IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 28, 2018, accepted February 21, 2018, date of publication February 28, 2018, date of current version March 13, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2810252

Hydrological Analysis Using Satellite Remote Sensing Big Data and CREST Model

JUN MA[1], WEIWEI SUN[1,2], (Member, IEEE), GANG YANG[1], AND DIANFA ZHANG[1]
[1]Department of Geography and Spatial Information Techniques, Ningbo University, Ningbo 315211, China
[2]State Key Laboratory of Information Engineering on Survey, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China

Corresponding author: Weiwei Sun (sunweiwei@nbu.edu.cn)

This work was supported in part by the Public Projects of Zhejiang Province under Grant 2016C33021, in part by the National Natural Science Foundation under Grant 41671342 and Grant U1609203, in part by the Chinese Post-Doctoral Science Foundation under Grant 2015M570668 and Grant2016T90732, in part by the Ningbo Natural Science Foundation under Grant 2017A610294, and in part by the K. C. Wong Magna Fund from Ningbo University.

ABSTRACT Hydrological modeling significantly contributes to the understanding of catchment water balance and water resource management and mitigates negative impacts of flooding. Considering the advantages of satellite remote sensing big data and the coupled routing and excess storage (CREST) model, this paper investigates the hydrological modeling in the Shehong basin during 2006–2013. The results show that humid Shehong basin has main rainfalls in summer (From May to September). For the monthly average rainfall and streamflow, there is a remarkable increase (+52%) in discharge and a smaller increase (+18%) in rainfall in the second period (2010–2013) relative to the first period (2006–2009). The CREST model was calibrated using China gauge-based daily precipitation analysis for the period of 2006–2009, followed by a favorable performance with Nash-Sutcliffe coefficient efficiency (NSCE) of 0.77, correlation coefficient (CC) up to 0.88 and −11% Bias. The model validation shows an error metric with NSCE of 0.74, CC of 0.87 and −11.7% Bias. In terms of water balance modeling results at Shehong basin, the runoff and rainfall estimates from CREST model coincide well with the gauge observations, indicating the model captures the appropriate signature of soil moisture variability. Therefore, the satellite-based precipitation product is feasible in hydrological prediction, and the CREST models the interaction between surface and subsurface water flow process in the Shehong basin.

INDEX TERMS Satellite remote sensing big data, hydrological analysis, TRMM, CREST, water balance.

I. INTRODUCTION

Floods are among the most common natural hazards which have caused great casualties and property losses over the past decades [1]. Hydrological modelling greatly contributes to flooding prediction and provides insights into the relationship between climate factors and hydrological cycle operating over a range of spatial and temporal scales [2]. Therefore, it is of great significance to model the hydrological features to predict the flooding events and to reduce the property losses.

Effective hydrological modelling requires a deep understanding of catchment water balance. The reason for that comes from two aspects. First, the catchment water balance impacts on the water yield of catchments [3]. Second, the climatic variability leads to conflicts that correlate with freshwater supply [4]. Hydrological modelling bridges the gap between climatic factors and catchment water balance, and they can be generally classified into two types: conceptual hydrological models and distributed physically-based models. Conceptual models deal with hydrological processes based on some simple mathematical formulas without explicit treatment of spatial heterogeneity in underlying surface [5]. Typical models are Storm Water Management Model (SWMM), Xin'an Jiang Model and Hydrologiska Byrans Vattenbalansavdelning (HBV) [6]. The distributed physically-based models represent each component of the hydrological cycle using their physical governing equations, with typical examples of Soil and Water Assessment Tool (SWAT), Variable Infiltration Capacity (VIC) and the Coupled Routing and Excess Storage (CREST) model [7]. Compared to conceptual hydrological models, the distributed models get more popularity and are more widely used in realistic applications. The distributed models simulate the spatio-temporal variations of hydrological cycle components on distributed grid cells and account for the intra-basin

variability of runoff-producing mechanisms [8]. Unfortunately, distributed hydrological models rely on sufficient reliable geospatial data as the model input. That adversely hinders their potential applications in ungauged or poorly gauged regions. In contrast, the CREST has good integration capability and can be well compatible with satellite remote sensing big data, making it advantageous to be used in ungauged or poorly gauged regions.

Another key point of hydrological modelling is to accurately estimate the precipitations in the region. Generally, three main approaches are available to measure precipitation, rain gauges, weather radars and satellite-based sensors [9]. Compared with conventional rain gauges and weather radars, satellites remote sensing is a more advantageous approach. Advances in remote sensing data with quasi-global coverage, continuity in time and easy access make it a cost-effective source of estimating precipitations for hydro-climatology studies in regions with sparse ground based networks [10]. One representative of satellite remote sensing big data product is the Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA) product, which is intended to be combined with passive microwave (PMW) and infrared (IR) observations to provide the ''best'' estimate of quasi-global precipitation [11]–[14]. TRMM has been becoming one of the most important and widely-used precipitation products of satellite remote sensing big data and provided tremendous information for hydrological applications [15]–[18].

The Shehong basin, located in Sichuan province of China, is hailed in the root of the Yangtze River Economic Belt and the highland of economic development in Western China. The combination of East Asian monsoon, complicated hydrological conditions and topography make the region a typical flood-vulnerable area of China [19]–[21]. Severe precipitations and the resulting frequent floods led to a formidable loss of life and property, making it urgent and critical to carry out a research on hydrological modeling in the region.

The purpose of the paper is to make a suitable hydrological analysis in the Shehong basin with satellite remote sensing big data and the CREST model. First, we investigate the hydro-climatology of Shehong basin at a range of time scales (e.g. daily, monthly and annual mean) based on observational data. After that, the CREST model is implemented to model the main processes of water balance. Finally, the hydrological capability of satellite remote sensing big data is explored with the reconstruction results from CREST model. The results shed light on subsequent hydrological investigations in the basins situated in subtropics.

The rest of this paper is organized as follows. Section II describes the study area and datasets. Sections III and IV introduces the CREST model, model calibration method and statistical metrics. Section V presents the information on model set-up and model validation. Section VI shows the reconstruction results of CREST model, and Section VII draws the conclusions.

II. STUDY AREA AND DATASET

A. STUDY AREA

The Shehong basin is located in southeast China, spanning from 30°50' N to 33°24' N in latitude and from 103°45' E to 105°38' E in longitude. The Shehong basin is a significant crop producing region of grain and wheat in Sichuan province and is also a typical flood-prone region causing by the combination of subtropical monsoon and sloping topography [22]. It includes the largest tributary of the Jialing River, with a drainage area of 24752km^2 and a length of 570km above Shehong hydrological station (Fig. 1). The tributary originates from the main peak of MinShan hills between Songpan County and Pingwu County in Sichuan province and drains into the Jialing river, which is the largest branch of the Yangtze River. The basin elevation ranges from 300m in the downstream plain areas to 5000m in the upstream mountainous areas and decreases rapidly from west to east. The land use is dominated by cultivated land (43%), woodland (36%) and grassland (18%). Referring to the China soil classification system, soils in the region are mainly composed of yellow soil (14%), yellow-brown soil (16%), purple soil (30%) and paddy soil (14%) [23], [24]. Yellow and yellow-brown soils are mostly distributed in the upper and middle reaches of the basin, whereas most purple and paddy soil exist in the downstream of the basin.

Due to alternate effects from ocean tropical air and continental polar air, the Shehong basin has the climate of subtropical humid monsoon with an average annual rainfall of ∼1200 mm that spreads over two rainfall peaks in summer. Extreme precipitation easily causes frequent floods in the region, and that poses a serious threat to personal property and national safety. In 2013, severe floods occurred in the Shehong basin and caused the death of 9 persons and a loss of ∼50 billion RMB (available at http://www.mca.gov.cn/). Therefore, it is significant to conduct a hydrological study across the Shehong basin to address the issues on water resources management and flood prediction [20], [25], [26].

B. IN-SITU AND SATELLITE DATASETS

Table 1 lists four kinds of datasets that are implemented in our study, including the digital elevation model (DEM), the land use, the discharge and the soil data. The DEM data is with 30 arc-second spatial resolution and was obtained from Hydrological data and maps based on Shuttle Elevation Derivatives at multiple Scales (HydroSHEDS) (available at https://hydrosheds.cr.usgs.gov/). The land use data was obtained from LANDSAT-7 remote sensing imagery via visual interpretation with 6 categories and 25 sub-categories. Both soils and land use datasets were obtained from Data Center for Resources and Environmental Sciences (RESDC), Chinese Academy of Sciences (available at http://www.resdc.cn) and are used to generate a prior parameter of CREST model [8].

Daily discharge data (in m^3/s) was obtained from the basin outlet of Shehong station for 2006 to 2013. The precipitation data of China Gauge-Based Daily Precipitation

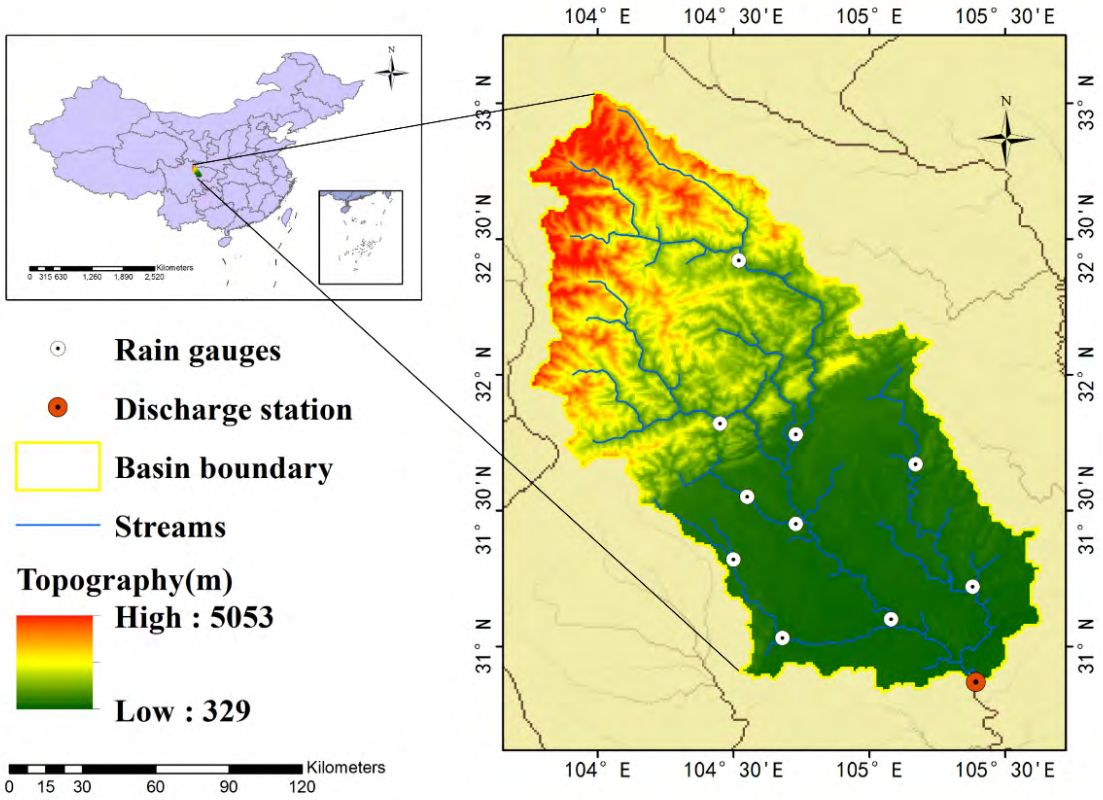


FIGURE 1. The location of Shehong basin.

TABLE 1. Description and sources of environmental database in the Shehong basin.

Sort	Format	Description	Source
DEM	RASTER	30 arc-second resolution digital elevation model (DEM) of the Shehong basin	https://hydrosheds.cr.usgs.gov/
Land use	RASTER	Cultivated land, woodland, grassland, water area, residential land, wasteland (2010)	Resources and Environmental Science Data Center (RESDC) http://www.resdc.cn
Meteorology	ASCII	Gauged precipitation (CGDPA), Satellite-based precipitation (3B42V7) and Potential evapotranspiration (PET)	CGDPA: China Meteorological Data Sharing Service System http://cdc.nmic.cn 3B42V7: NASA archive https://pmm.nasa.gov/dataccess/downloads/trmm PET: Famine Early Warning Systems Network https://earlywarning.usgs.gov/fews/search/Global
Discharge	Excel	Daily value of the discharge data in Shehong hydrological station (2006–2013)	Sichuan hydrological report
Soils	RASTER	Soil classification according to soil genesis classification system	Resources and Environmental Science Data Center (RESDC) http://www.resdc.cn

Analysis (CGDPA) product was downloaded from China Meteorological Data Sharing Service System (available at http://cdc.nmic.cn). This product, based on about 2400 gauge stations over Mainland China from 1955 to the present, is provided in real time at 0.25° × 0.25° spatial resolution using a modified interpolation method of climatology-based optimal interpolation (OI) with topographic corrections [27], [28]. We used it as in-situ data in the hydro-climatology analysis because it has been proven to be of high quality in both statistical and hydrologic assessments and has been used to be a benchmark for rainfall information in retrospective studies [1], [29], [30]. The TRMM was used as a forcing variable to drive the CREST model to simulate the rainfall-runoff relationship in the study area and to evaluate the hydrological capability of remote sensing data. TRMM is a joint mission of National Aeronautics and Space Administration (NASA) and the Japan Aerospace Exploration Agency, with its core observatory launched in 1997. The latest Version-7 product of the TRMM Multi-Satellite Precipitation Analysis (TMPA) is post-real-time 3B42V7 covering 50° N-50° S at

$0.25° \times 0.25°$ and daily resolution. The 3B42V7 dataset was obtained from NASA archive (available at https://pmm.nasa.gov/data-access/downloads/trmm). 3B42V7 incorporating Global Precipitation Climatology Center (GPCC) monthly precipitation gauge analysis was designed to maximize the data quality [1], [12].

Potential evapotranspiration (PET) data was achieved from the global daily database provided by the Famine Early Warning Systems Network (available at https://earlywarning.usgs.gov/fews/search/Global) at 1° spatial resolutions. It was extracted from the Global Data Assimilation System (GDAS) analysis fields using the Penman-Monteith equation with global coverage [31]–[33].

III. METHODS

A. CREST MODEL

The CREST model [8], [34] is a raster-based distributed hydrological model developed by the NASA SERVIR Project Team (available at www.servir.net) and the University of Oklahoma (available at http://hydro.ou.edu). It aims to simulate the spatio-temporal variation of water fluxes and storages on regular grid cells of arbitrary user-defined resolution [35]. The model accounts for main processes of water balance, i.e., runoff generation and infiltration. The mechanism of CREST model is briefly summarized as: (1) Use a variable infiltration curve [36] to represent the soil moisture storage capacity and to calculate runoff generation composition; (2) Use multi-linear reservoirs to simulate sub-grid cell routing of surface and subsurface runoff respectively; and (3) Couple the runoff generation and cell to cell routing components via feedback mechanisms to simulate the processes for water budget. This coupling allows the model to be applied at global, national and regional scales to make hydrological analysis.

Equations (1) and (2) represent variable infiltration curve which is formulated as follows:

$$i = i_m[1 - \left(1 - \frac{W}{W_m}\right)^{1/1+b_i}] \tag{1}$$

$$i_m = W_m(1 + b_i) \tag{2}$$

where i is the point infiltration capacity; i_m is the maximum infiltration capacity of a cell; W is the cell's total mean water depth of the three soil layers; W_m is the cell's maximum water capacity and is related to soil porosity; and b_i is the exponent of the curve. The amount of infiltration water (I) is then estimated as follows:

$$\begin{cases} if\ i + P_{soil} \geq i_m, & \mathrm{I} = W_m - W \\ otherwise, & \mathrm{I} = (W_m - W) + W_m[1 - \frac{i + P_{soil}}{i_m}]^{1+b_i} \end{cases} \tag{3}$$

Meanwhile, equations (4) and (5) show the process of runoff generation. Given the precipitation reaches the soil surface P_{soil} and Infiltration water(I), the excess rain (R) is then computed as:

$$R = P_{soil} - \mathrm{I} \tag{4}$$

A further partitioning of R into overland excess rain (R_O) and interflow excess rain (R_I) is defined as follows:

$$\begin{cases} if\ P_{soil} > K, & R_I = K\dfrac{R}{P_{soil}},\ R_O = R - R_I \\ Otherwise, & R_I = R,\ R_O = 0 \end{cases} \tag{5}$$

where K is the infiltration rate of the first soil layer (three in total) and is closely related to the soil saturated hydraulic conductivity. More information about evapotranspiration and runoff routing can be found in [8].The CREST model has been implemented successfully in previous studies, proving its high cost effectiveness in hydrological prediction [37], [38]. The latest version, CREST v2.1 with a upgrade using a fully distributed linear reservoir routing scheme (FDLRR) is employed as an approach to model the rainfall-runoff relationship in this study [39].

The input of CREST model includes 7 physical parameters and 5 conceptual parameters. The 7 physical parameters are the multiplier on the precipitation field (RainFact), the soil saturated hydraulic conductivity (Ksat), the Mean Water Capacity [13], the exponent of the variable infiltration curve (B), the impervious area ratio (IM), the factor to convert PET to local actual (KE), and the overland runoff velocity coefficient (CoeM). The 5 conceptual parameters are the overland flow speed exponent (ExpM), the multiplier to convert overland flow to channel flow speed (CoeR), the multiplier to convert overland flow to interflow flow speed (CoeS), the overland reservoir Discharge Parameter, and the interflow Reservoir Discharge Parameter (KI). The above parameters need to be calibrated automatically or manually, with the default value and the range of parameters estimated from field survey data like land use/cover data and soil surveys. Table 2 lists the core physically-based inputs of CREST model. The DEM are derived from HydroSHEDS and other basic parameters are derived from DEM.

B. CALIBRATION METHOD AND STATISTICAL METRICS

An auto-calibration method based on the shuffled complex evolution optimization algorithm (SCE-UA) [40] is used to calibrate the CREST model parameters. The SCE technique, which promises to be robust, stable and efficient for global optimization, has been frequently and successfully applied in the field of hydrological model calibration [41].

Table 3 lists several commonly used statistical indices to quantitatively evaluate the performance of CREST model simulation in the study. The Nash–Sutcliffe coefficient efficiency, NSCE [42] was utilized to evaluate the goodness of fit between the "test" field (simulated streamflow) and the "reference" field (gauged streamflow). Pearson correlation coefficient (CC) is a measure of correlation between the two fields, representing the degree of agreement. Three other popular metrics were chosen to demonstrate error and bias between model estimates and gauge observations. The mean error (ME) was selected to measure the average difference between the two fields while relative bias (BIAS) was expressed as a ratio to describe the systematic bias between

TABLE 2. Main physically-based inputs in CREST model.

Symbols	Brief description	Data sources	Unit
DEM	The Digital Elevation Model	Remote Sensing	m
FAC	Flow accumulation	Derived from DEM	N/A
FDR	Flow direction	Derived from DEM	N/A
Slope	Slope between cells	Derived from DEM	degree
Stream	Value in the river is 1, otherwise is 0	Derived from DEM	N/A

TABLE 3. List of the diagnostic statistics used in hydrologic comparison and evaluation.

Statistic metrics	Equation	Unit	Perfect value
Nash–Sutcliffe coefficient efficiency (NSCE)	$\mathrm{NSCE} = 1 - \frac{\sum_{n=1}^{N}(r_n - f_n)^2}{\sum_{n=1}^{N}(r_n - \bar{r})^2}$	N/A	1
Relative bias (BIAS)	$\mathrm{BIAS} = \frac{\frac{1}{N}\sum_{n=1}^{N}(f_n - r_n)}{\sum_{n=1}^{N} r_n} \times 100\%$	N/A	0
Correlation coefficient (CC)	$\mathrm{CC} = \frac{\sum_{n=1}^{N}(r_n - \bar{r})(f_n - \bar{f})}{\sqrt{\sum_{n=1}^{N}(r_n - \bar{r})^2 (f_n - \bar{f})^2}}$	N/A	1
Mean error (ME)	$\mathrm{ME} = \frac{1}{N}\sum_{n=1}^{N}(f_n - r_n)$	m³/s	0
Root-mean-square error (RMSE)	$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(f_n - r_n)^2}$	m³/s	0

Notation: N, number of samples; f_n, a ''test'' field f representing simulated streamflow; r_n, a reference field r corresponding gauged observed streamflow; $\bar{f}$, the average of all the simulated streamflow values, which is also fit for the observed streamflow with the $\bar{f}$ replaced by $\bar{r}$.

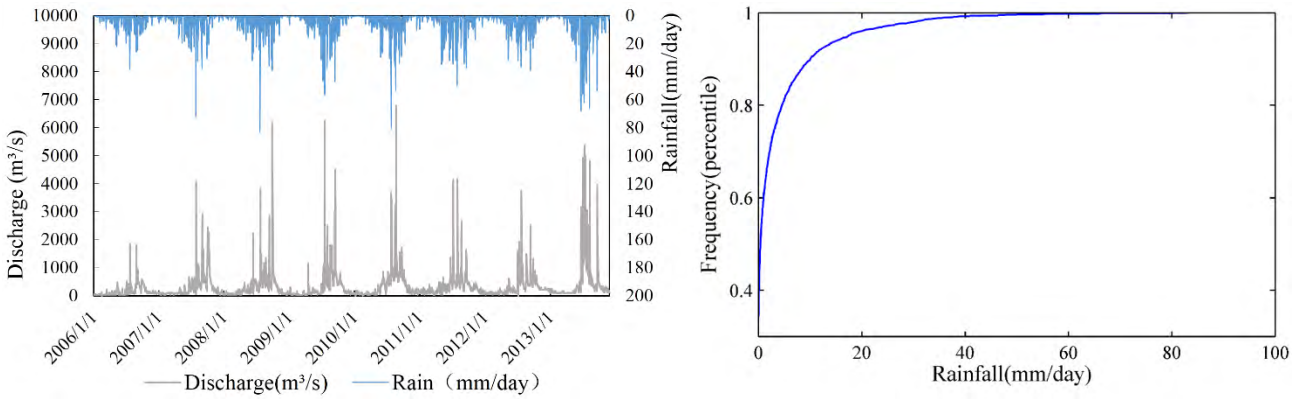


FIGURE 2. (a) Average daily discharge and rainfall time series of Shehong basin. (b) Cumulative distribution plot of observed basin average rainfall (mm/day) for 2006–2013.

the two fields. In addition, Root-mean-square error (RMSE) represents the average error magnitude.

IV. HYDRO-CLIMATOLOGY OF SHEHONG BASIN

A. RAINFALL AND DISCHARGE ANALYSIS

Table 4 lists the mean monthly rainfall and discharge over Shehong basin during the period of 2006-2013. It can be seen that the rainy season extends from May to September, with the maxima used to occur in July and September that formed dual peaks. Similarly, the maximum discharge occurred from July to September while the minimum discharge lasted from January to March, indicating that it was plenty of rain in summer that caused the short spell of high runoff, as depicted in Fig. 2a. Fig. 2b shows that approximately 90% recorded rainfalls are below 10 mm day^{-1}.

The discharge histogram in Fig. 3a mainly consists of low discharge, and more than half of recorded discharges were less than 170 m^3/s. Fig. 3b shows that the average discharge in 2006-2013 is 342.4 m^3/s and fluctuates from 165 m^3/s in 2006 to 511 m^3/s in 2013. All the wet years were accompanied by high monthly runoff (from July to September), whereas dry years resulted from continuously low monthly runoff throughout the whole year. The flow duration curve is shown in Fig. 3c, which indicates the number of times that a certain daily flow value is equal or above the average daily follow in the Shehong basin.

B. FOUR-YEAR MONTHLY TREND

The recorded data was also divided into two periods to further analyze the trend of discharge and rainfall. The first period

TABLE 4. Seasonal variation of rainfall and streamflow during 2006-2013.

	Periods	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Avg
Rainfall	2006-2009	0.36	0.53	1.16	2.35	3.15	4.10	8.11	5.98	6.88	3.04	0.68	0.25	
(mm/day)	2010-2013	0.39	0.35	0.89	1.99	6.20	6.27	11.70	6.05	6.75	2.14	0.91	0.46	
	Change	0.03	-0.18	-0.27	-0.36	3.05	2.17	3.59	0.07	-0.13	-0.90	0.23	0.21	
	% Change	8%	-34%	-23%	-15%	97%	53%	44%	1%	-2%	-30%	34%	84%	+18%
Streamflow	2006-2009	71	59	71	124	149	202	590	574	851	511	139	86	
(m^3/s)	2010-2013	116	99	111	122	230	433	1328	822	650	353	219	170	
	Change	45	40	40	-2	81	231	738	248	-201	-158	80	84	
	% Change	63%	68%	56%	-2%	54%	114%	125%	43%	-24%	-31%	57%	98%	+52%

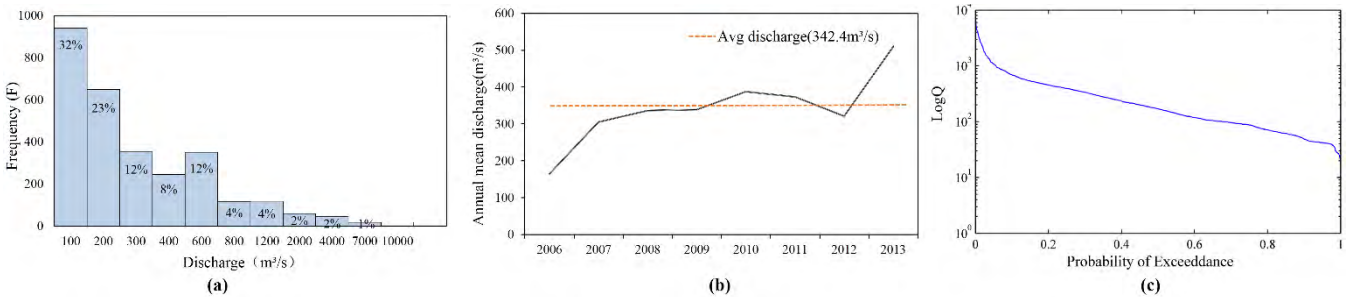


FIGURE 3. Observed daily discharge (m^3/s) for Shehong basin for 2006-2013 (a) histogram (b) Annual mean discharge and (c) Flow duration curve.

TABLE 5. Comparison of daily observed and simulated discharge using CREST model during the calibration period (2006-2009) and validation period (2010-2013).

Periods	Precipitation inputs	NSCE	BIAS (%)	CC	ME (m^3s^{-1})	RMSE (m^3s^{-1})
Calibration period	Gauged	0.77	-11.00	0.88	-33.16	243.94
	3B42V7	0.67	-16.00	0.80	-46.58	305.18
Validation period	Gauged	0.74	-11.70	0.87	-46.47	313.98
	3B42V7	0.61	-17.00	0.78	-62.71	395.83

is 2006-2009 (i.e., CREST calibration period) and the second period is 2010-2013 (i.e., CREST validation period). Compared with the first period, moderate increase (+18%) in rainfall existed in the second period and a significant increase (+52%) in discharge was also observed. Moreover, Table 4 provides detail information in monthly variation of two periods for both rainfall and discharge. For rainfall, there is a maximum increase in May (+97%) while a maximum decrease (−34%) in February. Similar with discharge, a maximum increase is witnessed in July (+125%) while a maximum drop (−31%) is discovered in October. The monthly trends of rainfall and discharge are generally consistent, indicating that precipitation is one of most important factors affecting runoff.

V. CREST MODEL SETUP, CALIBRATION AND VERIFICATION

A. CREST MODEL SETUP

Model initialization is important to produce actual and reasonable hydrologic status in the study area. We warmed up the model with 1-year (2005) data. The initialized CREST model was calibrated using a period of 4 years (2006-2009) including wet, dry and normal hydrological conditions over the Shehong basin. The period of 2010-2013 was finally validated using the parameters calibrated by corresponding precipitation inputs.

B. RESULTS OF CREST CALIBRATION AND VALIDATION

CREST calibration using SCE-UA resulted in favorable hydrological performance with NSCE of 0.77, BIAS of −11% and CC of 0.88. The performance of CREST in discharge simulation was also validated. As summarized in Table 5, there is a moderate decline of all the metrics in validation period compared to calibration period for both gauged precipitation and 3B42V7. The explanation for that is the increased human activities in the catchment area during recent years [35]. Fig. 4a shows that CREST model adequately captures a majority of peak flows, followed by good performance with NSCE of 0.74, BIAS of −11.7% and CC up to 0.87 in the validation period. The results show that CREST model reasonably and robustly reproduces the volume of daily observed discharge reported for Shehong station. It is accordingly suitable to use CGDPA as a benchmark for evaluating the hydrologic utility of the satellite-based products. However, it still misrepresented some extreme flood peaks in the rainy season (July to September). More specifically,

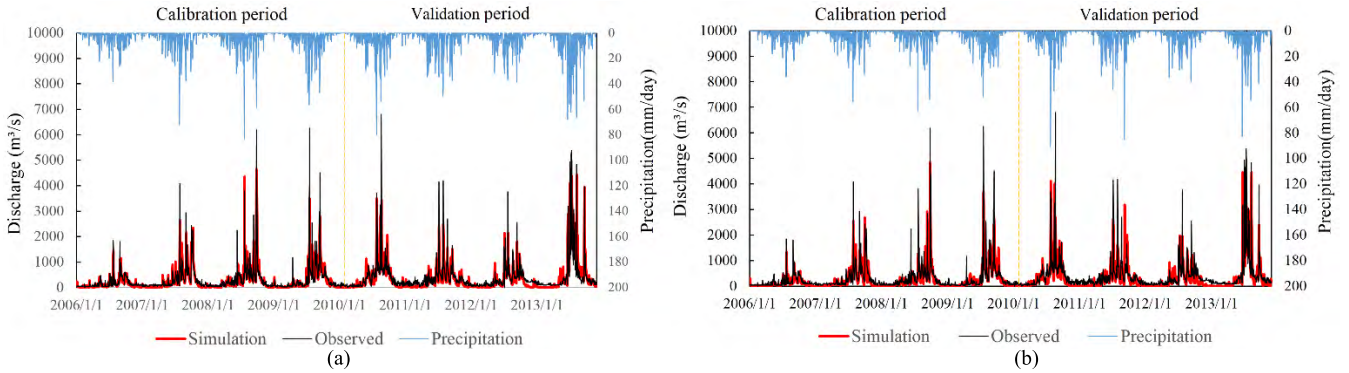


FIGURE 4. Comparison of CREST simulated daily discharge from (a) Gauged precipitation, (b) 3B42V7, with product-specific calibrated parameters and observed discharge in both calibration (from 1 Jan 2006 to 31 Dec 2009) and validation (from 1 Jan 2010 to 31 Dec 2013) periods for Shehong basin.

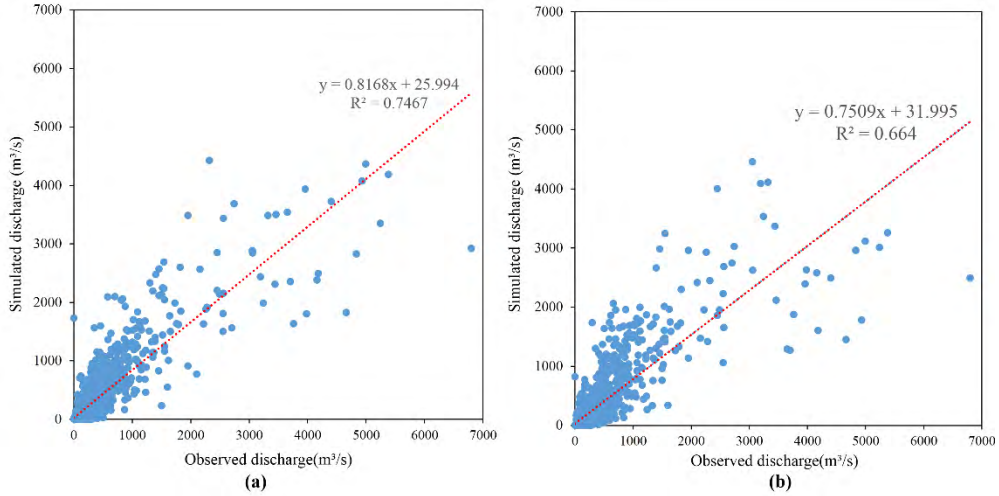


FIGURE 5. Scatterplots for the daily observed discharge versus CREST-simulated in Shehong basin using (a) Gauged (b) 3B42V7 as precipitation forcing in validation period.

based on the scatterplots, the magnitudes of flood peaks that exceeded 3000 m^3/s were nearly underestimated (Fig.5).

Satellite-based products are regarded as an alternative for application to ungauged basins where only satellite precipitation products are available for use [17], [34]. Fig. 4b shows that 3B42V7-driven model simulation matches well with the observed discharge, especially for the small flow peaks with NSCE of 0.61, BIAS of −17%, ME of 62.71 m^3/s, RMSE of 395.83 m^3/s and high CC (Table 5). Overall, 3B42V7 performs acceptable skill in hydrological application, but tends to be slightly worse than that of gauged precipitation. The reason is the uncertainty in satellite-based products leading by instrumental and systematic errors.

VI. CREST MODEL RECONSTRUCTION RESULTS

Basin-based models have contributed to advancements in quantitatively assessing the change on runoff process and impact of climate variability [3]. It is important to carry out catchment scale water balance modeling studies since they provide insights into hydrological process and the water amount. The water balance of a watershed is normally governed by dS/dt = P-ET-R, where dS/dt is the change of total water storage in the catchment, and P, ET, R represent precipitation, actual evapotranspiration and total runoff, respectively [43]–[45]. In the equation, precipitation is confirmed to be a dominant variable for accurate water balance estimation. The CREST model outputs many hydrological variables on distributed grid cells of arbitrary user-defined resolution. In the study, the water balance components like runoff and actual evapotranspiration are generated from CREST simulation from 2010-2013 at daily time step and 30 arc-second spatial resolution. We select 3B42V7 as precipitation input to force the CREST model. These four years were selected to build the reconstruction from the consideration of the data availability and the minimizing the run time of CREST model.

A. RUNOFF

Fig. 6 shows the basin average monthly water balance reconstruction results. Considerable agreements in magnitude and time evolution between monthly model runoff estimates and gauged observations can be observed. The monthly peak

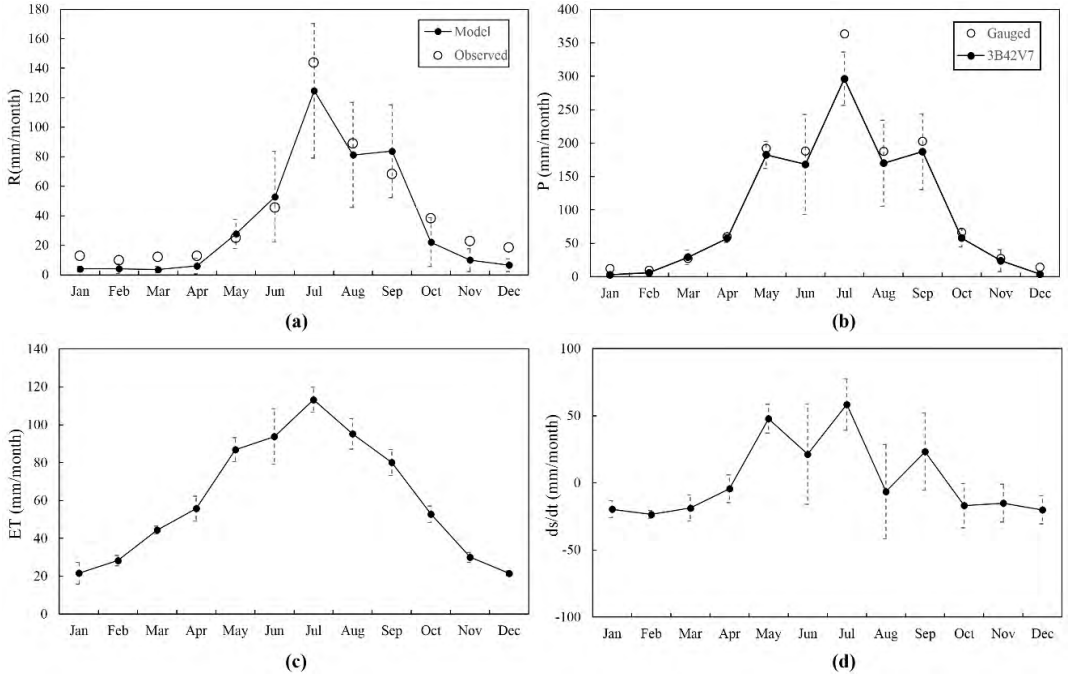


FIGURE 6. Monthly water balance analysis of mean annual cycle 2010–2013 (a) Model simulation versus observed runoff depth, (b) Precipitation, (c) Actual Evapotranspiration and (d) Change in storage. Error bars showing the ±1 std dev of monthly mean values.

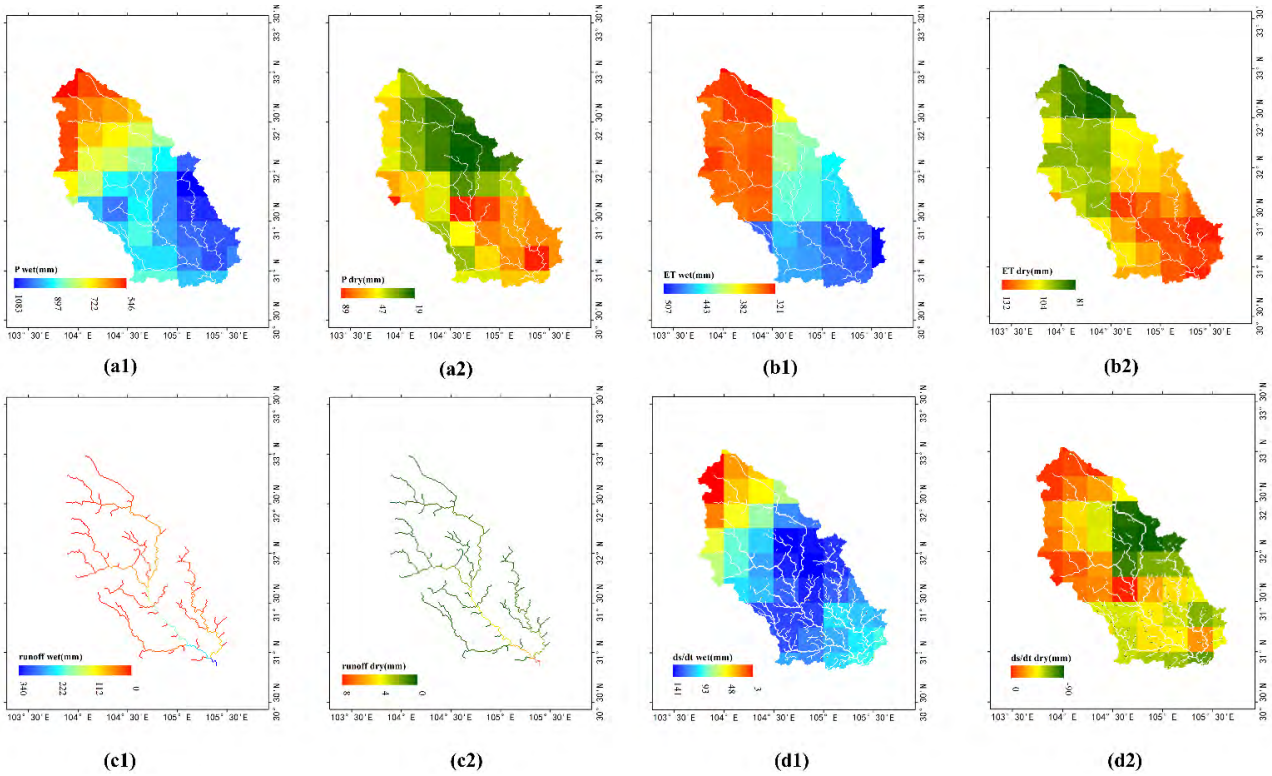


FIGURE 7. Spatially distributed Precipitation (P) (a1, a2), Actual evapotranspiration (ET) (b1, b2), Runoff depth (c1, c2),Change in storage (ds/dt) (d1, d2) for wet season (from June to September) and dry season (from December to March) respectively averaged over study period (2010–2013).

runoff exists in the months of July, August and September (Fig. 6b). There is an overall slight underestimation for model simulated runoff, especially for low values occurring in dry months and extreme high values in July, as depicted in Fig. 4b. The underestimation results from the accuracy of 3B42V7 product. Fig. 6a shows that 3B42V7 underestimates the

rainfall values relative to gauged precipitation that results in under prediction of runoff. The observation is consistent with the conclusions from other studies [16], [46]–[48]. Most runoff values are constrained within the ±1 standard deviation (std) of monthly mean values.

B. PRECIPITATION

Researchers can hardly conduct better simulations and acquire better cognition of water balance over catchments without accurate precipitation inputs [38]. In the study, TMPA 3B42V7 dataset was utilized as the hydrological variable of precipitation to characterize the water balance in Shehong basin. In Fig. 6a, 3B42V7 shows fairly good agreement with gauged precipitation and successfully captures the seasonality of precipitation over Shehong basin. Similar to runoff, a slight underestimation is observed throughout the year. It should be noted that model estimates overestimated the runoff values in June and September despite an underestimation of rainfall values. It may be attributed to the water management practices during that period. Similarly, almost all of the 3B42V7 estimates fall under the ±1 std dev of monthly mean values.

C. EVAPOTRANSPIRATION

Accurate estimation of evapotranspiration, a key component of hydrological cycle, is of great significance for understanding the relationships between climate and water balance. Furthermore, evapotranspiration plays an important role in determining the allocation of land surface precipitation in the hydrologic cycle consequently impacting the freshwater recharge [49]. In the catchment scale, it is difficult to measure sufficient quantities of reliable actual evapotranspiration. Hydrological modelling provides a practical way to estimate actual evapotranspiration. Fig. 6c exhibits the simulated actual evapotranspiration for the time period. Monthly ET shows similar variation trend as precipitation, but not exactly varies as much as rainfall does in the same period. The CREST model not only estimates hydrological variables at the basin outlet, but also outputs the runoff at any location of the Shehong basin. Figs. 7(a1) and (a2) show the distribution of precipitation in the dry and wet seasons across the basin is very unbalanced. Figs. 7 (b1) and (b2) illustrate that the evapotranspiration is similar to rainfall amounts in dry months. Figs. 7 (c1) and (c2) show the spatially distributed runoff averaged over 2010-2013.The high runoff values in the wet season coincide with flood events occurring across the basin.

VII. CONCLUSIONS

The study investigates hydrological modelling in Shehong basin, a humid region in southwest China utilizing CREST model and satellite remote sensing big data. The rainfall and discharge observations show that the main rainfall of Shehong basin is concentrated in summer (May to September) because of subtropical humid monsoon climate. There is an overall remarkable increase (+52%) in discharge and relatively less increase (+18%) in rainfall. Gauged and simulated discharges agree well with each other during the calibration and validation periods, with the gauged precipitation and 3B42V7 as precipitation forcing. TMPA-3B42V7 product shows comparable hydrologic utility with the reference precipitation in hydrological prediction at Shehong basin. Monthly model runoff estimates show considerable agreement with gauge observations recorded at Shehong station, indicating CREST model can capture the soil moisture storage variability. Change in storage patterns (the sum of change in storage is close to zero) indicate that CREST model reproduces the water cycle processes at Shehong basin with acceptable skills. In terms of water balance reconstruction results, the model is proven to be beneficial in addressing issues pertaining to sustainability of water resources within the catchment with respect to model robustness and effectiveness. Therefore, Satellite precipitation products help to understand the potential hydrological utility of remote sensing data for hydrological modelling studies in ungauged or poorly gauged basins. As retrospectively processed GPM-era IMERG datasets are released, more studies to simulate the key hydrological processes and reconstruct the water budget are required in the future.

ACKNOWLEDGMENT

The authors would like to thank the editor and referees for their suggestions that improve the paper.

REFERENCES

[1] Z. Gao, D. Long, G. Tang, C. Zeng, J. Huang, and Y. Hong, "Assessing the potential of satellite-based precipitation estimates for flood frequency analysis in ungauged or poorly gauged tributaries of China's Yangtze River basin," *J. Hydrol.*, vol. 550, pp. 478–496, Jul. 2017.
[2] C. Kidd and G. Huffman, "Global precipitation measurement," *Meteorol. Appl.*, vol. 18, no. 3, pp. 334–353, 2011.
[3] M. L. Shelton, *Hydroclimatology: Perspectives and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2009.
[4] D. Deus, R. Gloaguen, and P. Krause, "Water balance modeling in a semi-arid environment with limited *in situ* data using remote sensing in Lake Manyara, East African Rift, Tanzania," *Remote Sens.*, vol. 5, no. 4, pp. 1651–1680, 2013.
[5] B. Du *et al.*, "Exploring representativeness and informativeness for active learning," *IEEE Trans. Cybern.*, vol. 47, no. 1, pp. 14–26, Jan. 2017.
[6] S. Bergström, *The HBV Model—Its Structure and Applications*. Norrköping, Sweden: Swedish Meteorological and Hydrological Institute, 1992.
[7] B. Sivakumar and R. Berndtsson, Eds., *Advances in Data-Based Approaches for Hydrologic Modeling and Forecasting*. Singapore: World Scientific, 2010.
[8] J. Wang *et al.*, "The coupled routing and excess storage (CREST) distributed hydrological model," *Hydrol. Sci. J.*, vol. 56, no. 1, pp. 84–98, 2011.
[9] Z. Li, D. Yang, and Y. Hong, "Multi-scale evaluation of high-resolution multi-sensor blended global precipitation products over the Yangtze River," *J. Hydrol.*, vol. 500, pp. 157–169, Sep. 2013.
[10] Z. Li, D. Yang, B. Gao, Y. Jiao, Y. Hong, and T. Xu, "Multiscale hydrologic applications of the latest satellite precipitation products in the Yangtze River basin using a distributed hydrologic model," *J. Hydrometeorol.*, vol. 16, pp. 407–426, Feb. 2015.
[11] J. Meng, L. Li, Z. Hao, J. Wang, and Q. Shao, "Suitability of TRMM satellite rainfall in driving a distributed hydrological model in the source region of Yellow River," *J. Hydrol.*, vol. 509, pp. 320–332, Feb. 2014.
[12] G. J. Huffman and D. T. Bolvin, "TRMM and other data precipitation data set documentation," NASA, Greenbelt, MD, USA, Tech. Rep., 2013, vol. 28.

[13] G. J. Huffman, D. T. Bolvin, E. J. Nelkin, and D. B. Wolff, "The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales," *J. Hydrometeorol.*, vol. 8, no. 1, pp. 38–55, Feb. 2007.

[14] B. Du, Y. Zhang, L. Zhang, and D. Tao, "Beyond the sparsity-based target detector: A hybrid sparsity and statistics-based detector for hyperspectral images," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5345–5357, Nov. 2016.

[15] B. Yong *et al.*, "Assessment of evolving TRMM-based multisatellite real-time precipitation estimation methods and their impacts on hydrologic prediction in a high latitude basin," *J. Geophys. Res., Atmos.*, vol. 117, no. D9, p. D09108, 2012.

[16] B. Yong *et al.*, "Hydrologic evaluation of multisatellite precipitation analysis standard precipitation products in basins beyond its inclined latitude band: A case study in Laohahe basin, China," *Water Resour. Res.*, vol. 46, no. 7, pp. 759–768, 2010.

[17] K. Tong, F. Su, D. Yang, and Z. Hao, "Evaluation of satellite precipitation retrievals and their potential utilities in hydrologic modeling over the Tibetan Plateau," *J. Hydrol.*, vol. 519, pp. 423–437, Nov. 2014.

[18] L. Zhang, L. Zhang, D. Tao, X. Huang, and B. Du, "Compression of hyperspectral remote sensing images by tensor approach," *Neurocomputing*, vol. 147, no. 1, pp. 358–363, Jan. 2015.

[19] Q. Hu, D. Yang, Z. Li, A. K. Mishra, Y. Wang, and H. Yang, "Multi-scale evaluation of six high-resolution satellite monthly rainfall estimates over a humid region in China with dense rain gauges," *Int. J. Remote Sens.*, vol. 35, no. 4, pp. 1272–1294, 2014.

[20] G. Tang *et al.*, "Statistical and hydrological comparisons between TRMM and GPM level-3 products over a midlatitude basin: Is day-1 IMERG a good successor for TMPA 3B42V7?" *J. Hydrometeorol.*, vol. 17, pp. 121–137, Jan. 2016.

[21] B. Fu, Y. K. Wang, P. Xu, and K. Yan, "Mapping the flood mitigation services of ecosystems—A case study in the Upper Yangtze River Basin," *Ecol. Eng.*, vol. 52, pp. 238–246, Mar. 2013.

[22] L. Wu, T.-Y. Long, X. Liu, and J.-S. Guo, "Impacts of climate and land-use changes on the migration of non-point source nitrogen and phosphorus during rainfall-runoff in the Jialing River Watershed, China," *J. Hydrol.*, vol. 475, pp. 26–41, Dec. 2012.

[23] S. Chen *et al.*, "Evaluation of high-resolution precipitation estimates from satellites during July 2012 Beijing flood event using dense rain gauge observations," *PLoS ONE*, vol. 9, no. 4, p. e89681, 2014.

[24] X.-Z. Shi *et al.*, "Soil information system of China (SISChina) and its application," *Soils*, vol. 39, no. 3, pp. 329–333, 2007.

[25] U. Pfeifroth, J. Trentmann, A. H. Fink, and B. Ahrens, "Evaluating satellite-based diurnal cycles of precipitation in the African tropics," *J. Appl. Meteorol. Climatol.*, vol. 55, pp. 23–39, Jan. 2016.

[26] N. Li, G. Tang, P. Zhao, Y. Hong, Y. Gou, and K. Yang, "Statistical assessment and hydrological utility of the latest multi-satellite precipitation analysis IMERG in Ganjiang River basin," *Atmos. Res.*, vol. 183, pp. 212–223, Jan. 2017.

[27] P. Xie *et al.*, "A gauge-based analysis of daily precipitation over East Asia," *J. Hydrometeorol.*, vol. 8, pp. 607–626, Jun. 2007.

[28] Y. Shen and A. Xiong, "Validation and comparison of a new gauge-based precipitation analysis over mainland China," *Int. J. Climatol.*, vol. 36, no. 1, pp. 252–265, 2016.

[29] R. Sun, H. Yuan, X. Liu, and X. Jiang, "Evaluation of the latest satellite–gauge precipitation products and their hydrologic applications over the Huaihe River basin," *J. Hydrol.*, vol. 536, pp. 302–319, May 2016.

[30] Y. Wang, H. Yang, D. Yang, Y. Qin, B. Gao, and Z. Cong, "Spatial interpolation of daily precipitation in a high mountainous watershed based on gauge observations and a regional climate model simulation," *J. Hydrometeorol.*, vol. 18, pp. 845–862, Mar. 2017.

[31] J. Verdin, C. Funk, G. Senay, and R. Choularton, "Climate science and famine early warning," *Philos. Trans. Roy. Soc. B, Biol. Sci.*, vol. 360, no. 1463, pp. 2155–2168, 2005.

[32] B. Du, M. Zhang, L. Zhang, R. Hu, and D. Tao, "PLTD: Patch-based low-rank tensor decomposition for hyperspectral images," *IEEE Trans. Multimedia*, vol. 19, no. 1, pp. 67–79, Jan. 2017.

[33] B. Du, W. Xiong, J. Wu, L. Zhang, L. Zhang, and D. Tao, "Stacked convolutional denoising auto-encoders for feature representation," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 1017–1027, Apr. 2016.

[34] S. I. Khan *et al.*, "Satellite remote sensing and hydrologic modeling for flood inundation mapping in Lake Victoria basin: Implications for hydrologic prediction in ungauged basins," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 85–95, Jan. 2011.

[35] S. I. Khan *et al.*, "Hydroclimatology of Lake Victoria region using hydrologic model and satellite remote sensing data," *Hydrol. Earth Syst. Sci.*, vol. 15, pp. 107–117, Jan. 2011.

[36] R. J. Zhao, Y. L. Zhang, L. R. Fang, X. R. Liu, and Q. S. Zhang, "The Xinanjiang model," in *Proc. Hydrol. Forecast.*, 1980, pp. 351–356.

[37] Y. Zhang *et al.*, "Hydrometeorological analysis and remote sensing of extremes: Was the July 2012 Beijing flood event detectable and predictable by global satellite observing and global weather modeling systems?" *J. Hydrometeorol.*, vol. 16, pp. 381–395, Feb. 2015.

[38] X. Xue *et al.*, "Statistical and hydrological evaluation of TRMM-based multi-satellite precipitation analysis over the Wangchu Basin of Bhutan: Are the latest satellite precipitation products 3B42V7 ready for use in ungauged basins?" *J. Hydrol.*, vol. 499, pp. 91–99, Aug. 2013.

[39] X. Shen, Y. Hong, K. Zhang, and Z. Hao, "Refining a distributed linear reservoir routing method to improve performance of the CREST model," *J. Hydrol. Eng.*, vol. 22, no. 3, p. 04016061, 2016.

[40] Z. Ren-Jun, "The Xinanjiang model applied in China," *J. Hydrol.*, vol. 135, nos. 1–4, pp. 371–381, 1992.

[41] G. Kan *et al.*, "Accelerating the SCE-UA global optimization method based on multi-core CPU and many-core GPU," *Adv. Meteorol.*, vol. 2016, Mar. 2016, Art. no. 8483728.

[42] J. E. Nash and J. V. Sutcliffe, "River flow forecasting through conceptual models: Part I—A discussion of principles," *J. Hydrol.*, vol. 10, no. 3, pp. 282–290, 1970.

[43] C. W. Thornthwaite, "An approach toward a rational classification of climate," *Geograph. Rev.*, vol. 38, no. 1, pp. 55–94, 1948.

[44] C. J. Willmott, C. M. Rowe, and Y. Mintz, "Climatology of the terrestrial seasonal water cycle," *J. Climatol.*, vol. 5, no. 6, pp. 589–606, 1985.

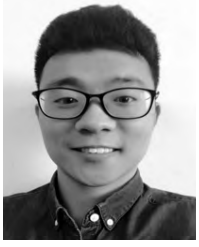
[45] L. Zhang, N. Potter, K. Hickel, Y. Zhang, and Q. Shao, "Water balance modeling over variable time scales based on the Budyko framework—Model development and testing," *J. Hydrol.*, vol. 360, pp. 117–131, Oct. 2008.

[46] G. Tang, Y. Ma, D. Long, L. Zhong, and Y. Hong, "Evaluation of GPM day-1 IMERG and TMPA version-7 legacy products over Mainland China at multiple spatiotemporal scales," *J. Hydrol.*, vol. 533, pp. 152–167, Feb. 2016.

[47] H. Guo *et al.*, "Early assessment of integrated multi-satellite retrievals for global precipitation measurement over China," *Atmos. Res.*, vols. 176–177, pp. 121–133, Jul./Aug. 2016.

[48] S. K. Siuki, B. Saghafian, and S. Moazami, "Comprehensive evaluation of 3-hourly TRMM and half-hourly GPM-IMERG satellite precipitation products," *Int. J. Remote Sens.*, vol. 38, no. 2, pp. 558–571, 2016.

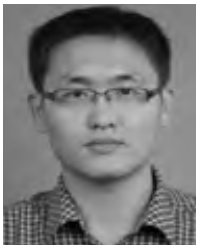
[49] J. Szilagyi and M. B. Parlange, "Defining watershed-scale evaporation using a normalized difference vegetation index," *J. Amer. Water Resour. Assoc.*, vol. 35, no. 5, pp. 1245–1255, 1999.



JUN MA received the B.S. degree in humanistic geography and urban-rural planning from Zhejiang Gongshang University in 2016. He is currently pursuing the master's degree with Ningbo University. His research interests include complex big data analysis and modeling in hydrology and remote sensing.

WEIWEI SUN (M'15) received the B.S. degree in surveying and mapping from Tongji University in 2007, and Ph.D. degree in cartography and geographic information engineering from Tongji University, Shanghai, China, in 2013.

From 2011 to 2012, he was a Visiting Scholar with the Department of applied mathematics, University of Maryland at College Park, with the famous Professor J. Benedetto to study on the dimensionality reduction of hyperspectral image. From 2014 to2016, he was a Post-Doctoral Researcher with the State Key Laboratory for Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, to study intelligent processing in hyperspectral imagery. He is currently an Associate Professor with Ningbo University, China, and is also a Visiting Scholar with the Department of Electrical and Computer Engineering, Mississippi State University. He has authored or co-authored over 40 journal papers. His current research interests include hyperspectral image processing with manifold learning, and anomaly detection and target recognition of remote sensing imagery using compressive sensing.



GANG YANG received the M.S. degree in geographical information system from the Hunan University of Science and Technology, Xiangtan, China, in 2012, and the Ph.D. degree from the School of Resource and Environmental Sciences, Wuhan University, Wuhan, China, in 2016. He is currently an Assistant Professor with Ningbo University, Ningbo, China.

His research interests include missing information reconstruction of remote sensing image, cloud removal of remote sensing image, and remote sensing time-series products temporal reconstruction.

DIANFA ZHANG received the Ph.D. degree from the School of Resource and Environment, Jilin University in 2000.

From 2000 to 2002, he was a Post-Doctoral Researcher with the Guiyang Institute of Geochemistry, Chinese Academy of Sciences. He is currently a Full Professor with Ningbo University, China. He has authored or co-authored over 50 journal papers. His current research interests include hyperspectral remote sensing processing, and the applications of remote sensing imagery in coastal land monitoring and coastal diseases.