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

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

[1]College of Water Resource & Hydropower, Sichuan University
[2]Sichuan University
[3]Institute of Geographic Sciences and Natural Resources Research
[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 24, 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 catchments and poorly gauged regions, a challenging area of research in hydrology. 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 in the Yalong River basin in China. To this end, seven RS data calibration schemes are explored, and compared to direct calibration against observed runoff and traditional regionalization using spatial proximity to predict runoff in ungauged catchments. The results show that using bias-corrected remotely sensed AET (bias-corrected PML-AET data) for constraining model calibration performs much better than using the raw 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 using lumped data, and outperforms the traditional regionalization approach especially in headwater 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 there is great potential in using bias-corrected RS-AET data to calibrating hydrological models (without the need for gauged streamflow data) to estimate daily and monthly runoff time series in ungauged catchments and sparsely gauged regions.

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 bias-corrected remote sensing data to calibrate hydrological model shows great potential especially in ungauged catchments.
- Compared to raw PML-AET, bias-corrected PML-AET improves runoff prediction noticeably and adding GRACE shows limited benefit.
- Gridded application performs better than lumped catchment modelling application for maximizing the benefit from the spatial PML-AET data.

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 catchments and poorly gauged regions, a challenging area of research in hydrology. 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 in the Yalong River basin in China. To this end, seven RS data calibration schemes are explored, and compared to direct calibration against observed runoff and traditional regionalization using spatial proximity to predict runoff in ungauged catchments. The results show that using bias-corrected remotely sensed AET (bias-corrected PML-AET data) for constraining model calibration performs much better than using the raw 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 using lumped data, and outperforms the traditional regionalization approach especially in headwater 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 there is great potential in using bias-corrected RS-AET data to

calibrating hydrological models (without the need for gauged streamflow data) to estimate daily and monthly runoff time series in ungauged catchments and sparsely gauged regions.

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

1. Introduction

Runoff Prediction in Ungauged Basins (PUB) is 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; Oudin et al., 2008; Zhang and Chiew, 2009; Hrachowitz et al., 2013; Li and Zhang, 2017). Oudin et al. (2008) compared classical regionalization schemes on 913 French catchments, and their result shows that regionalization based on spatial proximity provides the best solution among three regionalization methods (regression, spatial proximity and physical similarity). Therefore, spatial proximity is considered as a good approach for predicting runoff in ungauged catchments. However, the performance of the spatial proximity approach becomes gradually poorer with increase in regionalization distance (Li and Zhang, 2017), suggesting that the spatial proximity may not be suitable 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 with few stream gauges.

Remote sensing observation provides continuous data in both spatial and temporal scales, which makes it possible to estimate regional surface data in a quick and widely applicable way (Stewart and Finch, 1993; Sun et al., 2018). Therefore, remote sensing data has been widely applied and combined with hydrological models (Wanders et al., 2014; Beck et al., 2017a; Kittel et al., 2018; Kumar and Lakshmi, 2018). However, the quality of remote sensing data is not always guaranteed (Andersen et al., 2005; Liu et al., 2016; Beck et al., 2017b; Sun et al., 2018), and the accuracy varies across regions, which can have important regional implications (Hijmans et al., 2005; Wang et al., 2015). Thus, selection of the datasets should be done carefully. As inputs to hydrological models, the remote sensing data should be as accurate as possible. Studies show the bias correction of input data improves the runoff simulations under most conditions (Li et al., 2009b; 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 of Zhang et al. (2020) 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 quasi-runoff-free method (very limited runoff data) for hydrological model calibrations. Specifically, this work aims to improve calibration schemes by adding more remote sensing information (raw 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 (raw 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, China. 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, and the terrain mainly includes hilly plateaus, alpine canyons, and wide valley basins from north to south, respectively. All of these make the basin geography greatly different in both horizontal and vertical directions. In addition, the Yalong River basin covers a wide range of climate regimes varying from humid to semi humid climates and has contrasting dry and wet seasons. The mean annual precipitation is about 720 mm and the mean annual runoff is about 300-400 mm for the entire Yalong River basin. Half of the runoff in the Yalong River is formed by direct precipitation contribution, and the rest is replenished by groundwater and melting snow (ice) (Kang et al., 2001).

This study uses data from 30 catchments within the Yalong River basin. Figure 1 shows a map of the study area and information for 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 (simplified as CMFD-P). The CMFD-P has been shown to be a high 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 Penman-Monteith-Leuning model (abbreviated as PML_V1) was proposed by Leuning et al. (2008), and further improved by Zhang et al. (2010, 2016). The gridded actual evapotranspiration data used in this paper is obtained from the PML_V2 global evapotranspiration (simplified as PML-AET) product (Gan, 2018; Zhang et al., 2019). It is referred to as 'raw PML-AET' hereafter. In PML_V2, evaporation is divided into: transpiration from vegetation (E_c), direct evaporation from the soil (E_s) and evaporation of intercepted rainfall from vegetation (E_i). This study uses the PML-AET, equal to the sum of the three AET components defined above. 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 used in this paper is the 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). Three GRACE datasets come from three centers: the Jet Propulsion Laboratory (JPL), The University of Texas Center for Space Research (UTCSR) and the GeoForschungsZentrum Potsdam (GFZ), respectively. The GRACE data used in this study is the mean value of the three datasets. All the gridded datasets are split into 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 these 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° (approximately 11×11 km)	3-hour	1979-2018	https://data.tpdc.ac.cn/zh-hans/data/ 8028b944-daaa-4511-8769-965612652c49/	(He and Yang, 2011; Fan et al., 2017; He et al., 2020)
PML_V2	PML_V2 global evapotranspiration and gross primary production	0.05° (approximately 5×5 km)	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°(approximately 111×111 km)	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 the water balance method. As shown in Figure 1, the Xiaodeshi hydrological station and the Tongzilin hydrology station are 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. CMFD-P provides gridded data, and here it is validated against the observed rainfall data at ten rain gauges. The main idea is to verify the accuracy through daily precipitation detection 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
Detection 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*	$SI = 1 - \frac{\sum_{i=1}^{n}(P_i - G_i)^2}{\sum_{i=1}^{n}(\lvert G_i - \overline{G}\rvert + \lvert P_i - \overline{G}\rvert)^2}$	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 number of samples.

3.1.2. Bias correction of PML-AET

The PML-V2 is a global evapotranspiration and gross primary product dataset. To enhance its utility for this study, the mean annual PML-AET is bias corrected 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. Input data

To adhere to the principle of "essentially runoff-free calibration", only data from one single basin towards the downstream end of the system is used. This is the gauging station for the Daluo River Basin (Gauging station 21, see Figure 1) with streamflow data from 1999 to 2012. Mean annual precipitation comes from the CMFD-P gridded data. Mean annual potential evapotranspiration (E_p) is estimated using the Allen et al.

(2006) equation following Penman-Monteith method (Eq.(1)), using climate input data from the CMFD dataset (i.e. temperature, humidity, wind speed), 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 evapotranspiration (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/(m^2*d)); γ is the psychrometric constant (kPa/°C); T_{mean} is the daily mean 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. The Fu model

We used the classical Budyko framework – Fu model – to estimate mean annual Q (called Q_{fu} hereafter) (Fu, 1981; Zhang et al., 2004; Zhang et al., 2008). Q_{fu} is expressed as:

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

where Q_{fu} represents mean annual runoff (mm/year); P is mean annual precipitation (mm/year); E_p is mean annual potential evapotranspiration (mm/year); AI is the aridity index, calculated as E_p divided by P; α is a parameter that represents climate and physical characteristics. The value of the parameter α is estimated based on the basin-average mean annual precipitation and evapotranspiration, and the streamflow from the single Daluo Basin, and this value was 1.56. This α = 1.56 is then 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.

iii. Gridded 'real' mean annual PML-AET

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

iv. Scaling factor

A scaling factor SC at each grid cell is calculated as the 'real' mean annual AET divided by mean annual raw PML-AET; and

v. Bias-corrected PML-AET (8-day data)

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

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, developed by Zhao (1980). The model has been extensively used for runoff simulation and prediction across humid and semi-humid regions globally (Moore and Clarke, 1981; Zhao, 1992; Todini, 1996; Jayawardena and Zhou, 2000; Cheng et al., 2006; Ju et al., 2009; Li et al., 2009a; 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: gridded, regional and catchment. This means that the model (i.e. optimization of the 15 parameters in the model) is calibrated at each grid cell, each region, and each catchment, respectively. For the grid calibration, each grid cell has its own set of parameter values. For regional calibration (a region is defined as the contribution area between two gauges), all the grid cells within the region have the same set of parameter values. Therefore, the lowest-level tributary comprises one region, but higher lever catchments comprise multiple regions. For instance, the Ganzi, Xinlong and Gongke Basins have one, two and three regions, respectively. For catchment calibration, all the grid cells in the entire catchment have the same set of parameter values. The model therefore becomes more lumped as the scale increase from gridded 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 (Holland, 1992; Konak et al., 2006), is used to optimize the model parameters. Population size and the generations for the genetic algorithm are set as 400 and 100, respectively. The optimum point can normally be achieved after ~50 generations of searching (Li and Zhang, 2017). The selection of parameter sets is based on the fitness function (objective function). In scheme 1, the model is calibrated against observed daily runoff by using lumped catchment inputs, which represents the best possible model simulation or calibration. Scheme 2 is regionalization based on spatial proximity, that is, runoff is predicted using parameter values from the closest donor catchment with streamflow data to calibrate the model (Merz and Bloschl, 2004; Oudin et al., 2008; Li and Zhang, 2017). This scheme is the “traditional” regionalization approach to estimate runoff in ungauged catchments, regarded as the baseline for evaluating the performance of schemes 3-9. Scheme 3 uses the raw PML-AET output for model calibration. Schemes 4-6 uses the bias-corrected PML-AET for model calibration, and the difference among them is that scheme 4 is calibrated at each PML-AET grid cell, scheme 5 is calibrated at each region, and scheme 6 is calibrated 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 with equal weighting.

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 against observed runoff			1	CMFD-P, Ep, (Q at 30 stations)	Eq.(3)
Regionalization			2	CMFD-P, Ep, a set of parameters (at a neighbor station)	
Raw PML-AET runoff-free calibration approach	3			CMFD-P, Ep, (raw 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 Eqs. 3-6.

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

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

$$F_3 = 1 - NSE_{ET2} \text{ , } NSE_{ET2} = 1 - \frac{\sum_{i=1}^{N} (AET_{B-PML} - AET_{SIM})^2}{\sum_{i=1}^{N} (AET_{B-PML} - \overline{AET_{B-PML}})^2} \text{ ,} \quad (5)$$

$$F_4 = (1 - NSE_{ET2}) + (1 - NSE_{\Delta W}) \text{ , } \quad NSE_{\Delta W} = 1 - \frac{\sum_{i=1}^{N} (\Delta W_{GRACE} - \Delta W_{SIM})^2}{\sum_{i=1}^{N} (\Delta W_{GRACE} - \overline{\Delta W_{GRACE}})^2} \text{ ,} \quad (6)$$

where Q_{obs} represents the observed daily runoff, Q_{sim} represents the simulated daily runoff. AET_{SIM}, AET_{PML} and AET_{B-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. Eq. (3) is performed at daily scale, Eq. (4) and Eq. (5) are performed at 8-day scale, and Eq. (6) is performed at monthly scale. 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 the objective function, the better the simulation quality.

3.4. Evaluating the nine modelling schemes

The Kling–Gupta efficiency (*KGE*) (Gupta et al., 2009; Kling et al., 2012), Qualified Rate (*QR*) (Standardization Administration of the People's Republic of China, 2008), *NSE* and Log-transformed *NSE* (*LogNSE*) are used to evaluate the performance of the nine schemes at different temporal scales. The four metrics are defined as follows:

$$KGE = 1 - \sqrt{\left(\frac{\mathrm{cov}(Q_{obs}, Q_{sim})}{\sigma^2_{obs}\sigma^2_{sim}} - 1\right)^2 + \left(\frac{\mu_{sim}}{\mu_{obs}} - 1\right)^2 + \left(\frac{\sigma_{sim}}{\sigma_{obs}} - 1\right)^2} \tag{7}$$

$$QR = \frac{\mathrm{m}}{n} \tag{8}$$

$$NSE = 1 - \frac{\sum_{\mathrm{i}=1}^{N}(Q_{obs} - Q_{sim})^2}{\sum_{\mathrm{i}=1}^{N}(Q_{obs} - \overline{Q_{obs}})^2} \tag{9}$$

$$\mathrm{Log}NSE = 1 - \frac{\sum_{\mathrm{i}=1}^{N}(Log(Q_{obs}) - Log(Q_{sim}))^2}{\sum_{\mathrm{i}=1}^{N}(Log(Q_{obs}) - \overline{Log(Q_{obs})})^2} \tag{10}$$

where m represents the numbers of samples whose ABIAS (absolute bias) are less than 0.35, n is the total number of samples (total number of daily, or monthly streamflow data), cov is the covariance between observation and simulation, σ is the standard deviation, μ is the mean, Log is the log-transformed values. The subscripts obs and sim standing for observed and simulated, respectively. *KGE* combines the correlation, bias and coefficients of variation in a balanced way. *QR* is the qualified rate of modelled runoff whose absolute bias are less than 0.35. *KGE* and *QR* focus more on overall model performance. *NSE* indicates the ability to reproduce middle and high flows, and log-transformed *NSE* puts more weight on low flows. The value of *QR* varies from 0 to 1, the closer to 1 indicating better model performance (*QR*=1 means that the absolute bias from all samples is less than 0.35). The values of *KGE*, *NSE* and *LogNSE* vary from negative infinity to 1, the closer to 1 indicating better model performance. The temporal step is daily and monthly 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 detection 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 accuracy has increased significantly. CC, *NSE* and SI have 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 results of the hydrological modelling, 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 raw PML-AET and bias-corrected PML-AET are evaluated using their performance for estimating annual streamflow. The annual streamflow predicted from each of them is estimated by annual precipitation minus annual raw PML-AET (Q_1) and annual precipitation minus annual bias-corrected PML-AET (Q_2), respectively. If the agreement between Q_2 and Q_{obs} is better than between Q_1 and Q_{obs}, then it can be concluded that 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. The Daluo catchment has similar climate and physical characteristics to most catchments in the study area. Thus, it is reasonable to apply the calibrated parameter α in the study area, and it is also as expected that Q_2 is significantly better than Q_1 in Figure 4. In most basins, scatters of Q_{obs} against Q_2 distribute evenly on both sides of the 1: 1 line, which means the agreement and consistency between Q_2 and Q_{obs} is good, while Q_1 is severely biased. The mean *BIAS* values of Q_1 and Q_2 are -0.54 and -0.04 for 30 catchments, respectively; the mean *ABIAS* values of Q_1 and Q_2 are 0.55 and 0.18 for 30 catchments, respectively. This result demonstrates that the bias-corrected PML-AET achieves much better water balance (in terms of producing streamflow), compared to the raw 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 should also improve the performance of hydrological model in these catchments.

Figure 4 is about here

Figure 5 shows the mean annual spatial and seasonal distributions of CMFD-P, bias-corrected PML-AET and GRACE soil water storage change data. The mean annual precipitation and mean annual actual evapotranspiration are 721 mm and 359 mm, respectively. In the upper and middle reaches, the precipitation is lower than that in the lower reach, while the actual evapotranspiration in the upper and middle reaches is higher than that in the lower reach for spring, summer and autumn. In winter, the spatial distribution of precipitation varies little across the study area, and the actual evapotranspiration in the upper and middle reaches is lower than that in the lower reach. This indicates that the climates become drier from south to north at most times of the year. Figures 5i-5l indicate a greater water storage change in the lower reach than in the upper and middle reaches. Replenishing snow and ice might help to reduce the variation of water storage change in upper and middle streams. The water storage decreases in autumn and winter, and increases in spring and summer. Overall, the mean annual water storage change is close to 0 with a slightly negative value of about 1 mm. The mean annual precipitation data and mean annual actual evapotranspiration data follow similar seasonal patterns, and the simulated mean annual runoff (P minus bias-corrected PML-AET) matches the observed mean annual runoff reasonably well in different parts of the Yalong River basin. The two results suggest that the bias corrected PML-ET is suitable for calibrating a hydrological model in the Yalong river basin.

Figure 5 is about here

4.3. Runoff prediction

The plots in Figure 6 summarize the performance of nine modelling schemes in predicting daily runoff (6a, 6c, 6e, 6g) and monthly runoff (6b, 6d, 6f, 6h) across the 30 catchments in the Yalong River basin (to present patterns clearly, negative values are not shown here, but are shown later in Figure 7). In each scheme, the simulated monthly runoff is accumulated from the daily runoff, and monthly simulations are generally better than the daily runoff simulations. The annual runoff performance has not been analyzed because of the relatively short records. The *KGE* and *QR* focus more on the overall model performance, while *NSE* and *LogNSE* focus more on high flow and low flow, respectively. The range of the metrics above describes modelling stability and the model is more stable across the flow regime with a lower range of metrics.

Figure 6 is about here

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

The simulated streamflow obtained from scheme 3 (calibration using the raw 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. Table 5 shows mean values of metrics for scheme 3 and scheme 4, and their difference.

Table 5. Mean values of metrics for scheme 3 and scheme 4, and the differences between the two

	Scheme 3	Scheme 4	Scheme 4 - Scheme 3
KGE (daily)	0.13	0.65	0.51
QR (daily)	0.15	0.40	0.25
NSE (daily)	-0.08	0.39	0.47
LogNSE (daily)	-4.45	0.09	4.55
KGE (monthly)	0.19	0.74	0.55
QR (monthly)	0.15	0.45	0.31
NSE (monthly)	-0.01	0.65	0.66
LogNSE (monthly)	-3.84	0.15	3.99

As shown in Table 5, compared to scheme 3, the performance of scheme 4 is greatly improved. At daily scale, the improvement is 0.51 in mean *KGE*, 0.25 in mean *QR*, 0.47 in mean *NSE* and 4.55 in mean *LogNSE*; at monthly scale, the improvement is 0.55 in mean *KGE*, 0.31 in mean *QR*, 0.66 in mean *NSE* and 3.99 in mean *LogNSE*. Therefore, using the bias-corrected PML-AET data for constraining model calibration performs much better than using the raw 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. quasi-runoff-free calibration method, 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 that the model becomes more lumped with increasing scheme number. Schemes 7-9 repeat the spatial scale of schemes 4-6. Table 6 summarizes mean values of metrics for schemes 4-9.

Table 6. Mean values of metrics for schemes 4-9

	Scheme 4	Scheme 5	Scheme 6	Scheme 7	Scheme 8	Scheme 9
KGE (daily)	0.65	0.54	0.49	0.64	0.56	0.54
QR (daily)	0.40	0.37	0.29	0.40	0.40	0.29
NSE (daily)	0.39	0.32	0.26	0.39	0.31	0.29
LogNSE (daily)	0.09	-0.76	-1.55	0.00	-0.19	-2.05
KGE (monthly)	0.74	0.61	0.53	0.73	0.61	0.68
QR (monthly)	0.45	0.42	0.34	0.45	0.44	0.33
NSE (monthly)	0.65	0.51	0.47	0.62	0.50	0.48
LogNSE (monthly)	0.15	-0.64	-1.36	0.03	0.02	-1.74

As the spatial scale becomes greater from scheme 4 to scheme 6, the calibration performance becomes worse. Schemes 7-9 give a similar performance for spatial dependency. The median values in Figure 6 also show the same pattern with the mean values. These results indicate that the gridded model calibration schemes (scheme 4 and scheme 7) perform best. The reason that gridded calibration outperforms lumped calibration is that gridded remote sensing data provides more information, and therefore spatial heterogeneity of runoff can be better simulated and predicted using the parameter sets obtained from gridded calibrations. The bias-corrected PML-AET calibrations have a slightly improved performance with the increase in calibration resolution.

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

The mean *KGE*, mean *QR* and mean *NSE* of scheme 4 are relatively similar to those in scheme 7. This is also generally true for scheme 5 versus scheme 8 and for scheme 6 versus scheme 9, as shown in Table 6. The mean *LogNSE* of schemes 4 and 6 is relatively similar to those in schemes 7 and 9, respectively. but the mean *LogNSE* of scheme 8 is significantly increased compared to scheme 5. This result suggests that incorporating GRACE data could improve the low flow simulation in regional calibration. Comparing the results of gridded calibrations (scheme 7 and scheme 4) in Table 6 and Figure 5, the mean value of *LogNSE* of scheme 7 is smaller than that of scheme 4, but the mean values of *KGE*, *QR* and *NSE* are similar, and the range of *NSE* becomes slightly smaller, as indicated by noticeably higher *NSE* of daily runoff at the less than 25^{th} percentiles. This means that scheme 7 gives similar overall results, more stable high flow modelling results, but also negative influences on low flow. Similar patterns are also found at catchment scales (scheme 6 versus scheme 9). Regional and gridded calibrations give similar patterns of *KGE*, *QR*, and *NSE*, but the *LogNSE* of scheme 8 is larger than that of scheme 5, indicating improvements in predicting low flows. The reason for this may be that the resolution of GRACE data is closer to regional scale. Therefore, using GRACE together with PML-AET for model calibration has very limited benefit for gridded and catchment calibrations, but improves the performance of low flows at regional calibrations, 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 only 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 here. The results are also compared with scheme 1 which provides the best possible direct calibration results for catchment calibrations. Table 7 shows mean values of metrics for schemes 1, 2 and 4.

Table 7. Mean values of metrics for schemes 1, 2 and 4

	Scheme 1	Scheme 2	Scheme 4	Scheme 4 - Scheme 1	Scheme 4 - Scheme 2
KGE (daily)	0.70	0.59	0.65	-0.05	0.06
QR (daily)	0.33	0.30	0.40	0.07	0.10
NSE (daily)	0.58	0.45	0.39	-0.19	-0.06
LogNSE (daily)	-1.39	-1.93	0.09	1.48	2.02
KGE (monthly)	0.71	0.57	0.74	0.04	0.18
QR (monthly)	0.39	0.34	0.45	0.07	0.11
NSE (monthly)	0.72	0.54	0.65	-0.07	0.11
LogNSE (monthly)	-1.05	-2.63	0.15	1.19	2.78

The mean daily KGE, QR and NSE of scheme 4 are similar to those of scheme 2, and the mean daily LogNSE of scheme 4 is greater than that of scheme 2. The mean monthly metrics of scheme 4 are significantly larger than those of scheme 2. The results indicate scheme 4 performs slightly better than scheme 2 for daily calibrations, and performs significantly better than scheme 2 for monthly calibrations. The mean *NSE* and mean *QR* of scheme 4 are also close to those of scheme 1 especially in monthly simulations. The increase of *LogNSE* indicates a better low flow performance of quasi-runoff-free calibration method (schemes 4-9) These results provide confidence that model calibration against bias-corrected PML-ET at each grid cell can simulate ungauged catchments almost as well as or even better than traditional calibration and regionalization against streamflow data approaches to predict runoff in ungauged catchments.

4.3.5. Summary for runoff prediction

The results 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 methods. The RS-based model calibration framework performs better at gridded scale than at lumped scale, which reflects the advantage of remote sensing in that it is spatially and temporally explicit across the global land surface. However, combining GRACE water storage data with the bias-corrected PML-AET only improves model performance marginally for regional calibrations (especially in low flow prediction), with little benefit in the gridded and catchment calibrations.

4.4. Spatial characteristics of optimum model calibration schemes

Figure 7 shows spatial patterns of *KGE*, *QR*, *NSE* and *LogNSE* from schemes 4 and 7. The 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 four metrics of monthly runoff are generally larger or marginally larger than the metrics of daily runoff. This is expected because of the impacts of precipitation seasonality enhancing the performance statistics (Zhang et al., 2020). Another spatial feature is that the *KGE* and *NSE* values for mainstream catchments are generally larger than those for tributary catchments. The *KGE* values of schemes 4 and 7 for Nike (05) catchment are negative, and the *NSE* values of schemes 4 and 7 for Nike (05) and Lugu (24) catchments are negative, while the *QR* values for them are positive. The values of *LogNSE* for schemes 4 and 7 vary generally from 0.2 to 1.0, but there are also extreme negative values. All in all, the spatial patterns of schemes 4 and 7 are similar and 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 streamflow at Daluo station when correcting bias of the PML-AET.

Figure 7 is about here

Figure 8a, 8b, 8d, 8e, 8g, 8h, 8j, 8k, 8m, 8n, 8p, 8q, 8s, 8t, 8v, 8w, 8x further show spatial patterns of performance of scheme 4 by calculating the difference, compared to scheme 1 and scheme 2. Figure 8c, 8i, 8o, 8u, 8f, 8l, 8r, 8x shows spatial patterns of performance of scheme 7 by calculating the difference, compared to scheme 4. The difference of each metric is calculated as follows:

$$\Delta M = M_{\mathrm{a}} - M_{b} \tag{11}$$

where M is one of the four metrics (KGE, QR, NSE and LogNSE), a and b refer to the proposed scheme and benchmark scheme, respectively. The blue dots in Figure 8 indicate positive differences in that catchment, the grey dots indicate no obvious differences, and the red one indicate negative differences. The darker the color is, the greater the difference is.

Figure 8 is about here

Figures 8a-8x-show the daily and monthly distribution of ΔM. There are three main patterns for daily simulations, obtained from Figures 8a-8c, 8g-8i, 8m-8o and 8s-8u. The first pattern is that there are 5 out of 30 catchments with positive differences of all the 4 metrics for scheme 4 minus scheme 1. The difference for 3 out of the 5 catchments is not larger than 0.02, indicating a reasonable result compared to scheme 1. The result shows that although scheme 4 performs poorer than scheme 1 in most catchments, it outperforms scheme 1 in a couple of catchments (2 out of 30) which shows the advantage of incorporating remote sensing data and gridded calibration, even compared to calibration against stream gauge data.

The second pattern is that in all 11 main stream stations, the ΔM for scheme 4 minus scheme 2 are positive with grey, light blue or dark blue dots in daily simulations, which means scheme 4 performs better than scheme 2 for daily runoff simulation, in upstream and large catchments which are also in the main stream (e.g. Ganzi catchment). There are 13 out of 30 catchments with positive differences in all 4 metrics for scheme 4 minus scheme 2. The difference for 2 of them is not larger than 0.02, indicating a reasonable result compared to scheme 2 in these catchments. However, scheme 4 outperforms scheme 2 for 37% of catchments for all 4 metrics. These catchments are generally downstream and small catchments, indicating that this approach may perform better than traditional regionalization in these catchments.

The third pattern is that the inclusion of GRACE data shows only a marginal or no improvement in most catchments, with positive differences of four metrics in only 4 out of 30 catchments, and the positive differences are not larger than 0.02 for all 4 catchments. In downstream catchments, the values of the difference are negative for *LogNSE*, indicating weakness on low flow modelling in these catchments. All in all, scheme 7 has limited improvement on the model performance, compared to scheme 4.

In monthly runoff simulation (Figure 8d-8f, 8j-8l, 8p-8r, 8v-8x), there are 14 out of 30 (about 47%) catchments for scheme 4 minus scheme 1, and 21 out of 30 (about 70%) for scheme 4 minus scheme 2 having positive ΔM values for all the four metrics. Scheme 7 performs similar to scheme 4 in 24% of catchments, where ΔM values for scheme 7 minus scheme 4 are larger than -0.02. Furthermore, there is no catchment, where ΔM values for all the 4 metrics are all negative.

In summary, in daily runoff simulations, scheme 4 performs similarly to scheme 1, and outperforms scheme 1 in 7% of catchments, indicating the advantage of quasi-runoff-free calibration method. Scheme 4 also performs better than scheme 2 in upper catchments and mainstream large catchments. Scheme 4 and scheme 7 show similar performance in most catchments. In monthly runoff simulations, the model performance of scheme 4 against schemes 1 and 2 improved in upper and main stream large catchments, compared to daily runoff simulations. Scheme 4 outperforms scheme 1 and scheme 2 in 47% and 80% of catchments, respectively. Overall, scheme 7 has limited benefit for improving model performance of scheme 4, and scheme 4 performs close to scheme 1, or better than scheme 1 in a few regions. Scheme 4 also performs better than scheme 2 in upper catchments and mainstream large catchments.

4.5. Relationship between statistical metrics and catchment attributes

Figure 9 summarizes the relationships between statistical metrics (at daily and monthly scales) obtained from scheme 4 and seven catchment characteristics. Probability of significant test is conducted for each of the relationships. Most characteristics have no significant relationships to the metrics ($p > 0.1$). Among the seven catchment characteristics, catchment areas has strong positive impacts on most of the metrics; five catchment characteristics, including area ($p < 0.001$), elevation (p

< 0.001), normalized difference vegetation index (NDVI) (0.01 < p < 0.05), mean annual precipitation (0.001 < p< 0.01) and mean annual temperature (0.001< p < 0.01), have good relationships with daily NSE; mean annual precipitation has the best relationship (0.05 < p < 0.1) to daily *LogNSE* and monthly *LogNSE*. The result indicates that in the study region, the quasi-runoff-free calibration method does show the strong influence of catchment area on model performance, which agrees to the results of section 4.4. It is noted that the sample number of the relationship analysis is only 30, relatively small. More large-scale researches need to be conducted for the significant test and relationship analysis.

Figure 9 is about here

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 for the limited number of gauges in the Yalong River basin in its upstream alpine regions. However, this region contributes to the majority of water resources for the Jinsha River, which is a major tributary of the Yangtze River (Kang et al., 2001; Yang et al., 2006). Therefore, it is important to improve 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 *KGE* and mean *NSE* of daily runoff of schemes 4-9 are generally not larger than that obtained from traditional regionalization (scheme 2), the mean *QR* and mean *LogNSE* are occasionally larger than traditional regionalization. Thus, 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 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. The bias correction is crucial in the study area as demonstrated by comparing 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). With a single value of mean annual runoff data in a downstream gauge, the PML-AET based quasi-runoff-free calibration has been shown to have the potential for large scale application. Furthermore, using a single parameter of α in the Fu model can generate reasonable mean annual runoff estimates for most of the 30 catchments, demonstrating the applicability of using a downstream catchment for bias correction. Overall, the bias correction method of PML-AET is reasonable with a reliable gridded product and limited surface data.

5.3. Advantage of gridded model calibration

The remote sensing data provides a spatial coverage, and it has the potential to reduce uncertainty related to lumped calibrations through better parameterization for each grid (Arnold et al., 2010; Li and Zhang, 2017). In this study, the gridded hydrological modelling results are considerably better than the lumped hydrological modelling results. The gridded calibration schemes outperform lumped calibration schemes in all the four metrics. It is noted that the run time increases by about 170-fold from lumped calibration to gridded calibration. Therefore, a more efficient algorithm is needed to reduce model run time in the future, and if necessary, a compromise should be made between model accuracy and time consumption for practical applications.

5.4. Adding GRACE data has very limited benefit to improve predictions

Though available studies show GRACE water storage data has been effectively applied at 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 of including GRACE data for model calibration is negligible for gridded and catchment calibrations. This could be caused by the fact that the total water volume has been already properly considered by the bias-corrected PML-AET. However, adding GRACE data improves the performance of low flow in regional calibrations. This might be the result of the similar spatial resolution between GRACE (1° x 1°) and the region area. Furthermore, the resolution of GRACE data is spatially (1° x 1°) and temporally (monthly) coarse. It is probably not appropriate to incorporate GRACE data into the small and medium sized catchments located on the Yalong River Basin with complex terrains and large ranges in elevations (Kang et al., 2001).

5.5. Limitations and further directions

This study does not consider snow cover for model calibration even 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, snow cover should be incorporated into the runoff simulation in the upper catchments (Kang et al., 2001). However, to do this, hydrological models need to be modified, making sure the modified structure has a 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 the relatively small influences of reservoir dispatching during these years.

However, with the running of hydropower stations (such as Ertan hydropower station and Jinpin hydropower station) and land use change in recent years, human activity has increased dramatically, especially in downstream catchments (Liu et al., 2017; Liu et al., 2019b). For runoff simulation and prediction after 2012 in the Yalong River basin, a human-activity based hydrological model with accurate remote sensing data is essential and benefits both hydrology and management (Montanari et al., 2013).

The calibration schemes can still be further improved. Incorporating GRACE data improves the model stability across the flow regime in 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 need to be further developed.

6. Conclusion

In this study, nine modelling schemes are applied and assessed for runoff prediction in the Yalong River basin, an ideal location for testing the potential benefit of using remote sensing data, because of its complex terrain and wide-ranging climate conditions. The PML-AET datasets are first evaluated and then bias corrected against water-balance AET estimated using the Fu equation calibrated against streamflow data from a single gauging station. The performance of calibration schemes using the bias-corrected PML-

AET data is much better than the performance with raw PML-AET data. The performance of gridded modelling is much better than lumped modelling, albeit with an increase in model run times. The calibration schemes incorporating GRACE data provide very limited benefit to gridded and catchment calibrations, but slightly improves the performance of low flow in regional calibrations. Using bias-corrected PML-AET to constrain a gridded hydrological model outperforms lumped regionalization hydrological modelling especially in monthly runoff simulation for upstream and large catchments.

This study demonstrates that the quasi-runoff-free hydrological model calibration against bias-corrected remotely sensed PML-AET data (using only one gauged streamflow station data to calibrate the Fu equation to estimate water balance AET) can reliably estimate daily and monthly runoff. The performance metrics of the simulated runoff are similar to or better than the runoff estimated using parameter values from the closest calibration catchment. This method is therefore particularly suited for estimating runoff in ungauged catchment and large regions, particularly sparsely gauged regions.

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 in China for providing the daily streamflow data for 30 gauging stations. The PML evapotranspiration and the Climate Meteorological Forcing Dataset used in this study are provided by National Tibetan Plateau Data Center (http://data.tpdc.ac.cn). The GRACE water storage data are freely available from Data Catalog of the Google Earth Engine (https://developers.google.com/earth-engine/datasets). Daily dataset of China's surface climate data is available from the China Meteorological Data Service Center (http://data.cma.cn/). We thank the anonymous reviewers and editors for their critical and 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 the 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

Andersen, O.B., Seneviratne, S.I., Hinderer, J., & Viterbo, P. (2005). GRACE-derived terrestrial water storage depletion associated with the 2003 European heat wave. *Geophysical Research Letters*, *32*(18). DOI:10.1029/2005gl023574

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.

Beck, H.E., van Dijk, A., Levizzani, V., Schellekens, J., Miralles, D.G., Martens, B. , et al. (2017a). MSWEP: 3-hourly 0.25 degrees global gridded precipitation (1979-2015) by merging gauge, satellite, and reanalysis data. *Hydrology and Earth System Sciences*, *21*(1), 589-615. DOI:10.5194/hess-21-589-2017

Beck, H.E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A., Weedon, G.P. , et al. (2017b). Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrology and Earth System Sciences*, *21*(12), 6201-6217. DOI:10.5194/hess-21-6201-2017

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.

Gan, R., Zhang, Y., Shi, H., Yang, Y., Eamus, D., Cheng, L., Chiew, F.H.S., Yu, Q., . (2018). Use of satellite leaf area index estimating evapotranspiration and gross assimilation for Australian ecosystems. *Ecohydrology*, *e1974*. DOI:https://doi.org/10.1002/eco.1974

Gupta, H.V., Kling, H., Yilmaz, K.K., & Martinez, G.F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1), 80–91. DOI:https://doi.org/10.1016/j.jhydrol.2009.08.003

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

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*(15), 1965-1978. DOI:10.1002/joc.1276

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.

Kittel, C.M.M., Nielsen, K., Tottrup, C., & Bauer-Gottwein, P. (2018). Informing a hydrological model of the Ogooue with multi-mission remote sensing data. *Hydrology and Earth System Sciences*, *22*(2), 1453-1472. DOI:10.5194/hess-22-1453-2018

Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, *424-425*, 264-277.

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

Kumar, B., & Lakshmi, V. (2018). Accessing the capability of TRMM 3B42 V7 to simulate streamflow during extreme rain events: Case study for a Himalayan River Basin. *Journal of Earth System Science*, *127*(2). DOI:10.1007/s12040-018-0928-1

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).

Leuning, R., Zhang, Y.Q., Rajaud, A., Cleugh, H., & Tu, K. (2008). A simple surface conductance model to estimate regional evaporation using MODIS leaf area index and the Penman-Monteith equation. *Water Resources Research*, *44*(10), 240-256.

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. (2009a). Predicting runoff in ungauged catchments by using Xinanjiang model with MODIS leaf area index. *Journal of Hydrology*, *370*(1-4), 155-162.

Li, L., Hong, Y., Wang, J.H., Adler, R.F., Policelli, F.S., Habib, S. , et al. (2009b). Evaluation of the real-time TRMM-based multi-satellite precipitation analysis for an operational flood prediction system in Nzoia Basin, Lake Victoria, Africa. *Natural Hazards*, *50*(1), 109-123. DOI:10.1007/s11069-008-9324-5

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, W.B., Wang, L., Zhou, J., Li, Y.Z., Sun, F.B., Fu, G.B. , et al. (2016). A worldwide evaluation of basin-scale evapotranspiration estimates against the water balance method. *Journal of Hydrology*, *538*, 82-95. DOI:10.1016/j.jhydrol.2016.04.006

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

Moore, R.J., & Clarke, R.T. (1981). A DISTRIBUTION FUNCTION-APPROACH TO RAINFALL RUNOFF MODELING. *Water Resources Research*, *17*(5), 1367-1382. DOI:10.1029/WR017i005p01367

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

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

Standards Press of China.(2008).Standard for Hydrological Information And Hydrological forecasting.Standardization Administration of the People's Republic of China.

Stewart, J.B., & Finch, J.W. (1993). APPLICATION OF REMOTE-SENSING TO FOREST HYDROLOGY. *Journal of Hydrology*, *150*(2-4), 701-716. DOI:10.1016/0022-1694(93)90132-s

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

Sun, Q.H., Miao, C.Y., Duan, Q.Y., Ashouri, H., Sorooshian, S., & Hsu, K.L. (2018). A Review of Global Precipitation Data Sets: Data Sources, Estimation, and Intercomparisons. *Reviews of Geophysics*, *56*(1), 79-107. DOI:10.1002/2017rg000574

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

Todini, E. (1996). The ARNO rainfall-runoff model. *Journal of Hydrology*, *175*(1-4), 339-382. DOI:10.1016/s0022-1694(96)80016-3

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

Wang, S.S., Pan, M., Mu, Q.Z., Shi, X.Y., Mao, J.F., Brummer, C. , et al. (2015). Comparing Evapotranspiration from Eddy Covariance Measurements, Water Budgets, Remote Sensing, and Land Surface Models over Canada. *Journal of Hydrometeorology*, *16*(4), 1540-1560. DOI:10.1175/jhm-d-14-0189.1

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.

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

Zhang, Y., Leuning, R., Hutley, L.B., Beringer, J., McHugh, I., & Walker, J.P. (2010). Using long-term water balances to parameterize surface conductances and calculate evaporation at 0.05° spatial resolution. *Water Resources Research*, *46*(5). DOI:10.1029/2009wr008716

Zhang, Y., Peña-Arancibia, J.L., McVicar, T.R., Chiew, F.H.S., Vaze, J., Liu, C., , Lu, X. , et al. (2016). Multi-decadal trends in global terrestrial evapotranspiration and its components. *6*(1), 19124.

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

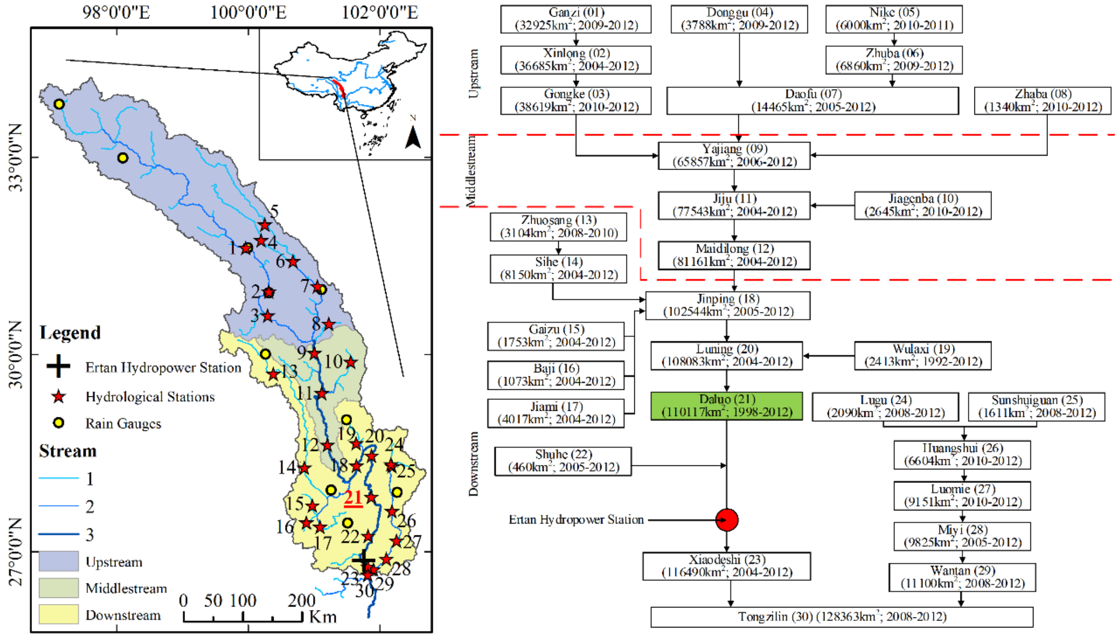


Figure 1. Information and location of study area. The station Daluo for constraining Fu model is labelled as 21.

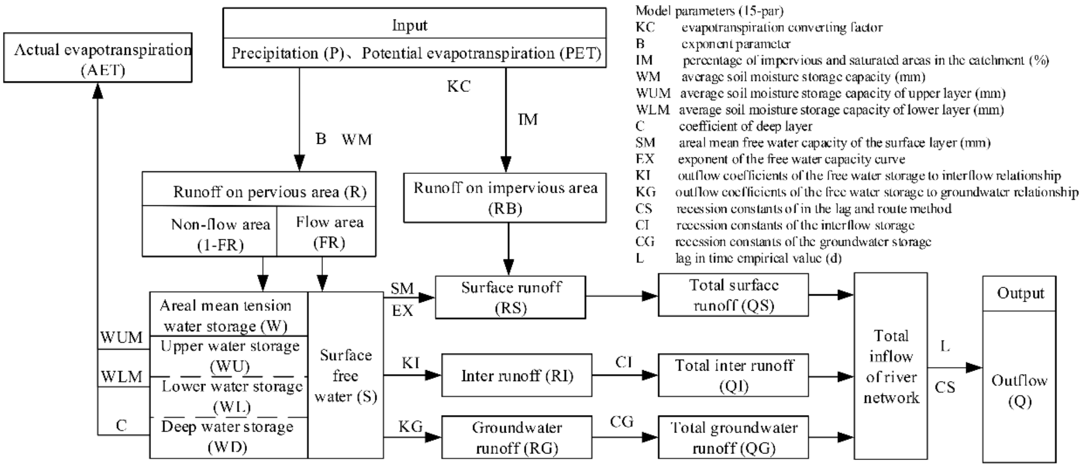


Figure 2. Model structure of Xinanjiang Model

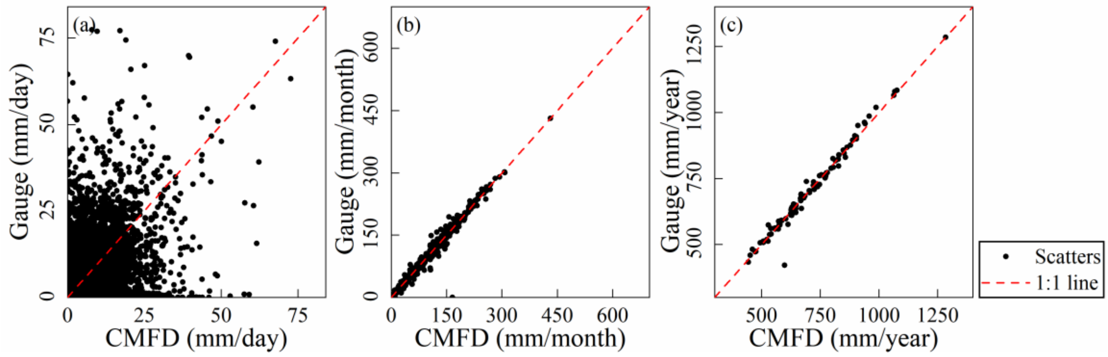


Figure 3. Comparison between observed precipitation and precipitation generated from CMFD data (CMFD-P)

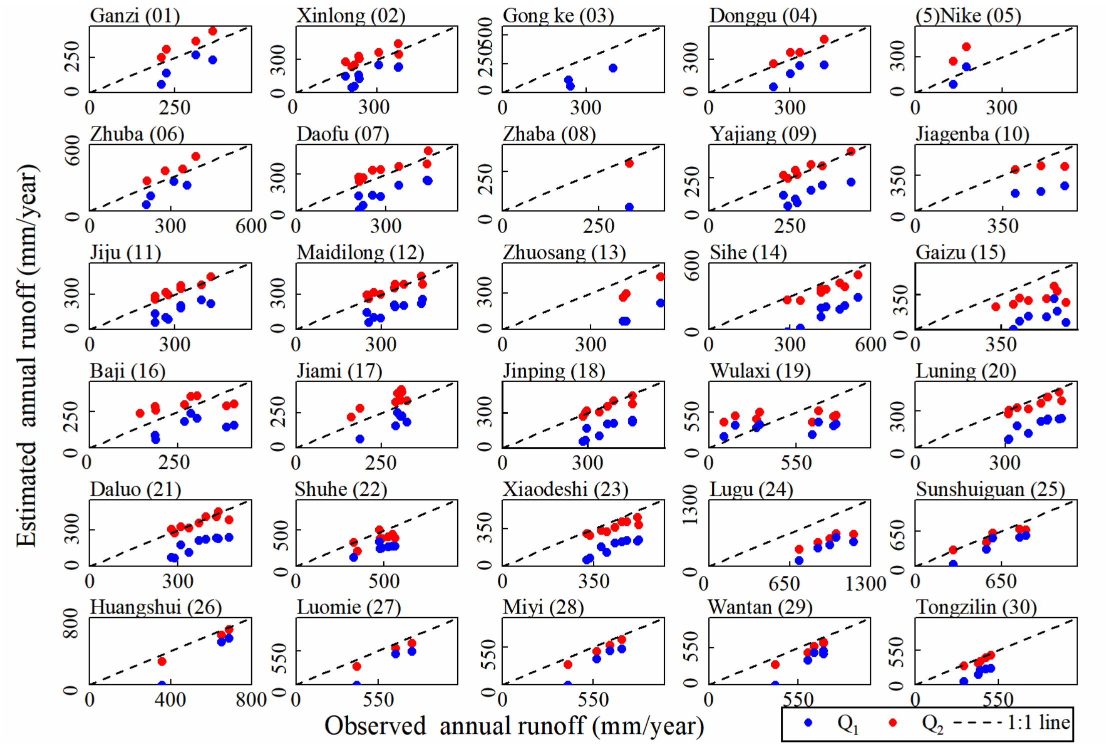


Figure 4. Evaluating annual runoff obtained from precipitation minus raw PML-AET (Q_1) and that (Q_2) obtained from precipitation minus bias-corrected PML-AET (The numbers in the bracket represent the watershed codes shown in Figure 1.)

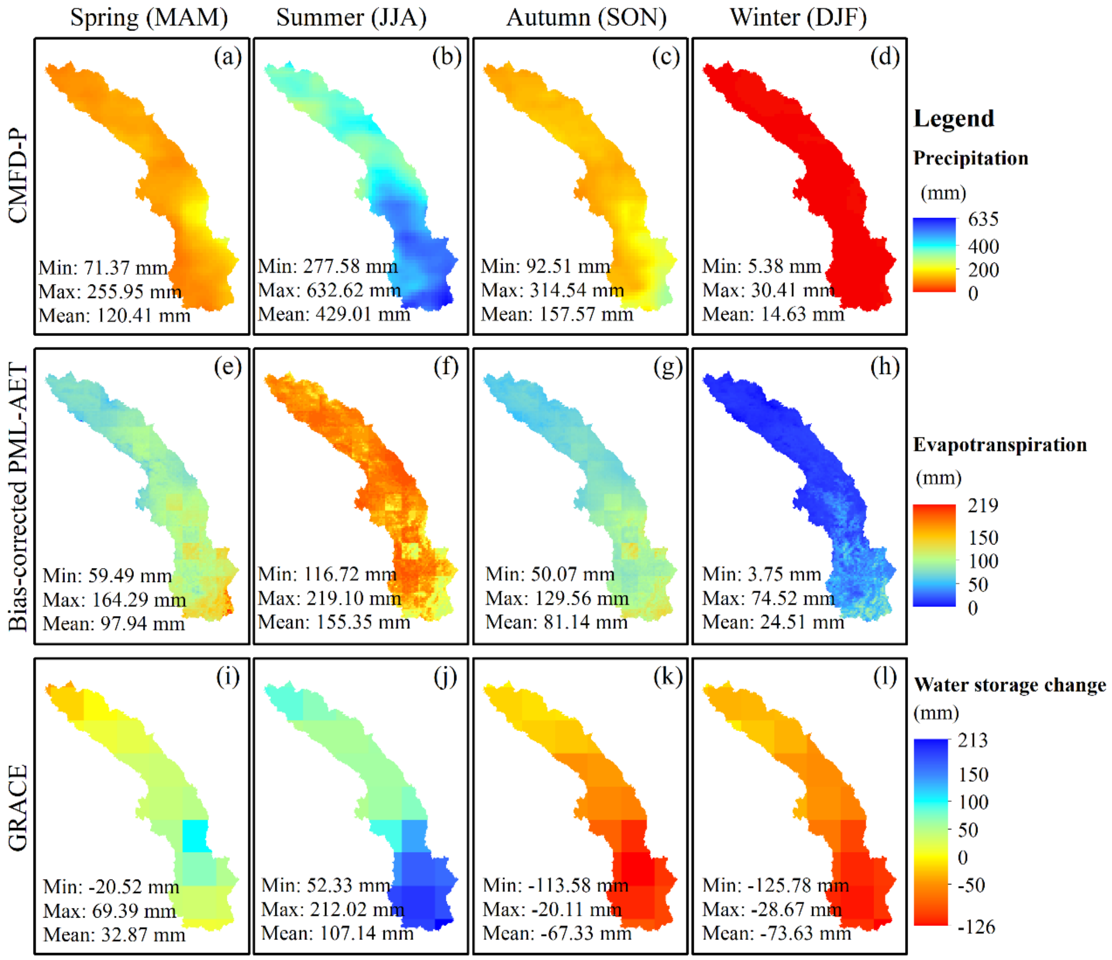


Figure 5. The mean annual spatial and seasonal distributions of CMFD-P, bias-corrected PML-AET and GRACE soil water storage change data

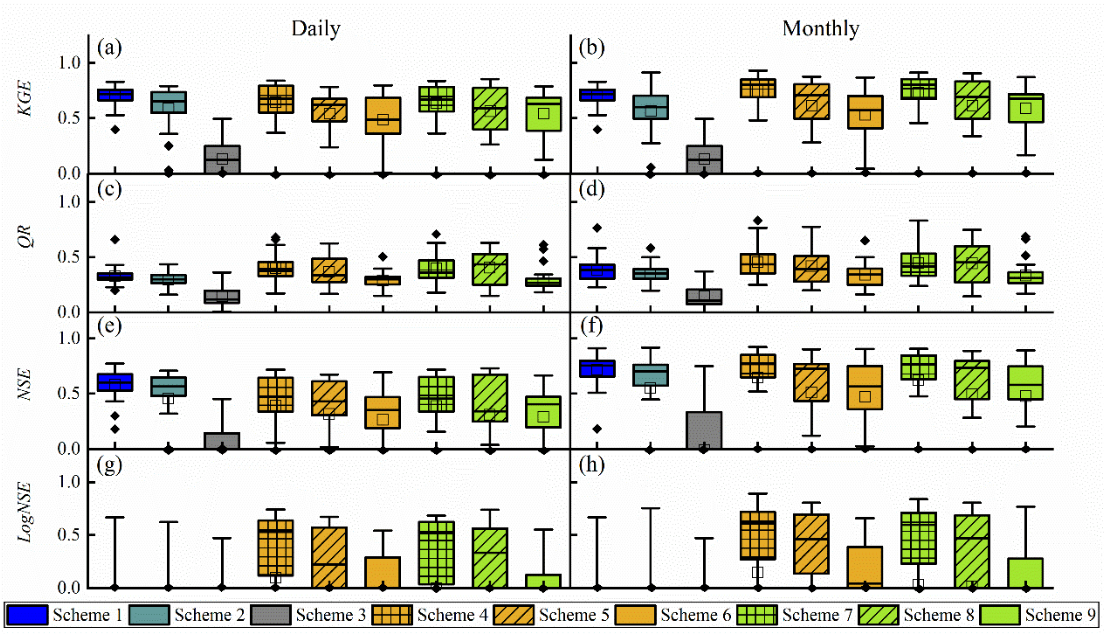


Figure 6. Comparison of performance of the nine calibration schemes for estimating streamflow. Noted that negative values are plotted as zero for better visualization (The boxes represent the values range from 25th to 75th percentiles, the lines in each plot from top to bottom represent upper boundary, median value and lower boundary, respectively. The square represents the mean value and the rhombus represents the outlier.)

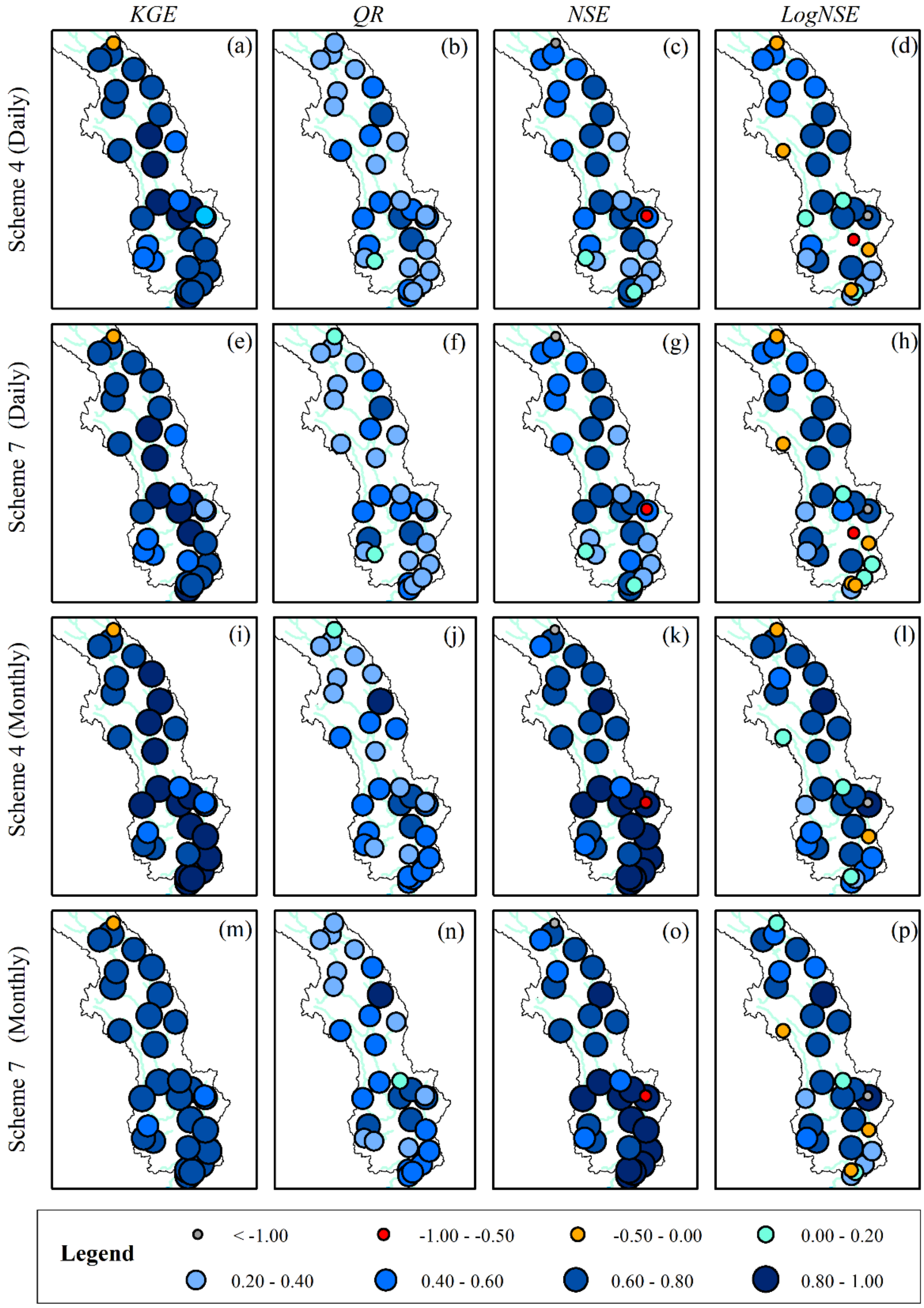


Figure 7. spatial patterns of metrics obtained from scheme 4 and scheme 7

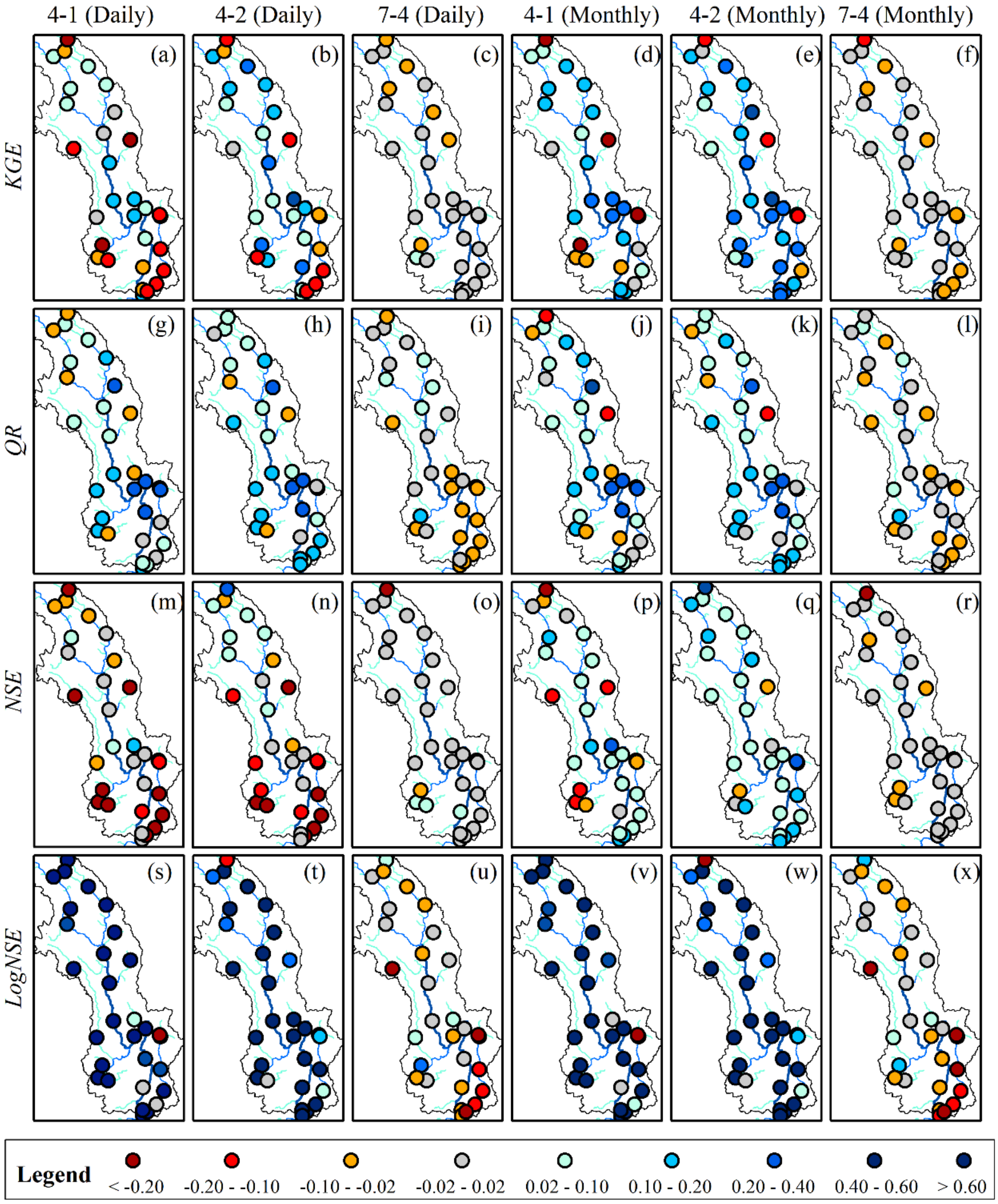


Figure 8. Spatial evaluation of scheme 4 against scheme 1, 2 and 7. The difference among them is obtained from Eq. (11). Having a range from -0.02 to 0.02, gray means the two perform similarly.

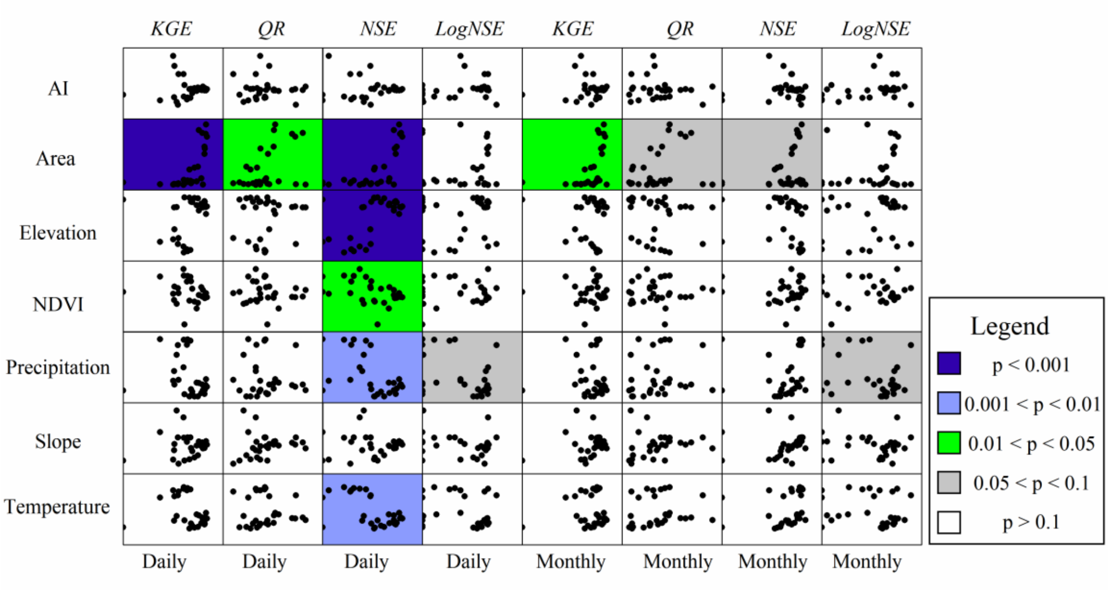
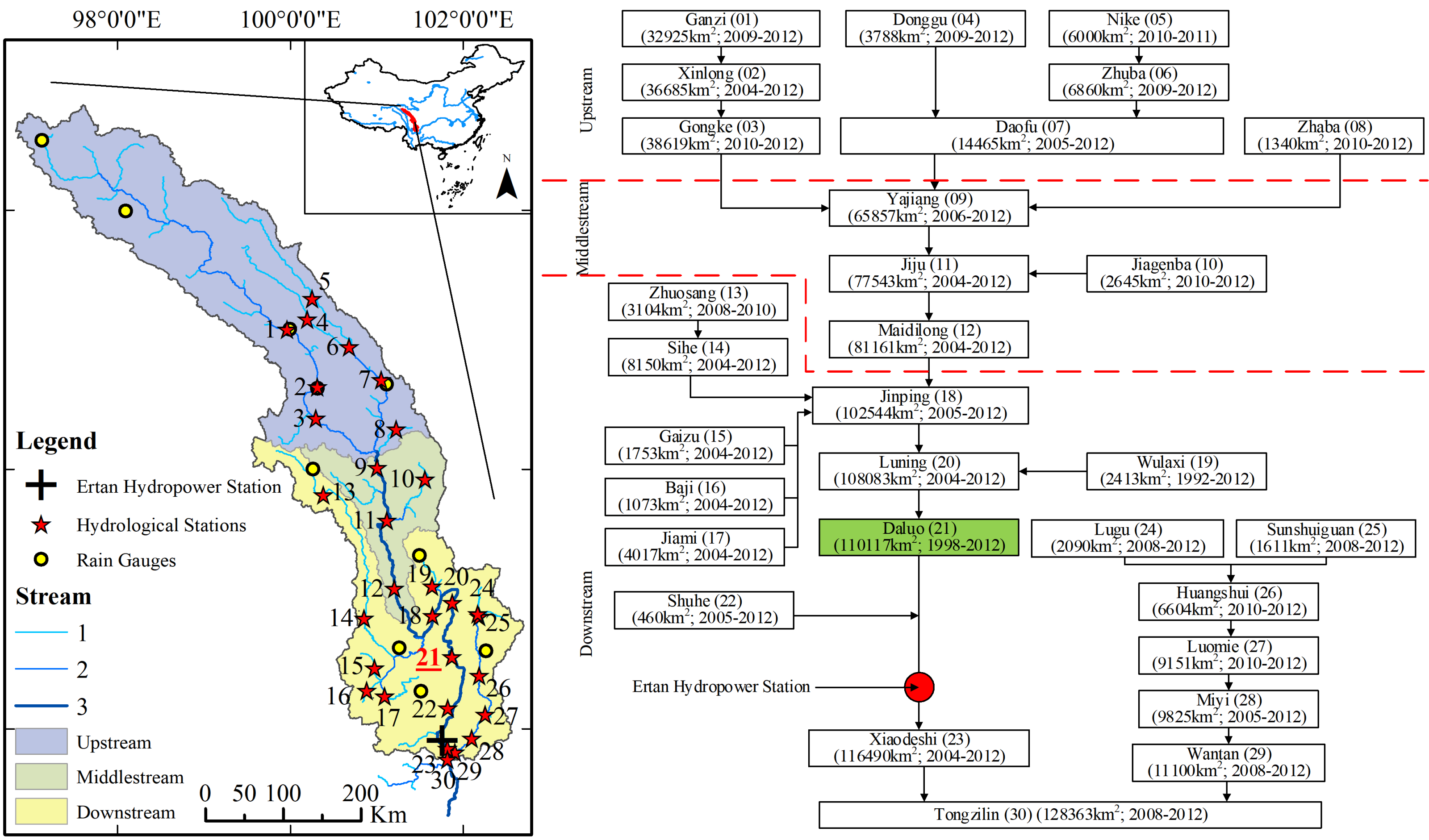


Figure 9. Scatterplots between metrics (obtained from scheme 4) and catchment characteristics (Each point represents one catchment. Negative values are set as zero to minimize the weight of negative metric values on significant test; p is the probability of significant test; AI is aridity index for each catchment; NDVI is mean annual normalized difference vegetation index.)

Figure 1.



98°0'0"E
100°0'0"E
102°0'0"E
33°0'0"N
30°0'0"N
27°0'0"N
N
Legend
Ertan Hydropower Station
Hydrological Stations
Rain Gauges
Stream
1
2
3
Upstream
Middlestream
Downstream
0 50 100 200
Km
21
Upstream
Middlestream
Downstream
Ganzi (01) (32925km2; 2009-2012)
Donggu (04) (3788km2; 2009-2012)
Nike (05) (6000km2; 2010-2011)
Xinlong (02) (36685km2; 2004-2012)
Zhuba (06) (6860km2; 2009-2012)
Gongke (03) (38619km2; 2010-2012)
Daofu (07) (14465km2; 2005-2012)
Zhaba (08) (1340km2; 2010-2012)
Yajiang (09) (65857km2; 2006-2012)
Jiju (11) (77543km2; 2004-2012)
Jiagenba (10) (2645km2; 2010-2012)
Zhuosang (13) (3104km2; 2008-2010)
Sihe (14) (8150km2; 2004-2012)
Maidilong (12) (81161km2; 2004-2012)
Jinping (18) (102544km2; 2005-2012)
Gaizu (15) (1753km2; 2004-2012)
Baji (16) (1073km2; 2004-2012)
Jiami (17) (4017km2; 2004-2012)
Luning (20) (108083km2; 2004-2012)
Wulaxi (19) (2413km2; 1992-2012)
Daluo (21) (110117km2; 1998-2012)
Lugu (24) (2090km2; 2008-2012)
Sunshuiguan (25) (1611km2; 2008-2012)
Huangshui (26) (6604km2; 2010-2012)
Shuhe (22) (460km2; 2005-2012)
Luomie (27) (9151km2; 2010-2012)
Ertan Hydropower Station
Miyi (28) (9825km2; 2005-2012)
Xiaodeshi (23) (116490km2; 2004-2012)
Wantan (29) (11100km2; 2008-2012)
Tongzilin (30) (128363km2; 2008-2012)

Figure 2.

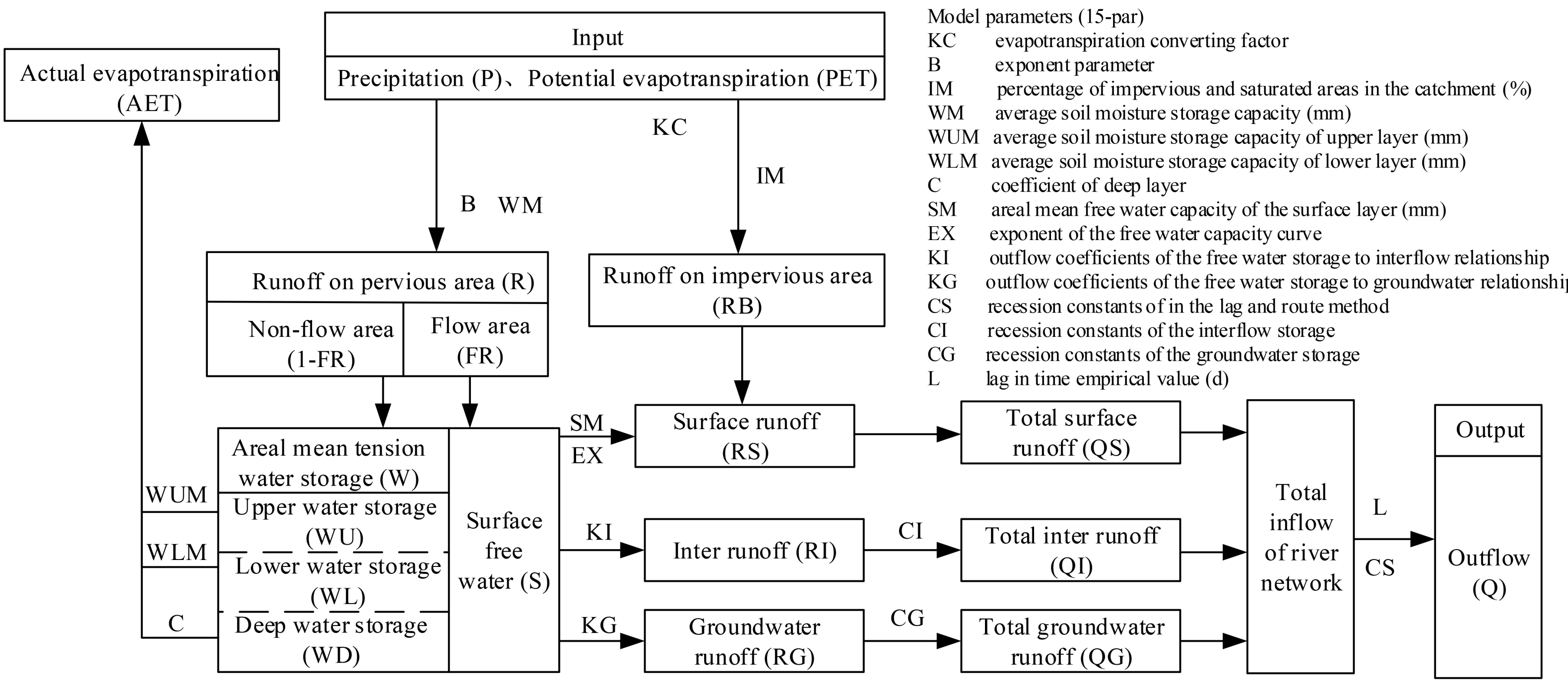
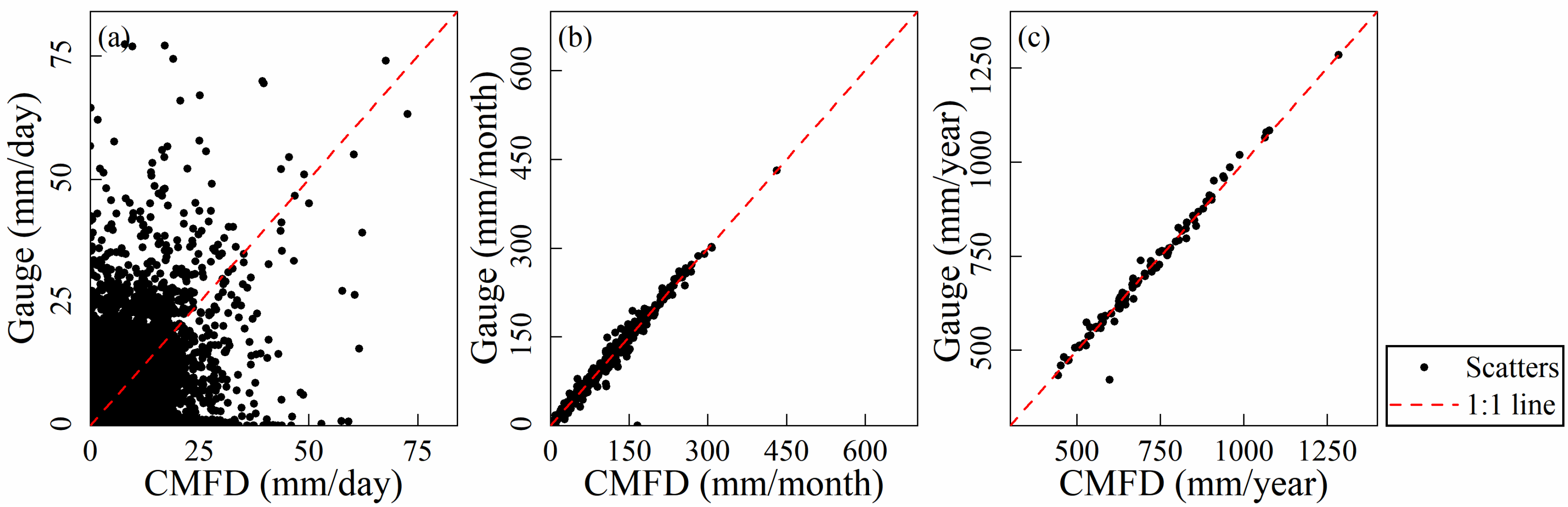
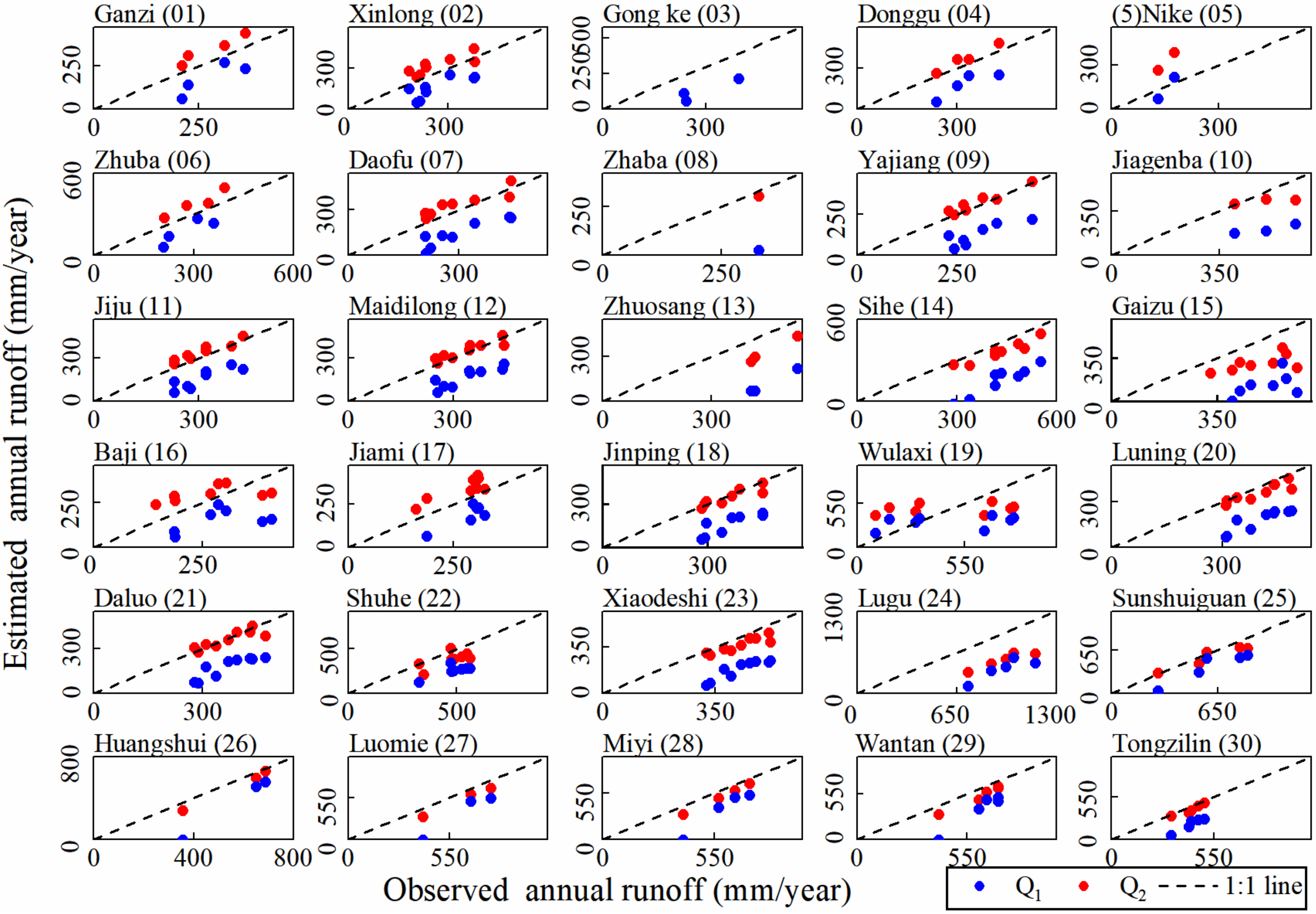

Model parameters (15-par)
KC evapotranspiration converting factor
B exponent parameter
IM percentage of impervious and saturated areas in the catchment (%)
WM average soil moisture storage capacity (mm)
WUM average soil moisture storage capacity of upper layer (mm)
WLM average soil moisture storage capacity of lower layer (mm)
C coefficient of deep layer
SM areal mean free water capacity of the surface layer (mm)
EX exponent of the free water capacity curve
KI outflow coefficients of the free water storage to interflow relationship
KG outflow coefficients of the free water storage to groundwater relationship
CS recession constants of in the lag and route method
CI recession constants of the interflow storage
CG recession constants of the groundwater storage
L lag in time empirical value (d)
Input
Precipitation (P)、Potential evapotranspiration (PET)
Actual evapotranspiration (AET)
KC
IM
B WM
Runoff on pervious area (R)
Non-flow area (1-FR)
Flow area (FR)
Runoff on impervious area (RB)
Areal mean tension water storage (W)
Upper water storage (WU)
Lower water storage (WL)
Deep water storage (WD)
WUM
WLM
C
Surface free water (S)
SM
EX
KI
KG
Surface runoff (RS)
Inter runoff (RI)
Groundwater runoff (RG)
CI
CG
Total surface runoff (QS)
Total inter runoff (QI)
Total groundwater runoff (QG)
Total inflow of river network
L
CS
Output
Outflow (Q)

Figure 3.



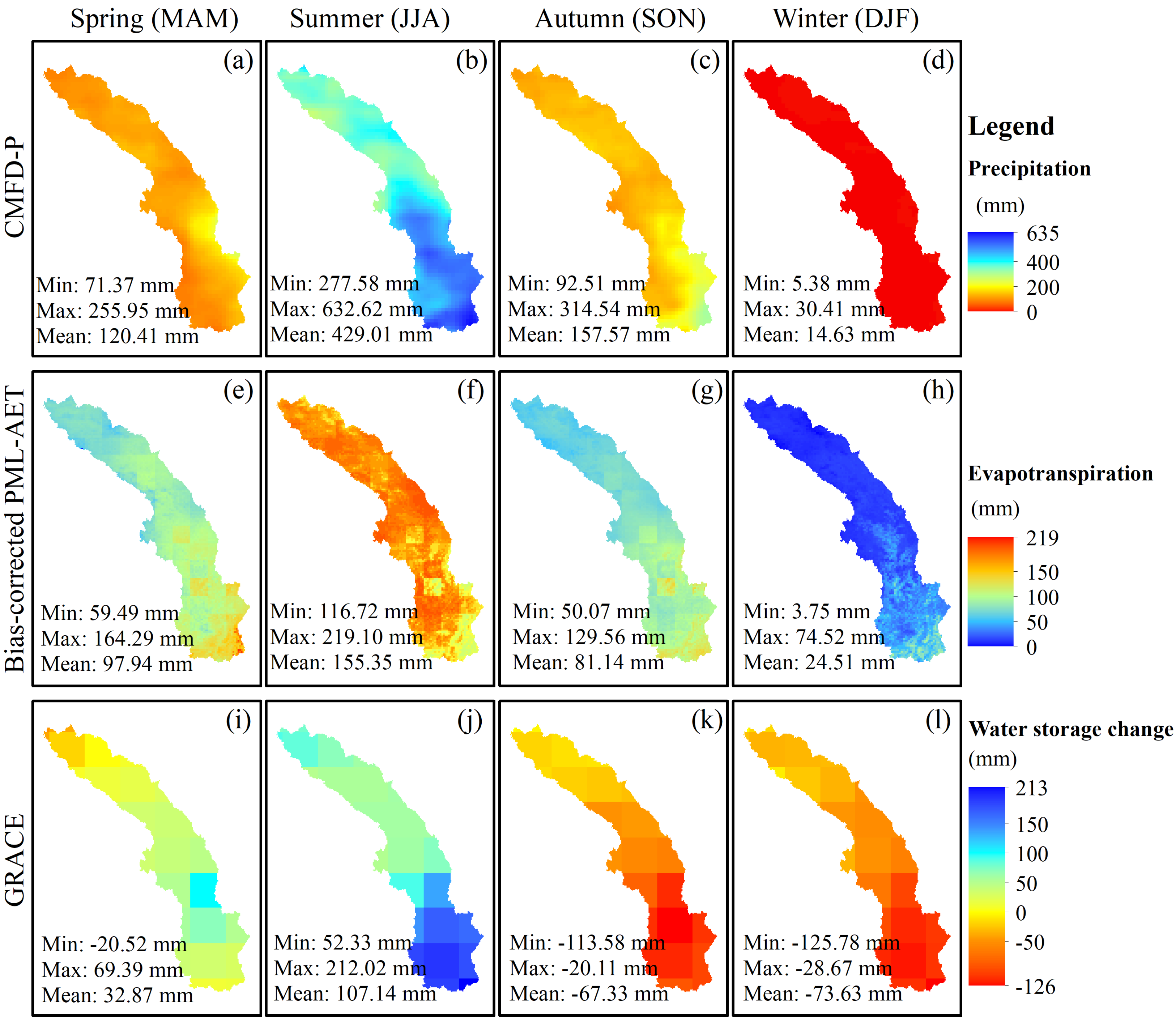
(a)
Gauge (mm/day)
CMFD (mm/day)
0
25
50
75
(b)
Gauge (mm/month)
CMFD (mm/month)
0
150
300
450
600
(c)
Gauge (mm/year)
CMFD (mm/year)
500
750
1000
1250
Scatters
1:1 line

Figure 4.



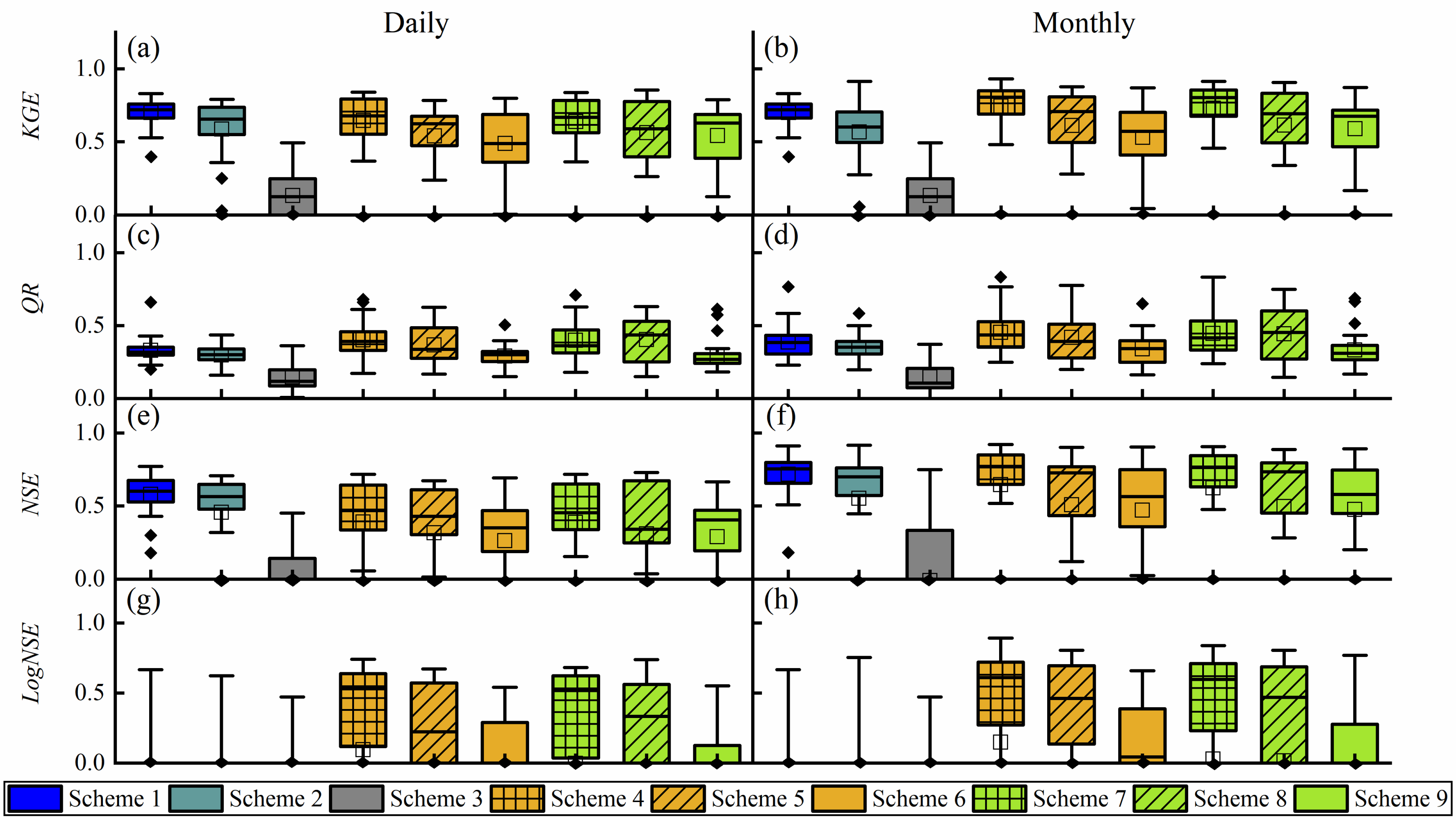
Ganzi (01)
Xinlong (02)
Gong ke (03)
Donggu (04)
(5)Nike (05)
Zhuba (06)
Daofu (07)
Zhaba (08)
Yajiang (09)
Jiagenba (10)
Jiju (11)
Maidilong (12)
Zhuosang (13)
Sihe (14)
Gaizu (15)
Baji (16)
Jiami (17)
Jinping (18)
Wulaxi (19)
Luning (20)
Daluo (21)
Shuhe (22)
Xiaodeshi (23)
Lugu (24)
Sunshuiguan (25)
Huangshui (26)
Luomie (27)
Miyi (28)
Wantan (29)
Tongzilin (30)
Estimated annual runoff (mm/year)
Observed annual runoff (mm/year)
Q_1
Q_2
1:1 line

Figure 5.



Spring (MAM)
Summer (JJA)
Autumn (SON)
Winter (DJF)
CMFD-P
Bias-corrected PML-AET
GRACE
(a)
Min: 71.37 mm
Max: 255.95 mm
Mean: 120.41 mm
(b)
Min: 277.58 mm
Max: 632.62 mm
Mean: 429.01 mm
(c)
Min: 92.51 mm
Max: 314.54 mm
Mean: 157.57 mm
(d)
Min: 5.38 mm
Max: 30.41 mm
Mean: 14.63 mm
(e)
Min: 59.49 mm
Max: 164.29 mm
Mean: 97.94 mm
(f)
Min: 116.72 mm
Max: 219.10 mm
Mean: 155.35 mm
(g)
Min: 50.07 mm
Max: 129.56 mm
Mean: 81.14 mm
(h)
Min: 3.75 mm
Max: 74.52 mm
Mean: 24.51 mm
(i)
Min: -20.52 mm
Max: 69.39 mm
Mean: 32.87 mm
(j)
Min: 52.33 mm
Max: 212.02 mm
Mean: 107.14 mm
(k)
Min: -113.58 mm
Max: -20.11 mm
Mean: -67.33 mm
(l)
Min: -125.78 mm
Max: -28.67 mm
Mean: -73.63 mm
Legend
Precipitation
(mm)
635
400
200
0
Evapotranspiration
(mm)
219
150
100
50
0
Water storage change
(mm)
213
150
100
50
0
-50
-126

Figure 6.



Daily
Monthly
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
KGE
QR
NSE
LogNSE
1.0
0.5
0.0
Scheme 1
Scheme 2
Scheme 3
Scheme 4
Scheme 5
Scheme 6
Scheme 7
Scheme 8
Scheme 9

Figure 7.

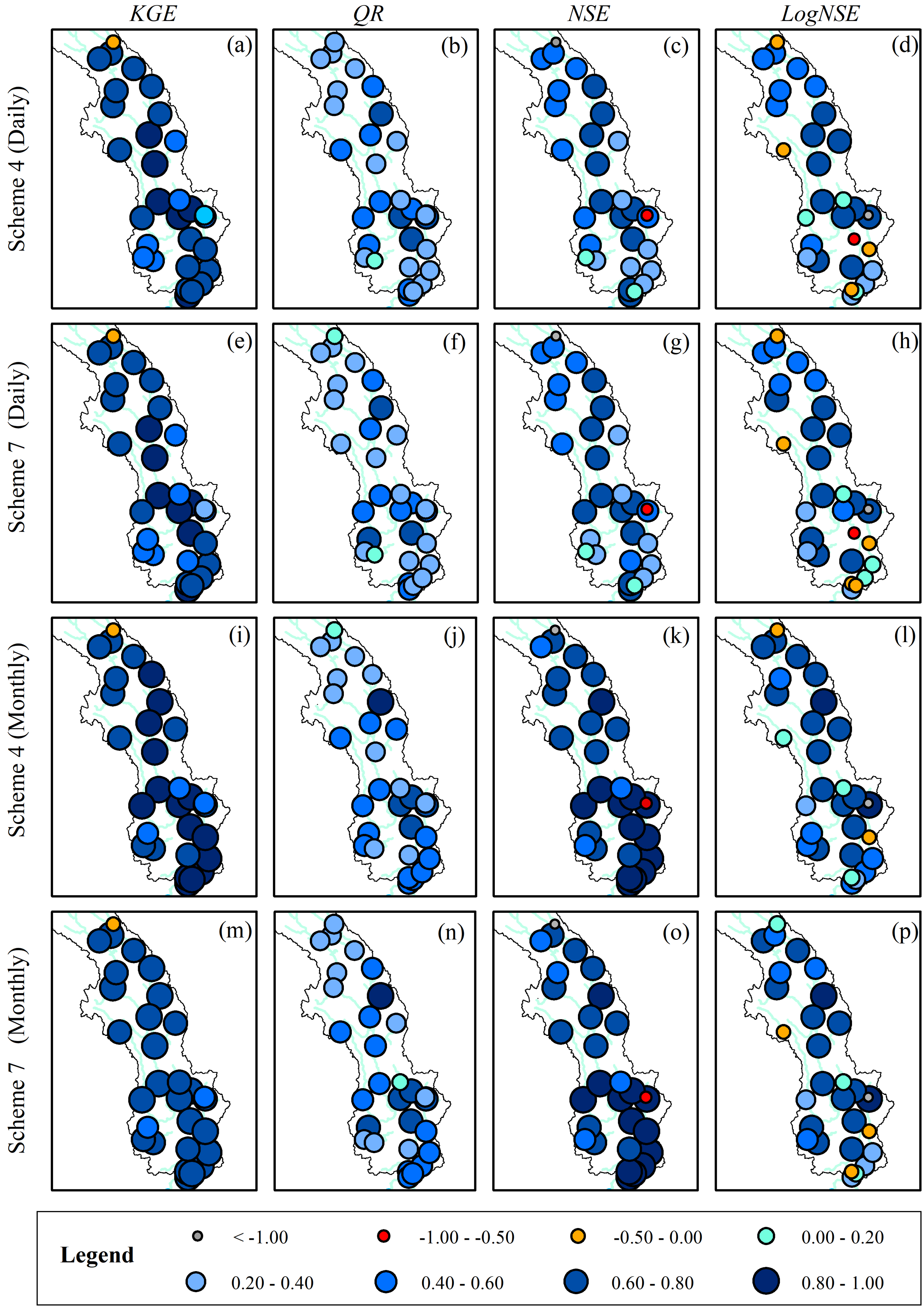
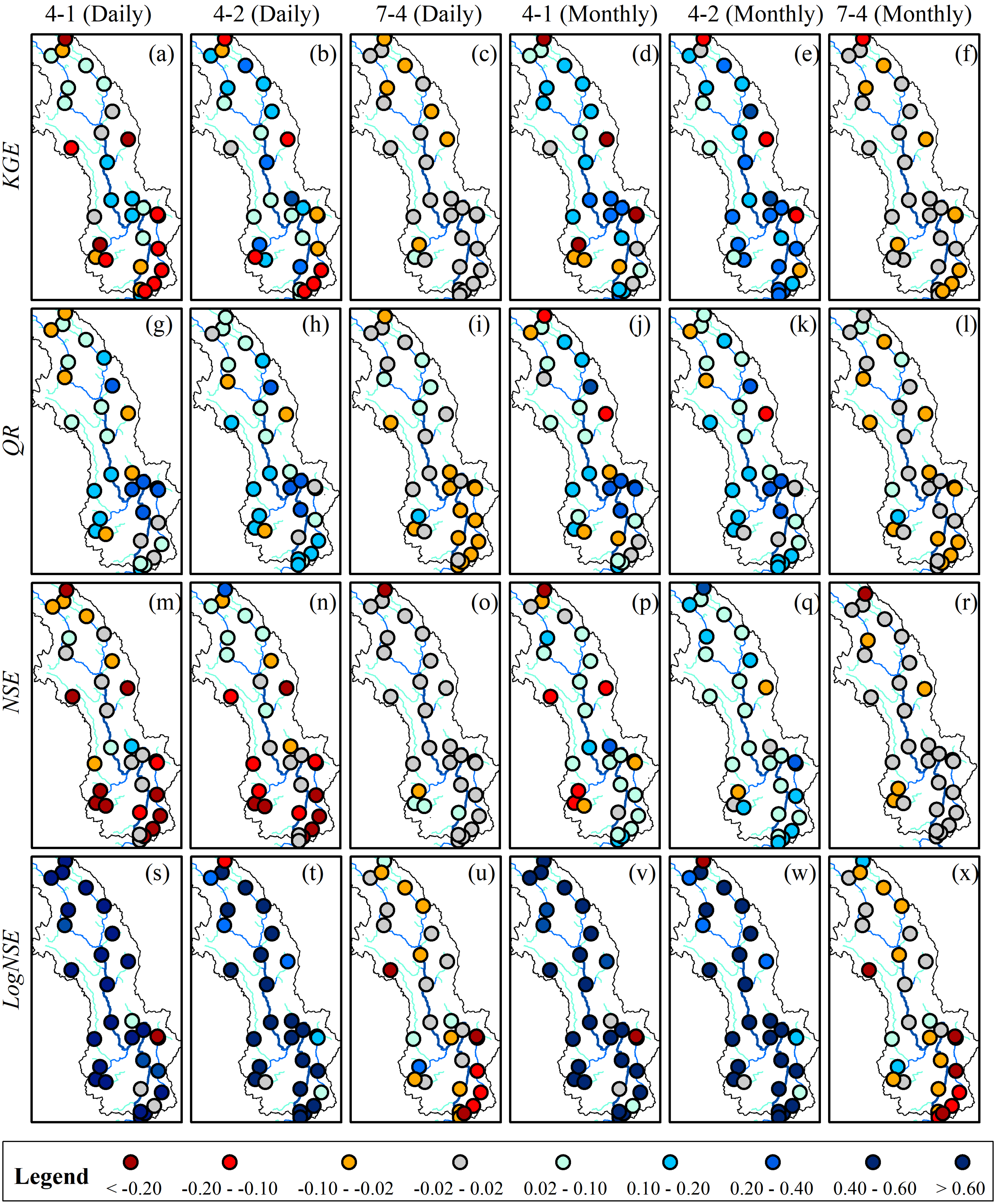
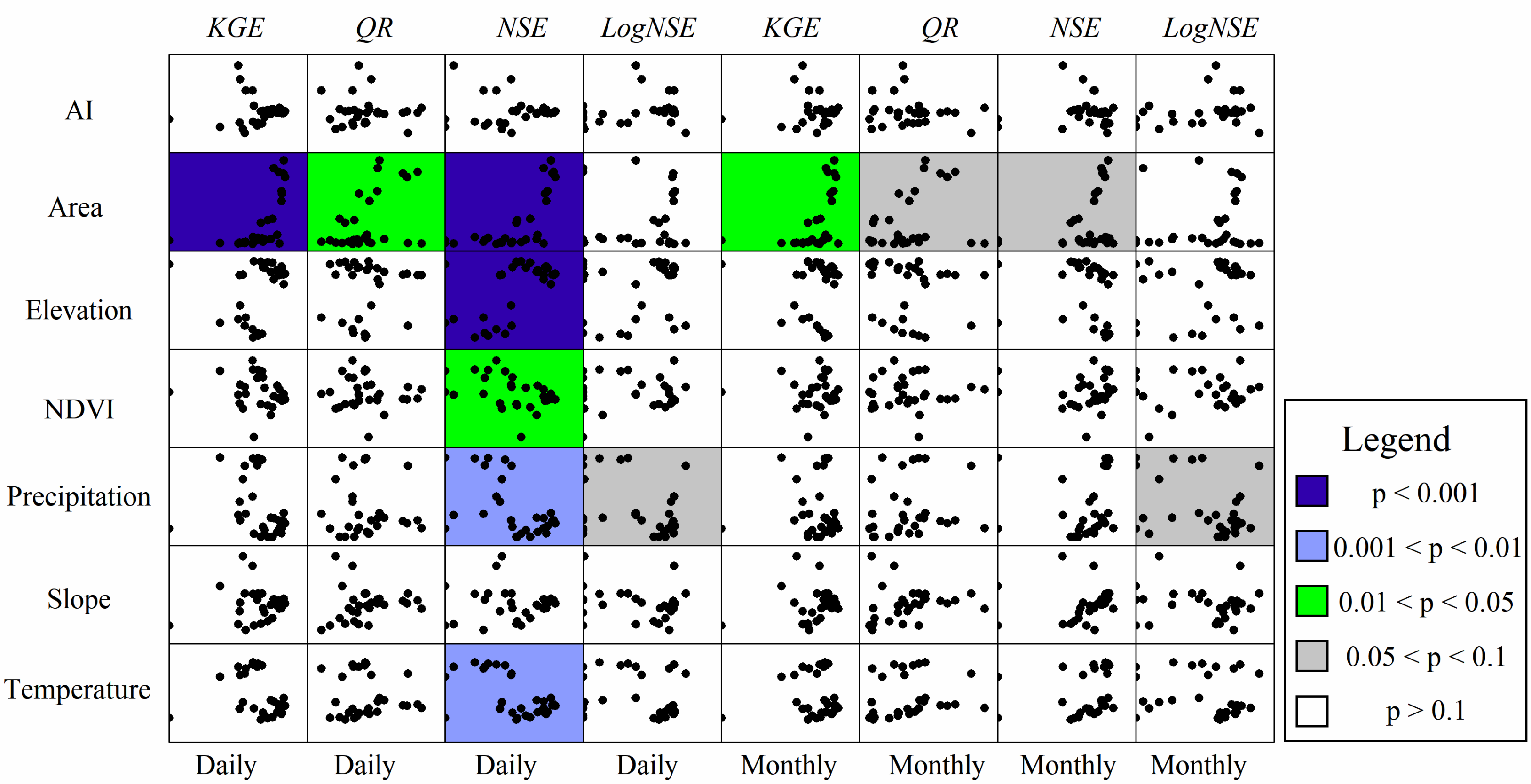

KGE
QR
NSE
LogNSE
Scheme 4 (Daily)
Scheme 7 (Daily)
Scheme 4 (Monthly)
Scheme 7 (Monthly)
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)
(m)
(n)
(o)
(p)
Legend
< -1.00
-1.00 - -0.50
-0.50 - 0.00
0.00 - 0.20
0.20 - 0.40
0.40 - 0.60
0.60 - 0.80
0.80 - 1.00

Figure 8.



4-1 (Daily)
4-2 (Daily)
7-4 (Daily)
4-1 (Monthly)
4-2 (Monthly)
7-4 (Monthly)
KGE
QR
NSE
LogNSE
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)
(m)
(n)
(o)
(p)
(q)
(r)
(s)
(t)
(u)
(v)
(w)
(x)
Legend
< -0.20
-0.20 - -0.10
-0.10 - -0.02
-0.02 - 0.02
0.02 - 0.10
0.10 - 0.20
0.20 - 0.40
0.40 - 0.60
> 0.60

Figure 9.


KGE
QR
NSE
LogNSE
KGE
QR
NSE
LogNSE
AI
Area
Elevation
NDVI
Precipitation
Slope
Temperature
Daily
Daily
Daily
Daily
Monthly
Monthly
Monthly
Monthly
Legend
p < 0.001
0.001 < p < 0.01
0.01 < p < 0.05
0.05 < p < 0.1
p > 0.1