the data into training and testing set. In KCV, the accuracy and computational time can depend on the number of folds. Thus, selection of fold also influences to computational burden. So, finding the optimum folds of this study, different *k*-folds are tested such as tenfold, sevenfold, fivefold and twofold, respectively (Figure 4). In Figure 4, clearly vivid that the model performance is directly proportional to a number of folds. With the initial design of the pseudo-model, fivefold shows effective performance when compared with remaining folds (Figure 4). Once, the model is opting for additional samples, the performance acts differently up to 250 samples. Later, it followed the same trend in the model fitting. At a stage, fivefold obtained the same performance as tenfold. While the initial stage of the pseudo-model showed a wide gap between different folds performance but finally it converged into a similar range. At the end of optimization, the model showed the best performances in all the folds. But, as modellers or model users depends on both accuracy as well as time. Where, fivefold CV found the accurate values with favourable computational time and it has the minor difference between tenfold CV and sevenfold CV. Hence, fivefold selected as a model accuracy analyser for judging the pseudo-models.

Once fixing the optimum crossvalidation fold, next step is to identify the best methods in sampling and pseudo-models. Figure 5(a–d), shows the variations of model outputs with different initial size and methods and it is also explained in this figure that the model output is depending on the initial design. Because the model score improved with an increase in the sample sizes. Overall, 50, 100 and 200 initial design size achieved good performances but for selecting one best method among all the methods, 200-initial setup identified accurate values with the lesser number of samples (i.e. 250 samples). While QSR sampling method is completely dominated other sampling methods of LHD and RND. Where QSR took 250 samples to optimize the model and LHD achieved the target but it took a greater number of samples (Figure 5(c)). In Figure 5(d) shows the comparison of best model performance in each stage of the initial design (i.e. 50, 100 and 200). Overall, 200 initial design provided the output with a lesser number of samples as well as higher accuracy. Hence, the initial design of QSR sampling with 200 sample size is selected for this study.

Next, target is to select the best pseudo-model with the specified initial design. To show the relation between the initial setting to the different pseudo-models, Figure 5 depicts the individual performance of each model with varying the sample size. GP and SVM tended to decrease the bias while varying the sample sizes, but ANN achieved the acceptable performances and it cannot follow any trend in its fitting. Among the three methods, GP provided the best implementations over the other methods in different initial setting due to its several hyperparameters and reinforcement learning. Therefore, the initial design (i.e. QSR with 200 initial set) and pseudo-model (i.e. GP) considered as best implementation methods. Our main objective is to show the most suitable pattern for the selection of methods to optimize SWAT parameters through pseudo-modelling. In Figure 6, it clearly shows the pattern with an effective solution for achieving the best SWAT performance.

3.2. Phase–2

The traditional PMO checks the model convergence without any future updates and control of sampling. The adaptive pseudo-modelling-based optimization restricts the samples and helps to computational burden. To show the excellence of adaptive strategy over existing approaches, compared with default model (uncalibrated SWAT model), SUFI-2 algorithm (which incorporated in SWAT-CUP software for calibration and uncertainty of SWAT models) (Abbaspour et al. 2015) and one-shot approach (traditional PMO) based on the performance, required number of samples, optimization results, computational time and prediction of flows to approximate SWAT model.

3.2.1. Performance-wise

Nash–Sutcliffe efficiency (NSE), Percentage of Bias (PBIAS) and Coefficient of determination (R^2) had taken to evaluate the watershed models for both calibration and validation. Table 2 shows the ratings of performance metrics for streamflow prediction and it is clearly showed that > 0.75 NSE, $\leq \pm 10$ PBIAS and > 0.75 R^2 indicates a very good model. So, the above criteria set as a threshold limit for the comparison of all the models. In Figure 7, one-shot approach and adaptive strategy (APMO) achieved the best performances in NSE and R^2. While SUFI-2 obtained good performances but it could not achieve the threshold limit in calibration and validation. While PBIAS shows the error percentage of a model, the default model (SWAT) contained a high range of error of approximately 170 on validation and it clearly explains the model requires the calibration to enhance the performance. Although the default SWAT model obtained the reasonable performances in

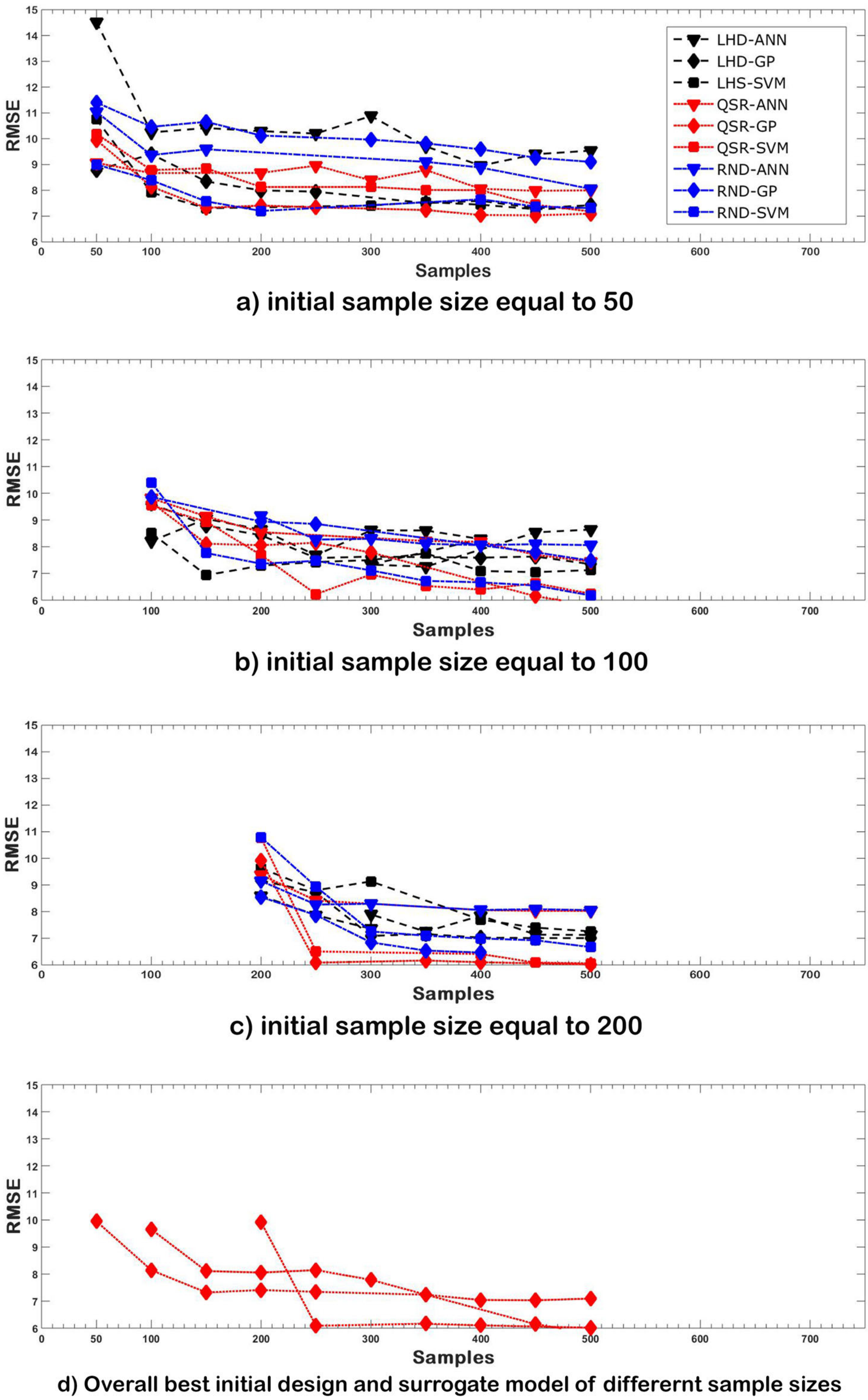

Figure 5. Different samples and pseudo-models performance with respect to RMSE, (a) with initial design size equal to 50; (b) with initial design size equal to 100; (c) with initial design size equal to 200; (d) overall, the best performances of QSR-GP in different initial sample size of 50, 100 and 200, respectively.

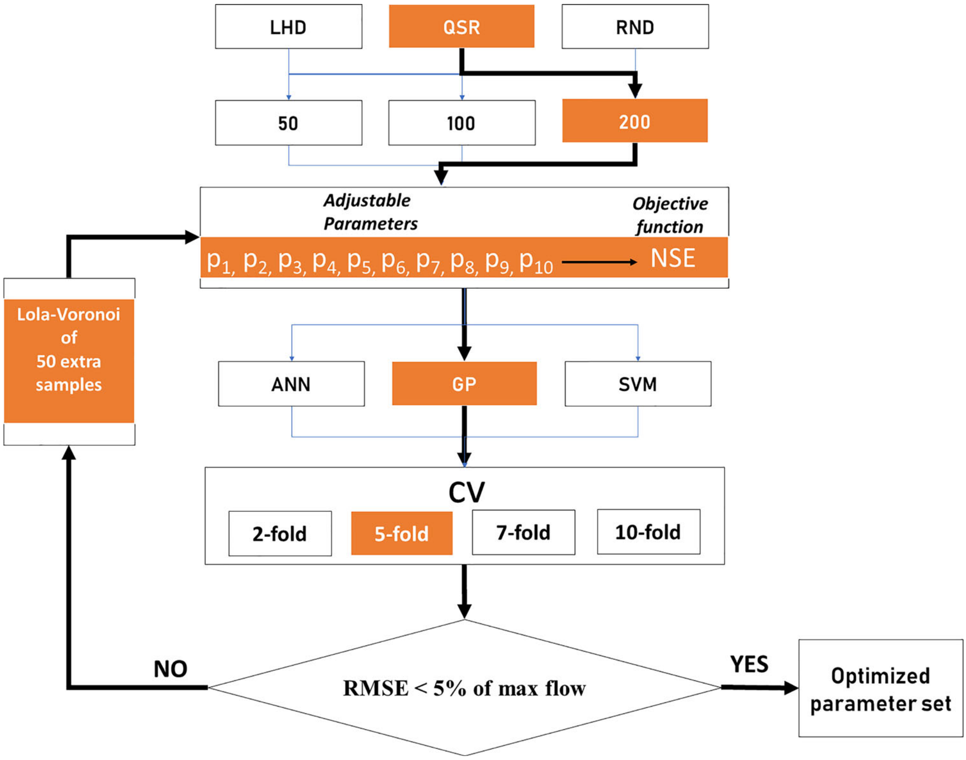

Figure 6. The best pattern to achieve an accurate model output using APMO framework for Kagna watershed.

NSE and R^2 due to its correlation of observed and predicted values, it contains a high range of bias. Based on the threshold limit of PBIAS, APMO achieved best performances and remaining methods are not able to provide near to threshold. Finally, it is vivid that adaptive strategy yields the best performances over one-shot approach and SUFI-2 for prediction of streamflow.

3.2.2. Sample-wise

For setting the maximum number of samples used 100 times of parameter dimensions, that is, $100 \times 10 = 1000$. Based on sub section 3.1, the adaptive strategy seized the samples within 250. While remaining models approached the maximum limits. It is not necessary to obtain for the whole set and enough to obtain necessary sets. As sample size increases computational burden can increase. For effective control of samples, the adaptive sampling able to restrict the parameter sets and make much more computationally efficient.

3.2.3. Computational-wise

The adaptive pseudo-model, one-shot pseudo-model and SUFI-2 are constructed in a MATLAB environment using Core i5 3.2 GHz processor with 8 GB RAM. For optimizing of a model parameter set follows a few steps such as the design of experiments, regression and optimization. Here, the design of experiments can take more time because it has to generate the parameter sets with the respective objective function using real world simulator. In the above section, it clearly explains that the computational time is directly proportional to a number of samples and it is clearly shown in Figure 8 that APMO achieved with fewer samples and ultimately reflects in lesser computational burden. In

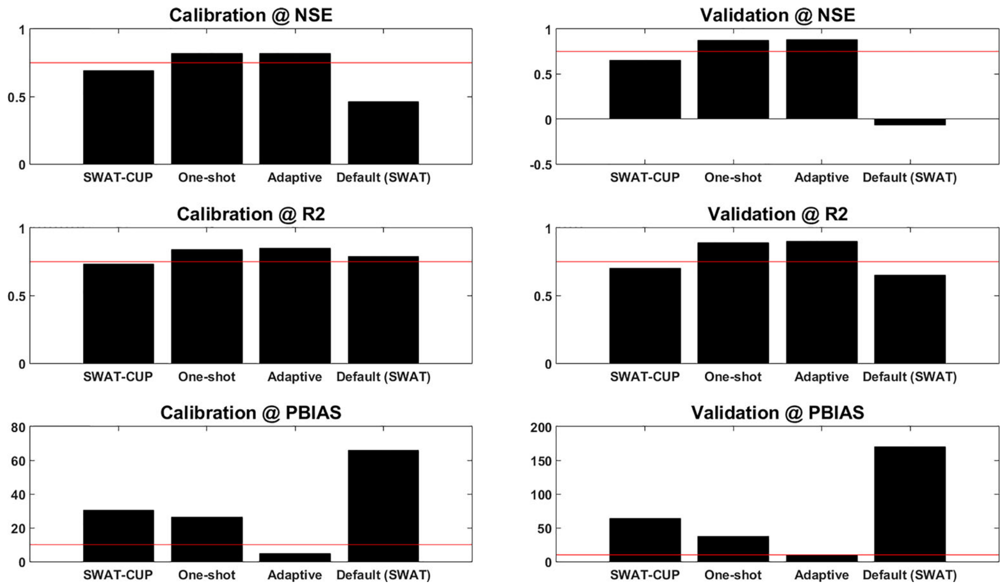

Figure 7. The comparison of different performance metrics for default (SWAT), SWAT-CUP, one-shot (traditional PMO) and adaptive pseudo-modelling optimization (APMO) with the threshold limit of 0.75 (NSE), 0.75 (R^2) and ±10 (PBIAS) in a red line.

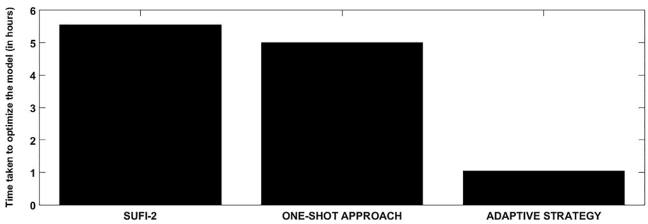

Figure 8. The computational time of SWAT model optimization for different methods.

Figure 8, SUFI-2 and One-shot approach required approximately 5 h to optimize, while adaptive strategy took only 20% of the computational time to that of remaining methods achieved.

3.2.4. Optimization results

Figure 8 shows the normalized optimal parameter set and compared different models based on the sensitivity rank. The parameters like CN2, EPCO, GWQMN, GW_DELAY and RCHRG_DP are most influential while remaining (SOL_AWC, GW_REVAP, REVAPMN, ALPHA_BF and CH_K2) are insensitive. In Figure 9, it is transparent that parameters near to default set which considers as sensitive, if it is so far and it indicates as insensitive. There is a strong correlation between the parameters which has been identified accurately by adaptive strategy, and it stated with less variation of parameters can affect the output. Finally, APMO replicated the original simulation model with sensitive and in-sensitive parameters.

3.2.5. Analyse of flows

For comparison of flows used two types of plots like hydrograph and flow-duration curves (FDC). These plots show accurately whether a model able to capture peaks and baseflow

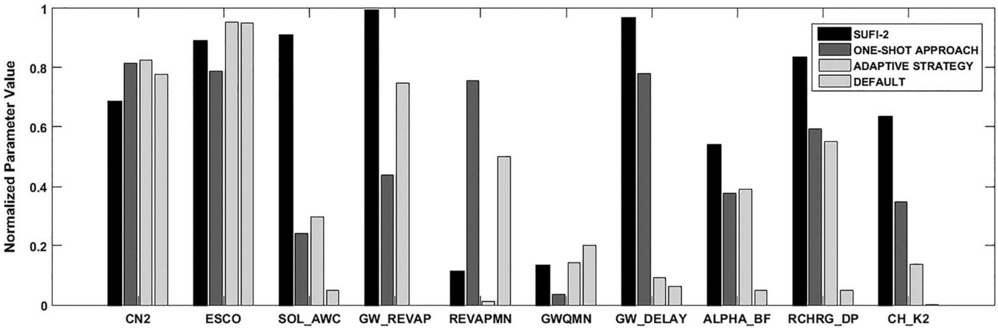

Figure 9. Comparison of the models based on the normalized optimal parameter value.

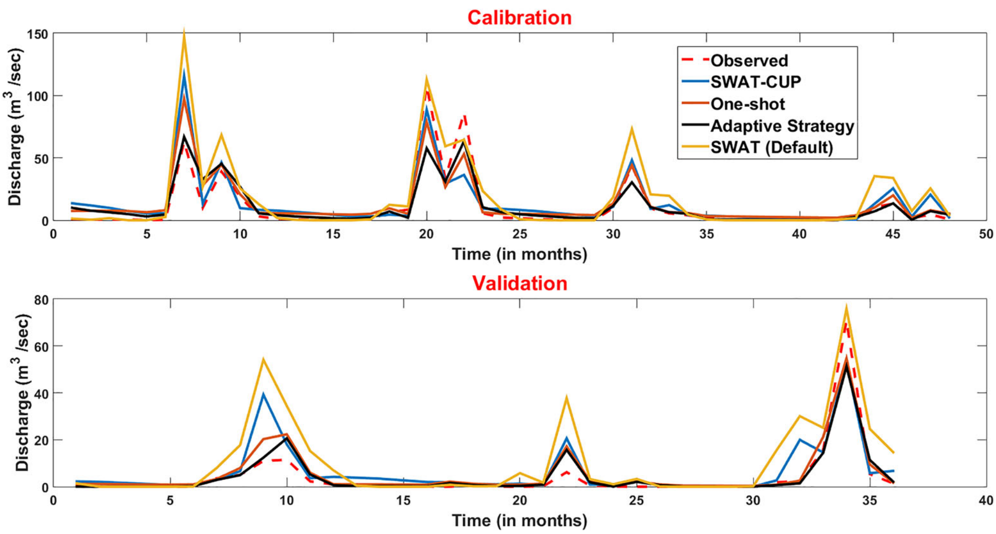

Figure 10. Hydrograph plot for analyzing of different flow predictions in calibration and validation.

or not. The accurate prediction of peaks helps for floods, and baseflow prediction useful for drought conditions. A hydrograph is a plot between time-period versus discharge for both calibration and validation (Figure 10). As observing in Figure 10, all models followed a similar trend, but adaptive strategy almost coincides with observed values in both peaks and base flow. FDC plots can divide into different categories like high flows (0 to Q20), medium flows (Q20 to Q75), low flows (Q75 to Q90) and very low flows (Q90 to Q100), this classification also called flow signatures. According to flow signatures in Figure 11, high flows contained little variation to the observed data of all models and remaining flows has captured the observed data. Moreover, an adaptive strategy is closer to observed data in all forms of flows compared to remaining models. Therefore, APMO proved its best in all sections of evaluation.

4. Conclusions

To show the effective framework to calibrate the computationally intensive hydrological models, we proposed a detailed analysis of APMO framework with model selection, comparison of existing methods and applicability. The complex conceptual models may run

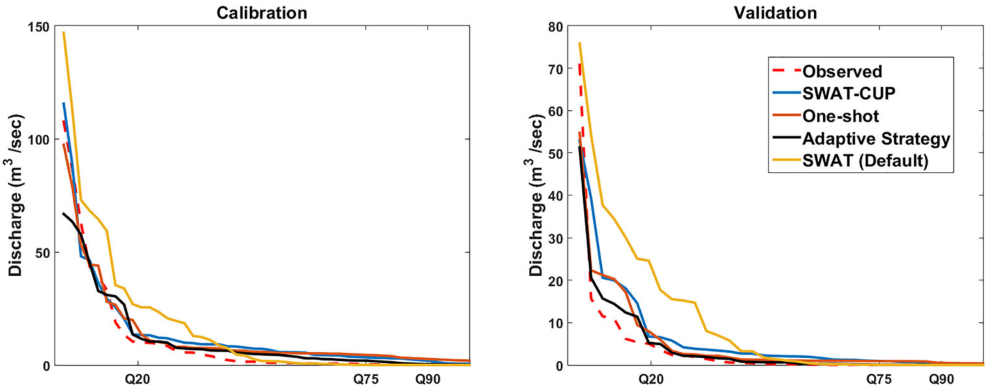

Figure 11. Comparison of flows for Pseudo-modelling-based optimization (Adaptive and One-shot approach), SWAT-CUP model and SWAT default values using Flow Duration Curve's (FDC) for both calibration and validation.

one simulation of nearly a few minutes or even hours. Where pseudo-models can replace the expensive simulation models and make the model computationally efficient. The developed pseud-model is effectively calibrated to SWAT hydrological model parameters. Following conclusions were justified:

- The GP model is adopted as the best pseudo-method when compared to ANN and SVM. SVM performed effectively but failed in a few cases, where GP performs exceptionally best in all cases. QSR of 200 sample size applied as initial design, and it showed effectively in space filling with 20 times of parameter dimensions. Here, it found that the initial design effects the optimization result through different sizes and methods. For additional sampling used Lola–Voronoi algorithm and it provided an adequate number of samples to optimize SWAT model parameters.
- GP model effectively optimized and restricted at 250 samples through adaptive sampling, where the one-shot approach optimized the whole surface. By adaptive sampling, effectively restricted the samples, nearly it controls 80% of the traditional pseudo-model and SUFI-2 computational time.
- The performance of APMO proved its best when compared to other methods. The precision of the optimized parameters in an adaptive strategy is closely related to sensitivity. The developed APMO assists optimization with more accuracy and much lesser computational time. Finally, it enhances the performance of optimization by smoothing the search space.

The accurate prediction of flow can show the behaviour of the watershed. In this study, the proposed model captured the observed values and showed the best modelling process for watershed analysis. The model can provide a basis for decision making of integrated watershed management such as drought management, stormwater and flood management, water quality, water conservation and recycling. Moreover, this proposed algorithm useful for any disciplines for optimization of geosciences and hydro-informatics applications.

Acknowledgement

Our sincere thanks to Centre for Disaster Mitigation and Management (CDMM), VIT Vellore for providing the lab resources.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Abbaspour KC, 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.

Abbaspour KC, Yang J, Maximov I, Siber R, Bogner K, Mieleitner J, Zobrist J, Srinivasan R. 2007. Modelling hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT. J Hydrol. 333(2–4):413–430.

Beven K, Binley A. 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrol Process. 6(3):279–298.

Crombecq K. 2009. A novel sequential desgin strategy for global surrogate modeling. In: Hill RR, Johansson B, Dunkin A, Ingalls RG, editors. Proceedings 2009 Winter Simulation Conference M. D. Rossetti.

Das SK, Ng AWM, Perera BJC. 2013. Development of a SWAT model in the Yarra River catchment. 20th International Congress on Modelling and Simulation; Aug 1–Dec 6; Adelaide. p. 2457–2463.

Dile YT, Daggupati P, George C, Srinivasan R, Arnold J. 2016. Introducing a new open source GIS user interface for the SWAT model. Environ Model Softw. 85:129–138.

Dile YT, Srinivasan R. 2014. Evaluation of CFSR climate data for hydrologic prediction in data- scarce watersheds: an application in the blue nile river basin 1. J Am Water Resour Assoc. 77845:1–13.

Forrester AIJ, Keane AJ. 2009. Recent advances in surrogate-based optimization. Prog Aerosp Sci. 45(1–3):50–79.

Forrester AIJ, Sóbester A, Keane AJ. 2008. Engineering design via surrogate modelling. Hoboken (NJ): Wiley.

Fuka DR, Walter MT, Macalister C, Degaetano AT, Steenhuis TS, Easton ZM. 2013. Using the climate forecast system reanalysis as weather input data for watershed models. Hydrol Process. 28(22): 5613–5623.

Gassman PW, Reyes MR, Green CH, Arnold JG. 2007. The soil and water assessment tool: historical development, applications, and future research directions. Am Soc Agric Biol Eng. 50:1211–1250.

Gong W, Duan Q, Li J, Wang C, Di Z, Dai Y, Ye A, Miao C. 2015. Multi-objective parameter optimization of common land model using adaptive surrogate modeling. Hydrol Earth Syst Sci. 19(5): 2409–2425.

Kennedy J, Eberhart R. 1995. Particle swarm optimization. In International Conference on Neural Networks. IEEE, 1942–1948.

Kuczera G, Parent E. 1998. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. J Hydrol. 211(1–4):69–85.

Moriasi DN, Arnold JG, Van LM, Bingner RL, Harmel RD, Veith TL. 2007. Model evaluation guidelines for sytematic quantification of accuracy in watershed simulations. Am Soc Agric Biol Eng. 50:885–900.

Nash JE, Sutcliffe JV. 1970. River flow forecasting through conceptual models part-1—a discussion of principles. J Hydrol. 10(3):282–290.

Noori N, Kalin L. 2016. Coupling SWAT and ANN models for enhanced daily streamflow prediction. J Hydrol. 533:141–151.

Razavi S, Tolson BA, Burn DH. 2012a. Review of surrogate modeling in water resources. Water Resour Res. 48:1–32.

Razavi S, Tolson BA, Burn DH. 2012b. Environmental modelling & software numerical assessment of metamodelling strategies in computationally intensive optimization. Environ Model Softw. 34:67–86.

Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, Tarantola S. 2010. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Comput Phys Commun. 181(2):259–270.

Uniyal B, Jha MK, Verma AK. 2015. Parameter identification and uncertainty analysis for simulating streamflow in a river basin of Eastern India. Hydrol Process. 29(17):3744–3766.

Van Der Herten J, Deschrijver D, Dhaene T. 2014. Fuzzy local linear approximation-based sequential design. IEEE. 14:2–6.

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 Model Softw. 60:167–179.

meeseva
easier, faster

మీ సేవ
సులభంగా, వేగంగా

FORM 5

00FF 01239135

ఆంధ్రప్రదేశ్ ప్రభుత్వము
GOVERNMENT OF ANDHRA PRADESH
DEPARTMENT OF MUNICIPAL ADMINISTRATION
వైద్య ఆరోగ్యశాఖ
MEDICAL & HEALTH DEPARTMENT
జనన ధృవ పత్రము
BIRTH CERTIFICATE

Certificate Id: 50012-B-82887

జనన మరణ నమోదు చట్టము 1969,12/17 విభాగము ప్రకారము ,ఆంధ్రప్రదేశ్ జనన మరణ నమోదు నిబంధనలు 1999,8/13 నిబంధన క్రింద జారీ చేయబడినది.

(Issued under Section 12/17 of the Registration of Births and Deaths Act 1969 and Rules 8/13 of the Andhra Pradesh Registration of Births and Deaths Rules 1999)

ఆంధ్రప్రదేశ్ రాష్ట్రము చిత్తూరు జిల్లా తిరుపతి నగరపాలక సంస్థ (స్థానిక ప్రదేశము) జనన రిజిష్టరు లోని జననానికి సంబంధించిన అసలు రికార్డు నుండి,క్రింది సమాచారము తీసుకొనబడినదని ధృవీకరించడమైనది.

This is to certify that the following information has been taken from the original record of birth which is in the register for(local area / local body) **THIRUPATHI MUNICIPAL CORPORATION OF CHITTOOR DISTRICT OF STATE OF ANDHRA PRADESH** .

పేరు / Name	**BUDAMALA VENKATESH**
లింగము/ Sex	**MALE**
పుట్టిన తేది / Date of Birth (DD/MM/YYYY)	**05/01/1994** **ZERO FIVE ZERO ONE ONE NINE NINE FOUR**
పుట్టిన స్థలము / Place of Birth	**GOVERNMENT MATERNITY HOSPITAL TIRUPATI -**
తల్లి పేరు / Name of Mother	**SARASWATHI**
తండ్రి/భర్త పేరు / Name of the Father/Husband	**MUNEENDRA BABU**
బిడ్డ జన్మించినపుడు తల్లి దండ్రుల చిరునామ / Address of the parents at the time of Birth of Child	**NA**
తల్లిదండ్రుల స్థిరనివాసపు చిరునామా / Permanent Address of parents	**NA**
నమోదు సంఖ్య/ Registration Number	**445**
నమోదు తేది/ Date of Registration (DD/MM/YYYY)	**04/02/1994**
రిమార్క్సులు/ Remarks	**Y**
జారీ చేసిన తేది / Date Of Issue (DD/MM/YYYY)	**21/10/2014**

NA - Not Available.

Application No:

CDMA011400908735

Date : 24/10/2014

Certified By

UshaKumari

Name : Dr.P.Usha Kumari
Registrar of Births & Deaths
THIRUPATHI MUNICIPAL CORPORATION
CHITTOOR DISTRICT

Note : This is Digitally Signed Certificate, does not require physical signature. And this certificate can be verified at http://www.ap.meeseva.gov.in/ by furnishing the application number mentioned in the Certificate.

Undertaking by the Principal Investigator

To

The Secretary
SERB, New Delhi

Sir

I **Venkatesh B**

herby certify that the research proposal titled **A robust skill verification of hindcast decadal data on hydroclimatic systems** submitted for possible funding by SERB, New Delhi is my original idea and has not been copied/taken verbatim from anyone or from any other sources. I further certify that this proposal has been checked for plagiarism through a plagiarism detection tool i.e. **TurnItin software** approved by the Institute and the contents are original and not copied/taken from any one or many other sources. I am aware of the UGCs Regulations on prevention of Plagiarism i.e. University Grant Commission (Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions) Regulation, 2018. I also declare that there are no plagiarism charges established or pending against me in the last five years. If the funding agency notices any plagiarism or any other discrepancies in the above proposal of mine, I would abide by whatsoever action taken against me by SERB, as deemed necessary.

01/08/2023
Signature of PI with date

Name / designation
Dr. Venkatesh B