

Hydrologic modeling by means of a hybrid downscaling approach: an application to the Sai Gon–Dong Nai Rivers Basin

T. Trinh^{a,b,c}, V. T. Nguyen^{d,*}, N. Do^e, K. Carr^a, D. H. Tran^d, N. V. Thang^b and T. H. Dang^d

^aHydrologic Research Laboratory, Department of Civil and Environmental Engineering, University of California, Davis, CA, USA

^bInstitute for Computational Science and Technology, SBI Building, Quang Trung Software City, Ho Chi Minh City, Vietnam

^cInstitute of Ecology and Works Protection, Vietnam Academy for Water Resources, Hanoi 116830, Vietnam

^dDepartment of Civil and Environmental Engineering, Seoul National University, Seoul, Republic of Korea

^eVietnam Academy for Water Resources, Hanoi, Vietnam

*Corresponding author. E-mail: vnguyen@snu.ac.kr

ABSTRACT

The spatial and temporal availability and reliability of hydrological data are substantial contributions to the accuracy of watershed modeling; unfortunately, such data requirements are challenging and perhaps impossible in many regions of the world. In this study, hydrological conditions are simulated using the hydrologic model-WEHY, whose data input are obtained from a hybrid downscaling technique to provide reliable and high temporal and spatial resolution hydrological data. The hybrid downscaling technique is coupled to a hydroclimate and machine learning models; wherein the global atmospheric reanalysis data, including ERA-Interim, ERA-20C, and CFSR are used for initial and boundary conditions of dynamical downscaling utilizing the Weather Research and Forecasting model (WRF). The machine learning model (ANN) then follows to further downscale the WRF outputs to a finer resolution over the studied watershed. An application of the combination of mentioned techniques is applied to the third-largest river basin in Vietnam, the Sai Gon–Dong Nai Rivers Basin. The validation of hybrid model is in the 'satisfactory' range. After the estimation of geomorphology and land cover within the watershed, WEHY's calibration and validation are performed based on observed rainfall data. The simulation results matched well with flow observation data with respect to magnitude for both the rising and recession time segments. In comparison, among the three selected reanalysis data sets, the best calibration and validation results were obtained from the CFSR data set. These results are closer to the observation data than those using only the dynamic downscaling technique in combination with the WEHY model.

Key words: artificial neural network (ANN), CFSR, ERA-20C, ERA-Interim, watershed environmental hydrology (WEHY), weather research and forecasting (WRF)

HIGHLIGHTS

- Data input for hydrologic model can obtain from a hybrid downscaling technique.
- Hybrid technique is coupled with a regional climate model with a machine learning model.
- Hybrid technique can provide better hydrological data in higher resolutions.

1. INTRODUCTION

High accuracy of hydrological data is considered a fundamental and challenging issue in the scientific research of water resources. Reliable estimations of hydrological data are not only required for designing large hydraulic structures, such as dams and flood protection structures, but also enable better decision making for irrigation strategies, water allocation planning, urban water supply, and water risk mitigation such as drought and flood (Dudley 1988; Ishida *et al.* 2015; Trinh *et al.* 2016; Hirpa *et al.* 2018; Bui *et al.* 2019; Liu *et al.* 2019; Zhang *et al.* 2019). According to Hirpa *et al.* (2018), high accuracy forecasted flow data were used to identify upcoming flood hazards ahead of their occurrence. Bui *et al.* (2019) took advantage of hydrological data in combination with other influencing variables (soil, land cover, topography, etc.) to predict flash flood in tropical typhoon areas. Trinh *et al.* (2016) applied a physically based modeling approach to obtain reliable flood frequency in the Cache Creek watershed in California over the 21st century. The results of that study were used for planning and design of water resources projects and for floodplain management in the downstream of the Cache Creek watershed (Tu *et al.* 2020). Such high-quality hydrological data are often estimated by means of reliable modeling, monitoring, and sensor technologies

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

which are not available in various regions around the world. Further studies are needed to determine a new standard approach to provide high accuracy hydrological data.

Hydrological data, such as flow data, can be obtained from hydrological stations (Hu *et al.* 2011; Malik *et al.* 2019). One critical issue with such data is that, for various reasons, it is typically limited or not available at the finer temporal and spatial scales on demand. Recently, flow data have been obtained through satellite remote sensing techniques (Smith 1997; Bjerklie *et al.* 2018); however, these techniques do not provide the high spatio-temporal resolution that is required for flood simulations. Another approach is stochastic hydrologic models (Salas 1993) which are based on the existing statistical properties from historical data over a study area. This method relies heavily on observations that may not be available and does not account for the ongoing change in the world's hydroclimate conditions (Milly *et al.* 2007). Recently, there have also been attempts to simulate streamflow data by means of hydrological models with input provided from estimated atmospheric data; Kavvas *et al.* (2013) and Trinh *et al.* (2020) developed and applied a physically based, distributed hydroclimate model to reconstruct and project flow data using global reanalysis climate data (Chen *et al.* 2011; Kure *et al.* 2013; Trinh *et al.* 2017; Gorguner *et al.* 2019). The global reanalysis climate data are generally too coarse for direct application to regional studies, such as water allocation and flow forecasting at the watershed scale; therefore, an application of downscaling techniques is required. There are two main downscaling techniques; the statistical and dynamical downscaling (DD) methods. Statistical downscaling (SD) methods take into account the empirical, spatial, and temporal relationships between global climate indicators (predictors) and regional-scale climate variables (predictands), and are trained based on historical periods. Since SD methods are based on the assumption of an unchanged statistical relationship, they do not consider the ongoing change in hydroclimate conditions. In order to provide a much more representative view of regional climate conditions, DD operates based on physical realism with complex local processes (Salathé 2003; Pierce *et al.* 2012; Walton *et al.* 2017); however, this approach requires considerable computational cost, simulation time, and output storage.

While SD and DD methods are widely used in climatology research, both face drawbacks that limit their applicability. Recently, the approach of combining DD with SD has been explored (Anh & Taniguchi 2018; Trinh *et al.* 2021; Tu *et al.* 2021). This new technique, called hybrid downscaling (HD), first uses large-scale atmospheric conditions as determined by a GCM for its lateral boundary conditions before being downscaled by a RCM model, then applies statistical methods to further downscale from selected RCM outputs to a finer spatial resolution (Trinh *et al.* 2021). According to Trinh *et al.* (2021), the HD technique is able to improve the accuracy of simulated data both in temporal and spatial scales. Another highlight of this approach is its inexpensive computational demand with respect to computer resources and time consumption.

In this context, an objective development of a methodology to simulate high accuracy hydrological data through coupling of a hydroclimate model with a machine learning technique is proposed. The proposed technique uses input provided from three different global reanalysis data sets, including ECMWF – Atmospheric Reanalysis coarse climate data of the 20th century (ERA-20C; Poli *et al.* 2016), ECMWF – Reanalysis Interim (ERA-Interim; Berrisford *et al.* 2009; Dee *et al.* 2011), and Climate Forecast System Reanalysis (CFSR; Saha *et al.* 2010; Wang *et al.* 2011). These reanalysis data were dynamically downscaled by a regional climate model followed by further downscaling with an artificial neural network (ANN) model to reach a finer resolution over the studied region. Furthermore, in order to make a comparison between both HD and DD techniques, a DD of data via WRF is also applied to obtain the same fine resolution as the HD technique. The outputs from both the DD and HD techniques are calibrated and validated over the study region (Trinh *et al.* 2021), then input into the hydrological watershed model-WEHY (Chen *et al.* 2004a, 2004b; Kavvas *et al.* 2004) for estimation of flow data. The Sai Gon–Dong Nai River Basin (SG-DN) is selected for this study, because it is the third-largest river basin in Vietnam including high value industrial zones; and it is therefore necessary to apply advanced technologies to investigate severe flood processes and model realistic historical flood events for this region.

2. STUDY AREA

The selected watershed, the Sai Gon–Dong Nai Rivers Basin (SG-DN) as shown in Figure 1, feeds the largest inland river in Vietnam, ranking third-largest in the country after the Mekong and Red River water systems. The SG-DN Rivers have become an important source of hydropower, including many hydropower plants and large amounts of water resources used for all southern provinces of Vietnam. Natural impacts from meteorological factors have caused many difficulties for socio-economic development activities in the basin. The SG-DN has a complex terrain system including mountainous and delta

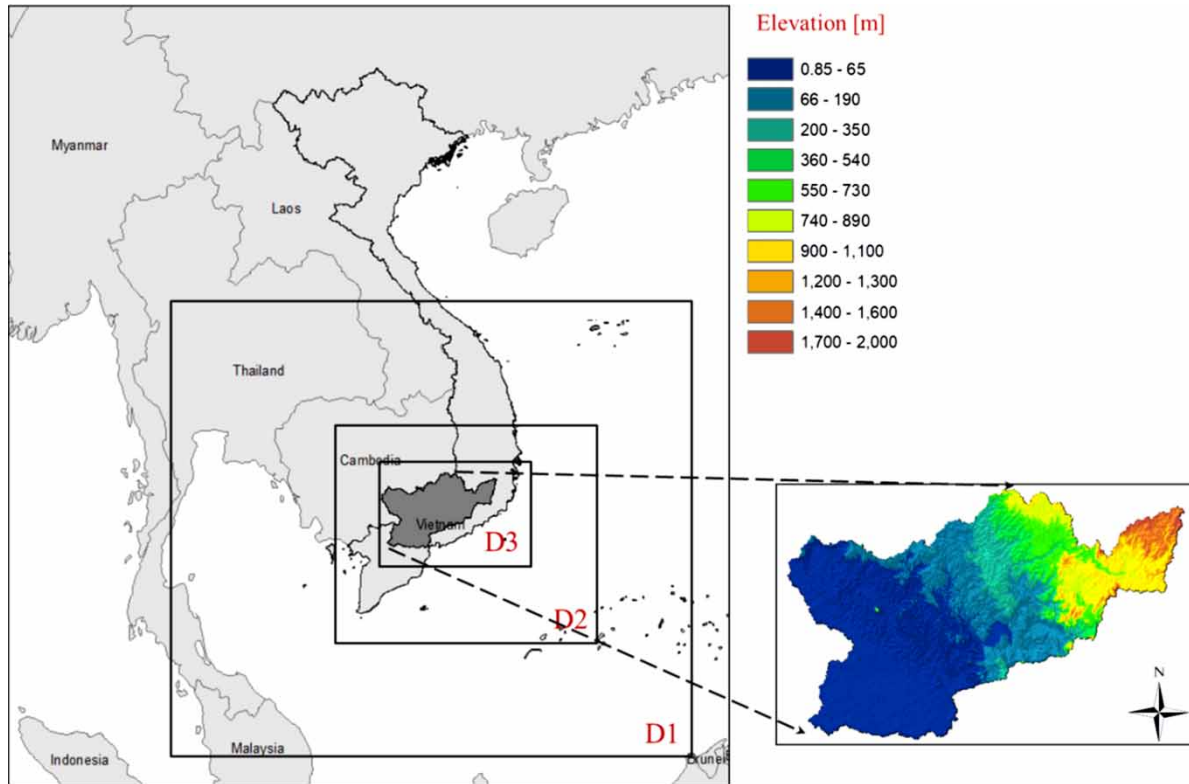


Figure 1 | Plan view of the Sai Gon–Dong Nai Rivers Basin in Vietnam.

regions with tropical heavy rainfall experienced from summer monsoon (SMS) and tropical cyclone (TC) systems (Yokoi & Matsumoto 2008; Nguyen-Thi *et al.* 2012).

The SG-DN Basin covers the provinces of Lam Dong, Binh Phuoc, Binh Duong, Dong Nai, Dak Nong, Long An, Tay Ninh, and Ho Chi Minh City, and parts of the provinces of Ninh Thuan, Binh Thuan, and Ba Ria-Vung Tau with a total catchment area of about 44,500 km². The SG-DN Basin includes two main river systems, the Sai Gon and Dong Nai Rivers. This area features complex terrain including mountainous and delta regions with elevations from 2–2,291 m. Along with an important source of hydropower, the SG-DN basin also includes a number of important industrial zones. The region's atmospheric condition falls in a tropical monsoon climate experiencing a wet summer from late May through early November with an average annual rainfall of about 1,800 mm, and humidity of 78–82%. The land use conditions of the watershed are varied, including agricultural, forested, and urban areas.

3. METHODOLOGY AND IMPLEMENTATION

This study introduces a new technique to provide reliable flow data by coupling physically based numerical atmospheric-hydrologic and machine learning models. Three global reanalysis data sets are used for this technique's input, including ECMWF – Atmospheric Reanalysis coarse climate data of the 20th century (ERA-20C, <https://rda.ucar.edu/datasets/ds626.0>; Poli *et al.* 2016), ECMWF – Reanalysis Interim (ERA-Interim, <https://rda.ucar.edu/datasets/ds627.0>; Berrisford *et al.* 2009; Dee *et al.* 2011), and Climate Forecast System Reanalysis (CFSR, <https://rda.ucar.edu/datasets/ds093.0>; Saha *et al.* 2010; Wang *et al.* 2011). These three data sets provide three-dimensional data and uniformly cover the globe at a spatial resolution of 1.25° (ERA-20C), 0.75° (ERA-Interim), and 0.5° (CFSR).

The Weather Research and Forecasting (WRF) model is selected as the regional climate model to dynamically downscale the three global analysis data sets. The dynamically downscaled atmospheric data are then downscaled to a finer resolution over the studied watershed using an ANN model. Before implementing the hydrological model, the combined downscaling technique needs to be calibrated and validated. After successful validation, the outputs of the downscaling models are used as the hydrological model's input, whose process is shown in Figure 2.

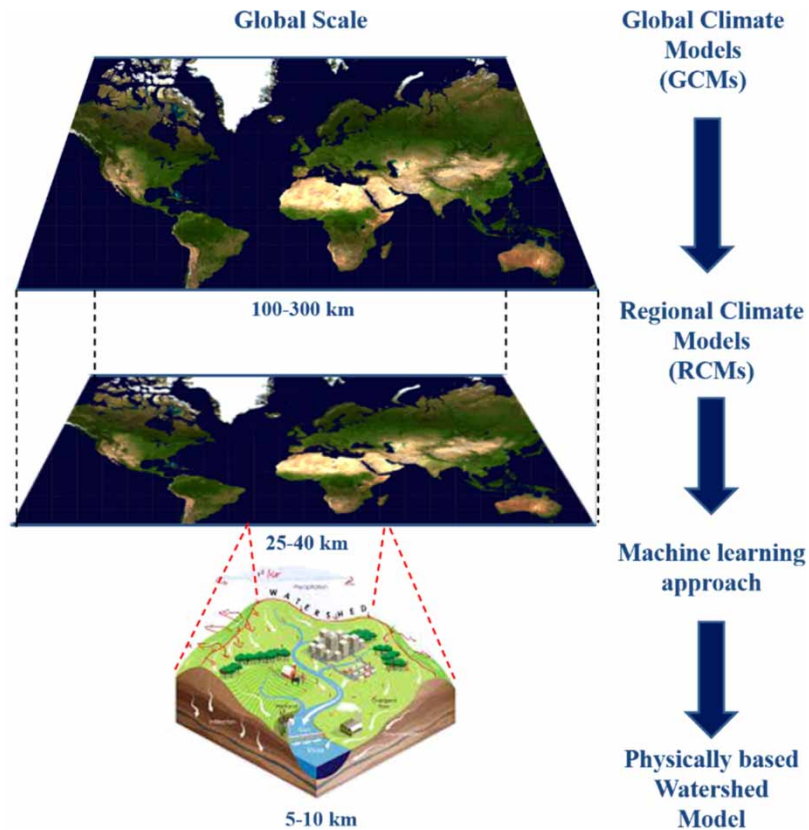


Figure 2 | Methodology for hybrid modeling flow data.

In summary, there are four main steps in developing this hybrid methodology:

1. Implementation and validation of the physically based numerical atmospheric model, WRF, over the target watershed for the three different reanalysis data sets;
2. Implementation and validation of the ANN model with its input obtained from WRF's outputs;
3. Implementation of the hydrology model over the target watershed with its input provided from the hybrid downscaling technology;
4. Calibration and validation of the hydrology model over the target watershed with its input provided by the hybrid downscaling technology;

In-depth description of each step is presented in the following sections.

3.1. Hybrid downscaling implementation and validation

According to *Trinh et al. (2021)*, the WRF model and ANN architecture with back-propagation algorithm were implemented over the SG-DN Rivers Watershed with the initial and lateral boundary conditions provided from three reanalysis data sets including ERA-Interim, ERA-20C, and CFSR. The implemented WRF and ANN models include the three domains as shown in *Figure 1*. Domain 1 (D1) is the outer domain with a spatial resolution of 81 km (21×18 horizontal grid points). Domain 2 (D2) is the inner domain with a resolution of 27 km (27×24 horizontal grid points). *Tables 1* and *2* show WRF's selected parameterization options for each reanalysis data set.

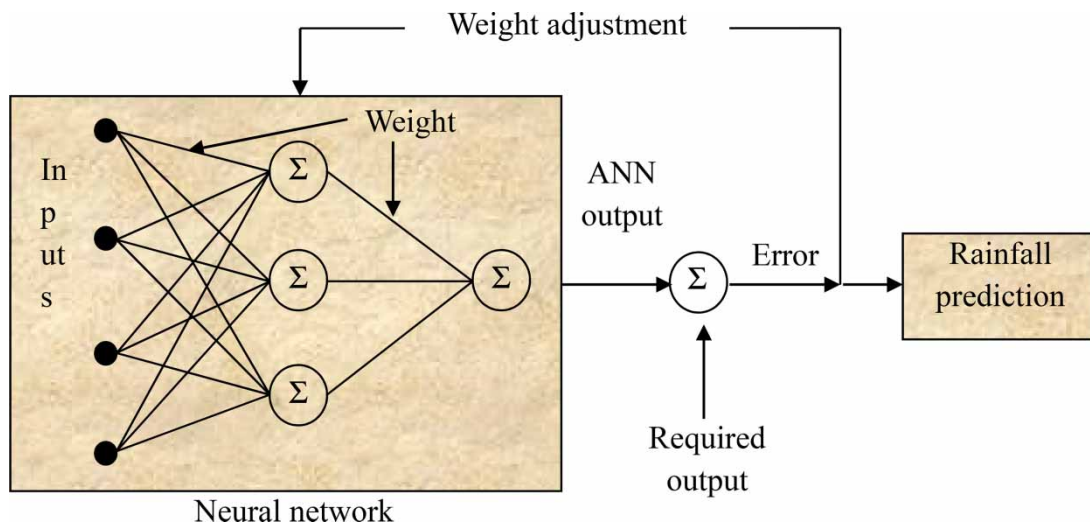
The downscaled atmospheric data from WRF were then input into the ANN model to further downscale to the innermost domain, D3, with spatial resolution of 9 km (48×33 horizontal grid points). The selected ANN architecture is comprised of three layers (input layer, hidden layer, and output layer) that are interconnected by synapse weights (see *Figure 3*). The number of nodes of the hidden layer was selected ranging from $(2n+1)$ to $(2n^{0.5}+m)$, where n is the number of input nodes and m is the number of output nodes (*Fletcher & Goss, 1993*).

Table 1 | WRF model configuration for ERA-20C and ERA-interim

WRF model configuration	Selected option
Microphysics processes	WSM3 Hong <i>et al.</i> (2004, MWR)
Cumulus parameterization	New SAS Han & Pan (2011, Weather Forecasting)
Planetary boundary layer scheme	BouLac scheme (Bougeault & Lacarrere 1989)
Radiation scheme	New Goddard scheme (Chou & Suarez 1999)
Surface scheme	RUC Land Surface Model (Benjamin <i>et al.</i> 2004)

Table 2 | WRF model configuration for CFSR

WRF model configuration	Selected option
Microphysics processes	SBU-YLin, Lin & Colle (2011, MWR)
Cumulus parameterization	New SAS Han & Pan (2011, Weather Forecasting)
Planetary boundary layer scheme	BouLac scheme (Bougeault & Lacarrere 1989)
Radiation scheme	New Goddard scheme (Chou & Suarez 1999)
Surface scheme	RUC Land Surface Model (Benjamin <i>et al.</i> 2004)

**Figure 3** | The architecture of ANN model.

The training phase of the ANN model serves to adjust the weights to minimize the difference between the network outputs (predictands) and the observation data. In this study, a gridded daily precipitation data set with 0.1° resolution (VNGP) is used for the ANN model training and validation. The Vietnam Gridded Precipitation (VNGP) data set has been published since 2016 and is widely used as a reliable observation data set (Nguyen *et al.* 2016). The development of VNGP was based on the Spheremap interpolation technique from 481 rain gauges covering the whole of Vietnam with a resolution of 0.1° . After interpolating, the VNGP was validated against gauge observations through correlations, mean absolute errors, root-mean-square errors, and spatial distribution. It is noted that VNGP is available from January 1980 to December 2010. VNGP is currently available at the Data Integration and Analysis System (DIAS, <https://diasjp.net/en>).

Table 3 shows the candidate predictors for the input layer of the ANN model. These are large-scale atmospheric variables that were simulated by the WRF model for domain D2.

To test the HD technique for SG-DN, a validation of the WRF and ANN application was performed over the target watershed. The validation is performed by comparing the model simulations against the corresponding observations.

3.2. Implementation of a physically based watershed model

This methodology can be implemented with any hydrologic model although it is recommended to use a physically based hydrologic model as mentioned in the Introduction. The WEHY model was selected in the application of this new method to provide reliable flow data when coupled with the HD technique. The WEHY model is a physically based model derived from the conservation equations of mass, momentum, and/or energy for water flows in various domains (Chen *et al.* 2004a, 2004b; Kavvas *et al.* 2004, 2006, 2013). Hydrologic modeling is carried out for the period from 1980 to 2010. The WEHY model's input requires both atmospheric data and physical surface information such as topography, soil, and land cover. The atmospheric data are the result of the HD method executed with three different global reanalysis data sets, as described in the previous section. Physical surface information is needed for the hydrologic model implementation. First, the ASTER Global DEM with the spatial resolution of 30 m (Tachikawa *et al.* 2011) is used to delineate -SG-DN Watershed. This step involves the delineation of hillslopes and streamflow networks within SG-DN. Based on the Geographical Information System (GIS) technique, the delineation provides 152 MCUs and 76 stream networks as shown in Figure 4.

Table 3 | Candidate predictors for the input layer of the ANN model (Trinh *et al.* 2021)

Variables	Unit	Pressure layer (hPa)
Precipitation flux	mm/day	Surface
Meridional wind velocity	m/s	700, 810, 910
Zonal wind velocity	m/s	700, 810, 910
Vertical pressure velocity	Pa/s	700, 810, 910
Total	Variables	10



Figure 4 | Delineated hillslopes, stream networks for SG-DN Watershed, and available flow observation stations.

Soil parameters and land use/cover were retrieved from Soilgrids 1 km (a global 3D soil information system at 1 km resolution) (Hengl *et al.* 2014; Trinh *et al.* 2018), and Global Land Cover Characterization (GLCC) data set (Loveland *et al.* 2000). Delineated soil parameters include mean soil hydraulic conductivity (cm/h), pore size distribution index, soil depth, and bubbling pressure. Delineated land cover parameters are vegetation root depth, roughness height, albedo, emissivity, and leaf area index. For a review of soil and land cover delineation, see Chen *et al.* (2004a) and Kavvas *et al.* (2004), respectively.

Based on the delineated MCUs for SG-DN Watershed, soil and land cover parameters were processed and estimated for each hillslope. WEHY soil parameter maps for soil hydraulic conductivity, mean pore size, bubbling pressure, and soil depth are depicted in Figure 5. Land parameter maps for vegetation root depth and leaf area index (July) are depicted in Figure 6.

4. RESULTS AND DISCUSSION

After the estimation of the geomorphologic, land use/cover, and soil parameters over SG-DN Watershed, the WEHY model is calibrated and validated. Calibration involves the evaluation of parameter sets used in the model. Parameter inputs are provided from HD under the three selected reanalysis data sets. For model validation, the simulated runoff is compared with the corresponding observations without a rainfall-runoff fitting exercise. It is noted that the WEHY model is fully physically based, and almost all parameters were estimated from GIS databases such as existing land, soil, and topography data. However, there are still a few parameters such as stream widths and roughness coefficients, soil moisture conditions, and slope at stream reach segments that need to be calibrated by comparing simulated results against available observation flow data over the target watershed. There are three available hydrologic stations in the SG-DN river system: Dau Tieng, Phuoc Hoa, and Tri An Stations, as shown in Figure 4. The ground stations of Dau Tieng and Tri An are important stations because they are located near the Dau Tieng and Tri An Dams which began operations in 1985 and 1987, respectively. The daily observation data at Tri An, Dau Tieng, and Phuoc Hoa Stations are available for the periods of 1980–2010, 1990–1999, and 1980–1990, respectively. However, it is noted that the observation at the Dau Tieng Station is the inflow data of the Dau Tieng Dam,

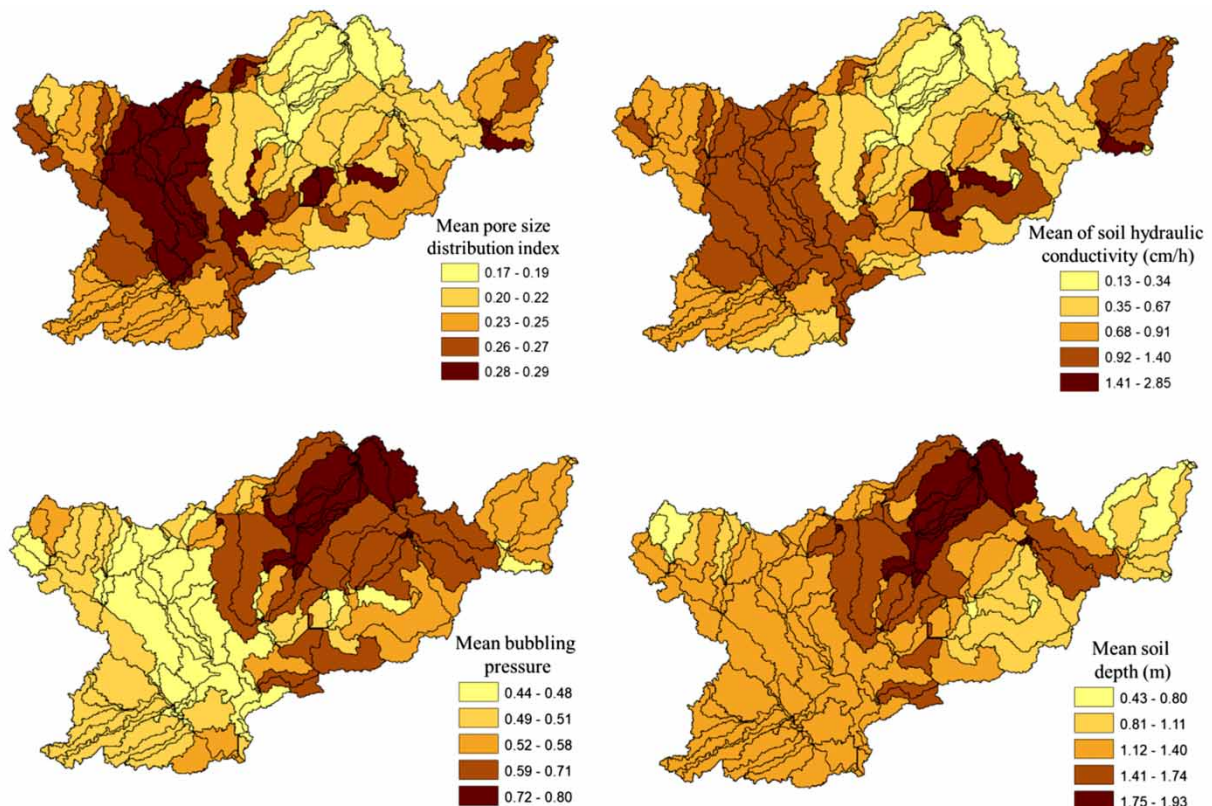


Figure 5 | Computed soil hydraulic parameter maps for SG-DN Watershed.

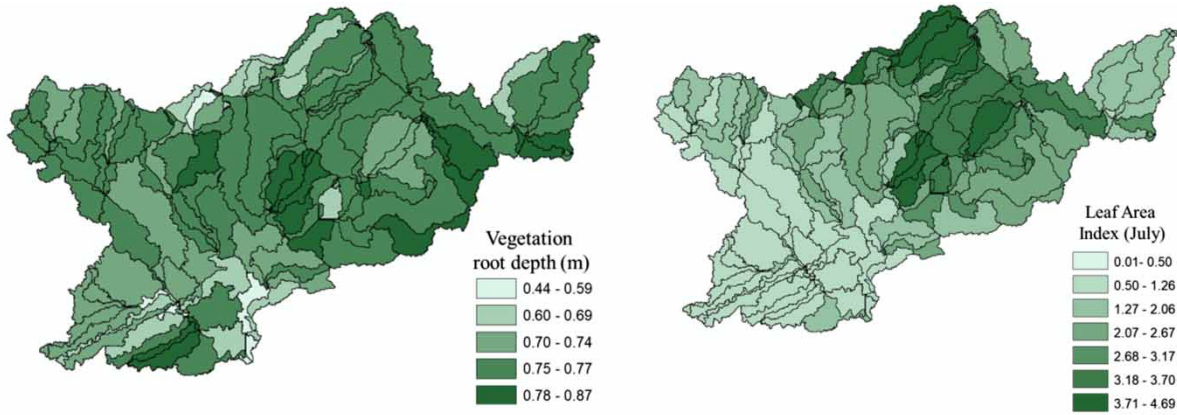


Figure 6 | Estimated land surface parameter maps for SG-DN river watershed.

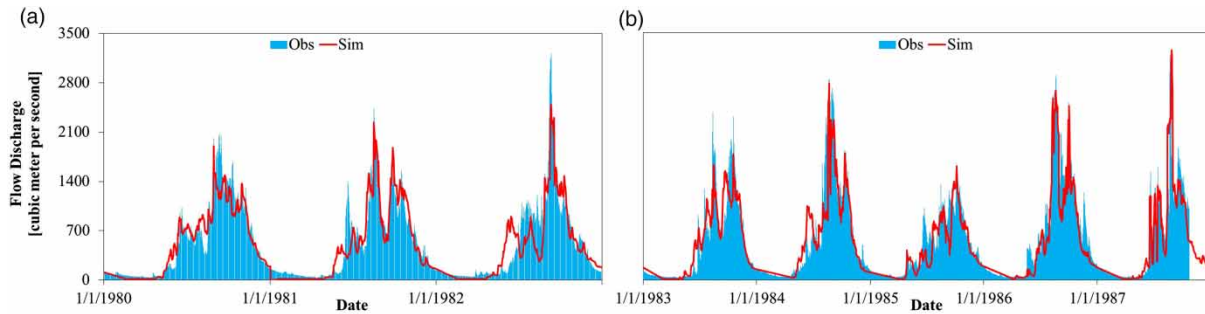


Figure 7 | Comparison of the daily mean discharge by WEHY using observation rainfall data and the corresponding observations at Tri An during (a) the calibration period, January 1980 to December 1982 and (b) the validation period, January 1983 to December 1987.

Table 4 | Summary statistics for comparison of the daily mean discharge at Tri An station using observation rainfall data

Evaluation statistics	Input from observation rainfall data	
	Calibration (1980–1982)	Validation (1983–1987)
Mean by Observation (m ³ /s)	512.41	552.61
Mean by Simulation (m ³ /s)	487.24	503.72
Standard Deviation by Observation (m ³ /s)	539.4	613.15
Standard Deviation by Simulation (m ³ /s)	496.15	561.72
Correlation Coefficient	0.92	0.91
Nash Coefficient	0.83	0.82

while the location of Tri An Station is downstream of its reservoir. Thus, the calibration and validation periods for Tri An and Phuoc Hoa Stations were selected before the construction of Tri An Reservoir. The calibration and validation of the daily mean discharges at Tri An Station are shown in Figure 7. Atmospheric inputs for the calibration and validation at Tri An used WRF’s outputs.

The visual comparison between the model simulations and corresponding observations shows that they match quite well at both the rising and recession time segments. The results of this analysis are presented in Table 4. The calibration and validation provide ‘very good’ range in correlation coefficients (0.91–0.92), and Nash–Sutcliffe efficiencies (0.82–0.83), based on daily mean discharge comparison (Moriasi *et al.* 2015).

From the validation of WEHY for the SG-DN Watershed, it shows a possibility to simulate flow conditions throughout the watershed with atmospheric input from DD and HD under ERA-20C, CFSR, and ERA-Interim. The comparisons between

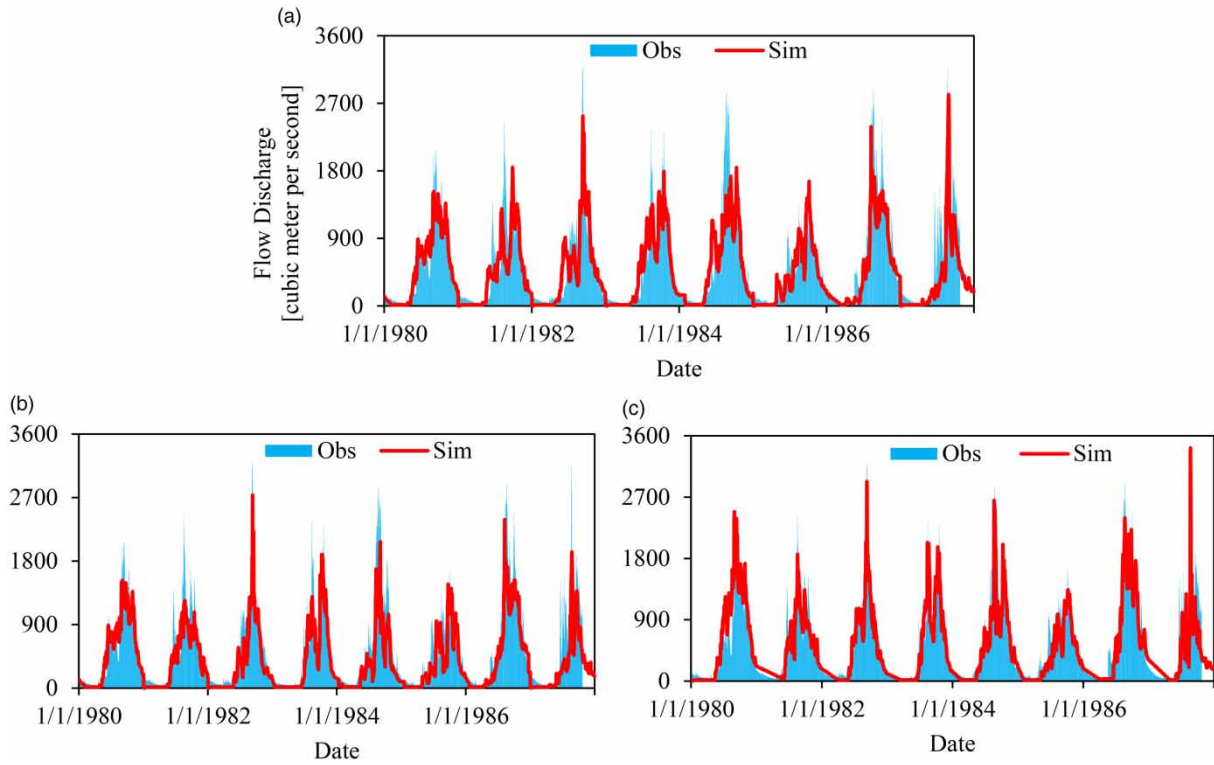


Figure 8 | Comparison of the daily mean discharge by WEHY simulations under HD using (a) ERA-Interim, (b) ERA-20C, and (c) CFSR, and the corresponding observations at Tri An during the period 1980–1987.

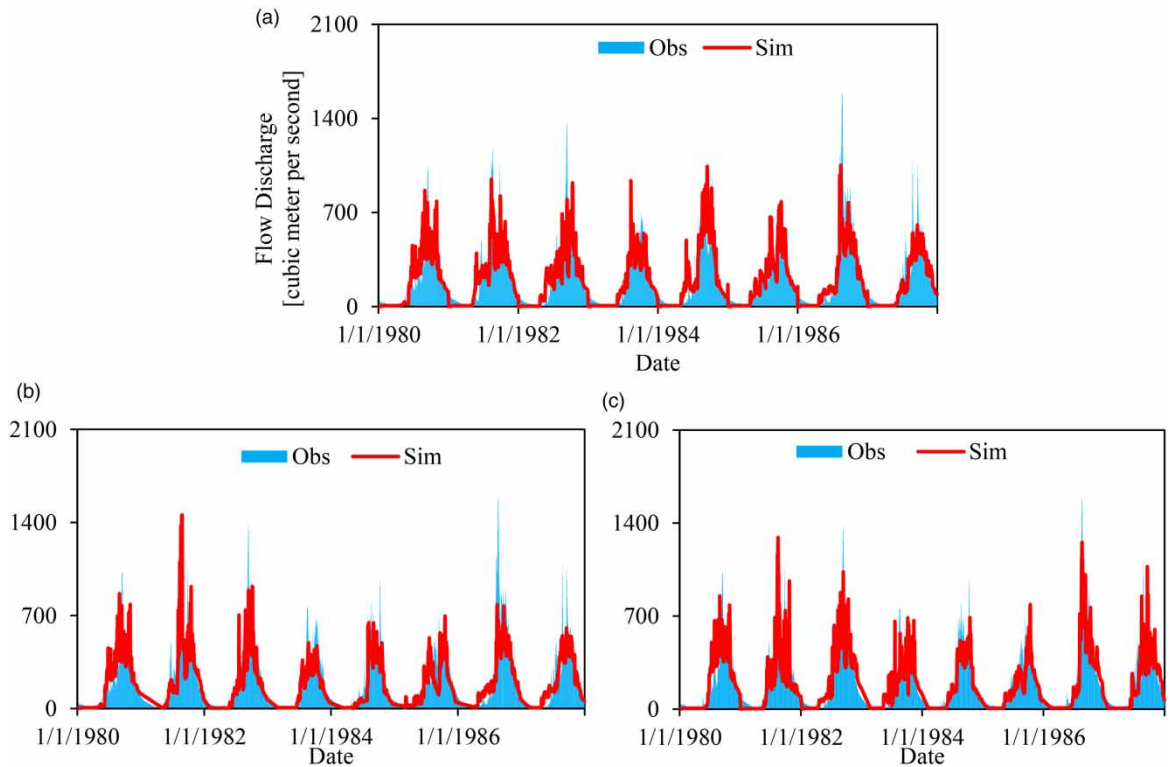


Figure 9 | Comparison of the daily mean discharge by WEHY simulations under HD using (a) ERA-Interim, (b) ERA-20C, and (c) CFSR, and the corresponding observations at Phuoc Hoa during the period 1980–1987.

simulation and observation are shown in Figures 8 and 9. These comparisons are products of the HD technique coupled with the WEHY model.

Tables 5–7 show evaluation statistics for comparison of the daily mean discharge at Tri An, Dau Tieng, and Phuoc Hoa Stations under DD using the ERA-Interim, ERA-20C, and CFSR data sets. The DD technique combined with the WEHY model provides ‘satisfactory’ results with correlation coefficients ranging from 0.63 to 0.91 (Moriassi *et al.* 2015). The model results under CFSR were closer to observations than the results based on the ERA-Interim and ERA-20C.

Tables 8–10 show evaluation statistics for comparison of the daily mean discharge at Tri An, Dau Tieng, and Phuoc Hoa stations under HD using ERA-Interim, ERA-20C, and CFSR data sets. Overall, the results obtained from the HD technique are better than those obtained from DD technique-based model performance parameters (correlation and NSE coefficients).

Table 5 | Summary statistics for comparison of the daily mean discharge at Tri An, Dau Tieng, and Phuoc Hoa stations under dynamical downscaling using ERA-interim

Evaluation statistics	Input from dynamical downscaling ERA-Interim		
	Tri An (1980–1987)	Phuoc Hoa (1980–1987)	Dau Tieng (1990–1999)
Mean by Observation (m ³ /s)	552.61	209.58	62.28
Mean by Simulation (m ³ /s)	445.72	180.88	45.09
Standard Deviation by Observation (m ³ /s)	586.15	244.65	68.76
Standard Deviation by Simulation (m ³ /s)	509.72	219.16	56.48
Correlation Coefficient	0.855	0.80	0.67
Nash Coefficient	0.71	0.63	0.35

Table 6 | Summary statistics for comparison of the daily mean discharge at Tri An, Dau Tieng, and Phuoc Hoa stations under dynamical downscaling using ERA-20C

Evaluation statistics	Input from dynamical downscaling ERA-20C		
	Tri An (1980–1987)	Phuoc Hoa (1980–1987)	Dau Tieng (1990–1999)
Mean by Observation (m ³ /s)	552.61	209.58	62.28
Mean by Simulation (m ³ /s)	502.19	202.51	52.41
Standard Deviation by Observation (m ³ /s)	586.15	244.65	68.76
Standard Deviation by Simulation (m ³ /s)	501.52	209.94	69.13
Correlation Coefficient	0.84	0.78	0.63
Nash Coefficient	0.70	0.60	0.22

Table 7 | Summary statistics for comparison of the daily mean discharge at Tri An, Dau Tieng, and Phuoc Hoa stations under dynamical downscaling using CFSR

Evaluation statistics	Input from dynamical downscaling CFSR		
	Tri An (1980–1987)	Phuoc Hoa (1980–1987)	Dau Tieng (1990–1999)
Mean by Observation (m ³ /s)	552.61	209.58	62.28
Mean by Simulation (m ³ /s)	478.90	159.94	50.56
Standard Deviation by Observation (m ³ /s)	586.15	244.65	68.76
Standard Deviation by Simulation (m ³ /s)	549.78	209.87	65.62
Correlation Coefficient	0.91	0.87	0.76
Nash Coefficient	0.80	0.71	0.51

As shown in the DD technique, these comparisons show that the model results under CFSR were closer to observations than the ERA-Interim and ERA-20C model results. It is noted that the validation result at the Dau Tieng Station is not as good as the results obtained at Tri An and Phuoc Hoa Stations, nevertheless it is still in a satisfactory range ($NSE > 0.5$ under CFSR data set). The estimation of inflow to the Dau Tieng Station is based on the calculation of the changes in water elevation and observed outflow. Therefore, it is suspected that the Dau Tieng's observed flow data may not be accurate data. In general, these results also confirm that applying the HD technique and the WEHY model (shown in Tables 8–10) gave better results than the DD technique and WEHY model (shown in Tables 5–7). Furthermore, the HD technique is also inexpensive computational demand with respect to computer resources and time consumption as shown in Table 11.

Table 8 | Summary statistics for comparison of the daily mean discharge at Tri An, Dau Tieng, and Phuoc Hoa stations under hybrid downscaling using ERA-interim

Evaluation statistics	Input from hybrid downscaling ERA-Interim		
	Tri An (1980–1987)	Phuoc Hoa (1980–1987)	Dau Tieng (1990–1999)
Mean by Observation (m^3/s)	552.61	209.58	62.28
Mean by Simulation (m^3/s)	450.75	186.57	38.54
Standard Deviation by Observation (m^3/s)	586.15	244.65	68.76
Standard Deviation by Simulation (m^3/s)	482.16	229.76	49.94
Correlation Coefficient	0.87	0.85	0.71
Nash Coefficient	0.74	0.71	0.40

Table 9 | Summary statistics for comparison of the daily mean discharge at Tri An, Dau Tieng, and Phuoc Hoa stations under hybrid downscaling using ERA-20C

Evaluation statistics	Input from hybrid downscaling ERA-20C		
	Tri An (1980–1987)	Phuoc Hoa (1980–1987)	Dau Tieng (1990–1999)
Mean by Observation (m^3/s)	552.61	209.58	62.28
Mean by Simulation (m^3/s)	424.15	189.08	46.51
Standard Deviation by Observation (m^3/s)	586.15	244.65	68.76
Standard Deviation by Simulation (m^3/s)	449.35	199.18	59.86
Correlation Coefficient	0.89	0.84	0.65
Nash Coefficient	0.73	0.67	0.35

Table 10 | Summary statistics for comparison of the daily mean discharge at Tri An, Dau Tieng, and Phuoc Hoa stations under hybrid downscaling using CFSR

Evaluation statistics	Input from hybrid downscaling CFSR		
	Tri An (1980–1987)	Phuoc Hoa (1980–1987)	Dau Tieng (1990–1999)
Mean by Observation (m^3/s)	552.61	209.58	62.28
Mean by Simulation (m^3/s)	481.58	179.91	48.38
Standard Deviation by Observation (m^3/s)	586.15	244.65	68.76
Standard Deviation by Simulation (m^3/s)	528.47	202.99	63.37
Correlation Coefficient	0.92	0.88	0.76
Nash Coefficient	0.82	0.74	0.51

Table 11 | Comparisons of computational demand with respect to computer resources and time consumption between the DD and HD techniques (based on the same CPU)

Intel(R) Xeon(R) CPU E5-2687 W v4 @ 3.00 GHz x 2 (24 cores, Memory: 64 GB, Storage: 3 T)	Dynamic downscaling method (DD)	Hybrid downscaling method (HD)
Time consumption for 1 year simulation (using CFSR)	~30 h	~6 h
Output storage for 1 year simulation (using CFSR)	2.5 GB	0.8 GB

5. CONCLUSION

This study introduces a possible methodology to improve the availability and reliability of simulated hydrological data using HD, and verifying by a hydrologic model. Based on the implemented model, it is possible to produce both atmospheric and hydrologic data at different time resolutions (hourly, daily, and monthly). This study requires calibration and validation processes for both atmospheric and hydrologic components. The global atmospheric input including ERA-Interim, ERA-20C, and CFSR are used for initial and boundary conditions for DD by means of the WRF model (Trinh *et al.* 2021). The ANN model follows to further downscale the WRF outputs to a finer resolution over the studied watershed. The validation of both the WRF and ANN models are in the 'good' range. The physically based watershed model-WEHY is implemented based on GIS databases. The simulations matched the flow observation data well with respect to magnitude at both the rising and recession time segments. In the comparisons among the three selected reanalysis data sets, the best calibration and validation results were obtained from the CFSR data set. Regarding data resolution, the order ranging from coarse to fine among the three selected data sets are the ERA-20C (1.25°), then ERA-Interim (0.75°) and CFSR (0.5°). Since CFSR has a better resolution than ERA-20C and ERA-Interim, it is more reliable in connection to the surface measurements. In addition, the temperature and precipitation data of CFSR (Saha *et al.* 2010) are based on the meteorological model in combination with data from satellite-based observing systems and surface observation. The direct assimilation of observations represents one of the major improvements of the CFSR data set.

Correlation coefficients (ranging from 0.87 to 0.92) and Nash–Sutcliffe efficiencies (0.73–0.82) are quite satisfactory when the proposed method is applied utilizing the CFSR data set according to daily comparisons at Tri An (Tables 8–10). These results are closer to the observation data than those using only the DD technique combined with the WEHY model. In general, these results also confirm that applying the HD technique and the WEHY model gave better results than the DD technique and the WEHY model. Furthermore, the HD technique also is inexpensive computational demand with respect to computer resources and time consumption as shown in Table 11. Based on the proposed method, it is possible to reconstruct, project, and forecast hydroclimate information over the implemented watershed. Furthermore, this application can also apply to flood and drought studies because the results are produced at different time resolutions (Trinh *et al.* 2016, 2017) (i.e., hourly, daily, and monthly). With the validated model, future studies will focus on modeling hydrologic conditions with inputs provided from future projections such as CMIP5 and CMIP6 scenarios. This new approach can be applied widely in many parts of the world where local observation data are sparse and ungauged.

ACKNOWLEDGEMENTS

This research was funded by Ho Chi Minh City's Department of Science and Technology (HCMC-DOST) and the Institute for Computational Science and Technology (ICST) under the grant number 16/2020/HĐ-QPTKHCN. The authors also would like to thank the anonymous reviewers for their valuable and constructive comments to improve our manuscript.

DATA AVAILABILITY STATEMENT

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

REFERENCES

- Anh, T. & Taniguchi, K. 2018 Coupling dynamical and statistical downscaling for high-resolution rainfall forecasting: case study of the Red River Delta, Vietnam. *Progress in Earth and Planetary Science* 5 (1), 1–18.
- Benjamin, S. G., D ev enyi, D., Weygandt, S. S., Brundage, K. J., Brown, J. M., Grell, G. A., Kim, D., Schwartz, B. E., Smirnova, T. G., Smith, T. L. & Manikin, G. S. 2004 An hourly assimilation–forecast cycle: the RUC. *Monthly Weather Review* 132 (2), 495–518.

- Berrisford, P., Dee, D. P. K. F., Fielding, K., Fuentes, M., Kallberg, P., Kobayashi, S. & Uppala, S. 2009 The ERA-interim archive. *ERA Report Series 1*, 1–16.
- Bjerklie, D. M., Birkett, C. M., Jones, J. W., Carabajal, C., Rover, J. A., Fulton, J. W. & Garambois, P. A. 2018 Satellite remote sensing estimation of river discharge: application to the Yukon River Alaska. *Journal of Hydrology* **561**, 1000–1018.
- Bougeault, P. & Lacarrere, P. 1989 Parameterization of orography-induced turbulence in a mesobeta-scale model. *Monthly Weather Review* **117** (8), 1872–1890.
- Bui, D. T., Hoang, N. D., Pham, T. D., Ngo, P. T. T., Hoa, P. V., Minh, N. Q., Tran, X. T. & Samui, P. 2019 A new intelligence approach based on GIS-based multivariate adaptive regression splines and metaheuristic optimization for predicting flash flood susceptible areas at high-frequency tropical typhoon area. *Journal of Hydrology* **575**, 314–326.
- Chen, Z. Q., Kavvas, M. L., Yoon, J. Y., Dogrul, E. C., Fukami, K., Yoshitani, K. & Matsuura, T. 2004a Geomorphologic and soil hydraulic parameters for watershed environmental hydrology (WEHY) model. *Journal of Hydrologic Engineering* **6** (465), 465–479.
- Chen, Z., Kavvas, M., Fukami, K., Yoshitani, J. & Matsuura, T. 2004b Watershed environmental hydrology (WEHY) model: model application. *Journal of Hydrologic Engineering* **9** (6), 480–490.
- Chen, Z. R., Kavvas, M. L., Ohara, N., Anderson, M. L. & Yoon, J. 2011 Coupled regional hydroclimate model and its application to the Tigris-Euphrates basin. *Journal of Hydrologic Engineering* **16** (12), 1059–1070.
- Chou, M. D. & Suarez, M. J. 1999 *A Solar Radiation Parameterization for Atmospheric Studies*, Vol. 15. National Aeronautics and Space Administration, Goddard Space Flight Center Greenbelt Maryland, Laboratory for Atmospheres, Climate and Radiation Branch, p. 38.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, D. P. & Bechtold, P. 2011 The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society* **137** (656), 553–597.
- Dudley, N. J. 1988 A single decision-maker approach to irrigation reservoir and farm management decision making. *Water Resources Research* **24** (5), 633–640.
- Fletcher, D. & Goss, E. 1993 Forecasting with neural networks: an application using bankruptcy data. *Information and Management* **24** (3), 159–167.
- Gorguner, M., Kavvas, M. L. & Ishida, K. 2019 Assessing the impacts of future climate change on the hydroclimatology of the Gediz Basin in Turkey by using dynamically downscaled CMIP5 projections. *Science of the Total Environment* **648**, 481–499.
- Han, J. & Pan, H. L. 2011 Revision of convection and vertical diffusion schemes in the NCEP global forecast system. *Weather and Forecasting* **26** (4), 520–533.
- Hengl, T., de Jesus, J. M., MacMillan, R. A., Batjes, N. H., Heuvelink, G. B., Ribeiro, E., SamuelRosa, A., Kempen, B., Leenaars, J. G., Walsh, M. G. & Gonzalez, M. R. 2014 Soilgrids1 km—global soil information based on automated mapping. *PLoS ONE* **9** (8), e105992.
- Hirpa, F. A., Salamon, P., Beck, H. E., Lorini, V., Alfieri, L., Zsoter, E. & Dadson, S. J. 2018 Calibration of the global flood awareness system (GloFAS) using daily streamflow data. *Journal of Hydrology* **566**, 595–606.
- Hong, S. Y., Dudhia, J. & Chen, S. H. 2004 A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. *Monthly Weather Review* **132** (1), 103–120.
- Hu, Y., Maskey, S., Uhlenbrook, S. & Zhao, H. 2011 Streamflow trends and climate linkages in the source region of the Yellow River, China. *Hydrological Processes* **25** (22), 3399–3411.
- Ishida, K., Kavvas, M. L., Jang, S., Chen, Z. Q., Ohara, N. & Anderson, M. L. 2015 Physically based estimation of maximum precipitation over three watersheds in Northern California: atmospheric boundary condition shifting. *Journal of Hydrologic Engineering* **20** (4), 04014052.
- Kavvas, M. L., Chen, Z. Q., Dogrul, C., Yoon, J. Y., Ohara, N., Liang, L., Aksoy, H., Anderson, M. L., Yoshitani, J., Fukami, K. & Matsuura, T. 2004 Watershed environmental hydrology (WEHY) model based on upscaled conservation equations: hydrologic module. *Journal of Hydrologic Engineering* **9** (6), 450–464.
- Kavvas, M. L., Yoon, J., Chen, Z. Q., Liang, L., Dogrul, E. C., Ohara, N., Aksoy, H., Anderson, M. L., Reuter, J. & Hackley, S. 2006 Watershed environmental hydrology model: environmental module and its application to a California watershed. *Journal of Hydrologic Engineering* **3** (261), 261–272.
- Kavvas, M. L., Kure, S., Chen, Z. Q., Ohara, N. & Jang, S. 2013 WEHY-HCM for modeling interactive atmospheric-hydrologic processes at watershed scale. I: model description. *Journal of Hydrologic Engineering* **18** (10), 1262–1271.
- Kure, S., Jang, S., Ohara, N., Kavvas, M. L. & Chen, Z. Q. 2013 WEHY-HCM for modeling interactive atmospheric-hydrologic processes at watershed scale. II: model application to ungauged and sparsely gauged watersheds. *Journal of Hydrologic Engineering* **18** (10), 1272–1281.
- Lin, Y. & Colle, B. A. 2011 A new bulk microphysical scheme that includes riming intensity and temperature-dependent ice characteristics. *Monthly Weather Review* **139** (3), 1013–1035.
- Liu, Y., You, M., Zhu, J., Wang, F. & Ran, R. 2019 Integrated risk assessment for agricultural drought and flood disasters based on entropy information diffusion theory in the middle and lower reaches of the Yangtze River, China. *International Journal of Disaster Risk Reduction* **38**, 101194.
- Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, Z., Yang, L. W. M. J. & Merchant, J. W. 2000 Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *International Journal of Remote Sensing* **21** (6–7), 1303–1330.
- Malik, A., Kumar, A. & Singh, R. P. 2019 Application of heuristic approaches for prediction of hydrological drought using multi-scalar streamflow drought index. *Water Resources Management* **33** (11), 3985–4006.

- Milly, P., Julio, B., Malin, F., Robert, M., Zbigniew, W., Dennis, P. & Ronald, J. 2007 Stationarity is dead. *Ground Water News Views* 4 (1), 6–8.
- Moriasi, D. N., Gitau, M. W., Pai, N. & Daggupati, P. 2015 Hydrologic and water quality models: performance measures and evaluation criteria. *Transactions of the ASABE* 2015 (58), 1763–1785.
- Nguyen-Thi, H. A., Matsumoto, J., Ngo-Duc, T. & Endo, N. 2012 A climatological study of tropical cyclone rainfall in Vietnam. *SOLA* 8, 41–44.
- Nguyen-Xuan, T., Ngo-Duc, T., Kamimera, H., Trinh-Tuan, L., Matsumoto, J., Inoue, T. & Phan-Van, T. 2016 The Vietnam gridded precipitation (VnGP) dataset: construction and validation. *The Scientific Online Letters on the Atmosphere (SOLA)* 12, 291–296.
- Pierce, D. W., Das, T., Cayan, D. R., Maurer, E. P., Miller, N. L., Bao, Y., Kanamitsu, M., Yoshimura, K., Snyder Mark, A., Sloan Lisa, C., Franco, G. & Tyree, M. 2012 Probabilistic estimates of future changes in California temperature and precipitation using statistical and dynamical downscaling. *Climate Dynamics* 40 (3–4), 839–856.
- Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., Laloyaux, P., Tan, D. G., Peubey, C., Thépaut, J. N. & Trémolet, Y. 2016 ERA-20C: an atmospheric reanalysis of the twentieth century. *Journal of Climate* 29 (11), 4083–4097.
- Saha, S., Moorthi, S., Pan, H. L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D. & Liu, H. 2010 The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological Society* 91 (8), 1015–1058.
- Salas, J. D. 1993 Analysis and modeling of hydrologic time series, Chapter 19. In: *Handbook of Hydrology* (Maidment, D. R., ed.). McGraw-Hill, New York.
- Salathé Jr., E. P. 2003 Comparison of various precipitation downscaling methods for the simulation of streamflow in a rainshadow river basin. *International Journal of Climatology: A Journal of the Royal Meteorological Society* 23 (8), 887–901.
- Smith, L. C. 1997 Satellite remote sensing of river inundation area, stage, and discharge: a review. *Hydrological Processes* 11 (10), 1427–1439.
- Tachikawa, T., Kaku, M., Iwasaki, A., Gesch, D., Oimoen, M., Zhang, Z., Danielson, J., Krieger, T., Curtis, B., Haase, J. & Abrams, M. 2011 ASTER Global Digital Elevation Model Version 2 – Summary of Validation Results August 31, 2011.
- Trinh, T., Ishida, K., Fischer, I., Jang, S., Darama, Y., Nosacka, J., Brown, K. & Kavvas, M. L. 2016 New methodology to develop future flood frequency under changing climate by means of physically based numerical atmospheric-hydrologic modeling. *Journal of Hydrologic Engineering* 21 (4), 04016001.
- Trinh, T., Ishida, K., Kavvas, M. L., Ercan, A. & Carr, K. 2017 Assessment of 21st century drought conditions at Shasta Dam based on dynamically projected water supply conditions by a regional climate model coupled with a physically-based hydrology model. *Science of the Total Environment* 586, 197–205.
- Trinh, T., Kavvas, M. L., Ishida, K., Ercan, A., Chen, Z. Q., Anderson, M. L., Ho, C. & Nguyen, T. 2018 Integrating global land-cover and soil datasets to update saturated hydraulic conductivity parameterization in hydrologic modeling. *Science of the Total Environment* 631, 279–288.
- Trinh, T., Ho, C., Do, N., Ercan, A. & Kavvas, M. L. 2020 Development of high-resolution 72 h precipitation and hillslope flood maps over a tropical transboundary region by physically based numerical atmospheric-hydrologic modeling. *Journal of Water and Climate Change* 11 (S1), 387–406.
- Trinh, T., Do, N., Nguyen, V. T. & Carr, K. 2021 Modeling high-resolution precipitation by coupling a regional climate model with a machine learning model: an application to Sai Gon-Dong Nai Rivers Basin in Vietnam. *Climate Dynamics* 57, 2713–2735.
- Tu, T., Ercan, A., Carr, K. J., Kavvas, M. L., Trinh, T., Ishida, K., Nosacka, J. & Brown, K. 2020 Coupling hydroclimate-hydraulic-sedimentation models to estimate flood inundation and sediment transport during extreme flood events under a changing climate. *Science of The Total Environment* 740, 140117.
- Tu, T., Ishida, K., Ercan, A., Kiyama, M., Amagasaki, M. & Zhao, T. 2021 Hybrid precipitation downscaling over coastal watersheds in Japan using WRF and CNN. *Journal of Hydrology: Regional Studies* 37, 100921.
- Walton, D. B., Hall, A., Berg, N., Schwartz, M. & Sun, F. 2017 Incorporating snow albedo feedback into downscaled temperature and snow cover projections for California's sierra Nevada. *Journal of Climate* 30 (4), 1417–1438.
- Wang, W., Xie, P., Yoo, S. H., Xue, Y., Kumar, A. & Wu, X. 2011 An assessment of the surface climate in the NCEP climate forecast system reanalysis. *Climate Dynamics* 37 (7), 1601–1620.
- Yokoi, S. & Matsumoto, J. 2008 Collaborative effects of cold surge and tropical depression-type disturbance on heavy rainfall in central Vietnam. *Monthly Weather Review* 136 (9), 3275–3287.
- Zhang, F., Guo, P., Engel, B. A., Guo, S., Zhang, C. & Tang, Y. 2019 Planning seasonal irrigation water allocation based on an interval multi-objective multi-stage stochastic programming approach. *Agricultural Water Management* 223, 105692.

First received 23 August 2021; accepted in revised form 24 November 2021. Available online 9 December 2021