Journal of Hydrology 588 (2020) 125016

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Streamflow prediction in "geopolitically ungauged" basins using satellite observations and regionalization at subcontinental scale

Tien L.T. Du[a,b], Hyongki Lee[a], Duong D. Bui[c,*], Berit Arheimer[d], Hong-Yi Li[a], Jonas Olsson[d], Stephen E. Darby[e], Justin Sheffield[e], Donghwan Kim[a], Euiho Hwang[f]

[a] Department of Civil and Environmental Engineering, University of Houston, Houston, TX, USA
[b] Da Nang Institute for Socio-Economic Development, Da Nang, Viet Nam
[c] National Center for Water Resources Planning and Investigation (NAWAPI), Ministry of Natural Resources and Environment (MONRE), Hanoi, Viet Nam
[d] Swedish Meteorological and Hydrological Institute (SMHI), Norrköping, Sweden
[e] School of Geography and Environmental Science, University of Southampton, Southampton, UK
[f] Water Resources Research Center, K-Water Institute, K-Water, Daejeon, South Korea

ARTICLE INFO

Keywords:
Catchment model
Regionalization
Flow correlation
Satellite observations
Altimetry
Mekong

ABSTRACT

A novel approach of combining regionalization and satellite observations of various hydrological variables were employed to significantly improve prediction of streamflow signatures at "geopolitically ungauged" basins. Using the proposed step-wise physiography and climate-based regionalization approach, the model performance at ungauged basins reached 80% of performance of locally calibrated parameters and significantly outperformed the global regionalization parameters. The proposed water level based flow correlation was found to help diagnose models and outperform the existing performance metrics of simulated water levels at ungauged basins. The study also set up the first multi-national, multi-catchment hydrological model in the Greater Mekong region, the top global biodiversity and major disaster risk hotspot in the world through sequential and iterative refinement of the existing global hydrological model. New model setup or existing models in the poorly-gauged and ungauged basins could benefit from the proposed approach to predict and evaluate models at ungauged basins.

1. Introduction

Adequate and reliable information about streamflows are imperative for effective management of water resources. Streamflow data are required for practical applications such as the design of drainage or water supply infrastructure, as well as planning short-term and long-term water use with respect to changes of land use and climate. However, only a small fraction of catchments in any part of the world, possess a stream gauge (Blöschl et al., 2013). Additionally, the number of actively gauged stations has in recent years declined significantly due to reduced government funds for monitoring networks (Ad Hoc Group et al., 2001; Shiklomanov et al., 2002). Given the scarcity of operational gauging stations, the availability of streamflow data is becoming increasingly limited.

In addition to the global trend of declining streamflow gauges, accessing existing data is often more difficult in transboundary river catchments. Unfollowing the human-defined political or administrative boundaries, transboundary river basins account for roughly one-half of the earth's land surface, generate about 60% of the global freshwater flow and are home to nearly 40% of the world's population (UNEP-DHI & UNEP, 2016). At least one transboundary water body exists in almost every non-island state in the world. Even if international agreements enabling data and information sharing among states exist in principle, in practice data sharing is often complex in transboundary waters (Gerlak et al., 2011). For example, in the Okavango River basin, although there is agreement between Namibia and Botswana on sharing river flow data, it is debated how to validate the accuracy of shared data (Turton et al., 2003). Or in the case of the Jordan river basin, where there are asymmetric power relations, intentionally ambiguous mechanisms were designed by stronger states to allow no actual data exchange while diffusing domestic opposition (Fischhendler, 2008). These transboundary river basins are thus considered as "geopolitically ungauged" where data observation networks may exist but data are unavailable for use due to geopolitical constraints (Kibler et al., 2014).

Since streamflow observations are not available and accessible for all locations, hydrological models often rely on regionalization

* Corresponding author.
E-mail address: duongdubui@gmail.com (D.D. Bui).

https://doi.org/10.1016/j.jhydrol.2020.125016

Available online 27 April 2020

approaches to transfer information from gauged to ungauged catchments (see Beck et al., 2016; Razavi and Coulibaly, 2012; Blöschl et al., 2013; Parajka et al., 2013; Hrachowitz et al., 2013; He et al., 2011 for reviews). There are different regionalization approaches with their respective advantages and limitations. In general, approaches that transfer calibrated parameter sets with respect to their climatic and physiographic similarity, and/or simultaneously calibrate multiple catchments with similar characteristics, performed better than other approaches (Arheimer et al., 2020; Beck et al., 2016; Donnelly et al., 2016; Garambois et al., 2015; Kim and Kaluarachchi, 2008; Parajka et al., 2007; Sellami et al., 2014). Nonetheless, it is of question if this approach would work in the case of process-based distributed hydrological models, which inevitably have a large number of parameters.

Process-based distributed hydrological models, which have parameters linked to physiography and/or climate in the context of multi-catchment modeling approach (including both gauged and ungauged basins), is a type of regionalization (Abbaspour et al., 2015; Arheimer et al., 2020; Donnelly et al., 2016; Hossain et al., 2017; Mohammed et al., 2018). Most of these studies used physiography-linked parameter sets, except the most recent study by Arheimer et al. (2020) that included similarity in climate characteristics. Among them, the study by Donnelly et al. (2016) explicitly evaluated physiography-linked parameter sets and concluded that they were useful for prediction at ungauged basins. Arheimer et al. (2020) included climate-linked parameter sets by assigning different potential evapotranspiration algorithms for catchments characterized according to Köppen climate classification. However, it is questionable if the choice of climate regions based on Köppen classification was optimal due to no explicit quantification of improvement in simulating streamflow.

The growing availability of spatially distributed remotely sensed data and open global data sources, together with better computational capacity and advanced methods to assure better data quality, has brought the possibility of macroscale hydrological modeling at the continental scale (e.g. Pechlivanidis and Arheimer, 2015; Abbaspour et al., 2015; Donnelly et al., 2016) and the global scale (Arheimer et al., 2020; Beck et al., 2016; Döll et al., 2003). However, it is known that global scale hydrological models do not always have satisfactory performance over all stations within their expansive domains, constraining their application for management purposes. Furthermore, the evaluation of model performance has been undertaken at gauged or pseudo-gauged stations. Accordingly, it is of question how to discern at which station model can satisfactorily capture the observed hydrological regimes. An innovative approach to evaluate model at ungauged basins without using streamflow data is thus required.

While streamflow data are less available and accessible, water level or stream stage data are more widely obtainable because there is less investment of people and equipment to measure them and increasingly more access and coverage of stage data derived from satellite altimetry become available (Okeowo et al., 2017; Lee et al., 2009). Water level observations either from in-situ observations or derived from satellite altimetry have been increasingly used to calibrate hydrological models towards replacing streamflow information for poorly and ungauged basins (Getirana, 2010; Sun et al., 2012; Lindström, 2016; Jian et al., 2017) and used innovatively to estimate important hydrological information for Mekong river basin (Chang et al., 2019; Kim et al., 2019). However, the evaluation of hydrological models using the existing performance metrics based on water level only can yield inaccurate results due to inherent numerical problems (Lindström, 2016; Jian et al., 2017). Therefore, it is vital to develop methods that more effectively utilize water level data, where available, for evaluation of model at ungauged basins as a surrogate for streamflow.

Similar flow dynamics (mean discharge, relative flow variability and catchment response rates) have been found between catchments having high spatial correlation of daily streamflow ($p > 0.9$), rather than catchments having spatial proximity (Archfield and Vogel, 2010; Betterle et al., 2017; Betterle et al., 2019). Instead of using streamflow for ungauged catchments (receptor), water level observations can be used to find the most highly correlated gauged catchments (donor). This approach is named as water level based flow correlation in this study. If a model can simulate similar correlation to the observed correlation patterns between gauged (using streamflow) and ungauged (using water level), it is hypothesized that performance of simulated ungauged catchments is as similar as gauged catchments. It is thus worth exploring whether this hypothesis is valid.

Accordingly, the overarching goal of this work is to develop and test a new method of using satellite observations and regionalization to improve the prediction of streamflow at "geopolitically ungauged" basins using Hydrological Predictions for the Environment (HYPE) semi-distributed hydrological model (Lindström et al., 2010). A first subcontinent-scale hydrological model would be setup for the Greater Mekong region, which is a global biodiversity and major disaster risk hotspot but poorly simulated in the existing global hydrological models, constraining their use for pressing management purposes (Tordoff et al., 2012; Dilley et al., 2005; Du et al., 2018). The region covers 13 river basins, of which six international river basins make up 90% of total area, passing the entire territory of Vietnam, Laos, Cambodia and parts of China, Thailand and Myanmar (Fig. 1). The specific objectives of this work are to: (i) examine how far a multi-catchment HYPE model using global open data sources including satellite observations can predict flow signatures for gauged catchments in the region; (ii) identify whether physiography and climate based regionalized parameters could improve prediction of streamflow signatures at ungauged catchments; (iii) determine whether water level based flow correlation could help to evaluate model performance at ungauged catchments.

2. Data

2.1. Input dataset for HYPE model

The study used HYPE semi-distributed hydrological model, which has been examined in extensive catchment types worldwide (Arheimer et al., 2020). In this study, HYPE for the Greater Mekong region is named as Greater Mekong HYPE (GM-HYPE), which has been developed incrementally, and the current final version 1.3 (GM-HYPE v1.3) was based on the first version (GM-HYPE v1). GM-HYPE v1 was the result of a long-standing collaboration between Swedish Meteorological Hydrological Institute (SMHI) and National Center for Water Resources Planning and Investigation (NAWAPI). To be comparable with the World-Wide HYPE model version 1.3 (WWH v1.3, Arheimer et al., 2020), the catchment model HYPE for the Greater Mekong region used the same topography and hydrological databases (Table 1). Additionally, supplementary forcing and gauging data were used. In addition to Hydrological Global Forcing Data (HydroGFD) precipitation and HydroGFD temperature (Berg et al., 2018), Multi-Source Weighted-Ensemble Precipitation (MSWEP) precipitation, Tropical Rainfall Measuring Mission (TRMM 3B42) precipitation and National Centers for Environmental Prediction Climate Forecast System version 2 (NCEP CFSv2) temperature, which have been examined to perform well in the region, were added (Mohammed et al., 2018; Tang et al., 2019). Since different forcing datasets have different spatial resolutions (Table 1), the nearest grid approach was used to assign the characteristics to each sub-catchment. Considering Vietnam to be the country that needs to monitor water resources outside of the country (given 60% of water is generated outside of the country; World Bank, 2019; Du et al., 2016), any streamflow observations inside Vietnam are named as gauged catchments (used for calibration) whereas the observations outside of Vietnam, where available, are named as "geopolitically ungauged" catchments (gauged but not used for calibration). Sources of the additional ground observations of streamflow, water level and precipitation were supplemented by project partners to calibrate and validate the

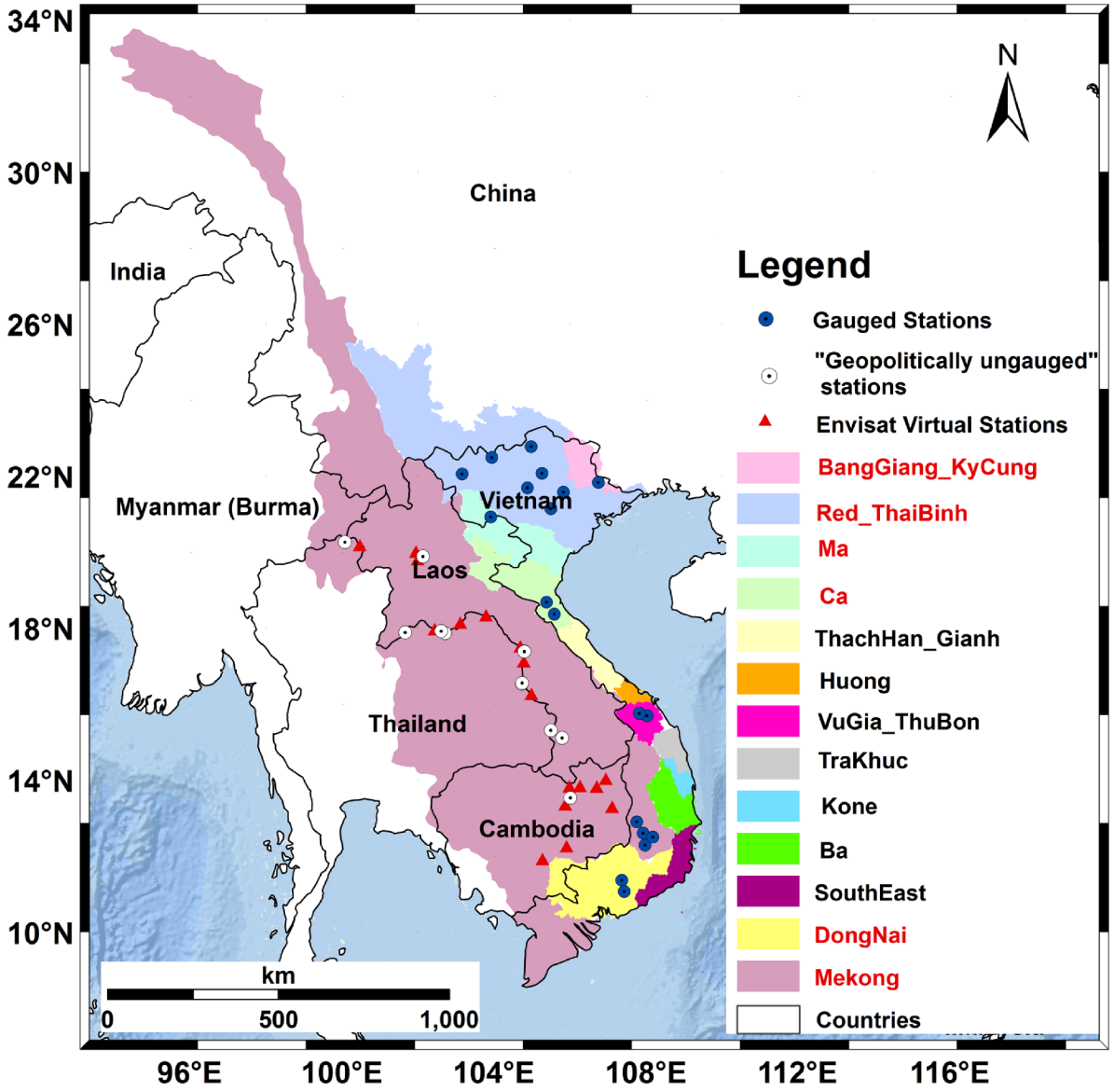

Fig. 1. The study domain of Greater Mekong, covering thirteen river basins, six of which are international transboundary river basins (red colored legends) (including the entire territory of Vietnam, Laos, Cambodia and part of China, Thailand and Myanmar) whereas the remaining river basins are located inside Vietnam. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

model performance in the region (see Fig. 1 for their locations. Details of stations' names, locations and basic information are provided in Table Supplementary 1).

2.2. Radar altimetry data

The heights of the earth surface every 35 days can be determined using the two-way travel time of radar pulses by Envisat Radar Altimeter 2 (RA 2) during period from August 2002 to October 2010 (see Fig. 2 and Table 2 for their locations). Altimetric along-track data v2.1 of the Envisat mission (CTOH_ENVISAT_2014_01) corrected by CTOH (Centre de Topographie des Océans et de l'Hydrosphère, LEGOS, France) were extracted and time series were generated using the automation algorithm developed in Okeowo et al. (2017). This algorithm was based on K-means clustering for the automatic detection of outliers. Their method was found to be computationally effective compared to other methods, such as Kalman filter approach by Schwatke et al. (2015) and applicable in the Mekong river basin (Kim et al., 2019b).

3. Methods

3.1. The multi-catchment hydrological model HYPE

HYPE is a process-oriented semi-distributed open-source model that is developed and used operationally to deliver high-resolution model predictions of water and nutrients (Lindström et al., 2010; Arheimer and Lindström, 2013). Initially developed for use in Sweden, it has more recently been used in applications in, for example, India (Pechlivanidis and Arheimer, 2015), Europe (Donnelly et al., 2016), and across the globe (Arheimer et al., 2020). The HYPE model code has been developed since 2005 with a flexible approach to start with simple process descriptions and further refine and increase complexity when necessary (Lindström et al., 2005; Bergström, 1991; Beven, 2011). The model structure is based on a multi-catchment approach allowing simultaneous modeling of multiple river basins, with each river basin divided into multiple subbasins and each subbasin further divided into hydrologic response units (HRUs). Each HRU can be divided vertically

Table 1
Data description and sources used in the Greater Mekong HYPE project.

Data type	Source and resolution	Reference
Topography (Flow accumulation, flow direction, digital elevation, river width)	SRTM (3 arcsec)	USGS
	HYDRO1k (30 arcsec)	UGGS
	GWD-LR (3 arcsec)	Yamazaki et al., 2014
Floodplains and Lake	Global Lake and Wetland Database (GLWD)	Lehner and Döll, 2004
Land Cover characteristics	ESA CCI Landcover v1.6.1 epoch 2010 (300 m)	Lehner et al., 2011; ESA Climate Change Initiative – Land Cover project
Precipitation	MSWEP (0.25° grid, 1979 – 2014)	Beck et al. 2017
	TRMM 3B42 (0.25° grid, 2001 – 2015)	Huffman et al., 2006
	HydroGFD (0.5° grid, 1961 – 2015)	Berg et al., 2018
	In-situ precipitation stations in Vietnam (176 stations, 1975 – 2006)	BIG DREAM project (VINIF.2019.DA17)
Temperature	HydroGFD (0.5° grid, 1961 – 2015)	Berg et al., 2018
	NCEP CFSv2 (0.25° grid,1979 – 2014)	Saha et al., 2011
Potential Evapotranspiration	MOD16A2 (8-day 1 km, 2001 – 2010)	Mu et al., 2011
Streamflow observations in Vietnam (Gauged) (used for calibration)	19 Stations (daily, 1980–2010)	BIG DREAM project (VINIF.2019.DA17)
Observations of streamflow and water level in Mekong ("geopolitically ungauged") (used for independent evaluation)	12 Stations (daily, 1980 – 2007)	Mekong-SERVIR project (ADPC)
Envisat-derived Water Level (Envisat-"ungauged"*) (used for independent evaluation)	17 Virtual Stations (daily every 35 days, 2002–2009)	Okeowo et al., 2017; Lee et al., 2009; Chang et al., 2019; Kim et al., 2019b; CTOH_ENVISAT_2014_01

*Envisat-"ungauged" catchments are catchments that have virtual stations of Envisat-derived water level but mostly have no observations of streamflow (except 3 catchments that are located in "geopolitically ungauged" catchments) (more explanations are provided in Table 2).

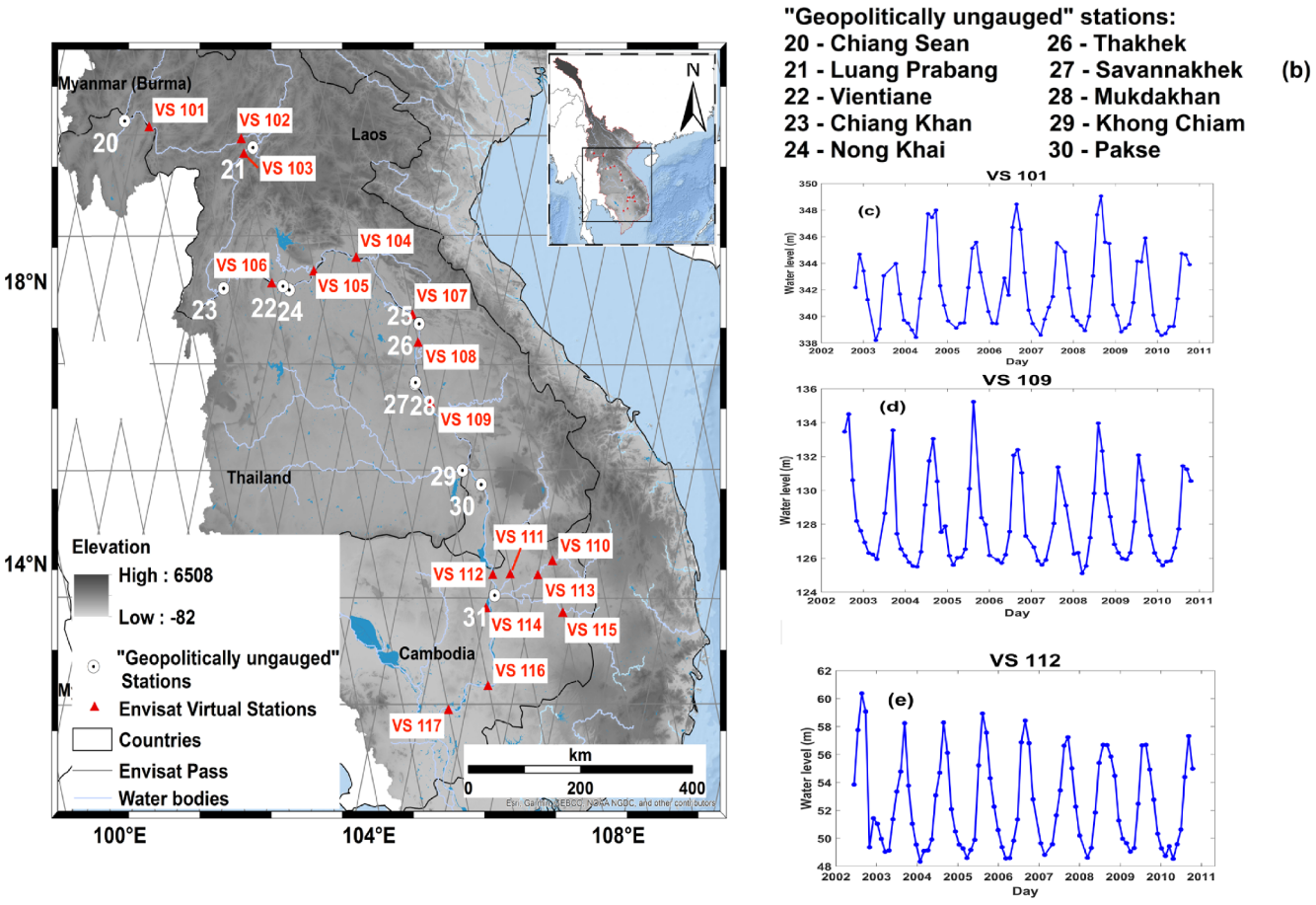

Fig. 2. Spatial distribution of virtual stations (red triangles) and "geopolitically ungauged" stations (white circles) employed in the study (Fig. 2a). The black lines denote the ground tracks of Envisat altimetry. Fig. 2b shows the names of "geopolitically ungauged" stations. Time series of river elevation at three VS's are shown in the panels on the right (Fig. 2c, 2d, 2e) (time series of all locations are not presented for reason of brevity). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
List of seventeen virtual stations (VSs) with Envisat pass numbers and their location.

No.	VS	Pass number	Location (Lat/Lon)	Located in "Geopolitically ungauged" catchments?
1.	VS 101	737	20.195°N/100.472°E	No – Envisat-"ungauged" catchment (EU101)
2.	VS 102	651	20.025°N/101.950°E	No – Envisat-"ungauged" catchment (EU102)
3.	VS 103	651	19.817°N/101.994°E	No – Envisat-"ungauged" catchment (EU103)
4.	VS 104	565	18.345°N/103.796°E	No – Envisat-"ungauged" catchment (EU104)
5.	VS 105	107	18.151°N/103.115°E	No – Envisat-"ungauged" catchment (EU105)
6.	VS 106	651	17.980°N/102.443°E	No – Envisat-"ungauged" catchment (EU106)
7.	**VS 107**	**21**	**17.531°N/104.699°E**	**Yes – (GU25) Nakhom Phanom Station** (EU107)
8.	VS 108	21	17.137°N/104.789°E	No – Envisat-"ungauged" catchment (EU108)
9.	**VS 109**	**21**	**16.279°N/104.990°E**	**Yes – (GU27) Savannakhek Station** (EU109)
10.	VS 110	937	14.044°N/106.944°E	No – Envisat-"ungauged" catchment (EU110)
11.	VS 111	479	13.856°N/106.269°E	No – Envisat-"ungauged" catchment (EU111)
12.	**VS 112**	**866**	**13.845°N/105.986°E**	**Yes – (GU31) Stung Treng Stations** (EU112)
13.	VS 113	322	13.842°N/106.709°E	No – Envisat-"ungauged" catchment (EU113)
14.	VS 114	866	13.372°N/105.881°E	No – Envisat-"ungauged" catchment (EU114)
15.	VS 115	939	13.310°N/107.111°E	No – Envisat-"ungauged" catchment (EU115)
16.	VS 116	21	12.270°N/105.911°E	No – Envisat-"ungauged" catchment (EU116)
17.	VS 117	565	11.933°N/105.276°E	No – Envisat-"ungauged" catchment (EU117)

Note: "Geopolitically ungauged" catchments are catchments that actually have historical observations of daily streamflow and water level for only cross-validating the proposed method (not used at all for calibration) (Table 1). Envisat-"ungauged" catchments are catchments that have virtual stations of Envisat-derived water level but mostly have no observations of streamflow (except 3 catchments that are located in "geopolitically ungauged" catchments). Among 17 Envisat-"ungauged" catchments, 3 of them are located in "geopolitically ungauged" (shown in bold font), which can be validated with the actual streamflow observations.

into three maximum distinct soil layers (normally the top layer has a thickness of around 25 cm, the second of 1–2 m and the third can be deeper to account for ground water) (Bui et al., 2011). The model is forced by precipitation and temperature at either daily or hourly temporal resolution, and its calculation starts at HRUs and is then aggregated to subbasin level. HYPE calculates flow paths in the soil based on snow melt, evapotranspiration, surface runoff, infiltration, percolation, macropore flow, tile drainage and outflow to the stream from soil layers when water content is above field capacity. Different algorithms are provided to calculate snow melt, evapotranspiration, and infiltration according to the physical characteristics of the modeled catchments. The runoff from the land classes is then routed through the network of rivers and lakes to generate river flow, which could be dampened due to effect of lakes and reservoirs. HYPE can also simulate the effect of floodplains, which is crucial for large river systems and their deltas (Andersson et al., 2017), and it can also simulate the transport and concentration of nutrients in both soil, rivers and lakes (Lindström et al., 2010). In addition to natural dynamics, the model can simulate simplified water management schemes, such as regulated reservoirs (hydropower), and irrigation. There are several parameters used in HYPE that can be constrained in a stepwise manner using different types of observed data (Arheimer and Lindström, 2013). The parameters may be soil type dependent (e.g., field capacity), land cover dependent (e.g., evapotranspiration coefficient) or general across the domain (e.g., river routing parameters). Parameters, which are linked to physiography and/or climate rather than to a specific catchment, are thus assumed to be transferable to ungauged sites. More details on the HYPE model, including visual schematic diagram, can be found in the web-based documentation (http://www.smhi.net/hype/wiki/) and Lindström et al. (2010).

The HYPE model has explicit lake routing, including two types of lakes, which are local lakes and outlet lakes. Local lakes, which are located inside the subbasin, only receive a portion of local surface runoff and then flows to main river of the same subbasin. Outlet lakes, which are located near the main river, receive both local runoff (after it has passed local lakes) and the river flow from upstream subbasins. Each lake can be set with an individually defined depth. The outflow from lakes (when water level is above a defined threshold) can be either determined by a general rating curve or a specific rating curve.

The rating curve for a lake outlet is written as:

$$q = k\,(w - w_0)^p$$

Thus, water level can be also seen as a transformation of streamflow:

$$(w - w_o) = (q/k)^{1/p}$$

where q is the outflow or streamflow (m^3/s), w is water level (m), w_0 a threshold (m), k p are rate and exponent of rating curve parameters (Lindström, 2016). When w_0 is known, $(w - w_0)$ is equal to water depth.

Lake water levels, which are easier and cheaper to be measured, are mainly used with the purpose of estimating streamflow through established rating curves. Meanwhile, there are a lot of basins that have no observations of streamflow. Therefore, Lindström (2016) tested if HYPE can be calibrated using water level data instead of streamflow. The study found that water levels could be useful for calibration of hydrological models without measuring streamflow by establishing a traditional rating curve but using a constant rating curve exponent. His suggestion of using $p = 2$ while adjusting k and w_0 appropriately for all lakes resulted in a reasonable agreement with observed daily water level records based on the assumption of parabolic lake outlets, which agreed with the previous study by Maidment (1992).

Accordingly, to integrate river elevation derived from Envisat altimetry into HYPE, modeled streamflow must be converted to water level. Using outlet lake routine from HYPE, negligibly small outlet lakes, which have inflow equal to outflow (no storage capacity to affect streamflow) were added in subbasin, where there is either in-situ observations of streamflow and water level or Envisat-derived river elevation. To reduce uncertainty of estimating water depth, in addition to constant p, constant $k = 100$ and $w_0 = 0$ were used, so equation (2) becomes $w = (q/100)^{1/2}$. Since the proposed method (water level based flow correlation, explained in section 3.3) emphasized the temporal dynamics rather than the true magnitude of a variable, it was not necessary to estimate the exact water depth. Simulated w would be compared with either in-situ observations of water level at "geopolitically ungauged" catchments or Envisat-derived water elevation at Envisat-"ungauged" catchments. Envisat-"ungauged" catchments are catchments that have virtual stations of Envisat-derived water level but mostly have no observations of streamflow (except 3 catchments that are located in "geopolitically ungauged" catchments) (Table 2).

3.2. Grouping catchments using climatic indexes

Similar seasonal water balance patterns between catchments, which could be explored based on three climatic indices alone, i.e., climatic

aridity, timing of seasonal precipitation, and a temperature-based measure of snowiness, was found to provide a useful backdrop to the signatures of streamflow variability over various time scales (daily to decadal) and states (low flow to floods) (Berghuijs et al., 2014). This study applied Berghuijs et al. (2014)'s approach to robustly group catchments based on their similarity in climatic characteristics. Accordingly, two dimensionless indices that account for similar water balances among catchments were calculated, namely the aridity index and the seasonality index (snowiness is not considered in this study since there is almost no snow impact in the study area).

Proposed by Budyko (1974), the aridity index is defined as:

$$\varphi = \frac{\bar{E}}{\bar{P}}$$

where $\bar{E}$ is the average potential evaporation rate (mm/day) and $\bar{P}$ is the average precipitation rate (mm/day). This average is calculated from 2002 to 2009, the same time period used to calibrate the model. φ can range from 0 to infinity (in theory) with higher values associated with more arid climate.

Here, it is assumed that the seasonal variability of precipitation and air temperature can be modeled as simple sine curves (Milly, 1994; Potter et al., 2005; Woods, 2009) as follows:

$$P(t) = \bar{P}\left[1 + \delta_P \sin\left(\frac{2\pi(t - s_P)}{\tau_P}\right)\right]$$

$$T(t) = \bar{T} + \Delta_T[\sin(\frac{2\pi(t - s_T)}{\tau_T})]$$

where t is the time (days), s is a phase shift (days), τ is the duration of the cycle under consideration (here, 365 days), $\bar{P}$ is the average precipitation (mm/day), $\bar{T}$ is the average temperature (°C/day) over same period 2002–2009, δ_P and Δ_T are dimensionless seasonal amplitudes, and the subscripts P and T stand for precipitation (mm/day) and temperature (°C/day) respectively. P(t) is the precipitation rate (mm/day) and T(t) is temperature (°C/day) as a function of t. Using a least squares optimization, δ_P and Δ_T were obtained for all individual 1120 catchments in the HYPE study domain.

Then, the seasonality index δ_P^* was calculated using Woods (2009):

$$\delta_P^* = \delta_P . sgn(\Delta_T) . \cos(\frac{2\pi(s_P - s_T)}{\tau})$$

where δ_P^* indicates whether precipitation is in phase with the potential evaporation and temperature regimes. The parameter δ_P^* can range from -1 to $+1$, with the former representing strongly winter-dominant precipitation (P out of phase with T) and the latter showing strongly summer-dominant precipitation (P in phase with T). $\delta_P^* = 0$ indicates the uniform precipitation throughout the year.

3.3. Water level based flow correlation between gauged and "ungauged" catchments

A measured correlation matrix (Pearson's r correlation coefficient) between daily in-situ water level of "geopolitically ungauged" catchments and daily streamflow of gauged catchments was calculated to find the most highly correlated reference gauged catchments to the study "ungauged" catchments. Similarly, a measured correlation matrix between Envisat-derived water level of "ungauged" catchments outside of Vietnam and the daily streamflow of gauged catchments was also computed. To examine the assumption that the correlation between two daily streamflow series was similar to water level based flow correlation between daily water level (either in-situ observations of water level or Envisat-derived water level) and streamflow, a corresponding correlation matrix between daily streamflow of "geopolitically ungauged" catchments and daily streamflow of gauged catchments was made. From previous studies on flow correlation, correlation coefficients larger than 0.9 were recommended to consider as being highly correlated catchments (Archfield and Vogel, 2010; Betterle et al., 2019). Because there were less catchments considered in the study, $r \geq 0.7$ was selected as the threshold correlation coefficient. Because r was smaller in the study, only catchments in the same climatic group (section 3.2) were examined, to ensure they have similar climate characteristics.

3.4. Step-wise physiography and climate based regionalization at gauged basins

For data-sparse regions, step-wise calibration approach was shown to be a useful method to reduce the problem of equifinality of the final model output (Andersson et al., 2017; Donnelly et al., 2016; Arheimer and Lindström, 2013; Strömqvist et al., 2012). At each key process, lumped calibration was carried out simultaneously for sub-groups of gauged basins (representative gauged basins - RGBs) with upstream areas dominated by a specific land-use or soil type. When calibration for a specific group of RGBs is deemed satisfactory, the parameters for that responding land-use or soil type can be kept constant and the next parameters for another group can be calibrated using another set of RGBs. The step-wise separation followed the hydrological pathways through the landscape, starting with climate inputs (precipitation, evapotranspiration), then subsequently moving downstream to soils (infiltration, storage, runoff), then the rivers and lakes (routing and storage). After each step, evaluation of model performance was undertaken for all 19 gauged stations and the best performance parameter set was used in the next step of the model refinement. The period 2002–2009 was selected as the calibrated period to analyze errors and refine the model. This period was chosen because it aligned with the availability of Moderate Resolution Imaging-Spectroradiometer (MODIS) – derived potential evapotranspiration (PET) and Envisat-derived water level. The earlier part of the simulation period (1991–2001) was retained for independent validation at the same stations.

A key objective in calibrating the Greater Mekong-HYPE model was to represent the main hydrological processes of all river basins. Therefore, model evaluation and refinements primarily focused on achieving satisfactory performance across the whole basin using consistent descriptions rather than excellent performance at few locations. The streamflow signatures to be evaluated in the study were the daily and monthly specific streamflow (mm/day and mm/month), high flow (5th percentile of daily specific flow in mm/day), low flow (95th percentile of daily specific flow in mm/day) and medium flow (50th percentile of daily specific flow in mm/day). These signatures were selected because they are the most important and widely used signatures of catchment runoff response to be applied in water resources planning and environmental studies (Arheimer et al., 2020; Donnelly et al., 2016) (Table 3).

The entire domain-scale performance was quantified by first calculating key performance criteria for each of the above flow signatures at each of the 19 streamflow gauges available inside Vietnam (Fig. 1), and then computing summary statistics to describe model performance across all locations. The model's ability to simulate daily and monthly streamflow at each gauge was quantified with standard metrics,

Table 3
Flow signatures evaluated in the study (Range estimated from 2002 to 2009 period).

Flow Signatures	Description
MeanDailyQ (QDD)	Mean daily specific flow in mm
MeanMonthlyQ (QMM)	Mean monthly specific flow in mm
Q5	5th percentile of daily specific flow in mm
Q50	50th percentile of daily specific flow in mm
Q95	95th percentile of daily specific flow in mm
MeanDailyW (WDD)	Mean daily water level in m
MeanMonthlyW (WMM)	Mean monthly water level in m

Table 4
Performance metrics used in the study.

Performance metrics	Equation/References	Range	Variables
KGE (Kling-Gupta Efficiency)	$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2}$ (Gupta et al., 2009)	Negative Infinity to 1 (the closer to 1, the better simulation)	QDD, QMM
RE (Relative error)	$\beta = \frac{\mu_s}{\mu_o}$; $RE = (\beta - 1).\ 100$ (Gupta et al., 2009)	Infinity to Infinity (the closer to 0, the better simulation)	QDD, QMM, Q5, Q50, Q95
RESD (Relative Error of Standard Deviation)	$\alpha = \frac{\sigma_s}{\sigma_o}$; $RESD = (\alpha).\ 100$ (Gupta et al., 2009)	Infinity to Infinity (the closer to 0, the better simulation)	QDD, QMM
Pearson's r Correlation Coefficient	$r = \frac{cov\,(x_o, x_s)}{\sigma_s \sigma_o}$	−1 to 1 (the closer to −1 or 1, the better simulation)	QDD, QMM, WDD, WMM
NSEW	$NSEW = NSE + \frac{(\beta-1)^2}{\sigma_o^2}$ (Lindstrom, 2016)	Negative Infinity to 1 (the closer to 1, the better simulation)	WDD, WMM
NSE_{anom}	$NSE_{anom} = 1 - \frac{\sum_{t=1}^{nt}\{[x_o(t) - \bar{x_o}] - (x_s(t) - \bar{x_s})]\}^2}{\sum_{t=1}^{nt}[x_o(t) - \bar{x_o}]^2}$ (Getirana, 2010)	Negative Infinity to 1 (the closer to 1, the better simulation)	WDD, WMM

Note. x represents the streamflow or water level time series. μ: the mean value of streamflow or water level time series. σ: the standard deviation of streamflow or water level time series. The sub-indexes o and s are observed and simulated streamflow or water level time series, respectively. t is the time step (one month for this application), nt is the total number of months.

Table 5
HYPE model parameter description, initial parameter range and Posterior parameter values.

Hydrological Process	Parameter and description	Initial Parameter Range	Posterior parameter values
Potential evapotranspiration	lb: threshold soil water for activation of PET	0.9	0.9
	kc5: crop coefficient for Penman-Monteith algorithm	**[0.9 – 1.4]**	**[1.2 – 1.9]**
	alb: albedo for PET algorithms	[0.12 – 0.23]	[0.12 – 0.23]
Soil water storage and flow path (for vegetated soil and land uses)	rrcs1: recession coefficient for uppermost soil layer	0.3	0.3
	rrcs2: recession coefficient for lowest soil layer	**0.03**	**0.015**
	rrcs3: recession coefficient for slope dependent	0.0002	0.0002
	mperc1: maximum percolation capacity from soil layer 1 to soil layer 2	20	20
	mperc2: maximum percolation capacity from soil layer 2 to soil layer 3	**20**	**50**
	macrate: fraction for macro-pore/subsurface flow	**0.3**	**0.4**
	mactrinf: threshold for macro-pore/subsurface flow	**10**	**6**
	mactrsm: threshold soil water for subsurface and surface runoff	**0.7**	**0.1**
	srrate: fraction for infiltration excess surface runoff (Horton overland flow)	**0.04**	**0**
	wcwp1: wilting point as a fraction for uppermost soil layer	0.2	0.2
	wcwp2: wilting point as a fraction for second soil layer	0.2	0.2
	wcwp3: wilting point as a fraction for lowest soil layer	0.2	0.2
	wcfc1: fraction of soil available for evapotranspiration for uppermost soil layer	0.15	0.15
	wcfc2: fraction of soil available for evapotranspiration for second soil layer	0.15	0.15
	wcfc3: fraction of soil available for evapotranspiration for lowest soil layer	0.15	0.15
	wcep1: effective porosity as a fraction, for uppermost soil layer	**0.04**	**0.015**
	wcep2: effective porosity as a fraction, for second soil layer	**0.04**	**0.3**
	wcep3: effective porosity as a fraction, for lowest soil layer	**0.04**	**0.4**
	srrcs: recession coefficient for saturated surface runoff (Dunne overland flow)	**[0.05 – 0.2]**	**[0 – 0.4]**
Seasonal water balances among catchment groups	cevpcorr: correction factor for PET	0	0
	rrcscorr: correction factor for soil recession coefficient	**0**	**[-0.5 – −0.2]**

Note. Posterior parameter values different from initial parameter range are shown in bold font.

including the Kling-Gupta Efficiency (KGE) and its components r, β, α, which are directly linked with Pearson's correlation coefficient, relative error (RE) and relative error of standard deviation (RESD, variability ratio) respectively (Gupta et al., 2009) (Table 4). For constraining PET parameter values, absolute value of RE was used to find the best agreements between modeled PET and MODIS-derived PET.

Both automatic and manual calibration approach were employed to take advantage of strengths of both methods. The advantage of the former is power and speed of computation and objective parameter constraints. Nevertheless, it is unlikely to provide physically acceptable parameter estimates, which are mostly addressed by highly labor-intensive manual calibration (Boyle et al., 2000). The automatic approach was the Differential Evolution Markov Chain (DEMC) method (Ter Braak, 2006). DEMC allowed to examine parameter sensitivity, probability based uncertainty estimate and a better convergence towards the global optimum. Two-step DEMC automatic calibration was undertaken. Firstly, short runs (around 400 iterations) were done to examine parameter sensitivity. Secondly, longer runs (with at least 1000 iterations) were undertaken for only sensitive parameters to allow convergence to global optimum values. DEMC automation was then followed by manual checks to ensure the physically acceptable parameter ranges and simulated hydrograph similar to the observed patterns. Table 5 describes the model parameters to be calibrated and lists the initial parameter values for each parameter. Other parameters were kept as default as the baseline parameters from the first Greater Mekong HYPE model version (GM-HYPE v1) (the same roughly calibrated parameter sets of the first World-Wide HYPE model version 1.0 (WWH v1.0, Arheimer et al., 2020).

Step-wise physiography and climate based regionalization framework for estimating different groups of model parameter values in each step were as follows:

(1) For precipitation and temperature, different datasets of precipitation and temperature were used with the baseline parameters from WWH v1.0 (roughly calibrated model at global scale) without undertaking any additional calibration to identify the optimal climate forcing datasets for the region. Daily KGE was used to evaluate this step. This model step after selecting the optimal climate data was named as the Greater Mekong HYPE model version 1.0 (GM-HYPE v1.0).
(2) PET parameter values (lb, kc5, alb: see Table 5 for description of parameter values) were constrained using the absolute value of RE between annually simulated PET and MODIS-derived PET. PET algorithm selected in the study was Food and Agriculture Organization (FAO) Penman-Monteith, which was integrated inside the HYPE model (Allen et al., 1998; Monteith, 1965). This algorithm was selected so that it was more comparable to MODIS-based PET, which was also based on Penman-Monteith logic (Mu et al., 2011). Two-step DEMC automatic calibration was undertaken to obtain the optimal values for each RGB of each land cover type (10 main land cover types were grouped from 36 European Space Agency (ESA) Climate Change Initiative CCI v1.6 data, see details of land cover description and grouping in Arheimer et al., 2020). Thirdly, manual checks of parameter values for each group were made to ensure their acceptable physical meaning. This model step after selecting optimal parameter values was named as the Greater Mekong HYPE model version 1.1 (GM-HYPE v1.1).
(3) Parameters related to soil storage, flow paths and runoff generation (19 parameters provided in Table 5) were first optimally tuned by two-step DEMC calibration with daily KGE used as the objective function. Because all gauged and ungauged basins were mainly vegetated areas, only parameters for vegetated soils were calibrated. The remaining parameters were kept as default. Following automatic calibration, manual check was done to examine the physical meaning of parameters, and hydrograph simulation of other signatures (daily streamflow, Q95, Q5 and Q50). This model step after selecting the optimal parameter values was named as the Greater Mekong HYPE model version 1.2 (GM-HYPE v1.2).
(4) Each catchment group (section 3.2) was evaluated separately and calibrated using regional correction parameters (cevpcorr, rrcscorr: see Table 5 for description of parameter values) (Hundecha et al., 2016). Two-step DEMC with daily KGE as the objective function and manual checks were done for all flow signatures (daily streamflow, Q95, Q5 and Q50). This model step after selecting the optimal regional correction parameter values was named as the Greater Mekong HYPE model version 1.3 (GM-HYPE v1.3).

3.5. Performance of regionalized parameters at ungauged basins

The performance of physiography and climate based regionalized parameters was assessed with the following approach. At 12 "geopolitically ungauged" to be used for independent evaluation (Fig. 3), KGE, RE for daily streamflow, RE for Q95, Q5, Q50 were obtained by using the following sets of parameters:

(1) Step-wise physiography and climate based regionalization parameters transfer from gauged catchments (par GM-HYPE v1.3) (section 3.4).
(2) Global regionalization parameters from the WWH v1.3 (Arheimer et al., 2020). This parameter set was forced with the same climate data as the WWH v1.3 model (HydroGFD Precipitation and Temperature).
(3) Locally calibrated parameters in the ideal situation where observed streamflow were available for calibration (Step one to three of section 3.4 without manual calibration so that selected parameters can be objective).

The performance metrics from the three parameter sets were compared to address research objective 2 (section 1) if the proposed regionalization method could help improve prediction of streamflow signatures at ungauged basins.

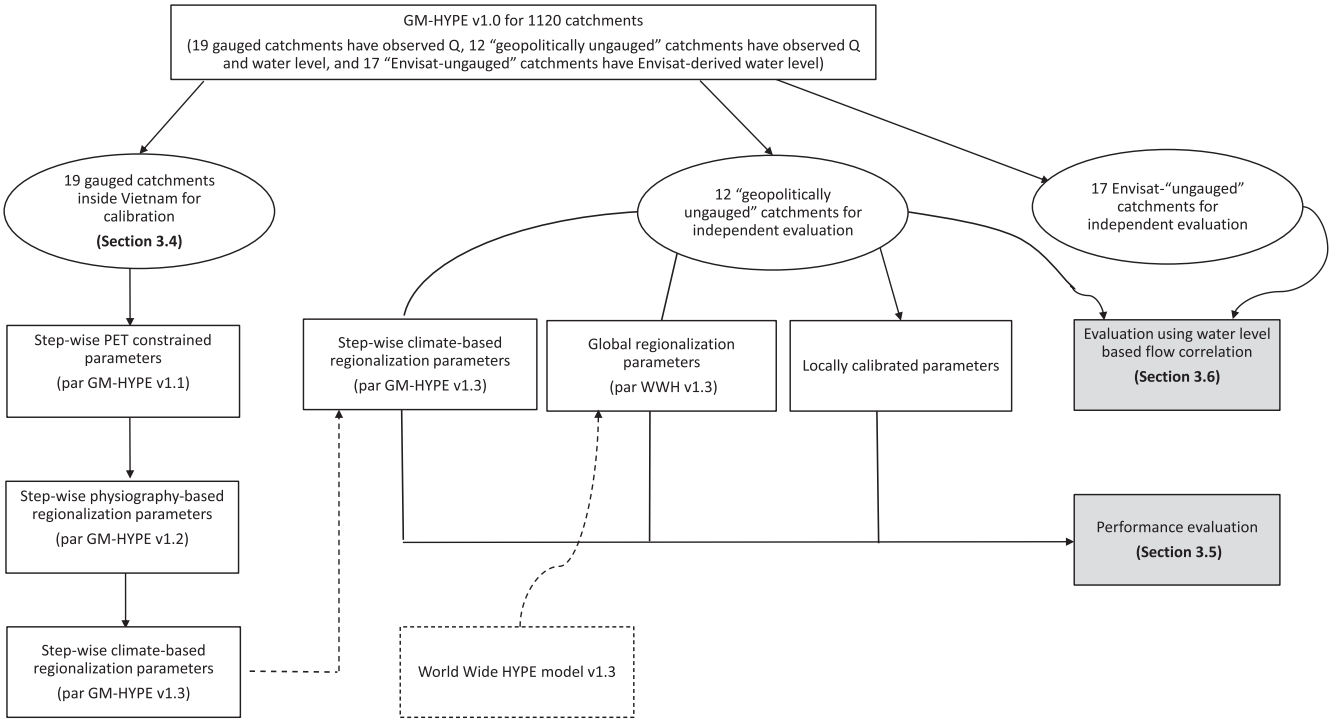

Fig. 3. Flow chart summarizing steps of proposed method in the study.

3.6. Model evaluation at ungauged basins using water-level based flow correlation

At 12 "geopolitically ungauged" and 17 Envisat-"ungauged" catchments, which were assumed to have no observations of streamflow but only water level, different performance metrics were used to examine if water levels can be modelled with a satisfactory level of performance. Water levels (e.g. modeled, in-situ, altimetry derived) are water heights that include both water depths and mean heights of river beds with respect to different vertical references. Mean heights of river beds, however, are mostly unknown unless being surveyed with large financial resources and manpower or estimated with high uncertainty. Although Pearson's correlation coefficient can capture temporal variation between modeled and recorded water levels, it could not evaluate the bias errors between them. Thus, different performance metrics have been developed to remove this bias. Instead of estimating differences between observed and modeled variables, Getirana (2010) proposed Nash-Sutcliffe efficiency for anomalies (NSEanom), which is a modified NSE metrics to to calculate the differences between their anomalies. Similarly, Lindström (2016) introduced Nash-Sutcliffe efficiency adjusted for bias (NSEW) to eliminate the bias between them (Table 4). Nevertheless, due to numerical problems, these modified formulations can yield inaccurate results because modeled water levels without unknown mean heights of river beds are within ranges of few meters whereas recorded water levels are within several hundred meters above sea levels (Lindström, 2016). Therefore, in practice, it is recommended to evaluate water depths. However, in ungauged or "geopolitically ungauged" basins, in-situ data are scare.

Accordingly, this study proposed applying hydrologic similarity theory (Betterle et al., 2017; Archfield and Vogel, 2010) by assuming that the most highly correlated reference gauged catchments (using daily streamflow) also have similar performance to that of the study "ungauged" catchments (using water level). To use this method, first, the modeled water levels of the "ungauged" catchments were evaluated against recorded water levels using Pearson's correlation coefficient, NSEanom and NSEW. For in-situ water levels, evaluation was undertaken at both daily and monthly time steps. For Envisat-derived water levels, evaluation was performed at any day step that has recorded data (one daily observation every 35 days). NSEanom and NSEW were used only to examine if numerical problems of evaluating models based on water levels existed. When only modeled water level had good correlation with recorded water level ($r \geq 0.7$), following steps were undertaken. This condition ensured that the temporal variation of modeled water levels against observations was captured. Secondly, the modeled correlation between modeled water levels of the "ungauged" catchments and modeled streamflow of the reference most highly correlated catchments was computed. If there was similar result between modeled correlation and measured correlation (modeled correlation can range from 0.5 to 0.9 compared to measured correlation), performance of reference gauged catchment was assumed to be the performance of "ungauged" catchment. To cross-validate this assumption, performance of "ungauged" catchments against the historical observations of streamflow, where available, was evaluated and compared with the assumption (Fig. 4).

4. Results

4.1. Catchment delineation and characteristics

The World Hydrological Input Set-up Tool (WHIST) developed by SMHI (Swedish Meteorological and Hydrological Institute, developer of HYPE model) was used to delineate catchment borders (Arheimer et al., 2020). Consistent with the WWH v1.3, catchment delineation was defined using the same approach according to the locations of gauging stations in the river network (including 19 "gauged" stations and 12 "geopolitically ungauged"), the outlets of large lakes/reservoirs, and

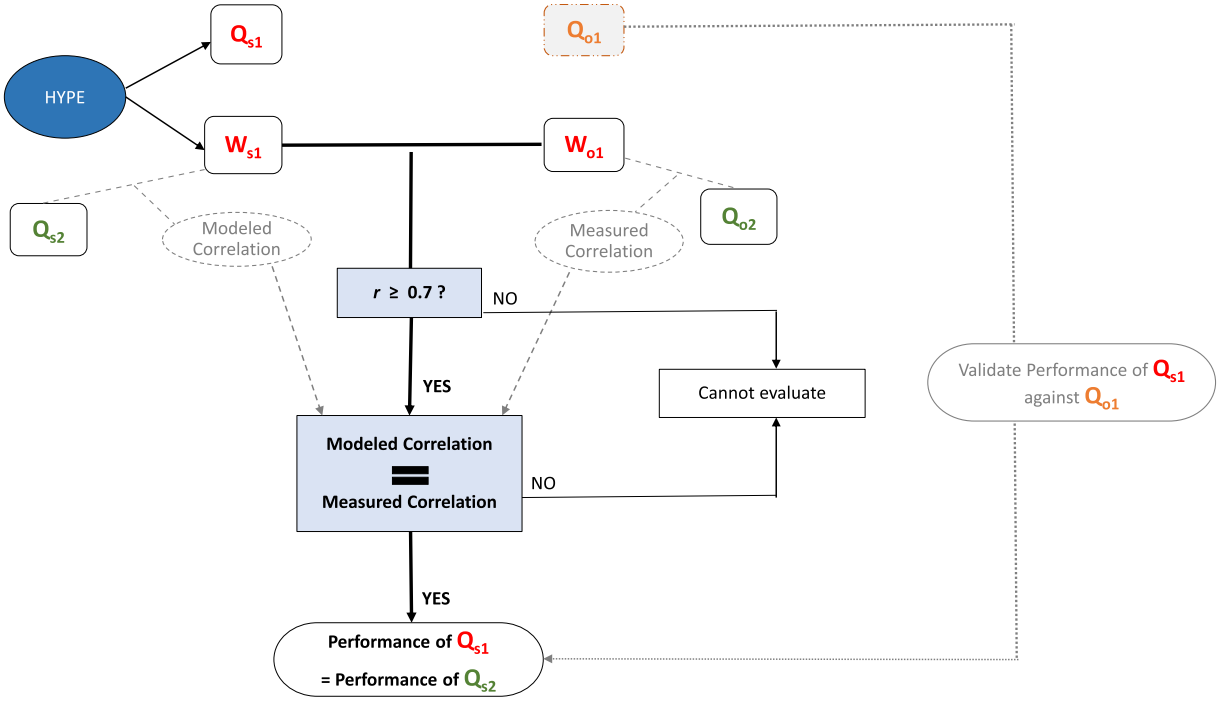

Fig. 4. Model evaluation framework at "ungauged" catchments using water level based flow correlation method. Q and W represents daily or monthly streamflow and water level time series, respectively. The subscripts s and o are simulated and observed time series respectively. The second subscripts 1 and 2 are "ungauged" (either "geopolitically ungauged" or Envisat-"ungauged") and their most highly correlated gauged catchments, respectively. Q_{o1} (if available) is not used in model setup or calibration, but only used to cross-validate the assumption that performance of "ungauged" catchments is similar to that of the reference most highly correlated gauged catchments. The expression "=" is understood as between +/- 0.2 (so modeled correlation can range from 0.5 to 0.9 compared to measured correlation).

seeking to reach an average catchment size of ~1,000 km^2 (Arheimer et al., 2020). As a result, the Greater Mekong region (~1,2 million km^2) was divided into 1,120 sub-catchments with an average size of 1,047 km^2. Sub-catchments within low-lying areas with extensive floodplains tended to have a larger size (average 3,600 km^2), among which the TonleSap basin had the largest size of 10,000 km^2. The outputs of catchment delineation were quality checked with station metadata (obtained from governmental reports). 100% of the estimated catchment areas were found to fall within +/- 5% of the areas reported by these metadata. For lakes and reservoirs, in total, 15 lakes and 18 reservoirs (only lakes and reservoirs larger than 10 km^2 recorded by GLWD and GRanD were considered in this version) were identified.

Similar to WWH v1.3, HRUs represented a combination of land cover characteristics and elevation, resulting in 169 HRUs (details of HRUs can be found in Arheimer et al., 2020). Different hydrological active soil depths were assigned for the HRUs, based on the variability in vegetation, and elevation they represented as suggested by Troch et al. (2009) and Gao et al. (2014) and currently used in WWH v1.3 (Arheimer et al., 2020). Similar to WWH v1.3, detailed description of soil properties was not included in HYPE model to reduce number of parameters. Nevertheless, five general distinct soil classes (including (i) no soil (water), (ii) urban soil, (iii) rock (no texture), (iv) vegetated soil and (v) irrigated soil) based on impermeable conditions and infiltration of land covers were identified to describe the hydrological processes in the region.

4.2. Grouping catchments using climatic indexes

Across all 1,120 catchments and during the 2002 to 2009 study period, the aridity index ranged from 0.4 to 1.7 whereas seasonality index ranges from −0.3 to 1. Accordingly, consistent with Berghuijs et al. (2014), four catchment groups were made, including group (1): Humidity ($\varphi \leq 0.75$) with Mild seasonality ($\delta_P^* \leq 0.5$); group (2): Humidity ($\varphi \leq 0.75$) with High seasonality ($\delta_P^* > 0.5$); group (3): Sub-humidity ($\varphi > 0.75$) with Mild seasonality ($\delta_P^* \leq 0.5$); and group (4): Sub-humidity ($\varphi > 0.75$) with High seasonality ($\delta_P^* > 0.5$) (Table 6). Fig. 5 shows the geographic spread and organization of four catchment groups obtained from this classification approach (See Figure Supplementary 2 for spatial distribution of all catchments based on group classification). Most catchments having historical streamflow observations (both gauged and ungauged) were classified as group 3 or group 4. Catchments of group 1 were mostly located near the coastal area with stronger humidity and more wet-season dominant precipitation. Catchments of group 3 were located mostly in the southwest of the region with less humidity and less seasonal water variability. Catchments of group 4 were located mostly in the northwest of the region with less humidity and more dry-season dominant precipitation.

Grouping catchments using climatic indexes could provide a robust reference to further regionalize parameters for each climate group. This study adapted a simple regional calibration approach, following Hundecha et al. (2016). After step-wise calibration of the model for all catchments, we evaluated the model for each catchment group to find out which signatures need to be refined and then performed regional calibration separately by using group-specific correction parameters. It should be noted that only catchment group 1, 3 and 4 could be calibrated and validated whereas catchment group 2 had no validation because there were no gauged stations for this group.

4.3. Water level based flow correlation between gauged and "ungauged" catchments

Fig. 6 validates the assumption that water level based flow correlation using daily observed in-situ water levels and streamflow had similar results to the correlation using both daily observed streamflow observations. In the case of Envisat-derived water levels, because there were less observations (one daily observation every 35 days), the correlation coefficient became slightly smaller but the difference was negligible. Accordingly, for the "geopolitically ungauged" and Envisat-"ungauged" catchments in catchment group 3, Talai (G18) was found to be the reference most highly correlated gauged catchment. For the "geopolitically ungauged" and Envisat-"ungauged" catchments in catchment group 4, LaoCai (G2), LaiChau (G4), YenBai (G6), Xala (G9) were the reference most highly correlated gauged catchments.

4.4. Step-wise physiography and climate-based regionalization at gauged basins

4.4.1. Baseline model performance (GM-HYPE v1.0)

Six sets of precipitation and temperature data were used to identify the most appropriate climate inputs for the model. Among them, HydroGFD had the coarsest resolution (0.5° grid) whereas MSWEP, TRMM and NCEP were gridded at had 0.25° resolution. There were 176 in-situ precipitation stations to examine the quality of different climate data inputs of the model. The period 2000–2006 was selected to examine their correlation as it was the period that all datasets were available. In terms of magnitudes, it was found that HydroGFD and TRMM precipitation datasets overestimated during wet months (5% and 7% respectively) and underestimated during dry months (13% and 5% respectively) compared to the in-situ precipitation, resulting in weaker correlation with in-situ precipitation (0.65 and 0.53 respectively) (Table 7). MSWEP had smaller bias (less than 1% for the entire year) and stronger correlation with in-situ precipitation. There was, unfortunately, no in-situ temperature dataset to compare with HydroGFD and NCEP. Monthly average temperature from HydroGFD was larger than monthly average NCEP (Fig. 7).

Using the initial default parameter set WWH v1.0 with different sets of climate data, no significant difference in model performance was found between them. Any temperature dataset combined with the same precipitation dataset resulted in almost similar performance. Among the precipitation datasets, MSWEP led to the highest model performance, followed by HydroGFD and TRMM. Since the MSWEP precipitation and NCEP temperature datasets had better resolution (both at

Table 6
Catchment groups using climatic indexes.

Group	Description	Total catchments	Available observations within		
			Gauged catchments	"Geopolitically ungauged" catchments	Envisat-"ungauged" catchments
1	Humidity with Mild Seasonality	54	4	0	0
2	Humidity with High Seasonality	68	0	0	1
3	Sub-humidity with Mild Seasonality	322	5	3	9
4	Sub-humidity with High Seasonality	680	10	9	7

Note: Humidity ($\varphi \leq 0.75$); Sub-humidity ($\varphi > 0.75$); Mild seasonality ($\delta_P^* \leq 0.5$); High seasonality ($\delta_P^* > 0.5$).

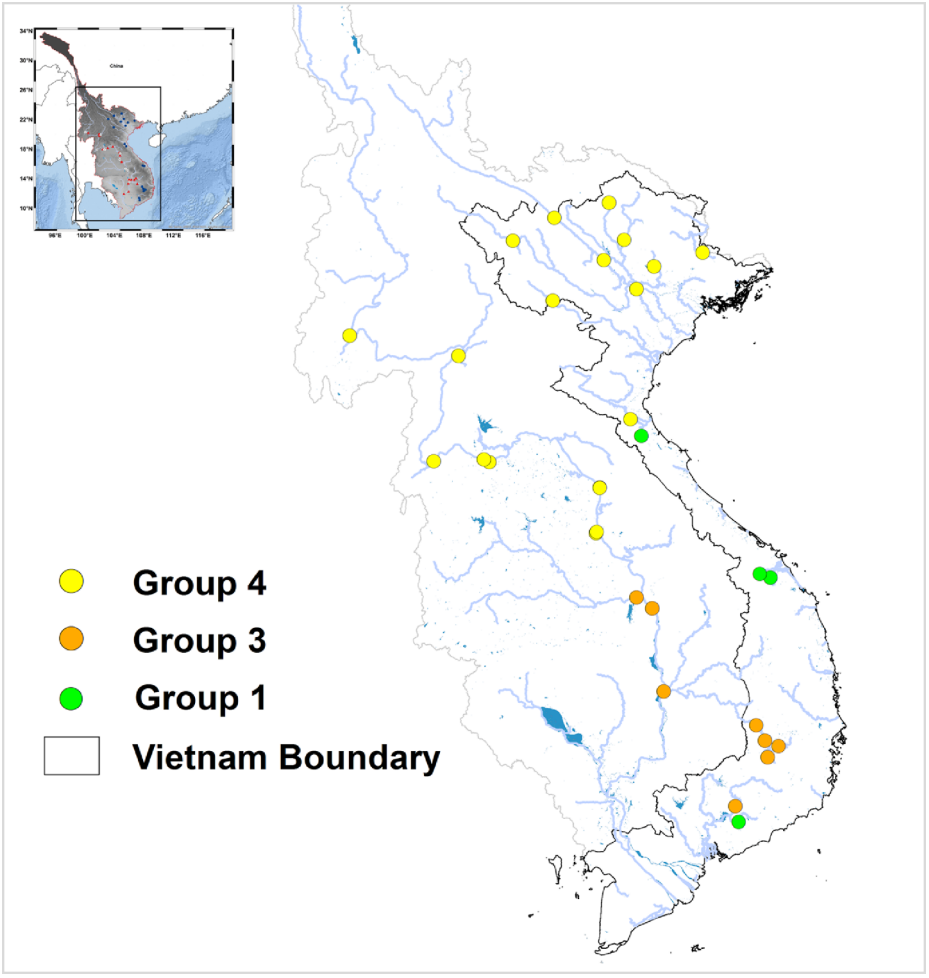

Fig. 5. The geographic distribution of all 31 evaluated catchments (gauged catchments inside Vietnam boundary and "geopolitically ungauged" catchments outside of Vietnam) into 4 climatic catchment groups.

0.25°), this set of forcing data was selected as the climate input data to be used for the baseline model (GM-HYPE v1.0) (Table 8).

4.4.2. Refining potential evapotranspiration (GM-HYPE v1.1)

Evaporation is a significantly important process in all river basins in Vietnam, accounting for around 50% of precipitation on average (Nguyen, 2005). Given the importance of evaporation in the region and large errors of streamflow variability in many locations of the baseline model, calibration was undertaken to estimate PET – the upper limit of evaporation in the model. Among three PET related parameters, land use dependent parameter (*kc5*) was found to be sensitive. The posterior *kc5* was found to reduce relative volumetric errors (RE) between modeled PET and MODIS-derived PET by 40% compared to initial *kc5* value (Table 5). With this posterior *kc5* values, model performance for all flow signatures significantly improved over all stations, particularly for KGE (from 0.3 to 0.47 for daily streamflow). In this model version, nevertheless, low flows were significantly underestimated while high flow were overestimated, requiring refinement of the soil storage and flow paths process (Fig. 8).

4.4.3. Refining soil storage and flow paths (GM-HYPE v1.2)

The GM-HYPE v1.1 model displayed a quick and peaky response of streamflow to rainfall events, resulting in the underestimation of low flows and the overestimation of high flows (Fig. 9). DEMC automation found the sensitive parameters that needed to be calibrated. They were parameters governing the soil porosities (wcep1, wcep2, wcep3), percolation (mperc2), subsurface runoff and surface runoff (macrate, mactrinf, mactrsm, rrcs2). Parameters of the GM-HYPE v1.1 model represented significantly little soil storage, and high recession coefficients. Therefore, soil related parameters were adjusted to increase soil storage capacity, more infiltration and lower recession coefficients for subsurface runoff. In addition, runoff components (Horton overland Flow, Dunne overland flow, subsurface flow) for different soil classes were unreasonable compared to Dunne theory (Dunne, 1978; Li et al., 2014). Accordingly, srrcs (Dunne overland flow related parameter) and srrate (Horton overland flow related parameter) were manually calibrated so that Horton overland flow dominated in urban and bare soil class whereas subsurface runoff and Dunne overland flow dominated in vegetated soil class. Refining these descriptions helped to maintain physical meaning of parameters, while significantly improving overall simulated flow signatures for all gauged catchments during both the calibration and validation period. For instance, for the calibration period, compared to the GM-HYPE v1.1 model, the KGE for daily streamflow improved from 0.47 to 0.7. On the other hand, volumetric errors of low flows significantly reduced to −27% from −95% and high flow from 35% to 1.7% (Fig. 8).

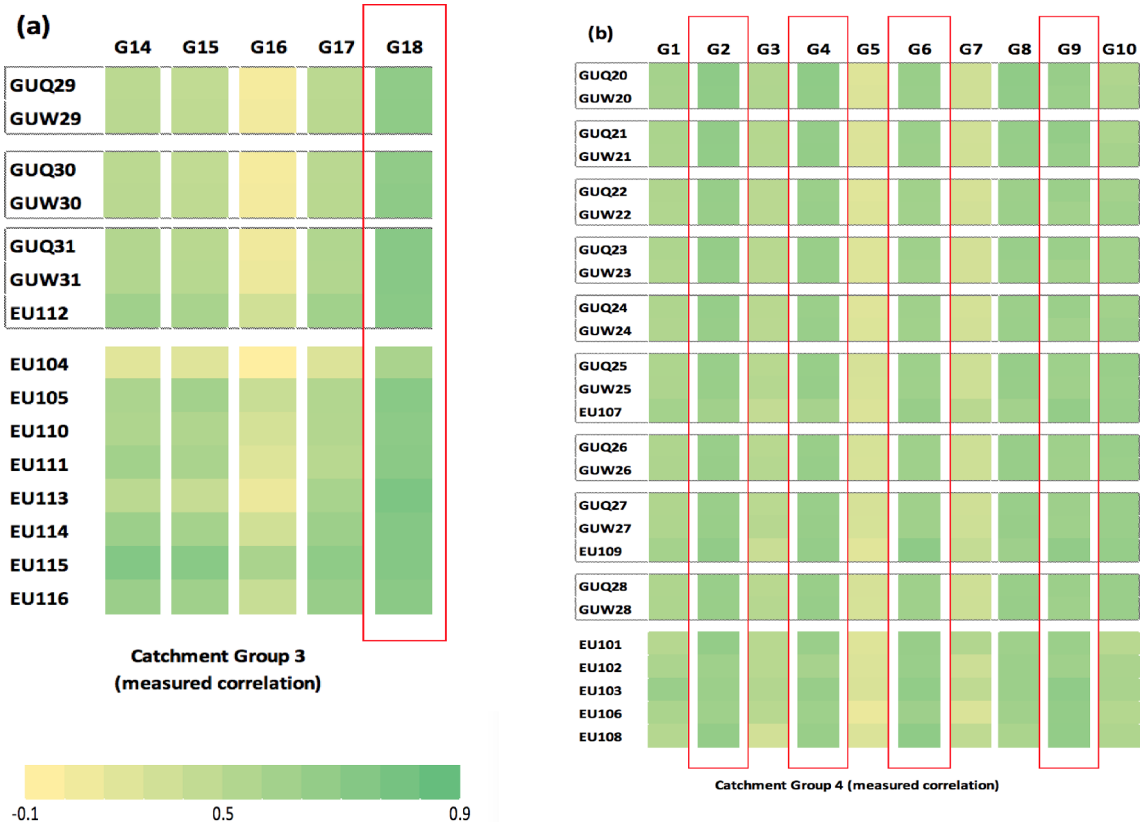

Fig. 6. Matrices show Pearson' correlation coefficient between gauged catchments (horizontal positions: G1 → G18 using daily streamflow) and "ungauged" catchments (vertical positions, including "geopolitically ungauged" catchments: GUQ20 → GUQ31 using daily streamflow; GUW20 → GUW31 using daily in-situ water levels; and Envisat-"ungauged" catchments EU101 → EU109 using daily Envisat-derived water level). Each dotted box shows the same "ungauged" catchment using different datasets (either streamflow GUQ or in-situ water level GUW or Envisat-derived water level EU) correlated with same gauged catchments. Fig. 6a is correlation matrix of "ungauged" catchment group 3 and Fig. 6b is correlation matrix of "ungauged" catchment group 4 (Table 6). Red color box highlights the most highly correlated gauged catchments with "ungauged" catchments ($r \geq 0.7$). For details of the location and name of catchments, see Table S1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 7
Correlation between different precipitation datasets.

Precipitation	HydroGFD	MSWEP	TRMM
MSWEP	0.78		
TRMM	0.61	0.85	
In-situ Precipitation	0.65	**0.75**	0.53

Note. Precipitation dataset has the highest correlation with in-situ precipitation are shown in bold font.

4.4.4. Refining seasonal water balances among catchment groups (GM-HYPE v1.3)

The model GM-HYPE v1.2 had an overall satisfactory performance for both daily and monthly streamflow time series in both the calibration and validation periods (KGE for daily streamflow was above 0.5). However, the low flow signature (Q95) was underestimated for few stations. In the model GM-HYPE v1.2, the global physiography-based parameters, which were based on soil and land cover characteristics of catchments, were used for all catchments. Evaluating the model GM-HYPE v1.2 for each catchment group (only 3 groups having gauged stations), the global physiography-based parameters were more suitable for catchment group 3, whereas low flow signatures for both catchment groups 1 and 4 were still underestimated (Fig. 10). Catchment group 1 with humidity and mild seasonality has more wet-season dominant storage variation whereas catchment group 4 with sub-humidity and high seasonality has more dry-season dominant storage variation (Berghuijs et al., 2014). Accordingly, various correction factors (evapotranspiration and recession coefficients) were used to simultaneously consider the variety of climate characteristics between catchments (Hundecha et al., 2016). In a trial and error, the correction factor for the soil recession coefficient (rrcscorr) has resulted in improvement for the low flow signatures for group 1 and group 4 for both calibration and validation periods whereas other parameter (cevpcorr) did not result in any improvement.

4.5. Performance of regionalized parameters at ungauged basins

Table 9 summarizes the performance in terms of KGE, RE for daily streamflow, RE for Q5, Q95, Q50 obtained in 12 "geopolitically ungauged" evaluation catchments using physiography-based regionalized parameters (par GM-HYPE v1.2), physiography and climate based regionalized parameter sets from gauged catchments (par GM-HYPE v1.3), global regionalization parameter sets (par WWH v1.3) and locally calibrated parameter sets. Similar to Arheimer et al. (2020), although global regionalization parameters could characterize spatial variability of flow signatures across the globe, they had difficulties in capturing low flows, particularly in tropical catchments. The difference between physiography-based regionalized parameters and physiography and climate-based regionalized parameters was not significant.

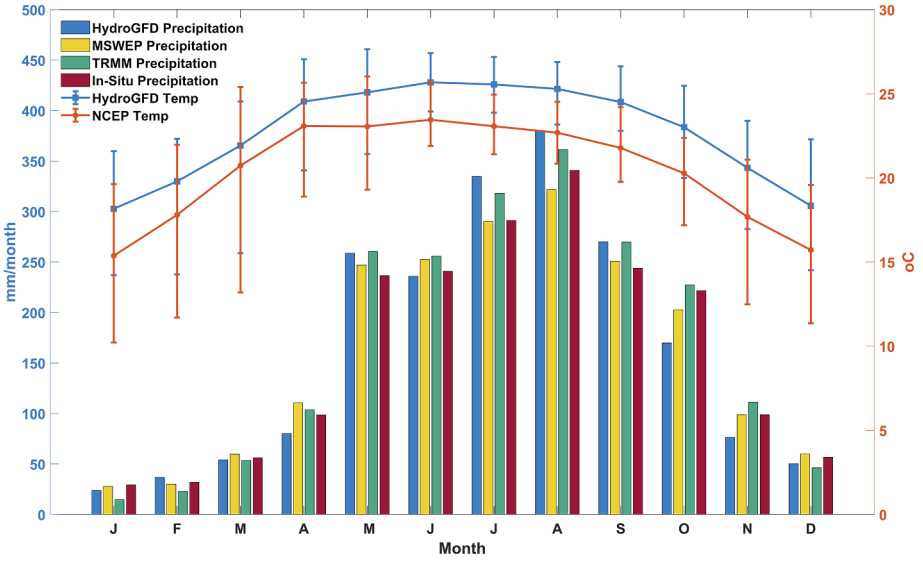

Fig. 7. Monthly time series of different climate datasets.

Table 8
Model performance using different climate datasets.

Precipitation	Temperature	KGE	Absolute RE (%)	Absolute RESD (%)	r
HydroGFD	HydroGFD	0.25	**16.8**	53.55	0.59
HydroGFD	NCEP	**0.32**	22.61	54.75	0.59
MSWEP	HydroGFD	**0.31**	21.57	**50.16**	**0.74**
MSWEP	NCEP	0.29	**19.34**	**47.59**	**0.74**
TRMM	HydroGFD	0.25	25.27	69.15	0.73
TRMM	NCEP	0.20	25.56	73.64	0.73

Note. Table presents median performance metrics for 19 gauged Vietnamese stations. For clarity, in each column, the two best values are shown in bold font (the highest values (the better) for KGE and r, the smallest values (the better) for absolute RE and absolute RESD).

Nevertheless, the later could significantly reduce volumetric errors of low flow because only one extra parameter was used in the WWH v1.3 compared to previous model version. Compared to locally calibrated parameters, physiography and climate-based regionalized parameters reached nearly 80% in terms of KGE for daily streamflow and was even a slightly better in terms of volumetric errors for low flow, medium and high flow.

Fig. 11 shows performance of daily simulated streamflow of all catchments in terms of KGE compared to their historical observations of streamflow during validation period (1991 – 2001) for two model versions, including baseline GM-HYPE v1.0 and final GM-HYPE v1.3. The catchments presented in Fig. 11 include both gauged catchments located inside black Vietnamese boundary and "geopolitically ungauged" catchments located outside of black Vietnamese boundary. Most of simulated catchments using the final model version have better captured hydrological processes of the region, resulting in a substantial improvement (mostly blue dots in Fig. 11b). In both model versions, streamflow in Lang Son (located in Bang Giang Ky Cung basin, Northeast of Vietnam, the only yellow dot in Fig. 11b) was not well simulated. The reason could be the underrepresented spatial variation of precipitation in the catchment owing to the its small size (the smallest size 1,500 km^2 in all evaluated catchments in the study). Future research could be further improved by using the average of the nearest

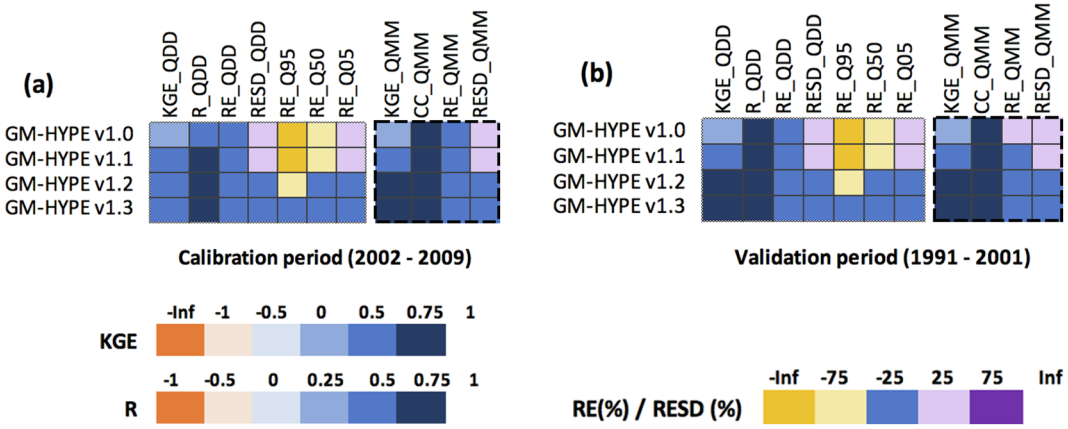

Fig. 8. Model performance (all values are median values) of all model versions at gauged stations in both calibration (Fig. 8a) and validation periods (Fig. 8b). Dotted box for daily flow signatures and dashed box for monthly flow signatures. Color interpretation of the Figure: blue is good and yellow/red/purple is poor performance. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

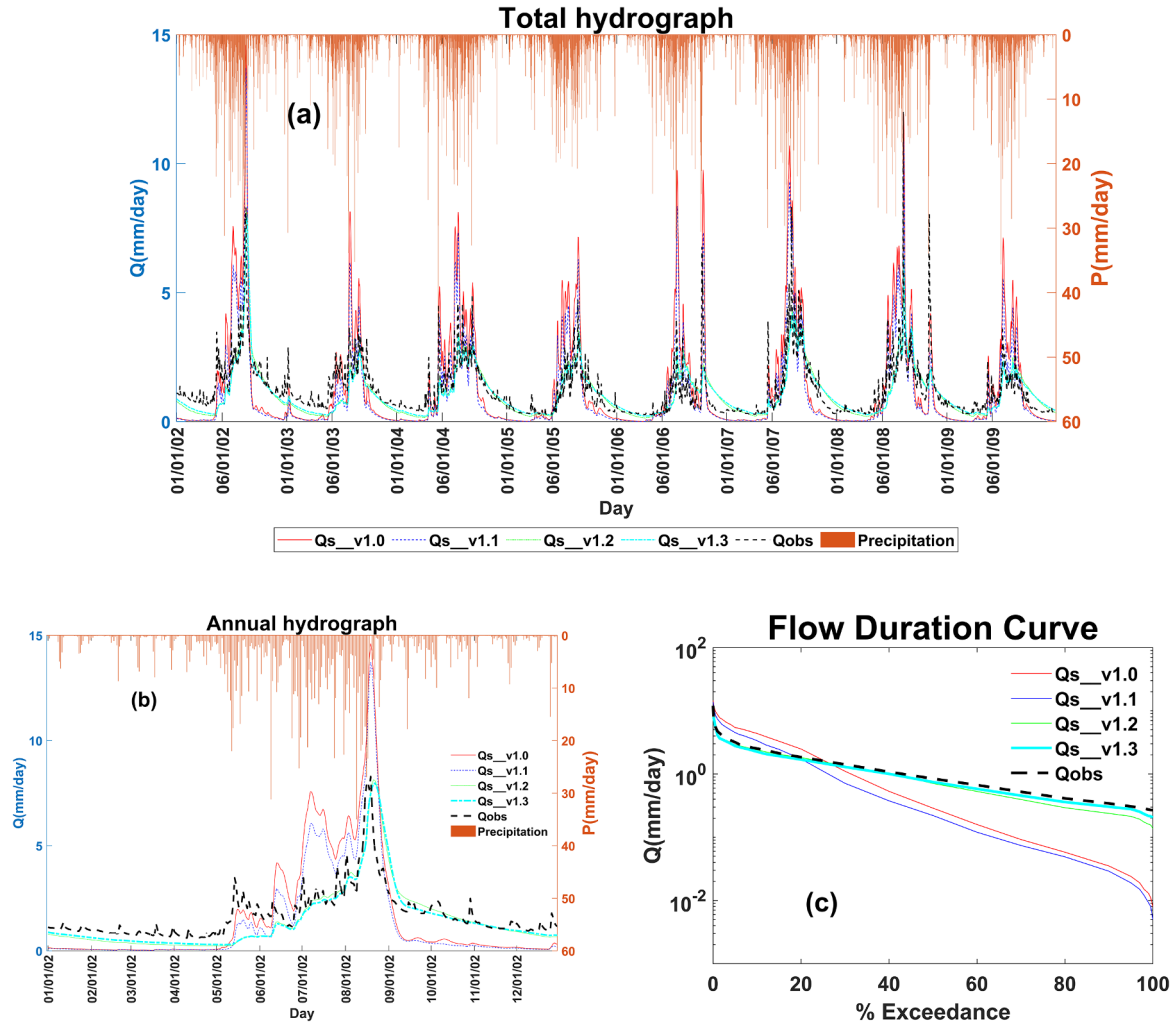

Fig. 9. Total hydrograph of simulated streamflow of all model versions against observed streamflow (Fig. 9a), annual hydrograph (Fig. 9b) and flow duration curve of simulated streamflow of all model versions against observed streamflow (Fig. 9c) at one sample gauged location at Lao Cai (located in Red River basin). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

precipitation grids or higher resolution precipitation datasets like NASA Global Precipitation Measure integrated multi-satellite retrievals with 0.1° resolution (GPM IMERGF-V6) (Le et al., 2020).

4.6. Model evaluation at ungauged basins using water level based flow correlation

To evaluate the performance of model at ungauged catchments that have observed water levels, evaluation framework using water level based flow correlation (Fig. 4) and existing performance metrics of simulated water levels was used. This section applied the framework and performance metrics for both baseline and final model versions to examine if this method and/or performance metrics can work for both scenarios (Fig. 12 and Fig. 13). Accordingly, firstly, the daily and/or monthly simulated water levels were evaluated against the recorded water levels using the existing performance metrics for simulated water levels, including Pearson's correlation coefficients, NSEanom and NSEW. From Fig. 12 and Fig. 13, conflicting performance results of simulated water levels compared to observed water levels were found. In any row of both figures, inconsistent colors between r, NSEanom and NSEW for simulated water levels (especially for Envisat-derived water level) were shown. For example, from Fig. 12d, at EU_107 station, Pearson's correlation coefficient (dark blue color – good result) showed that simulated water level had good temporal correlation with observed water levels; NSEanom (blue color – acceptable result) informed that they had acceptable magnitude bias; NSEW (orange color – bad result) advised that they had significantly high magnitude bias. Comparing simulated streamflow of this station with observed streamflow (blue color of KGE – acceptable result), the model was found to simulate daily flow at acceptable level but could not capture high flows (light purple color of Q5 – overestimated) and low flows (yellow color of Q95 – highly underestimated). This finding confirmed previous studies that model performance based on only water levels could yield inaccurate results in modelling streamflow signatures (Lindström, 2016; Jian et al., 2017). Additionally, unlike water levels that have limited performance metrics and derived hydrological signatures, there is a high variety of performance metrics to evaluate various signatures of streamflow that could help diagnose model problems and inform where to improve. For example, KGE metric can inform whether temporal pattern or variation or magnitude of daily flow is not good (Gupta et al., 2009) whereas relative volumetric errors of flow signatures from flow duration curve (high, low and medium) can inform which part of runoff (surface or subsurface runoff) is not well represented (Yokoo and Sivapalan, 2011). Accordingly, it raised a question how to have extra important model diagnostic information if only observations of water levels are available.

Water level based flow correlation was found to possibly address the above question. Using water level based flow correlation evaluation

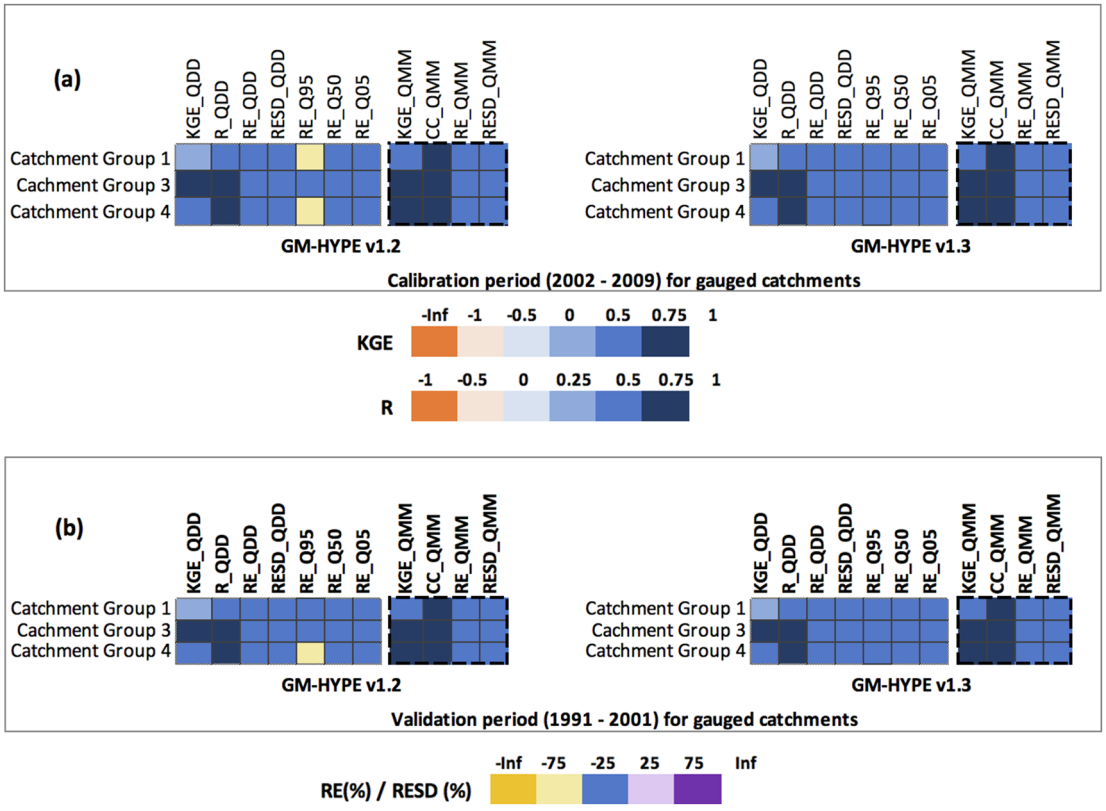

Fig. 10. Model performance by different catchment groups for 19 gauged catchments for calibration (Fig. 10a) and validation (Fig. 10b). Color interpretation of the Figure: blue is good and yellow/red/purple is poor performance. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 9
Model performance using multiple performance metrics of different flow signatures with various parameter sets for the 12 "geopolitically ungauged" evaluation catchments for the period 2002 – 2009.

Performance metrics	Flow Signatures	Par GM-HYPE v1.2 (section 4.4.3)	Par GM-HYPE v1.3 (section 4.4.4)	Par WWH v1.3 (global regionalization parameters)	Locally calibrated parameters
KGE	QDD	**0.68**	**0.68**	0.32	**0.88**
RE	QDD	**−1.3**	−1.7	−61	**1**
KGE	QMM	**0.76**	**0.76**	0.21	**0.87**
RE	QMM	**−1.3**	−1.7	−61	**0.99**
RE	Q95	−32.35	**−10.6**	−98.17	**−14.12**
RE	Q5	**−4.43**	**4.06**	−48.23	6.45
RE	Q50	**1.54**	**4.51**	−80.82	10.61

framework, firstly, temporal patterns of simulated water levels were examined against observations using correlation coefficients, which were all above 0.7 for both model versions (Fig. 12, Fig. 13). Secondly, modeled correlation between simulated water levels of "ungauged" catchments and simulated streamflow of gauged catchments were compared against measured correlation between observed water levels of "ungauged" catchments and observed streamflow of gauged catchments. The difference between them was within +/- 0.2 for both model versions, thus the performance of "ungauged" catchments was similar to the performance of the most highly correlated gauged catchments (see Figure Supplementary 4). Accordingly, in both baseline and final model versions, performance of "ungauged" catchments were similar to performance of the most highly correlated gauged catchments for all flow signatures. It was then validated with any "ungauged" catchments that have historical observed streamflow to cross-validate the hypothesis. Consistent results were found for all flow signatures between the reference most highly correlated gauged catchments and "ungauged" catchments having observed streamflow (where available for cross-validation) to accept the hypothesis. For catchments having only Envisat-derived water level without observed streamflow, its performance cannot be validated. Nevertheless, since this method worked for both in-situ water levels and 3 catchments having both Envisat-derived water level and observed streamflow, performance of the remaining 14 catchments with Envisat-derived water level could be evaluated using the reference most highly correlated gauged catchments.

Accordingly, it showed that water level based flow correlation method could be used to evaluate the model performance at ungauged catchments having only observations of water levels. Furthermore, compared to previous studies that used water levels to evaluate model performance, this approach can not only overcome numerical problems of existing performance metrics for water levels but also provide important model diagnostic information on how to improve model performance without streamflow observations.

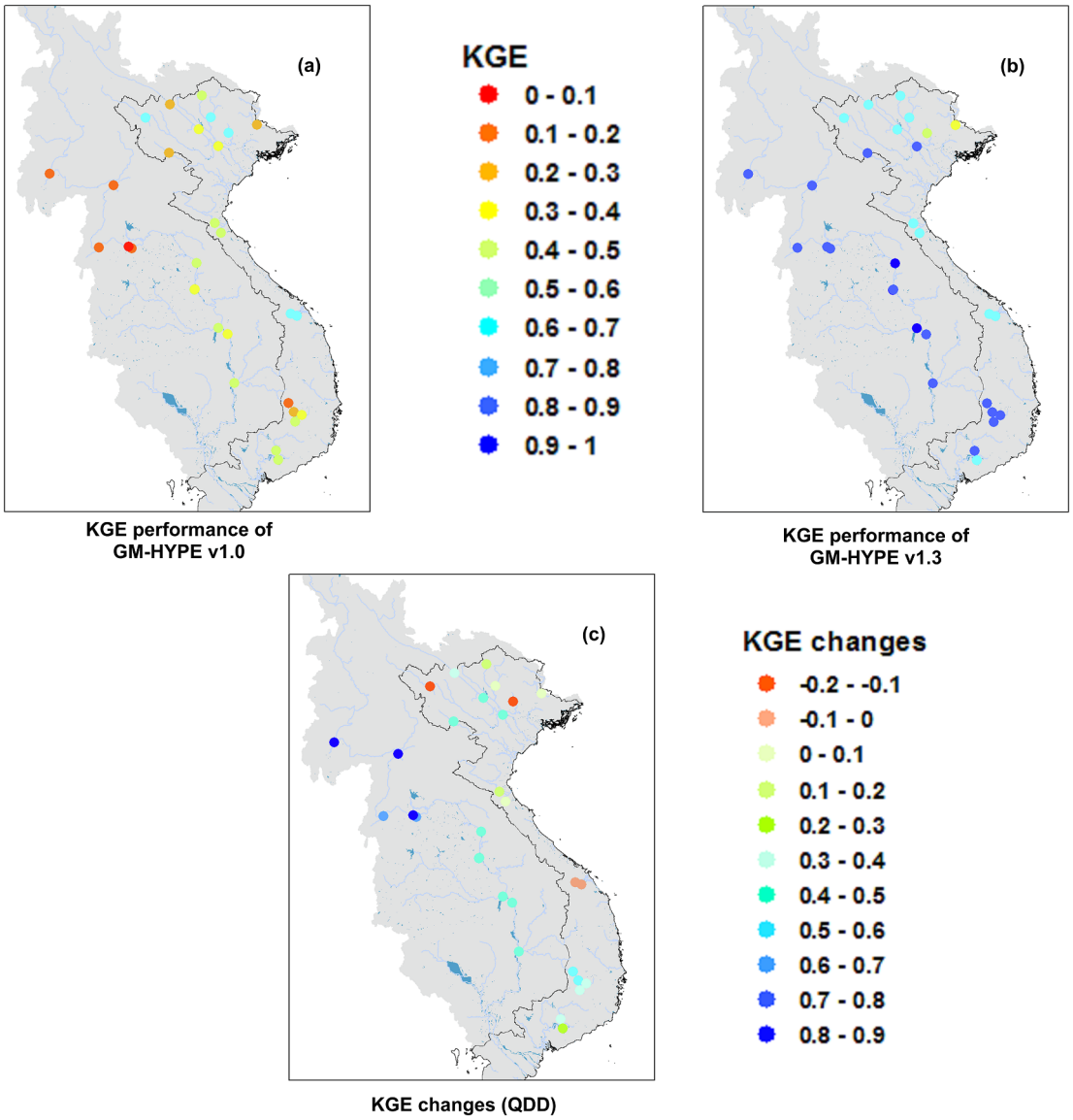

Fig. 11. Spatial overview of the model performance for GM-HYPE v1.0 (Fig. 11a), GM-HYPE v1.3 (Fig. 11b) and their changes from GM-HYEPE v1.0 to GM-HYPE v1.3 (Fig. 11c) in terms of KGE for daily streamflow time series. Model performance for both gauged (inside boundary of Vietnam) and "ungauged" catchments for validation periods (1991–2001). See Figure Supplementary 3 for model performance of simulated low flows and high flows.

5. Discussion

5.1. Step-wise physiography and climate-based regionalization at gauged basins

Catchment models are important tools to support decision makers in sustainable planning of water resources. The Greater Mekong region is the top global biodiversity hotspot but increasingly facing urgent socio-economic development and climate change impacts. Accordingly, it is imperative to have a multi-national and multi-catchment model to support river basin authorities. It could thus help predict river flows across administrative borders and allocate water resources among water users in a harmonized manner. For the first time, a multi-national and multi-catchment Greater Mekong HYPE was set up in this important region. The analysis of the final model GM-HYPE v1.3 version (KGE of daily and monthly streamflow is 0.7 and 0.8 respectively) shows that the model is useful for water authorities in managing water related issues. The model has been setup on the foundation of the World-Wide HYPE model and successfully refined to capture the hydrological processes for the region. It shows that global hydrological model, in this case the World-Wide HYPE model, could be a useful starting point as a time-saving alternative for other regions to further refine it with local expert knowledge, so that it could be useful in supporting decision makers for water management. Additionally, further refining an existing model would allow critical knowledge and experiences shared between research groups and practitioners, thus increasing full transparency in the research process, further understanding of general hydrological patterns, process and functions between catchments. It can thus ultimately advance hydrological sciences toward a unified theory of hydrology at catchment scale (Sivapalan, 2005) and better predict flow signatures at ungauged basins (Blöschl et al., 2013).

The approach of sequentially and iteratively (both automatically and manually) refining inadequately described hydrological processes, together with local knowledge can substantially improve the appropriateness of model application in a new region. Calibration is inevitable in process-based distributed model because of impossibility to measure all required model parameters at the model simulation scale (Beven, 1989; Blöschl and Sivapalan, 1995). This study combined both automatic and manual calibration to combine the strengths of both

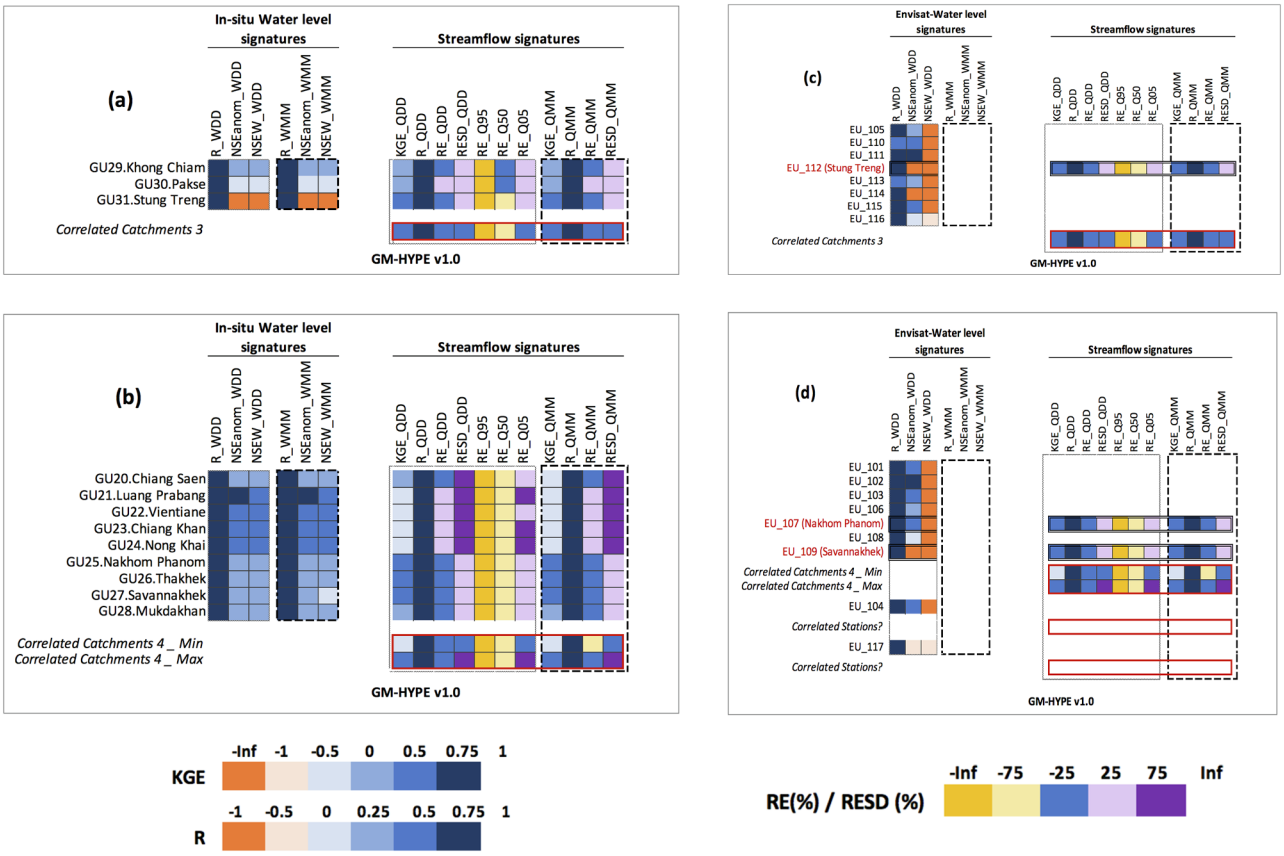

Fig. 12. Evaluation of the baseline model GM-HYPE v1.0 for "geopolitically ungauged" catchment group 3 (Fig. 12a), "geopolitically ungauged" catchment group 4 (Fig. 12b), Envisat-"ungauged" catchment group 3 (Fig. 12c) and Envisat-"ungauged" catchment group 4 (Fig. 12d). Simulated water levels were evaluated against in-situ water level (left images) and Envisat-derived water level (right images). Since modeled correlation was similar to measured correlation, simulated streamflow of "geopolitically ungauged" or Envisat-"ungauged" catchments were similar to that of the reference most highly correlated gauged catchments (red highlight box). This simulation was then validated against observed streamflow for any "ungauged" catchments that have historical observations (observed streamflow of "ungauged" catchments were only used for cross-validation, not used in calibration). For the reference most highly correlated gauged catchment 4, there were 4 catchments, thus both minimum and maximum performance metrics were presented. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

methods to achieve more physically acceptable parameters at a timely efficient manner at each step in hydrological processes. The study area has various seasonal water variability due to its substantial precipitation and evaporation variability from tropical monsoon effect. Therefore, using climatic indexes (aridity index and seasonality index) is a useful approach to group catchments so that all catchments can be simulated in the same modeling domain. Adding one simple step (step 4 in the step-wise calibration approach in Section 3.4) into the common step-wise physiography-based parameters helped reduce the underestimation of low flow of two catchment groups. In this study, simple regionalized parameter approach (correction parameters) were used. More substantial model improvement could be made if other regionalization approaches could be employed, such as linear parameter estimation based on catchment descriptors (Hundecha et al., 2016). Future studies could examine this hypothesis.

5.2. Performance of regionalized parameters at ungauged basins

HYPE with physiography and climate based regionalized parameters appears to perform as good as locally calibrated parameters and outperform global regionalization parameters in all flow signatures. This result confirms findings of the previous studies that similarity in catchment characteristics and climate characteristics can lead to similarity in rainfall-runoff responses (Beck et al., 2016; Berghuijs et al., 2014). Climatic indexes based on observations of precipitation and temperature during the same period with observations of streamflow could provide more dynamically agreeing characteristics of each catchment rather than using Köppen climate classification, which has different timeline with streamflow observation (Kottek et al., 2006). The evaluated catchments are all vegetated (either forest or agricultural lands) catchments so the difference in physiography is not significant. Nevertheless, physiography-based regionalized parameters are demonstrated to predict well flow signatures in ungauged basins across Europe (Donnelly et al., 2016). This approach could be helpful for existing model using physiography-based regionalized parameters to be further improved without altering the current parameter sets. Since this is a poorly-gauged region, obtaining more streamflow observations would be challenging. Therefore, it is important to develop more approaches to validate the simulated streamflow from model for ungauged catchments. The next section is one of those attempts. Another approach to cross-validate simulated streamflow for ungauged catchments could be using ensemble learning regression combining satellite altimetry data and a hydrologic model, which could be HYPE model in this case (Kim et al., 2019c).

5.3. Model evaluation at ungauged basins using water level based flow correlation

To evaluate model performance at ungauged basins, both existing performance metrics of water levels and proposed water level based flow correlation were adopted. Inconsistent and even conflicting performance results using different performance metrics happened for both

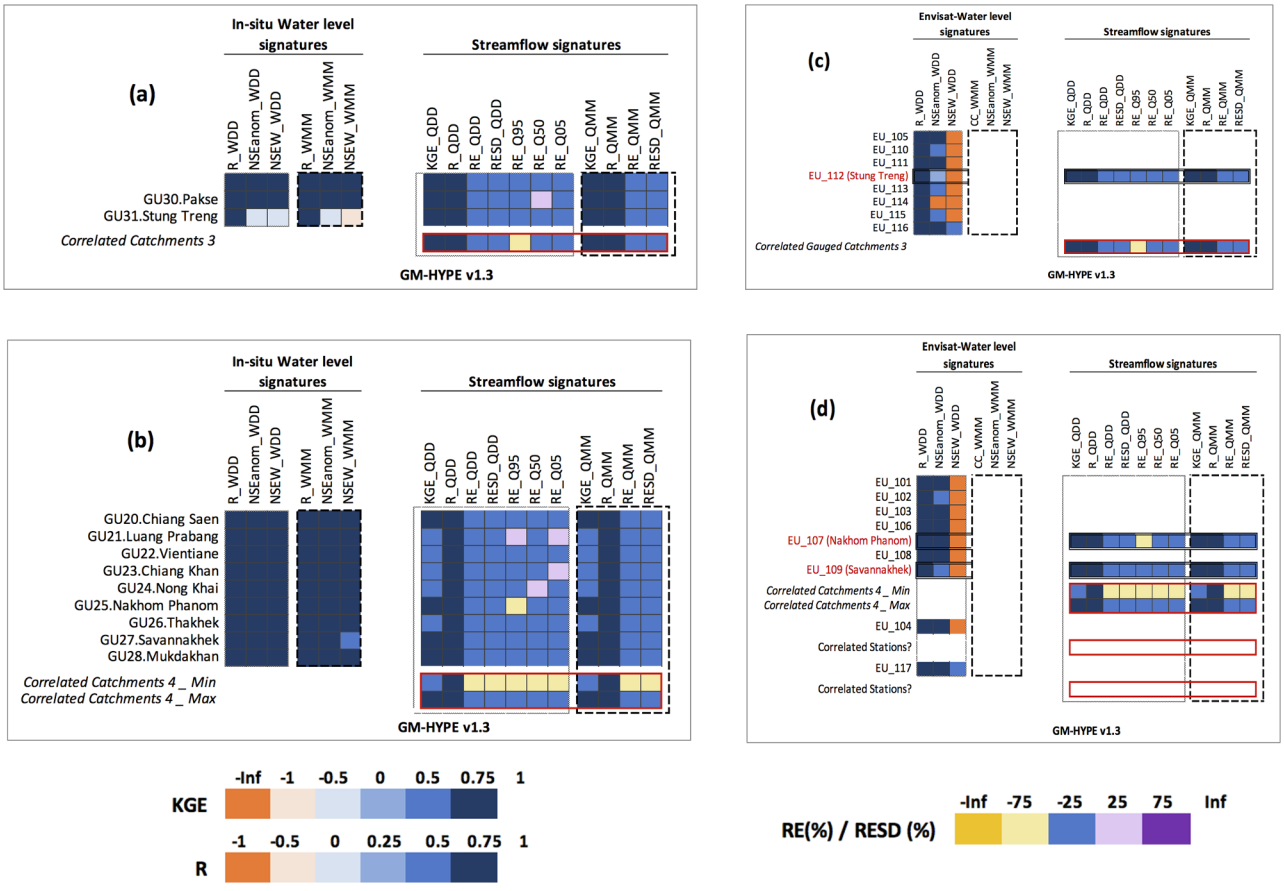

Fig. 13. Evaluation of the final model GM-HYPE v1.3 for "geopolitically ungauged" catchment group 3 (Fig. 13a), "geopolitically ungauged" catchment group 4 (Fig. 13b), Envisat-"ungauged" catchment group 3 (Fig. 13c) and Envisat-"ungauged" catchment group 4 (Fig. 13d). Simulated water levels were evaluated against in-situ water level (left images) and Envisat-derived water level (right images). Since modeled correlation was similar to measured correlation, simulated streamflow of "geopolitically ungauged" or Envisat-"ungauged" catchments were similar to that of the reference most highly correlated gauged catchments (red highlight box). This simulation was then validated against observed streamflow for any "ungauged" catchments that have historical observations (observed streamflow of "ungauged" catchments were only used for cross-validation, not used in calibration). For the reference most highly correlated gauged catchment 4, there were 4 catchments, thus both minimum and maximum performance metrics were presented. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

baseline and final models, which make diagnosing and evaluating the model at ungauged basins difficult (Figs. 12 and 13). Meanwhile, using water level based flow correlation method (both in-situ and Envisat derived water levels) can provide more details regarding model diagnostics of which signature needs to be further refined. For instance, in the baseline model (Fig. 12), using the performance of the reference most highly correlated gauged catchments, it informed that model could not capture low flow for both catchment group 3 and catchment group 4. For this study, the threshold for identifying the most correlation catchments were only 0.7 because of limited ground observations. Data access in this region is particularly arduous, therefore correlation threshold was lower than other studies (Archfield and Vogel, 2010; Betterle et al., 2017; Betterle et al., 2019). Lower correlation threshold could have reduced the matching performance between the reference most highly correlated gauged catchments and "ungauged" catchments although the difference is not significant. Future research could further examine this hypothesis. Findings show that flow correlation method with in-situ water level can be used to evaluate the performance of ungauged catchments through the most highly correlated gauged catchments. For Envisat-derived water level, since there are only three virtual stations located in catchments having ground observations, three out of 17 Envisat-"ungauged" catchments have been validated. Nevertheless, since this method was found to work with both in-situ water level and three Envisat virtual stations, it is assumed that the remaining 14 Envisat-"ungauged" catchments could have similar satisfactory simulation to the reference gauged catchments.

It is expected that not only sub-continental multi-catchment hydrological models but also multi-continental multi-catchment hydrological models would benefit from this approach if water level-based flow correlation was found between altimetry-derived water level in ungauged catchments of a poorly gauged continent and streamflow in gauged catchments of another excessively gauged continent. For instance, if the study area of GM-HYPE v1.3 could be extended, along with water level based flow correlation,the performance of the current non-validated catchment group 2 of model could be validated. Meanwhile, global-scale model could more satisfactorily capture the full range of variability of hydrological regimes that actually exist within their large domains. Thus, it can further increase the ability of hydrological models to be employed routinely and with confidence in ungauged basins. More altimetry satellite missions with denser coverage in the future could further advance this approach to improve predictions of flow regimes in ungauged basins.

6. Conclusion

The study uses a novel approach to combine regionalization and satellite observations of various hydrological variable to improve prediction of streamflow signatures at "geopolitically ungauged" basins. Using the proposed step-wise physiography and climate-based regionalization approach, the model performance at ungauged basins

reached 80% of performance of the ideal situation, where observed streamflow data were available for calibration, and significantly outperformed the global regionalization parameters using the Köppen climate classification. This approach would be helpful for both new model setup and existing process-based distributed models because it is flexible and does not change the current parameter values of existing models. Additionally, the proposed water level based flow correlation was found to help diagnose models and outperform the existing performance metrics of simulated water levels at ungauged basins. It is expected that more satellite altimetry missions with a denser coverage in the future, together with macroscale hydrological model, either at continental scale or global scale with a wide variety of observed streamflow patterns (Alemaw and Chaoka, 2003; Arheimer et al., 2020; Beck et al., 2016; Döll et al., 2003) could benefit from this approach to further evaluate model performance in ungauged basins.

The study also helps to setup the first multi-national, multi-catchment hydrological model in the Greater Mekong region, the top global biodiversity and major disaster risk hotspot in the world. This model version would be useful for water authorities to monitor and plan sustainable use of water resources across administrative boundaries under rapid changing development activities and climate impacts. Using a common hydrological model concept and setup approach compared to the global hydrological model would allow critical sharing of knowledge and experiences to advance toward a unified theory of hydrology at catchment scale and better predict flow signatures at ungauged basins. Nevertheless, knowledge gaps in aquifers, floodplain effect, and water extraction by human have not been addressed. Future model version could be further improved, such as using average of the nearest precipitation grids (for better reproducing regimes in small catchments), incorporating other hydrological data (e.g. groundwater level, total terrestrial storage change, soil moisture), and adding water management modules (e.g. regulated reservoirs, irrigation, water quality) to explore impacts of various changing scenarios from climate and human activities on the vital water, food and energy security in the region. Web data portal could be developed to allow more data accesses and knowledge sharing of water status in this important region (McDonald et al., 2019; Biswas and Hossain, 2018).

Credit authorship contribution statement

Tien L.T. Du: Conceptualization, Methodology, Investigation, Writing - original draft, Writing - review & editing. **Hyongki Lee:** Supervision, Resources. **Duong D. Bui:** Supervision, Resources. **Berit Arheimer:** Resources, Writing - review & editing. **Hong-Yi Li:** Writing - review & editing. **Jonas Olsson:** Writing - review & editing. **Stephen E. Darby:** Writing - review & editing. **Justin Sheffield:** Writing - review & editing. **Donghwan Kim:** Writing - review & editing. **Euiho Hwang:** Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This study is supported by NASA's Applied Sciences Program for SWOT Science Team (NNX16AQ33G); NASA's GEO Program (80NSSC18K0423); Vingroup Innovation Foundation (VINIF.2019.DA17); the Greater Mekong HYPE Study of the ECMWF COPERNICUS contract (C3S_422_Lot1_SMHI); NASA's SERVIR Program (80NSSC20K0152); Vietnam National Foundation for Science and Technology Development (NAFOSTED) and the United Kingdom's Natural Environment Research Council (NERC) (NE/S002847/1); USAID's PEER Project (AID-OAA-A-11-00012); and South Korea Ministry of Environment's Demand Responsive Water Supply Service Program (2019002650004). Especially, the authors would like to express our gratitude to SMHI hydrological research unit and water monitoring and forecasting team at NAWAPI, where various prior common works over multiple years on the regional modelling platform were done to make this study possible. We would like to thank Dr. Phil Graham, Dr. René Capell and Johan Strömqvist at SMHI for supporting with HYPE model setup, Kel Markert at the University of Alabama in Huntsville for helping with Google Earth Engine coding, and data providers listed in Table 1 for providing us important resources and data to undertake this work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2020.125016.

References

Abbaspour, K.C., Rouholahnejad, E., Vaghefi, Srinivasan, R., Yang, H., Kløve, B., 2015. A continental-scale hydrology and water quality model for Europe: calibration and uncertainty of a high-resolution large-scale SWAT model. J. Hydrol. 524, 733–752.

Ad Hoc Group, Vörösmarty, C., Askew, A., Grabs, W., Barry, R.G., Birkett, C., Döll, P., Goodison, B., Hall, A., Jenne, R., Kitaev, L., 2001. Global water data: a newly endangered species. Eos, Trans. Am. Geophys. Union 82 (5), 54–58.

Alemaw, B.F., Chaoka, T.R., 2003. A continental scale water balance model: a GIS-approach for Southern Africa. Phys. Chem. Earth, Parts A/B/C 28 (20–27), 957–966.

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. Fao, Rome 300 (9), D05109.

Andersson, J.C., Arheimer, B., Traoré, F., Gustafsson, D., Ali, A., 2017. Process refinements improve a hydrological model concept applied to the Niger River basin. Hydrol. Process. 31 (25), 4540–4554.

Archfield, S.A., Vogel, R.M., 2010. Map correlation method: Selection of a reference streamgage to estimate daily streamflow at ungaged catchments. Water Resour. Res. 46 (10).

Arheimer, B., Lindström, L., 2013. Implementing the EU Water Framework Directive in Sweden. Chapter 11.20 in: Bloeschl, G., Sivapalan, M., Wagener, T., Viglione, A. and Savenije, H. (Eds). Runoff Predictions in Ungauged Basins – Synthesis across Processes, Places and Scales. Cambridge University Press, Cambridge, UK. (p. 465) pp. 353-359.

Arheimer, B., Pimentel, R., Isberg, K., Crochemore, L., Andersson, J.C.M., Hasan, A., Pineda, L., 2020. Global catchment modelling using World-Wide HYPE (WWH), open data and stepwise parameter estimation. Hydrol. Earth Syst. Sci. 535–559. https://doi.org/10.5194/hess-24-535-2020.

Beck, H.E., van Dijk, A.I., De Roo, A., Miralles, D.G., McVicar, T.R., Schellekens, J., Bruijnzeel, L.A., 2016. Global-scale regionalization of hydrologic model parameters. Water Resour. Res. 52 (5), 3599–3622.

Beck, H.E., Van Dijk, A.I., Levizzani, V., Schellekens, J., Gonzalez Miralles, D., Martens, B., De Roo, A., 2017. MSWEP: 3-hourly 0.25 global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. Hydrol. Earth Syst. Sci. 21 (1), 589–615.

Berg, P., Donnelly, C., Gustafsson, D., 2018. Near-real-time adjusted reanalysis forcing data for hydrology. Hydrol. Earth Syst. Sci. 22 (2), 989–1000.

Berghuijs, W.R., Sivapalan, M., Woods, R.A., Savenije, H.H., 2014. Patterns of similarity of seasonal water balances: a window into streamflow variability over a range of time scales. Water Resour. Res. 50 (7), 5638–5661.

Bergström, S., 1991. Principles and confidence in hydrological modelling. Hydrol. Res. 22 (2), 123–136.

Betterle, A., Radny, D., Schirmer, M., Botter, G., 2017. What do they have in common? Drivers of streamflow spatial correlation and prediction of flow regimes in ungauged locations. Water Resour. Res. 53 (12), 10354–10373.

Betterle, A., Schirmer, M., Botter, G., 2019. Flow dynamics at the continental scale: Streamflow correlation and hydrological similarity. Hydrol. Process. 33 (4), 627–646.

Beven, K.J., 2011. Rainfall-runoff Modelling: The Primer. John Wiley & Sons.

Beven, K., 1989. Changing ideas in hydrology—the case of physically-based models. J. Hydrol. 105 (1–2), 157–172.

Biswas, N.K., Hossain, F., 2018. A scalable open-source web-analytic framework to improve satellite-based operational water management in developing countries. J. Hydroinf. 20 (1), 49–68.

Blöschl, G., Sivapalan, M., Savenije, H., Wagener, T., Viglione, A., 2013. Runoff Prediction in Ungauged Basins: Synthesis Across Processes, Places and Scales. Cambridge University Press.

Blöschl, G., Sivapalan, M., 1995. Scale issues in hydrological modelling: a review. Hydrol. Process. 9 (3–4), 251–290.
Boyle, D.P., Gupta, H.V., Sorooshian, S., 2000. Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. Water Resour. Res. 36 (12), 3663–3674.
Budyko, M.I., 1974. Climate and Life. Academic press, New York.
Bui, D., Kawamura, A., Tong, T.N, Amaguchi, H., Nakagawa, N., Iseri, Y., 2011. Identification of aquifer system in the whole Red River Delta, Vietnam. Geosci. J. 15 (3), 323. https://doi.org/10.1007/s12303-011-0024-x.
Chang, C.H., Lee, H., Hossain, F., Basnayake, S., Jayasinghe, S., Chishtie, F., Saah, D., Yu, H., Sothea, K., Du Bui, D., 2019. A model-aided satellite-altimetry-based flood forecasting system for the Mekong River. Environ. Modell. Software 112, 112–127.
CTOH_ENVISAT_2014_01. doi: 10.6096.
Dilley, M., Chen, R.S., Deichmann, U., Lerner-Lam, A.L., Arnold, M., 2005. Natural Disaster Hotspots: A Global Risk Analysis. The World Bank.
Döll, P., Kaspar, F., Lehner, B., 2003. A global hydrological model for deriving water availability indicators: model tuning and validation. J. Hydrol. 270 (1–2), 105–134.
Donnelly, C., Andersson, J.C., Arheimer, B., 2016. Using flow signatures and catchment similarities to evaluate the E-HYPE multi-basin model across Europe. Hydrol. Sci. J. 61 (2), 255–273.
Du, T.L.T., Bui, D.D., Quach, X.T. and Robins, L., 2016. Water governance in a changing era: Perspectives on Vietnam. Water governance dynamics in the Mekong region, pp. 241–248.
Du, T., Bui, D., Nguyen, M., Lee, H., 2018. Satellite-based, multi-indices for evaluation of agricultural droughts in a highly dynamic tropical catchment, Central Vietnam. Water 10 (5), 659.
Dunne, T., 1978. Field studies of hillsope flow processes. Hillslope Hydrol. 227–293.
Fischhendler, 2008. Ambiguity in transboundary environmental dispute resolution: The Israeli—Jordanian water agreement. J. Peace Res. 45 (1), 91–109.
Gao, H., Hrachowitz, M., Schymanski, S.J., Fenicia, F., Sriwongsitanon, N., Savenije, H.H.G., 2014. Climate controls how ecosystems size the root zone storage capacity at catchment scale. Geophys. Res. Lett. 41 (22), 7916–7923.
Garambois, P.A., Roux, H., Larnier, K., Labat, D., Dartus, D., 2015. Parameter regionalization for a process-oriented distributed model dedicated to flash floods. J. Hydrol. 525, 383–399.
Gerlak, A.K., Lautze, J., Giordano, M., 2011. Water resources data and information exchange in transboundary water treaties. Int. Environ. Agreements: Politics Law Econ. 11 (2), 179–199.
Getirana, A.C., 2010. Integrating spatial altimetry data into the automatic calibration of hydrological models. J. Hydrol. 387 (3–4), 244–255.
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. J. Hydrol. 377 (1–2), 80–91.
He, Y., Bárdossy, A., Zehe, E., 2011. A review of regionalisation for continuous streamflow simulation. Hydrol. Earth Syst. Sci. 15 (11), 3539–3553.
Hossain, F., Sikder, S., Biswas, N., Bonnema, M., Lee, H., Nguyen, L.D, Nguyen, H.N, Bui, D.D, Long, D., 2017. Predicting Water Availability of the Regulated Mekong River Basin Using Satellite Observations and a Physical Model. Asian J. Water Environ. Pollut. 14 (3), 39–48. https://doi.org/10.3233/AJW-170024.
Hrachowitz, M., Savenije, H.H.G., Blöschl, G., McDonnell, J.J., Sivapalan, M., Pomeroy, J.W., Arheimer, B., Blume, T., Clark, M.P., Ehret, U., Fenicia, F., 2013. A decade of Predictions in Ungauged Basins (PUB)—a review. Hydrol. Sci. J. 58 (6), 1198–1255.
Huffman, G.J., Adler, R.F., Bolvin, D.T., Gu, G., Nelkin, E.J., Bowman, K.P., Stocker, E., Wolff, D.B., 2006. The TRMM Multi-satellite Precipitation Analysis (TMPA): quasi-global precipitation estimates at fine scales.
Hundecha, Y., Arheimer, B., Donnelly, C., Pechlivanidis, I., 2016. A regional parameter estimation scheme for a pan-European multi-basin model. J. Hydrol.: Reg. Stud. 6, 90–111.
Jian, J., Ryu, D., Costelloe, J.F., Su, C.H., 2017. Towards hydrological model calibration using river level measurements. J. Hydrol.: Reg. Stud. 10, 95–109.
Kibler, K.M., Biswas, R.K., Juarez Lucas, A.M., 2014. Hydrologic data as a human right? Equitable access to information as a resource for disaster risk reduction in transboundary river basins. Water Policy 16 (S2), 36–58.
Kim, D., Lee, H., Beighley, E., Tshimanga, R.M., 2019c. Estimating discharges for poorly gauged river basin using ensemble learning regression with satellite altimetry data and a hydrologic model. Adv. Space Res.
Kim, U., Kaluarachchi, J.J., 2008. Application of parameter estimation and regionalization methodologies to ungauged basins of the Upper Blue Nile River Basin, Ethiopia. J. Hydrol. 362 (1–2), 39–56.
Kim, D., Lee, H., Chang, C.H., Bui, D.D., Jayasinghe, S., Basnayake, S., Chishtie, F., Hwang, E., 2019b. Daily river discharge estimation using multi-mission radar altimetry data and ensemble learning regression in the lower Mekong River Basin. Remote Sensing 11 (22), 2684.
Kim, D., Yu, H., Lee, H., Beighley, E., Durand, M., Alsdorf, D.E., Hwang, E., 2019a. Ensemble learning regression for estimating river discharges using satellite altimetry data: Central Congo River as a test-bed. Remote Sens. Environ. 221, 741–755.
Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World map of the Köppen-Geiger climate classification updated. Meteorol. Z. 15 (3), 259–263.
Le, M.H., Lakshmi, V., Bolten, J., Du Bui, D., 2020. Adequacy of Satellite-derived Precipitation Estimate for Hydrological modeling in Vietnam Basins. Journal of Hydrology, 124820.
Lee, H., Shum, C.K., Yi, Y., Ibaraki, M., Kim, J.W., Braun, A., Kuo, C.Y., Lu, Z., 2009. Louisiana wetland water level monitoring using retracked TOPEX/POSEIDON altimetry. Mar. Geod. 32 (3), 284–302.
Lehner, B., Döll, P., 2004. Development and validation of a global database of lakes, reservoirs and wetlands. J. Hydrol. 296 (1–4), 1–22.
Lehner, B., Liermann, C.R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson10, C., 2011. Global reservoir and dam (grand) database. Technical Documentation, Version, 1, pp. 1–14.
Li, H.Y., Sivapalan, M., Tian, F., Harman, C., 2014. Functional approach to exploring climatic and landscape controls of runoff generation: 1. Behavioral constraints on runoff volume. Water Resour. Res. 50 (12), 9300–9322.
Lindström, G., Rosberg, J., Arheimer, B., 2005. Parameter precision in the HBV-NP model and impacts on nitrogen scenario simulations in the Rönneå River, Southern Sweden. AMBIO J. Human Environ. 34 (7), 533–538.
Lindström, G., Pers, C., Rosberg, J., Strömqvist, J., Arheimer, B., 2010. Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales. Hydrol. Res. 41 (3–4), 295–319.
Lindström, G., 2016. Lake water levels for calibration of the S-HYPE model. Hydrol. Res. 47 (4), 672–682.
Maidment, D.R., 1992. Handbook of Hydrology. McGraw-Hill, New York.
McDonald, S., Mohammed, I.N., Bolten, J.D., Pulla, S., Meechaiya, C., Markert, A., Nelson, E.J., Srinivasan, R., Lakshmi, V., 2019. Web-based decision support system tools: The Soil and Water Assessment Tool Online visualization and analyses (SWATOnline) and NASA earth observation data downloading and reformatting tool (NASAaccess). Environ. Modell. Software 120, 104499.
Milly, P.C.D., 1994. Climate, soil water storage, and the average annual water balance. Water Resour. Res. 30 (7), 2143–2156.
Monteith, J.L., 1965. Evaporation and environment. In: Symposia of the Society for Experimental Biology (Vol. 19, pp. 205–234). Cambridge University Press (CUP) Cambridge.
Mohammed, I.N., Bolten, J.D., Srinivasan, R., Lakshmi, V., 2018. Satellite observations and modeling to understand the Lower Mekong River Basin streamflow variability. J. Hydrol. 564, 559–573.
Mu, Q., Zhao, M., Running, S.W., 2011. Improvements to a MODIS global terrestrial evapotranspiration algorithm. Remote Sens. Environ. 115 (8), 1781–1800.
Okeowo, M.A., Lee, H., Hossain, F., Getirana, A., 2017. Automated generation of lakes and reservoirs water elevation changes from satellite radar altimetry. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10 (8), 3465–3481.
Parajka, J., Viglione, A., Rogger, M., Salinas, J.L., Sivapalan, M., Blöschl, G., 2013. Comparative assessment of predictions in ungauged basins–Part 1: Runoff-hydrograph studies. Hydrol. Earth Syst. Sci. 17 (5), 1783–1795.
Parajka, J., Blöschl, G., Merz, R., 2007. Regional calibration of catchment models: potential for ungauged catchments. Water Resour. Res. 43 (6).
Pechlivanidis, I., Arheimer, B., 2015. Large-scale hydrological modelling by using modified PUB recommendations: the India-HYPE case. Hydrol. Earth Syst. Sci. 19 (11), 4559–4579.
Potter, N.J., Zhang, L., Milly, P.C.D., McMahon, T.A., Jakeman, A.J., 2005. Effects of rainfall seasonality and soil moisture capacity on mean annual water balance for Australian catchments. Water Resour. Res. 41 (6).
Razavi, T., Coulibaly, P., 2012. Streamflow prediction in ungauged basins: review of regionalization methods. J. Hydrol. Eng. 18 (8), 958–975.
Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.T., Chuang, H.Y., Iredell, M., Ek, M., 2011. NCEP climate forecast system version 2 (CFSv2) 6-hourly products. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory.
Sellami, H., La Jeunesse, I., Benabdallah, S., Baghdadi, N., Vanclooster, M., 2014. Uncertainty analysis in model parameters regionalization: a case study involving the SWAT model in Mediterranean catchments (Southern France).
Schwatke, C., Dettmering, D., Bosch, W., Seitz, F., 2015. DAHITI–an innovative approach for estimating water level time series over inland waters using multi-mission satellite altimetry. Hydrol. Earth Syst. Sci. 19 (10), 4345–4364.
Shiklomanov, A.I., Lammers, R.B., Vörösmarty, C.J., 2002. Widespread decline in hydrological monitoring threatens pan-Arctic research. Eos, Trans. Am. Geophys. Union 83 (2), 13–17.
Sivapalan, M., 2005. Pattern, process and function: elements of a unified theory of hydrology at the catchment scale. Encyclopedia of Hydrological Sciences.
Strömqvist, J., Arheimer, B., Dahné, J., Donnelly, C., Lindström, G., 2012. Water and nutrient predictions in ungauged basins: set-up and evaluation of a model at the national scale. Hydrol. Sci. J. 57 (2), 229–247.
Sun, W., Ishidaira, H., Bastola, S., 2012. Calibration of hydrological models in ungauged basins based on satellite radar altimetry observations of river water level. Hydrol. Process. 26 (23), 3524–3537.
Tang, X., Zhang, J., Gao, C., Ruben, G.B., Wang, G., 2019. Assessing the uncertainties of four precipitation products for swat modeling in Mekong River Basin. Remote Sensing 11 (3), 304.
Tordoff, A.W., Bezuijen, M.R., Duckworth, J.W., Fellowes, J.R., Koenig, K., Pollard, E.H. B., Royo, A.G., ZUN, P., 2012. Ecosystem profile: Indo-Burma biodiversity hotspot 2011 update. Critical Ecosystem Partnership Fund, Arlington, Virginia.
Ter Braak, C.J., 2006. A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces. Statistics and Computing 16 (3), 239–249.
Troch, P.A., Martinez, G.F., Pauwels, V.R., Durcik, M., Sivapalan, M., Harman, C., Brooks, P.D., Gupta, H., Huxman, T., 2009. Climate and vegetation water use efficiency at catchment scales. Hydrol. Processes Int. J. 23 (16), 2409–2414.
Turton, A., Ashton, P., Cloete, E., 2003. An introduction to the hydropolitical drivers in

the Okavango River basin. Transboundary Rivers, Sovereignty and Development: Hydropolitical Drivers in the Okavango River Basin. Pretoria & Geneva: AWIRU & Green Cross International.
UNEP-DHI, U.N.E.P., 2016. Transboundary river basins: status and trends (summary for policy makers). Nairobi: United Nations Environment Programme (UNEP).
Woods, R.A., 2009. Analytical model of seasonal climate impacts on snow hydrology: Continuous snowpacks. Adv. Water Resour. 32 (10), 1465–1481.
World Bank, 2019. Vietnam: Toward a Safe, Clean and Resilient Water System. The World Bank, Washington DC.
Yamazaki, D., O'Loughlin, F., Trigg, M.A., Miller, Z.F., Pavelsky, T.M., Bates, P.D., 2014. Development of the global width database for large rivers. Water Resour. Res. 50 (4), 3467–3480.
Yokoo, Y., Sivapalan, M., 2011. Towards reconstruction of the flow duration curve: development of a conceptual framework with a physical basis. Hydrol. Earth Syst. Sci. 15 (9).