Using Remote Sensing Data-based Hydrological Model Calibrations for Predicting Runoff in Ungauged or Poorly Gauged Catchments

Yongqiang Zhang[1], Qi Huang[2], Guanghua Qin[3], Qiuhong Tang[4], Changming Liu[4], Jun Xia[5], Francis Hock Soon Chiew[6], and David A. Post[6]

[1]Institute of Geographic Sciences and Natural Resources Research
[2]College of Water Resource & Hydropower, Sichuan University
[3]Sichuan University
[4]Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences
[5]State key laboratory of Water Resources and Hydropower Engineering Science, Wuhan University
[6]Commonwealth Scientific and Industrial Research Organisation (CSIRO)

November 26, 2022

Abstract

Because remote sensing (RS) data are spatially and temporally explicit and available across the globe, they have the potential to be used for predicting runoff in ungauged or poorly gauged catchments, a challenging area of research in hydrology over the last several decades. There is potential to use remotely sensed data for calibrating hydrological models in regions with limited streamflow gauges. This study conducts a comprehensive investigation on how to incorporate gridded remotely sensed-evapotranspiration (AET) and water storage data for constraining hydrological model calibration in order to predict daily and monthly runoff in 30 catchments of Yalong River basin, China. To this end, seven RS data calibration schemes are explored, compared to traditional calibration against observed runoff and traditional regionalization using spatial proximity. Our results show that using bias-corrected remotely sensed AET (bias-corrected PML-AET data) for constraining model calibration performs much better than using the non bias-corrected remotely sensed AET data (non bias-corrected AET obtained from PML model estimate). Using the bias-corrected PML-AET data in a gridded way is much better than that in a lumped way, and outperforms the traditional regionalization approach especially at upstream and large catchments. Combining the bias-corrected PML-AET and GRACE water storage data performs similarly to using the bias-corrected PML-AET data only. This study demonstrates that and there is great potential to use RS-AET based data for calibrating hydrological models in order to predict runoff in data sparse regions with complex terrain conditions.

Using Remote Sensing Data-based Hydrological Model Calibrations for Predicting Runoff in Ungauged or Poorly Gauged Catchments

Qi Huang[1,2,6], Guanghua Qin[1,3,6], Yongqiang Zhang[2]*, Qiuhong Tang[2], Changming Liu[2], Jun Xia[4], Francis H.S. Chiew[5], and David Post[5]

[1]College of Water Resource & Hydropower, Sichuan University, Chengdu 610065, China;

[2]Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

[3]State Key Laboratory of Hydraulics and Mountain River Engineering, Sichuan University, Chengdu 610065, China

[4]State Key Laboratory of Water Resources & Hydropower Engineering Sciences，Wuhan University, 430072, Wuhan, China

[5]CSIRO Land and Water, Black Mountain, Canberra, ACT 2601, Australia

[6]These authors contributed equally to this work.

*Corresponding author: Yongqiang Zhang

(yongqiang.zhang2014@gmail.com;zhangyq@igsnrr.ac.cn)

Key points

- Using remote sensing data to calibrate hydrological model shows great potential especially at upstream and large areas in Yalong River basin
- Compared to raw PML-AET, bias-correction of PML-AET improves runoff prediction noticeably and adding GRACE shows limited benefit
- Gridded modelling calibrated at each grid performs noticeably better than lumped modelling calibrated at each catchment

Abstract

Because remote sensing (RS) data are spatially and temporally explicit and available across the globe, they have the potential to be used for predicting runoff in ungauged or poorly gauged catchments, a challenging area of research in hydrology over the last several decades. There is potential to use remotely sensed data for calibrating hydrological models in regions with limited streamflow gauges. This study conducts a comprehensive investigation on how to incorporate gridded remotely sensed-evapotranspiration (AET) and water storage data for constraining hydrological model calibration in order to predict daily and monthly runoff in 30 catchments of Yalong River basin, China. To this end, seven RS data calibration schemes are explored, compared to traditional calibration against observed runoff and traditional regionalization using spatial proximity. Our results show that using bias-corrected remotely sensed AET (bias-corrected PML-AET data) for constraining model calibration performs much better than using the non bias-corrected remotely sensed AET data (non bias-corrected AET obtained from PML model estimate). Using the bias-corrected PML-AET data in a gridded way is much better than that in a lumped way, and outperforms the traditional regionalization approach especially at upstream and large catchments. Combining the bias-corrected PML-AET and GRACE water storage data performs similarly to using the bias-corrected PML-AET data only. This study demonstrates that and there is great potential to use RS-AET based data for

calibrating hydrological models in order to predict runoff in data sparse regions with complex terrain conditions.

Key words: Remote sensing, evapotranspiration, PML, runoff prediction, grid, bias correction

1. Introduction

Runoff Prediction in Ungauged Basins (PUB) is materially important for accounting and managing water resources, and flood disaster risk management (Montanari et al., 2013). A widely used approach for PUB is regionalization that transfers calibrated model parameters from a gauged catchment (or a donor) to an ungauged catchment (Post and Jakeman, 1999; Hundecha and Bardossy, 2004; Merz and Bloschl, 2004; Yadav et al., 2007; Oudin et al., 2008; Zhang and Chiew, 2009; Hrachowitz et al., 2013; Pechlivanidis and Arheimer, 2015; Li and Zhang, 2017). However, the performance of the regionalization approach becomes gradually poorer with an increasing regionalization distance (Li and Zhang, 2017), suggesting that regionalization is highly uncertain in regions with very limited or sparsely distributed streamflow gauges. The data scarcity, and hence the regionalization challenge, is prominent especially in alpine and complex-terrain regions.

Remote sensing observation provides continuous data in both spatial and temporal scales, which makes it possible to estimate regional surface meteorological data in a quick, accurate and widely applicable way. Therefore, remote sensing data has been widely applied and combined with hydrological models in ungauged catchments. As inputs to hydrological model, the remote sensing data should be relatively accurate, otherwise it needs to be bias corrected (Stisen and Sandholt, 2010; Habib et al., 2014; Zhang and Tang, 2015). What's more, it has also been shown that constraining multiple variables such as soil moisture and water storage data from remote sensing can improve the performance of hydrological models (Sutanudjaja et al., 2014; Wanders et al., 2014; Li et al., 2016; Kundu et al., 2017; Yassin et al., 2017; Pomeon et al., 2018). Nevertheless, practically all studies calibrate the models against observed streamflow data, which is limited in poorly gauged regions. Zhang et al. (2020) proposed a remotely sensed actual evapotranspiration (RS-AET) calibration approach based on PML evapotranspiration products (PML-AET）, and showed that this approach is potentially useful in the relatively wet regions of Australia. Nevertheless, there are several limitations in the study that can be improved upon. First, Zhang et al. (2020) did not consider the potential for improving the quality of the remote sensing actual evapotranspiration data that was used for hydrological model calibration. Second, the study used a lumped catchment-average rainfall-runoff modelling approach and does not take advantage of the spatial continuity of remote sensing data. Third, the research does not consider the potential to combine remote sensing actual evapotranspiration with remote sensing water storage data.

To further advance the study of Zhang et al. (2020), this paper proposes a more comprehensive framework that uses runoff-free or very limited runoff data for hydrological model calibrations. Specifically, this work aims to improve calibration schemes by adding more remote sensing information (non bias-corrected PML-AET, bias-corrected PML-AET, GRACE water storage) into model calibrations, and calibrating the hydrological model both in lumped and gridded ways. Nine modelling schemes (seven are based on RS-data calibrations; one is based on runoff-data calibration; one is based on spatial proximity regionalization) are tested on the Yalong River Basin, the upper reach of which is located on the southeastern Tibetan Plateau and the northwest of Yunnan-Guizhou Plateau, with complex terrain conditions. The major objectives of this study are to:

i. Evaluate the merit of using limited runoff data for bias correcting remote sensing evapotranspiration data

ii. Investigate the performance of calibrations with different remote sensing data (non bias-corrected PML-AET, bias-corrected PML-AET, GRACE water storage);

iii. Evaluate the performance of calibrations at different spatial scales (gridded and lumped); and

iv. Investigate the spatial characteristics of optimum model calibration schemes.

2. Study area and data

2.1. Study area

The study area is located in the Yalong River basin. The Yalong River, the largest tributary on the left bank of the Jinsha River, originates from the southern foot of the Bayankala Mountains in Yushu County, Qinghai Province. The river flows from the northwest to the southeast, and the length of the mainstream is around 1570 km. The whole basin area is around 1.36 x 10^5 km^2, shaped like a north-south stripe (96°52′E-102°48′E,26°32′N-33°58′N) and located on the southeastern Tibetan Plateau and the northwest of Yunnan-Guizhou Plateau. The river basin spans more than seven degrees of latitude from north to south, and the geographic characteristics in the basin are complex. The altitude varies greatly from 5,400 m to 980 m from the north to the south, with a reduction of 4,420 m, and the terrain mainly includes hilly plateaus, alpine canyons, and wide valley basins from north to south, respectively. All of these make the climate and geography of the basin greatly different in both horizontal and vertical directions.

This study uses data from 30 catchments within the Yalong River basin. Fig.1 shows the geographical map of the study area and the information of the 30 catchments. It also summarizes the flow path through the 30 catchments.

Figure 1 is about here

2.2. Data

The Climate Meteorological Forcing Dataset (simplified as CMFD) is used to drive the hydrological model. The CMFD is a reanalysis product of near-surface meteorological and environmental elements in China. The gridded precipitation data used here is the CMFD-Precipitation. The CMFD-P has been shown to be a high data quality dataset (Yang et al., 2017; Ren et al., 2018; Wu et al., 2019; He et al., 2020), and is also further evaluated here against daily gauged precipitation in the study area (see sections 3.1.1 and 4.1).

The gridded actual evapotranspiration data used in this paper is obtained from PML_V2 global evapotranspiration (simplified as PML-AET) product (Zhang et al., 2019). It is referred as 'non bias-corrected PML-AET' thereafter. Since this is a global product, it is necessary for bias correction to be applied in order to improve its usability for hydrological modelling applications (see Sections 3.1.2 and 4.2).

The water storage data applied in this paper is Gravity Recovery and Climate Experiment's total water storage anomaly data (simplified as GRACE) and has been corrected by officially provided scale factors (Swenson and Wahr, 2006; Landerer and Swenson, 2012). All the gridded datasets were resampled to 0.05° to match the PML resolution. The daily runoff data is obtained from hydrological observed records, and used here as the reference data for model validation. Table 1 gives more information on those data.

Table 1. Detailed information for research data used in this study

Short name	Detailed Name	Spatial Resolution	Temporal Resolution	Temporal Coverage	Data source	Key references
CMFD	Climate Meteorological Forcing Dataset	0.1°	3-hour	1979-2018	http://westdc.westgis.ac.cn/data/ 7a35329c-c53f-4267-aa07-e0037d913a21	(He and Yang, 2011; Fan et al., 2017; He et al., 2020)
PML_V2	PML_V2 global evapotranspiration and gross primary production	0.05°	8-day	2002.07-2019.08	http://www.tpdc.ac.cn/zh-hans/data/ 48c16a8d-d307-4973-abab-972e9449627c/	(Zhang et al., 2019)
GRACE_ RL05	Gravity Recovery and Climate Experiment	1°	1-month	2002.04-2017.02	https://grace.jpl.nasa.gov/data/ get-data/monthly-mass-grids-land/	(Swenson and Wahr, 2006; Landerer and Swenson, 2012)
Meteorological gauge Data	Daily dataset of China's surface climate data	-	1-day	1951-2019	http://data.cma.cn/data/cdcdetail/dataCode/ SURF_CLI_CHN_MUL_DAY_V3.0.html	-
Hydrological station Data	Daily mean runoff of hydrological stations in Yalong River	-	1-day	2004-2012 (Varying across stations)	The information and data of stations are provided by Yalong River hydropower-development company	

It should be noted that there are two downstream catchments (Xiaodeshi catchment and Tongzilin catchment) impacted by the Ertan reservoir regulation during 2004-2012. To obtain the 'natural flow' for these catchments, streamflow series is restored through reservoir dispatching data based on water balance method. As shown in Fig.1, the Xiaodeshi hydrology station and the Tongzilin hydrology station are in the downstream of the Ertan hydropower station and are both in the mainstream of Yalong River. Ignoring other human activities along the river, the ‘natural flow’ series of Xiaodeshi and Tongzilin catchment is obtained by adding the value of the Ertan Hydropower Station inflow minus the outflow.

3. Methodology

3.1. Data Processing

3.1.1. Evaluation of CMFD-P

As shown in Figure 1, the available rain gauges are few and sparsely distributed. Thus, a set of accurate gridded precipitation dataset is needed. The accuracy of CMFD-P data is evaluated against ten surface meteorological precipitation gauges in the Yalong River Basin. The main idea is to verify the accuracy through detect ability and accuracy indicators. The evaluation indicators are listed in Table 2, together with their descriptions.

Table 2. Evaluation indicators for precipitation

Type of Indicators	Evaluation Indicators	Short name	Formula	Ideal Value
Detect Ability Indicators	Probability Of Detection	*POD*	$POD = \frac{n_{11}}{n_{11} + n_{01}}$	1
	Frequency Of Hit	*FOH*	$FOH = \frac{n_{11}}{n_{11} + n_{10}}$	1

	Heidke's Skill Score	*HSS*	$HSS = \frac{2(n_{11}n_{00} - n_{10}n_{01})}{(n_{11}+n_{01})(n_{01}+n_{00})+(n_{11}+n_{10})(n_{10}+n_{00})}$	1
Accuracy Indicators	Correlation coefficient	*CC*	$CC = \frac{\sum_{i=1}^{n}(P_i-\bar{P})(G_i-\bar{G})}{\sqrt{\sum_{i=1}^{n}(P_i-\bar{P})^2\sum_{i=1}^{n}(G_i-\bar{G})^2}}$	1
	Nash-Sutcliffe Efficiency	*NSE*	$NSE = 1 - \frac{\sum_{i=1}^{n}(P_i-G_i)^2}{\sum_{i=1}^{n}(G_i-\bar{G})^2}$	1
	Similarity indicator	*SI*	$SI = 1 - \frac{\sum_{i=1}^{n}(P_i-G_i)^2}{\sum_{i=1}^{n}(\lvert G_i-\bar{G}\rvert+\lvert P_i-\bar{G}\rvert)^2}$	1
	Mean error	*ME/(mm)*	$ME = \sum_{i=1}^{n}(G_i-P_i)/n$	0
	Mean absolute error	*MAE/(mm)*	$MAE = \sum_{i=1}^{n}\lvert G_i-P_i\rvert/n$	0
	Bias	*BIAS*	$BIAS = \sum_{i=1}^{n}(G_i-P_i)/\sum_{i=1}^{n}G_i$	0
	Absolute bias	*ABIAS*	$ABIAS = \sum_{i=1}^{n}\lvert G_i-P_i\rvert/\sum_{i=1}^{n}G_i$	0

* n_{11} represents the frequency of precipitation detected by both CMFD and the rainfall gauges; n_{10} represents the frequency of precipitation detected by CMFD but not the rainfall gauges; n_{01} represents the frequency of precipitation detected by the gauges but not CMFD; n_{00} represents the frequency of precipitation detected by neither CMFD nor the rainfall gauges. P represents precipitation in CMFD, G represents gauged precipitation, and n is the amount of samples.

3.1.2. Bias correction of PML-AET

The PML-V2 is a global evapotranspiration and gross primary product. It therefore needs to be bias corrected for application at small spatial scales and local regions. In addition, the actual evapotranspiration can be directly validated and bias corrected using the water balance method. The mean annual PML-AET is bias corrected here to match the mean annual precipitation minus mean annual runoff estimated by the Fu model (the Fu model is an adaption of the Budyko framework) (Fu, 1981; Zhang et al., 2004; Zhang et al., 2008). The bias correction is carried out as follows:

i. In order to use minimum possible observed runoff data for the bias correction and to maximize the usability of the PML-AET model calibrations, mean annual observed runoff data (Q_{obs}) of a downstream basin, Daluo (Gauging station 21, see Fig.1) for the period of 1999 to 2012 is used for the method inputs (this was the length of streamflow data available at Daluo). What’s more, mean annual gridded precipitation data (CMFD-P) and mean annual gridded potential evaporation E_p are also used for the method inputs. E_p is estimated using the Allen et al. (2006) equation following Penman-Monteith method (*Eq.1*). The input data comes from the CMFD dataset (i.e. temperature, humidity, wind speed), digital elevation model (latitude and longitude), and daily dataset of China's surface sunshine duration data that was spatially interpolated by kriging method (Delhomme, 1978)). E_p is calculated using the following equation:

$$E_p = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T_{mean} + 273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)}, \quad (1)$$

where E_p is the potential evaporation (mm/d); Δ is the slope of the saturation vapor pressure versus temperature curve (kPa/°C); R_n is the net radiation flux density at the surface (MJ/(m*d)); G is the sensible heat flux from the surface to the soil (MJ/(m²*d)); γ is the psychometrics constant (kPa/°C); T_{mean} is the daily temperature (°C); u_2 is the wind speed at 2-m height (m/s); e_s is the saturation vapor pressure at air temperature (kPa); e_a is the actual vapor pressure of the air (kPa).

ii. A single value of the parameter α in the Fu model is 1.56, estimated based on the basin-average mean annual precipitation and potential evaporation at Daluo catchment from 1999 to 2012. This α value of 1.56 is used to calculate Q_{fu} at each (0.05° x 0.05°) of 5170 grid cells within the study area for the period of 2004 to 2012. Q_{fu} is expressed as:

$$Q_{fu} = P[1+(AI)^{\partial}]^{1/\partial} - E_P, \qquad (2)$$

where Q_{fu} represents mean annual runoff (mm/year). P is mean annual rainfall (mm/year). E_p is mean annual potential evapotranspiration (mm/year). AI is the aridity index, calculated as E_p divided by P.

iii. The 'real' value of mean annual AET (2004-2012) at each grid is calculated as P minus Q_{fu};

iv. A scaling factor SC at each grid cell is calculated as the 'real' mean annual AET divided by mean annual Non bias-corrected PML-AET; and

v. Finally, the bias-corrected PML-AET for each grid is obtained by multiplying the non bias-corrected PML-AET by the scaling factor at each grid.

In summary, this study uses mean annual streamflow data from one downstream gauge of Daluo and from an independent period of 1999-2012 to parameterize the Fu model, and then uses Fu mean annual runoff estimate to bias correct PML-AET at each grid cell in 2004-2012.

3.2. Xinanjiang Model

The Xinanjiang model is a lumped conceptual model, was developed by Zhao (1980). The model has been extensively used for runoff simulation and prediction across humid and semi-humid regions globally (Zhao, 1992; Jayawardena and Zhou, 2000; Cheng et al., 2006; Ju et al., 2009; Li et al., 2009; Yao et al., 2009). The model is driven by daily precipitation and potential evapotranspiration for the period of 2004-2012. The model outputs include daily runoff and daily actual evapotranspiration. Daily water storage is one of state variables in this model and is used in the calibration functions in this study. The model structure is shown in Figure 2.

Figure 2 is about here

3.3. Model calibration schemes

The RS-ET runoff free calibration method is developed by Zhang et al. (2020) and its objective function is calibrated only against PML-AET. It has been shown that water storage data can also enhance hydrological model calibration (Yassin et al., 2017). This study will therefore explore the model calibration against both remotely sensed (and bias corrected) PML-AET and water storage data. This study also assesses the model calibrations at three spatial scales: grid, region and catchment. This means that the model is calibrated at each grid, region, and catchment, respectively. As a result, for grid calibration, each grid cell has one calibration parameter set; for region calibration, each grid has one calibration parameter set; for catchment calibration, each catchment has one calibration parameter set. The model becomes more lumped with the scale increase from grid to catchment.

Altogether, nine calibration schemes are considered (Table 3), seven of which are based on PML-AET calibration methods and two of which are based on streamflow calibration. A global optimizer, the genetic algorithm built in MATLAB (Holland, 1992; Konak et al., 2006), is used to optimize model parameters. Scheme 1 is calibrated against observed daily runoff by using lumped catchment inputs, which represents the best simulation capability in each catchment. Scheme 2 is regionalization based on spatial proximity (i.e. selecting a donor catchment with minimum Euclidian distance between centroids of the 'ungauged' catchment and the donor) (Merz and Bloschl, 2004; Oudin et al., 2008; Li and Zhang, 2017). This scheme is the traditional regionalization approach, regarded as the baseline for evaluating the performance of schemes 3-9. Scheme 3 uses the non bias-corrected PML-AET output for model calibration. Schemes 4-6 apply the bias-corrected PML-AET for model calibration, but the difference among them is scheme 4 for calibration at each PML-AET grid cell, scheme 5 for calibration at each region (The region is defined as the contribution area between two gauges. Therefore, the lowest-level tributary comprises 1 region, but higher lever catchments comprise multiple regions. For instance, Ganzi (1), Xinlong (2) and Gongke (3) have one, two, three regions, respectively), and scheme 6 for calibration at each catchment. Schemes 7-9 are similar to schemes 4-6, respectively, but with the model calibrated against both the bias-corrected PML-AET data and the GRACE water storage data for model calibration.

Table 3 summarizes the nine schemes for model calibration and provides the objective function used for calibration in each scheme.

Table 3. Summary of nine model calibration schemes 1-9.

The numbers 1-9 represent scheme numbers, respectively. Eq. (3)- Eq. (6) represent objective functions.

Calibration Method	At grids	At regions	At catchment	Model input data (and calibration data)	Objective functions
Calibration with observed runoff Regionalization			1	CMFD-P, Ep, (Q at 30 stations)	Eq.(3)
			2	CMFD-P, Ep, a set of parameters (at a neighbor station)	Eq. (3)
Non bias-corrected PML-AET runoff-free calibration approach	3			CMFD-P, Ep, (non bias-corrected PML-AET)	Eq. (4)
Bias corrected PML-AET calibration approach	4	5	6	CMFD-P, Ep, (bias-corrected PML-AET)	Eq. (5)
Bias corrected PML-AET combined with GRACE storage data runoff-free calibration approach	7	8	9	CMFD-P, Ep, (bias-corrected PML-AET, GRACE)	Eq. (6)

The widely used Nash-Sutcliffe Efficiency (NSE) (Nash and Sutcliffe, 1970) is used as the objective functions defined in Equations 3-6.

$$F_1 = 1 - NSE_Q \, , \; NSE_Q = 1 - \frac{\sum_{i=1}^{N} (Q_{obs} - Q_{sim})^2}{\sum_{i=1}^{N} (Q_{obs} - \overline{Q_{obs}})^2} \, , \tag{3}$$

$$F_2 = 1 - NSE_{ET1} \, , \; NSE_{ET1} = 1 - \frac{\sum_{i=1}^{N} (AET_{PML} - AET_{SIM})^2}{\sum_{i=1}^{N} (AET_{PML} - \overline{AET_{PML}})^2} \, , \tag{4}$$

$$F_2 = 1 - NSE_{ET2} \, , \; NSE_{ET2} = 1 - \frac{\sum_{i=1}^{N} (AET_{B-PML} - AET_{SIM})^2}{\sum_{i=1}^{N} (AET_{B-PML} - \overline{AET_{B-PML}})^2} \, , \tag{5}$$

$$F_3 = (1 - NSE_{ET2}) + (1 - NSE_{\Delta W}) \, , \quad NSE_{\Delta W} = 1 - \frac{\sum_{i=1}^{N} (\Delta W_{GRACE} - \Delta W_{SIM})^2}{\sum_{i=1}^{N} (\Delta W_{GRACE} - \overline{\Delta W_{GRACE}})^2} \, , \tag{6}$$

where Q_{obs} represents the observed daily runoff, Q_{sim} represents the simulated daily runoff. AET_{SIM}, AET_{PML} and $AET_{B\text{-}PML}$ represent modeled actual evapotranspiration, the raw PML-AET output and bias-corrected PML-AET with a temporal step of eight days, respectively. ΔW_{GRACE} and ΔW_{SIM} with a temporal step of one month represent the water storage change estimated by GRACE and calculated by Xinanjiang model, respectively. It is noted that Q_{sim} generated from grid and regional calibrations, is aggregated to catchment scale to compare to Q_{obs}. The smaller the value of objective function is, the better the simulation quality.

3.4. Evaluating the nine modelling schemes

The NSE_Q and Qualified Rate (QR) (Standardization Administration of the People's Republic of China, 2008) is used to evaluate the performance of the nine schemes at different temporal scales. The model performance is mainly decided by NSE_Q and QR is considered as an assistant indicator. The QR is defined as:

$$QR = \frac{m}{n} \quad (7)$$

where m represents the numbers of samples whose ABIAS are less than 0.35, n is the total number of samples (total number of daily, or monthly streamflow data). The value of NSE_Q varies from negative infinity to 1, the closer to 1 indicating better model performance. The value of QR varies from 0 to 1, the closer to 1 indicating better model performance (QR=1 means that the bias from all samples is less than 0.35). The temporal step is one day and one month for daily runoff and monthly runoff, respectively. The model evaluation period is the period of available observed runoff series in each catchment.

4. Results

4.1. Evaluation of CMFD-P

Figure 3 evaluates CMFD-P, the 0.05°×0.05° reanalysis precipitation product of China, against ten precipitation gauges at different time scales. Table 4 shows the performance of the CMFD-P using statistical indices summarized from the ten gauges. At daily scale, the values of POD, FOH, and HSS are 0.93, 0.67, and 0.62, respectively. This indicates that the detect ability of CMFD-P is relatively good. The CMFD-P is able to detect most of the daily precipitation events between 2004 and 2012. The accuracy of CMFD-P is also relatively good at the daily scale with high SI (0.75) and low BIAS (-0.002). On the other hand, the low frequency of hits leads to low NSE (0.26) and high ABIAS (0.83). At the monthly scale, the consistency between the CMFD-P and the station's precipitation has increased significantly compared to the daily scale. The value of accuracy indicators has increased significantly. CC, NSE and SI has increased to 0.99, 0.99 and 1.00, respectively, and ABIAS has decreased dramatically to 0.06. Compared to monthly performance, the performance of CMFD-P at annual scale is slightly degraded, indicated by smaller NSE and SI, but ABIAS at annual scale is 0.02, noticeably smaller than that at monthly scale. In summary, CMFD-P has overall quite good quality in this region. Furthermore, it performs best at monthly scale, followed by annual and daily scales. The poor performance of daily precipitation might bring more uncertainties to the hydrological models, but the high *SI* and low *BIAS* might show positive influence in the modelling.

Figure 3 is about here

Table 4. Evaluation of CMFD-P (precipitation in CMFD). The definition of each index is given in Table 2

	POD	FOH	HSS	ME/mm	BIAS	MAE/mm	ABIAS	CC	NSE	SI
Daily	0.93	0.67	0.62	-0.001	-0.002	1.61	0.83	0.59	0.26	0.75
Monthly	-	-	-	-0.153	-0.002	3.22	0.06	0.99	0.99	1.00
Annual	-	-	-	-0.366	-0.002	13.40	0.02	0.99	0.98	0.99

4.2. Bias-corrected PML-AET

The non bias-corrected PML-AET and bias-corrected PML-AET are evaluated using their performance for estimating annual streamflow. The annual streamflow of them is estimated by annual precipitation minus annual non bias-corrected PML-AET (Q_1) and annual precipitation minus annual bias corrected PML-AET (Q_2), respectively. If the consistency between Q_2 and Q_{obs} is much better than it between Q_1 and Q_{obs}, the bias correction improves the accuracy of the AET estimation.

Figure 4 summarizes the performance of Q_1 and Q_2 at annual scale for all 30 streamflow gauges. It is clear that Q_2 is noticeably better than Q_1. In most basins, scatters of Q_{obs} against Q_2 distribute evenly on both sides of the 1: 1 line, which means the consistency between Q_2 and Q_{obs} is good, while Q_1 is severely biased. This result demonstrates that the bias-corrected PML-AET achieves much better water balance (in terms of producing streamflow), compared to the non bias-corrected PML-AET. It should be noted that the Q_{obs} at Daluo station was used to bias correct PML-AET. Therefore, the performance of bias correction of mainstream catchments in the upper reach of Daluo catchment (Daluo, Luning, Jinping, Maidilong, Jiju and Yajiang) is better than that in other catchments. The better bias correction might improve the performance of hydrological model in these catchments.

Figure 4 is about here

4.3. Runoff prediction

Figure 5 summarizes the performance of nine modelling schemes in predicting daily runoff (5a,5c) and monthly runoff (5b,5d) across 30 catchments in the Yalong River basin (to present patterns clearly, several negative values are not shown here, but are shown latter in Fig.6). In each scheme, the simulated monthly runoff is accumulated by daily runoff and generally performs better than daily runoff with a higher mean value. The annual runoff performance has not been analyzed because of short records of yearly observed runoff. NSE describes modeling ability, QR describes the modeling quality. The range of NSE and QR describes modeling stability and the model performs better with a lower range of NSE and QR.

Figure 5 is about here

4.3.1. Non bias-corrected PML-AET calibration versus bias-corrected PML-AET calibration

The simulated streamflow obtained from scheme 3 (calibration using the non bias-corrected PML-AET data) and from scheme 4 (calibration using the bias-corrected PML-AET data) are evaluated against observed streamflow at daily and monthly scales. Scheme 3 gives a result of mean NSE of -0.08 and mean QR of 0.15 for daily runoff simulation, while mean NSE of -0.01 and mean QR of 0.15 for monthly runoff simulation, indicating very poor accuracy in the modelling. For daily runoff in scheme 4, the mean NSE is 0.39 and the mean QR is 0.40, while for monthly runoff in scheme 4, the mean NSE is 0.65 and the mean QR is 0.45. Compared to scheme 3, the performance of scheme 4 is greatly improved noticeably with increment of 0.47 in mean NSE and 0.25 in mean QR for daily runoff, and an increment of 0.66 in mean NSE and 0.30 in mean QR for monthly runoff. Therefore, using the bias-corrected PML-AET data for constraining model calibration performs much better than using the non bias-corrected PML-AET data, and the improvement in monthly runoff simulation is larger than that in daily runoff simulation. Therefore, in the following sections of 4.3.2-4.3.4, we only show the relative merits related to bias-corrected PML-AET (i.e. schemes 4-9).

4.3.2. Lumped calibration versus gridded calibration

The bias-corrected PML-AET data, as well as its combination with the GRACE data are used to calibrate model parameters in Schemes 4-6 and Schemes 7-9, respectively. The difference in Schemes 4-6 is the spatial scale becomes more lumped with the increase of the scheme number and scheme 7-9 repeat the spatial scale of schemes 4-6. For daily runoff simulation, the mean NSE of schemes 4-9 is 0.39, 0.32, 0.26, 0.39, 0.31 and 0.27, respectively; the mean QR of schemes 4-9 is 0.40, 0.37, 0.29, 0.40, 0.40 and 0.29, respectively. For monthly runoff simulation, the mean NSE of schemes 4-9 is 0.65, 0.51, 0.47, 0.62, 0.50 and 0.48, respectively; the mean QR of schemes 4-9 is 0.45, 0.42, 0.34, 0.45, 0.44 and 0.33, respectively. As the spatial scale is expanded from scheme 4 to scheme 6, the calibration performance becomes worse. Schemes 7-9 give the similar performance for spatial dependency. These results indicate that the gridded model calibration schemes (scheme 4 and scheme 7) perform best.

4.3.3. Bias-corrected ET calibration versus calibration of bias-corrected PML-AET combined with GRACE data

Both the mean NSE and mean QR of scheme 4 is relatively similar to the NSE and QR in scheme 7. This is also generally true for scheme 5 versus scheme 8 and for scheme 6 versus scheme 9, as show in section 4.3.2. Comparing the results of scheme 7 with scheme 4 in Figure 5, the mean value of NSE and QR are similar, but the range of NSE becomes smaller slightly, indicated by noticeably higher NSE of daily runoff at the less than 25th percentiles. This means the scheme 7 gives more stable results. Similar patterns also fit at region (scheme 5 versus scheme 8) and catchment scales (scheme 6 versus scheme 9). Therefore, using GRACE together with PML-AET for model calibration has very limited benefit for both daily and monthly runoff prediction, compared to using PML-AET solely.

4.3.4. RS model calibration versus traditional regionalization

Scheme 7 is marginally better than scheme 4, and scheme 4 is noticeably superior to other PML-AET based calibration schemes. Therefore, scheme 4 is selected as the best candidate to compare with scheme 2, the traditional regionalization that is considered as the benchmark. In addition, the results of scheme 1, calibrated against the observed runoff, are considered as the best possible result.

In the daily runoff simulation, schemes 1 and 2 give a mean NSE of 0.58, 0.45 and mean QR of 0.33, 0.30, respectively. The mean NSE of scheme 4 decreases by 0.06 and mean QR of scheme 4 increases by 0.10, when compared to scheme 2. In monthly runoff simulation, schemes 1 and 2 give a mean NSE of 0.72, 0.56 and mean QR of 0.39, 0.34, respectively. The mean NSE of scheme 4 increases by 0.09 and mean QR of scheme 4 increases by 0.11, when compared with scheme 2. The mean NSE and mean QR of scheme 4 are also close to those of scheme 1 especially in monthly simulations. These results provide confident that model calibration against bias-corrected PML-ET at each grid cell can simulate ungauged catchment and regional runoff almost as well as traditional calibration and regionalization against streamflow data approaches.

4.3.5. Summary

Results shown in sections 4.3.1 to 4.3.4 indicate that bias correction of PML-AET is critical for improving the runoff prediction/simulation in ungauged or poorly gauged catchments comparing to traditional regionalization method. The RS-based model calibration framework performs better at gridded scale than at lumped scale, which reflects the advantage of remote sensing in that is spatially and temporally explicit across the global land surface. However, combining GRACE water storage data with the bias-corrected PML-AET has very limited benefit to further improve the predictions.

4.4. Spatial characteristics of optimum model calibration schemes

Figure 6 shows the NSE and QR spatial patterns from schemes 4 and 7. The NSE and QR spatial patterns of schemes 4 and 7 are very similar with a difference of less than 0.1 in most catchments. For both schemes, the NSE of monthly runoff is generally larger than the NSE of daily runoff. This is expected because of impacts of precipitation seasonality (Zhang et al., 2020). Another spatial feature is that the NSE values in mainstream catchments are generally larger than those in tributary catchments. The NSE values of schemes 4 and 7 for Nike (05) and Lugu (24) catchments are negative, but the QR values of them are positive. All in all, the NSE and QR spatial patterns of schemes 4 and 7 are similar and both indicate better runoff simulations in mainstream catchments than in small catchments. The result in Daluo station is always good, this might be the result of the application of Q at Daluo station when correcting bias of the PML-AET.

Figure 6 is about here

Figures 7a-7l further shows spatial patterns of performance of scheme 4 by calculating the increment, compared to scheme 1 and scheme 2. Figures 7i-7l shows spatial patterns of performance of scheme 7 by calculating the increment, compared to scheme 4. The increments are calculated as follows:

$$\Delta NSE = NSE_a - NSE_b, \quad \Delta QR = QR_a - QR_b \tag{8}$$

where *a* and *b* refer to the proposed scheme and benchmark scheme, respectively. The blue dots in Figure 7 indicate positive increments in that basin, the grey dots indicate no obvious increments or decrements, and the red ones indicate negative increments. The darker the color is, the greater the deference is.

Figure 7 is about here

Figures 7a-7d, 7e-7h, 7i-7l show the daily and monthly distribution of ΔNSE and ΔQR. There are three main patterns from Figures 7a, 7e, 7i, 7c, 7g and 7k in which the daily simulations are described.

The first pattern is that the ΔNSE of scheme 4 is greater than it of scheme 1 in 10% catchments with positive ΔQR in most regions. The result shows that though scheme 4 performs no better than scheme 1 in most catchments but it outperforms scheme 1 in certain catchment which shows the advantage of remote sensing data and gridded calibration.

The second pattern is that in all 11 main stream stations the ΔNSE values for scheme 4 minus scheme 2 are positive with grey or light blue dots in daily simulations, which means scheme 4 performs slightly better than scheme 2 for daily runoff simulation, in upstream and large catchments which are also in main stream (e.g. Ganzi catchment). Considering the model performance decided by both ΔNSE and ΔQR, scheme 4 performs reasonably in simulating daily runoff in downstream and small catchments, compared to scheme 2.

The third pattern is that the combination of GRACE shows a marginal improvement in most catchments according to ΔNSE, but shows slightly decrement of ΔQR in downstream catchment. All in all, scheme 7 marginally improve the model performance of scheme 4.

In monthly runoff simulation (Fig.7b, 7f, 7j, 7d, 7h, 7l), there are about 73% and 77% of the catchments with no negative ΔNSE and ΔQR values for scheme 4 minus scheme 1, respectively; and about 16% of catchments with negative values for both ΔNSE and ΔQR for scheme 4 minus scheme 1. There are about 87% and 77% of the catchments with positive ΔNSE and ΔQR values for scheme 4 minus scheme 2, respectively; and only about 7% of catchments with negative values for both ΔNSE and ΔQR for scheme 4 minus scheme 2. Scheme 7 performs similar to scheme 4 with 6% of catchments with negative values for both ΔNSE and ΔQR for scheme 7 minus scheme 4.

In summary, in daily runoff simulations, scheme 4 performs close to scheme 1 in large and main stream catchments, and outperforms scheme 1 in few catchments. Scheme 4 also perform better than scheme 2 in upper catchments and mainstream large catchments. Scheme 4 and scheme 7 show similar performance in most regions. In monthly runoff simulations, the model performance of scheme 4 against schemes 1 and 2 improved in upper and main stream large catchments. Scheme 4 performs no worse than scheme 1 in 84% catchments and outperforms scheme 2 in 28 out of the

30 catchments. Overall, scheme 7 marginally improve model performance of scheme 4, and scheme 4 performs close to scheme 1, or better than scheme 1 in few regions. scheme 4 also performs better than scheme 2 in upper catchments and mainstream large catchments.

5. Discussion

5.1. Potential for using RS data calibration methods

The climate and topography of the Yalong River is complex and covers a wide range, ranging from alpine mountains to humid basins. The complex topography and climate is one of the reasons of few gauges in the Yalong River basin in its upstream alpine regions. However, this region contributes majority of water resources for Jinsha River, which a major tributary for the Yangtze River (Kang et al., 2001; Yang et al., 2006). Therefore, it has important implication to use RS-data to calibrate hydrological modelling for improving prediction skills in this region or other similar regions.

This study explores the performance of seven RS-data based calibration schemes in 30 catchments of the Yalong River basin. Though the mean NSE and QR of daily runoff of schemes 4-9 are no better than that obtained from traditional regionalization result (scheme 2), the performance of scheme 4 is slightly better than scheme 2 in upstream and large catchments and the results of monthly runoff simulation of certain schemes (schemes 4 and 7) are superior to the those obtained from scheme 2. Scheme 4 even outperforms scheme 1 for simulating daily runoff in a couple of catchments, which demonstrates the advantage for model calibration against PML-AET at each grid cell, and the advantage is more noticeable at monthly scale. This indicates that the proposed approaches, especially for scheme 4, have great potential in data sparse regions.

5.2. Why the bias-corrected PML-AET works better

Our results demonstrate that it is necessary to bias correct PML-AET data for more reliable model calibration in Yalong River Basin. Since the raw PML-AET is not bias corrected using water balance method, therefore, it is inevitable to get noticeable biases in some areas, such as Yalong River Basin (Fig.4). The bias correction is crucial in the study area as demonstrated by comparing the calibration schemes 3 and 4. It is noted that this study aims to improve the PML-AET model calibrations in ungauged or poorly gauged catchments (Zhang et al., 2020), we use a single value of mean annual runoff data in a downstream gauge and in an independent period, which guarantees that the PML-AET model calibration approach having the potential for large scale application. Furthermore, using a single parameter of α in Fu model can generate reasonable mean annual runoff for most of the 30 catchments, demonstrating that the applicability of using the downstream catchment for bias correction. All in all, the bias correction method of PML-AET is reasonable with reliable gridded product and limited surface data.

5.3. Advantage and disadvantage of gridded model calibration

Remote sensing data is gridded, and it can reduce uncertainty related to lumped calibrations and gives detailed parameters for each grid (Arnold et al., 2010; Li and Zhang, 2017). In this study, the gridded hydrological modelling shows great advantage compared to lumped hydrological modelling. The gridded calibration schemes outperform in terms of both NSE and QR compared with lumped calibration schemes. It is noted that the time consumption increases by about 170 folds from lumped calibration to gridded calibration. But, this cost should be paid for achieving better predictions. Therefore, a more efficient algorithm is needed to reduce time consumption in the future, and if necessary, a compromise should be made between model accuracy and time consumption for practical application.

5.4. Adding GRACE data has very limited benefit to prediction

Though available studies show GRACE water storage data has been properly applied in basin scales (Rodell et al., 2004), and the snow storage at high latitudes is also considered in GRACE water storage data (Syed et al., 2008), this study found that the benefit to include GRACE data for model calibration is limited. This could be caused by the fact that the total water volume has been already properly considered by bias-corrected PML-AET. Furthermore, the resolution of GRACE data is spatially (1° x 1°) and temporally (monthly) coarse. It is probably not appropriate to apply GRACE data into the small and median sized catchments located on the Yalong River

Basin with complex terrains and large ranges in elevations (Kang et al., 2001). Therefore, more researches are needed to better utilize GRACE data for model calibration in the catchments with complex terrains.

5.5. Limitations and further directions

This study does not consider snow cover for model calibration though the recharge ratio of snowmelt runoff is relatively large, and it is the main component of runoff in the upper reach of Yalong River basin (Kang et al., 2001). In addition, spring runoff has a strong response to climate warming in alpine areas of Yalong River basin (Deng and Hou, 1996; Liu et al., 2019a). In the future, the snow cover should be considered into the upper reach catchments runoff simulation (Kang et al., 2001). However, hydrological models need to modified, making sure the modified structure has physically meaningful conceptualization for appropriately assimilating remote sensing data, such as snow cover and soil moisture.

The ‘natural flow’ is obtained by ignoring irrigation and other human-activity consumption of water volume in this study. The method is reasonable during 2004-2012 due to less influences of reservoir dispatching. But with the running of hydropower stations (such as Ertan hydropower station, Jinpin hydropower station) and land use change in recent years, the human activity is severe especially in downstream catchments (Liu et al., 2017; Liu et al., 2019b). For runoff simulation and prediction after 2012 in Yalong River basin, a human-activity based hydrological model with accurate remotely sensing data is essential and benefits both hydrology and management (Montanari et al., 2013).

The calibration schemes still need to be improved in some aspects. Running model at grids brings not only increased accuracy but also increased time consumption. In addition, incorporating GRACE data improves the model stability across the selected catchments though the overall improvement is marginal. Furthermore, the main challenge of applying remote sensing data into rainfall-runoff modelling includes choosing proper products, reducing the uncertainty of the products and matching remote sensing data with model variables (Li et al., 2016). Therefore, the model structure and constraining variables still need to be further developed.

6. Conclusion

In this study, nine modelling schemes are applied and examined for runoff prediction in the Yalong River basin, an idea region for testing the benefit of using remote sensing data since it has complex terrain conditions and sparse streamflow observation network. The PML-AET datasets are first evaluated and then bias corrected using very limited number of streamflow data. The performance of calibration schemes is noticeably better after bias correction of non bias-corrected PML-AET. The performance of gridded modelling is much better than lumped modelling, albeit with a large increase in model run times. The calibration schemes incorporating GRACE data is very limited benefit to schemes solely calibrated by bias-corrected PML-AET. Using bias-corrected PML-AET to constrain gridded hydrological model outperforms lumped regionalization hydrological modelling (sometimes even lumped calibration against observed streamflow) especially in monthly runoff simulation at upstream and large catchments. Utilizing quasi-runoff free method in gridded way might improve the performance of lumped calibration against observed streamflow.

This study implies that there is great potential to utilize and improve the runoff-free (or very limited runoff) calibration method in data sparse region. In future research, other remote sensing datasets (such as snow cover dataset) in both high resolution and high quality should be examined to constrain model variables along with incorporating human activity in ungauged catchments.

Acknowledgements

This study was supported by the CAS Talent Program, the National Natural Science Foundation of China (Grant No. 41971032 and 51879172) and the Second Tibetan Plateau Scientific Expedition and Research (2019QZKK0208). We acknowledge the Yalong River hydropower development company, China, for providing daily runoff data for 30 hydrology stations. PML evapotranspiration and GRACE water storage data are freely available from Data Catalog of the Google Earth Engine (https://developers.google.com/earth-engine/datasets). The Climate Meteorological Forcing Dataset used in this study is provided by National Tibetan Plateau Data Center (http://data.tpdc.ac.cn). Daily dataset of China's surface climate data is available from the China Meteorological Data Service Center (http://data.cma.cn/). We thank anonymous reviewers and editors for their critical but constructive comments on this paper.

Declaration of competing interest

The authors declare no conflicts of interest.

Author contributions

YQZ conceived this study. QH prepared and performed data analysis and prepared figures. QH, GHQ and YQZ wrote the paper and other authors contributed discussion and interpretations of the results and manuscript revision.

References

Allen, R.G., Pruitt, W.O., Wright, J.L., Howell, T.A., Ventura, F., Snyder, R. , et al. (2006). A recommendation on standardized surface resistance for hourly calculation of reference ETO by the FAO56 Penman-Monteith method. *Agricultural Water Management*, *81*(1-2), 1-22. DOI:10.1016/j.agwat.2005.03.007

Arnold, J.G., Allen, P.M., Volk, M., Williams, J.R., & Bosch, D.D. (2010). Assessment of different representations of spatial variablility on SWAT model performance. *Transactions of the Asabe*, *53*(5), 1433-1443.

Cheng, C., Zhao, M., Chau, K., & Wu, X. (2006). Using genetic algorithm and TOPSIS for Xinanjiang model calibration with a single procedure. *Journal of Hydrology*, *316*(1-4), 129-140.

Delhomme, J.P. (1978). Kriging in the Hydrosciences. *Advances in Water Resources*, *1*(5), 251-266. DOI:10.1016/0309-1708(78)90039-8

Deng, Y., & Hou, Y. (1996). Climatic Warming and its Impact on the Water Resources of the Yalong

River, China, Regional Hydrological Response to Climate Change. *Springer*, pp. 381-387.

Fan, Y., Lu, H., Yang, K., He, J., Wang, W., Wright Jonathon, S. , et al. (2017). Evaluation of multiple forcing data sets for precipitation and shortwave radiation over major land areas of China. *Hydrology & Earth System Sciences*, *21*(11), 5805-5821. DOI:10.5194/hess-21-5805-2017

Fu, B.P. (1981). On the calculation of the evaporation from land surface. *Sci. Atmos. Sin*, *5*(1), 23-31.

Habib, E., Haile, A.T., Sazib, N., Zhang, Y., & Rientjes, T. (2014). Effect of Bias Correction of Satellite-Rainfall Estimates on Runoff Simulations at the Source of the Upper Blue Nile. *Remote Sensing*, *6*(7), 6688-6708. DOI:10.3390/rs6076688

He, J., & Yang, K. (2011). China Meteorological Forcing Dataset, Cold and Arid Regions Science Data Center at Lanzhou. *Cold and Arid Regions Science Data Center at Lanzhou*. DOI:10.3972/westdc.002.2014.db

He, J., Yang, K., Tang, W., Lu, H., Qin, J., Chen, Y. , et al. (2020). The first high-resolution meteorological forcing dataset for land process studies over China. *Scientific Data*, *7*(1). DOI:10.1038/s41597-020-0369-y

Holland, J.H. (1992). Genetic Algorithms. *Scientific American*, *267*(1), 66-72. DOI:10.1038/scientificamerican0792-66

Hrachowitz, M., Savenije, H.H.G., Bloeschl, G., McDonnell, J.J., Sivapalan, M., Pomeroy, J.W. , et al. (2013). A decade of Predictions in Ungauged Basins (PUB)a review. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, *58*(6), 1198-1255. DOI:10.1080/02626667.2013.803183

Hundecha, Y., & Bardossy, A. (2004). Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model. *Journal of Hydrology*, *292*(1-4), 281-295. DOI:10.1016/j.jhydrol.2004.01.002

Jayawardena, A., & Zhou, M. (2000). A modified spatial soil moisture storage capacity distribution curve for the Xinanjiang model. *Journal of Hydrology*, *227*(1-4), 93-113.

Ju, Q., Yu, Z., Hao, Z., Ou, G., Zhao, J., & Liu, D. (2009). Division-based rainfall-runoff simulations with BP neural networks and Xinanjiang model. *Neurocomputing*, *72*(13-15), 2873-2883.

Kang, E., Cheng, G., Lan, Y., & Chen, X. (2001). Alpine runoff simulation of the Yalong River for the south-north water diversion. *J Glaciol Geocryol*, *23*(1), 139-148.

Konak, A., Coit, D.W., & Smith, A.E. (2006). Multi-objective optimization using genetic algorithms: A tutorial. *Reliability Engineering & System Safety*, *91*(9), 992-1007. DOI:10.1016/j.ress.2005.11.018

Kundu, D., Vervoort, R.W., & van Ogtrop, F.F. (2017). The value of remotely sensed surface soil moisture for model calibration using SWAT. *Hydrological Processes*, *31*(15), 2764-2780. DOI:10.1002/hyp.11219

Landerer, F.W., & Swenson, S. (2012). Accuracy of scaled GRACE terrestrial water storage estimates. *Water resources research*, *48*(4).

Li, H., & Zhang, Y. (2017). Regionalising rainfall-runoff modelling for predicting daily runoff: Comparing gridded spatial proximity and gridded integrated similarity approaches against their lumped counterparts. *Journal of Hydrology*, *550*, 279-293. DOI:10.1016/j.jhydrol.2017.05.015

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

Li, Y., Grimaldi, S., Walker, J.P., & Pauwels, V.R.N. (2016). Application of Remote Sensing Data to Constrain Operational Rainfall-Driven Flood Forecasting: A Review. *Remote Sensing*, *8*(6). DOI:10.3390/rs8060456

Liu, X., Chen, R., Liu, J., Wang, X., Zhang, B., Han, C. , et al. (2019a). Effects of snow-depth change on spring runoff in cryosphere areas of China. *Hydrological Sciences Journal*, *64*(7), 789-797.

Liu, X., Peng, D., & Xu, Z. (2017). Identification of the impacts of climate changes and human activities on runoff in the Jinsha River Basin, China. *Advances in Meteorology*, *2017*.

Liu, X., Yang, M., Meng, X., Wen, F., & Sun, G. (2019b). Assessing the impact of reservoir parameters on runoff in the Yalong River Basin using the SWAT Model. *Water*, *11*(4), 643.

Merz, R., & Bloschl, G. (2004). Regionalisation of catchment model parameters. *Journal of Hydrology*, *287*(1-4), 95-123. DOI:10.1016/j.jhydrol.2003.09.028

Montanari, A., Young, G., H, H.G.S., D, H., T, W., L, L., Ren , et al. (2013). “Panta Rhei—Everything Flows”: Change in hydrology and society—The IAHS Scientific Decade 2013–2022. *Hydrological Sciences Journal*, *58*(6), 1256-1275. DOI:10.1080/02626667.2013.809088

Nash, J.E., & Sutcliffe, J.V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, *10*(3), 282-290. DOI:https://doi.org/10.1016/0022-1694(70)90255-6

Oudin, L., Andreassian, V., Perrin, C., Michel, C., & Le Moine, N. (2008). Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments. *Water Resources Research*, *44*(3). DOI:10.1029/2007wr006240

Pechlivanidis, I.G., & Arheimer, B. (2015). Large-scale hydrological modelling by using modified PUB recommendations: the India-HYPE case. *Hydrology and Earth System Sciences*, *19*(11), 4559-4579. DOI:10.5194/hess-19-4559-2015

Pomeon, T., Diekkrueger, B., Springer, A., Kusche, J., & Eicker, A. (2018). Multi-Objective Validation of SWAT for Sparsely-Gauged West African River Basins-A Remote Sensing Approach. *Water*, *10*(4). DOI:10.3390/w10040451

Post, D.A., & Jakeman, A.J. (1999). Predicting the daily streamflow of ungauged catchments in SE

Australia by regionalising the parameters of a lumped conceptual rainfall-runoff model. *Ecological Modelling*, *123*(2-3), 91-104. DOI:10.1016/s0304-3800(99)00125-8

Ren, M., Xu, Z., Pang, B., Liu, W., & Liu, J. (2018). Accuracy Evaluation of A Variety of Satellite-Derived Precipitation Products in Beijing City. *AGUFM*, *2018*, H33I-2211.

Rodell, M., Famiglietti, J.S., Chen, J., Seneviratne, S.I., Viterbo, P., Holl, S. , et al. (2004). Basin scale estimates of evapotranspiration using GRACE and other observations. *Geophysical Research Letters*, *31*(20). DOI:10.1029/2004gl020873

Standardization Administration of the People's Republic of China, & General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China. (2008).Standard for Hydrological Information And Hydrological forecasting. Standards Press of China.

Stisen, S., & Sandholt, I. (2010). Evaluation of remote-sensing-based rainfall products through predictive capability in hydrological runoff modelling. *Hydrological Processes*, *24*(7), 879-891. DOI:10.1002/hyp.7529

Sutanudjaja, E.H., van Beek, L.P.H., de Jong, S.M., van Geer, F.C., & Bierkens, M.F.P. (2014). Calibrating a large-extent high-resolution coupled groundwater-land surface model using soil moisture and discharge data. *Water Resources Research*, *50*(1), 687-705. DOI:10.1002/2013wr013807

Swenson, S., & Wahr, J. (2006). Post-processing removal of correlated errors in GRACE data. *Geophysical Research Letters*, *33*(8).

Syed, T.H., Famiglietti, J.S., Rodell, M., Chen, J., & Wilson, C.R. (2008). Analysis of terrestrial water storage changes from GRACE and GLDAS. *Water Resources Research*, *44*(2). DOI:10.1029/2006wr005779

Wanders, N., Bierkens, M.F.P., de Jong, S.M., de Roo, A., & Karssenberg, D. (2014). The benefits of using remotely sensed soil moisture in parameter identification of large-scale hydrological models. *Water Resources Research*, *50*(8), 6874-6891. DOI:10.1002/2013wr014639

Wu, Y., Guo, L., Zheng, H., Zhang, B., & Li, M. (2019). Hydroclimate assessment of gridded precipitation products for the Tibetan Plateau. *Science of The Total Environment*, *660*, 1555-1564.

Yadav, M., Wagener, T., & Gupta, H. (2007). Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. *Advances in Water Resources*, *30*(8), 1756-1774. DOI:10.1016/j.advwatres.2007.01.005

Yang, F., Lu, H., Yang, K., Wang, W., Li, C., Han, M. , et al. (2017). Evaluation and comparison among multiple forcing data sets for precipitation and shortwave radiation over mainland China. *Hydrology and Earth System Sciences Discussions*, *21*(11), 1-32.

Yang, Z., Wang, H., Saito, Y., Milliman, J.D., Xu, K., Qiao, S. , et al. (2006). Dam impacts on the

Changjiang (Yangtze) River sediment discharge to the sea: The past 55 years and after the Three Gorges Dam. *Water Resources Research*, *42*(4). DOI:10.1029/2005wr003970

Yao, C., Li, Z., Bao, H., & Yu, Z. (2009). Application of a developed Grid-Xinanjiang model to Chinese watersheds for flood forecasting purpose. *Journal of Hydrologic Engineering*, *14*(9), 923-934.

Yassin, F., Razavi, S., Wheater, H., Sapriza-Azuri, G., Davison, B., & Pietroniro, A. (2017). Enhanced identification of a hydrologic model using streamflow and satellite water storage data: A multicriteria sensitivity analysis and optimization approach. *Hydrological Processes*, *31*(19), 3320-3333. DOI:10.1002/hyp.11267

Zhang, L., Hickel, K., Dawes, W., Chiew, F.H., Western, A., & Briggs, P. (2004). A rational function approach for estimating mean annual evapotranspiration. *Water resources research*, *40*(2).

Zhang, L., Potter, N., Hickel, K., Zhang, Y., & Shao, Q. (2008). Water balance modeling over variable time scales based on the Budyko framework–Model development and testing. *Journal of Hydrology*, *360*(1-4), 117-131.

Zhang, X., & Tang, Q. (2015). Combining satellite precipitation and long-term ground observations for hydrological monitoring in China. *Journal of Geophysical Research-Atmospheres*, *120*(13), 6426-6443. DOI:10.1002/2015jd023400

Zhang, Y., & Chiew, F.H.S. (2009). Relative merits of different methods for runoff predictions in ungauged catchments. *Water Resources Research*, *45*(7). DOI:10.1029/2008wr007504

Zhang, Y., Chiew, F.H.S., Liu, C., Tang, Q., Xia, J., Tian, J. , et al. (2020). Can Remotely Sensed Actual Evapotranspiration Facilitate Hydrological Prediction in Ungauged Regions Without Runoff Calibration? *Water Resources Research*, *56*(1). DOI:10.1029/2019wr026236

Zhang, Y., Kong, D., Gan, R., Chiew, F.H.S., McVicar, T.R., Zhang, Q. , et al. (2019). Coupled estimation of 500 m and 8-day resolution global evapotranspiration and gross primary production in 2002–2017. *Remote Sensing of Environment*, *222*, 165-182. DOI:https://doi.org/10.1016/j.rse.2018.12.031

Zhao, R.J. (1980). The xinanjiang model, Proceedings of the Oxford Symposium.

Zhao, R.J. (1992). The Xinanjiang model applied in China. *Journal of Hydrology*, *135*(1-4), 371-381.

Figures and figure captions

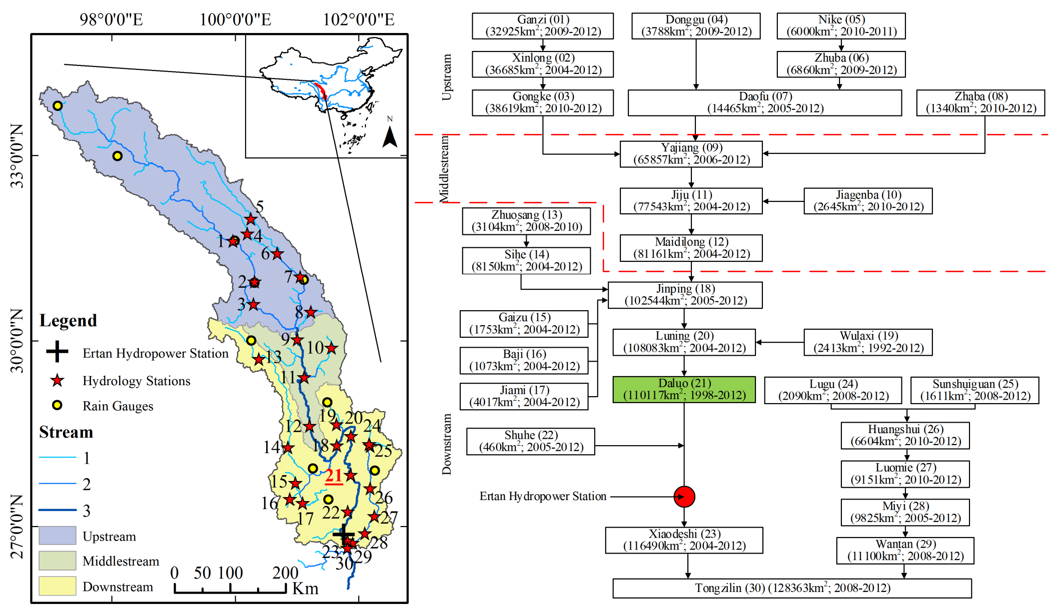

Figure 1. Information and location of study area. The station Daluo for constraining Fu model is labelled as 21.

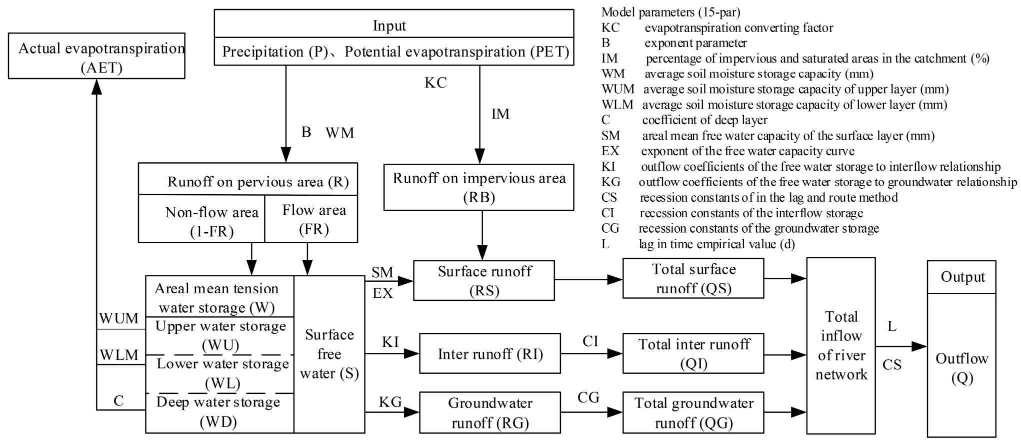

Figure 2. Model structure of Xinanjiang Model

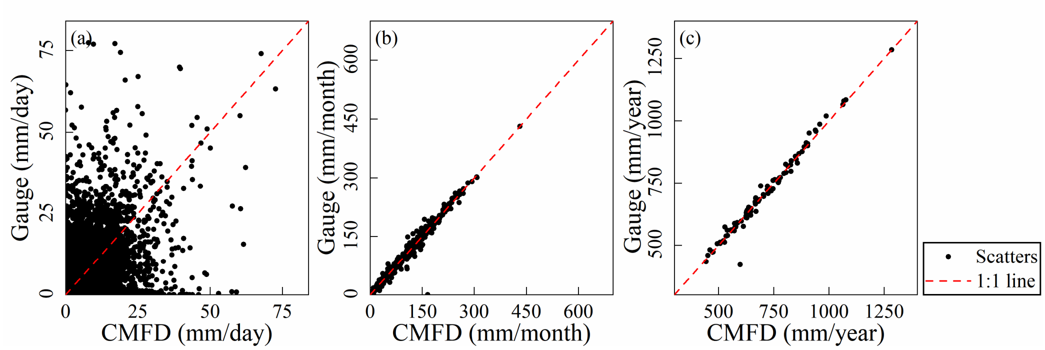

Figure 3. Comparison between observed precipitation and precipitation generated from CMFD data (CMFD-P)

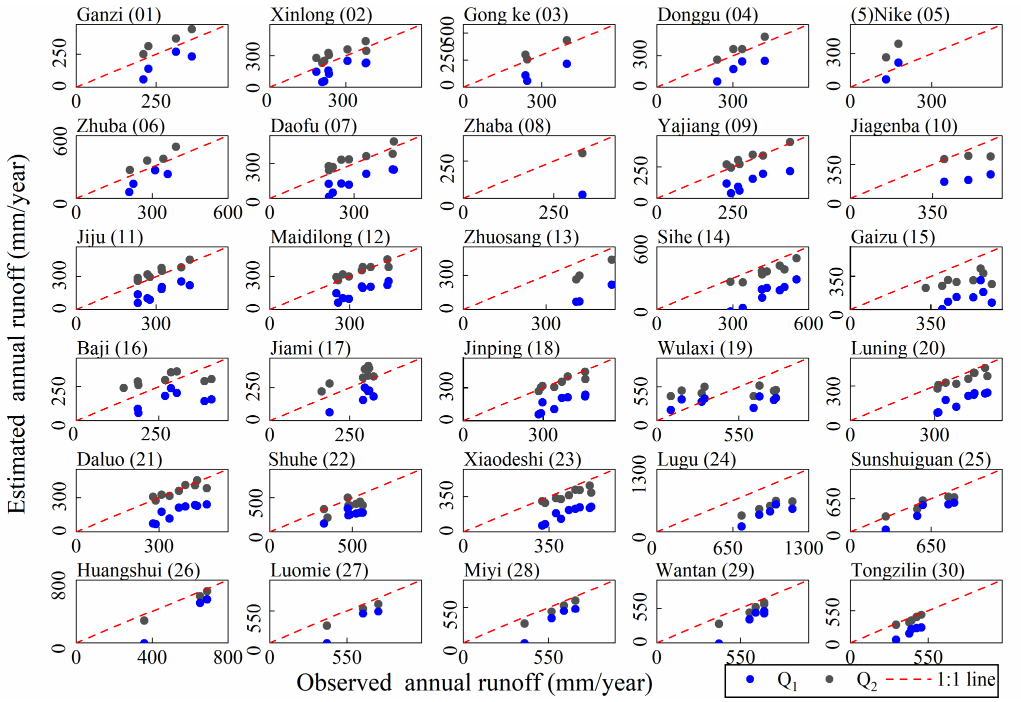

Figure 4. Evaluating annual runoff obtained from precipitation minus non bias-corrected PML-AET (Q_1) and that (Q_2) obtained from precipitation minus bias corrected PML-AET (The numbers in the bracket represent the watershed codes shown in Figure 1.)

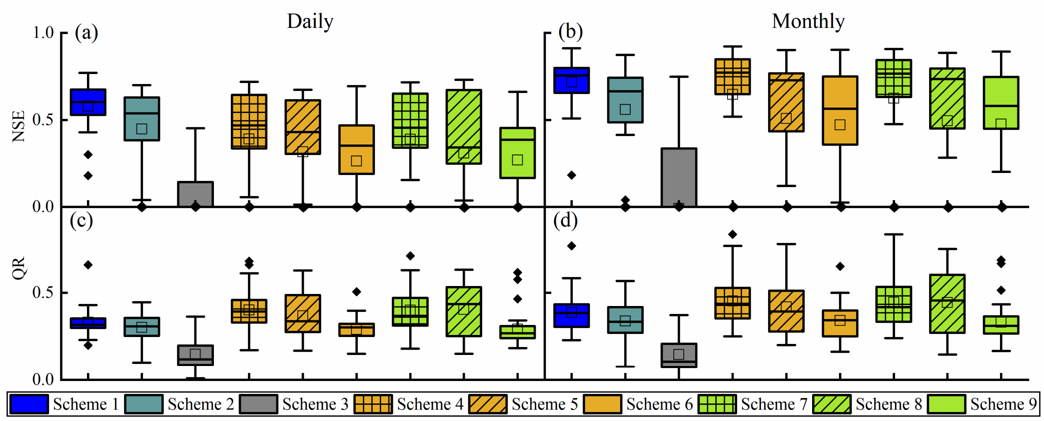

Figure 5. Comparison of performance of the nine calibration schemes for estimating streamflow. Noted that negative NSE values are plotted as zero for better visualization (The boxes represent the values range from 25th to 75th percentiles, the lines in each plot from top to bottom represent upper boundary, median value and lower boundary, respectively. The square represents the mean value and the rhombus represents the outlier.)

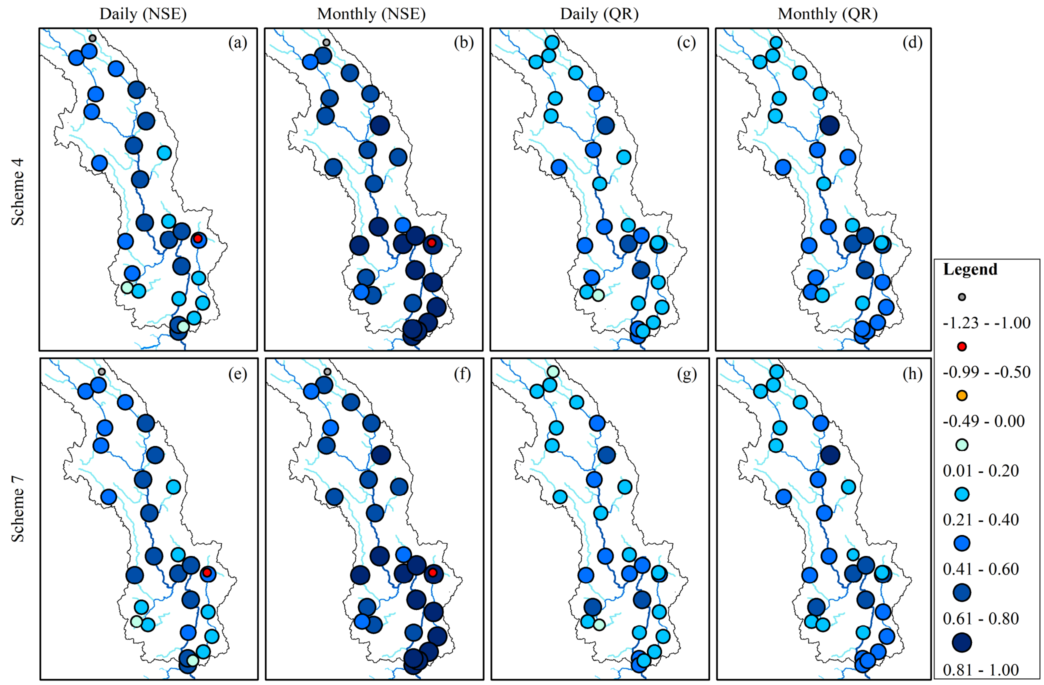

Figure 6. NSE and QR spatial patterns obtained from scheme 4 and scheme 7

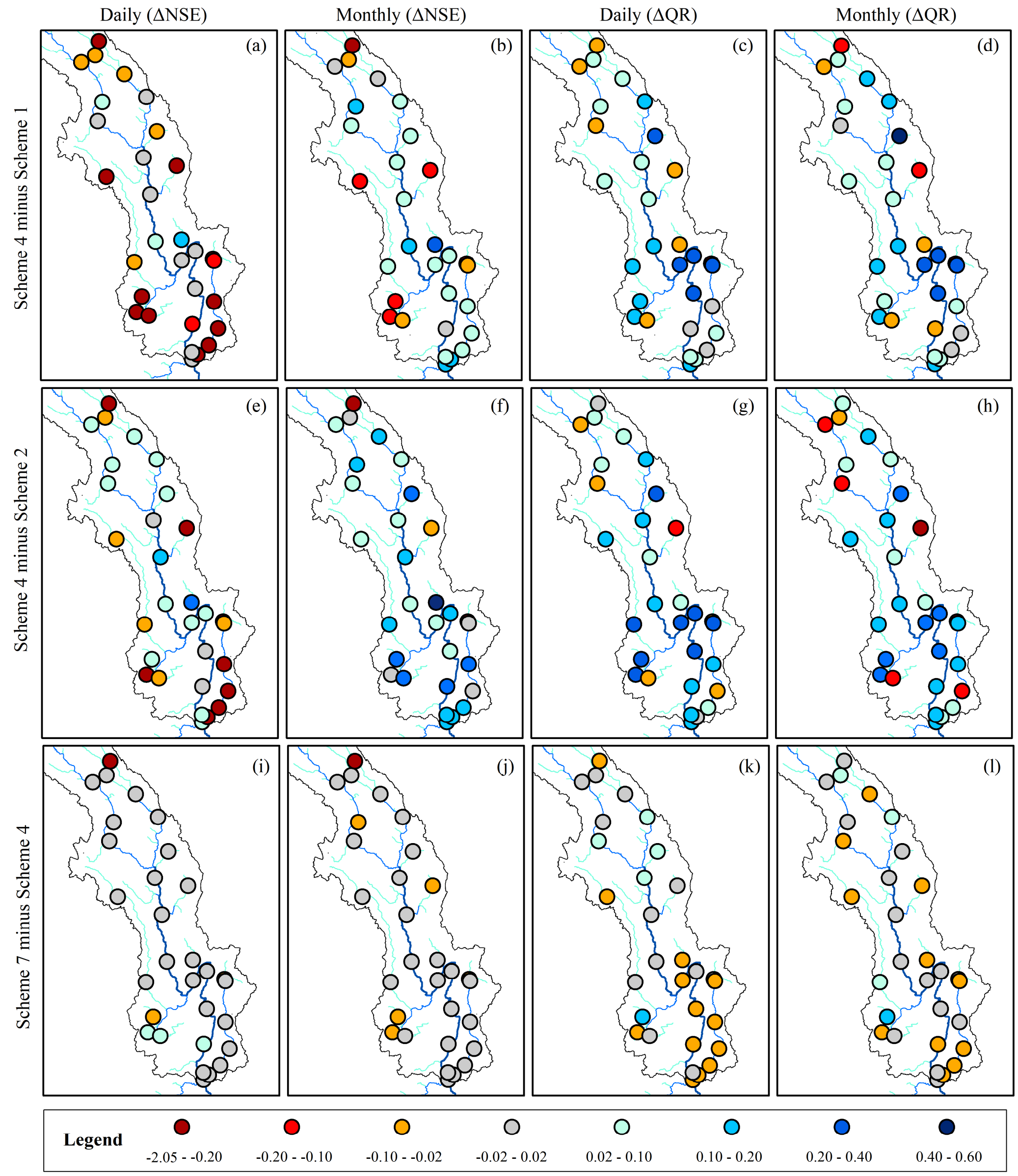

Figure 7. Spatial evaluation of scheme 4 against scheme 1, 2 and 7. The difference among them is obtained from Equation (8). Having a range from -0.02 to 0.02, gray means the two perform similarly.