Hydrology Research

Hydrology Research Vol 00 No 0, 1 doi: 10.2166/nh.2022.118

Regionalization of catchment hydrological model parameters for global water resources simulations

Wen-yan Qi[a,b], Jie Chen[a,c,*], Lu Li[d], Chong-Yu Xu[e], Jingjing Li[a], Yiheng Xiang[a] and Shaobo Zhang[a]

[a] State Key Laboratory of Water Resources & Hydropower Engineering Science, Wuhan University, 299 Bayi Road, Wuhan 430072, China
[b] School of Civil Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China
[c] Hubei Provincial Key Lab of Water System Science for Sponge City Construction, Wuhan University, Wuhan, China
[d] NORCE Norwegian Research Centre, Bjerknes Centre for Climate Research, Bergen, Norway
[e] Department of Geosciences, University of Oslo, P.O. Box 1047, Blindern, 0316 Oslo, Norway
*Corresponding author. E-mail: jiechen@whu.edu.cn

ABSTRACT

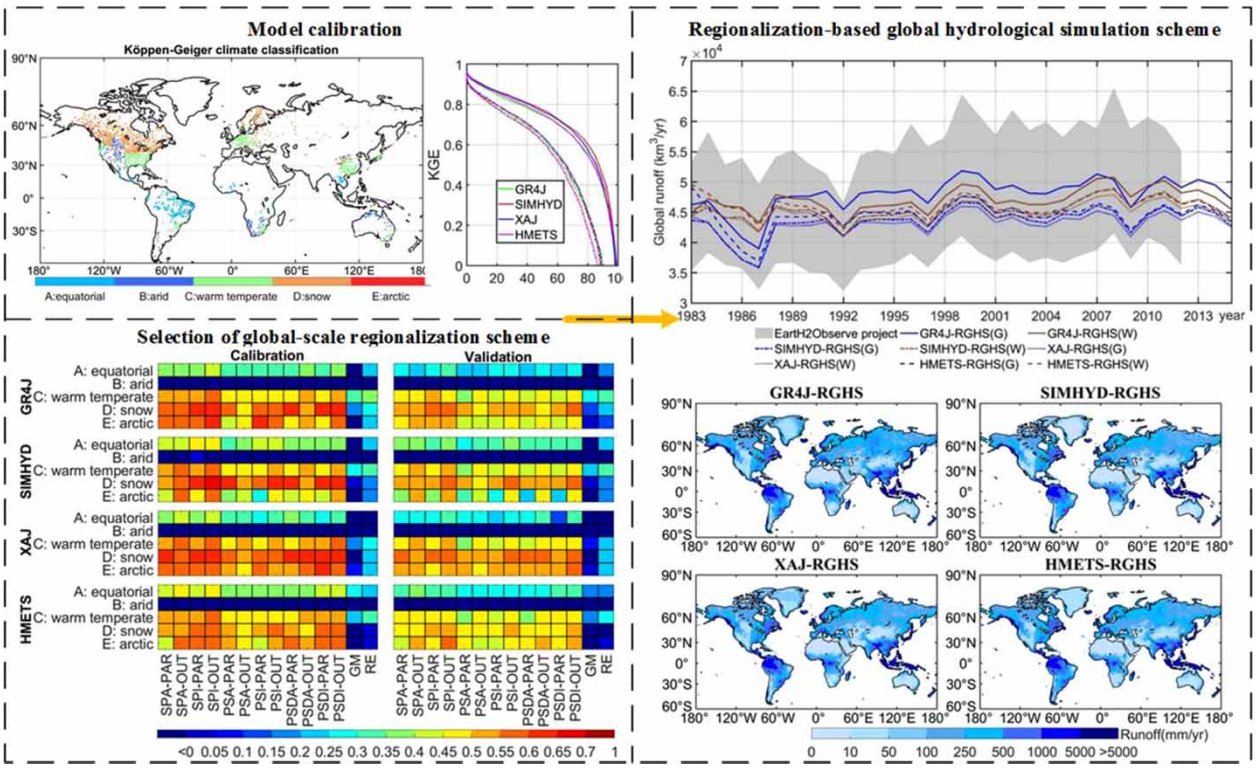
Parameter regionalization of hydrological models is one of the most commonly used methods for hydrological prediction over ungauged catchments. Although there were many regional studies, there is no clear conclusion on the best-performed regionalization method for global hydrological modelling. The objective of this study is to determine an appropriate global-scale regionalization scheme (GSRS) for global hydrological modelling. To this end, the performance of five regionalization methods with two different average options, two weighting approaches, and seven efficiency thresholds (i.e. Kling-Gupta efficiency (KGE) values to measure hydrological model performances) was compared over thousands of catchments based on four conceptual hydrological models. Results of nine global models from the Global Earth Observation for Integrated Water Resource Assessment (EartH2Observe) project were selected to validate the accuracy of GSRS in estimating global runoff. The results show that: (1) Spatial proximity method with the Inverse Distance Weighting method and the output average option offers the best regionalization result when using the KGE ≥ 0.5 as an efficiency threshold for all four hydrological models, (2) the regionalization-based global hydrological simulation schemes (RGHSs), i.e. the proposed GSRS combining with four hydrological models, consistently performs better than the nine global models from EartH2Observe project in the estimation of runoff for most catchments, with varying degrees of improvement in the median, upper and lower quartiles, and whiskers of each performance metric, and (3) the global long-term annual water resources estimated by RGHSs range between 42,592 and 46,810 km^3/yr.

Key words: global hydrological modelling, global water resources estimates, regionalization

HIGHLIGHTS

- The performance of regionalization methods is evaluated with multiple hydrological models globally.
- A global-scale regionalization scheme (GSRS) is proposed for global hydrological modelling.
- The robust regionalization-based global hydrological simulation schemes (RGHSs), i.e. the proposed GSRS combining with four hydrological models, can produce reliable simulations of global water resources.

GRAPHICAL ABSTRACT



1. INTRODUCTION

Water resource is one of the most important natural resources that significantly influences the social and economic development of a region or globe (Parajka *et al.* 2007; Petelet-Giraud *et al.* 2018; Grill *et al.* 2019; Behboudian *et al.* 2021; Naghdi *et al.* 2021). With the development of society and economy, the conflict of water resources supply and demand has been intensified (Müller Schmied *et al.* 2016; Vargas Godoy *et al.* 2021). In addition, the changing climate and increasing water demand due to the growing population lead to a series of environmental and social problems, e.g. frequent floods and droughts, food crisis, and soil salinization (Zhang *et al.* 2020; Li *et al.* 2021; Ridolfi *et al.* 2021; Roushangar *et al.* 2021; Zhou *et al.* 2022). Efficient water resource allocation could help to address water shortage problems around the world, which should be based on a full understanding of the spatial and temporal variation of global water resources (Shrestha *et al.* 2017; Zhang *et al.* 2020). However, due to the lack of observational data around the world, especially in poorly instrumented regions (e.g. the Middle East and West Africa), it is difficult to directly obtain global water information. In addition, problems caused by water conflicts in multinational catchments and virtual water trade both reflect the requirements of continental and global-scale hydrological simulations (Döll *et al.* 2003; Oki & Kanae 2006; Widén-Nilsson *et al.* 2007, 2009).

Global water resources are commonly approximated by using runoff coefficient (Liang & Greene 2020), gauge-based observations (Harris *et al.* 2014; Ghiggi *et al.* 2019), precipitation minus evaporation (Syed *et al.* 2010; Chandanpurkar *et al.* 2017) and water balance modelling, among which water balance modelling is one of the most popular solutions for global water resource estimations (Döll *et al.* 2003; Oki & Kanae 2006; Widén-Nilsson *et al.* 2007, 2009; Beck *et al.* 2016, 2020). Models developed to simulate continental or global water resources can be roughly classified into dynamic global vegetation models (DGVMs), land surface models (LSMs), and global hydrological models (GHMs). Most DGVMs do not include lateral water flows or surface water bodies and can therefore only be used to assess runoff generation but not streamflow discharge (Döll *et al.* 2015). The LSM is commonly used as a component of climate models in simulating the energy and water balance at soil, atmosphere, and vegetation interfaces (Haddeland *et al.* 2011; Bierkens 2015). However, global climate models are considerably biased in global runoff simulations (Sellers *et al.* 1986; Xu *et al.* 2005; Sood & Smakhtin 2015). Hence, GHMs focusing on the simulation of water resources have been developed to simulate (sub-) surface water fluxes

and storages. Some of the widely used GHMs include Variable Infiltration Capacity model (VIC, Liang *et al.* 1994), Water Balance Model–Water Transport Model (WBM-WTM, Vörösmarty *et al.* 1989), PCRaster GLOBal Water Balance model (PCR-GLOBWB, Van Beek & Bierkens 2008; Van Beek *et al.* 2012), and Water And Snow balance MODeling system (WASMOD-M, Widén-Nilsson *et al.* 2007, 2009).

Although great progresses have been made in global hydrological modelling during the past decades, several studies highlighted the considerable discrepancies among these simulation results, which is partly due to the difficulty of parameter estimation for hydrological models (Widén-Nilsson *et al.* 2007; Schellekens *et al.* 2016). The lack of observational data in many regions of the world (especially in poorly instrumented regions) and the low quality or 'disinformative' data make it difficult to directly calibrate global hydrological model parameters (Kauffeldt *et al.* 2013; Schellekens *et al.* 2016). Therefore, the majority of macroscale models (e.g. GHMs) were inclined to get attached to prior parameterizations, which may result in insufficient streamflow simulations (Beck *et al.* 2016). For example, the parameter values of the WBM-WTM were tuned by an adjustment factor, rather than optimization (Vörösmarty *et al.* 1989). Thus, the parameter values of WGHM were globally uniform or related to land cover and its associate properties, except the runoff coefficient which was tuned against the time series of measured annual discharges (Döll *et al.* 2003). The parameter values of Macro-PDM were set based on literature review or previous model applications, and 6 out of 13 parameters were globally uniform (Arnell 1999, 2003).

Considering the restriction of the observational data, parameter regionalization, the most common method to solve the Predictions in Ungauged Basins (PUB) problem, has been used in global hydrological modelling. For example, the WASMOD-M transferred the calibrated parameter sets from catchment scale to grid cells by searching for the most commonly occurring parameter set within a rectangular window and found that regionalized parameters produced better streamflow estimates than spatially uniform parameters (Widén-Nilsson *et al.* 2007). Beck *et al.* (2016) transferred the calibrated parameter sets of the HBV model from the selected donor catchments to 0.5° grid cell with the most similar climatic and physiographic characteristics and found that HBV with regionalized parameters outperformed nine state-of-the-art macroscale models. Liang & Greene (2020) employed the regression method to transfer the runoff coefficient and estimated global runoffs. The above studies concluded that regionalization methods have a great potential to be used for global hydrological modelling.

Various regionalization methods were proposed and numerous studies have compared these methods over different regions during the past decades (Abdulla & Lettenmaier 1997; Hundecha & Bárdossy 2004; Oudin *et al.* 2008; Pokhrel *et al.* 2008; Jin *et al.* 2009; Samaniego *et al.* 2010; He *et al.* 2011; Razavi & Coulibaly 2013; Li & Zhang 2017; Yang *et al.* 2019; Yang *et al.* 2020). However, there is still no clear conclusion on the best-performed regionalization method for use in ungauged catchments. In particular, there is a lack of detailed information regarding the performance of regionalization methods on the global scale and the guidance for the use of regionalization methods for global hydrological modelling. Therefore, the objectives of this study are to (1) evaluate the performance of the most widely used regionalization methods over thousands of catchments in the world to determine an appropriate global-scale regionalization scheme (GSRS) for global hydrological modelling and (2) propose a robust regionalization-based global hydrological simulation scheme (RGHS) to estimate global runoff.

2. DATA AND METHODS

2.1. Observed streamflow data

The observed daily streamflow data used in this study were obtained from three sources, i.e. the Global Runoff Data Centre (GRDC; http://www.bafg.de/GRDC/, Lehner 2012), the Canadian model parameter experiment (CANOPEX) database (Arsenault *et al.* 2016, 2020), and some catchments of China (Gu *et al.* 2018). The GRDC dataset comprises river discharge data for more than 9,500 stations from 161 countries. Basin boundaries and flow paths were taken from HYDRO1 K (Gong *et al.* 2009). Continent boundaries were taken from STN-30p (Vörösmarty *et al.* 2000).

There are some uncertainties that may impact the results of global hydrological modelling. For example, small catchments might result in less reliable simulations since the model resolution is low (i.e. 0.5°). Therefore, the following three criteria were used to choose catchments for our analysis:

(1) The streamflow record length was not shorter than 5 years (not necessarily consecutive) during the 1982–2015 period.
(2) The catchment size is over 2,500 km^2 to ensure that each catchment covers at least one 0.5° grid cells.
(3) The upper limit of catchment size was set as 50,000 km^2 to minimize the effects of river regulation and channel routing effects (Gericke & Smithers 2014; Beck *et al.* 2016, 2020).

Based on the above criteria, 2,277 catchments were selected in total for the comparison of regionalization methods and GSRS. Figure 1 shows the distribution of these catchments and climate classification extracted from the World Map of Köppen-Geiger climate classification (Kottek *et al.* 2006). The majority of those catchments are located in North America, Europe, Southeast Asia, and central South America. However, catchments located in the Middle East, North Africa, the central Australian, and the Russian Far East and Siberia regions are much fewer.

2.2. Meteorological data

The meteorological data used in this study are daily precipitation, air temperature, and potential evaporation. The potential evaporation data at the global scale were obtained from the Global Land Evaporation Amsterdam Model (GLEAM V3) potential evaporation dataset (1980–2015) at the 0.5° resolution, which is calculated by using the Priestley and Taylor equation based on observations of surface net radiation and near-surface air temperature (Miralles *et al.* 2010; Martens *et al.* 2017). The daily temperature data were obtained from the European Centre for Medium range Weather Forecasts (ECMWF)–Interim Reanalysis (ERA-Interim) at the 0.5° resolution for the 1979–2019 period (Dee *et al.* 2011). The precipitation data were obtained from the Global Precipitation Climatology Centre (GPCC) V.2018 precipitation dataset at the 0.5° resolution for the 1982–2016 period (Fuchs *et al.* 2009), which was interpolated using gauged precipitation data provided by national meteorological and hydrological services, regional and global data collections, as well as the World Meteorological Organization (WMO) GTS-data (GPCC, http://gpcc.dwd.de).

In addition, comprehensive reviews of regionalization methods were made by He *et al.* (2011) and Razavi & Coulibaly (2013), in which the number of times that catchment descriptors used in other studies were counted after reviewing the regionalization methods. Based on this information, 13 catchment descriptors, classified as climate index, terrain characteristics, land use, and soil characteristics, were selected for regionalization (Table 1). Those datasets with spatial resolutions $<0.5°$ were resampled to 0.5° using bilinear averaging.

2.3. Hydrological models

Four conceptual hydrological models, i.e. Génie Rural à 4 paramètres Journalier model (GR4 J), simple lumped conceptual daily rainfall-runoff model (SIMHYD), Xinanjiang model (XAJ), and Hydrological Model of École de technologie supérieure (HMETS), were used to simulate runoff at the daily time step. They were chosen because of their proven effectiveness around the world and successful application in regionalization studies (Perrin *et al.* 2003; Oudin *et al.* 2008; Zhang & Chiew 2009; Li

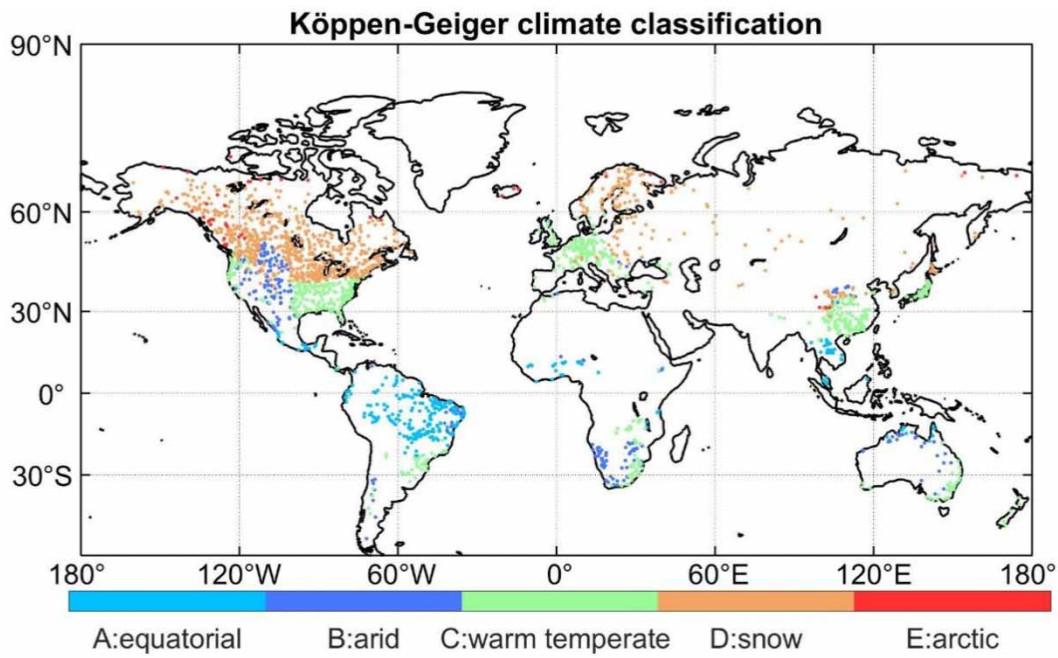


Figure 1 | Location of the catchments used in this study (each point represents the outlet of the catchment).

Table 1 | The statistical information of catchment descriptors used in regionalization methods

Indices	Mean	Median	Minimum	Maximum
Climate index				
Aridity index	0.85	0.8	0.04	3.33
Mean annual potential evaporation (mm)	1,169	1,089	301	3,060
Terrain characteristics				
Mean slope (°)	2.49	1.78	0.01	20.95
Mean elevation (m)	645	545	1.94	4,719
Area (km^2)	12,016	7,486	2,500	50,000
Land use				
Forest (%)	41.67	41.01	0	98.79
Water body (%)	2.6	0.07	0	69.73
Built-up land (%)	1.49	0.31	0	62.35
Total cultivated land (%)	16.98	6.41	0	96.2
Soil index				
Topsoil Clay Fraction (% wt.)	16.83	15.97	0	73.26
Subsoil Clay Fraction (% wt.)	18.38	17.35	0	78.22
Soil thickness (m)	11.82	3.74	0	50
Water holding capacity (mm/m)	42.36	43.83	0	100

Data from Harmonized World Soil Database (version 1.1); GlobCover Land Cover Maps, http://due.esrin.esa.int/page_globcover.php; and Global Aridity and PET Database, http://www.cgiar-csi.org.

et al. 2014; Zhang *et al.* 2014, 2016; Li & Zhang 2017; Chen *et al.* 2018; Shen *et al.* 2018; Yang *et al.* 2020; Qi *et al.* 2021b). In addition, the four hydrological models used in this study have different numbers of parameters and vary in the generalization of the natural hydrological processes and physical mechanisms (Zhao 1980, 1992; Chiew *et al.* 2002; Perrin *et al.* 2003; Chiew 2010; Boumenni *et al.* 2017; Martel *et al.* 2017; Shen *et al.* 2018). For example, the model structure for representing the hydrological process is more complicated and detailed in SIMHYD, XAJ, and HMETS than the most parsimonious GR4 J. For the simulation of evaporation, SIMHYD and HMETS use a one-layer evaporation model, while a three-layer model is used in XAJ. As for the streamflow generation, XAJ and HMETS use the saturation excess runoff generation mechanism based on the soil moisture content of the aeration zone reaching its field capacity, while SIMHYD considers both infiltration-excess and saturation-excess runoffs by using an interception store and a soil moisture store. In addition, GR4 J and HMETS take into account the groundwater exchange based on surface water–groundwater interaction functions, while this is not the case for SIMHYD and XAJ. Since there is no snow module in GR4 J, SIMHYD, and XAJ, the snow module-CemaNeige (Valéry 2010) was incorporated into the original models in this study. CemaNeige allows estimating the snowmelt and simulating the snowpack evolution by using two parameters, and the coupling of these models and CemaNeige has been tested in some other studies (e.g. Valéry 2010; Coron *et al.* 2014; Hublart *et al.* 2015; Guo *et al.* 2020; Wang *et al.* 2020; Yang *et al.* 2020). Table 2 shows the main structures of the four hydrological models. More detailed information about model parameters can be found in Supplementary Appendix Table A1.

2.4. Hydrological model calibration and evaluation methods

For each catchment, the record of observed streamflow data was split into a calibration period (consisting of the first 70% of the record) and a validation period (consisting of the remaining 30% of the record). The shuffled complex evolution method optimization algorithm (SCE-UA, Duan *et al.* 1992, 1993) was used to optimize the hydrological model parameters using KGE (Gupta *et al.* 2009) as the objective function. KGE has been introduced as an improvement of the widely used Nash–Sutcliffe efficiency (NSE), which considers different types of model errors, namely the error in the mean, the variability, and the dynamics (Vis *et al.* 2015). In addition, the accuracy of volume estimate (AVE), which is defined here as one minus volume error, and the commonly used NSE were also used to evaluate the performance of hydrological simulations. Different

Table 2 | Overview of the four hydrological models

Model	Parameters	Snow module	Characteristics of the model
GR4 J	6	CemaNeige	The effective rainfall is divided into direct runoff and delayed runoff according to the ratio of 90:10 A nonlinear production reservoir with two-unit hydrographs A routing reservoir
SIMHYD	11	CemaNeige	Precipitation loss calculation Two linear reservoirs for the calculation of interflow and base flow A nonlinear routing reservoir
XAJ	17	CemaNeige	Three-layer evapotranspiration system Linear reservoirs for surface flow routing Two recession coefficients for interflow and groundwater flow routing
HMETS	21	HMETS	A snowmelt module and an evapotranspiration module Generation of surface and delayed runoff after evapotranspiration and infiltration Generation of hypodermic flow and groundwater flow with two reservoirs A routing module

metrics emphasize different aspects of hydrological streamflow properties. For example, the KGE puts more emphasis on the simulation of flow variability (Santos *et al.* 2018; Knoben *et al.* 2019; Qi *et al.* 2021b). The NSE gives more weights to high flows (Mizukami *et al.* 2019; Wan *et al.* 2021), while the AVE emphasizes more on the total water balance (Vis *et al.* 2015).

$$\mathrm{KGE} = 1 - \sqrt{(R-1)^2 + \left(\frac{\overline{Q}_{sim}}{\overline{Q}_{\mathrm{obs}}} - 1\right)^2 + \left(\frac{\mathrm{CV}_{\mathrm{sim}}}{\mathrm{CV}_{\mathrm{obs}}} - 1\right)^2} \tag{1}$$

$$\mathrm{NSE} = 1 - \frac{\sum (Q_{\mathrm{sim}} - Q_{\mathrm{obs}})^2}{\sum (Q_{\mathrm{obs}} - \overline{Q}_{\mathrm{obs}})^2} \tag{2}$$

$$\mathrm{AVE} = 1 - \mathrm{VE} = 1 - \frac{|\sum (Q_{\mathrm{obs}} - Q_{\mathrm{sim}})|}{\sum (Q_{\mathrm{obs}})} \tag{3}$$

where Q_{obs} represents the observed runoff and Q_{sim} represents the simulated runoff. $\overline{Q}_{\mathrm{obs}}$ is the mean of observed runoff and $\overline{Q}_{\mathrm{sim}}$ is the mean of simulated runoff. R is the Pearson correlation coefficient between observed and simulated runoffs. $\mathrm{CV}_{\mathrm{sim}}$ and $\mathrm{CV}_{\mathrm{obs}}$ represent variation coefficients of observed and simulated streamflows, respectively. KGE, NSE, and AVE values range from $-\infty$ to 1 and the closer the value to one, the better the simulation.

2.5. Regionalization methods

The regionalization methods used in this study are the most commonly used step-wise regionalization methods, i.e. the global mean method, the regression-based methods, and the distance/attribute-based methods (Jin *et al.* 2009; He *et al.* 2011; Razavi & Coulibaly 2013; Yang *et al.* 2018, 2019, 2020). Moreover, the great performance of these methods has been observed for the ungauged runoff prediction in many studies (Merz & Blöschl 2004; Oudin *et al.* 2008; Yang *et al.* 2018, 2019).

The global mean method is a simple regionalization method. Generally, the arithmetic mean of the parameters of all gauged catchments in the region is directly applied to the ungauged catchments.

The regression-based methods assume that a well-behaved relationship exists in the observable catchment characteristics and model parameters (Burn & Boorman 1993). Parameters for an ungauged catchment are derived by using the relationship between catchment descriptors and model parameters for the donor catchments (Xu 1999b, 2003; Muller-Wohlfeil *et al.* 2003; Arsenault & Brissette 2014; Yang *et al.* 2018, 2019). The multiple linear regression method was used in this study, in which the relationships among model parameters and the selected 13 catchment descriptors were established using multiple linear regressions and these functions were then used to estimate model parameters for ungauged catchments.

The distance/attribute-based methods assume that the parameter sets of hydrological models on gauged catchments can be transferred to nearby or physically similar ungauged catchments following various procedures. The key of these methods is to find the closest or the most similar donor (gauged) catchments to the ungauged catchments. The distance/attribute-based

methods usually include spatial proximity, physical similarity, and physical similarity considering distance (Yang *et al.* 2018, 2020). The spatial proximity method assumes that nearby catchments should have similar behaviour for climate and catchment conditions (features) varying uniformly in space (Tobler 1970; Pool *et al.* 2017). The Haversine formula was used to calculate the distance D_{ud} between the donor and ungauged catchments' outlets, the same as Qi *et al.* (2021a). The Haversine formula determines the great-circle distance between two points on a sphere given their longitudes and latitudes (Abebe *et al.* 2020). The D_{ud} is calculated by Equation (4):

$$D_{ud} = 2 \times r \times \sin^{-1}\left(\sqrt{\sin^2\left(\frac{\phi_u - \phi_d}{2}\right) + \cos(\phi_u) \times \cos(\phi_d) \times \sin^2\left(\frac{\lambda_u - \lambda_d}{2}\right)}\right) \tag{4}$$

where r is the average radius of the Earth (i.e. 6,378.137 km); t and d represent the target and donor catchments, respectively; ϕ_u, ϕ_d and λ_u, λ_d are catchment outlet latitude and longitude values of the target and donor catchments (in radians).

The physical similarity method is based on the assumption that catchments with similar attributes show similar hydrological behaviours. The core of the physical similarity method is the selection of the physical similarity metric (Yang *et al.* 2020). Many studies have focused on the selection of proper similarity index between the donor and ungauged catchments and the proper catchment attributes for similarity index calculation (Burn & Boorman 1993; Yang *et al.* 2018). The similarity index in this study is calculated by Equation (5) (Burn & Boorman 1993):

$$\mathrm{SI}_{ud} = \sum_{i=1}^{k} \frac{|\mathrm{CD}_{d,i} - \mathrm{CD}_{u,i}|}{\Delta \mathrm{CD}_i} \tag{5}$$

where CD is the catchment descriptor; u and d represent the ungauged and donor catchments, respectively; k is the total number of catchment descriptors; and $\Delta \mathrm{CD}_i$ represents the range of ith catchment descriptor.

Considering the limitation of the two distance/attribute-based methods mentioned earlier, some studies (e.g. Samuel *et al.* 2011; Viviroli & Seibert 2015; Yang *et al.* 2018) integrated spatial proximity with physical similarity to improve the regionalization ability. In the present study, the physical similarity method considering distance in which the distance was considered as one of the catchment descriptors was used and then the similarity index was calculated.

For distance/attribute-based methods, there are two different averaging options to transfer the model parameter sets from donor catchments: (1) parameter average option, which transfers the averaged model parameters from donor catchments to ungauged catchments; and (2) output average option, which averages runoff simulations calculated by using individual parameter sets from donor catchments to ungauged catchments (Oudin *et al.* 2008; Yang *et al.* 2018). In addition, there are two different weighting approaches used to combine the model parameters or model outputs: (1) Arithmetic Mean (AM) method and (2) Inverse Distance Weighted (IDW) method (Parajka *et al.* 2007; Yang *et al.* 2018). All regionalization methods used in this study are summarized in Table 3.

To find suitable donor catchments and the optimal regionalization scheme (i.e. combination of different optional regionalization methods), the performance of regionalization methods under different thresholds of model efficiency in terms of KGE was tested. Threshold values of model efficiency (all, $\geq$0, $\geq$0.5, $\geq$0.6, $\geq$0.7, $\geq$0.8, or $\geq$0.9) were determined for the calibration period. In other words, when the efficiencies of the catchments were below a threshold, these catchments were not used as donor catchments to simulate runoffs for ungauged catchments. The threshold named 'all' means that all catchments are used as donor catchments regardless theirs performance in the calibration stage. What is more, all catchments, whether poorly or well modelled, were all considered as pseudo-ungauged under different efficiency thresholds for cross-validation.

2.6. Global water resources estimation

The commonly used leave-one-out cross-validation strategy was used to evaluate the performance of the regionalization methods. Specifically, the performance of five parameter regionalization methods with two different average options, two weighting approaches (14 regionalization schemes, Table 3), and seven efficiency thresholds was compared over thousands of catchments based on four conceptual hydrological models. The best-performed regionalization method (with the highest KGE value) was then selected as the GSRS to regionalize parameters of four hydrological models to the spatial resolution of 0.5°×0.5° grid cell all over the world except for Antarctica. This procedure is based on an assumption that the parameters at

Table 3 | Summary of regionalization methods used in this study

Regionalization methods	Averaging options	Weighting approaches method	Abbreviation
Global mean (GM)			GM
Regression (RE)			RE
Spatial Proximity (SP)	Parameter Averaging	Arithmetic Mean	SPA-PAR
		Inverse Distance Weighted	SPI-PAR
	Output Averaging	Arithmetic Mean	SPA-OUT
		Inverse Distance Weighted	SPI-OUT
Physical Similarity (PS)	Parameter Averaging	Arithmetic Mean	PSA-PAR
		Inverse Distance Weighted	PSI-PAR
	Output Averaging	Arithmetic Mean	PSA-OUT
		Inverse Distance Weighted	PSI-OUT
Physical Similarity considering distance (PSD)	Parameter Averaging	Arithmetic Mean	PSDA-PAR
		Inverse Distance Weighted	PSDI-PAR
	Output Averaging	Arithmetic Mean	PSDA-OUT
		Inverse Distance Weighted	PSDI-OUT

the catchment scale can be transferred to the grid scale. In other words, the 0.5°×0.5° grid cell was treated as a catchment (Widén-Nilsson *et al.* 2007; Beck *et al.* 2016).

In the RGHS, the gridded version of hydrological models driven by precipitation, air temperature, potential evaporation, and the global gridded parameters was used to estimate global water resources. For calculating catchment streamflow, runoff routing algorithms are required to converge the grid runoff to catchment streamflow (Vörösmarty *et al.* 1989; Döll *et al.* 2003). There are several large-scale runoff routing algorithms have been developed (Graham *et al.* 1999; Vörösmarty *et al.* 2000; Döll & Lehner 2002; Gong *et al.* 2009, 2011). The improved Network-response Routing Function (NRF) method was selected in this study, since this method transfers high-resolution delay dynamics, instead of networks, to any lower spatial resolution where runoff is generated (Gong *et al.* 2009; Li *et al.* 2020). The formula of this method is given by Equation (6):

$$vi = v_{45} * (\tan c_i)^b \tag{6}$$

where vi is the wave velocity of the grid and c_i is the slope of the grid. v_{45} is the wave velocity in the grid with the slope of 45° and b is a parameter that reflects how sensitive is the wave velocity to the slope.

The values of these parameters for calibration are [4,5,6,7,8,9,10] for v_{45} and [0.2,0.3,0.4,0.5,0.6] for b, respectively. They were chosen based on computer capability limitations and the physical meaning of each parameter. Therefore, there are 35 routing parameter sets, composed of five b values and seven v_{45} values. In this study, the enumeration method was used to calibrate these two routing parameters. The catchment streamflow was generated by converging grid runoff from all grids in the catchment based on each routing parameter set. Since two routing parameters are calibrated rather than regionalized, it is hard to obtain streamflow time series for real ungauged catchments. To overcome this problem, the commonly used aggregation was used to obtain the simulation performance of ungauged catchments without calibration of routing parameters (Widén-Nilsson *et al.* 2007; Beck *et al.* 2020).

2.7. Performance evaluation for RGHSs

The performance of the RGHSs was evaluated in four ways (Figure 2).

(1) The effectiveness of RGHSs was first evaluated by comparing the evaluation metrics of the 2,277 catchments obtained by RGHSs (RGHS-NRF and RGHS-aggregation) with the results obtained by calibration and regionalization at the catchment scale.
(2) The second evaluation was done by comparing the evaluation metrics of the 2,277 catchments obtained by RGHSs with the results obtained from nine state-of-the-art macroscale models, coming from the EartH2Observe project (Schellekens *et al.* 2016), which were driven by WATCH Forcing Data ERA-Interim (WFDEI) for the period 1979–2012 at the 0.5° resolution. The catchments' streamflow time series of these nine models were calculated by using the NRF method in

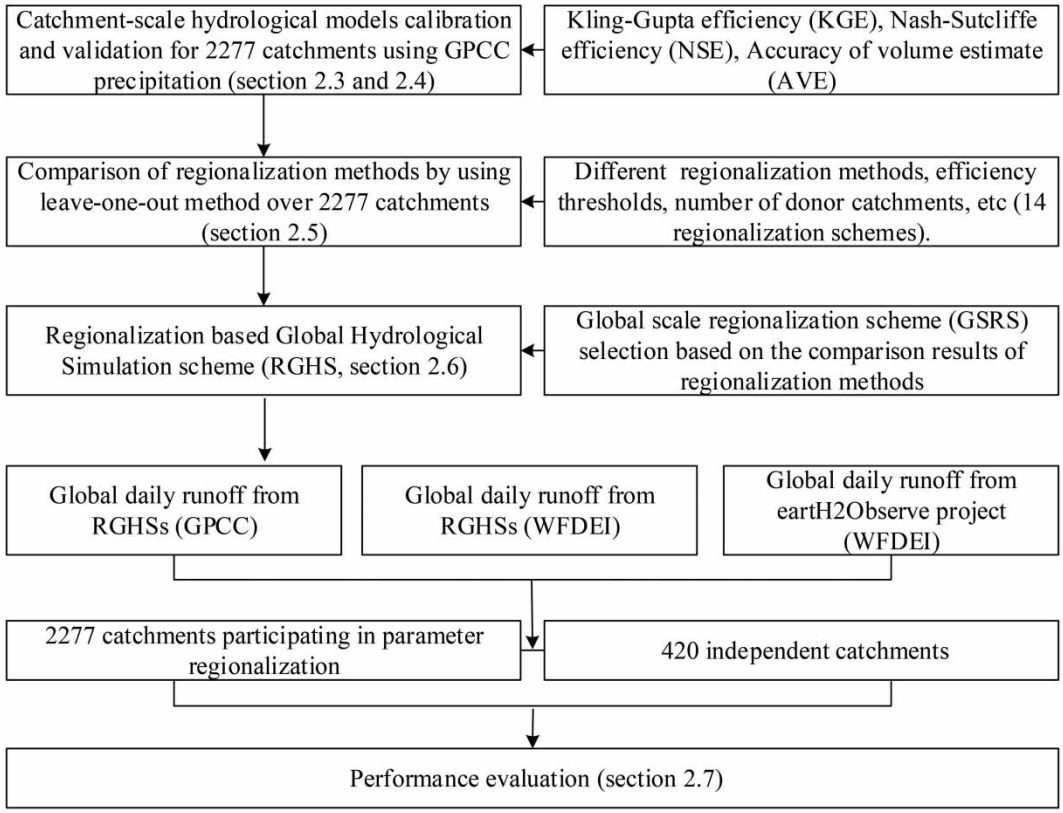


Figure 2 | The schematic diagram of the main steps in this study.

this study. For the comparison of RGHSs versus nine models from the EartH2Observe project, RGHSs were driven by WFDEI (RGHS-WFDEI).

(3) In addition, 420 independent catchments with surface area ranging between 50,000 and 100,000 km^2 and with streamflow records being longer than 5 years were used as validation catchments. The KGE, NSE, and AVE values of the validation catchments obtained by RGHSs and the results obtained by nine models were further compared to assess the value of RGHSs.

(4) The last way to evaluate the effectiveness of RGHSs was by comparing the long-term averaged annual total runoff (1983–2015, 1982 is considered as the warm-up period) from global land and each continent obtained by RGHSs with the results in literatures.

3. RESULTS

3.1. Hydrological model performance at the catchment scale

Figure 3 shows the cumulative density function curves of the percentage of catchments with KGE, NSE, and AVE values exceeding the given value (the value of *y*-axis) for four hydrological models over the calibration and validation periods. Generally, the performances of the four hydrological models are close to each other. For all hydrological models, the KGE value is higher than 0.7 for more than 60% catchments at the calibration period and 0.5 for the validation period. Specifically, SIMHYD shows the best performance, followed by XAJ, GR4 J, and HMETS. It is the same when using NSE and AVE as evaluation metrics. Figure 4 presents the spatial distribution of the model efficiency for the calibration period in terms of KGE values. The results show that the KGE values are above 0.8 for most catchments in eastern Canada, the USA, southern China, and along the Atlantic Coast of Europe. The catchments in the American tropics, the Andes (South America), and northwest China show low KGE values, which might be because of the complex topography and insufficiency of precipitation observations for the calibration of hydrological models (Demirel *et al.* 2015; Vetter *et al.* 2015; Beck *et al.* 2016). Among the

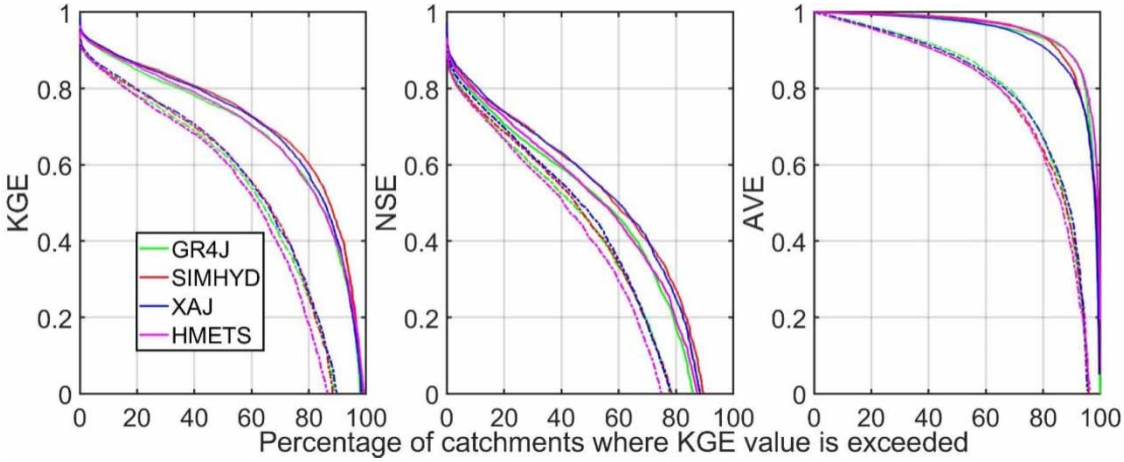


Figure 3 | The performance of hydrological models by split-sample test evaluated by the KGE, NSE, and AVE values. The solid and dash lines show the performance for the calibration and validation periods, respectively.

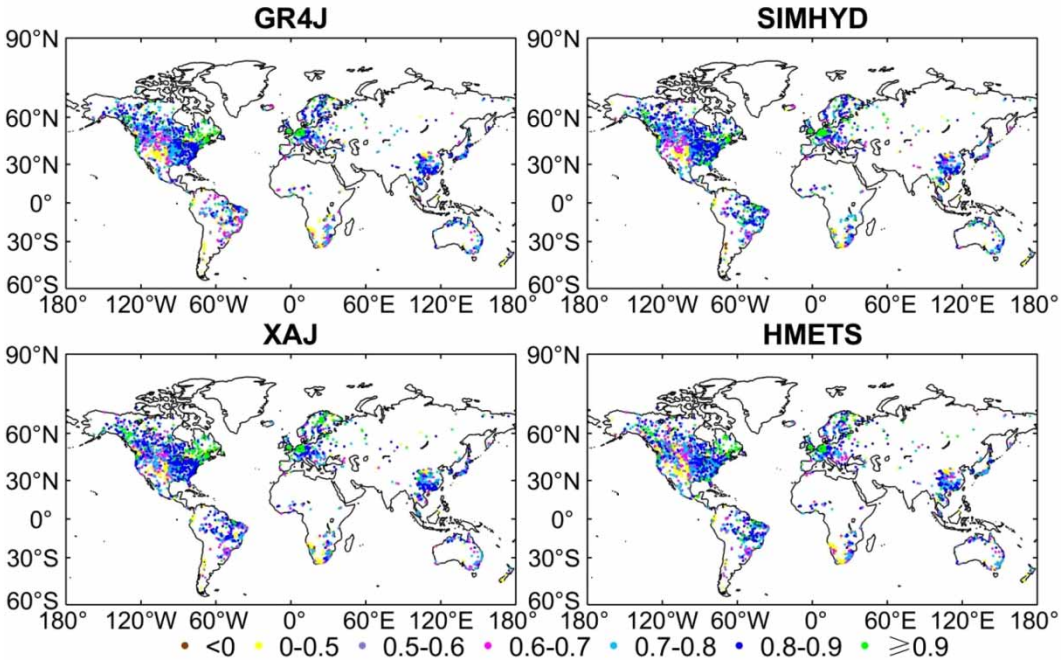


Figure 4 | Spatial distribution of model efficiency (i.e. KGE).

five different climate regions, the highest median KGE values are obtained for the cold regions, and the lowest for the arid regions (Figure 5). The good performance in the cold regions may be because the hydrological processes are not that sensitive to precipitation variability. The large different performance among four hydrological models in the equatorial and arid regions is likely due to their different skills in simulating the highly nonlinear process from precipitation to runoff.

The median KGE values of these hydrological models are 0.748 for GR4 J, 0.774 for SIMHYD, 0.766 for XAJ, and 0.750 for HMETS in the calibration period and 0.632 for GR4 J, 0.644 for SIMHYD, 0.641 for XAJ, and 0.617 for HMETS in the validation period. The performance of hydrological models used in this study is comparable to or higher than those from previous studies. For example, Beck *et al.* (2020) calibrated the HBV model for 4,229 catchments globally and obtained median daily KGE values of 0.770 and 0.690 for the calibration and validation periods, respectively. Alfieri *et al.* (2020) calibrated the LISFLOOD model for 1,126 catchments and obtained median daily KGE values of 0.670 and 0.610 for the calibration and validation periods, respectively.

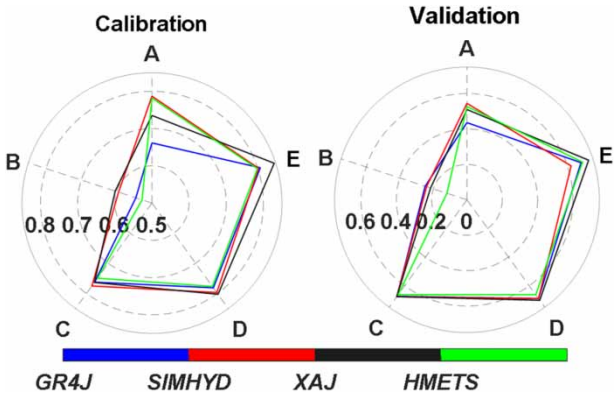


Figure 5 | Median values of KGE over all catchments for four hydrological models under different climate regions. A, B, C, D, and E represent equatorial, arid, warm temperate, snow, and polar regions, respectively.

3.2. Comparison of regionalization methods

3.2.1. The performance of different regionalization schemes

For illustrative purpose, the model performance (GR4 J) of distance/attribute-based methods using different numbers of donor catchments with different weighting and averaging options under each KGE threshold is shown in Figure 6 (similar results can be found for the other three models, which are shown in Supplementary Appendix Figure A1). Results show that the IDW approach performs better than the AM approach for all distance/attribute-based regionalization methods and all hydrological models. This is consistent with the results from other studies (e.g. Samuel *et al.* 2011; Arsenault & Brissette 2014; Li *et al.* 2014). The worse performance of the AM approach is probably caused by the large difference of distance or similarity among our studied catchments. However, the weighting scheme of IDW minimizes the negative impact caused by the farthest distance or the least similar donors. Furthermore, the optimal number of donor catchments for the parameter

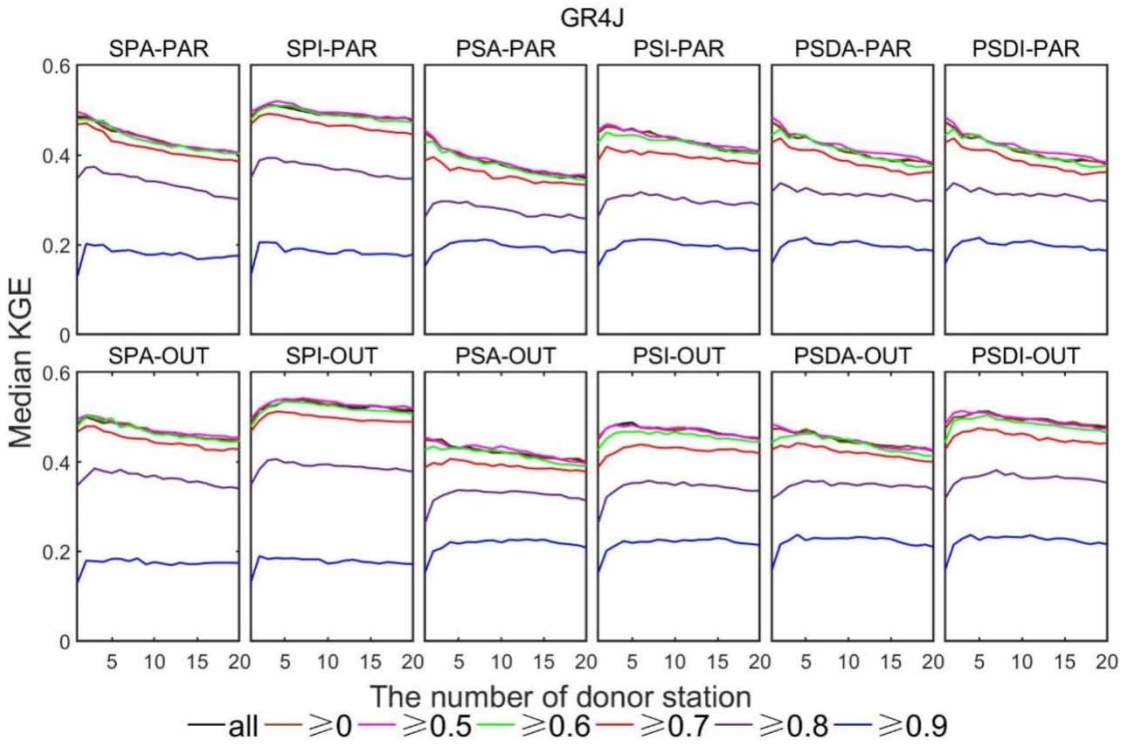


Figure 6 | The median KGE value for distance/attribute-based regionalization methods with an increasing number of donor catchments.

averaging option always ranges between 1 and 5. For the output averaging option, it is between 4 and 6 (see Figure 6). In general, the output averaging option outperforms the parameter averaging option globally in this study, which is also consistent with previous regional studies (e.g. Samuel *et al.* 2011; Arsenault & Brissette 2014; Li *et al.* 2014; Yang *et al.* 2018, 2020) at the regional scale. Therefore, to balance the effect and the amount of computation, five donor catchments are suggested to use for the output averaging method.

Poorly modelled catchments may yield higher uncertain model parameter values, but they may add some diversities for modelling ungauged catchments as well (Oudin *et al.* 2008). Therefore, it is worth evaluating the necessity to consider poorly modelled catchments in regionalization. In this study, we compared the performance of regionalization methods under different KGE thresholds for all regionalization methods (including regression method, not shown) and found that the KGE threshold of 0.5 is the best (see Figure 6). Including catchments with KGE values below 0.5 as donors has little effect on the performance of hydrological models for ungauged catchments. However, when using the efficiency threshold of 0.9, the hydrological model performance remarkably drops for ungauged catchments. This indicated that there is a loss of information between those thresholds, which is attributable to the loss of donor catchments. When using the threshold of 0.9, only 170 (for GR4 J), 211 (for SIMHYD), 236 (for XAJ), and 193 (for HMETS) donor catchments out of 2,277 were available and most of them are located in the USA, southern China, and along the Atlantic Coast of Europe. This result is consistent with those in Neri *et al.* (2020), as they concluded that the performance of regionalization methods strongly depends on the informative content of the dataset of available donor catchments. Therefore, the threshold of 0.5 is taken as the best performance threshold, and the total number of donor catchments for the global regionalization scheme ranges from 1902 (for GR4 J) to 1985 (for SIMHYD) for different hydrological models.

3.2.2. Comparison of regionalization methods

The performance differences among regionalization methods are observed in Figure 7. Generally, the SPI-OUT method outperforms the others, followed by PSDI-OUT and PSI-OUT, and the GM method performs the worst. Overall, the performances of different regionalization methods are consistent in each hydrological model. This indicates that the ranking of regionalization methods is independent of model structures. As for hydrological models, the two parsimonious hydrological models (i.e. GR4 J and SIMHYD) slightly outperform the other two more complex models (i.e. XAJ and HMETS) in most situations. Figure 8 summarizes the median KGE values of all regionalization methods for five climate regions. The results show that SPI-OUT outperforms other regionalization methods for all climate regions, followed by PSDI-OUT. GM and RE perform worse than other regionalization methods for all climate regions. In general, the performance of SPI-OUT is relatively stable (except for arid regions) for different climate regions. However, all regionalization methods perform much worse for arid regions than for the other four climate regions.

Based on the comparison of regionalization methods, it can be concluded that the SPI-OUT outperforms others. However, according to the principle of this method (i.e. the nearby catchments should have similar behaviours), the performance of the SPI-OUT method may deteriorate with the increase of the mean distance between the donor and ungauged catchments. Therefore, it is important to explore how the mean distance impacts the performance of regionalization methods. Figure 9 shows the percentage of the best-performed regionalization method under different mean distances (between the donor and ungauged catchments) under the threshold of 0.5 (the SPI-OUT, PSI-OUT, and PSDI-OUT were selected for their outstanding performance). For all hydrological models, when the mean distance between donors and ungauged catchments is smaller than 700 km, the average proportion of catchments that the SPI-OUT method outperformed others is the largest. It is hard to decide which method outperforms others when the mean distance is larger than 700 km for GR4 J and HMETS, 900 km for SIMHYD, and 1,000 km for XAJ. The advantage of the SPI-OUT method is reduced with the increased mean distance between donors and ungauged catchments. Overall, the greatest gains of the SPI-OUT method in performance are achieved for ungauged catchments with mean distances no more than 700 km from the donors for most hydrological models.

3.3. Regionalization-based global hydrological simulation schemes

3.3.1. Global-scale regionalization scheme

The GSRS was selected based on the best-performed catchment scale regionalization methods (Section 3.2). That is, only catchments with KGE value being greater than 0.5 at the calibration period were used as donor catchments. For grid cells with a mean distance less than 700 km to donors, the calibrated parameter sets of the 5 nearest donor catchments were

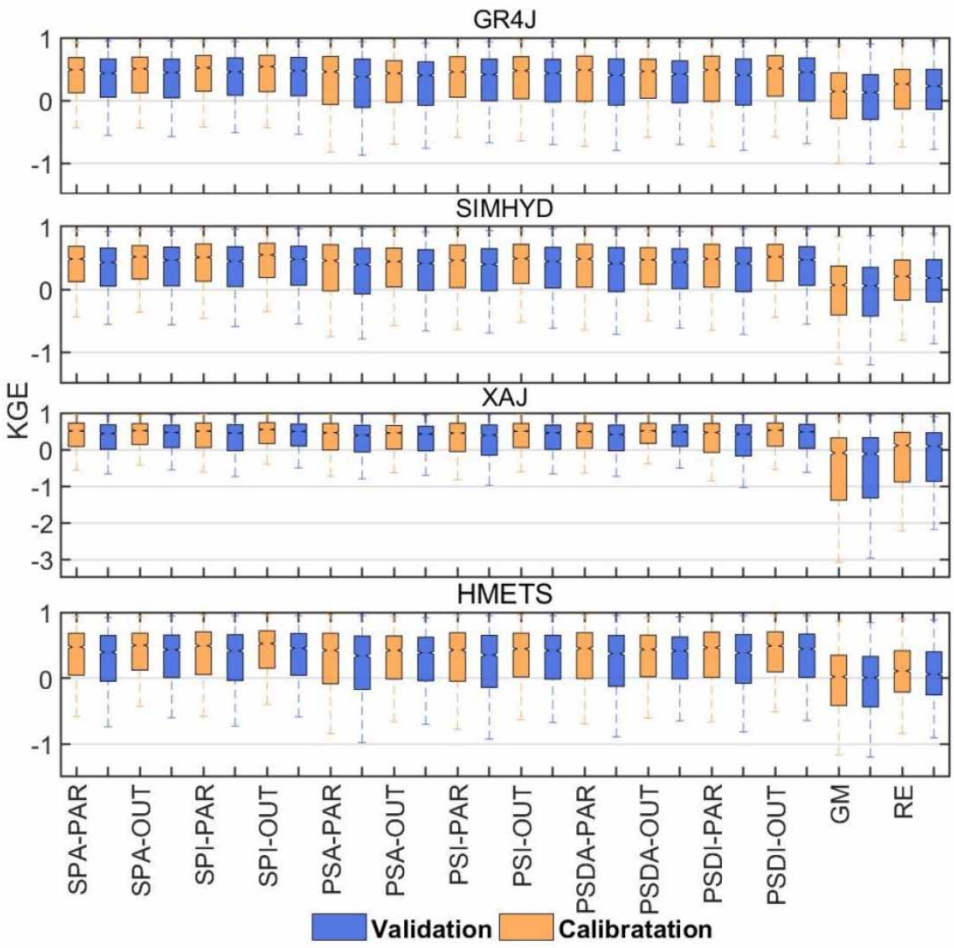


Figure 7 | Comparison of model efficiencies on ungauged catchments using several regionalization schemes.

transferred by using the SPI-OUT method. For grid cells with a mean distance larger than 700 km, the parameters were extracted from the PSDI-OUT method. Figure 10 shows the mean distance to the 5 nearest donor catchments. The mean distances are generally $\leq$700 km for most parts of the world. However, there were still some regions where the mean distance is $>$700 km, like Greenland, central Siberia, southern South America, Southeast Asia, and Northeast Africa.

3.3.2. RGHSs performance at the catchment scale

By using the GSRS, the runoff was first calculated for all global grid cells except for Antarctica. The NRF approach was then used to converge gird runoff to catchment streamflows. The 2,277 catchments were used to evaluate the performance of the RGHSs, since all of the grids were regionalized (Table 4). The median KGE value of SIMHYD-RGHS is the largest (0.385) and that of HMETS-RGHS is the smallest (0.374). When using NSE, HMETS-RGHS has the largest value (0.283) and GR4 J-RGHS has the smallest value (0.231). In terms of AVE, SIMHYD-RGHS performs the best (0.686) and XAJ-RGHS performs the worst (0.672). All above-mentioned differences are considered not large.

Generally, RGHSs perform better than the catchment scale regression and global mean methods, but worse than other regionalization methods. The differences between the median KGE values of RGHSs and that using the best-performed catchment scale regionalization method (i.e. SPI-OUT) are 0.166 (for GR4 J), 0.173 (for SIMHYD), 0.166 (for XAJ), and 0.155 (for HMETS). The differences between the median KGE value of RGHSs and that using the calibrated parameters are 0.372 (for GR4 J), 0.389 (for SIMHYD), 0.386 (for XAJ), and 0.376 (for HMETS). The performance of RGHSs is about half of that obtained using calibrated parameters in terms of the median value of KGE. Similar results are also observed when using NSE. However, the differences between RGHSs and those using the calibrated parameters are small when using AVE. In

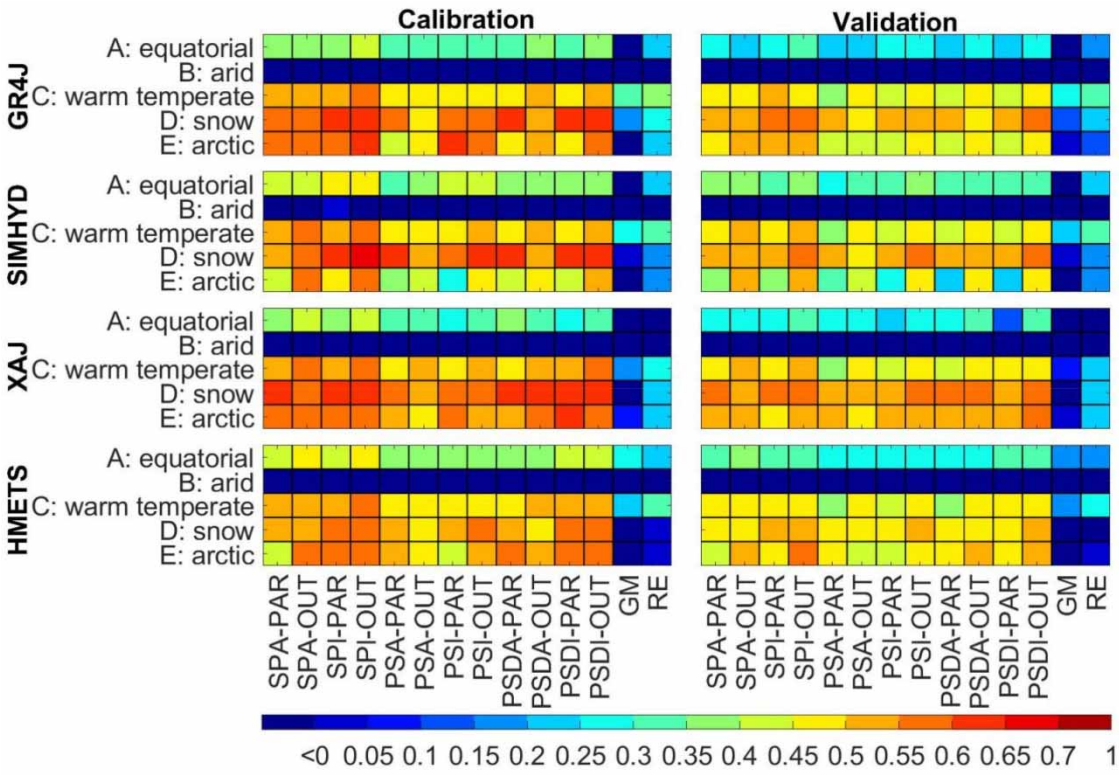


Figure 8 | Median KGE values of each regionalization method in different climate regions for both calibration period and validation period.

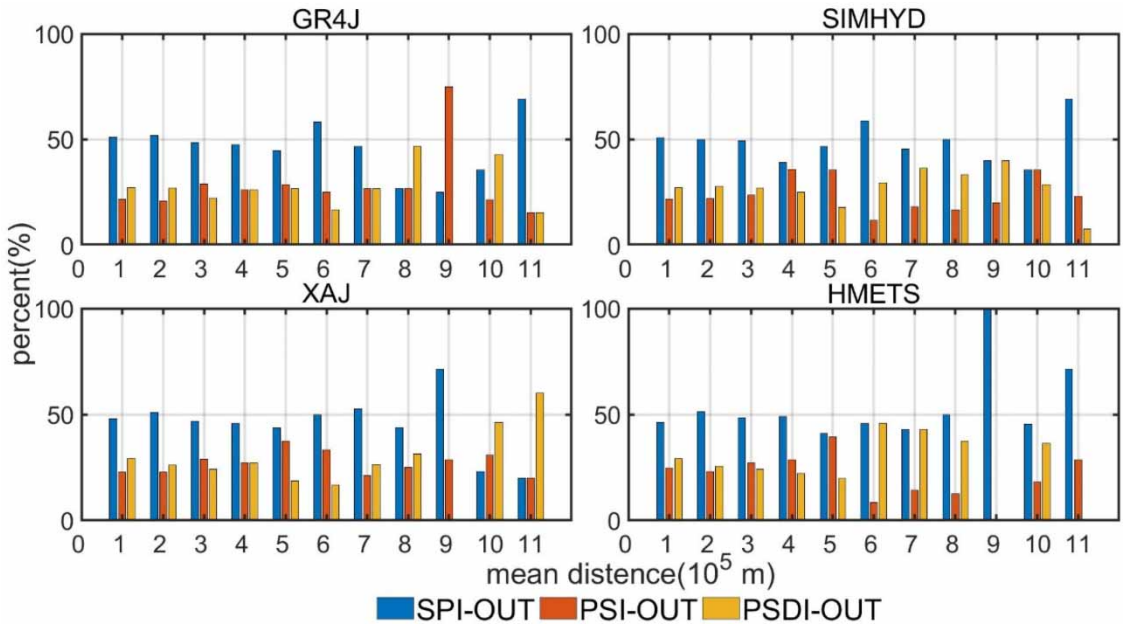


Figure 9 | The proportion of outperformed regionalization method over 2,277 catchments with increasing mean distances between the donor and ungauged catchments.

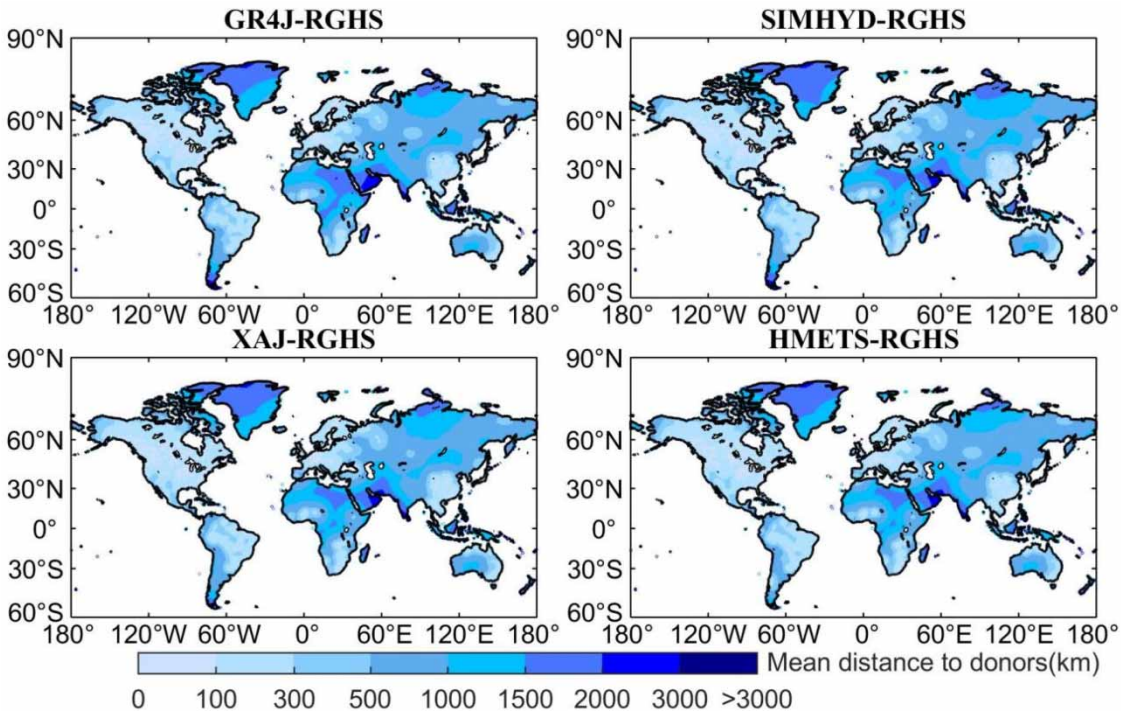


Figure 10 | Mean distance to the 5 nearest donor catchments.

Table 4 | The median of KGE, NSE, and AVE values of RGHSs over 2,277 catchments

	Calibration	Validation	GM	SPI-OUT	RGHS	RGHS-aggregation
KGE						
GR4 J	0.748	0.632	0.141	0.542	0.376	0.342
SIMHYD	0.774	0.644	0.068	0.558	0.385	0.365
XAJ	0.766	0.641	−0.093	0.546	0.380	0.360
HMETS	0.750	0.617	0.019	0.529	0.374	0.356
NSE						
GR4 J	0.533	0.450	−0.052	0.386	0.231	0.192
SIMHYD	0.566	0.448	−0.249	0.400	0.256	0.240
XAJ	0.568	0.471	−0.857	0.400	0.256	0.222
HMETS	0.528	0.407	−0.420	0.380	0.283	0.252
AVE						
GR4 J	0.977	0.891	0.633	0.842	0.678	0.677
SIMHYD	0.985	0.875	0.516	0.840	0.686	0.686
XAJ	0.979	0.883	0.500	0.843	0.672	0.668
HMETS	0.982	0.875	0.618	0.833	0.680	0.678

addition, the median KGE and NSE values of 2,277 catchments of RGHS-aggregation are smaller than that of RGHS-NRF. However, when using AVE, the difference between these two methods becomes small. This result indicates that the aggregation method is proper to be used in the water balance analysis. However, effective estimation of streamflow time series and extreme flows in ungauged catchments still needs to be further studied.

Figure 11 shows the spatial distribution of KGE for RGHSs. For most of Europe and the east coast of North America, KGE is generally above 0.5, and even above 0.7 over many catchments. In contrast, most catchments in southwestern Africa and northwestern Australia perform much worse. To further evaluate the effectiveness of the RGHSs in various climate regimes, Table 5 summarizes the median KGE values over 2,277 catchments for five Köppen-Geiger climate types. Generally, all RGHSs perform much worse for arid climate regions than for the other four climate regions. Previous studies (e.g. Oki *et al.* 2001; Döll *et al.* 2003; Beck *et al.* 2016) indicated that most of the existing GHMs overestimate runoff in arid basins. This is because the high evaporative losses in arid regions, the highly nonlinear response of runoff to rainfall, and the flashy nature of the streamflow time series make it difficult to simulate streamflow time series (Pilgrim *et al.* 1988; Widén-Nilsson *et al.* 2007; Beck *et al.* 2016).

3.3.3. Comparison of RGHSs with models from EarthH2Observe project

To further evaluate the effectiveness of the RGHSs, the KGE, NSE, and AVE values obtained from four RGHSs are compared to those obtained from nine macroscale hydrologic models from the EartH2Observe project. Figures 12 and 13 present evaluation metrics over 2,277 catchments that participated in global-scale regionalizations and 402 independent validation catchments, respectively. The overall performance of RGHS-GPCC is better than that of the nine models, with varying

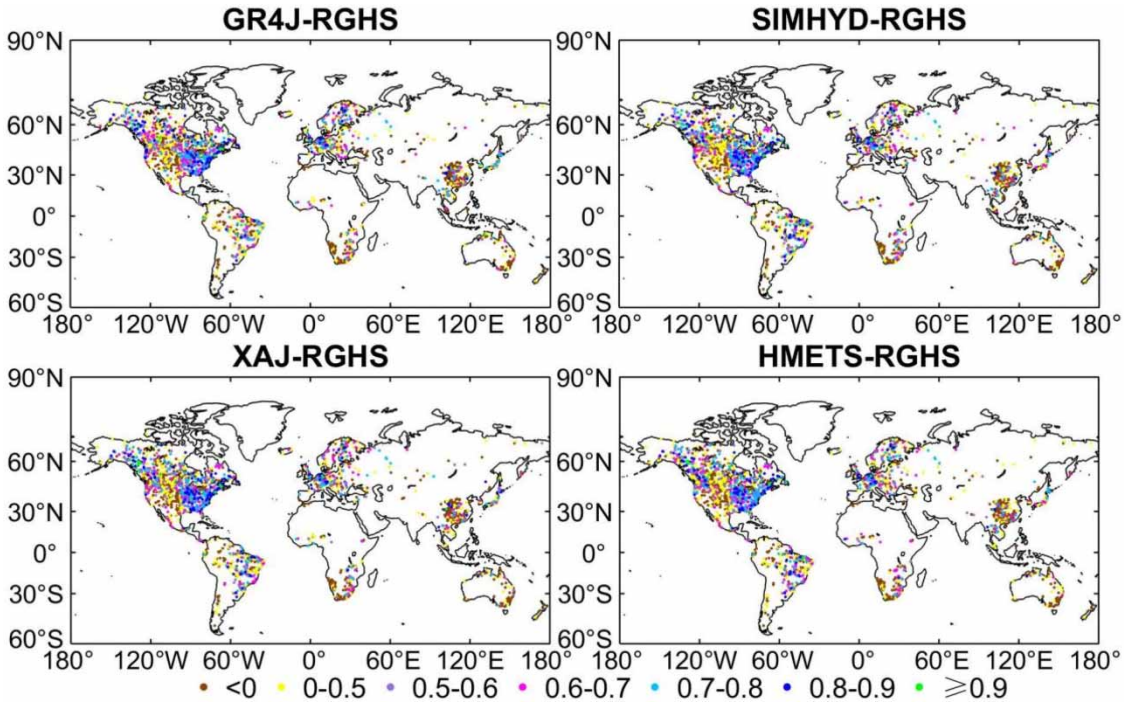


Figure 11 | The distribution of KGE value for four RGHSs.

Table 5 | The median of KGE values of RGHSs for different climate regions

Climate Type	GR4 J-RGHS	SIMHYD-RGHS	XAJ-RGHS	HMETS-RGHS
All (n=2,277)	0.376	**0.385**	0.380	0.374
A: equatorial (n=293)	0.183	**0.280**	0.270	0.266
B: arid (n=247)	−0.347	−0.341	**−0.313**	−0.324
C: warm temperate (n=717)	0.448	0.450	**0.453**	0.452
D: snow (n=970)	**0.470**	0.466	0.467	0.468
E: polar (n=50)	0.446	0.449	**0.517**	0.440

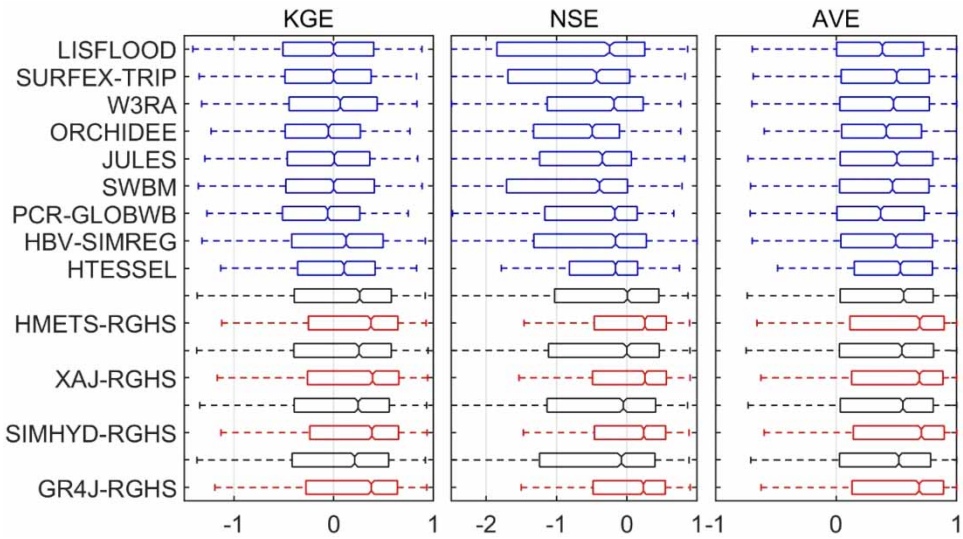


Figure 12 | Comparison of RGHSs with nine models over 2,277 catchments. The red and black boxes represent RGHS driven by GPCC and WFDEI, respectively. Please refer to the online version of this paper to see this figure in colour: https://doi.org/10.2166/nh.2022.118|0|0|2022.

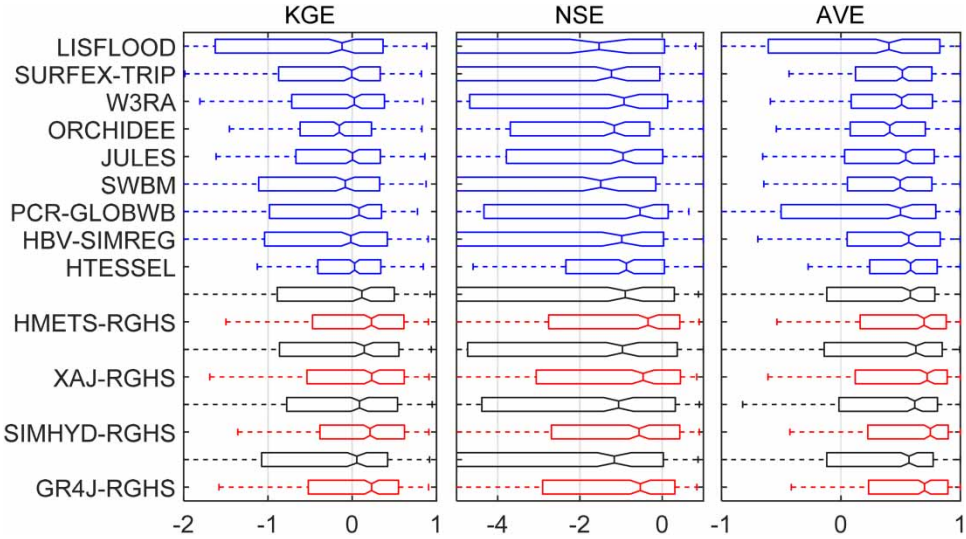


Figure 13 | Comparison of RGHSs with nine models over validation catchments. The red and black boxes represent RGHS driven by GPCC and WFDEI, respectively. Please refer to the online version of this paper to see this figure in colour: https://doi.org/10.2166/nh.2022.118|0|0|2022.

degrees of improvement in the median, upper and lower quartiles, and whiskers of each performance metric (except for the HTESSEL model for the lower quartiles and whiskers). When comparing the models in terms of the median of each performance metric, the median KGE of RGHS-GPCC ranges between 0.374 (HMETS-RGHS) and 0.385 (SIMHYD-RGHS), while that of nine models ranges between −0.061 (PCR-GLOBWB) and 0.124 (HBV-SIMREG) over 2,277 catchments. As for the independent validation catchments, the median KGE of RGHSs is between 0.213 (HMETS-RGHS) and 0.231 (XAJ-RGHS), while that of nine models is between −0.154 (ORCHIDEE) and 0.082 (PCR-GLOBWB). Similar results are observed when using NSE and AVE as evaluation metrics.

Compared to RGHS-GPCC, the performance of RGHSs driven by WFDEI (RGHS-WFDEI) somewhat deteriorates. For example, the median KGE of RGHS-WFDEI is between 0.209 (GR4 J-RGHS) and 0.260 (HMETS-RGHS) over 2,277 catchments, and 0.056 (GR4 J-RGHS) and 0.143 (SIMHYD-RGHS) for the independent validation catchments, which is lower

than that of RGHS-GPCC. However, RGHS-WFDEI still outperforms the other nine models from the EartH2Observe project in most cases. For example, over 2,277 catchments, the 25% quantile and median evaluation metrics of RGHS-WFDEI are always bigger than the other nine models and only the 75% quantile evaluation metrics of RGHS-WFDEI are lower than that of the HTESSEL. As for the independent validation catchments, the 25% quantile and median KGE and NSE of RGHS-WFDEI (GR4 J-RGHS) are only lower than that of the PCR-GLOBWB. The median AVE does not vary greatly among the models and the median AVE of RGHS-WFDEI is only lower than that of the HTESSEL. However, the 75% quantile NSE and AVE of RGHS-WFDEI have no advantages.

3.3.4. Global and continental water resources

The spatial variability of long-term (1983–2015) averaged annual runoff is presented in Figure 14, which shows a very high spatial variability around the world. However, the spatial distributions of the long-term annual runoff from four RGHSs in this study are consistent. High long-term annual runoffs are obtained for the Amazon, maritime Southeast Asia, central Africa, and the southern section of the Andes in Patagonia. The long-term averaged annual runoffs in central Australia, North Africa, the Middle East, and the northwest of North America are low. The global long-term annual runoffs estimated by RGHSs are 46,810 km^3/yr for GR4 J-RGHS, 42,733 km^3/yr for SIMHYD-RGHS, 42,592 km^3/yr for XAJ-RGHS, and 45,100 km^3/yr for HMETS-RGHS. Overall, the simulated continental runoffs from the GR4 J-RGHS and HMETS-RGHS are larger than those from the SIMHYD-RGHS and XAJ-RGHS, especially for Asia and South America.

The global and continental water resources are calculated using RGHSs and compared to those calculated using the other five GHMs in literatures (Table 6). The six individual simulations (including the results of this study) of global long-term average water resources encompass a range between 29,485 and 46,810 km^3/yr. The difference between these individual global water resource simulations is 17,325 km^3/yr, which is even much higher than the global consumptive water use estimated using Global Water Use Model of WaterGAP 2 (1,250 km^3/yr in 1995). A comparative study of seven models has shown that the range of predicted global runoff from different models was about 45% of the mean simulated runoff, which means that only limited agreement exists between GHMs (Haddeland *et al.* 2011). In fact, it is difficult to compare different GHMs, for there are different periods, data quality, spatial and temporal resolution, and so on of the data used in the building of the GHMs.

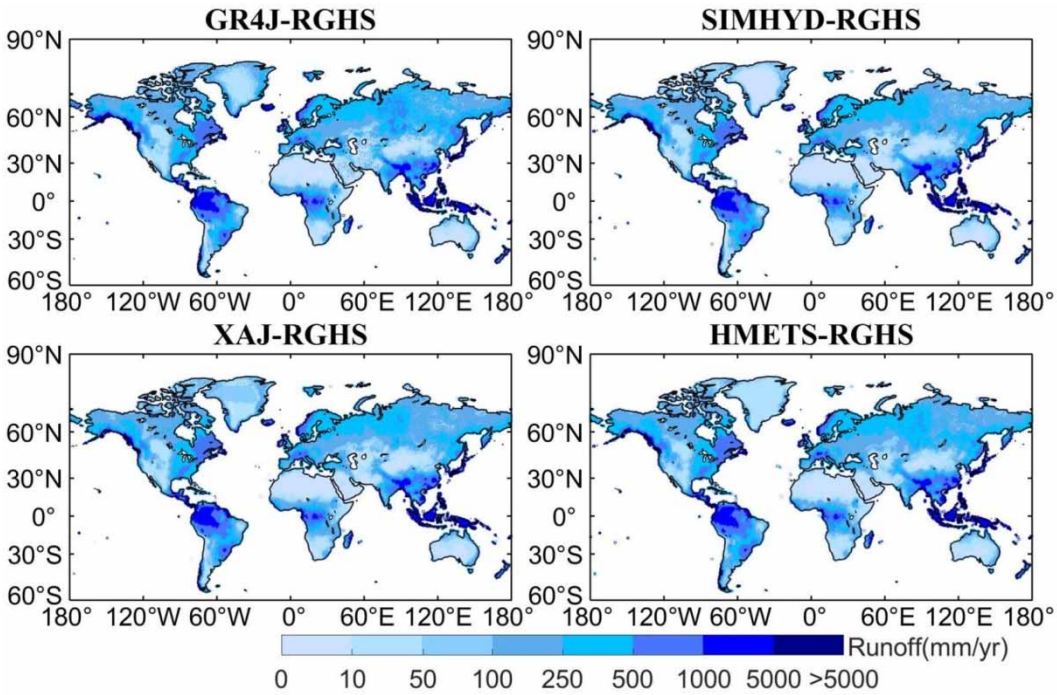


Figure 14 | Long-term averaged annual total runoff from land and open water fraction of cell (time period 1983–2015), in mm/yr.

Table 6 | Global and continental runoff estimates in km^3/yr

Long-term average annual runoff	K78	O01	D03	GRDC	W07	GR4 J-RGHS	SIMHYD-RGHS	XAJ-RGHS	HMETS-RGHS
Global (except Antarctica)	44,560	29,485	36,687	40,533	38,605	46,810	42,733	42,592	45,100
Europe	2,970	2,191	2,763	3,083	3,669	3,396	3,219	3,278	3,377
Asia	14,100	9,385	11,234	13,848	13,611	16,146	14,210	13,864	15,369
Africa	4,600	3,616	3,529	3,690	3,738	5,383	4,732	4,938	4,668
North America[a]	8,180	3,824	5,540	6,294	7,009	6,927	6,256	6,429	6,877
South America	12,200	8,789	11,382	11,897	9,448	12,263	11,686	11,563	12,142
Oceania[b]	2,510	1,680	2,239	1,722	1,129	2,695	2,630	2,521	2,668

The last four columns are from this study.
K78, Korzun *et al.* (1978), Table 157, time period not specified.
O01, Oki *et al.* (2001), Table 2, land surface models and TRIP routing model, time period 1987–1988.
D03, Döll *et al.* (2003), Table 1, model WGHM, time period 1961–1990.
GRDC (2004), time period 1961–1990.
W07, Widén-Nilsson *et al.* (2001), Table 2, model WASMOD-M, time period 1915–2000.
[a]Includes Greenland, except W07 who only simulated a minor part of Greenland.
[b]Oceania is defined as Australia, New Zealand, Papua New Guinea, and some small Islands.

Therefore, we further compared the inter-annual variation of global runoff from RGHSs and the EartH2Observe project for the same period (1983–2012, see Figure 15). The results show that inter-annual variability of global runoff is relatively consistent. However, there is a large difference in magnitudes. In particular, large differences exist for nine global models from the EartH2Observe project, even if they were all driven by the same precipitation dataset (i.e. WFDEI). However, the annual global runoff from RGHSs driven by GPCC and WFDEI are similar and in the range of nine global models, probably due to the fact that the same GSRS were used for global runoff simulations.

4. DISCUSSION

4.1. The impact of equifinality on regionalization

The regionalization methods used in this study are called step-wise regionalization, namely, the model parameters are calibrated, and then transferred to the ungauged regions (Samaniego *et al.* 2017). The equifinality problem of hydrological models may be one of the problems when using this method (Bárdossy 2007; Götzinger & Bárdossy 2007), because only

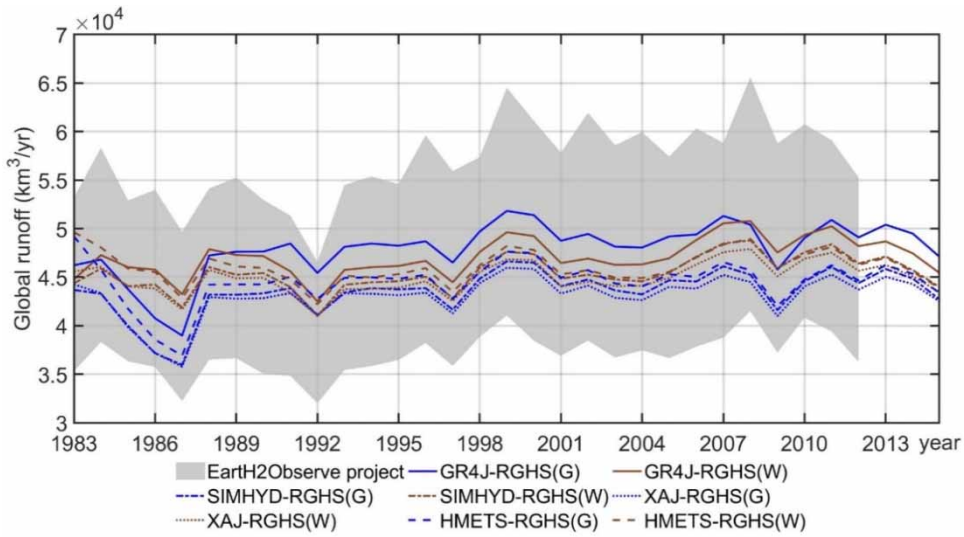


Figure 15 | The inter-annual variation of global runoff from RGHSs and the EartH2Observe project. G represents RGHSs driven by GPCC and W represents RGHSs driven by WFDEI.

one parameter set is used for the regionalization, even though the hydrological model may have multiple sets of parameters that lead to equally acceptable model performance (Beven & Freer 2001; Samaniego *et al.* 2010). To illustrate the influence of equifinality on the regionalization performance, another nine calibration sets were generated during model calibration with different initial random seeds. The difference of median KGE values between these sets is no more than 0.05. For each of 2,277 catchments, one of the nine calibrated parameter sets was then randomly selected for the catchment scale regionalization. Figure 16 shows the performance of the randomly selected parameter set and the original parameter set under the threshold of 0.5. The results show that the regionalization performance of the randomly selected parameter set is consistent with the results of the original parameter set. This indicates that the equifinality does not contribute significantly to the overall uncertainty in the applications of regionalization methods. A similar conclusion was also drawn by Arsenault & Brissette (2014).

4.2. Uncertainties in global water resource estimations

The comparison of different models showed that great uncertainties exist in global water resources simulation. The uncertainties are usually induced by meteorological input, hydrological model structures and parameters (Nijssen *et al.* 2001; Tuo *et al.* 2016; Sawunyama & Hughes 2018). This study preliminarily investigated the impacts of precipitation inputs on global water resources by running RGHSs by using two different precipitation datasets (i.e. GPCC and WFDEI). The global long-term

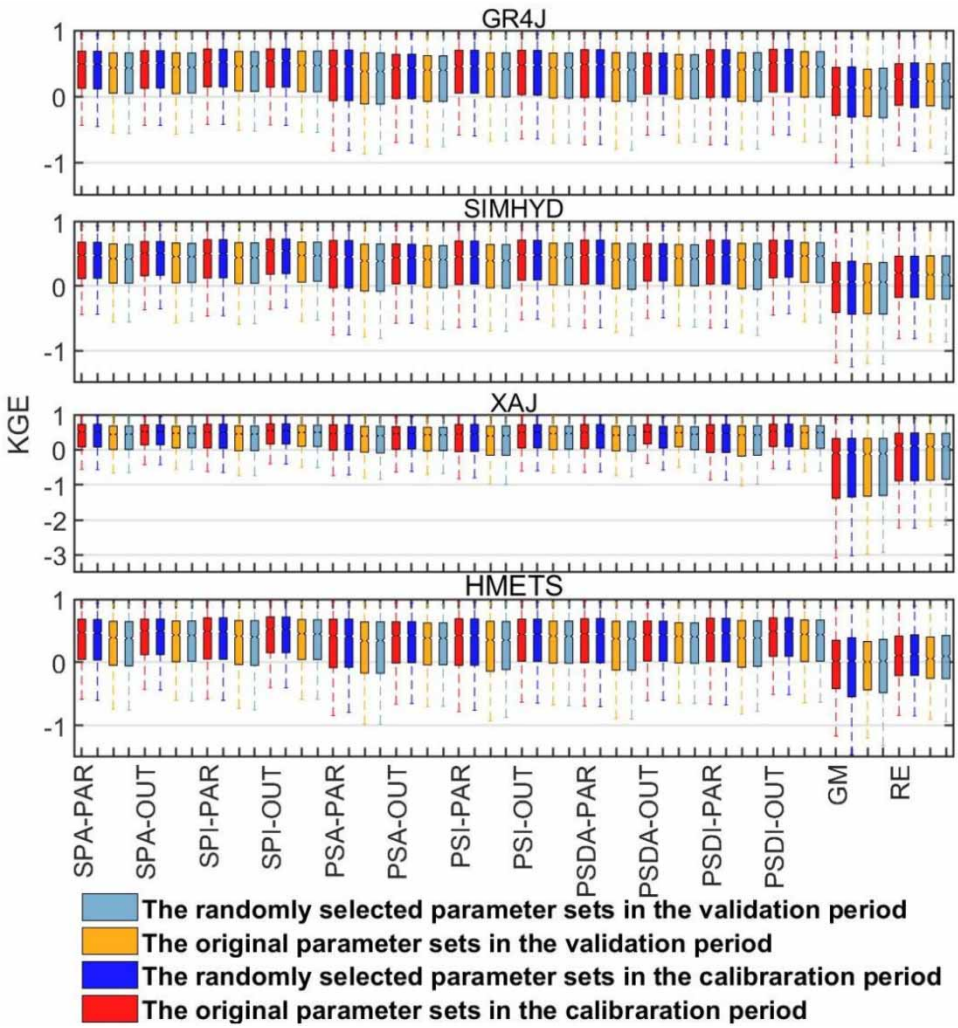


Figure 16 | Comparison of model efficiencies on ungauged catchments using several regionalization schemes between original parameter sets and randomly selected parameter sets.

annual runoffs estimated by RGHS-GPCC range from 42,592 to 46,810 km^3/yr and those estimated by RGHS-WFDEI range from 48,132 to 53,365 km^3/yr. This indicates that the precipitation input is one of the largest uncertainty sources for global water resources estimation.

In addition, the difference of the global long-term average annual runoff is 4,218 km^3/yr between RGHS-GR4 J and RGHS-XAJ (driven by GPCC), indicating that the instability of hydrological models could also result in uncertainty in global water resources estimation. Similar conclusions were also drawn in Schellekens *et al.* (2016). They simulated the global water budget by using ten global models driven by the same precipitation data and found significant differences in global runoff among these models. In fact, due to the complexity of water cycle and the lack of precise understanding of natural hydrological rules, each hydrological model emphasizes different aspects and has its own conceptualization of hydrological processes, which leads to uncertainty in water resources estimation. In addition, multiple objective functions and various optimization algorithms in hydrological models would result in different parameter sets and may lead to a certain level of uncertainty in water resource modelling (Kunnath-Poovakka *et al.* 2021). Therefore, the reduction of uncertainty for global water resources estimation is an avenue for future studies.

Moreover, the observed discharge rather than natural discharge was used to calculate the parameter maps and global water resources. Almost all large rivers are regulated (Vörösmarty *et al.* 2004; Nilsson *et al.* 2005), which may result in modified high-flow and low-flow and affect the performance of nature catchment streamflow simulation and the performance of regionalization of hydrological models (Widén-Nilsson *et al.* 2007; Lehner *et al.* 2011). Therefore, catchments larger than 50,000 km^2 were discarded in this study, as regulation by dams exists in larger catchments with high probability (Vörösmarty *et al.* 2004; Nilsson *et al.* 2005). In general, considering the impact caused by regulated catchments is a challenge in global hydrological modelling due to the lack of global information to an adequate accuracy. Many GHMs, therefore, do not include regulation effects. For example, Widén-Nilsson *et al.* (2009) assumed that regulation has less influence on average flow volumes and used long-term average runoff instead of time series to minimize the effect of the regulation problem on model calibration. Beck *et al.* (2016) used the Global Reservoir and Dam (GRanD) database (v1.1) (Lehner *et al.* 2011) to exclude catchments influenced by reservoirs. Arheimer *et al.* (2020) concluded the potential and difficulty to improve the global model by individual modelling river regulation. Therefore, further studies need to be carried out to consider the influence of human activities (e.g. regulation) to improve the performance of regionalization and GHMs.

5. CONCLUSION

This study first evaluated the performance of multiple regionalization methods over 2,277 catchments distributed around the world based on four catchment hydrological models. A GSRS was then selected to regionalize hydrological parameters at 0.5° grid cell, and four RGHSs (GR4 J-RGHS, SIMHYD-RGHS, XAJ-RGHS, and HMETS-RGHS) were presented for global water resource estimation. Finally, nine global models from the EartH2Observe project were selected to further validate the performance of RGHSs in global water resources estimation. The following conclusions can be drawn.

(1) The SPI-OUT method offered the best results with the highest KGE value when using the 0.5 efficiency threshold for selecting donor catchments, and the optimal number of donor catchments ranges between 3 and 6 for the output averaging option.
(2) Compared to the models from the Earth2Observe project, the presented RGHSs provide better simulation results. For example, the median KGE values of RGHSs were from 0.374 (HMETS-RGHS) to 0.385 (SIMHYD-RGHS). However, the median KGE values of EartH2Observe were −0.06 (PCR-GLOBWB) to 0.124 (HBV-SIMREG). Similar results were also observed when using NSE and AVE as evaluation metrics. On the whole, RGHSs offered great performance in the cold and temperate regions, while the poor performance of the RGHSs was observed in the arid climate regions.
(3) The RGHSs showed reasonable estimations of global water resources. The global long-term average annual runoff estimated by RGHSs are 46,810 (for GR4 J-RGHS), 42,733 (for SIMHYD-RGHS), 42,592 (for XAJ-RGHS), and 45,100 (for HMETS-RGHS) km^3/yr, which is in the good range of those estimated by other GHMs in literatures.

ACKNOWLEDGEMENTS

This work was partially supported by the National Key Research and Development Program of China (No. 2017YFA0603704), the National Natural Science Foundation of China (Grant Nos. 51779176, 51539009), the Overseas Expertise Introduction Project for Discipline Innovation (111 Project) funded by the Ministry of Education of China and

State Administration of Foreign Experts Affairs P.R. China (Grant No. B18037), the Research Council of Norway (FRINATEK Project 274310), and the Thousand Youth Talents Plan from the Organization Department of CCP Central Committee (Wuhan University, China).

DATA AVAILABILITY STATEMENT

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

REFERENCES

Abdulla, F. A. & Lettenmaier, D. P. 1997 Development of regional parameter estimation equations for a macroscale hydrologic model. *Journal of Hydrology* **197** (1–4), 230–257.

Abebe, M. A., Tekli, J., Getahun, F., Chbeir, R. & Tekli, G. 2020 Generic metadata representation framework for social-based event detection, description, and linkage. *Knowledge-Based Systems* **188**, 104817.

Alfieri, L., Lorini, V., Hirpa, F. A., Harrigan, S., Zsoter, E., Prudhomme, C. & Salamon, P. 2020 A global streamflow reanalysis for 1980–2018. *Journal of Hydrology X* **6**, 100049.

Arnell, N. W. 1999 A simple water balance model for the simulation of streamflow over a large geographic domain. *Journal of Hydrology* **217** (3–4), 314–335.

Arnell, N. W. 2003 Effects of IPCC SRES* emissions scenarios on river runoff: a global perspective.

Arsenault, R. & Brissette, F. P. 2014 Continuous streamflow prediction in ungauged basins: the effects of equifinality and parameter set selection on uncertainty in regionalization approaches. *Water Resources Research* **50** (7), 6135–6153.

Arsenault, R., Bazile, R., Ouellet Dallaire, C. & Brissette, F. 2016 CANOPEX: A Canadian hydrometeorological watershed database. *Hydrological Processes* **30** (15), 2734–2736.

Arsenault, R., Brissette, F., Chen, J., Guo, Q. & Dallaire, G. 2020 *NAC2H: The North American Climate Change and Hydroclimatology Data Set*. p. e2020WR027097.

Bárdossy, A. 2007 Calibration of hydrological model parameters for ungauged catchments. *Hydrology and Earth System Sciences* **11** (2), 703–710.

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.

Beck, H. E., Pan, M., Lin, P., Seibert, J., Dijk, A. I. J. M. & Wood, E. F. 2020 Global fully distributed parameter regionalization based on observed streamflow from 4,229 headwater catchments. *Journal of Geophysical Research: Atmospheres* **125** (17), e2019JD0 31485.

Behboudian, M., Kerachian, R., Motlaghzadeh, K. & Ashrafi, S. 2021 Evaluating water resources management scenarios considering the hierarchical structure of decision-makers and ecosystem services-based criteria. *Science of the Total Environment* **751**, 141759.

Beven, K. & Freer, J. 2001 Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology* **249** (1–4), 11–29.

Bierkens, M. F. P. 2015 Global hydrology 2015: state, trends, and directions. *Water Resources Research* **51** (7), 4923–4947.

Boumenni, H., Bachnou, A. & Alaa, N. E. 2017 The rainfall-runoff model GR4 J optimization of parameter by genetic algorithms and Gauss-Newton method: application for the watershed Ourika (High Atlas, Morocco). *Arabian Journal of Geosciences* **10** (5), 343.

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

Chandanpurkar, H. A., Reager, J. T., Famiglietti, J. S. & Syed, T. H. 2017 Satellite- and reanalysis-based mass balance estimates of global continental discharge (1993–2015). *Journal of Climate* **30** (21), 8481–8495.

Chen, J., Li, C., Brissette, F. P., Chen, H., Wang, M. & Essou, G. R. C. 2018 Impacts of correcting the inter-variable correlation of climate model outputs on hydrological modeling. *Journal of Hydrology* **560**, 326–341.

Chiew, F. H. 2010 Lumped conceptual rainfall-runoff models and simple water balance methods: overview and applications in ungauged and data limited regions. *Geography Compass* **4** (3), 206–225.

Chiew, F. H., Peel, M. C. & Western, A. W. 2002 Application and testing of the simple rainfall-runoff model SIMHYD. In: *Mathematical Models of Small Watershed Hydrology and Applications* (Singh, V. P. & Frevert, D., eds), pp. 335–367.

Dee, D. P., Uppala, S. M., Simmons, A. J. *et al.* 2011 The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society* **137** (656), 553–597.

Demirel, M. C., Booij, M. & Hoekstra, A. 2015 The skill of seasonal ensemble low-flow forecasts in the Moselle River for three different hydrological models. *Hydrology and Earth System Sciences*.

Döll, P. & Lehner, B. 2002 Validation of a new global 30-min drainage direction map. *Journal of Hydrology* **258** (1–4), 214–231.

Döll, P., Kaspar, F. & Lehner, B. 2003 A global hydrological model for deriving water availability indicators: model tuning and validation. *Journal of Hydrology* **270** (1–2), 105–134.

Döll, P., Douville, H., Güntner, A., Müller Schmied, H. & Wada, Y. 2015 Modelling freshwater resources at the global scale: challenges and prospects. *Surveys in Geophysics* **37** (2), 195–221.

Duan, Q., Sorooshian, S. & Gupta, V. 1992 Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research* **28** (4), 1015–1031.

Duan, Q., Gupta, V. K. & Sorooshian, S. 1993 Shuffled complex evolution approach for effective and efficient global minimization. *Journal of Optimization Theory and Applications* **76** (3), 501–521.

Fuchs, T., Schneider, U. & Rudolf, B. 2009 The Global Precipitation Climatology Centre (GPCC)-in situ observation based precipitation climatology on regional and global scale. *EGUGA*, 10519.

Gericke, O. J. & Smithers, J. C. 2014 Review of methods used to estimate catchment response time for the purpose of peak discharge estimation. *Hydrological Sciences Journal* **59** (11), 1935–1971.

Ghiggi, G., Humphrey, V., Seneviratne, S. & Gudmundsson, L. 2019 GRUN: an observation-based global gridded runoff dataset from 1902 to 2014. *Earth System Science Data* **11**, 1655–1674.

Gong, L., Widén-Nilsson, E., Halldin, S. & Xu, C. Y. 2009 Large-scale runoff routing with an aggregated network-response function. *Journal of Hydrology* **368** (1–4), 237–250.

Gong, L., Halldin, S. & Xu, C.-Y. 2011 Global-scale river routing – an efficient time-delay algorithm based on HydroSHEDS high-resolution hydrography. *Hydrological Processes* **25**, 1114–1128.

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

Graham, S. T., Famiglietti, J. S. & Maidment, D. R. 1999 Five-minute, 1/2°, and 1° data sets of continental watersheds and river networks for use in regional and global hydrologic and climate system modeling studies. *Water Resources Research* **35** (2), 583–587.

Grill, G., Lehner, B., Thieme, M. *et al.* 2019 Mapping the world's free-flowing rivers. *Nature* **569** (7755), 215–221.

Gu, L., Chen, J., Xu, C.-Y., Wang, H.-M. & Zhang, L. 2018 Synthetic impacts of internal climate variability and anthropogenic change on future meteorological droughts over China. *Water* **10** (11).

Gupta, H. V., Kling, H., Yilmaz, K. K. & Martinez, G. F. 2009 Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. *Journal of Hydrology* **377** (1–2), 80–91.

Haddeland, I., Clark, D. B., Franssen, W. *et al.* 2011 Multimodel estimate of the global terrestrial water balance: setup and first results. *Journal of Hydrometeorology* **12** (5), 869–884.

Harris, I., Jones, P., Osborn, T. & Lister, D. 2014 Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 dataset. *International Journal of Climatology* **34**, 623–642.

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

Hundecha, Y. & Bárdossy, A. 2004 Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model. *Journal of Hydrology* **292** (1–4), 281–295.

Jin, X., Xu, C.-y., 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 (in Chinese).

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.

Knoben, W. J. M., Freer, J. E. & Woods, R. A. 2019 Technical note: inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrology and Earth System Sciences* **23** (10), 4323–4331.

Kottek, M., Grieser, J., Beck, C., Rudolf, B. & Rubel, F. 2006 World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift* **15** (3), 259–263.

Kunnath-Poovakka, A., Ryu, D., Eldho, T. I. & George, B. 2021 Parameter uncertainty of a hydrologic model calibrated with remotely sensed evapotranspiration and soil moisture. *Journal of Hydrologic Engineering* **26** (3).

Lehner, B. 2012 *Derivation of Watershed Boundaries for GRDC Gauging Stations Based on the HydroSHEDS Drainage Network*. Global Runoff Data Centre in the Federal Institute of Hydrology (BFG).

Lehner, B., Liermann, C. R., Revenga, C. *et al.* 2011 High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Frontiers in Ecology and the Environment* **9** (9), 494–502.

Li, H. & Zhang, Y. 2017 Regionalising rainfall-runoff modelling for predicting daily runoff: comparing gridded spatial proximity and gridded integrated similarity approaches against their lumped counterparts. *Journal of Hydrology* **550**, 279–293.

Li, F., Zhang, Y., Xu, Z., Liu, C., Zhou, Y. & Liu, W. 2014 Runoff predictions in ungauged catchments in southeast Tibetan Plateau. *Journal of Hydrology* **511**, 28–38.

Li, J., Zhao, H., Zhang, J. *et al.* 2020 An improved routing algorithm for a large-scale distributed hydrological model with consideration of underlying surface impact. *Hydrology Research* **51** (5), 834–853.

Liang, S. & Greene, R. 2020 A high-resolution global runoff estimate based on GIS and an empirical runoff coefficient. *Hydrology Research* **51** (3), 1238–1260.

Liang, X., Lettenmaier, D. P., Wood, E. F. & Burges, S. J. 1994 A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research: Atmospheres* **99** (D7), 14415–14428.

Martel, J.-L., Demeester, K., Brissette, F. P., Arsenault, R. & Poulin, A. 2017 HMET: a simple and efficient hydrology model for teaching hydrological modelling, flow forecasting and climate change impacts. *The International Journal of Engineering Education* **33** (4), 1307–1316.

Martens, B., Miralles, D. G., Lievens, H. *et al.* 2017 GLEAM v3: satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development* **10** (5), 1903–1925.

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

Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A. & Dolman, A. J. 2010 Global land-surface evaporation estimated from satellite-based observations. *Hydrology and Earth System Sciences Discussions* **7** (5), 8479–8519.

Mizukami, N., Rakovec, O., Newman, A. J., Clark, M. P., Wood, A. W., Gupta, H. V. & Kumar, R. 2019 On the choice of calibration metrics for 'high-flow' estimation using hydrologic models. *Hydrology and Earth System Sciences* **23** (6), 2601–2614.

Müller Schmied, H., Adam, L., Eisner, S. *et al.* 2016 Variations of global and continental water balance components as impacted by climate forcing uncertainty and human water use. *Hydrology and Earth System Sciences* **20** (7), 2877–2898.

Muller-Wohlfeil, D.-I., Xu, C.-Y. & Iversen, H. L. 2003 Estimation of monthly river discharge from Danish catchments. *Nordic Hydrology* **34** (4), 295–320.

Naghdi, S., Bozorg-Haddad, O., Khorsandi, M. & Chu, X. 2021 Multi-objective optimization for allocation of surface water and groundwater resources. *Science of the Total Environment* **776**, 146026.

Neri, M., Parajka, J. & Toth, E. 2020 Importance of the informative content in the study area when regionalising rainfall-runoff model parameters: the role of nested catchments and gauging station density. *Hydrology and Earth System Sciences* **24** (11), 5149–5171.

Nijssen, B., O'Donnell, G. M., Lettenmaier, D. P., Lohmann, D. & Wood, E. F. 2001 Predicting the discharge of global rivers. *Journal of Climate* **14** (15), 3307–3323.

Nilsson, C., Reidy, C. A., Dynesius, M. & Revenga, C. 2005 Fragmentation and flow regulation of the world's large river systems. *Science* **308** (5720), 405–408.

Oki, T. & Kanae, S. 2006 Global hydrological cycles and world water resources. *Science* **313** (5790), 1068–1072.

Oki, T., Agata, Y., Kanae, S., Saruhashi, T., Yang, D. & Musiake, K. 2001 Global assessment of current water resources using total runoff integrating pathways. *Hydrological Sciences Journal* **46** (6), 983–995.

Oudin, L., Andréassian, V., Perrin, C., Michel, C. & Le Moine, N. 2008 Spatial proximity, physical similarity, regression and ungaged catchments: a comparison of regionalization approaches based on 913 French catchments. *Water Resources Research* **44** (3), W03413.

Parajka, J., Blöschl, G. & Merz, R. 2007 Regional calibration of catchment models: potential for ungauged catchments. *Water Resources Research* **43** (6), W06406.

Perrin, C., Michel, C. & Andréassian, V. 2003 Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology* **279** (1–4), 275–289.

Petelet-Giraud, E., Cary, L., Cary, P. *et al.* 2018 Multi-layered water resources, management, and uses under the impacts of global changes in a southern coastal metropolis: when will it be already too late? crossed analysis in Recife, NE Brazil. *Science of the Total Environment* **618**, 645–657.

Pilgrim, D., Chapman, T. & Doran, D. 1988 Problems of rainfall-runoff modelling in arid and semiarid regions. *Hydrological Sciences Journal* **33** (4), 379–400.

Pokhrel, P., Gupta, H. V. & Wagener, T. 2008 A spatial regularization approach to parameter estimation for a distributed watershed model. *Water Resources Research* **44** (12), W12419.

Pool, S., Viviroli, D. & Seibert, J. 2017 Prediction of hydrographs and flow-duration curves in almost ungauged catchments: which runoff measurements are most informative for model calibration? *Journal of Hydrology* **554**, 613–622.

Qi, W.-y., Chen, J., Li, L., Xu, C.-Y., Xiang, Y.-h., Zhang, S.-b. & Wang, H.-M. 2021a Impact of the number of donor catchments and the efficiency threshold on regionalization performance of hydrological models. *Journal of Hydrology* **601**.

Qi, W.-y., Chen, J., Xu, C.-Y. & Wan, Y. 2021b Finding the optimal multimodel averaging method for global hydrological simulations. *Remote Sensing* **13**, 2574.

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

Roushangar, K., Ghasempour, R., Kirca, V. S. O. & Demirel, M. C. 2021 Hybrid point and interval prediction approaches for drought modeling using ground-based and remote sensing data. *Hydrology Research* **52** (6), 1469–1489.

Samaniego, L., Kumar, R. & Attinger, S. 2010 Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research* **46** (5), W05523.

Samaniego, L., Kumar, R., Thober, S. *et al.* 2017 Toward seamless hydrologic predictions across spatial scales. *Hydrology and Earth System Sciences* **21** (9), 4323–4346.

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.

Santos, L., Thirel, G. & Perrin, C. 2018 Technical note: Pitfalls in using log-transformed flows within the KGE criterion. *Hydrology and Earth System Sciences* **22** (8), 4583–4591.

Sawunyama, T. & Hughes, D. A. 2018 Application of satellite-derived rainfall estimates to extend water resource simulation modelling in South Africa. *Water SA* **34** (1), 1–10.

Schellekens, J., Dutra, E., Martínez-de la Torre, A. *et al.* 2016 A global water resources ensemble of hydrological models: the eartH2Observe Tier-1 dataset. *Earth System Science Data Discussions*, 1–35.

Sellers, P. J., Mintz, Y., Sud, Y. C. & Dalcher, A. J. J. o. A. S. 1986 A simple biosphere model (SiB) for use within general-circulation models. **43** (6), 505–531.

Shen, M., Chen, J., Zhuan, M., Chen, H., Xu, C.-Y. & Xiong, L. 2018 Estimating uncertainty and its temporal variation related to global climate models in quantifying climate change impacts on hydrology. *Journal of Hydrology* **556**, 10–24.

Shrestha, N. K., Du, X. & Wang, J. 2017 Assessing climate change impacts on fresh water resources of the Athabasca River Basin, Canada. *Science of the Total Environment* **601–602**, 425–440.

Sood, A. & Smakhtin, V. 2015 Global hydrological models: a review. *Hydrological Sciences Journal* **60** (4), 549–565.

Syed, T. H., Famiglietti, J. S., Chambers, D. P., Willis, J. K. & Hilburn, K. 2010 Satellite-based global-ocean mass balance estimates of interannual variability and emerging trends in continental freshwater discharge. *Proceedings of the National Academy of Sciences of the United States of America* **107** (42), 17916–17921.

Tobler, W. R. 1970 A computer movie simulating urban growth in the Detroit region. *Economic Geography* **46** (Suppl. 1), 234–240.

Tuo, Y., Duan, Z., Disse, M. & Chiogna, G. 2016 Evaluation of precipitation input for SWAT modeling in Alpine catchment: a case study in the Adige river basin (Italy). *Science of the Total Environment* **573**, 66–82.

Van Beek, L. & Bierkens, M. 2008 *The Global Hydrological Model PCR-GLOBWB: Conceptualization, Parameterization and Verification Department of Physical Geography*. Utrecht University, Utrecht, the Netherlands.

Van Beek, L. P., Eikelboom, T., van Vliet, M. T. & Bierkens, M. F. 2012 A physically based model of global freshwater surface temperature. *Water Resources Research* **48** (9), W09530.

Vargas Godoy, M. R., Markonis, Y., Hanel, M., Kyselý, J. & Papalexiou, S. M. 2021 The global water cycle budget: a chronological review. *Surveys in Geophysics* **42** (7), 1075–1107.

Vetter, T., Huang, S., Aich, V., Yang, T., Wang, X., Krysanova, V. & Hattermann, F. 2015 Multi-model climate impact assessment and intercomparison for three large-scale river basins on three continents. *Earth System Dynamics* **6** (1), 17–43.

Vis, M., Knight, R., Pool, S., Wolfe, W. & Seibert, J. 2015 Model calibration criteria for estimating ecological flow characteristics. *Water* **7** (12), 2358–2381.

Viviroli, D. & Seibert, J. 2015 Can a regionalized model parameterisation be improved with a limited number of runoff measurements? *Journal of Hydrology* **529**, 49–61.

Vörösmarty, C. J., Moore, B., Grace, A. L. *et al.* 1989 Continental scale models of water balance and fluvial transport: an application to South America. *Global Biogeochemical Cycles* **3** (3), 241–265.

Vörösmarty, C., Fekete, B. M., Meybeck, M. & Lammers, R. B. 2000 Global system of rivers: its role in organizing continental land mass and defining land-to-ocean linkages. *Global Biogeochemical Cycles* **14** (2), 599–621.

Vörösmarty, C., Lettenmaier, D., Leveque, C. *et al.* 2004 Humans transforming the global water system. *Eos, Transactions American Geophysical Union* **85** (48), 509–514.

Wan, Y., Chen, J., Xu, C.-Y., Xie, P., Qi, W., Li, D. & Zhang, S. 2021 Performance dependence of multi-model combination methods on hydrological model calibration strategy and ensemble size. *Journal of Hydrology* **603**, 127065.

Widén-Nilsson, E., Halldin, S. & Xu, C.-y. 2007 Global water-balance modelling with WASMOD-M: parameter estimation and regionalisation. *Journal of Hydrology* **340** (1–2), 105–118.

Widén-Nilsson, E., Gong, L., Halldin, S. & Xu, C. Y. 2009 Model performance and parameter behavior for varying time aggregations and evaluation criteria in the WASMOD-M global water balance model. *Water Resources Research* **45** (5), W05418.

Xu, C.-Y. 1999a Estimation of parameters of a conceptual water balance model for ungauged catchments. *Water Resources Management* **13** (5), 353–368.

Xu, C.-Y. 1999b From GCMs to river flow: a review of downscaling techniques and hydrologic modeling approaches. *Progress in Physical Geography* **23** (2), 229–249.

Xu, C.-Y. 2003 Testing the transferability of regression equations derived from small sub-catchments to large area in central Sweden. *Hydrology and Earth System Sciences* **7** (3), 317–324.

Xu, C.-Y., Widen, E. & Halldin, S. 2005 Modelling hydrological consequences of climate change – progress and challenges. *Advances in Atmospheric Sciences* **22** (6), 789–797.

Yang, X., Magnusson, J., Rizzi, J. & Xu, C.-Y. 2018 Runoff prediction in ungauged catchments in Norway: comparison of regionalization approaches. *Hydrology Research* **49** (2), 487–505.

Yang, X., Magnusson, J. & Xu, C.-Y. 2019 Transferability of regionalization methods under changing climate. *Journal of Hydrology* **568**, 67–81.

Yang, X., Magnusson, J., Huang, S., Beldring, S. & Xu, C.-Y. 2020 Dependence of regionalization methods on the complexity of hydrological models in multiple climatic regions. *Journal of Hydrology* **582**, 124357.

Zhang, Y. & Chiew, F. 2009 Evaluation of regionalisation methods for predicting runoff in ungauged catchments in southeast Australia. In: *18th World IMACS/MODSIM Congress*, Cairns, Australia, pp. 3442–3448.

Zhang, Y., Vaze, J., Chiew, F. H., Teng, J. & Li, M. 2014 Predicting hydrological signatures in ungauged catchments using spatial interpolation, index model, and rainfall–runoff modelling. *Journal of Hydrology* **517**, 936–948.

Zhang, Y., Zheng, H., Chiew, F. H., Arancibia, J. P. & Zhou, X. 2016 Evaluating regional and global hydrological models against streamflow and evapotranspiration measurements. *Journal of Hydrometeorology* **17** (3), 995–1010.

Zhang, Y. F., Li, Y. P., Sun, J. & Huang, G. H. 2020 Optimizing water resources allocation and soil salinity control for supporting agricultural and environmental sustainable development in Central Asia. *Science of the Total Environment* **704**, 135281.

Zhou, Y. L., Guo, S. L., Xu, C.-Y., Xiong, L. H., Chen, H., Ngongondo, C. & Li, L. 2022 Probabilistic interval estimation of design floods under non-stationary conditions by an integrated approach. *Hydrology Research*. doi:10.2166/nh.2021.007.

First received 10 November 2021; accepted in revised form 6 February 2022. Available online 18 February 2022