Application of a newly developed large-scale conceptual hydrological model in simulating streamflow for credibility testing in data scarce condition

Pranesh K. Paul1,2 | Babita Kumari2 | Srishti Gaur2 | Ashok Mishra2 | Niranjan Panigrahy3 | Rajendra Singh2

1Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China
2Agricultural and Food Engineering Department, Indian Institute of Technology, Kharagpur, India
3Co-ordination Unit, Odisha University of Agriculture & Technology, Bhubaneswar, India

Correspondence
Pranesh K. Paul, Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, 100101 Beijing, China.
Email: praneshpaul5@gmail.com and praneshkp@igsnrr.ac.cn

Abstract
Regionalization approach is adopted to test the credibility of a newly developed, large-scale, conceptual hydrological model, namely, satellite-based hydrological model (SHM), under data-scarce conditions. SHM has a modular structure. There are five modules: surface water module (SW), which uses natural resources conservation service-curve number and Hargreaves equation; forest module (F), which uses water balance and subsurface dynamics; snow module (S), which uses temperature index and radiation-temperature index algorithms; groundwater module, which considers contributions from SW, F, and S modules and uses water level variation process; and routing module, which uses modified spatial distributed direct hydrograph model. Four neighboring basins of varying size are selected to form the reference (Subarnarekha) and the test (Brahmani, Baitarani, and Kangsabati) basins. The similarity index and hierarchical agglomerative cluster analysis show that Brahmani is the most homogeneous basin with respect to the Subarnarekha basin. The model simulations are run at 1 km resolution. The values of statistical indices Nash-Sutcliffe efficiency, coefficient of determination, ...
and percent BIAS, are found to be 0.82%, 0.84%, and 13.99%, respectively, during calibration (1993–1999; 1993 as a warm-up period), and 0.61%, 0.63%, and 14.13%, respectively, during validation (2000–2004) at Jenapur gauging station of the Brahmani river basin. Subsequently, the calibrated parameters are transferred in the model setup of the Subarnarekha basin for simulating streamflow over a randomly chosen period 1998–2000. The results, obtained with both Brahmani and Subarnarekha parameters, confirm that the model performs well with the regionalized parameters, and is suitable for hydrologically homogeneous ungauged basins. The results show that the newly developed model may be used for the water balance simulations in ungauged or scantily gauged basins, leading to the development of water resources plans for better utilization of this vital resource. The results from the sensitivity analysis and uncertainty analysis show that the nonsensitive parameters have a negligible impact on the model results. Further study, however, may be needed to evaluate the model performance and to establish its universal credibility.

**Recommendations for Water Resource Managers and The Scientific Community**

1. This successful testing of the newly developed hydrological model implies that it may be a useful tool for water resources assessment in data-scarce conditions.

2. The encouraging results support the further application of the model for data-scarce basins in different parts of India and the world.

3. This study, for the first time, applies a tool namely “cluster 3.0” (generally used in biotechnology) for hierarchical agglomerative cluster analysis to perform homogeneity test. Any modeler and policymakers may use the tool for homogeneity test of a basin with respect to a reference basin without learning the mathematical theory, in details.
4. The results of parametric sensitivity analysis would help the future modelers in applying this model by providing ideas of the value ranges of model parameters.

**KEYWORDS**
1 km resolution, cluster analysis, credibility testing, regionalization, SHM, similarity index

# INTRODUCTION

Extensive hydrological measurements with spatio-temporal variation are scarce for varying ecosystems and climatic conditions. Therefore, large-scale hydrological modeling is considered as one of the powerful tools to develop required scenarios for sustainable development and management of water resources (Beven, 2012; Beven & Young, 2013; Kauffeldt, Halldin, Rodhe, Xu, & Westerberg, 2013; Singh & Frevert, 2006). Hydrological models, however, are classified by the involvement of physical processes as physically based and conceptual (Refsgaard, 1996). Physically based models use equations of mass, momentum, and energy conservation to represent the physical processes in a deterministic way (Refsgaard, 1996). However, extensive data and computational requirements, and unavoidable uncertainties due to input data, model structure, and model parameters create significant constraints on the use of physically based models (Beven, 1989, 2012; Kebede, Diekkrüger, & Moges, 2014; Singh & Frevert, 2006; Singh & Woolhiser, 2002). This leads to the use of conceptual models, which require comparably lesser computational efforts and time (Singh & Frevert, 2006; Singh & Woolhiser, 2002; Srivastava, Deb, & Kumari, 2019). Keeping this in mind, Space Application Centre (SAC), Ahmedabad has launched the “Programme on Climate change Research In Terrestrial environment-2” (PRACRITI-2). Under the programme, a large-scale conceptual hydrological model, satellite-based hydrological model (SHM), has been developed (Paul, Gaur, et al., 2019; Paul, Kumari, Panigrahi, Mishra, & Singh, 2018). A larger objective behind the model development is to utilize the relevant products such as soil moisture and evapotranspiration from the Indian remote sensing satellites.

Since the prediction of flow for gauged and ungauged basins is one of the fundamental goals in hydrology (Deckers, Booij, Rientjes, & Krol, 2010), robust testing of any model for gauged as well as ungauged basins is of utmost importance for establishing the credibility of the model. To test the credibility of a model for the ungauged basins, regionalization of parameters that involves the transfer of the calibrated parameter values from a gauged basin to an ungauged basin has been the subject of several studies (Ali, Tetzlaff, Soulsby, McDonnell, & Capell, 2012; Athira, Sudheer, Cibin, & Chaubey, 2016; Beck et al., 2016; Dornes et al., 2008; Götzinger & Bárdossy, 2007; He, Bárdossy, & Zehe, 2011; Huang, Liang, & Liang, 2003; Visessri & McIntyre, 2016), there appears to be no universally accepted regionalization method (Deckers et al., 2010; Jin, Xu, Yu, Zhang, & Chen, 2009). However, regionalization approaches that are based on principles of similarity by spatial proximity and similarity of basin characteristics are popularly used (Deckers et al., 2010). The first approach centers on the rationale that since climatic, topographic and physiographic
characteristics are comparable, adjacent basins would exhibit a similar flow regime. The second approach is based on the hypothesis that the calibrated parameters representing specific basin characteristics may also be applied in other basins having similar characteristics.

A considerable effort has been put into studying regionalization approach using physical similarity (Kay, Jones, Crooks, Calver, & Reynard, 2006; Li, Zhang, Chiew, & Xu, 2009; Masih, Uhlenbrook, Maskey, & Ahmad, 2010; McIntyre, Lee, Wheater, Young, & Wagener, 2005; Parajka, Merz, & Blöschl, 2005; Samuel, Coulibaly, & Metcalfe, 2011; Young, 2006; Zhang & Chiew, 2009). However, definitive conclusions on method for physical similarity determination between reference basin and test basins, based on basin descriptors, are not yet universally accepted. In general, soil type, land use, and topographic characteristics are the most widely used basin descriptors in regionalization approach (Merz & Blöschl, 2004; Razavi & Coulibaly, 2013). Using these basin descriptors, researchers use varying methodologies to find out the physical similarity of test basins with respect to reference basins, including, similarity index (Burn & Boorman, 1993), Kohonen networks, and fuzzy c-means techniques (Hall & Minns, 1999), and various agglomerative clustering algorithms (Nathan & McMahon, 1990). The clustering algorithms include k-means clustering (Burn, 1989), fuzzy c-means ward’s cluster method and hybrid cluster analysis (Ramachandra Rao & Srinivas, 2006). In addition to these, self-organized maps (Jingyi & Hall, 2004), hierarchical clustering and hierarchical agglomerative clustering (Ouyang, Ren, Cheng, & Zhou, 2010), and fuzzy clustering and Kohonen method (Abdolhay, Saghafian, Soom, & Ghazali, 2012) are also used.

In recent years, a few studies focus on parameter regionalization and credibility testing using multiscale parameter regionalization (MPR), the similarity between gauged and ungauged catchment attributes, differential split sample (DSS) test, and temporal, spatial and spatio-temporal techniques (Gaborit, Ricard, Lachance-Cloutier, Anctil, and Turcotte (2015); R. Kumar, Livneh, & Samaniego (2013); Patil & Stieglitz (2015); Rakovec et al., 2016; Seiller, Hajji, & Anctil, 2015; Sellami, Jeunesse, Benabdallah, Baghdadi, & Vanclooster, 2014). However, regionalization studies remain rare in data-scarce Indian Territory. Besides, SHM is already tested successfully for data available conditions (Paul, Gaur, et al., 2019). Since testing of SHM in data-scarce conditions is still not done, this study focuses on testing SHM for ungauged basins. Besides, SHM is applied with 1 km resolution (details are provided in Section 2.2). Regionalization study at such finer resolution is unique, at least in India, if not in other parts of the globe. The specific objective of this study is to test a newly developed, large-scale, conceptual hydrological model for data-scarce conditions using regionalization approach. A sensitivity analysis is performed to identify the sensitive parameter. Also, the uncertainty analysis is performed to find the contribution of the nonsensitive parameters to the overall model simulation uncertainty and to ascertain the suitability of SHM for ungauged basins.

2 | STUDY AREA

The study area (Figure 1) comprises of four river basins, namely, Brahmani, Baitarani, Subarnarekha, and Kangsabati, located in Eastern India between 83° 52′ and 86° 15′ east longitudes, and 21° 15′ and 23° 35′ north latitudes. The basins are situated in subtropical climate zone having southwest monsoon season from June to October. Table 1 presents a description of
The study area in detail. Agricultural land, water bodies, and forests cover 69.7%, 7.8%, and 17.5% of the study area, respectively. Figure 1 also shows gauging stations for streamflow measurement along with the reservoirs present in the area.

2.1 Data

Daily rainfall and maximum and minimum temperatures are obtained from India Meteorological Department (IMD), Pune at 1° × 1° resolution. Data are interpolated to 1 km × 1 km resolution by using bilinear interpolation technique to use as input into SHM. Soil and land use land cover (LULC) maps are collected from the Food and Agriculture Organisation (FAO) website (http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/en/) at 1 km × 1 km scale. Digital Elevation Model (DEM) of 30 m × 30 m resolution is taken from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) website (https://asterweb.jpl.nasa.gov/gdem.asp). DEM is resampled into 1 km × 1 km resolution for use in SHM. Observed streamflow data, for various gauging stations of the study area, are collected from the Central Water Commission (CWC), Bhubaneswar, India. The observed streamflow is available for 1995–2004, 1992–2004, 1987–2005, and 1987–2005 for Jenapur, Muri, Jamshedpur, and Ghatshila, respectively. The observed streamflow at Muri and Jamshedpur gauging stations are naturalized to reduce the effect of reservoirs. The detailed description of the naturalization process is provided below.
| Name of basin | Drainage area (km²) | Elevation (m above mean sea level) | Mean annual rainfall (mm) | Mean annual temperature (minimum and maximum) (°C) | Mean annual runoff (mm) | Gauging station | Distinct characteristics | Major land-use attributes (%) |
|---------------|---------------------|----------------------------------|---------------------------|--------------------------------------------------|-------------------------|----------------|--------------------------|-----------------------------|
| Brahmani      | 39,033              | 1 to 1041                        | 1,305                     | 4 and 47                                         | 465                     | Jenapur       | Rain fed agriculture and flood-prone | 63.68 10.68 16.39 |
| Subarnarekha  | 29,196              | 6 to 803                         | 1,800                     | 8 and 45                                         | 224                     | Muri, Jamshedpur, Ghatshila | Richest coal, iron, and bauxite deposits | 55.34 14.63 14.73 |
| Baitarani     | 10,982              | 10 to 750                        | 1,450                     | 5 and 47.5                                       | 446                     | Anandpur     | Rich red loamy and gravely detritus soil | 55.88 17.46 14.85 |
| Kangsabati    | 5,796               | 11 to 656                        | 1,400                     | 10.6 and 40                                      | 397                     | Kangsabati dam site | Drought-prone | 43.47 27.98 23.85 |
2.1.1 | Naturalization of streamflow

There are two reservoirs, Getalsud and Chandil, in Subarnarekha basin. Since we have gauged flow data (at three gauging stations of Subarnarekha basin) for the post-reservoir period, these may represent a different hydrologic regime of the basin from the prereservoir period (Biemans et al., 2011; Magilligan & Nislow, 2005; Risley, Constantz, Essaid, & Rounds, 2010). As SHM presently does not take into account the effect of the reservoirs during model simulation, we need to reconstruct the flow data to minimize the impact of reservoir storage on streamflow simulations. There are several techniques available for streamflow naturalization, for example, drainage area ratio method (Hirsch, 1979; Wurbs, 2006), modified soil conservation service (SCS)-curve number (CN) technique (Wurbs, Muttitah, & Felden, 2005), generic equations (Wurbs, 2006), and regional statistical technique (Hirsch, 1979). Among these, regional statistical technique and drainage area ratio methods are well-suited when the observed flow data for the prereservoir period are not available. The regional statistical technique, however, performs better if one or more base stations have long continuous streamflow records available (Hirsch, 1979). Thus, we have used this technique to naturalize the flow at Muri and Jamshedpur gauging stations. The most downstream, and hence the least hydrologically affected gauging station, Ghatshila has been as the base station.

2.1.2 | Regional statistical technique

The technique assumes that the standardized flow at the base station and the gauging station, chosen for data reconstruction, are equal for every month. The reconstruction equation is as follows:

\[
Q_{ij} = A(q_j) + B(q_j) \frac{P_{ij} - A(p_i)}{B(p_j)},
\]

where \(A(q_j)\) is the estimated mean flow for month \(j\) at the gauging station; \(B(q_j)\) is the estimated standard deviation of flow for month \(j\) at the gauging station; \(A(p_j)\) is the mean flow for month \(j\) at the base station; \(B(p_j)\) is the standard deviation of flow for month \(j\) at the base station; \(P_{ij}\) is the observed streamflow (day \(i\), month \(j\)); and \(Q_{ij}\) is the reconstructed streamflow (day \(i\), month \(j\)).

The regional regression equations between streamflow and basin characteristics are used to estimate means and standard deviations of the monthly flows at Muri and Jamshedpur gauging stations (Thomas & Benson, 1970). The regression equations used are of the following form:

\[
X = aA_{h1}S_{b1}L_{b2}S_{i1}E_{b5}F_{b6}(P - 20)^{b7}I_{24,2}^{b8},
\]

where \(X\) is the dependent variables (\(A(q_j)\) and \(B(q_j)\), stated in Equation (1)); \(A\) is the drainage area; \(S\) is the slope of the main channel; \(L\) is the length of the main channel; \(S_i\) is the percentage area of water bodies; \(E\) is the mean basin elevation; \(F\) is the percentage of forest area in the basin; \(P\) is the mean annual precipitation; \(I_{24,2}\) is the 2-year, 24-hr maximum rainfall intensity; and \(a, b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8\) are the regression coefficients. However, to estimate the values of regression coefficients, \(A(p_j)\) and \(B(p_j)\) are used on the left-hand side of Equation (2) as dependent variables. The values of independent variables, on
the right-hand side of Equation (2), are considered for the Ghatshila subbasin. The estimated values of the regression coefficients are subsequently used to determine means and standard deviations of monthly flows at Muri and Ghatshila gauging stations. From the regression equation, used in Hirsch (1979), terms related to temperature and snowfall, are neglected as snowfall does not occur in the study area.

2.2 Description of SHM

SHM works on 1 km × 1 km spatial grid resolution and properties at the center of a cell are assumed to be the properties of the cell. SHM has five modules: surface water (SW), forest (F), snowmelt (S), groundwater (GW), and routing (ROU). SHM grid cells corresponding to forest and snow land cover are modeled using the F and S modules, respectively; whereas other grid cells are modeled using the SW module. SHM uses MySQL (open source software) as a relational database management system (RDBMS). The details of the model are given below.

2.2.1 Model structure

Figure 2 presents the model structure of SHM in a particular grid cell.

2.2.2 Core theory

In the SW module, NRCS-CN method (Chow, Maidment, & Mays, 2005) and Hargreaves method (Hargreaves & Samani, 1985) are used to estimate the surface runoff and the potential evapotranspiration (PET), respectively. A water balance method is used to determine soil moisture. The soil profile is considered as a single-layered zone of 300 mm depth, and moisture-holding and moisture-transmitting characteristics of the soil and underlying layer are considered while accounting for the soil moisture. The wetting of the soil layer depends on the amount of infiltration. The amount of water, more than the maximum capacity of the zone, percolates, and contributes to the groundwater. The soil moisture is depleted by evapotranspiration, at a potential rate or actual rate, depending on the moisture condition. The potential rate is considered as long as sufficient moisture is available.

The F module functions based on water balancing and the dynamics of the subsurface to provide output in the form of runoff, soil moisture, evapotranspiration, and contribution to groundwater using the technique and parameters stated in Das, Islam, Dutta, Dubey, and Sarkar (2014). Subsurface is considered of having soil matrix and macropores. There are two types of macropores: main bypass and internal catchments, with the main bypass being responsible for direct contribution to groundwater. Soil matrix is assumed to have three layers, which are crucial for water balance and changes in the soil moisture. After infiltration, the saturation of three layers gets started from the top in batch, and after complete saturation of the three layers, the excess water goes to the groundwater. Runoff generation, after a precipitation event, occurs according to the antecedent conditions in the subsurface.

The S module determines the snow density from snow albedo (Smith & Halverson, 1979) for estimating snowmelt depth by using two different algorithms, namely, temperature index
algorithm and radiation-temperature index algorithm. Since the study area does not have any snow-covered land, S module is not considered in this study.

GW module uses the contribution from SW, F, and S modules and generates baseflow, following the water level variation process described in Sekhar and Rasmi (2004). The resultant baseflow along with the surface runoff generated from other modules is routed up to the outlet as streamflow.

For routing of generated runoff and baseflow, modified spatial distributed direct hydrograph (SDDH; Paul et al., 2018) is used. This is a cell-to-cell routing technique. The details of the modification of ROU module are described in Paul et al. (2018).

2.2.3 Model parameters

There are 23 parameters in the model, which are listed in Table 2.

2.3 Model applications

SHM is applied in three applications so far. SHM is applied in the Kabini basin for testing the compatibility of the ROU module (Paul et al., 2018). Subsequently, SHM is tested using hierarchical operational testing of Klemes (1986) (Paul, Gaur, et al., 2019). The performance of SHM is further compared with the widely used Soil and Water Assessment Tool (SWAT) model (Paul, Zhang, Mishra, Panigrahi, & Singh, 2019).
| Sl no. | Name of parameters                                      | Types of parameter                                                                 | Range   | Unit    | Module | Symbol |
|--------|--------------------------------------------------------|------------------------------------------------------------------------------------|---------|---------|--------|--------|
| 1      | Curve number                                           | Spatially distributed-based on LULC, soil type and slope                           | 0–100   | SW      | CN     |        |
| 2      | Interception parameter                                 | Static                                                                            | 0.01–0.1| SW      | C_{int}|        |
| 3      | Manning’s roughness coefficient (surficial depression) | Spatially distributed-based on LULC                                              | 0.01–0.5| m^{−(1/3)} | SW    | N      |
| 4      | Fraction of macropores ended at bottom of A-horizon    | Spatially distributed-based on LULC                                              | 0–1     | F       | fmA    |        |
| 5      | Shape factor for frequency distribution curve IC domain | Spatially distributed-based on LULC                                              | 0–1     | F       | Sfi    |        |
| 6      | Volume fraction of static macropores at the soil surface| Spatially distributed-based on LULC                                              | 0–0.5   | F       | Vfm    |        |
| 7      | Proportion of IC domain at the soil surface            | Spatially distributed-based on LULC                                              | 0–1     | F       | Pid    |        |
| 8      | Maximum diameter of soil polygons                       | Static                                                                            | 0.1–1,000 cm | F | max_Dsp |        |
| 9      | Minimum diameter of soil polygons                       | Static                                                                            | 0.1–1,000 cm | F  | min_Dsp |        |
| 10     | Surface runoff generation threshold                     | Spatially distributed-based on soil type                                         | 0.15–3 cm | F | thold   |        |
| 11     | Full limit of geological storage                        | Static                                                                            | 5–10 cm | F       | Geo_lim|        |
| 12     | Upper limit of perched aquifer thickness               | Static                                                                            | 3–5 m   | F       | D_lim  |        |
| 13     | Transmissivity                                         | Spatially distributed-based on aquifer type                                       | 12–1,200 m²/day | GW | T       |        |
| 14     | Specific yield                                         | Spatially distributed-based on aquifer type                                       | 0.01–0.3 | GW    | S_y   |        |
| 15     | Groundwater recharge factor                            | Static                                                                            | 0.2–0.9 | GW      | Grf    |        |
| 16     | Minimum groundwater table                              | Spatially distributed-based on aquifer type                                       | 200–2,000 m above MSL | GW | H_{min}|        |
| Sl no. | Name of parameters                              | Types of parameter                          | Range  | Unit   | Module | Symbol |
|-------|------------------------------------------------|---------------------------------------------|--------|--------|--------|--------|
| 17    | Discharge parameter                            | Static                                      | 0.1–0.5| GW     | Pd     |
| 18    | Manning’s roughness coefficient (overland flow) | Spatially distributed-based on LULC         | 0.01–0.5| m<sup>−(1/3)</sup> | ROU    | n_o    |
| 19    | Manning’s roughness coefficient (channel flow)  | Static                                      | 0.01–0.5| m<sup>−(1/3)</sup> | ROU    | n_c    |
| 20    | Width of channel                               | Static                                      | 0.5–1,000| m      | ROU    | B      |
| 21    | Critical temperature                           | Static                                      | 1–1.6  | °c     | S      | Ct     |
| 22    | Runoff coefficient for snow melt               | Static                                      | 0–1    |        | S      | Rcs    |
| 23    | Runoff coefficient for rain                    | Static                                      | 0–1    |        | S      | Rcr    |

Abbreviation: LULC, land use land cover.
2.4 | Methodology

To test the capability of SHM for data-scarce conditions, Subarnarekha river basin, a gauged basin, is selected as the reference (data-scarce) basin. The Brahmani, Baitarani, and Kangsabati are used as the test basins (for details please see Section 2) to check the regionalization capability of the model. Homogeneity test is conducted for identifying the most homogeneous test basin with respect to the reference basin, Subarnarekha. For homogeneity test, we have used “similarity index” (Burn & Boorman, 1993) technique because of its simplicity, and “hierarchical agglomerative cluster analysis” (Ouyang et al., 2010) because of its pattern discovering capability of the attributes. Though Subarnarekha is assumed as the data-scarce basin, we calibrated and validated SHM setup for the basin at Muri (calibration period: 1991–1998 [1991 as a warm-up]; validation period: 1999–2004), Jamshedpur (calibration period: 1986–1996 [1986 as a warm-up]; validation period: 1997–2004) and Ghatshila (calibration period: 1986–1996 [1986 as a warm-up]; validation period: 1998–2004) gauging stations. Testing of the credibility of the model is performed during a randomly chosen period (1998–2000). SHM simulation results obtained using the transferred parameters from the most homogeneous test basin are compared with the SHM simulation results with Subarnarekha parameters at Muri, Jamshedpur, and Ghatshila stations. Finally, to understand the role of sensitive SHM parameters in regionalization process and model competency, sensitivity analysis, and parametric uncertainty analysis are performed for the reference basin and the most homogeneous basin.

2.4.1 | Regionalization approach

For regionalization, the homogeneity test is performed first to identify the most homogeneous basin among the test basins (Brahmani, Baitarani, and Kangsabati) with respect to the reference basin (Subarnarekha). After determining the most homogeneous basin, the model is calibrated and validated for the chosen basin. Subsequently, the calibrated parameters are transferred to the model setup of the reference basin for streamflow simulation.

2.4.2 | Homogeneity test

Homogeneity test is performed using the “similarity index” and the “agglomerative hierarchical cluster analysis technique.” The techniques used in the homogeneity test are described below.

2.4.3 | Similarity index technique

Burn and Boorman (1993) defined similarity index as the “sum of absolute differences of k selected physiographic attributes of the gauged $X^G$ basin and the ungauged $X^U$ basin, normalized by the range of attributes ($\Delta X$)”.

Mathematically, it is expressed as follows:

$$
\Phi = \sum_{i=1}^{k} \frac{|X^G_i - X^U_i|}{\Delta X},
$$

where $\Phi$ is the similarity index.
Here, for performing the similarity index test, we have used the combination of attributes based on LULC (FAO data), soil type (FAO data) and ten equal intervals of elevation, ranging from 0 m to 1,041 m, for all four basins. The test basin with the lowest value of the similarity index is chosen as homogeneous to the reference basin. Besides LULC, soil and elevation, land management features related to agricultural management, urban management, and hydraulic structures may also play a role in driving the basin similarity to a certain extent. We, however, did not consider these features here due to unavailability of the detailed information.

2.4.4 Cluster analysis

In this study, we applied the hierarchical agglomerative clustering algorithm, considering complete linkage dissimilarity and the Euclidean distance measure (Ouyang et al., 2010). We used an open-source Cluster 3.0 software package for the purpose (de Hoon, Imoto, Nolan, & Miyano, 2002, 2004; Eisen, Spellman, Brown, & Botstein, 1998). The Euclidean distance measurement is performed using the following equation:

\[ d(x, y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)^2. \]  

(4)

The distance \(d\) between \(x\) and \(y\) is computed directly, where \(x\) and \(y\) are values of the same attribute for reference and test basins, respectively. While using the Euclidean distance measure, the data, however, is suitably normalized.

The software contains Cluster and TreeView programs to provide a computational and a graphical environment, respectively, for analysing data. The LULC, soil type, and 10 equal intervals of elevation ranging from 0 m to 1,041 m, covering the study area, are combined to form hydrological response units (HRUs) for four basins and used as input to the Cluster 3.0 program. TreeView program is subsequently run with output files from cluster 3.0 program as input, and the output is visualized and analysed.

Figure 3 presents the essential components of TreeView (Status Panel, Array tree, Array names, Global, and zoom pixels) in Cluster 3.0. The visual interpretation of the Global pixels of Array tree (a zoomed-in view is presented in the right panel of the figure) helps us in identifying the most homogeneous basin with respect to the reference basin. For example, based on the oval-marked similar regions of the global pixels in the figure, Basin D would be chosen as the most homogeneous basin with respect to the reference Basin A.

2.4.5 Transfer of parameters from the homogeneous basin to reference basin

There are two types of parameters in SHM—spatially distributed and static (Table 2). Both types of parameters are calibrated in the homogeneous basin and transferred to the reference basin. For sensitive spatially distributed parameters, the calibration is performed for various combinations of watershed characteristics. For example, we explain the calibration of the parameter \(n_o\). “\(n_o\)” values are dependent on different LULC classes. Therefore, “\(n_o\)” values are calibrated over the specified range for different LULC classes in the homogeneous basin. Later, the calibrated values are transferred from the database of the homogeneous basin to the database of
reference basin for respective LULCs. The details are described in Figure S1. On the other hand, the static parameters are lumped over the basin. For nonsensitive parameters, we use developers’ default values/values from standard references.

2.4.6 | Evaluation criteria

Nash-Sutcliffe efficiency (NSE; Nash & Sutcliffe, 1970) and coefficient of determination ($R^2$; Moriasi et al., 2007) are used as the goodness-of-fit parameters to assess the model simulation accuracy, along with percent bias (PBIAS; Moriasi et al., 2007) for measuring the performance for the water balance.

2.4.7 | Sensitivity analysis

Sensitivity analysis is conducted to identify the model output variability with respect to changes in the values of 20 SHM parameters (Table 2) for the period 1998–2000 in Jenapur subbasin of Brahmani basin and three subbasins of the Subarnarekha basin (Baird, 1989; Paul, Gaur, et al., 2019; Paul et al., 2018). Three parameters of the S module (Table 2) are excluded from the sensitivity analysis exercise as the study basins are free from snow. The model output variability is analysed against variations in the value of a single parameter around its calibrated value. NSE

FIGURE 3  TreeView containing various clusters formed in Cluster 3.0

| Cluster Analysis | Elevation class | Soil classes | LULC classes |
|------------------|----------------|--------------|--------------|
|                  | Class 1       | Class 2      | Class 3      | Class 1       | Class 2      | Class 3      | Number of grid cells in different basins | Number of grid cells in different basins | Number of grid cells in different basins |
| Basin A          | X1            | XY1          | XY21         | P1            | PQ1          | PQR1         | A1            | AB1            | ABC1 |
| Basin B          | X2            | XY2          | XY22         | P2            | PQ2          | PQR2         | A2            | AB2            | ABC2 |
| Basin C          | X3            | XY3          | XY23         | P3            | PQ3          | PQR3         | A3            | AB3            | ABC3 |
| Basin D          | X4            | XY4          | XY24         | P4            | PQ4          | PQR4         | A4            | AB4            | ABC4 |
is chosen as the objective function for this analysis. Parameters are perturbed by ±5% and ±10% from their calibrated values, one at a time, while keeping the other parameters constant.

### 2.4.8 Uncertainty analysis

For each parameter, 95 percent predictive uncertainty (PPU) band is developed to analyse the uncertainty. For the purpose, simulated streamflows obtained using the calibrated values and the four perturbed values, that is at ±5% and ±10% of the calibrated values, of the parameters are considered. Consequently, 97.5% upper band and 2.5% lower band are estimated to develop to 95 PPU band. Then, values of P (measured data bracketed by the 95 PPU band) and R (the relative length of the 95 PPU band with respect to the model simulated values) are determined to quantify the uncertainty of each parameter, following the procedure described in Paul, Zhang, et al. (2019). Consequently, the p value versus R value scatter plots are used for individual subbasins to quantify the uncertainty of individual parameters.

The contribution in uncertainties from different sources (sensitive parameters, nonsensitive parameters, their interactions and other sources) in the simulated streamflow is obtained through analysis of variance (ANOVA). Though, it is a widely used technique in climate projection analysis of streamflow (Bosshard et al., 2013; Gaur, Bandyopadhyay, & Singh, 2020; Yip, Ferro, Stephenson, & Hawkins, 2011); it is rarely used for identifying uncertainty sources in hydrological simulations. In the process, the difference between simulated values and references values, $\Delta Y$ (Equation 5) is considered as the input in the ANOVA method for the partitioning of uncertainty due to individual sources as:

$$\Delta Y = Y_{\text{sim}} - Y_{\text{obs}} \forall Y,$$

where $Y_{\text{sim}}$ and $Y_{\text{obs}}$ represent the simulated and observed values for a given variable $Y$.

An ANOVA model is constructed with different contributing sources, that is, sensitive parameters (represented by $\alpha$), nonsensitive parameters (represented by $\beta$), and the error term ($\varepsilon$) representing the internal variability (Equation 6). The nonlinear term ($\alpha \beta$) shows the interaction between these two sources and is also considered a factor for uncertainty.

$$\Delta Y = \mu + \alpha + \beta + \alpha \beta + \varepsilon.$$  

ANOVA partitions the total variance into components due to different sources of variation. Therefore, the total variance in simulated streamflow is represented by the sum of squares from contributions and interactions as follows:

$$SS_{\text{Total}} = SS_{\text{Contribution}} + SS_{\text{Interaction}} + SS_{\text{Other sources}},$$

where $SS_{\text{Total}}$ (Equation 7) presents the total sum of squares and $SS_{\text{Contribution}}$ and $SS_{\text{Interaction}}$ present the sum of squares from contributors and interactions, respectively.

Equation (7) may also be expressed as follows:

$$SS_{\text{Total}} = SS_{\alpha} + SS_{\beta} + SS_{\alpha \beta} + SS_{\varepsilon},$$

where $\forall$ Contribution $\in (\alpha, \beta)$, $\forall$ interaction $\in (\alpha \beta)$, $\forall$ Other Sources $\in (\varepsilon)$. 

PAUL ET AL. | Natural Resource Modeling | 15 of 27
Uncertainty is estimated by the ratio of variance contribution due to individual factors to total variance (\(SS_{\text{Total}}/SS_{\text{Contribution}}\) or \(SS_{\text{Total}}/SS_{\text{Interaction}}\) or \(SS_{\text{Total}}/SS_{\text{Other Sources}}\)). Therefore, the variance contribution due to each term (Equation 8) is estimated.

3 | RESULTS AND DISCUSSION

Results of the regionalization approach and sensitivity analysis for both reference and the most homogeneous basins are presented and discussed in this section.

3.1 | Calibration and validation of SHM in Subarnarekha basin

Figures S2 and S3 in the Supporting Information section present the time series plots of the simulated and observed streamflows and scatter plots of simulated versus observed streamflows at the gauging stations of Subarnarekha basin, respectively. The values of statistical indices, NSE, \(R^2\), and PBIAS are found to be 0.65%, 0.61%, and 13.12%, respectively, during Muri calibration, and 0.5%, 0.51%, and 16.37%, respectively, during Muri validation. Similarly, NSE, \(R^2\), and PBIAS values are 0.8%, 0.82%, and 14.31%, respectively, during Jamshedpur calibration, and 0.67%, 0.77%, and 14.15%, respectively, during Jamshedpur validation. For Ghatshila, NSE, \(R^2\), and PBIAS are 0.66%, 0.61%, and 13.13%, respectively, during calibration, and 0.5%, 0.5%, and 16.38%, respectively, during validation. The values of statistical indices show that the performance of SHM is satisfactory (i.e., NSE > 0.36, \(R^2 > 0.5\), and \(-25% \leq \text{PBIAS} \leq 25%\)) during the calibration and validation periods for Muri, Jamshedpur, and Ghatshila gauging stations (Moriasi et al., 2007; Van Liew, Veith, Bosch, & Arnold, 2007). The results observed here are better than those reported in Paul, Gaur, et al. (2019). The improved model performance may be attributed to a finer grid resolution of one km grid cell resolution used here as compared to 5 km grid cell resolution in Paul, Gaur, et al. (2019). However, as evident from PBIAS values, underestimation persists throughout the model simulation (though within acceptable limit).

3.2 | Outcome of regionalization approach

3.2.1 | Outcome of homogeneity test

Results of the homogeneity test to identify the most homogeneous test basin with respect to the reference basin are presented below.

3.2.2 | Outcome using similarity index technique

Table 3 presents the results of the homogeneity test using the similarity index technique. As evident, Brahmani is the most homogeneous basin with respect to Subarnarekha basin as the similarity index is the lowest (1.04) for this basin as compared to Baitarani and Kangsabati basins.
3.2.3 | Outcome from hierarchical agglomerative cluster analysis

Figure 4 presents the zoom pixels extracted from TreeView of Cluster 3.0 software, obtained using the data of the reference (Subarnarekha) and test basins (Brahmani, Baitarani, and Kangsabati). The similar regions of the pixels in Figure 4, marked with blue-colored ovals and circles, show that Brahmani basin is the most homogeneous basin with respect to Subarnarekha basin. Thus, both the Similarity Index technique and Hierarchical Agglomerative Cluster Analysis technique produce similar results.

3.2.4 | Model simulation results with regionalization approach

SHM setup for Brahmani river basin is calibrated over 1993–1999 (1993 as a warm-up) and validated over 2000–2004 at Jenapur gauging station. The values of statistical indices, NSE, $R^2$, and PBIAS, are found to be 0.82%, 0.84%, and 13.99%, respectively, during calibration, and 0.61%, 0.63%, and 14.13%, respectively, during validation. The values of the statistical indices show that SHM is calibrated satisfactorily at Jenapur gauging station of Brahmani basin (Moriasi et al., 2007; Van Liew et al., 2007). We have, however, calibrated the model against discharge at the outlet only. Thus, our calibration technique includes only the temporal variability of the system and neglects the spatial component as the aggregated discharge at the outlet neglects the sensitivity to spatially explicit hydrological states and fluxes. Figure 5 presents the time series plot of the simulated and observed streamflows, and the scatter plot of simulated versus observed streamflow at Jenapur gauging station. Figure 5a depicts that the simulated flows capture the overall temporal pattern of the observed flow well. It is also evident from Figure 5b that the linear relationship between the observed and simulated streamflows is acceptable. Subsequently, the calibrated parameters of Brahmani basin are transferred in the

| Attributes | Absolute difference of the attributes normalized by respective ranges |
|------------|---------------------------------------------------------------|
| LULCs      | Brahmani 0.47 Baitarani 0.61 Kangsabati 0.64                |
| Elevation  | Brahmani 0.41 Baitarani 0.68 Kangsabati 0.54                |
| Soil type  | Brahmani 0.16 Baitarani 0.22 Kangsabati 0.41                |
| Similarity index ($\Phi$) | Brahmani 1.04 Baitarani 1.52 Kangsabati 1.60 |

*Note: Bold value is to show the lowest total similarity Index for Brahmani basin in comparison to other basins.*

Name of basins for cluster analysis

**FIGURE 4** Zoom pixels of TreeView window in Cluster 3.0
SHM model setup of Subarnarekha basin for simulating the streamflow over a randomly chosen period of 1998–2000.

Figure 6a presents the time series plot of simulated and observed streamflows at three gauging stations of Subarnarekha basin for two cases: (a) with calibrated Brahmani parameters,
and (b) with calibrated parameters of the subbasins of Subarnarekha basin. The figure indicates that the simulated flow captures the overall temporal pattern of the observed flow well. It is clear from 1:1 scatter plots (Figure 6b) that the linear relationship between the observed and simulated streamflows is acceptable. Besides, Figure 7 summarizes the statistical indices of simulation results at three gauging stations of Subarnarekha basin with Brahmani parameters. The statistical indices obtained using previously calibrated parameters for Subarnarekha basin are also included for better comparison. As evident, NSE and $R^2$ values are satisfactory at all three gauging stations with both Brahmani and Subarnarekha parameters (Moriasi et al., 2007; Van Liew et al., 2007). PBIAS values show flow underestimation at all three stations, though, within the acceptable limit (Moriasi et al., 2007). However, at Muri gauging station, medium flows are overestimated at a few instances Brahmani parameters (Figure 6a(i)), whereas, high flows are overestimated with Subarnarekha parameters (Figure 6a(ii)). At Ghatshila gauging station, however, medium and high flows are underestimated with both Brahmani and Subarnarekha parameters (Figure 6a(v), (vi)). The flow underestimation at Ghatshila gauging station may be due to nonnaturalized observed flow (Paul, Gaur, et al., 2019). Since regionalized model parameters cannot account for all the variations in the basin, under- or over-estimation of the observed flow to a certain extent are expected. Overall results, however, show that SHM performs credibly with regionalized parameters at 1 km grid cell resolution. Thus, it is proved to be a useful modeling tool for flow simulations in data-scarce conditions.

**FIGURE 6** (a) Simulation results (as daily time series plot) with (i) Brahmani parameters at Muri gauging station, (ii) Subarnarekha parameters at Muri gauging station, (iii) Brahmani parameters at Jamshedpur gauging station, (iv) Subarnarekha parameters at Jamshedpur gauging station, (v) Brahmani parameters at Ghatshila gauging station, and (vi) Subarnarekha parameters at Ghatshila gauging station. (b) Simulation results (as scatter plot) with (i) Brahmani parameters at Muri gauging station, (ii) Subarnarekha parameters at Muri gauging station, (iii) Brahmani parameters at Jamshedpur gauging station, (iv) Subarnarekha parameters at Jamshedpur gauging station, (v) Brahmani parameters at Ghatshila gauging station, and (vi) Subarnarekha parameters at Ghatshila gauging station.
Sensitivity analysis

Sensitivity analysis is carried out following the procedure described in Section 2.4, and only three parameters, CN, $n_o$ and $n_c$ are found to be sensitive (Figure 8). Figure 8 presents the results of the sensitivity analysis. The results show that CN is the most sensitive parameter, followed by $n_o$ and $n_c$, for both Subarnarekha and Brahmani basins. In Jenapur subbasin, streamflow simulations are more sensitive to negative perturbation than positive perturbation, specifically for perturbation of $-5\%$ for CN and $-10\%$ for $n_o$ and $n_c$. In Muri subbasin, however, streamflow simulations are sensitive to both positive and negative perturbations in the calibration parameters, with higher sensitivity to negative perturbations. Besides, CN appears to be sensitive at Jamshedpur gauging station for negative perturbation and at Ghatshila gauging stations for both positive and negative perturbations. However, $n_c$ and $n_o$ show limited sensitivity at Jamshedpur gauging station. The figure also depicts that at Ghatshila gauging station, $n_c$ and $n_o$ show limited sensitivity for positive perturbations, and negligible sensitivity for negative perturbations. Besides, it is also visible that parameter sensitivities are unidirectional at Jenapur and Muri as compared to the other two subbasins. Overall, the variability of NSE, however, is not uniform due to complex interactions between parameter sets (Table 2) and the hydrological responses for the three subbasins of

FIGURE 6 (Continued)
Subarnarekha basin and Jenapur subbasin of Brahmani basin (Beven, 2006). The non-sensitivity of F module parameters may be due to the low forest cover in the study area. The lower sensitivity of GW module parameters shows that the groundwater contribution to the streamflow generation is low. However, the nonsensitive behavior of the ROU module parameter “B” is contradictory to the earlier finding of Paul et al. (2018). The larger size of the study area may be responsible for this behavior.

3.4 | Uncertainty analysis

The 95 PPU bands, along with the observed data, for all parameters and all the subbasins, are produced in Figure S4. Figure 9 presents the p value versus R value plots for various parameters in different subbasins. It is clear from Figure 9 that the sensitive parameters show more uncertainty as compared to the nonsensitive parameters. Figure 10 presents the uncertainty contribution of the sensitive parameters, nonsensitive parameters, their interactions and other sources to the total
model simulation uncertainty for different subbasins. It is evident from Figure 10 that non-sensitive parameters have the lowest contribution to the total model uncertainty. The sensitive parameters are the primary source of uncertainty in all subbasins, followed by the other sources. The other sources may include model structure and input data uncertainty. The next contributor is, however, the interaction between sensitive and nonsensitive parameters. Though the order of contribution from different sources of uncertainty to the total uncertainty remains the same for all subbasins, the magnitudes vary over the subbasin. The lower contribution of nonsensitive parameters to the total model uncertainty, obtained from the ANOVA analysis corroborates the findings from the $p$ value versus $R$ value plots (Figure 9).

4 CONCLUSIONS

Regionalization approach is adopted to test the credibility of a newly developed large-scale, conceptual, hydrological model, SHM, for data-scarce conditions. Model is applied at 1 km grid cell resolution. The similarity index and hierarchical agglomerative cluster analysis show that Brahmani is the most homogeneous basin with respect to the Subarnarekha basin. Overall results of the regionalization approach demonstrate that SHM performs credibly in simulating streamflow with regionalized parameters in a hydrologically homogeneous region at 1 km grid.
cell resolution. The sensitivity analysis shows that three parameters, CN, \( n_o \), and \( n_c \), are sensitive. The uncertainty analysis shows that the sensitive parameters are the major contributors to the total model uncertainty, with contributions of nonsensitive parameters being negligible. Therefore, SHM may be adopted in ungauged basins of India and other parts of the world for studying the streamflow characteristics, leading to better management of water resources. Further applications, however, may establish the finding with more details.

**FIGURE 9**  
\( P \) value vs. \( R \) value for various satellite-based hydrological model parameters in different subbasins of the study area. "The names of the nonsensitive parameters are not shown to keep the figure less hazy"

**FIGURE 10**  
Contribution of different sources into uncertainty in streamflow simulation for 1998–2000
ACKNOWLEDGMENTS
This study is financially supported by Space Application Centre (SAC), Ahmedabad (Grant no. IIT/SRIC/AGFE/DWI/2013-14/124). The technical support of programmer Mr. Partha Samanta in developing the model is acknowledged. We also acknowledge the technical support of Dr. R. P. Singh and Dr P. K. Gupta from SAC, Ahmedabad, and our other project partners from NERIST, Itanagar, IIT Guwahati and IISc, Bangalore.

AUTHOR CONTRIBUTIONS
Pranesh K. Paul is responsible for conceptualization (lead), data curation (lead), formal analysis (lead), investigation (equal), methodology (equal), resources (equal), software (equal), validation (equal), visualization (equal), writing-original draft (equal), and writing-review and editing (equal). Babita Kumari was involved in data curation (equal), formal analysis (equal), methodology (equal), software (equal), validation (equal), writing-original draft (equal). Srishti Gaur is involved in conceptualization (equal), data curation (equal), formal analysis (equal), investigation (equal), methodology (equal), software (equal), validation (equal), writing-original draft (equal). Ashok Mishra is involved in conceptualization (equal), data curation (equal), funding acquisition (equal), methodology (equal), project administration (equal), resources (equal), software (equal), supervision (equal), visualization (equal), writing-review and editing (equal). Niranjan Panigrahy is responsible in conceptualization (equal), data curation (equal), investigation (equal), methodology (equal), project administration (equal), resources (equal), software (equal), supervision (equal), visualization (equal), writing-review & editing (equal). Rajendra Singh is responsible for conceptualization (equal), data curation (equal), formal analysis (equal), investigation (equal), project administration (equal), resources (equal), software (equal), supervision (equal), visualization (equal), writing-review and editing (equal).

ORCID
Pranesh K. Paul http://orcid.org/0000-0002-8504-3477

REFERENCES
Abdolhay, A., Saghafian, B., Soom, M. A. M., & Ghazali, A. H. B. (2012). Identification of homogenous regions in Gorganrood basin (Iran) for the purpose of regionalization. Natural Hazards, 61(3), 1427–1442.
Ali, G., Tetzlaff, D., Soulsby, C., McDonnell, J. J., & Capell, R. (2012). A comparison of similarity indices for catchment classification using a cross-regional dataset. Advances in Water Resources, 40, 11–22.
Athira, P., Sudheer, K. P., Cibin, R., & Chaubey, I. (2016). Predictions in ungauged basins: An approach for regionalization of hydrological models considering the probability distribution of model parameters. Stochastic Environmental Research and Risk Assessment, 30(4), 1131–1149.
Baird, B. F. (1989). Managerial Decisions under uncertainty: An introduction to the analysis of decision making. New York, NY: Wiley.
Beck, H. E., van Dijk, A. I. J. M., de Roo, A., Miralles, D. G., McVicar, T. R., Schellekens, J., & Bruijnzeel, L. A. (2016). Global-scale regionalization of hydrologic model parameters. Water Resources Research, 52(5), 3599–3622.
Beven, K. (1989). Changing ideas in hydrology—The case of physically-based models. Journal of Hydrology, 105, 157–172.
Beven, K. (2006). A manifesto for the equifinality thesis. Journal of Hydrology, 320(1–2), 18–36.
Beven, K., & Young, P. (2013). A guide to good practice in modeling semantics for authors and referees. Water Resources Research, 49(8), 5092–5098.
Beven, K. (2012). Rainfall-runoff modelling: The primer, Rainfall-runoff modeling: The primer (2nd ed., pp. 1–18). Chichester: Wiley-Blackwell.
Biemans, H., Haddeland, I., Kabat, P., Ludwig, F., Hutjes, R. W. A., Heinke, J., ... Gerten, D. (2011). Impact of reservoirs on river discharge and irrigation water supply during the 20th century. *Water Resources Research, 47*(3), 1–15. https://doi.org/10.1029/2009WR008929

Bossard, T., Carambia, M., Goerring, K., Kotlarski, S., Krahe, P., Zappa, M., & Schar, C. (2013). Quantifying uncertainty sources in an ensemble of hydrological climate-impact projections. *Water Resources Research, 49*(3), 1523–1536.

Burn, D. H. (1989). Cluster analysis as applied to regional flood frequency. *Journal of Water Resources Planning and Management, 115*(5), 567–582.

Burn, D. H., & Boorman, D. B. (1993). Estimation of hydrological parameters at ungauged catchments. *Journal of Hydrology, 143*(3–4), 429–454.

Chow, V. T., Maidment, D. R., & Mays, L. W. (2005). *Applied hydrology* (pp. 147–155). New York, NY: McGraw-Hill.

Das, P., Islam, A., Dutta, S., Dubey, A. K., & Sarkar, R. (2014). *Estimation of runoff curve numbers using a physically-based approach of preferential flow modelling*. Hydrology in a Changing World: Environmental and Human Dimensions. Proceedings of FRIEND-Water 2014, Montpellier, France, IAHS publication, p. 363.

Deckers, D. L. E. H., Booij, M. J., Rientjes, T. H. M., & Krol, M. S. (2010). Catchment variability and parameter estimation in multi-objective regionalisation of a rainfall-runoff model. *Water Resources Management, 24*(14), 3961–3985.

Dornes, P. F., Tolson, B. A., Davison, B., Pietroniro, A., Pomeroy, J. W., & Marsh, P. (2008). Regionalisation of land surface hydrological model parameters in subarctic and arctic environments. *Physics and Chemistry of the Earth, 33*(17–18), 1081–1089.

Eisen, M. B., Spellman, P. T., Brown, P. O., & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America, 95*(25), 14863–14868.

Gaborit, É., Ricard, S., Lachance-Cloutier, S., Ancilt, F., & Turcotte, R. (2015). Comparing global and local calibration schemes from a differential split-sample test perspective. *Canadian Journal of Earth Sciences, 52*(11), 990–999.

Gaur, S., Bandyopadhyay, A., & Singh, R. (2020). Modelling potential impact of climate change and uncertainty on streamflow projections: A case study. *Journal of Water and Climate Change*. https://doi.org/10.2166/wcc.2020.254

Götzinger, J., & Bárpossy, A. (2007). Comparison of four regionalisation methods for a distributed hydrological model. *Journal of Hydrology, 333*(2–4), 374–384.

Hall, M. J., & Minns, A. W. (1999). The classification of hydrologically homogeneous regions. *Hydrological Sciences Journal, 44*(April), 693–704.

Hargreaves, G. H., & Samani, Z. A. (1985). Reference crop evapotranspiration from ambient air temperature. *American Society of Agricultural Engineers, 1*, 96–99.

He, Y., Bárpossy, A., & Zehe, E. (2011). A review of regionalisation for continuous streamflow simulation. *Hydrology and Earth System Sciences, 15*(11), 3539–3553.

Hirsch, R. M. (1979). An evaluation of some record reconstruction techniques. *Water Resources Research, 15*(6), 1781–1790. https://doi.org/10.1029/WR015i006p01781

de Hoon, M. J. L., Imoto, S., Nolan, J., & Miyano, S. (2002). Statistical analysis of a small set of time-ordered gene expression data using linear splines. *Bioinformatics, 18*(11), 1477–1485.

de Hoon, M. J. L., Imoto, S., Nolan, J., & Miyano, S. (2004). Open source clustering software. *Bioinformatics, 20*(9), 1453–1454.

Huang, M., Liang, X., & Liang, Y. (2003). A transferability study of model parameters for the variable infiltration capacity land surface scheme. *Journal of Geophysical Research, 108*(D22), 8864.

Jin, X., Xu, C., Yu, Zhang, Q., & Chen, Y. D. (2009). Regionalization study of a conceptual hydrological model in Dongjiang basin, south China. *Quaternary International, 208*(1–2), 129–137.

Jingyi, Z., & Hall, M. J. (2004). Regional flood frequency analysis for the Gan-Ming River basin in China. *Journal of Hydrology, 296*(1–4), 98–117.

Kauffeldt, A., Halldin, S., Rodhe, A., Xu, C. Y., & Westerberg, I. K. (2013). Disinformative data in large-scale hydrological modelling. *Hydrology and Earth System Sciences, 17*(7), 2845–2857.

Kay, A. L., Jones, D. A., Crooks, S. M., Calver, A., & Reynard, N. S. (2006). A comparison of three approaches to spatial generalization of rainfall-runoff models. *Hydrological Processes, 20*, 3953–3973.
Kebede, A., Diekkrüger, B., & Moges, S. A. (2014). Comparative study of a physically based distributed hydrological model versus a conceptual hydrological model for assessment of climate change response in the Upper Nile, Baro-Akobo basin: A case study of the Sore watershed, Ethiopia. *International Journal of River Basin Management, 12*(4), 299–318.

Klemes, V. (1986). Operational testing of hydrological simulation models. *Hydrological Sciences Journal, 31*(1), 13–24.

Kumar, R., Livneh, B., & Samaniego, L. (2013). Toward computationally efficient large-scale hydrologic predictions with a multiscale regionalization scheme. *Water Resources Research, 49*(9), 5700–5714.

Li, H., Zhang, Y., Chiew, F. H. S., & Xu, S. (2009). Predicting runoff in ungauged catchments by using Xinanjiang model with MODIS leaf area index. *Journal of Hydrology, 370*(1–4), 155–162.

Van Liew, M. W., Veith, T. L., Bosch, D. D., & Arnold, J. G. (2007). Suitability of SWAT for the conservation effects assessment project: Comparison on USDA agricultural research service service watersheds. *Journal of Hydrologic Engineering, 12*(2), 173–189.

Magilligan, F. J., & Nislow, K. H. (2005). Changes in hydrologic regime by dams. *Geomorphology, 71*(1–2), 61–78. https://doi.org/10.1016/j.geomorph.2004.08.017

Masih, I., Uhlenbrook, S., Maskey, S., & Ahmad, M. D. (2010). Regionalization of a conceptual rainfall-runoff model based on similarity of the flow duration curve: A case study from the semi-arid Karkheh basin, Iran. *Journal of Hydrology, 391*(1–2), 188–201.

McIntyre, N., Lee, H., Wheater, H., Young, A., & Wagener, T. (2005). Ensemble predictions of runoff in ungauged catchments. *Water Resources Research, 41*(12), 1–14.

Merz, R., & Blöschl, G. (2004). Regionalisation of catchment model parameters. *Journal of Hydrology, 287*(1–4), 95–123.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE, 50*(3), 885–900.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology, 10*(3), 282–290.

Nathan, R. J., & McMahon, T. A. (1990). Identification of homogeneous regions for the purposes of regionalisation. *Journal of Hydrology, 121*, 217–238.

Ouyang, R., Ren, L., Cheng, W., & Zhou, C. (2010). Similarity search and pattern discovery in hydrological time series data mining. *Hydrological Processes, 24*(9), 1198–1210.

Parajka, J., Merz, R., & Blöschl, G. (2005). A comparison of regionalisation methods for catchment model parameters. *Hydrology and Earth System Sciences, 9*(3), 157–171.

Patil, S. D., & Stieglitz, M. (2015). Comparing spatial and temporal transferability of hydrological model parameters. *Journal of Hydrology, 525*, 409–417.

Paul, P. K., Gaur, S., Yadav, B., Panigrahy, N., Mishra, A., & Singh, R. (2019). Diagnosing credibility of a large-scale conceptual hydrological model in simulating streamflow. *Journal of Hydrologic Engineering, 24*(4), 4019004. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001766

Paul, P. K., Kumari, N., Panigrahi, N., Mishra, A., & Singh, R. (2018). Implementation of cell-to-cell routing scheme in a large scale conceptual hydrological model. *Environmental Modelling & Software, 101*, 23–33.

Paul, P. K., Zhang, Y., Mishra, A., Panigrahi, N., & Singh, R. (2019). Comparative study of two state-of-the-art semi-distributed hydrological models. *Water, 11*, 871. https://doi.org/10.3390/w1105087

Rakovec, O., Kumar, R., Mai, J., Cuntz, M., Thober, S., Zink, M., ... Samaniego, L. (2016). Multiscale and multivariate evaluation of water fluxes and states over European river basins. *Journal of Hydrometeorology, 17*(1), 287–307.

Ramachandra Rao, A., & Srinivas, V. V. (2006). Regionalization of watersheds by hybrid-cluster analysis. *Journal of Hydrology, 318*(1–4), 37–56.

Razavi, T., & Coulibaly, P. (2013). Streamflow prediction in ungauged basins: Review of regionalization methods. *Journal of Hydrologic Engineering, 18*(8), 958–975.

Refsgaard, J. C. (1996). Terminology, modeling protocol and classification of hydrologic model codes. In M. B. Abbott & J. C. Refsgaard (Eds.), *Distributed Hydrologic Modelling* (pp. 17–40). Dordrecht: Kluwer Academic Publishers.

Risley, J. C., Constantz, J., Essaid, H., & Rounds, S. (2010). Effects of upstream dams versus groundwater pumping on stream temperature under varying climate conditions. *Water Resources Research, 46*(6), 1–32. https://doi.org/10.1029/2009WR008587
Samuel, J., Coulibaly, P., & Metcalfe, R. A. (2011). Estimation of continuous streamflow in Ontario ungauged basins: comparison of regionalization methods. *Journal of Hydrologic Engineering, 16*(5), 447–459.

Seiller, G., Hajji, I., & Anctil, F. (2015). Improving the temporal transposability of lumped hydrological models on twenty diversified U.S. watersheds. *Journal of Hydrology: Regional Studies, 3*, 379–399.

Sekhar, M., & Rasmi, S. (2004). Groundwater flow modeling of Gundal sub-basin in Kabini river basin, India. *Asian Journal of Water Environment and Pollution, 1*(1), 65–77.

Sellami, H., La Jeunesse, I., Benabdallah, S., Baghdadi, N., & Vanclouster, M. (2014). Uncertainty analysis in model parameters regionalization: A case study involving the SWAT model in Mediterranean catchments (Southern France). *Hydrology and Earth System Sciences, 18*(6), 2393–2413.

Singh, V. P. & Frevert, D. K. (Eds.), (2006). *Watershed models* (pp. 3–14). Boca Raton, FL: Taylor and Francis Group.

Singh, V. P., & Woolhiser, D. A. (2002). Mathematical Modeling of Watershed Hydrology. *Journal of Hydrologic Engineering, 7*(4), 270–292.

Smith, J. L., & Halverson, H. G. (1979). *Estimating snowpack density from Albedo measurement*. Research Paper PSW-RP-136 U.S. Department of Agriculture, Forest Service, Pacific Southwest Forest and Range Experiment Station, Berkeley, CA

Srivastava, A., Deb, P., & Kumari, N. (2019). Multi-model approach to assess the dynamics of hydrologic components in a tropical ecosystem. *Water Resources Management, 34*, 327–341. https://doi.org/10.1007/s11269-019-02452-z

Thomas, D. M., & Benson, M. A. (1970). *Generalization of streamflow characteristics from drainage basin characteristics*. USGS WaterSupply Paper 1975, Washington, DC: US Government Printing Office.

Vissessri, S., & McIntyre, N. (2016). Regionalisation of hydrological responses under land-use change and variable data quality. *Hydrological Sciences Journal, 61*(2), 302–320.

Wurbs, R. A. (2006). Methods for developing naturalized monthly flows at gauged and ungauged sites. *Journal of Hydrologic Engineering, 11*(1), 55–64. https://doi.org/10.1061/(ASCE)1084-0699(2006)11:1(55)

Wurbs, R. A., Muttiah, R. S., & Felden, F. (2005). Incorporation of climate change in water availability modeling. *Journal of Hydrologic Engineering, 10*(5), 375–385. https://doi.org/10.1061/(ASCE)1084-0699(2005)10:5(375)

Yip, S., Ferro, C. A. T., Stephenson, D. B., & Hawkins, E. (2011). A simple, coherent framework for partitioning uncertainty in climate predictions. *Journal of Climate, 24*(17), 4634–4643.

Young, A. R. (2006). Stream flow simulation within UK ungauged catchments using a daily rainfall-runoff model. *Journal of Hydrology, 320*(1–2), 155–172.

Zhang, Y., & Chiew, F. H. S. (2009). Relative merits of different methods for runoff predictions in ungauged catchments. *Water Resources Research, 45*(W07412), 1–13. https://doi.org/10.1029/2008WR007504

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section.

---

**How to cite this article:** Paul PK, Kumari B, Gaur S, Mishra A, Panigrahy N, Singh R. Application of a newly developed large-scale conceptual hydrological model in simulating streamflow for credibility testing in data scarce condition. *Natural Resource Modeling*. 2020;33:e12283. https://doi.org/10.1111/nrm.12283