

ASSESSING THE CLIMATE FORECAST SYSTEM REANALYSIS WEATHER DATA DRIVEN HYDROLOGICAL MODEL FOR THE YANGTZE RIVER BASIN IN CHINA

LU, J. Z.¹ – ZHANG, L.¹ – CUI, X. L.¹ – ZHANG, P.² – CHEN, X. L.^{1*} – SAUVAGE, S.³ – SANCHEZ-PEREZ, J. M.³

¹*State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China*

²*State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China*

³*ECOLAB, Université de Toulouse, CNRS, INPT, UPS, 31400 Toulouse, France*

**Corresponding author*

e-mail: xiaoling_chen@whu.edu.cn; phone: +86-27-68778321

(Received 10th Dec 2018; accepted 20th Feb 2019)

Abstract. It is a challenge to adequately represent meteorological data for river basin modelling in data-scarce regions. Hydrological models may benefit from alternative climate data sources such as the Climate Forecast System Reanalysis (CFSR) weather data. A proposed method was applied to evaluate the applicability of CFSR data in the Soil and Water Assessment Tool (SWAT) eco-hydrological model for the 1.8×10^6 km² Yangtze River Basin (YRB) in China. The CFSR data were first validated by ground-based meteorological station (GMS) data to confirm sufficient accuracy to predict the hydrology for the YRB. It was determined that the CFSR simulation was able to generate acceptable accuracy with calibrated parameters of GMS-based model. The R^2 and ENS of the monthly results from both GMS- and CFSR-SWAT simulations were > 0.70 in both calibration and validation periods. The monthly streamflow hydrograph from GMS and CFSR simulations showed two different patterns: distinct differences in baseflow and less discrepancy for the peak flows, such as the patterns at Cuntan and Yichang gauge stations, whereas the results are characterized as close matches the peaks and the recession periods at Hankou and Datong gauge stations. The CFSR simulation generated better results in the middle and lower reaches than in the upper reach of the basin. To a certain extent, the CFSR provides an alternative to quickly build hydrological model in data scarce area for large basins.

Keywords: *SWAT model, data scarce area, model calibration, meteorological forcing data, large-scale watershed*

Introduction

Traditionally, weather data collected at ground-based meteorological stations (GMS) are utilized to drive eco-hydrological models; therefore, it is important to collect adequate and representative weather data. It is difficult and time-consuming to obtain adequate meteorological data for hydrological models to accurately represent climate patterns for some watersheds, especially those located in data-scarce regions, due to inadequate overall coverage of the watershed area provided by ground-based meteorological station (GMS) data. In developing countries or ungauged areas (Lu et al., 2018), weather data are also unavailable because of financial limitations and/or inadequate technology (Vu et al., 2012). In addition, there are shortcomings related to GMS data in regards to supporting rapid development of watershed models. For a small-scale watershed, GMS data are sometimes located outside of the investigated watershed and consequently, the collected records from GMS may not accurately represent the actual weather that occurred at the

watershed (Fuka et al., 2014). The scarcity of GSM, especially in developing countries, increases the uncertainty in precipitation inputs for each model sub-basin (Worqlul et al., 2014). In addition, because of the data management policies of governments for state security in some countries, it is usually difficult to collect sufficient data for hydrological models.

The availability of other sources of data (i.e., satellite precipitation and reanalysis data) where there is no conventional GSM has recently attracted the interest of hydrologists. The hydrological model can benefit from satellite observation data and global reanalyzed weather data such as TRMM (Tropical Rainfall Measuring Mission), ECMWF (European Centre for Medium-Range Weather Forecasting), ERA-40 (40-year reanalysis), CFSR (Climate Forecast System Reanalysis) produced by National Center of Environmental Prediction (NCEP), APHRODITE (Asian Precipitation-Highly-Resolved Observational Data Integration towards Evaluation of Water Resources), and PERSIANN-CDR (Precipitation Estimation from Remotely Sensed Information using Artificial Neural Network-Climate Data Record) to overcome data shortages (Behrangi et al., 2011; Ward et al., 2011; Worqlul et al., 2015; Tan et al., 2017).

Concerns have been increasing on the subject of climate change as a driver of hydrological changes within the Yangtze River Basin (YRB) in China (Zhang et al., 2008; Zhu et al., 2011) based on satellite observation data and global reanalysis weather data. Previous studies were carried out to evaluate the remotely sensed weather products and their reanalysis datasets by application in different basins. Yang et al. (2014) found that gridded precipitation products poorly reproduced hydrologic processes except that the APHRODITE and trend surface data can give stable and desirable results in a relatively flat test basin, in an evaluation of a gridded precipitation-driven Soil and Water Assessment Tool (SWAT) eco-hydrological model application for the upper reaches of Three Gorges Reservoir. Sun et al. (2013) investigated the spatiotemporal rainfall patterns of the YRB using TRMM/PR (TRMM/ Precipitation Radar) data in a decadal scale, which reveals that there are possible linkages between major precipitation patterns and hydrological extremes. Li et al. (2015) conducted water balance and streamflow studies for multiple time scales using three global satellite rainfall products [TMPA (TRMM Multi-satellite Precipitation Analysis) 3B42 V7, TMPA 3B42 Real Time (RT), and CMORPH (Climate Prediction Center morphing technique)] during 2003 to 2012 in the YRB. They found that the 3B42 V7 data generally resulted in improved annual water budgeting and monthly streamflow performance. The 3B42 RT tends to overestimate streamflow in the upper Yangtze River, and, in contrast, CMORPH shows serious underestimation in downstream sub-basins. Worqlul et al. (2014, 2015) compared to rainfall estimations by TRMM 3B42, MPEG (Multi-Sensor Precipitation Estimate–Geostationary) and CFSR with ground-observed data for the Lake Tana basin and Upper Blue Nile in Ethiopia. The MPEG and CFSR satellites provided the most accurate rainfall estimates, and both the gauged data and the CFSR precipitation estimates were able to reproduce the hydrology well for both models and both watersheds. Tan et al. (2017) analyzed the accuracy of three long-term gridded climate products (APHRODITE, PERSIANN-CDR and NCEP-CFSR) for two tropical river basins in Malaysia. The APHRODITE data resulted in strong results for streamflow simulations, while NCEP-CFSR data showed unsatisfactory performance.

The satellite-based precipitation products and their reanalysis weather data provide an alternative data source to study hydrology and climate change. The application of remotely-sensed data and corresponding reanalysis precipitation data within eco-

hydrological models is gradually increasing in many regions; however, use of such remains limited in many areas of the globe including the YRB. Therefore, gridded precipitation products must be evaluated to improve the performance of the watershed model in this data-scarce area. In addition, the SWAT eco-hydrological model (Arnold et al., 1998; Arnold et al., 2012; Gassman et al., 2007, 2014) has recently been applied to parts or all of the YRB to predict the water balance, sediment transport, and nutrients. Ouyang et al. (2008) used SWAT to evaluate the impact on streamflow and nonpoint source pollutants from the conversion of cropland to forest in Bazhong City, located in the central YRB region. Zhang et al. (2011) investigated the water balance of Three Gorges Reservoir in the upper reach of the Yangtze River with the aid of SWAT model. Sun et al. (2016) mapped the water vulnerability of the YRB using SWAT for 1994 to 2013; the statistical results showed that the simulated output accurately replicated the observed data during both the calibration and validation periods. Chen et al. (2017) and Su et al. (2017) assessed the impacts of climate change on river discharge in the upper reaches of the YRB based on hydrological models under different climate projections. Several other studies reported predicted water budget, water quality, and/or the impacts of climate change or land use management using SWAT for various YRB sub-regions (Yang et al., 2014; Huang et al., 2009, 2014; Zhou et al., 2013). The applicability of global reanalysis climate data for hydrological model predictions has not so far been adequately investigated in such a large watershed: the entire Yangtze River basin.

Therefore, the aim of this study was to assess the performance of CFSR weather data versus GMS data for predicting YRB hydrology using SWAT. To evaluate the applicability of CFSR weather data in the watershed model, the same hydrological parameters of the model calibrated using GMS weather data were used in both models, and then both the results were validated against observed data. The simulated results of streamflow from the SWAT model driven by CFSR weather data with and without calibration were compared. Finally, the monthly averaged streamflow simulated by the models driven by CFSR and GMS weather data were investigated to evaluate the performance of CFSR as input weather data.

Materials and Methods

Study area

The Yangtze River, also called the Changjiang, which means ‘Long River’ in Chinese, is the longest river in China and the third longest in the world, with a length of approximately 6,380 km and a drainage area of 1.8×10^6 km². Originating from the Tibetan Plateau, the Yangtze River flows across the Qinghai-Tibet Plateau in west China, Sichuan basin in the middle China, and east China Plain with an average elevation of 3000 m, 1000 m, and 100 m, respectively (Gong et al., 2006). Generally, the Yangtze River is divided into three reaches: upper, which drains the area controlled by the Yichang gauge station, middle, which captures drainage between the Yichang and Hukou gauge stations, and lower, which covers the area downstream of the Hukou gauge station to the estuary (*Figure 1*).

The YRB is located in the Asian monsoon region, which manifests as two independent components: the Indian summer monsoon and the East Asian summer monsoon which affect the upper and mid-lower YRB sub-regions, respectively (Ding et al., 2005). Because of the large area and extensive range of elevation, precipitation varies considerably between the different climatic zones. The annual precipitation in most areas

of the YRB is 800-1600 mm, whereas it is less than 400 mm in the river source areas in the arid zone. The area with precipitation larger than 1600 mm is mainly located in the Sichuan, Jiangxi, Hunan and Hubei provinces. Large floods are frequent in the summer due to the heavy monsoon rainfall influence. Precipitation declines across the basin from east to west, ranging on average annually from 1528 mm in the east to 859 mm in the west. The upper Yangtze region is characterized by the highest elevation and the least precipitation (Chen et al., 2014). The weather data of GMS over the YRB for 65 weather stations (Figure 1) were collected for model setup. The mean annual temperature in the middle and the lower Yangtze basin is 19 and 15°C from the west to east respectively (Zhang et al., 2006). The annual average streamflow of the Yangtze River is $9.5 \times 10^{11} \text{ m}^3$, and the average runoff depth is 526 mm, which is the largest streamflow in China, comprising 35% of the streamflow of all of China's rivers.

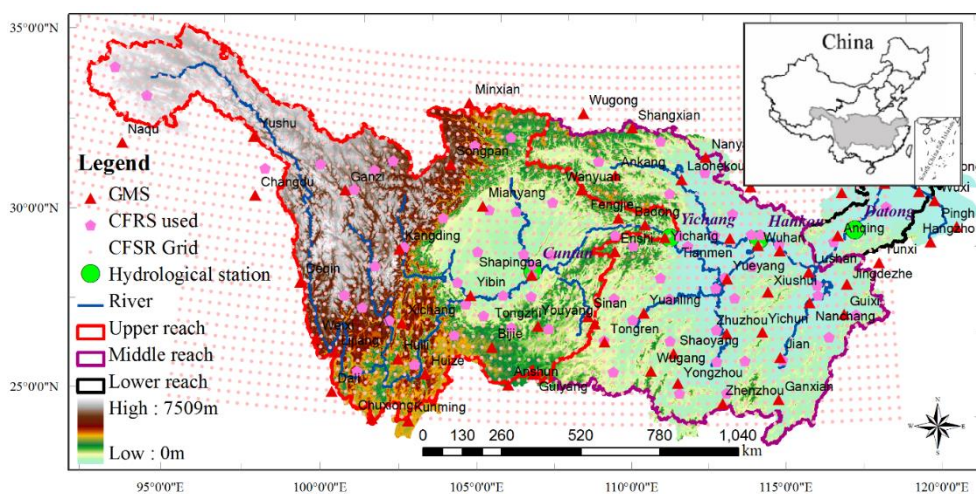


Figure 1. The Digital Elevation Model (DEM) of the YRB, its river networks, hydrological stations, Ground-based Meteorological Station (GMS) and Climate Forecast System Reanalysis (CFSR) grids. The location of the basin with respect to China is shown in the upper right panel

Datasets and model setup for the YRB

Spatial data

The spatial databases used to establish the SWAT model included the Digital Elevation Model (DEM), land use and soil type. The DEM describes the spatial distribution of regional topography and serves a fundamental role in the definition of the basin boundary, division of sub-basin, generation of river network, extraction of various hydrological parameters and division of HRUs (Tao et al., 2015; Lu et al., 2017). The DEM (Figure 1) with 90-m resolution used in this study was generated from the Shuttle Radar Topography Mission (SRTM) DEM (<http://srtm.csi.cgiar.org/SELECTION/inputCoord.asp>) processed with ArcGIS 9.3 software. Land use affects the runoff process on the land surface and has an important influence on the results of the simulation. The land use data used here were produced by the Department of Geography, University of Maryland (Hansen et al., 2000). This land use / cover products were generated using the AVHRR (Advanced Very High Resolution Radiometer) images which were analyzed to distinguish fourteen land use and cover classes with a spatial scale of 1 km pixel resolution. These images were analyzed to distinguish fourteen land use and cover classes with a spatial

scale of 1 km pixel resolution. The land use was then reclassified to nine classes for the SWAT model input (*Figure 2a*). The main land use type is forest, accounting for 36% of the area, whereas pasture (23.5%) and crop land (23%) make up the next largest proportions of the YRB.

The original China soil data that are parts of the HWSD (Harmonized World Soil Database) dataset were produced by the Institute of Soil Science, Chinese Academy of Sciences based on the Second National Soil Survey of China (Shi et al., 2006). Soil properties were extracted by 1:1,000,000 HWSD version 1.1 (*Figure 2b*). The main soil type is Haplic Luvisols with higher organic matter, Cumulic Anthrosols with human improvement of water and fertility conservation, and Gelic Leptosols with thin layer that are sensitive to the temperature with the percentage of 16.1%, 13.8%, and 11.3% respectively in the YRB.

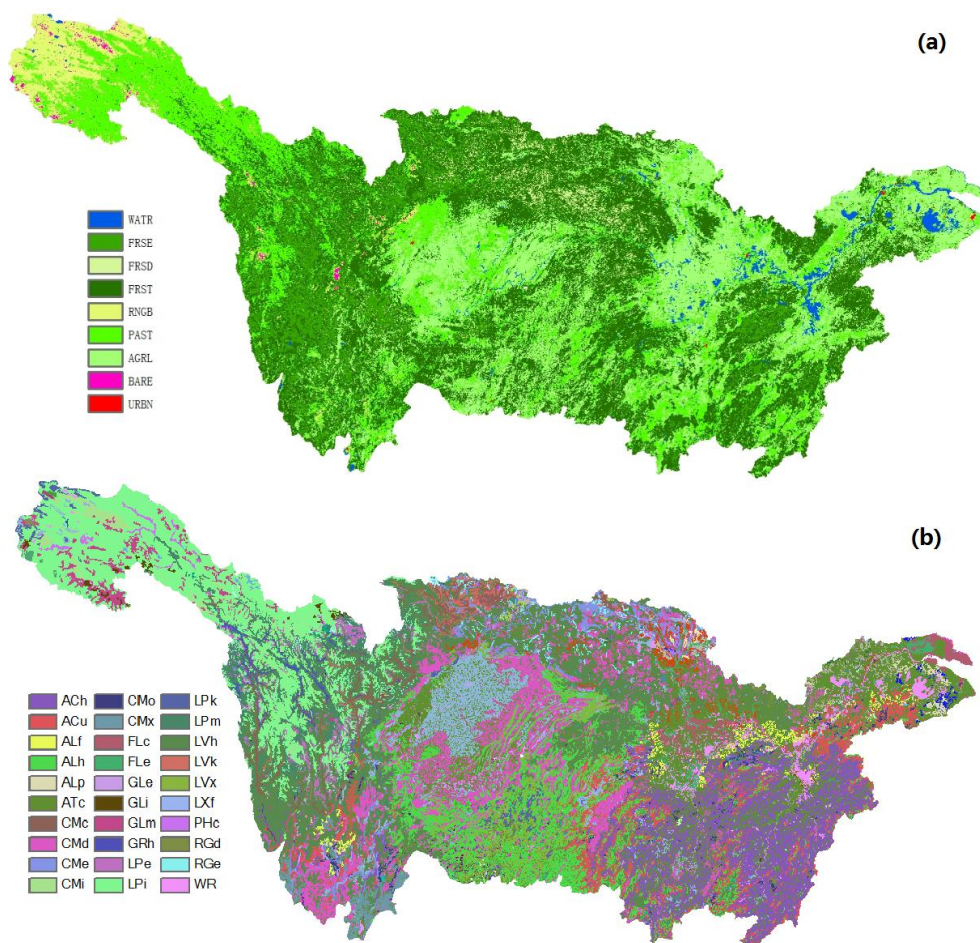


Figure 2. Land use (a) and soil type data (b) for the YRB

Hydro-meteorological data

Four main hydrological gauge stations (Cuntan, Yichang, Hankou, and Datong), which controls the runoff from different tributaries in the upper, middle and lower reaches of the YRB, were selected for measured streamflow data (*Figure 1*). Monthly streamflow were acquired for these four hydrological gauge stations for the time period 1981-2002 from the Bureau of Hydrology, Changjiang (Yangtze River) Water Resources Commission for China

(<http://www.cjh.com.cn/>). The GMS weather data covered over the YRB were downloaded for 1981 to 2005 from the China Meteorological Data Sharing Service System (<http://data.cma.cn/>). Weather data for each station included observed daily precipitation, average temperature, minimum and maximum temperature, relative humidity, wind speed and direction and solar radiation data that were all the input data required for the hydrological model.

The CFSR weather data are produced using cutting-edge data assimilation techniques, (which reflects both conventional meteorological station observations and satellite irradiances) as well as highly advanced atmospheric, oceanic, and surface-modelling components. These data are available globally at an hourly time step since 1979. The CFSR dataset consists of hourly weather forecasts generated by the NCEP Global Forecast System of the U.S. National Weather Service. The horizontal resolution of the CFSR is ~38 km (Saha et al., 2010). CFSR data (precipitation, wind, relative humidity, and solar) over the YRB for the period of 1981 to 2005 were downloaded in SWAT file format (Dile and Srinivasan, 2014) from the website (<https://globalweather.tamu.edu/>). More than 2,000 CFSR gridded points were initially imported (*Figure 1*), and data for 69 points were selected automatically for the SWAT simulations based on the sub-basin delineations.

Model setup

The SWAT model, a river basin or watershed scale eco-hydrological model, was developed by United States Department of Agriculture (USDA) and Texas A & M University (Gassman et al., 2007; Williams et al., 2008). SWAT is able to predict water, sediment, nutrient, pesticide, and fecal bacteria yields using spatial distributed data on topography, soils, land cover, land management, and weather. The hydrological parameters are calculated by SWAT model based on the minimum research units, hydrologic response units (HRUs) which consist of homogeneous slope, land cover, and soil characteristics. Since SWAT has already been widely applied to watersheds around the world for stream flow simulation, it will also present sophisticated performance on the simulation in the YRB. One of the SWAT versions ArcSWAT 2009.93.3 integrated in ESRI ArcGIS 9.3 was used in this study which is one the major releases of SWAT version 2009.

In this study, the YRB was divided into 69 sub-basins and 613 HRUs according to land use, soil type and slope characteristics in the watershed. The performance of the SWAT model for the YRB was evaluated by the following three statistical indexes: The coefficient of determination (R^2), Nash-Sutcliffe Efficiency Coefficient (E_{NS}) and Percent bias (PBias) (Moriasi et al., 2007, 2015). In the period of sensitivity analysis, the Automated Sensitivity Analysis tool integrated in SWAT was applied to the YRB, which is based on the LH-OAT (Latin Hypercube Sampling-One at A Time) analysis method (van Griensven et al., 2006). SWAT usually must be calibrated before being applied to analyze the hydrology of a given watershed, considering the complex hydrological processes and limitations of the model (Gassman et al., 2007). Therefore, following a sensitivity analysis, model calibration was carried out and was executed using external SWAT-CUP software in which the SUFI-2 (Abbaspour et al., 2007) procedure was applied to calibrate SWAT in this study.

The SWAT parameter sensitivity analysis was initially conducted as a function of the GMS weather data, to compare the predicted streamflow for each hydrological gauge station to select the most sensitive parameters for hydrology prediction at each station. In order to maintain consistency, both the GMS- and CFSR-driven models were calibrated separately within SWAT-CUP at the four hydrological gauge stations (Cuntan, Yichang, Hankou and Datong). The parameters identified as being adjusted were based on the most

sensitive aspects of predicted hydrographs in comparison with observed monthly streamflow from 1981 to 1992, and the simulated results were validated with observed data from 1993 to 2002. The calibrated parameters in the GMS simulation were then applied in the CFSR-driven SWAT model in order to compare to the result of GMS simulation at the same condition. Finally, the CFSR-driven SWAT model was completely calibrated and validated to investigate the differences of the hydrological parameters and modelling results between the GMS- and CFSR-based models.

Results

CFSR Precipitation and temperature validation by GMS weather data

The CFSR data at grid points that involved in SWAT model computation was firstly validated by the GMS data including precipitation and maximum and minimum temperature. CFSR grid points near the GMSs Huize (2110 m.a.s.l. (meters above sea level)), Mianyang (522 m.a.s.l.), Badong (334 m.a.s.l.), Ji'an (78 m.a.s.l.) and Yuanling (143 m.a.s.l.) (Their locations can be found in *Figure 1*), which represent different elevations in this area, were selected to compare the precipitation and temperature between the two data sources.

The scatter plots of the monthly precipitation and the linear regression lines at Huize, Mianyang, Badong, Ji'an and Yuanling for the GMS versus CFSR weather data from 1981 to 2005 are shown in *Figure 3*.

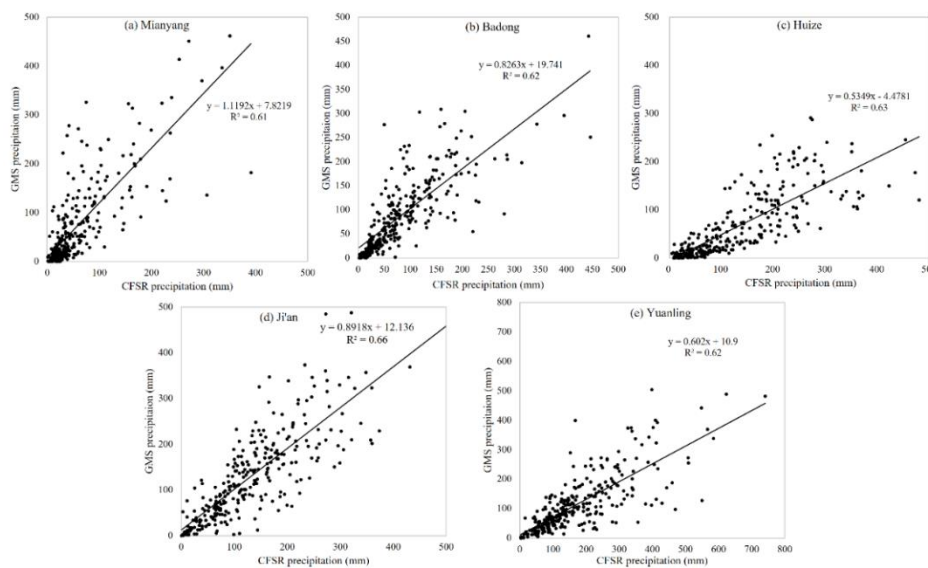


Figure 3. Comparison of monthly precipitation between the CFSR and GMS climate data sources during 1981 to 2005 at: (a) Badong, (b) Mianyang, (c) Huize, (d) Ji'an, and (e) Yuanling meteorological stations

The CFSR precipitation underestimated the GMS observed precipitation, based on the regression comparison for the Mianyang station. The precipitation was underestimated when it was < 113 mm and overestimated when > 113 mm. The most extreme underestimations of CFSR precipitation, as compared to the GMS data, occurred at the Huize station. The Huize station is located at much higher altitude than the Mianyang and

Badong stations, which may be the reason for the greater discrepancy. According to the coefficient of the regression equation the accuracy of the estimated precipitation at the Mianyang station, which is characterized by high elevation, was lower than the computed accuracy for the Ji'an station. However, the R^2 between the GMS and CFSR monthly average precipitation was larger than 0.60 at all five meteorological stations from 1981 to 2005, which showed that the CFSR weather data provide sufficient accuracy to build the hydrological model.

To validate the CFSR temperature, the monthly average minimum and maximum temperature were calculated to compare with the corresponding GMS temperature. *Figure 4* shows the comparison between the monthly average minimum and maximum temperature at Mianyang, Badong, and Huize meteorological stations and corresponding CFSR grid points during 1981 to 2005. The CFSR monthly average minimum temperature was slightly underestimated at all five meteorological stations, especially at the Badong and Huize stations, although the R^2 at both meteorological stations was > 0.96 . Both the monthly average maximum and minimum temperature was underestimated by the CFSR weather data at the Badong meteorological station, and the monthly average minimum temperature was underestimated at the Mianyang, Huize and Yuanling meteorological stations. However, the estimated temperature from the CFSR weather data was highly correlated with that from the GMS weather data, with an R^2 larger than 0.9 at all five meteorological stations.

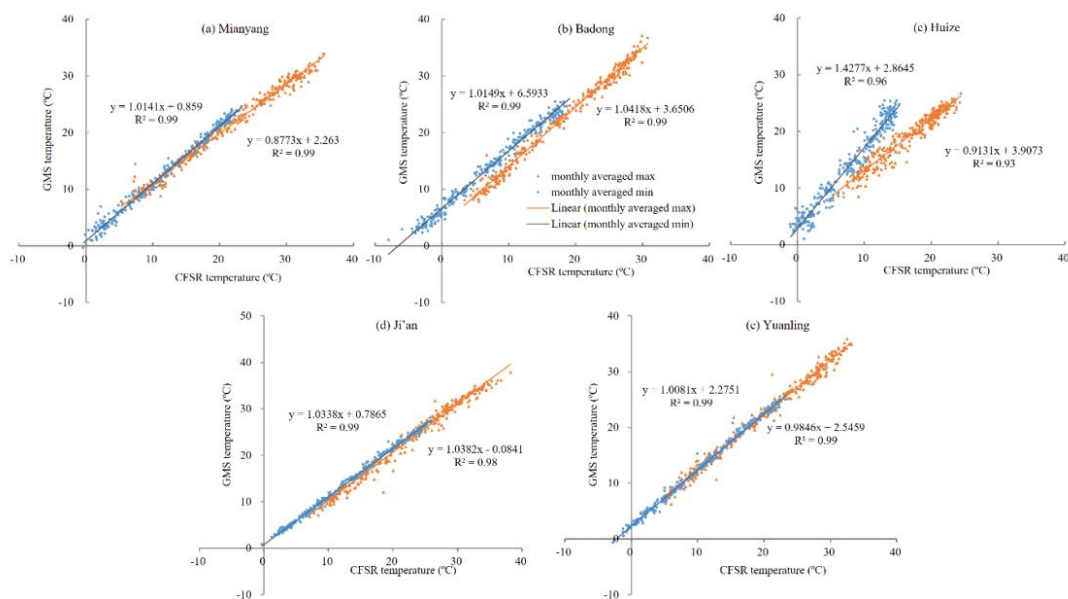


Figure 4. Comparison of the monthly average maximum and minimum temperature from the CFSR and GMS weather data at (a) Badong, (b) Mianyang, (c) Huize, (d) Ji'an and (e) Yuanling meteorological stations in the years 1981-2005

Based on the validation of the CFSR weather data with the GMS weather data at these five representative meteorological stations from the year 1981 to 2005 in the YRB, the CFSR weather data have reasonable accuracy to represent the weather conditions occurring in the watershed. Therefore, to a certain degree, the CFSR weather datasets representing the weather occurring over the watershed are able to drive hydrological model.

Parameter sensitivity analysis and calibration in the GMS- and CFSR-driven models

To decrease model uncertainty, streamflow was calibrated at each of the four gauge stations. For each simulation, with models driven by GMS and CFSR weather data, parameter sensitivity analysis and parameter calibration were conducted at each gauge station separately. After sensitivity analysis was performed, the most sensitive parameters were selected as calibration parameters at each gauge station. In this study, the following parameters were calibrated in both models using SWAT-CUP: CN2, CH_K2, ESCO, CANMX, RCHRG_DP, ALPHA_BF, GW_DELAY, HRU_SLP, SLSUBBSN, SFTMP, and CH_N2 (their descriptions can be found in *Table 1*). According to the parameter calibration by SWAT-CUP, the values of the most sensitive parameters were determined in the different models. The most sensitive parameters calibrated in each simulation at each stations and their values are listed in *Table 1*.

Table 1. Hydrological parameters calibrated during 1981 to 1992 in the SWAT models driven by GMS and CFSR weather data for simulations at different hydrological gauge stations

Parameters	Parameter description	GMS-driven model				CFSR-driven model			
		Cuntan	Yichang	Hankou	Datong	Cuntan	Yichang	Hankou	Datong
CN2	runoff curve number for moisture condition II	94.00	89.00	97.87	35.00	87.00	87.50	82.55	44.38
CH_K2	effective hydraulic conductivity in the main channel alluvium, unit: mm/h	10.12	4.02	—	—	15.91	6.57	—	12.49
ESCO	soil evaporation compensation factor	0.95	0.85	0.09	—	0.80	0.35	—	0.725
CANMX	maximum canopy storage, unit: mm	—	—	—	—	1.08	6.47	—	42.50
RCHRG_DP	deep aquifer percolation fraction	0.26	0.54	0.29	—	0.25	0.15	0.91	0.33
ALPHA_BF	baseflow alpha factor, unit: day	0.03	0.04	0.63	—	—	—	—	—
GW_DELAY	groundwater delay, unit: days	—	—	313.75	—	—	—	466.25	—
HRU_SLP*	average slope steepness, unit: m/m	—	—	—	0.48	—	—	0.53	—
SLSUBBSN*	average slope length, unit: m	—	—	—	-0.09	—	—	0.35	—
SFTMP	snowfall temperature, unit: °C	—	—	—	1.68	—	—	—	—
CH_N2	Manning's "n" value for the main channel	—	—	—	0.01	—	—	—	—

* multiply original value by (1+ the value shown in table)

Calibration and validation of the SWAT model driven by GMS weather data

The SWAT model for the YRB was firstly driven by GMS weather data. After parameter sensitivity analysis, the observed streamflow and simulated results at Cuntan, Yichang, Hankou and Datong hydrological gauge stations were presented during the calibration (1981-1992) and validation period (1993-2002) in *Figure 5*. For the calibration period, there was general agreement between the simulated and observed monthly streamflow from 1981 to 2002. From the evaluation indexes presented in *Table 2*, the R^2 are greater than 0.80, while the E_{NS} are greater than 0.70 during both calibration and validation periods. The results can be rated as “good” according to suggested performance ratings for a monthly time step (Moriasi et al., 2007, 2015). There were several months in which the peak simulated streamflow under predicted the observed peak streamflow at the Datong and Cuntan stations, respectively. In contrast, at Hankou and Cuntan stations, the peak simulated streamflow replicated the observed peak streamflow. There were fewer discrepancies of the peak streamflow between the simulated and observed data at Hankou and Datong stations than that at Cuntan and Yichang stations.

As the purpose of the study was to compare the performance of CFSR simulation in relation to conventional GMS simulation, all parameters calibrated in the GMS-driven

model were hold constant and used in the CFSR-driven model. The results from the CFSR-driven SWAT model with Calibrated parameters of GMS model (hereafter called CSWCG) were obtained. The observed streamflow was used to validate both modelling results: the results from the GMS and CSWCG simulations and the CSWCG results. The SWAT simulation with the GMS data generated good results for baseflow conditions, whereas the CFSR results replicated observed peak streamflow compared with the observed data, particularly at Hankou and Datong gauge stations. There were larger discrepancies between the observed and simulated results at Cuntan and Yichang gauge stations than at Hankou and Datong gauge stations. However, the evaluation indexes (Table 2) at Yichang gauge station were superior to that at Hankou and Datong stations in both the calibration and validation periods. From Figure 5, the simulated streamflow from CSWCG closely match peak streamflow for the lower reaches of the Yangtze River, which was demonstrated by the comparison of the simulated results and the observed streamflow at the four gauge stations in the upper, middle, and lower reaches of the Yangtze River.

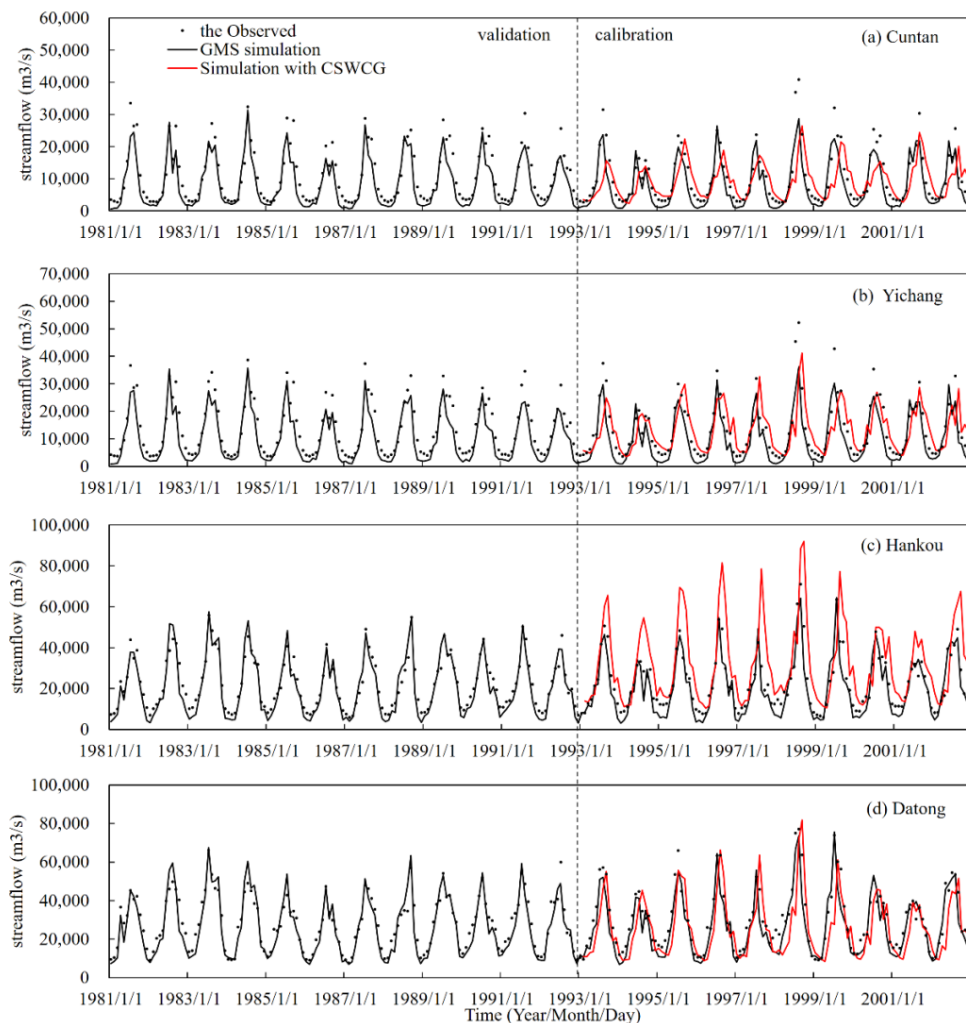


Figure 5. Calibration (1981-1992) and validation (1993-2002) of the ground-based meteorological station (GMS) simulation and comparison to the results from the CFSR-driven SWAT model with calibrated parameter of GMS model (CSWCG) using the monthly observed streamflow at (a) Cuntan, (b) Yichang, (c) Hankou and (d) Datong gauge stations

Table 2. Performance evaluation for the SWAT model driven by the ground-based meteorological station (GMS) weather data in the calibration and validation period and comparison to the results from the CFSR-driven SWAT model with calibrated parameters of GMS model (CSWCG)

	Calibration			Validation			Validation of CSWCG		
	R ²	E _{NS}	PBias/%	R ²	E _{NS}	PBias/%	R ²	E _{NS}	PBias/%
Cuntan	0.91	0.81	23.30	0.83	0.73	24.44	0.78	0.71	11.22
Yichang	0.92	0.78	25.00	0.85	0.71	27.03	0.85	0.83	1.43
Hankou	0.92	0.76	6.00	0.88	0.82	12.5	0.89	0.82	-12.01
Datong	0.89	0.74	5.00	0.88	0.82	12.6	0.88	0.81	-5.82

To explicitly present the performance of the GMS simulation in the YRB, the evaluation indexes for the SWAT model driven by GMS weather data in the calibration and validation periods and comparison to the results from the CFSR simulation with calibrated parameters of GMS model (CSWCG) are listed in *Table 2*. The R² and E_{NS} were greater than 0.80 in both the calibration and validation periods except Cuntan station. The model underestimated streamflow at Cuntan and Yichang stations while overestimated at Hankou and Datong. The comparison between the observed data and the results from the CFSR simulation showed rather high accuracy, with R² and E_{NS} larger than 0.80, except that at Cuntan gauge station. As for the values of PBias, it is slightly underestimated at Yichang gauge station and overestimated at Hankou and Datong gauge stations. The E_{NS} and R² of the validation of the simulation with GMS data are all larger than 0.80, and PBias is less than 10%; therefore, the results in this validation period can be rated as “very good” according to the research on performance ratings for a monthly time step (Moriassi et al., 2007, 2015). However, the results of validation in CSWCG at Cuntan and Hankou gauge stations are rated as “good” because their absolute values of PBias are larger than 10%. The E_{NS} value is less than 0.75 only at Cuntan station, and it is “good” at Yichang and Datong gauge stations.

Calibration and validation of the SWAT model driven by CFSR weather data

The sufficiency of the modelling performance is highly dependent on the calibrated parameters. Therefore, it is necessary to investigate the performance of the SWAT model setup with CFSR global weather data in the YRB. The CFSR weather data were used to establish a SWAT model for the YRB, which was calibrated and validated in CFSR-driven SWAT model by the discharge data at Cuntan, Yichang, Hankou, and Datong hydrological gauge stations (*Figure 6*). From the tendency of monthly streamflow comparison from 1981 to 2002, the simulated results from the CFSR-driven SWAT model fit closely to the observed data, especially at Hankou and Datong gauge stations, with E_{NS} and R² greater than 0.75. The simulated baseflow closely matched observed, and there was little discrepancy between the peak streamflow, especially in July and August of 1998, during which heavy floods occurred along the Yangtze River. The daily discharge of 87,000 m³/s, with the largest water level in history, was recorded at Datong gauge station on July 27, 1998, and exceeded the average monthly discharge of 56,800 m³/s in the wet season. From the validation of the CFSR precipitation data at Huize, Mianyang and Badong stations, the CFSR precipitation was underestimated in the period of heavier rainfall especially at the peak rainfall at Huize and Mianyang station compared with the GMS observation. The underestimated precipitation in the upper reaches with higher elevation led to

difficulty to reach the observed peak streamflow. The simulated peak streamflow nearly reached the observation at Hankou and Datong gauge stations in the middle and lower reaches of the Yangtze River. From the illustration of the *Figure 6*, the CFSR-driven SWAT model driven matches the observations in the lower streamflow period for the whole area and generates better results in the middle and lower reaches of the Yangtze River than in the upper reaches, because the CFSR precipitation has better accuracy for representing weather occurring in the middle and lower reaches than in the upper reaches of the Yangtze River basin.

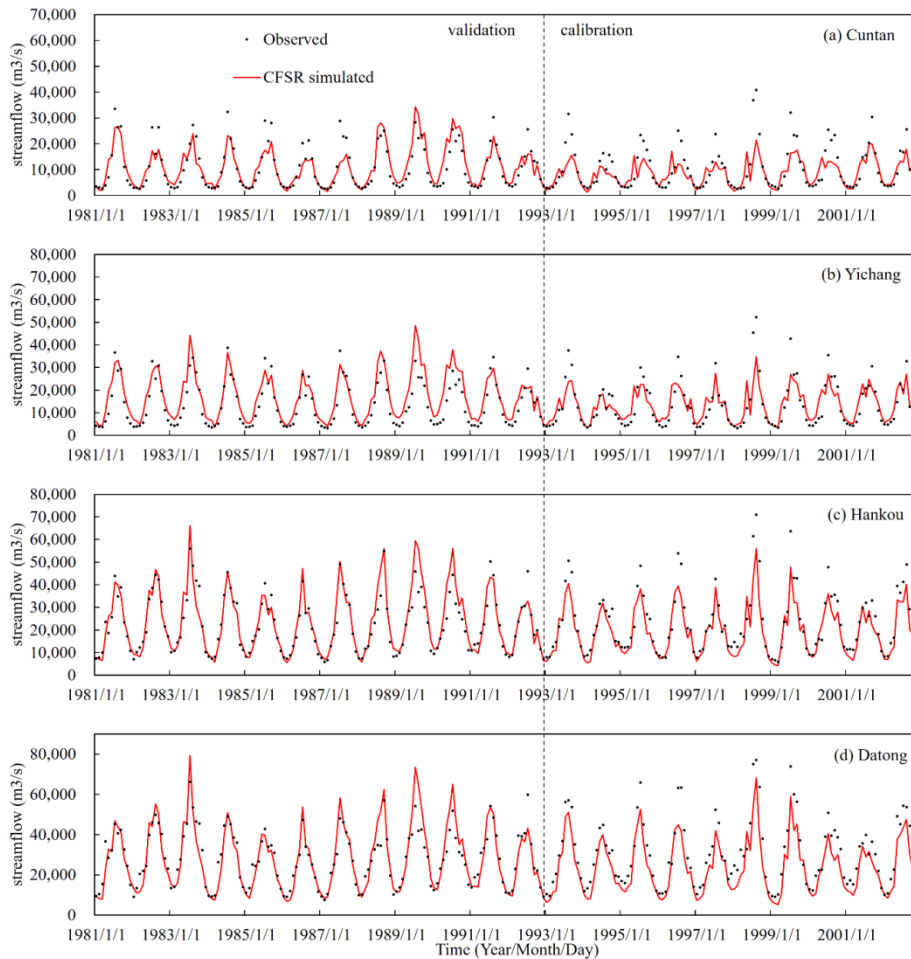


Figure 6. The Climate Forecast System Reanalysis (CFSR) driven SWAT model calibration (1981-1992) and validation (1993-2002) using monthly observed streamflow at (a) Cuntan, (b) Yichang, (c) Hankou, and (d) Datong gauge stations from 1981 to 2002

To demonstrate the model performance, *Table 3* shows the evaluation indexes of the SWAT model driven by CFSR weather data in calibration and validation at Cuntan, Yichang, Hankou, and Datong hydrological gauge stations. The R^2 and E_{NS} are larger than 0.70 in both calibration and validation periods, which are superior to the simulated result based on NCEP (National Centers for Environmental Prediction of United States) data in upper reach of Yangtze River basin by the study (Yang et al., 2014). In addition, the model tended to overestimate the streamflow, except at Datong gauge station in calibration period, and it underestimated the streamflow in validation except slightly overestimated

at Yichang station. According to performance ratings (Moriassi et al., 2007, 2015), the calibration results at all of the hydrological stations can be rated as “good” because E_{NS} and R^2 are larger than 0.75.

Comparing *Table 2* and *Table 3* for the model performances of the GMS- and CFSR-driven SWAT model calibrated and validated separately alone, the performance of the model driven by GMS weather data is better than that of the CFSR data for calibration, as well as for validation. The comparison shows the GMS-driven SWAT model is more robust than the CFSR model. However, the CFSR-driven model can also generate rather satisfactory results.

Table 3. Performance evaluation for the SWAT model driven by the Climate Forecast System Reanalysis (CFSR) weather data in calibration and validation period at different hydrological gauge stations

	Calibration			Validation		
	R^2	E_{NS}	PBias/%	R^2	E_{NS}	PBias/%
Cuntan	0.75	0.75	-1.91	0.71	0.70	14.9
Yichang	0.83	0.71	-23.6	0.76	0.72	-0.14
Hankou	0.88	0.84	-4.8	0.89	0.80	15.34
Datong	0.88	0.79	0.00	0.91	0.74	21.05

Although the R^2 of the precipitation validation was approximately 0.6 between from CFSR and GMS weather data, the CFSR-driven SWAT model produced a streamflow with R^2 and E_{NS} greater than 0.70 in both calibration and validation periods. The established SWAT model provides parameter adjustment to improve the accuracy of hydrology prediction through parameter sensitivity analysis and model calibration. The streamflow simulated with GMS weather data is better than that simulated with CFSR weather data, and the results simulated with CFSR weather data are reasonable.

From the perspective of the whole large-scale watershed, the Datong gauge station records the discharge for the whole basin. The simulation with the CFSR weather data produced better results for the watershed scale than the simulation with the GMS data based on the indexes of E_{NS} and R^2 during validation period at Hankou and Datong station. From the perspective of the sub-watershed-scale, the discharge recorded at Cuntan, Yichang and Hankou gauge stations represented the hydrological regimes of different sub-watersheds. Comparing the results in *Table 2* and *Table 3*, the GMS weather data simulation generated better results than the simulation with CFSR weather data at Cuntan and Yichang stations during both calibration and validation periods. Therefore, the GMS weather data are superior to the CFSR weather data for predicting hydrology at Cuntan and Yichang gauge stations at the sub-watershed-scale. In addition, a group of parameters calibrated in the GMS-driven model in future years with the scarce data for the Cuntan and Yichang gauge stations were proposed because the E_{NS} , R^2 and PBias in these simulations are all superior to the simulation with parameters calibrated with the CFSR model.

Comparison of the monthly average streamflow hydrograph (1981-2002) among the observed data, simulated results by CFSR- and GMS-driven model

Hydrological regime of the monthly average streamflow from 1981 to 2002 were simulated to investigate the performance of the GMS- and CFSR-driven SWAT model which were calibrated and validated separately, respectively (*Figure 7*). For Cuntan gauge

station, the simulation with GMS underestimated the streamflow during March to July, and it was a better match to the observations than the CFSR weather simulation at the recession curve from August to December. The results from the CFSR weather-driven model more closely matched the observations during January to June, whereas the GMS data simulation more closely matched the peaks than the CFSR simulation in July and better replicated the recession curve during September to December at Yichang gauge station. The GMS weather model generated better results and replicated the observations better than the CFSR model throughout the year at Hankou and Datong gauge stations.

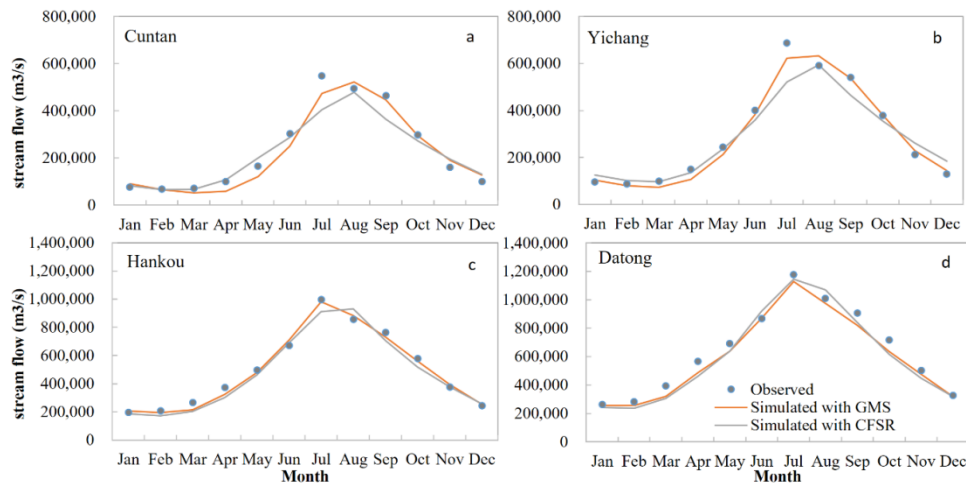


Figure 7. Average monthly streamflow hydrograph (1981-2002) for observed and simulated results from the SWAT model driven by the Climate Forecast System Reanalysis (CFSR) and the ground-based meteorological station (GMS) weather data at (a) Cuntan, (b) Yichang, (c) Datong and (d) Hankou gauge stations

From these four hydrological gauge stations, there were two patterns in the comparison curves. Firstly, there was a distinct difference in the lower streamflow between the results from the GMS and CFSR simulations, and there was less discrepancy for the peak streamflow with the GMS simulation. Secondly, the results from the GMS weather simulation more closely matched the peaks and the recession period than the CFSR weather model. From the perspective of gauge station distribution in the basin, the results from the GMS and CFSR simulations in the upper reaches of the Yangtze River are characterized as distinct difference in the lower streamflow and less discrepancy for the peak streamflow, such as the patterns at Cuntan and Yichang gauge stations, whereas the results are characterized as close match the peaks and the recession period in the middle and lower reaches of the Yangtze River, such as the patterns at Hankou and Datong gauge stations.

Discussion

From the investigation of the CFSR-driven SWAT model with Calibrated parameters of GMS model (CSWCG), although, the R^2 and E_{NS} were greater than 0.80 in both the calibration and validation periods except Cuntan station, The CSWCG resulted in acceptable results. It proves that it is feasible to apply the CFSR global data to the establishment of a SWAT model in a large-scale watershed. In addition, the CFSR and

GMS data driven SWAT model can share the same hydrological parameters to a certain extent for hydrology prediction in the YRB. That is to say, the CFSR dataset can be used to quickly established hydrology model for real time flood prediction when conventional weather data are absent in urgently, as well as the former calibrated hydrological parameters is also applicative.

Performance of the CFSR data using as hydrological model driven data is reasonable accepted, which is in agreement with the findings of Dile et al. (2014), Fuka et al. (2014), and Yang et al. (2014). However, the results found here conflict with the findings of Tan et al. (2017), who report that the original CFSR weather data is not suitable to apply for streamflow simulations in Malaysia. Differences in climate and geographical conditions are the most likely explanation for such differences between different studies. In any case, data accuracy and representativeness is more important (Yang et al., 2014). Therefore, It is important to validate the data representativeness of the weather occurred in study area before hydrological prediction.

The input weather data density is an influence, and alters the hydrologic model results. the GMSs and CFSR grids is show in *Figure 1*, in which the spatial distribution patterns of GMSs, CFSR grids and used CFSR were clearly demonstrated. The gridded data usually contain weather data of high density for hydrology simulation in the large-scale basins. Although, the density of CFSR grids is superior to the GMSs available in this study, the performances of the hydrologic simulations in this study strongly suggest the spatial density has no major influence on model simulations of this study area. Because of the automatically selected by SWAT model, the number of used CFSR data grids happened to be the same of GMS data in this study. The advantage of intensive distributions of the CFSR grids will be take effects in many large-scale and data-scarce regions.

Conclusions

In this study, the performance of CFSR global data for quickly building a hydrological model was investigated for the Yangtze River, the largest river basin in China. The study demonstrated that the CFSR weather data could modestly replicate weather conditions occurring in the YRB to a certain degree. In addition, the accuracy of the CFSR meteorological data was stronger at low elevation conditions versus high altitude. With reasonable accuracy of CFSR data validated by GMS observed weather data, it is feasible to predict hydrology with the CFSR weather datasets to drive the hydrological model of the YRB.

The R^2 and E_{NS} were larger than 0.8 in both the calibration and validation periods for the GMS weather data driven SWAT model in the YRB. However, the model showed a tendency to underestimate streamflow at all four gauge stations. Because of the reasonable accuracy of the CFSR-driven SWAT model without calibration, the CFSR- and GMS driven SWAT models can share the same hydrological parameters for hydrology prediction in the YRB, and the prediction results were better in the middle and lower reaches than in the upper reaches of the Yangtze River. Using only the CFSR global weather data, the SWAT model R^2 and E_{NS} are larger than 0.70 in both calibration and validation periods. Although the R^2 of the precipitation validation is only 0.6 between the CFSR and GMS weather data, the established SWAT model provides parameter adjustment to improve the accuracy of hydrology prediction through parameter sensitivity analysis and model calibration. Comparing the monthly average streamflow hydrograph,

the results from the GMS and CFSR simulations are characterized by two patterns in the upper, middle, and lower reaches.

This paper presented the performance of the CFSR driven hydrological model in the Yangtze River basin during 1981-2002. As we know, the Three Gorges Dam commenced impoundment since 2003, which would influence hydrology regimes of the Yangtze River. It will be an interesting topic for the further research based on CFSR driven hydrological model. Although, the CFSR-driven hydrological model generated results are less accurate than that of GMS-driven model, the CFSR global data are an alternative data source for quickly building a SWAT model for hydrology prediction in such a large-scale river basin. We can conclude that the CFSR weather data provides new opportunities to meet the challenges of driving the hydrological model in data-unavailable river basin. The CFSR weather data can be a viable option for simulating the hydrology where traditional GMS networks do not satisfied the given spatiotemporal resolution, or the GMS weather data are unavailable.

Acknowledgements. This work was funded by the National Key Research and Development Program (2017YFB0504103), the National Natural Science Funding of China (NSFC) (41331174), the Open Foundation of Jiangxi Engineering Research Center of Water Engineering Safety and Resources Efficient Utilization (OF201601), the Fundamental Research Funds for the Central Universities (2042018kf0220) and the LIESMARS Special Research Funding.

REFERENCES

- [1] Abbaspour, K. C., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J., Zobrist, J., Srinivasan, R. (2007): Spatially-distributed modelling of hydrology and water quality in the prealpine/alpine Thur watershed using SWAT. – *Journal of Hydrology* 333(2-4): 413-430.
- [2] Arnold, J. G., Srinivasan, R., Muttiah, R. S., Williams, J. R. (1998): Large area hydrologic modeling and assessment part I: Model development. – *Journal of the American Water Resources Association* 34(1): 73-89.
- [3] Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., Santhi, C., Harmel, R. D., van Griensven, A., van Liew, M. W., Kannan, N., Jha, M. K. (2012): SWAT: Model use, calibration, and validation. – *Transactions of the ASABE* 55(4): 1491-1508.
- [4] Behrangi, A., Khakbaz, B., Jaw, T. C., Agha, K. A., Hsu, K., Sorooshian, S. (2011): Hydrologic evaluation of satellite precipitation products over a mid-size basin. – *Journal of Hydrology* 397: 225-237.
- [5] Chen, J., Wu, X., Finlayson, B. L., Webber, M., Wei, T., Li, M., Chen, Z. (2014): Variability and trend in the hydrology of the Yangtze River, China: Annual precipitation and runoff. – *Journal of Hydrology* 513: 403-412.
- [6] Chen, J., Gao, C., Zeng, X., Xiong, M., Wang, Y., Jing, C., Krysanova, V., Huang, J., Zhao, N., Su, B. (2017): Assessing changes of river discharge under global warming of 1.5°C and 2°C in the upper reaches of the Yangtze River Basin: Approach by using multiple- GCMs and hydrological models. – *Quaternary International* 453(25): 63-73.
- [7] Dile, Y. T., Srinivasan, R. (2014): Evaluation of CFSR climate data for hydrologic prediction in data-scarce watersheds: an application in the Blue Nile River Basin. – *Journal of the American Water Resource Association (JAWRA)* 50(5): 1226-1241.
- [8] Ding, Y., Chan, J. C. L. (2005): The East Asian summer monsoon: an overview. – *Meteorology and Atmospheric Physics* 89(1): 117-142.

- [9] Fuka, D. R., Walter, M. T., MacAlister, C., Degaetano, A. T., Steenhuis, T. S., Easton, Z. M. (2014): Using the Climate Forecast System Reanalysis as weather input data for watershed models. – *Hydrological Processes* 28: 5613-5623.
- [10] Gassman, P. W., Reyes, M. R., Green, C. H., Arnold, J. G. (2007): The Soil and Water Assessment Tool: historical development, applications, and future research directions. – *Transactions of the ASABE* 50(4): 1211-1250.
- [11] Gassman, P. W., Sadeghi, A. M., Srinivasan, R. (2014): Application of the SWAT model special section: overview and insights. – *Journal of Environmental Quality* 43(1): 1-8.
- [12] Gong, L., Xu, C., Chen, D., Halldin, S., Chen, Y. D. (2006): Sensitivity of the Penman–Monteith reference evapotranspiration to key climatic variables in the Changjiang (Yangtze River) basin. – *Journal of Hydrology* 329: 620-629.
- [13] Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Sohlberg, R. (2000): Global land cover classification at 1 km resolution using a decision tree classifier. – *International Journal of Remote Sensing* 21(6-7): 1331-1365.
- [14] Huang, Z., Xue, B., Pang, Y. (2009): Simulation on stream flow and nutrient loadings in Gucheng Lake, Low Yangtze River Basin, based on SWAT model. – *Quaternary International* 208(1-2): 109-115.
- [15] Huang, F. Q., Huang, C. M., Di, B. F., Yin, Q. Y. (2014): Spatial and temporal soil and water loss in the lower reaches of the Jinsha River based on the SWAT model. – In: Yeh, J. (ed.) *WIT Transactions on The built Environment* 157: 259-267. WIT Press, Southampton, UK: the Advanced Civil, Urban and Environmental Engineering.
- [16] Li, Z., Yang, D., Gao, B., Jiao, Y., Hong, Y., Xu, T. (2015): Multiscale Hydrologic Applications of the Latest Satellite Precipitation Products in the Yangtze River Basin using a Distributed Hydrologic Model. – *Journal of Hydrometeorology* 16(1): 407-426.
- [17] Lu, J., Cui, X., Chen, X., Sauvage, S., Sánchez-Pérez, J. M. (2017): Evaluation of hydrological response to climate variability using SWAT model: application to the Fuhe Basin of Poyang Lake watershed, China. – *Hydrology Research* 48(6): 1730-1744.
- [18] Lu, J., Chen, X., Zhang, L., Sauvage, S., Sánchez-Pérez, J. M. (2018): Water balance assessment of an ungauged area in Poyang Lake watershed using a spatially distributed runoff coefficient model. – *Journal of Hydroinformatics* 20(5): 1009-1024.
- [19] Moriasi, D. N., Arnold, J. G., van Liew, M. W., Bingner, R. L., Harmel, R. D., Veith, T. L. (2007): Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. – *Transactions of the ASABE* 50(3): 885-900.
- [20] Moriasi, D. N., Gitau, M. W., Pai, N., Daggupati, P. (2015): Hydrologic and water quality models: performance measures and evaluation criteria. – *Transactions of the ASABE* 58(6): 1763-1785.
- [21] Ouyang, W., Hao, F., Wang, X., Cheng, H. (2008): Nonpoint Source Pollution Responses Simulation for Conversion Cropland to Forest in Mountains by SWAT in China. – *Environment Management* 41(1): 79-89.
- [22] Saha, S., Moorthi, S., Pan, H. (2010): The NCEP Climate Forecast System Reanalysis. – *Bulletin of the American Meteorological Society* 91(8): 1015-1057.
- [23] Shi, X., Yu, D., Yang, G., Wang, H., Sun, W., Du, G., Gong, Z. (2006): Cross-Reference Benchmarks for Translating the Genetic Soil Classification of China into the Chinese Soil Taxonomy. – *Pedosphere* 16(2): 147-153.
- [24] Su, B., Huang, J., Zeng, X., Gao, C., Jiang, T. (2017): Impacts of climate change on streamflow in the upper Yangtze River basin. – *Climatic Change* 141(3): 533-546.
- [25] Sun, Z., Chang, N. B., Huang, Q., Opp, C. (2013): Precipitation patterns and associated hydrological extremes in the Yangtze River basin, China, using TRMM/PR data and EOF analysis. – *Hydrological Sciences Journal* 57(7): 1315-1324.
- [26] Sun, F., Kuang, W., Xiang, W., Che, Y. (2016): Mapping water vulnerability of the Yangtze River Basin: 1994–2013. – *Environment Management* 58(5): 857-872.
- [27] Tan, M. L., Gassman, P. W., Cracknell, A. P. (2017): Assessment of three long-term gridded climate products for hydro-climatic simulations in tropical river basins. – *Water* 9: 229.

- [28] Tao, C., Chen, C., Lu, J., Sauvage, S., Sánchez-Pérez, J. M. (2015): Assessing impacts of different land use scenarios on water budget of Fuhe River, China using SWAT model. – *International Journal of Agricultural Biological Engineering* 8(3): 95-109.
- [29] van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M., Srinivasan, R. (2006): A global sensitivity analysis tool for the parameters of multi-variable catchment models. – *Journal of Hydrology* 324(1-4): 10-23.
- [30] Vu, M. T., Raghavan, S. V., Liong, S. Y. (2012): SWAT use of gridded observations for simulating runoff – a Vietnam river basin study. – *Hydrology and Earth System Sciences* 16: 2801-2811.
- [31] Ward, E., Buytaert, W., Peaver, L., Wheeler, H. (2011): Evaluation of precipitation products over complex mountainous terrain: A water resources perspective. – *Advances in Water Resources* 34: 1222-1231.
- [32] Williams, J. R., Arnold, J. G., Kiniry, J. R., Gassman, P. W., Green, C. H. (2008): History of model develop at Temple, Texas. – *Hydrological Sciences Journal* 53(5): 948-960.
- [33] Worqlul, A. W., Maathuis, B., Adem, A. A., Demissie, S. S., Langan, S., Steenhuis, T. S. (2014): Comparison of rainfall estimations by TRMM 3B42, MPEG and CFSR with ground-observed data for the Lake Tana basin in Ethiopia. – *Hydrology and Earth System Sciences* 18: 4871-4881.
- [34] Worqlul, A. W., Collick, A. S., Tilahun, S. A., Langan, S., Rientjes, T. H. M., Steenhuis, T. S. (2015): Comparing TRMM 3B42, CFSR and ground-based rainfall estimates as input for hydrological models, in data scarce regions: the Upper Blue Nile Basin, Ethiopia. – *Hydrology and Earth System Sciences Discussion* 12: 2081-2112.
- [35] Yang, Y., Wang, G., Wang, L., Yu, J., Xu, Z. (2014): Evaluation of gridded precipitation data for driving SWAT model in area upstream of Three Gorges Reservoir. – *PLoS ONE* 9(11): e112725.
- [36] Zhang, Q., Xu, C., Jiang, T., Wu, Y. (2006): Possible influence of ENSO on annual maximum streamflow of the Yangtze River, China. – *Journal of Hydrology* 333: 265-274.
- [37] Zhang, Q., Xu, C., Zhang, Z., Chen, Y. D., Liu, C., Lin, H. (2008): Spatial and temporal variability of precipitation maxima during 1960–2005 in the Yangtze River basin and possible association with large-scale circulation. – *Journal of Hydrology* 353(3): 215-227.
- [38] Zhang, N., He, H. M., Zhang, S. F., Jiang, X. H., Xia, Z. Q., Huang, F. (2011): Influence of Reservoir operation in the upper reaches of the Yangtze River (China) on the inflow and outflow regime of the TGR-based on the improved SWAT Model. – *Water Resources Management* 26(3): 691-705.
- [39] Zhou, F., Xu, Y., Chen, Y., Xu, C. Y., Gao, Y., Du, J. (2013): Hydrological response to urbanization at different spatio-temporal scales simulated by coupling of CLUE-S and the SWAT model in the Yangtze River Delta Region. – *Journal of Hydrology* 485: 113-125.
- [40] Zhu, Y., Wang, H., Zhou, W., Ma, J. (2011): Recent changes in the summer precipitation pattern in East China and the background circulation. – *Climate Dynamics* 36: 1463-1473.