Evaluation of Routed-Runoff from Land Surface Models and Reanalyses Using Observed Streamflow in Chinese River Basins

Yue MIAO1,2 and Aihui WANG*

1 Nansen–Zhu International Research Centre, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029
2 University of Chinese Academy of Sciences, Beijing 100049

(Received July 28, 2019; in final form November 26, 2019)

ABSTRACT

Previous studies have demonstrated that offline land surface models (LSMs) and global hydrological models (GHMs) can reasonably reproduce streamflow in large river basins. Global reanalyses supply fine spatiotemporal runoff estimates, but they are not fully intercompared and evaluated in China. This study assesses the routed-runoff from five offline LSM/GHM runs (VIC-CN05.1, CLM-CFSR, CLM-ERAI, CLM-MERRA, and CLM-NCEP) and three reanalysis datasets (ERAI/Land, JRA55, and MERRA-2) against the gauged streamflow (26 stations) in major Chinese river basins during 1980–2008. The Catchment-based Macro-scale Floodplain model (CaMa-Flood) is employed to route those runoff datasets to the hydrological stations. Four statistical quantities, including the correlation coefficient (R), standard deviation (STD), Nash–Sutcliffe efficiency coefficient (NSE), and relative error (RE), along with a ranking method, are used to quantify the quality of those products. The results show that the spatial patterns of both modeled and observed streamflow in summer are similar, but their magnitudes are different. Except for MERRA-2, the other products can reproduce well the interannual variability of streamflow in both the Yangtze and Yellow River basins. All products generally underestimate the magnitude and variance of monthly streamflow, while VIC-CN05.1 and JRA55 are closer to observations compared to other products. The correlation coefficients for all products are overall larger than 0.61, with the highest value (0.85) from VIC-CN05.1. In addition to CLM-MERRA, MERRA-2, and CLM-NCEP with relatively small precipitation, other products can simulate peak flow well with positive NSEs up to 0.41 (ERAI/Land). Considerable uncertainties exist among the eight products at the Yellow River outlet, which might be because the LSMS ignore frequent human activities. Based on the above statistics, performances of the eight runoff products are ranked in descending order as follows: VIC-CN05.1, ERAI/Land, JRA55, CLM-CFSR, CLM-ERAI, MERRA-2, CLM-MERRA, and CLM-NCEP, which provides a reference for flood/hydrological drought warning and hydroclimatic research in the future.

Key words: runoff, land surface model (LSM), reanalysis data, Catchment-based Macro-scale Floodplain model (CaMa-Flood), observed streamflow

Citation: Miao, Y., and A. H. Wang, 2020: Evaluation of routed-runoff from land surface models and reanalyses using observed streamflow in Chinese river basins. J. Meteor. Res., 34(1), 73–87, doi: 10.1007/s13351-020-9120-z.

1. Introduction

The climate in China is dominated by the East Asian monsoon system, and it displays clear seasonal variations, with approximately 56.5% of the total precipitation occurring in summer (Jiang et al., 2015; Yao et al., 2017). According to national statistics in 2018, a total of 39 heavy precipitation events occurred in China, and floods increased in the north while decreasing in the south compared to the previous five years. It is reported that more than 35 million people nationwide were affected by floods and associated geological hazards, 338 people died, 64,000 houses collapsed, and the direct economic losses were over 106 billion RMBs (http://www.zaihai.cn/a/zuixinzaihai/guonazaihai/2019/0109/1239.html). Streamflow is usually regarded as an indicator of

Supported by the National Key Research and Development Program of China (2016YFA0602401) and National Natural Science Foundation of China (41875106).

*Corresponding author: wangaihui@mail.iap.ac.cn.

©The Chinese Meteorological Society and Springer-Verlag Berlin Heidelberg 2020
floods and can be accurately gauged at hydrological stations. However, hydrological stations are relatively sparse in river channels, streamflow records often contain many missing data, and their length is usually too short to perform long-term analysis. For the above reasons, the simulated runoff from global hydrological models (GHMs) or land surface models (LSMs) coupled with a routing model are often used as proxies for flood and hydrological drought monitoring and forecasting and water resources management (Li et al., 2013; Wu et al., 2014; Scanlon et al., 2018; Liu et al., 2019).

To monitor floods and manage water resources, it is necessary to accurately simulate and forecast streamflow. The GHMs are designed to predict streamflow or address water scarcity concerns through empirical water balance approaches at regional or global scales, and their parameters need to be calibrated (Sood and Smakhtin, 2015). The LSMs are initially developed as the lower boundary of general circulation models (GCMs) to simulate land–atmosphere interactions (e.g., water and energy exchanges) based on physical parameterization schemes. Previous studies have indicated that GHMs and LSMs driven by the prescribed atmospheric forcings can reproduce streamflow well at multiple timescales (from monthly to interannual) (Maurer et al., 2002; Zhu and Lettenmaier, 2007; Zhang et al., 2014). For instance, Zhu and Lettenmaier (2007) discovered that simulated monthly streamflow from the offline Variable Infiltration Capacity (VIC) model was compatible with observations over relatively small river basins (less than 10⁴ km²) in Mexico. In addition to GHMs and LSMs, traditional rainfall–runoff models (RRMs) are also indispensable tools for hydrological simulations. In general, the RRM, which aim to reproduce the measured streamflow via calibration techniques, are relatively simple and widely applied to small catchments (Nasonova et al., 2009). Compared to the RRM, the GHMs and LSMs showed lower skills in simulating runoff, although they had the same precipitation input (Zhou et al., 2012; Zhang et al., 2016). However, the performances of GHMs and LSMs can be significantly improved through appropriate calibration, comparable to that of the RRM on daily and monthly timescales (Nasonova et al., 2009).

The runoff estimates in the reanalysis products [such as the Modern-Era Retrospective analysis for Research and Applications version 2 (MERRA-2) and ECMWF reanalysis interim/Land data (ERA-Interim/Land)] also have potential for flood/drought warning and water resources management. Reichle et al. (2017) assessed the runoff data from MERRA-2, MERRA-Land, MERRA, and ERA-Interim/Land with observed monthly naturalized streamflow over 18 river basins in the United States, and the results showed that the correlation coefficients of monthly streamflow anomalies from four reanalysis products were above 0.6 over most of the river basins. The results also indicated that the MERRA-2 streamflow skill was better than that of the MERRA and MERRA-Land and was comparable to that of the ERA-Interim/Land in terms of the monthly streamflow anomaly correlation coefficients. Compared to the ERA-Interim, the offline ERA-Interim/Land product is derived from the LSM with improved hydrological parameterization schemes (e.g., infiltration and runoff processes) and a bias-corrected ERA-Interim-based forcing dataset.

To monitor floods and manage water resources, it is necessary to accurately simulate and forecast streamflow. The GHMs are designed to predict streamflow or address water scarcity concerns through empirical water balance approaches at regional or global scales, and their parameters need to be calibrated (Sood and Smakhtin, 2015). The LSMs are initially developed as the lower boundary of general circulation models (GCMs) to simulate land–atmosphere interactions (e.g., water and energy exchanges) based on physical parameterization schemes. Previous studies have indicated that GHMs and LSMs driven by the prescribed atmospheric forcings can reproduce streamflow well at multiple timescales (from monthly to interannual) (Maurer et al., 2002; Zhu and Lettenmaier, 2007; Zhang et al., 2014). For instance, Zhu and Lettenmaier (2007) discovered that simulated monthly streamflow from the offline Variable Infiltration Capacity (VIC) model was compatible with observations over relatively small river basins (less than 10⁴ km²) in Mexico. In addition to GHMs and LSMs, traditional rainfall–runoff models (RRMs) are also indispensable tools for hydrological simulations. In general, the RRM, which aim to reproduce the measured streamflow via calibration techniques, are relatively simple and widely applied to small catchments (Nasonova et al., 2009). Compared to the RRM, the GHMs and LSMs showed lower skills in simulating runoff, although they had the same precipitation input (Zhou et al., 2012; Zhang et al., 2016). However, the performances of GHMs and LSMs can be significantly improved through appropriate calibration, comparable to that of the RRM on daily and monthly timescales (Nasonova et al., 2009).

The runoff estimates in the reanalysis products [such as the Modern-Era Retrospective analysis for Research and Applications version 2 (MERRA-2) and ECMWF reanalysis interim/Land data (ERA-Interim/Land)] also have potential for flood/drought warning and water resources management. Reichle et al. (2017) assessed the runoff data from MERRA-2, MERRA-Land, MERRA, and ERA-Interim/Land with observed monthly naturalized streamflow over 18 river basins in the United States, and the results showed that the correlation coefficients of monthly streamflow anomalies from four reanalysis products were above 0.6 over most of the river basins. The results also indicated that the MERRA-2 streamflow skill was better than that of the MERRA and MERRA-Land and was comparable to that of the ERA-Interim/Land in terms of the monthly streamflow anomaly correlation coefficients. Compared to the ERA-Interim, the offline ERA-Interim/Land product is derived from the LSM with improved hydrological parameterization schemes (e.g., infiltration and runoff processes) and a bias-corrected ERA-Interim-based forcing dataset.

To monitor floods and manage water resources, it is necessary to accurately simulate and forecast streamflow. The GHMs are designed to predict streamflow or address water scarcity concerns through empirical water balance approaches at regional or global scales, and their parameters need to be calibrated (Sood and Smakhtin, 2015). The LSMs are initially developed as the lower boundary of general circulation models (GCMs) to simulate land–atmosphere interactions (e.g., water and energy exchanges) based on physical parameterization schemes. Previous studies have indicated that GHMs and LSMs driven by the prescribed atmospheric forcings can reproduce streamflow well at multiple timescales (from monthly to interannual) (Maurer et al., 2002; Zhu and Lettenmaier, 2007; Zhang et al., 2014). For instance, Zhu and Lettenmaier (2007) discovered that simulated monthly streamflow from the offline Variable Infiltration Capacity (VIC) model was compatible with observations over relatively small river basins (less than 10⁴ km²) in Mexico. In addition to GHMs and LSMs, traditional rainfall–runoff models (RRMs) are also indispensable tools for hydrological simulations. In general, the RRM, which aim to reproduce the measured streamflow via calibration techniques, are relatively simple and widely applied to small catchments (Nasonova et al., 2009). Compared to the RRM, the GHMs and LSMs showed lower skills in simulating runoff, although they had the same precipitation input (Zhou et al., 2012; Zhang et al., 2016). However, the performances of GHMs and LSMs can be significantly improved through appropriate calibration, comparable to that of the RRM on daily and monthly timescales (Nasonova et al., 2009).

The runoff estimates in the reanalysis products [such as the Modern-Era Retrospective analysis for Research and Applications version 2 (MERRA-2) and ECMWF reanalysis interim/Land data (ERA-Interim/Land)] also have potential for flood/drought warning and water resources management. Reichle et al. (2017) assessed the runoff data from MERRA-2, MERRA-Land, MERRA, and ERA-Interim/Land with observed monthly naturalized streamflow over 18 river basins in the United States, and the results showed that the correlation coefficients of monthly streamflow anomalies from four reanalysis products were above 0.6 over most of the river basins. The results also indicated that the MERRA-2 streamflow skill was better than that of the MERRA and MERRA-Land and was comparable to that of the ERA-Interim/Land in terms of the monthly streamflow anomaly correlation coefficients. Compared to the ERA-Interim, the offline ERA-Interim/Land product is derived from the LSM with improved hydrological parameterization schemes (e.g., infiltration and runoff processes) and a bias-corrected ERA-Interim-based forcing dataset. Balsamo et al. (2015) compared the monthly runoff values in the ERA-Interim and ERA-Interim/Land against observed river discharge worldwide and found that the ERA-Interim/Land outperformed the ERA-Interim in terms of correlation coefficients. For instance, the correlations for monthly runoff from the ERA-Interim/Land were more than 0.7 in over half of the river basins in Asia.

Runoff is an important variable in the surface water balance equation and represents water flow out of a grid cell or at a station location. The streamflow measured at a hydrological station represents the sum of the runoff in the upstream catchment area. To compare the simulated runoff at the grid cell with the gauged streamflow at the hydrological station, a routing model is usually used. Different river routing processes have great impacts on the quantity and timing of the simulated streamflow (Sheng et al., 2017). There are many different river-routing schemes, such as the linear reservoir (Lawrence et al., 2011), the unit hydrograph and linearized Saint–Venant equation (SVE; Lohmann et al., 1996, 1998), and the 2D-diffusion wave (Senatore et al., 2015). In this study, we adopt the Catchment-based Macro-scale Floodplain model (CaMa-Flood; Yamazaki et al., 2011) based on the diffusive wave equation with backwater effect consideration in comparatively flat watersheds. The diffusive wave equation is one simplified form of the SVE, which represents river velocity explicitly (Li et al., 2013). In addition to streamflow, the CaMa-Flood model can also simulate water depth and inundation area using floodplain topography at the subgrid scale. Moreover, the computational efficiency of the CaMa-Flood model is high with the local inertial flow equation and adaptive time step scheme implemented (Yamazaki et al., 2013). The CaMa-Flood model has been successfully applied to flood risk assessment, streamflow projection under climate change, and biogeochemistry research (Pappenberger et al., 2012; Koirala et al., 2014; Lu et al., 2016).
Many studies have assessed simulated runoff from multiple models at continental or global scales to seek the differences and similarities among them (Lohmann et al., 2004; Xia et al., 2012; Beck et al., 2017). For example, Xia et al. (2012) validated and intercompared the modeled streamflow from four LSMs in the North American Land Data Assimilation System. Beck et al. (2017) evaluated runoff simulations from six GHMs and four LSMs globally in the European Union project. Nevertheless, the uncertainties of those runoff estimates have not been well evaluated in Chinese river basins. In this study, we aim to quantify the performances of simulated runoff from LSMs and reanalysis datasets against the gauged streamflow in Chinese river basins and provide a reference when choosing appropriate datasets to monitor floods/drought and perform hydroclimatic research in the future.

The remaining of the paper is organized as follows. Section 2 introduces the CaMa-Flood routing model, while Section 3 outlines runoff products, gauged streamflow data, and analysis methods. Section 4 presents the evaluation results, and conclusions and discussion are given in Section 5.

2. CaMa-Flood routing model

The CaMa-Flood model (Yamazaki et al., 2011, 2012) is used here to route the daily total runoff (surface and subsurface runoff) in each grid to hydrological stations along a prescribed drainage network. The horizontal water movement is described by a diffusive wave equation with the backwater effect considered. The floodplain inundation dynamic variables (e.g., water level, water storage, and inundation area) are modeled realistically by subgrid floodplain topography extracted from the 90-m Digital Elevation Model (DEM) and the flow direction maps (HydroSHEDS). The river width and depth parameters are estimated by empirical equations based on a climatological estimated daily runoff dataset (Kim et al., 2009). Furthermore, the river width parameter has been modified by a satellite width dataset suitable for large rivers (Yamazaki et al., 2014). In this study, the model resolution is set to $0.25^\circ \times 0.25^\circ$, which is the same as the prescribed river network. To intercompare the routed runoff from different products, the river width and depth parameters are not calibrated against the gauged streamflow. Although no calibration is performed, the characteristics of the river channels in the routing model are generally in line with real rivers in China (Fig. 1).

3. Data and methods

3.1 Runoff products

The runoff estimates from five retrospective datasets based on the offline LSMs (VIC-CN05.1, CLM-CFSR, CLM-EARAI, CLM-MERRA, and CLM-NCEP) and three reanalysis datasets [ERA-Interim/Land hereafter ERAI/Land, Japanese 55-yr reanalysis (JRA55), and MERRA-2] are used in this study. Table 1 summarizes the detailed information of the eight runoff products. Since their durations are different from each other, routing simulations are operated during the common time period of 1980–2009.

The VIC-CN05.1 runoff product is based on VIC 4.2.d, driven by a pure daily station-based atmospheric forcing dataset (precipitation, maximum and minimum temperature, and wind speed) called CN05.1 (Wu and Gao, 2013). The CN05.1 atmospheric forcing dataset was constructed on the basis of more than 2400 stations in China and has been widely used in model evaluation and...
Table 1. Summary of simulated runoff products from offline land surface models (LSMs) and reanalysis datasets

| No. | Name            | LSM Forcing | Precipitation | Resolution | Duration   | Source                      |
|-----|-----------------|-------------|---------------|------------|------------|-----------------------------|
| 1   | VIC-CN05.1      | VIC4.2      | CN05.1        | –          | 0.25° × 0.25° | 1961–2017                  |
| 2   | CLM-CFSR        | CLM4.5      | CFSR          | GPCP       | 0.5° × 0.5° | 1979–2009                  |
| 3   | CLM-ERA1        | CLM4.5      | ERAI          | GPCP       | 0.5° × 0.5° | 1979–2009                  |
| 4   | CLM-MERRA       | CLM4.5      | MERRA         | GPCP       | 0.5° × 0.5° | 1979–2009                  |
| 5   | CLM-NCEP        | CLM4.5      | NCEP–NCAR     | CRU TS     | 0.5° × 0.5° | 1979–2009                  |
| 6   | ERAI/Land       | HTESSEL     | ERAI          | GPCP       | 0.75° × 0.75° | 1979–2010                  |
| 7   | JRA55           | SiB         | JRA55         | –          | T319 (~55 km) | 1958–2012                  |
| 8   | MERRA-2         | Catchment   | MERRA-2       | CPCU, CMAP | 0.5° × 0.625° | 1980–present               |

CLM: NCAR Community Land Model; CFSR: NCEP Climate Forecast System Reanalysis; CPCU: NOAA Climate Prediction Center (CPC) unified gauge analysis; CMAP: CPC Merged Analysis of Precipitation. Other acronyms can be found in the main text of this paper.

The other four retrospective runoff products (CLM-CFSR, CLM-ERA1, CLM-MERRA, and CLM-NCEP) developed by Wang et al. (2016) were based on the Community Land Model (CLM) version 4.5, driven by the CFSR, ERA-Interim, MERRA, and NCEP atmospheric reanalysis forcing datasets with bias-corrected precipitation, respectively. The model-generated precipitation in MERRA-2 is adjusted by the Catchment LSM. Compared to MERRA (the early version), the model-generated precipitation in MERRA-2 is adjusted by the station-based CPCU product (Xie et al., 2007; Chen et al., 2008) and merged CMAP product (Xie and Arkin, 1997) over Africa within the coupled system.

3.2 Gauged streamflow

Figure 2 shows the locations of 26 hydrological stations over 9 river basins in China, at which the monthly streamflow records are available from 1980 to 2008. In this study, we concentrate on the Huai River, the Yellow River, and the Yangtze River basins because they are prone to floods. Their drainage areas are 270,000, 750,000, and 1,800,000 km², respectively. In each river basin, we select two stations with one at the upper reach of the stream (e.g., 1-Huai_Wangjiaba, 3-Yangtze_Zhimenda, and 7-Yellow_Tangnaihai in Fig. 2) and the other at the outlet (2-Huai_Bengbu, 6-Yangtze_Datong, and 8-Yellow_Huayuankou in Fig. 2) to perform extensive evaluations. Because the Yangtze River is the longest river with the largest drainage area in China, two more stations (4-Yangtze_Pingshan and 5-Yangtze_Yichang in Fig. 2) in the middle reach are also selected for detailed analysis. These stations are chosen partially because they have relatively complete streamflow observations during the study period.

3.3 Analysis methods

In addition to errors from the routing model and DEM-related parameters, the accuracy of simulated streamflow is also determined by the runoff input. Thus, eight runoff products are first evaluated by a composite runoff dataset and intercompared to show the differences among them. The spatial pattern, seasonal cycle, and in-
terannual variability in simulated streamflow are then assessed against the gauged streamflow. The mean magnitude and standard deviation (STD) of runoff are computed to depict spatial features quantitatively, while the interannual variability is described by the coefficient of variation (CV). The CV (STD/mean) reflects the dispersion degree among the datasets and facilitates intercomparison over different hydrological regimes. To quantify the performances of the eight runoff products, we calculated the correlation coefficient ($R$), standard deviation (STD), Nash–Sutcliffe efficiency coefficient (NSE), and relative error (RE) for the monthly simulated and observed streamflow at all the hydrological stations. The NSE, sensitive to high values, reveals the abilities of the models in simulating the magnitude and timing of the peak flow (Moriasi et al., 2007). The perfect simulation appears when the NSE equals 1. An NSE less than 0 indicates an unreliable simulation. The values of $R$ and normalized STD (the ratio of the simulations to observations) closer to 1 and the RE closer to 0 correspond to better performances.

To intercompare the overall performances of the eight products, a simplified ranking method is introduced (Wang and Zeng, 2012). The ranking score ranges from 1 to 8, with 1 indicating the best performance. First, the statistical metrics of each product at all hydrological stations are averaged (see columns 3–6 in Table 2). According to each averaged metric, the eight products are scored from the best (1) to the worst (8). Then, the scores of four statistical metrics are averaged as the mean score of each product (see column 7 in Table 2). Finally, the mean scores of the eight products are ranked to determine their relative performances.

4. Results

4.1 Runoff

The climatological (monthly and annual) composite

![Fig. 2. Locations of 26 hydrological stations in China with records available from 1980 to 2008. Eight stations without missing data in the Huai River, Yangtze River, and Yellow River basins are selected and labeled as 1, 2, ..., 8 (in red) in the map. The corresponding stations are referred to as 1-Huai_Wangjiaba, 2-Huai_Bengbu, 3-Yangtze_Zhimenda, 4-Yangtze_Pingshan, 5-Yangtze_Yichang, 6-Yangtze_Datong, 7-Yellow_Tangnaihai, and 8-Yellow_Huayuankou, with a naming convention of “Number-RiverName_LocationName.” Note that the stations 2-Huai_Bengbu, 6-Yangtze_Datong, and 8-Yellow_Huayuankou are in the outlet of each river.](image)

**Table 2.** The average (across eight hydrological stations) statistical quantities (NSE, RE, STD, and $R$) for eight runoff products based on the monthly streamflow observations. All $R$ values are significant at $p = 0.02$, and the numbers in the parentheses represent the ranking scores of each statistical quantity.

| No. | Name      | NSE     | RE    | STD   | $R$     | Mean score |
|-----|-----------|---------|-------|-------|---------|------------|
| 1   | VIC-CN05.1| 0.24 (4)| 0.00 (1)| 0.72 (2)| 0.85 (1)| 2.00       |
| 2   | CLM-CFSR  | 0.26 (2)| −0.19 (4)| 0.49 (4)| 0.65 (7)| 4.25       |
| 3   | CLM-ERA5 | 0.26 (3)| −0.18 (3)| 0.48 (5)| 0.61 (8)| 4.75       |
| 4   | CLM-MERRA | −0.19 (6)| −0.51 (6)| 0.26 (7)| 0.67 (6)| 6.25       |
| 5   | CLM-NCEP  | −0.53 (8)| −0.74 (8)| 0.20 (8)| 0.75 (4)| 7.00       |
| 6   | ERAI/Land | 0.41 (1)| −0.24 (5)| 0.55 (3)| 0.78 (3)| 3.00       |
| 7   | JRA55     | 0.09 (5)| 0.10 (2)| 0.87 (1)| 0.72 (5)| 3.25       |
| 8   | MERRA-2   | −0.29 (7)| −0.69 (7)| 0.39 (6)| 0.82 (2)| 5.50       |
runoff field from the Global Runoff Data Centre (GRDC) on a 0.5° × 0.5° resolution is adopted to validate those simulated runoff products. The GRDC composite runoff dataset is the combination of gauged discharge and outputs of a water balance model, and it keeps the accuracy of the observations and the spatiotemporal pattern of simulations at the same time (Fekete et al., 2002). There are 21 hydrological stations in China included in the GRDC, and the length of record varies with the stations but continuously updates. Although the GRDC is not based purely on observation, it is widely used as a reference runoff dataset for regional and global research (Wang et al., 2016; Lv et al., 2018). Figure 3 shows the mean spatial patterns of the GRDC and eight runoff products in summer (June–July–August) of 1980–2009. The magnitudes of the runoff values from all the products are relatively large in southeastern China (> 5 mm day⁻¹) and small in other regions (< 1 mm day⁻¹).

![Fig. 3. Spatial patterns of (a) the Global Runoff Data Centre (GRDC) composite runoff field and (b–i) the eight runoff products in summer during 1980–2009. The mean values (mean) and spatial standard deviations (STD) over China are indicated above each panel.](image)
of precipitation in summer during 1980–2009. Although precipitation values in the six datasets have assimilated more or less observations (in situ/remote sensing), the CN05.1 dataset is the most accurate and can be regarded as the ground truth because it was derived solely from abundant station observations in China. The spatial average precipitation and STD of the CN05.1 dataset over China are 3.53 and 2.30 mm day\(^{-1}\), respectively. Among the eight datasets, the mean and STDs of precipitation are largest in JRA55, ERAI, and GPCP. Except for the CRU (3.48 mm day\(^{-1}\)) and MERRA-2 (3.18 mm day\(^{-1}\)), the mean precipitation of the other datasets is larger than that of the CN05.1. Only MERRA-2 (2.23 mm day\(^{-1}\)) has a smaller spatial STD than that of the CN05.1. Hence, likely underestimated runoff in the CLM-NCEP and MERRA-2 datasets might result from small amounts of precipitation. Since the precipitation for the CLM-MERRA dataset has been adjusted by the GPCP, small runoff values are caused by the other forcing variables in the MERRA dataset and the LSM.

The CV (STD/mean) of the mean precipitation for six datasets is 0.06, whereas the CV of the mean runoff for eight products is 0.30. Greater consistency among the precipitation datasets than that of runoff from the eight products demonstrates that simulated runoff is also affected by other atmospheric forcing variables and model parameterization schemes. For example, the only difference between the CLM-EraI and ERAI/Land products is the LSMS, as they share the same combination of reanalysis forcing data and precipitation. The biases in the CLM-CFSR, CLM-EraI, and CLM-MERRA products with identical precipitation and LSM illustrate the effect of other meteorological forcing variables on the runoff simulations (Wang et al., 2016).

### 4.2 Streamflow

The runoff from the LSMS and reanalyses are routed to the river channels by using the CaMa-Flood model. We compare the spatial patterns of climatological simulated streamflow against observations in summer during 1980–2008 (Fig. 5). In most areas of China, the distributions of simulated and observed streamflow match well with each other. In the middle and lower reaches of the Yangtze River basin, only the VIC-CN05.1 and CLM-CFSR products can capture the magnitude of the observed streamflow (> 30,000 m\(^3\) s\(^{-1}\)). The streamflow in the Yellow River basin is obviously underestimated by the CLM-MERRA, CLM-NCEP, and MERRA-2 products (< 500 m\(^3\) s\(^{-1}\)).

The seasonal cycles of the simulated and observed streamflow during 1980–2008 are presented in Fig. 6. The performance of the simulated seasonal streamflow

![Fig. 4. Spatial patterns of climatological summer precipitation for each runoff product of (a) CN05.1, (b) GPCP, (c) CRU, (d) ERA-Interim, (e) JRA55, and (f) MERRA-2, during 1980–2009. The mean values (mean) and spatial standard deviations (STD) over China are indicated above each panel.](image-url)
varies with the source runoff products, station locations (upper/lower), and different river basins. The peak flow is significantly underestimated by most products, especially at the upstream stations in the Yangtze and Yellow River basins (Figs. 6c, g). However, the peak flow is noticeably overestimated by VIC-CN05.1 and JRA55 at Huayuankou station at the outlet of the Yellow River basin (Fig. 6h). The simulated streamflow at the upper stream stations has clear seasonal variations in peak flow in summer, while the peak timing of all the products at the downstream stations lags behind the peak observations (Figs. 6b, f, h). Except for the CLM-MERRA, CLM-NCEP, and MERRA-2 products, the seasonal cycles of the other products match well with measurements at the Yichang and Datong sites in the middle reach of the Yangtze River (Figs. 6d, e). In general, the seasonal cycles of the simulations from the VIC-CN05.1, JRA55, and ERAI/Land products track closer than the simulations from the other products to the observed seasonal cycles. The seasonal streamflow in the Huai River and Yangtze River basins are simulated better than that in the Yellow River basin.

The interannual variabilities described by the coefficients of variation (CVs) of the simulated and observed streamflow during 1980–2008 are compared in Fig. 7. In the Huai River basin, the CV of observation is only caught by the MERRA-2, VIC-CN05.1, and ERAI/Land products at Wangjiaba station (Fig. 7a) but are completely underestimated by simulations at Bengbu station (Fig. 7b). In the Yangtze and Yellow River basins (Fig. 7c–h), the CVs of the simulated streamflow agree well with the observations except for MERRA-2, which shows apparent overestimation.

Monthly streamflow values at Yichang and Datong stations in the middle and lower reaches of the Yangtze River basin during 1980–2008 are taken as an example (Fig. 8). The amplitude of the observed streamflow is largely underestimated by most simulations except that JRA55 can sometimes catch the magnitude of the peak flow at Yichang station. Obviously, the simulated
streamflow from the MERRA-2 and CLM-NECP products are much smaller than the observed streamflow, which is attributed to small input runoff. Nevertheless, the timing of the simulated peak flow from all the products is consistent with that in the observations. The normalized STDs and $R$ values between the simulated and observed monthly streamflow during 1980–2008 are summarized in the Taylor diagram (Fig. 9).
Most normalized STDs are less than 1, particularly for the CLM-NCEP and CLM-MERRA products, and most of the R values are within 0.60–0.90. The simulated streamflow at Huayuankou station in the Yellow River basin is the worst (see number 8 in Fig. 9), with all R values smaller than 0.60. Of all the products, VIC-CN05.1, with more points close to the reference point, has better performance than the other products. To quantitatively compare the performance of each runoff product, the STDs and R values are averaged for the eight hydrological stations in Table 2. In terms of the mean STD, the JRA55 product performs best (0.87), followed by the VIC-CN05.1 (0.72) and ERAI/Land (0.55), while the MERRA-2 (0.39), CLM-MERRA (0.26), and CLM-NCEP (0.20) products show poor performances. The smaller variations in the monthly simulated streamflow from the eight products may be caused by the absence of human activities in the LSMs and the limitations of the routing parameterization scheme (e.g., diffusive wave) in the CaMa-Flood model. The mean R value of the VIC-CN05.1 product is the highest at 0.85, while that of the CLM-ERAI product is the lowest at 0.61. The MERRA-2 product obtains the second largest R value of 0.82. The mean R values of CLM-MERRA (0.67) and CLM-NCEP (0.75) are not bad.

Figure 10 illustrates histograms of the RE and NSE for monthly simulated and observed streamflow during 1980–2008. The NSE values for the CLM-NCEP product at all stations are smaller than 0, which means that the CLM-NCEP product has no skills to simulate peak flow. MERRA-2 performs relatively better in the Huai River basin, with positive NSE at Wangjiaba and Bengbu stations (0.27 and 0.22). Large uncertainties exist in the NSE and RE values from the eight products at Huayuankou station. For example, most NSE values range from −1 to 1 except for VIC-CN05.1 and JRA55 at
Huayuankou station (−2.86 and −3.81). The RE values of VIC-CN05.1 and JRA55 at Huayuankou station are 1.23 and 1.17, respectively. The overestimations and negative NSEs of VIC-CN05.1 and JRA55 at Huayuankou station may result from the frequent anthropogenic activities (e.g., dams and water withdrawals) in the Yellow River basin, which are not included in the LSMs. In terms of the mean NSE in Table 2, the ERAI/Land product simulates peak flow the best with the highest NSE value of 0.41, followed by the CLM-CFSR product (0.26). Due to the large negative values at Huayuankou station, the mean NSE values for VIC-CN05.1 (0.24) and JRA55 (0.09) are no longer satisfactory. The peak flow skills of CLM-MERRA, MERRA-2, and CLM-NCEP are still poor, with a negative mean NSE. On the whole, the simulated streamflow is smaller than the observed streamflow with most RE values less than 0 (Fig. 10b). The systematic underestimation of the monthly streamflow by all products may be related to inaccurate parameters in the CaMa-Flood model, such as river width and depth (Yamazaki et al., 2012). The seemingly perfect performances of VIC-CN05.1 and JRA55 with mean RE values near 0 are unreliable because the large positive values at Huayuankou station compensate for the negative values at the other hydrological stations. The performances of CLM-MERRA (−0.51), MERRA-2 (−0.69), and CLM-NCEP (−0.74) are still unsatisfactory, with the largest three negative biases.

In summary, the overall performances of the eight runoff products are ranked based on the mean scores of four statistical quantities (Table 2), which are displayed in descending order as follows: VIC-CN05.1, ERAI/Land, JRA55, CLM-CFSR, CLM-ERAI, MERRA-2, CLM-MERRA, and CLM-NCEP.

We also rank the simulations at eight hydrological stations based on the four statistical quantities averaged for the eight runoff products (Table 3): Pingshan, Yichang, Wangjiaba, Bengbu, Huayuankou, Datong, Tangnaihai, and Zhimenda. The ranking score for Huayuankou station is unreliable because large uncertainties among the eight products counteract each other. The Yellow River basin in the semiarid and arid regions has relatively small discharge but high evaporation, and it is heavily affected by human activities (Liu et al., 2019; Wang et al., 2019). Hence, simulated streamflow at Huayuankou station in the Yellow River outlet suffers more errors from the lack of human activities in the LSMs and routing process. According to the Chinese Glacier Inventory, glacier runoff contributes to 9.2% of the streamflow above Zhimenda station in the Yangtze River (Wu et al., 2013). Wu and Gao (2013) illustrated that precipitation stations were sparse in western China, especially in the northern part of the Tibetan Plateau. Glacial meltwater supplementation that is ignored in the LSMs and inaccurate precipitation in the source region of the Yangtze River may account for poor simulations at Zhimenda station. On average, all products outperform at Pingshan and Yichang stations, which are located in the middle reach of the Yangtze
River basin with abundant water resources. With the exception of Huayuankou and Zhimenda stations, the simulations at the upstream stations are better than those at the downstream stations because of relatively fewer human activities in the upper reaches.

5. Conclusions and discussion

In this study, we aim to quantify the overall performances of eight current state-of-the-art runoff estimates in China from LSMs and reanalysis datasets. The CaMa-Flood model is adopted to route the runoff in each grid cell to 26 hydrological stations in the major Chinese rivers. Four statistical quantities (STD, R, NSE, and RE) are calculated for monthly streamflow during 1980–2008 and a ranking method is used to intercompare the qualities of the eight runoff products.

The spatial patterns of climatological modeled streamflow in summer are similar to the observed streamflow in most areas of China. However, only VIC-CN05.1 and CLM-CFSR can catch the magnitude of the streamflow in the middle and lower reaches of the Yangtze River basin, and the simulations of CLM-MERRA, CLM-NCEP, and MERRA-2 significantly underestimate the streamflow in the Yellow River basin. The seasonal cycles of the simulated streamflow at the upper stream stations are relatively better than those at the downstream stations. Among all the products, the simulated seasonal streamflow from VIC-CN05.1, JRA55, and ERAI/Land are better than that of the others. The interannual variabilities in streamflow are poorly simulated by all the products in the Huai River basin; while in the Yangtze River and Yellow River basins, except for the MERRA-2 product, the interannual variabilities in the other products are compatible with observations. The seasonal and monthly streamflow are generally underestimated by all the products, especially MERRA-2, CLM-MERRA, and CLM-NCEP. The normalized STDs of the monthly streamflow from all the products range from 0.20 to 0.87, while the R values vary within 0.61–0.85. In addition to MERRA-2, CLM-MERRA, and CLM-NCEP, other products can reasonably reproduce the peak flow well with positive NSEs.

The overall performances of the eight runoff products are ranked in descending order as follows: VIC-CN05.1, ERAI/Land, JRA55, CLM-CFSR, CLM-ERAI, MERRA-2, CLM-MERRA, and CLM-NCEP. It is expected that VIC-CN05.1 outperforms the other products because it is derived from the recent VIC model driven solely by observational atmospheric forcing data with abundant stations in China. The VIC model has demonstrated good performance in simulating hydrological processes in previous studies (Zhang et al., 2014; Xia et al., 2018). Compared to the JRA55 from the coupled assimilation system, the offline ERAI/Land benefits from more accurate precipitation and a more sophisticated LSM. JRA55 is the first reanalysis produced by the four-dimensional variational (4D-Var) assimilation method, which may explain its better performance than that of most other products (Ebita et al., 2011). The precipitation values in ERAI/Land, CLM-CFSR, CLM-ERAI, and CLM-MERRA are comparable, so the worse performances of CLM-CFSR and CLM-MERRA result from the LSM and other forcing variables. The better quality of ERAI/Land than that of CLM-ERAI indicates that the HTESSEL model might be superior to the CLM model in runoff simulations. The different skills of CLM-CFSR, CLM-ERAI, and CLM-MERRA are determined by other forcing variables in the corresponding reanalysis forcing datasets. Nearly all metrics for the MERRA-2, CLM-MERRA, and CLM-NCEP products are worse than those of the other products, except that the R values of the MERRA-2 and CLM-NCEP products are relatively high. We also find that inaccurate precipitation results in the poor performances of MERRA-2 and CLM-NCEP. Moreover, the first-generation NCEP product is worse than the more recent reanalysis atmospheric forcing datasets in atmospheric models and assimilation techniques (Rienecker et al., 2011). Compared to MERRA-2, the lower ability of CLM-MERRA with more accurate pre-

Table 3. The average (across eight runoff products) statistical quantities (NSE, RE, STD, and R) for eight hydrological stations based on the monthly streamflow observations. All R values are significant at $p = 0.02$, and the numbers in the parentheses represent the ranking scores of each statistical quantity.

| Station                | NSE   | RE     | STD    | R      | Mean score |
|------------------------|-------|--------|--------|--------|------------|
| 1-Huai_Wangjiaba       | 0.28 (2) | -0.43 (6) | 0.45 (6) | 0.73 (5) | 4.75       |
| 2-Huai_Bengbu          | 0.27 (3) | -0.17 (2) | 0.38 (7) | 0.68 (7) | 4.75       |
| 3-Yangtze_Zhimenda     | 0.03 (5) | -0.61 (8) | 0.34 (8) | 0.85 (3) | 6.00       |
| 4-Yangtze_Pingshan     | 0.39 (1) | -0.29 (4) | 0.66 (1) | 0.85 (1) | 1.75       |
| 5-Yangtze_Yichang      | 0.25 (4) | -0.32 (5) | 0.55 (3) | 0.85 (2) | 3.50       |
| 6-Yangtze_Datong       | -0.07 (7) | -0.28 (3) | 0.54 (4) | 0.72 (6) | 5.00       |
| 7-Yellow_Tangnaihai    | -0.04 (6) | -0.44 (7) | 0.46 (3) | 0.76 (4) | 5.50       |
| 8-Yellow_Huayuankou    | -0.86 (8) | 0.11 (1) | 0.59 (2) | 0.40 (8) | 4.75       |
pcipitation may be attributed to other forcing variables in the MERRA product and LSM.

Large uncertainties exist in the runoff estimates from the eight products at Huayuankou station located at the outlet of the Yellow River basin. The runoff simulations at Zhimenda station are poor partially because of unaccounted supplementation by glacial runoff in the LSMs and sparse precipitation stations in the upstream Yangtze River. With the exception of Huayuankou and Zhimenda stations, the simulations at the upper reach stations generally perform better than those at the downstream stations, which are more affected by human activities. Human activities have direct and indirect impacts on streamflow. For example, one important function of a reservoir is preventing floods through weakening the magnitude of streamflow and delaying the timing of the peak flow (Li et al., 2015). However, agricultural irrigation increases soil moisture and evapotranspiration in farmland areas and reduces the streamflow (Kang and Eltahir, 2018). Moreover, land use/cover can significantly change terrestrial characteristics and local climate and then affect the generation and routing of runoff (Gordon et al., 2005). All the runoff products in this study do not consider human activities, while the gauged streamflow is not naturalized, which leads to large biases between simulations and observations, especially in river basins with limited water and intense human impacts. In this case, VIC-CN05.1 calibrated by the naturalized streamflow can be used as a proxy to evaluate the other products. Thus, it is necessary to consider human activities and other water sources (such as snow, glaciers, and lakes) in the LSMs to obtain realistic streamflow estimates. The results of this work can provide a reference for studies based on simulated runoff, such as flood/drought warning and water management.

Acknowledgments. We are thankful for the two anonymous reviewers’ constructive comments and suggestions.

REFERENCES

Adler, R. F., G. J. Huffman, A. Chang, et al., 2003: The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present). J. Hydrometeor., 4, 1147–1167, doi: 10.1175/1525-7541(2003)004<1147:TVGP-CP>2.0.CO;2.

Balsamo, G., C. Albergel, A. Beljaars, et al., 2015: ERA-Interim/Land: A global land surface reanalysis data set. Hydrol. Earth Syst. Sci., 19, 389–407, doi: 10.5194/hess-19-389-2015.

Beck, H. E., A. I. J. M. van Dijk, A. de Roo, et al., 2017: Global evaluation of runoff from 10 state-of-the-art hydrological models. Hydrol. Earth Syst. Sci., 21, 2881–2903, doi: 10.5194/hess-21-2881-2017.

Chen, M. Y., W. Shi, P. P. Xie, et al., 2008: Assessing objective techniques for gauge-based analyses of global daily precipitation. J. Geophys. Res. Atmos., 113, D04110, doi: 10.1029/2007JD009132.

Dai, Y. J., W. Shangguan, Q. Y. Duan, et al., 2013: Development of a China dataset of soil hydraulic parameters using pedotransfer functions for land surface modeling. J. Hydrometeor., 14, 869–887, doi: 10.1175/Jhm-D-12-0149.1.

Ebita, A., S. Kohyashi, Y. Ota, et al., 2011: The Japanese 55-year reanalysis “JRA-55”: An interim report. SOLA, 7, 149–152, doi: 10.2151/sola.2011-038.

Fekete, B. M., C. J. Vörösmarty, and W. Grabs, 2002: High-resolution fields of global runoff combining observed river discharge and simulated water balances. Glob. Biogeochem. Cycles, 16, 1042, doi: 10.1029/1999gb001254.

Fekete, B. M., C. J. Vörösmarty, J. O. Roads, et al., 2004: Uncertainties in precipitation and their impacts on runoff estimates. J. Climate, 17, 294–304, doi: 10.1175/1525-7541(2004)017<0294:Upatip>2.0.CO;2.

Gao, X. J., M. L. Wang, and G. Filippo, 2013: Climate change over China in the 21st century as simulated by BCC-CSM1.1-RegCM4.0. Atmos. Ocean. Sci. Lett., 6, 381–386, doi: 10.3878/j.issn.1674-2834.13.0029.

Gelaro, R., W. McCarty, M. J. Suárez, et al., 2017: The Modern-Era Retrospective analysis for Research and Applications, version 2 (MERRA-2). J. Climate, 30, 5419–5454, doi: 10.1175/Jcli-D-16-0758.1.

Gordon, L. J., W. Steffen, B. F. Jönsson, et al., 2005: Human modification of global water vapor flows from the land surface. Proc. Natl. Acad. Sci. USA, 102, 7612–7617, doi: 10.1073/pnas.0500208102.

Huffman, G. J., R. F. Adler, D. T. Bolvin, et al., 2009: Improving the global precipitation record: GPCP version 2.1. Geophys. Res. Lett., 36, L17808, doi: 10.1029/2009gl040000.

Jiang, J., D. B. Jiang, and Y. H. Lin, 2015: Monsoon area and precipitation over China for 1961–2009. Chinese J. Atmos. Sci., 39, 722–730, doi: 10.3878/j.issn.1006-9895.1410.14191. (in Chinese)

Kang, S., and E. A. B. Eltahir, 2018: North China Plain threatened by deadly heatwaves due to climate change and irrigation. Nat. Commun., 9, 2894, doi: 10.1038/s41467-018-05252-y.

Kim, H., P. J.-F. Yeh, T. Oki, et al., 2009: Role of rivers in the seasonal variations of terrestrial water storage over global basins. Geophys. Res. Lett., 36, L17402, doi: 10.1029/2009gl039006.

Koirala, S., Y. Hirabayashi, R. Mahendran, et al., 2014: Global assessment of agreement among streamflow projections using CMIP5 model outputs. Environ. Res. Lett., 9, 064017, doi: 10.1088/1748-9326/9/6/064017.

Lawrence, D. M., K. W. Oleson, M. G. Flanner, et al., 2011: Parameterization improvements and functional and structural advances in version 4 of the Community Land Model. J. Adv. Model. Earth Syst., 3, M03001, doi: 10.1029/2011MS00045.

Li, H. Y., M. S. Wigmosta, H. Wu, et al., 2013: A physically based runoff routing model for land surface and earth system models. J. Hydrometeor., 14, 808–828, doi: 10.1175/Jhm-d-12-015.1.

Li, H. Y., L. R. Leung, A. Getirana, et al., 2015: Evaluating global streamflow simulations by a physically based routing model
coupled with the Community Land Model. *J. Hydrometeor.*, **16**, 948–971, doi: 10.1175/jhm-d-14-0079.1.

Liu, X. C., W. F. Liu, H. Yang, et al., 2019: Multimodel assessments of human and climate impacts on mean annual streamflow in China. *Hydrol. Earth Syst. Sci.*, **23**, 1245–1261, doi: 10.5194/hess-23-1245-2019.

Lohmann, D., R. Nolte-Holube, and E. Raschke, 1996: A large-scale horizontal routing model to be coupled to land surface parametrization schemes. *Tellus A*, **48**, 708–721, doi: 10.3402/tellusa.v48i5.12200.

Lohmann, D., E. Raschke, B. Nijssen, et al., 1998: Regional scale hydrology: I. Formulation of the VIC-2L model coupled to a routing model. *Hydrol. Sci. J.*, **43**, 131–141, doi: 10.1080/0262666980942107.

Lohmann, D., K. E. Mitchell, P. R. Houser, et al., 2004: Streamflow and water balance intercomparisons of four land surface models in the North American Land Data Assimilation System project. *J. Geophys. Res. Atmos.*, **109**, D07S91, doi: 10.1029/2003jd003517.

Lu, X. L., Q. L. Zhuang, Y. L. Liu, et al., 2016: A large-scale methane model by incorporating the surface water transport. *J. Geophys. Res. Biogeosci.*, **121**, 1657–1674, doi: 10.1002/2015JG003321.

Lv, M. Z., H. Lu, K. Yang, et al., 2018: Assessment of runoff components simulated by GLDAS against UNH-GRDC dataset at global and hemispheric scales. *Water*, **10**, 969, doi: 10.3390/w10080969.

Maurer, E. P., A. W. Wood, J. C. Adam, et al., 2002: A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States. *J. Climate*, **15**, 3237–3251, doi: 10.1175/1520-0442(2002)015<3237:Althbd>2.0.Co;2.

Mitchell, T. D., and P. D. Jones, 2005: An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. Climatol.*, **25**, 693–712, doi: 10.1002/joc.1181.

Moriasi, D. N., J. G. Arnold, M. W. van Liew, et al., 2007: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE*, **50**, 885–900, doi: 10.13031/2013.23153.

Nasonova, O. N., Y. M. Gusev, and Y. E. Kovalev, 2009: Investigating the ability of a land surface model to simulate streamflow with the accuracy of hydrological models: A case study using MOPEX materials. *J. Hydrometeor.*, **10**, 1128–1150, doi: 10.1175/2009jhm1083.1.

Pappenberger, F., E. Dutra, F. Wetterhall, et al., 2012: Deriving global flood hazard maps of fluvial floods through a physical model cascade. *Hydrol. Earth Syst. Sci.*, **16**, 4143–4156, doi: 10.5194/hess-16-4143-2012.

Peng, D. D., and T. J. Zhou, 2017: Why was the arid and semiarid Northwest China getting wetter in the recent decades? *J. Geophys. Res. Atmos.*, **122**, 9060–9075, doi: 10.1002/2016JD026424.

Reichle, R. H., C. S. Draper, Q. Liu, et al., 2017: Assessment of MERRA-2 land surface hydrology estimates. *J. Climate*, **30**, 2937–2960, doi: 10.1175/JCLI-D-16-0720.1.

Rienecker, M. M., M. J. Suarez, R. Gelaro, et al., 2011: MERRA: NASA’s Modern-Era Retrospective analysis for Research and Applications. *J. Climate*, **24**, 3624–3648, doi: 10.1175/JCLI-D-11-00015.1.

Scanlon, B. R., Z. Z. Zhang, H. Save, et al., 2018: Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data. *Proc. Natl. Acad. Sci. USA*, **115**, E1080–E1089, doi: 10.1073/pnas.1704665115.

Senatore, A., G. Mendicino, D. J. Goichis, et al., 2015: Fully coupled atmosphere–hydrology simulations for the central Mediterranean: Impact of enhanced hydrological parameterization for short and long time scales. *J. Adv. Model. Earth Syst.*, **7**, 1693–1715, doi: 10.1002/2015MS000510.

Sheffield, J., G. Goteti, and E. F. Wood, 2006: Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *J. Climate*, **19**, 3088–3111, doi: 10.1175/Jcli3790.1.

Sheng, M. Y., H. M. Lei, Y. Jiao, et al., 2017: Evaluation of the runoff and river routing schemes in the Community Land Model of the Yellow River basin. *J. Adv. Model. Earth Syst.*, **9**, 2993–3018, doi: 10.1002/2017ms001026.

Sood, A., and V. Smakhtin, 2015: Global hydrological models: A review. *Hydrol. Sci. J.*, **60**, 549–565, doi: 10.1002/hysj.2014.90580.

Wang, A. H., and X. B. Zeng, 2011: Sensitivities of terrestrial water cycle simulations to the variations of precipitation and air temperature in China. *J. Geophys. Res. Atmos.*, **116**, D02107, doi: 10.1029/2010JD014659.

Wang, A. H., and X. B. Zeng, 2012: Evaluation of multireanalysis products with in situ observations over the Tibetan Plateau. *J. Geophys. Res. Atmos.*, **117**, D05102, doi: 10.1029/2012JD016553.

Wang, A. H., X. B. Zeng, and D. L. Guo, 2016: Estimates of global surface hydrology and heat fluxes from the Community Land Model (CLM4.5) with four atmospheric forcing datasets. *J. Hydrometeor.*, **17**, 2493–2510, doi: 10.1175/JHM-D-16-0041.1.

Wang, Y. P., W. W. Zhao, S. Wang, et al., 2019: Yellow River water rebalanced by human regulation. *Sci. Rep.*, **9**, 9707, doi: 10.1038/s41598-019-46063-5.

Wei, S. G., Y. J. Dai, B. Y. Liu, et al., 2013: A China data set of soil properties for land surface modeling. *J. Adv. Model. Earth Syst.*, **5**, 212–224, doi: 10.1002/jame.20026.

Wu, H., R. F. Adler, Y. D. Tian, et al., 2014: Real-time global flood estimation using satellite-based precipitation and a coupled land surface and routing model. *Water Resour. Res.*, **50**, 2693–2717, doi: 10.1002/2013wr014710.

Wu, J., and X. J. Gao, 2013: A gridded daily observation dataset over China region and comparison with the other datasets. *Chinese J. Geophys.*, **56**, 1102–1111, doi: 10.6038/cjg20130406. (in Chinese)

Wu, S. S., Z. J. Yao, H. Q. Huang, et al., 2013: Glacier retreat and its effect on stream flow in the source region of the Yangtze River. *J. Geogr. Sci.*, **23**, 849–859, doi: 10.1007/s11442-013-1048-0.

Xia, Y. L., K. Mitchell, M. Ek, et al., 2012: Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *J. Geophys. Res. Atmos.*, **117**, D03109, doi: 10.1029/2011JD016048.

Xia, Y. L., D. M. Mocko, S. G. Wang, et al., 2018: Comprehens-
ive evaluation of the Variable Infiltration Capacity (VIC) model in the North American land data assimilation system. *J. Hydrometeor.*, 19, 1853–1879, doi: 10.1175/jhm-d-18-0139.1.

Xie, P. P., and P. A. Arkin, 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.*, 78, 2539–2558, doi: 10.1175/1520-0477(1997)078<2539:Gpaya>2.0.Co;2.

Xie, P. P., A. Yatagai, M. Y. Chen, et al., 2007: A gauge-based analysis of daily precipitation over East Asia. *J. Hydrometeor.*, 8, 607–626, doi: 10.1175/Jhm583.1.

Yamazaki, D., S. Kanae, H. Kim, et al., 2011: A physically based description of floodplain inundation dynamics in a global river routing model. *Water Resour. Res.*, 47, W04501, doi: 10.1029/2010wr009726.

Yamazaki, D., H. Lee, D. E. Alsdorf, et al., 2012: Analysis of the water level dynamics simulated by a global river model: A case study in the Amazon River. *Water Resour. Res.*, 48, W09508, doi: 10.1029/2012wr011869.

Yamazaki, D., G. A. M. de Almeida, and P. D. Bates, 2013: Improving computational efficiency in global river models by implementing the local inertial flow equation and a vector-based river network map. *Water Resour. Res.*, 49, 7221–7235, doi: 10.1002/wrcr.20552.

Yamazaki, D., F. O’Loughlin, M. A. Trigg, et al., 2014: Development of the global width database for large rivers. *Water Resour. Res.*, 50, 3467–3480, doi: 10.1002/2013wr014664.

Yao, S. B., D. B. Jiang, and G. Z. Fan, 2017: Seasonality of precipitation over China. *Chinese J. Atmos. Sci.*, 41, 1191–1203, doi: 10.3878/j.issn.1006-9895.1703.16233. (in Chinese)

Zhang, X. J., Q. H. Tang, M. Pan, et al., 2014: A long-term land surface hydrologic fluxes and states dataset for China. *J. Hydrometeor.*, 15, 2067–2084, doi: 10.1175/Jhm-D-13-0170.1.

Zhang, Y. Q., H. X. Zheng, F. H. S. Chiew, et al., 2016: Evaluating regional and global hydrological models against streamflow and evapotranspiration measurements. *J. Hydrometeor.*, 17, 995–1010, doi: 10.1175/jhm-d-15-0107.1.

Zhou, X. Y., Y. Q. Zhang, Y. P. Wang, et al., 2012: Benchmarking global land surface models against the observed mean annual runoff from 150 large basins. *J. Hydrol.*, 470–471, 269–279, doi: 10.1016/j.jhydrol.2012.09.002.

Zhu, C. M., and D. P. Lettenmaier, 2007: Long-term climate and derived surface hydrology and energy flux data for Mexico: 1925–2004. *J. Climate*, 20, 1936–1946, doi: 10.1175/JCLI4086.1.

Tech & Copy Editor: Zhirong CHEN