Large uncertainties in runoff estimations of GLDAS versions 2.0 and 2.1 in China

Wei Qi[1], Junguo Liu[1*], Hong Yang[2, 3], Xueping Zhu[4], Yong Tian[1], Xin Jiang[5], Xu Huang[6], Lian Feng[1]

[1] School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China

[2] Eawag, Swiss Federal Institute of Aquatic Science and Technology, Duebendorf, Switzerland

[3] Department of Environmental Science, MGU, University of Basel, Basel, Switzerland

[4] College of Water Resources Science and Engineering, Taiyuan University of Technology, Taiyuan 030024, China

[5] Water Resources Research Institute of Shandong Province, Jinan 250013, China

[6] Bureau of Hydrology (Information Center), Songliao Water Resources Commission, Jilin 130021, China

[*] Email: junguo.liu@gmail.com; liujg@sustech.edu.cn

Highlights:

Both GLDAS2.0 and GLDAS2.1 have large uncertainties, but GLDAS2.1 is better: absolute values of relative bias are above 39%

In the Tibetan Plateau, absolute value of relative bias is 64% for GLDAS2.0 and is larger than in the middle and northeast China

Abstract:

Gauge observed runoff can reflect influences of both natural hydrological cycle and human intervention. The Global Land Data Assimilation System (GLDAS) 2.0 and 2.1 provide abundant runoff which are useful for water resources assessment in ungauged/poorly gauged regions. However, GLDAS2.0 and GLDAS2.1 runoff have only been validated and inter-compared in very limited regions. In this study, they are evaluated and inter-compared utilizing gauge observation in 11 large river basins in China. Results show their runoff have large uncertainties: absolute values of Relative Bias (|RB|) being above 39% and Nash-Sutcliffe Efficiency lower than 0.15 on average, but GLDAS2.1 is better. Both of them have large uncertainty in the Tibetan Plateau: |RB| are higher than 40%. The gap between GLDAS runoff and observations could attribute to both GLDAS system uncertainty and the fact that GLDAS does not consider human intervention. Therefore, cautions should be taken when using them in coupled human-natural systems.

Key words: China; GLDAS; Runoff; The Tibetan Plateau; Uncertainty

Plain Language Summary:

Runoff data of the Global Land Data Assimilation System (GLDAS) 2.0 and 2.1 have only been validated and inter-compared in very limited regions. We evaluated and inter-compared GLDAS2.0 and GLDAS2.1 runoff data utilizing gauge observation in 11 large river basins in China. We found they have large uncertainties, especially in the Tibetan Plateau. The uncertainty may result from the system itself, such as forcing data and models used. In addition, human intervention may also contribute to the uncertainty, because the system did not consider human influence. We suggest that cautions must be taken when using them.

1 Introduction

Water resources management in coupled human-natural systems is of importance to human well-being. Human activities can influence natural hydrological cycle, and natural systems can provide service to human activities. With rapidly increasing population in the anthropogenic era, the interactions between human and natural systems becoming stronger than ever before [Liu et al., 2007].

Runoff is one of the most integrative indicators of basin scale hydrology [Liang et al., 1994; Chen et al., 1996; Sellers, 1997; Soroosh et al., 2005; Zaitchik et al., 2010; Qi et al., 2015; Wang et al., 2017; Qi et al., 2020]. The Global Land Data Assimilation System (GLDAS) [Rodell et al., 2004] provide abundant runoff data which is useful for water resources assessment in ungauged/poorly gauged regions. GLDAS estimates runoff on the basis of global Land Surface Models (LSMs) and global scale forcing data, such as the Princeton Global meteorological Forcing (PGF) dataset [Sheffield et al., 2006]. Currently, the LSMs used in GLDAS do not consider human intervention [Scanlon et al., 2018], such as human water abstraction. Therefore, the datasets in GLDAS could be problematic when studying coupled human-natural systems. Hydrological gauge observations are results of natural hydrological processes with influences from human activities. Investigating the gap between GLDAS runoff and hydrological gauge observations could provide valuable information on the strengths and weaknesses of GLDAS systems under coupled human-nature systems; thereby can provide information on where cautions should be taken when using them and guidance for future development of the systems.

The quality of GLDAS1.0 data has been investigated in many studies [Kato et al., 2007; Zaitchik et al., 2010; Wang et al., 2011; Chen et al., 2013; Qi et al., 2015; Lv et al., 2018]. For

example, Qi et al. [2015] evaluated GLDAS1.0 input data in Northeast China, and found GLDAS1.0 overestimated downward solar radiation; Bai et al. [2016] evaluated runoff of GLDAS1.0 in the Tibetan Plateau, and found Noah model overestimated runoff. Regarding GLDAS version 2, studies on its quality remain few. For example, Wang et al. [2016a] assessed the soil temperature estimation of GLDAS2.0 and found good agreement with in situ measurement; Wang et al. [2016b] studied the applicability of GLDAS2.0 in terms of precipitation, evapotranspiration, air temperature, water storage and runoff, and found that runoff is underestimated. For GLDAS2.1, Lv et al. [2018] compared GLDAS2.1 runoff with the University of New Hampshire and Global Runoff Data Centre composite dataset, and found considerable uncertainties. However, studies to investigate GLDAS2.1 uncertainty in runoff remain limited on large scales on the basis of hydrological gauge runoff observation. In addition, studies to compare runoff estimation of GLDAS2.0 and GLDAS2.1 based on hydrological gauge observation on large scales also remain limited.

The overall objective is to evaluate and inter-compare the GLDAS2.0 and GLDAS2.1 runoff data on the basis of hydrological gauge observation from 2000 to 2010 (their only overlapping period) in China. The investigated GLDAS data are GLDAS_NOAH025_3H_2.0 [Matthew and Hiroko Kato, 2015] and GLDAS_NOAH025_3H_2.1 [Matthew and Hiroko Kato, 2016]. The GLDAS2.0 and GLDAS2.1 runoff are calculated on the basis of the Noah land surface model (Noah 3.3), considering vegetation transportation, surface water and energy balance, snow melt water, etc. The GLDAS system uncertainty refers to the input data uncertainty and model uncertainty hereafter. Three regions with different hydro-climate characteristics are chosen for the analysis, i.e., the Tibetan Plateau region, Middle China and Northeast China. Uncertainties in the GLDASs data and the possible factors causing the uncertainties are discussed. Some cautions in the use of the GLDASs data in China are also discussed. This

study provides unique insights into the strengths and weaknesses of GLDAS2.0 and GLDAS2.1 runoff data in China.

2 Study regions and methods

In this study, three regions of China are investigated: Northeast China, Middle part of China (Middle China) and the Tibetan Plateau. They cover an area of 3,511,244 km^2 in total. The selected basins in the three regions are shown in Fig. 1. Northeast China has a long winter and warm summer, with multi-year mean monthly temperature being 4.26 degrees Celsius. The basins in Tibetan Plateau have high altitude (above 4,000 meters on average) and multi-year mean monthly temperature is around zero degree Celsius. The Middle China covers a large area with various land covers, e.g., forest, desert, short vegetation, etc. More details of the study regions and data are shown in Table 1, including basin area, mean altitude, mean annual rainfall, mean air temperature and temporal coverage of monthly runoff which is from hydrological gauge observation.

Runoff in a grid (q_{grid}) is calculated as the sum of GLDAS overland flow, interflow and groundwater

$$q_{grid} = q_{sur} + q_{g+inter} \tag{1}$$

where q_{sur} represents overland flow (m^3/s); $q_{g+inter}$ represents sum of interflow and groundwater (m^3/s). In this study, because the time of runoff concentration is less than one month [Wang et al., 2016b; Allen et al., 2018], the monthly runoff of GLDAS2.0 and GLDAS2.1 at the gauging sites are calculated by summing up the upstream monthly runoff

$$q_{gauge} = \sum q_{grid} \tag{2}$$

where q_{gauge} represents the calculated monthly runoff at gauging sites. This routing method is commonly used in previous studies (e.g., Crooks et al. [2014]; Li et al. [2015]; Wang et al. [2016b]), because it is a nonparametric approach (as shown in Eq. 2) and therefore does not introduce parameter uncertainty. Therefore, it is acceptable to use this approach in this study.

Three evaluation criteria are used: Correlation Coefficient (CC), Relative Bias (RB) and Nash-Sutcliffe Efficiency (NSE) because they are commonly utilized in uncertainty evaluations especially for runoff [Beven and Binley, 1992; Beven and Freer, 2001; Wang et al., 2011; Qi et al., 2016a; Qi et al., 2016b; Qi et al., 2016c; Yang et al., 2017; Qi et al., 2018; Qi et al., 2019]:

$$CC = \frac{\sum_{i=1}^{n}\left(X_{oi}-\overline{X_o}\right)\cdot\left(X_{si}-\overline{X_s}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_{oi}-\overline{X_o}\right)^2\cdot\sum_{i=1}^{n}\left(X_{si}-\overline{X_s}\right)^2}} \tag{3}$$

$$NSE = 1-\frac{\sum_{i=1}^{n}\left(Q_{pi}-Q_{ti}\right)^2}{\sum_{i=1}^{n}\left(Q_{ti}-\overline{Q_t}\right)^2} \tag{4}$$

$$RB = \frac{\sum_{i=1}^{n}Q_{pi}-\sum_{i=1}^{n}Q_{ti}}{\sum_{i=1}^{n}Q_{ti}}\times100\% \tag{5}$$

where Q_{pi} and Q_{ti} represent simulated and observed data respectively at time i; $\overline{Q_t}$ represents average of observed data; n represents the total number of data points. The higher the CC and NSE values, the better the GLDAS data performance in replicating monthly variations. To compare with uncertainty in runoff, the three criteria are also utilized for rainfall evaluations. The rainfall data used in this study was generated based on over 2,400 in situ rain gauges in China with a resolution of $0.5^\circ \times 0.5^\circ$ by the National Meteorological Information Center (http://data.cma.cn/data/) [Zhao and Zhu, 2015; Shen and Xiong, 2016]. In this study, we regrid

the rainfall data into $0.25° \times 0.25°$ grids which are the same resolutions as GLDAS2.0 and GLDAS2.1 girds, and the data located inside the study regions was utilized.

3 Results and discussion

3.1 Relative Bias

Fig. 2 shows the evaluations of runoff and rainfall in terms of RB. GLDAS2.0 greatly overestimates runoff in the Liao river (21%), and underestimates in other regions with the exception of the Yarlung Tsangpo river. In the Upper Yellow river, GLDAS2.0 has the highest uncertainty with RB being up to -93%, whereas it has the lowest uncertainty in the Yarlung Tsangpo river (1%). The average of absolute values of RB is 53% for GLDAS2.0. Therefore, GLDAS2.0 runoff estimation has large uncertainty. Similarly, GLDAS2.1 greatly overestimates runoff in the Liao river with RB being 30%, and it largely underestimates runoff in the Upper Yellow river with RB being -66%. The RB value is 26% in the Yarlung Tsangpo river, which is much higher than the value for GLDAS2.0. In addition, similar to GLDAS2.0, GLDAS2.1 underestimates runoff in most of the basins with the exceptions of the Liao river, Middle Yellow river and Yarlung Tsangpo river. The average of absolute values of RB is 39% for GLDAS2.1. Therefore, GLDAS2.1 is closer to observation than GLDAS2.0 on average but still has large uncertainty. In addition, based on the runoff evaluation results, the absolute values of RB are 9% for GLDAS2.1 in the middle Yellow river and Lancang river, and the RB value is 1% for GLDAS2.0 in the Yarlung Tsangpo river. Therefore, GLDAS2.1/GLDAS2.0 could be used in these rivers if the total volume of runoff is of interest. Root Mean Square Error (RMSE) values are shown in supporting information Table S1 and S2. The RMSE values are the largest in the Datong station on both monthly and yearly scales. The differences between Fig. 2 and supporting information Table S1 and S2 are because the uncertainty criteria investigated are different.

Regarding RB for rainfall (Fig. 2b), GLDAS2.0 overestimates rainfall with RB being 27% in the Yarlung Tsangpo river, whereas it seriously underestimates rainfall in Heihe with RB being -65%. The average of absolute RB values is 23% for GLDAS2.0. Similarly, GLDAS2.1 seriously overestimates rainfall in the Yarlung Tsangpo river and underestimate rainfall in the Heihe. The average of absolute RB values of GLDAS2.1 is 20%, which is lower than that of GLDAS2.0. Therefore, GLDAS2.1 is better than GLDAS2.0 on average, which is the same as the results of runoff evaluations. Differences between the two version datasets also exist. For example, GLDAS2.0 underestimates rainfall in most basins except in the Yarlung Tsangpo and Liao rivers. However, GLDAS2.1 overestimates rainfall in most basins except in the Upper Yangtze river and Heihe. It should be noted that, although GLDAS2.1 underestimates rainfall in the Middle Yellow river, the RB value is very small (-1%).

Comparing the evaluations between runoff and rainfall, GLDAS2.0 underestimates runoff when it underestimates rainfall. However, GLDAS2.1 underestimates runoff even though it overestimates rainfall in the Lancang river, Lower and Upper Songhua river, Middle and Lower Yangtze river and Upper Yellow river, and it overestimates runoff even though it underestimates rainfall a little in the Middle Yellow river. These uncertainty may be caused by both the rainfall uncertainty and Noah model uncertainty (including model structures and parameterization) [Swenson et al., 2012; Chen et al., 2013; Yang et al., 2018]. Niu et al. [2011] pointed out that Noah model may result in large uncertainty in surface hydrological processes (including the vegetation photosynthesis and transpiration processes, and groundwater simulation). The Noah model uses a combined surface layer of vegetation and soil surface which cannot explicitly compute photosynthesis and transpiration processes. The Noah model uses a shallow soil column and free drainage scheme at the bottom of the soil column, which could lead to large

uncertainty in groundwater simulation. In addition, interactions between Noah model uncertainty and rainfall data uncertainty may also influence the results as shown in the study by Qi et al. [2016a] in which it is revealed that the interactions can contribute to a similar magnitude of uncertainty as models. Further, human activities (such as irrigation, domestic and industrial water use) can consume large amount of water resources [Tang et al., 2008], which may also contribute the difference between GLDAS simulation and hydrological gauge observation because LSMs used in GLDAS do not consider human activities in runoff simulation. For example, the water resources utilization percentage for irrigation, household and industry is higher than 40% according to the water resources reports of the Songliao Water Conservancy Committee of the Ministry of Water Resources of China (http://www.slwr.gov.cn/szy2011/); in the Yellow river, water resources utilization rate is also above 20% (http://www.yrcc.gov.cn/other/hhgb/), in the Heihe river, it is above 50% [Wu et al., 2005]. The four river basins on the Tibetan Plateau are covered by a large amount of snow and glacier. However, the Noah model used in GLDAS2.0 and GLDAS2.1 does not include a glacier sheet model, which may also contribute to the uncertainty in the simulated runoff.

3.2 Nash-Sutcliffe Efficiency

Fig. 3 shows the evaluations of runoff and rainfall in terms of monthly scale NSE. GLDAS2.0 runoff has the highest NSE in Yarlung Tsangpo river (0.56), and the lowest value in the Upper Yellow river (-1.39). The average NSE is -0.21 for GLDAS2.0. Similarly, GLDAS2.1 runoff also has the lowest NSE in the Upper Yellow river (-0.41); but its highest NSE value is in the Lancang river (0.66), which is different from GLDAS2.0. The average NSE is 0.15 for GLDAS2.1. Therefore, GLDAS2.1 is better than GLDAS2.0 on average, but the NSE value still shows very low accuracy. Because GLDAS2.0 and 2.1 use the same LSM (i.e., Noah 3.3), the differences here could attribute to their rainfall uncertainty, model uncertainty [Barlage et

al., 2010] and interactions between the model uncertainty and rainfall data uncertainty.

Regarding the rainfall estimation (Fig. 3b), for GLDAS2.0, the NSE values are the highest in the Middle and Lower Yangtze river (0.95), and NSE is the lowest in the Heihe (0.14). The average NSE is 0.71 for GLDAS2.0. Similarly, for GLDAS2.1, NSE is the highest in the Middle Yangtze river (0.97), where the NSE value is the same as the Middle Yellow river. The average NSE is 0.82 for GLDAS2.1, which is better than GLDAS2.0. Comparing with the results in Fig. 2, GLDAS2.0 and GLDAS2.1 have large uncertainties in runoff estimation in terms of both RB (above 39%) and NSE (lower than 0.15), but GLDAS2.1 is better than GLDAS2.0. This result implies that GLDAS2.1 should be preferable to GLDAS2.0 in water related research, and more confidence should be given to analysis results using GLDAS2.1 than GLDAS2.0.

3.3 Time series analysis

Runoff data of GLDAS2.0 and GLDAS2.1 are compared with observation in Fig. 4 in time series plots. GLDAS2.0 and GLDAS2.1 underestimate most of the peak value in the Lower and Upper Songhua river, Middle and Upper Yangtze river, Upper Yellow River and Heihe. In the Upper Yangtze river, Upper Yellow River and Heihe, the underestimation may result from the lower rainfall estimation (Supplementary Information Fig. S1g, h and i). In addition, the human activities also have large influence on river runoff in the Heihe [Zang and Liu, 2013], and therefore the runoff uncertainty of GLDAS2.0 and GLDAS2.1 in the Heihe could also attribute to the fact that GLDAS2.0 and GLDAS2.1 do not consider human activities influence. In the other basins, it may be due to model uncertainty and interactions between rainfall data uncertainty, Noah model uncertainty and human intervention, because the peak rainfall estimation is not largely underestimated (Supplementary Information Fig. S1a, b and e). In the

Lancang river, GLDAS2.0 underestimates all the peak runoff, which may be because GLDAS2.0 underestimates all the peak rainfall (Supplementary Information Fig. S1j). Compared with GLDAS2.0, peak runoff of GLDAS2.1 replicates observation relatively well in the Lancang river. In the Lower and Middle Yangtze river and the Yarlung Tsangpo river, both GLDAS2.0 and GLDAS2.1 could replicate runoff seasonal variations well. However, in other basins, the CC values are low. The average CC values are 0.59 and 0.63 for GLDAS2.0 and GLDAS2.1, respectively. Thus, GLDAS2.1 is better than GLDAS2.0.

In the middle and lower Yangtze river and the middle Yellow river, the rainfall data of GLDAS2.0 and GLDAS2.1 replicate observation well, with NSE and CC being above 0.88 and 0.97, respectively. In the middle and lower Yangtze river, the absolute values of RB are less than 9% for GLDAS2.0 and GLDAS2.1. In the middle Yellow river, RB of GLDAS2.1 is -1%, and RB of GLDAS2.0 is -18%. Therefore, the rainfall data may not be a main uncertainty source of runoff in the middle and lower Yangtze river. In the middle Yellow river, rainfall uncertainty may be not a main uncertainty source of runoff for GLDAS2.1, but it may be a main source for GLDAS2.0.

As pointed out by Qi et al. [2016a], models can amplify/decrease input data uncertainty in the uncertainty propagation chain from input to model output because of imperfect/incomplete model structures, which is often termed as interactions/non-linear uncertainty propagation [Bosshard et al., 2013]. For example, the monthly rainfall from GLDAS2.1 is very close to the observation in the Upper Yellow river (Fig. S1), but the monthly runoff from GLDAS2.1 is much lower than observation (Fig. 4); in the Lower and Upper Songhua river, the monthly rainfall from GLDAS2.1 are higher than observation (Fig. S1), but the monthly runoff from GLDAS2.1 is much lower than observation (Fig. 4). In this kind of cases, the results may imply

that models and interactions between models and input data have more significant influence on output than input data uncertainty.

3.4 Regional average analysis

Fig. 5 shows regional averages of NSE and absolute values of RB. Absolute RB of GLDAS2.0 runoff is higher in the Tibetan Plateau (64%) than in Middle and Northeast China. Different from GLDAS2.0, absolute RB value of GLDAS2.1 is the highest in Northeast China (43%), and is 40% in the Tibetan Plateau (which is still very high). These results indicate that both GLDAS2.0 and GLDAS2.1 have large uncertainty in the Tibetan Plateau. Similarly, absolute RB for rainfall also show the uncertainty is the highest in the Tibetan Plateau (Fig. 5c) with absolute RB being 36% and 24% for GLDAS2.0 and GLDAS2.1, respectively.

Regarding NSE, Fig. 5b shows NSE of GLDAS2.0 is the lowest in the Tibetan Plateau (-0.46), which is similar to rainfall (Supplementary Information Fig. S1d): NSE of GLDAS2.0 rainfall is also the lowest in the Tibetan Plateau (0.60). Different from GLDAS2.0, NSE of GLDAS2.1 runoff is the highest in the Tibetan Plateau. Although NSE of GLDAS2.1 runoff is higher than GLDAS2.0 in the Tibetan Plateau, but still shows very low accuracy with a value of only 0.30 (Fig. 5b). For GLDAS2.1, different from runoff, NSE of rainfall is the lowest in the Tibetan Plateau (0.73 in Fig. 5d). Overall, both GLDAS2.0 and GLDAS2.1 have large uncertainties in runoff estimation in the Tibetan Plateau, with GLDAS2.0 having higher uncertainty in terms of both RB and NSE. The large uncertainty may be because the rainfall data of both GLDAS2.0 and GLDAS2.1 have the highest uncertainty in the Tibetan Plateau among the three regions.

Fig. 5 also shows the NSE values of runoff are decreasing from Northeast China to the Tibetan Plateau for GLDAS2.0, which is consistent with the changes in its rainfall (Fig. 5d). However,

the NSE values are increasing for GLDAS2.1 from Northeast China to the Tibetan Plateau, which reverses the changing patterns of its rainfall (Fig. 5d), i.e., GLDAS2.1 rainfall has the highest NSE (0.92) in Northeast China, the second highest (0.83) in Middle China and the lowest (0.73) in the Tibetan Plateau, whereas runoff has the highest value (0.30) in the Tibetan Plateau, the second highest (0.11) in Middle China and the lowest value (0.01) in Northeast China. Few studies have reported this opposite changing pattern between runoff and rainfall uncertainties.

4 Conclusions and future work

This study evaluates and inter-compares the GLDAS2.0 and GLDAS2.1 runoff simulations on large scales based on the observation data in China. We found that both GLDAS2.0 and GLDAS2.1 have large uncertainties in runoff simulations, with absolute values of RB being above 39% and NSE values being lower than 0.15 on average. Therefore, they are not reliable enough to underpin water related studies in China. GLDAS2.0 has larger uncertainties in the Tibetan Plateau than in the Middle China and Northeast China. We also found that NSE of GLDAS2.1 runoff has an opposite changing pattern to that of rainfall. Thus, cautions should be taken when using them.

Most of hydrological gauges do not monitor snow- and/or glacier- melt flow. Frozen ground distribution is also not well monitored in the regions studied. Therefore, we did not evaluate the influence of snow- and/or glacier- melt flow and frozen ground in the Tibetan Plateau and Northeast China specifically. Future research is encouraged when well monitored snow- and/or glacier- melt flow and frozen ground data are available. In addition to uncertainty from input data and parameterization, the runoff uncertainty may result from the absence of consideration of glacier melt and human intervention to some degree, because the models GLDAS2.0 and

GLDAS2.1 used do not consider glacier melt and human water abstraction. Future research is encouraged to validate improvement of the system performance when GLDAS system model is updated to include them.

(http://www.lancang-mekong.net/).

References:

Allen, G.H., David, C.H., Andreadis, K.M., Hossain, F., Famiglietti, J.S., (2018). Global Estimates of River Flow Wave Travel Times and Implications for Low‐Latency Satellite Data. *Geophysical Research Letters*, 45(15): 7551-7560. DOI:10.1029/2018GL077914

Bai, P., Liu, X., Yang, T., Liang, K., Liu, C., (2016). Evaluation of streamflow simulation results of land surface models in GLDAS on the Tibetan plateau. *Journal of Geophysical Research: Atmospheres*, 121(20). DOI:10.1002/2016JD025501

Barlage, M., Chen, F., Tewari, M., Ikeda, K., Gochis, D., Dudhia, J., Rasmussen, R., Livneh, B., Ek, M., Mitchell, K., (2010). Noah land surface model modifications to improve snowpack prediction in the Colorado Rocky Mountains. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 115(D22). DOI:10.1029/2009JD013470

Beven, K.J., Binley, A., (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, 6(3): 279-298. DOI:10.1002/hyp.3360060305

Beven, K.J., Freer, J.E., (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology*, 249(1-4): 11–29. DOI:10.1016/S0022-1694(01)00421-8

Bosshard, T., Carambia, M., Goergen, K., Kotlarski, S., Krahe, P., Zappa, M., Schär, C., (2013). Quantifying uncertainty sources in an ensemble of hydrological climate-impact projections. *Water Resources Research*, 49(3): 1523-1536. DOI:10.1029/2011wr011533

Chen, F., Mitchell, K., Schaake, J., Xue, Y., Pan, H.-L., Koren, V., Duan, Q.Y., Ek, M., Betts, A., (1996). Modeling of land surface evaporation by four schemes and comparison with FIFE observations. *Journal of Geophysical Research: Atmospheres*, 101(D3): 7251. DOI:10.1029/95jd02165

Chen, Y., Yang, K., Qin, J., Zhao, L., Tang, W., Han, M., (2013). Evaluation of AMSR‐E retrievals and GLDAS simulations against observations of a soil moisture network on the central Tibetan Plateau. *Journal of Geophysical Research: Atmospheres*, 118(10): 4466-4475. DOI:10.1002/jgrd.50301

Crooks, S.M., Kay, A.L., Davies, H.N., Bell, V.A., (2014). From Catchment to National Scale Rainfall-Runoff Modelling: Demonstration of a Hydrological Modelling Framework. *Hydrology*, 1(1): 63-88. DOI:10.3390/hydrology1010063

Kato, H., Rodell, M., Beyrich, F., Cleugh, H., Gorsel, E.v., Liu, H., Meyers, T.P., (2007). Sensitivity of Land Surface Simulations to Model Physics, Land Characteristics, and Forcings, at Four CEOP Sites. *Journal of the Meteorological Society of Japan*, 85A: 187-204. DOI:10.2151/jmsj.85A.187

Li, H., Xu, C.Y., Beldring, S., (2015). How much can we gain with increasing model complexity with the same model concepts? *Journal of Hydrology*, 527: 858-871. DOI:10.1016/j.jhydrol.2015.05.044

Liang, X., Lettenmaier, D.P., Wood, E.F., Burges, S.J., (1994). A simple hydrologically based model of land-surface water and energy fluxes for general-circulation models. *Journal of Geophysical Research: Atmospheres*, 99(D7): 14415-14428. DOI:10.1029/94jd00483

Liu, J., Dietz, T., Carpenter, S.R., Folke, C., Alberti, M., Redman, C.L., Schneider, S.H., Ostrom, E., Pell, A.N., Lubchenco, J., Taylor, W.W., Ouyang, Z., Deadman, P., Kratz, T., Provencher, W., (2007). Coupled Human and Natural Systems. *Ambio*, 40(8): 639-

649. DOI:10.1579/0044-7447(2007)36[639:CHANS]2.0.CO;2

Lv, M., Lu, H., Yang, K., Xu, Z., Lv, M., Huang, X., (2018). Assessment of Runoff Components Simulated by GLDAS against UNH–GRDC Dataset at Global and Hemispheric Scales. *Water*.

Matthew, R., Hiroko Kato, B., 2015. GLDAS Noah Land Surface Model L4 3 hourly 0.25 x 0.25 degree V2.0. NASA/GSFC/HSL, Goddard Earth Sciences Data and Information Services Center (GES DISC), Greenbelt, Maryland, USA. DOI:10.5067/342OHQM9AK6Q

Matthew, R., Hiroko Kato, B., 2016. GLDAS Noah Land Surface Model L4 3 hourly 0.25 x 0.25 degree V2.1. NASA/GSFC/HSL, Goddard Earth Sciences Data and Information Services Center (GES DISC), Greenbelt, Maryland, USA. DOI:10.5067/E7TYRXPJKWOQ

Niu, G., Yang, Z., Mitchell, K.E., Chen, F., Ek, M.B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., (2011). The community Noah land surface model with multiparameterization options (Noah MP): 1. Model description and evaluation with local scale measurements. *Journal of Geophysical Research: Atmospheres*, 116(D12). DOI:10.1029/2010JD015139

Qi, W., Zhang, C., Fu, G., Zhou, H., (2015). Global Land Data Assimilation System data assessment using a distributed biosphere hydrological model. *Journal of Hydrology*, 528: 652-667. DOI:10.1016/j.jhydrol.2015.07.011

Qi, W., Zhang, C., Fu, G., Sweetapple, C., Zhou, H., (2016a). Evaluation of global fine-resolution precipitation products and their uncertainty quantification in ensemble discharge simulations. *Hydrology and Earth System Sciences*, 20(2): 903-920. DOI:10.5194/hess-20-903-2016

Qi, W., Zhang, C., Fu, G., Zhou, H., (2016b). Quantifying dynamic sensitivity of optimization

algorithm parameters to improve hydrological model calibration. *Journal of Hydrology*, 533: 213-223. DOI:10.1016/j.jhydrol.2015.11.052

Qi, W., Zhang, C., Fu, G., Zhou, H., Liu, J., (2016c). Quantifying Uncertainties in Extreme Flood Predictions under Climate Change for a Medium-Sized Basin in Northeastern China. *Journal of Hydrometeorology*, 17(12): 3099-3112. DOI:10.1175/JHM-D-15-0212.1

Qi, W., Liu, J., Chen, D., (2018). Evaluations and Improvements of GLDAS2.0 and GLDAS2.1 Forcing Data's Applicability for Basin Scale Hydrological Simulations in the Tibetan Plateau. *Journal of Geophysical Research: Atmospheres*. DOI:10.1029/2018JD029116

Qi, W., Liu, J., Leung, F., (2019). A framework to quantify impacts of elevated CO2 concentration, global warming and leaf area changes on seasonal variations of water resources on a river basin scale. *Journal of Hydrology*, 570: 508-522. DOI:10.1016/j.jhydrol.2019.01.015

Qi, W., Liu, J., Xia, J., Chen, D., (2020). Divergent sensitivity of surface water and energy variables to precipitation product uncertainty in the Tibetan Plateau. *Journal of Hydrology*, 581: 124338. DOI:10.1016/j.jhydrol.2019.124338

Rodell, M., Houser, P.R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J.K., Walker, J.P., Lohmann, D., Toll, D., (2004). The Global Land Data Assimilation System. *Bulletin of the American Meteorological Society*, 85(3): 381-394. DOI:10.1175/bams-85-3-381

Scanlon, B.R., Zhang, Z., Save, H., Sun, A.Y., Schmied, H., van Beek, L., Wiese, D.N., Wada, Y., Long, D., Reedy, R.C., Longuevergne, L., Döll, P., Bierkens, M.F.P., (2018). Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data. *Proceedings of the National Academy of Sciences*, 115(6): 201704665. DOI:10.1073/pnas.1704665115

Sellers, P.J., (1997). Modeling the Exchanges of Energy, Water, and Carbon Between Continents and the Atmosphere. *Science*, 275(5299): 502-509. DOI:10.1126/science.275.5299.502

Sheffield, J., Goteti, G., Wood, E.F., (2006). Development of a 50-Year High-Resolution Global Dataset of Meteorological Forcings for Land Surface Modeling. *Journal of Climate*, 19(13): 3088-3111. DOI:10.1175/JCLI3790.1

Shen, Y., Xiong, A., (2016). Validation and comparison of a new gauge‐based precipitation analysis over mainland China. *International Journal of Climatology*, 36(1): 252-265. DOI:10.1002/joc.4341

Soroosh, S., Lawford, R., Try, P.L., Rossow, W.B., Roads, J., Polcher, J., Sommeria, G., Schiffer, R.W., (2005). Water and energy cycles: Investigating the links. *Water and energy cycles: Investigating the links*.

Swenson, S.C., Lawrence, D.M., Lee, H., (2012). Improved simulation of the terrestrial hydrological cycle in permafrost regions by the Community Land Model. *Journal of Advances in Modeling Earth Systems*, 4(3). DOI:10.1029/2012MS000165

Tang, Q., Oki, T., Kanae, S., Hu, H., (2008). Hydrological Cycles Change in the Yellow River Basin during the Last Half of the Twentieth Century. *Journal of Climate*, 21(8): 1790-1806. DOI:10.1175/2007JCLI1854.1

Wang, F., Wang, L., Koike, T., Zhou, H., Yang, K., Wang, A., Li, W., (2011). Evaluation and application of a fine-resolution global data set in a semiarid mesoscale river basin with a distributed biosphere hydrological model. *Journal of Geophysical Research: Atmospheres*, 116(D21). DOI:10.1029/2011jd015990

Wang, L., Li, X., Chen, Y., Yang, K., Chen, D., Zhou, J., Liu, W., Qi, J., Huang, J., (2016a). Validation of the global land data assimilation system based on measurements of soil temperature profiles. *Agricultural and Forest Meteorology*, 218(Hydrol. Earth Syst. Sci.

12 2008): 288-297. DOI:10.1016/j.agrformet.2016.01.003

Wang, L., Zhou, J., Qi, J., Sun, L., Yang, K., Tian, L., Lin, Y., Liu, W., Shrestha, M., Xue, Y., Koike, T., Ma, Y., Li, X., Chen, Y., Chen, D., Piao, S., Lu, H., (2017). Development of a land surface model with coupled snow and frozen soil physics. *Water Resources Research*, 53. DOI:10.1002/2017WR020451

Wang, W., Cui, W., Wang, X., Chen, X., (2016b). Evaluation of GLDAS-1 and GLDAS-2 Forcing Data and Noah Model Simulations over China at the Monthly Scale. *Journal of Hydrometeorology*, 17(11): 2815-2833. DOI:10.1175/JHM-D-15-0191.1

Wu, J., Ding, Y., Chen, R., Wang, G., Shen, Y., 2005. The variation and utilization of water resources in the Heihe River basin. In: Cunha, M., Brebbia, C. (Eds.), WIT TRANSACTIONS ON ECOLOGY AND THE ENVIRONMENT, Algarve, PORTUGAL, pp. 331-339.

Yang, H., Zhou, F., Piao, S., Huang, M., Chen, A., Ciais, P., Li, Y., Lian, X., Peng, S., Zeng, Z., (2017). Regional patterns of future runoff changes from Earth system models constrained by observation. *Geophysical Research Letters*: 5540-5549. DOI:10.1002/2017GL073454

Yang, K., Wang, C., Li, S., (2018). Improved Simulation of Frozen‐Thawing Process in Land Surface Model (CLM4.5). *Journal of Geophysical Research: Atmospheres*, 123(23). DOI:10.1029/2017JD028260

Zaitchik, B.F., Rodell, M., Olivera, F., (2010). Evaluation of the Global Land Data Assimilation System using global river discharge data and a source‐to‐sink routing scheme. *Water Resources Research*, 46(6). DOI:10.1029/2009wr007811

Zang, C., Liu, J., (2013). Trend analysis for the flows of green and blue water in the Heihe River basin, northwestern China. *Journal of Hydrology*, 502: 27-36. DOI:10.1016/j.jhydrol.2013.08.022

Zhao, Y., Zhu, J., (2015). Assessing Quality of Grid Daily Precipitation Datasets in China in Recent 50 Years (in Chinese). *PLATEAU METEOROLOGY*, 34(1): 50-58.

Table 1 Details of the study regions

| Region	Basin name	Gauge name	Area (km²)	Mean altitude (m)	Mean annual precipitation (mm)	Mean monthly air temperature (ºC)	Monthly runoff
Middle China	Lower Yangtze river	Datong	1,694,410	1717	989	10.3	2000-2010
Middle China	Middle Yangtze river	Three Gorges Dam	980,213	2624	811	6.4	2000-2008
Middle China	Middle Yellow river	Huayuankou	780,662	1957	461	6.0	2002-2010
Middle China	Heihe	Langxinshan	59,120	2489	327	5.0	2000-2010
Northeast China	Lower Songhua river	Jiamusi	539,089	392	480	2.8	2000-2005
Northeast China	Upper Songhua river	Dalai	210,460	477	435	1.4	2000-2010
Northeast China	Liao river	Liaozhong	179,160	507	377	5.7	2000-2004, 2006-2010
Tibetan Plateau	Yarlung Tsangpo river	Nuxia	205,147	4762	559	0.1	2000-2010
Tibetan Plateau	Upper Yangtze river	Zhimenda	137,371	4762	471	1.1	2000-2010
Tibetan Plateau	Upper Yellow river	Tangnaihai	123,580	4125	550	-1.7	2000-2010
Tibetan Plateau	Lancang river	Changdu	53,656	4557	641	-0.8	2000-2010
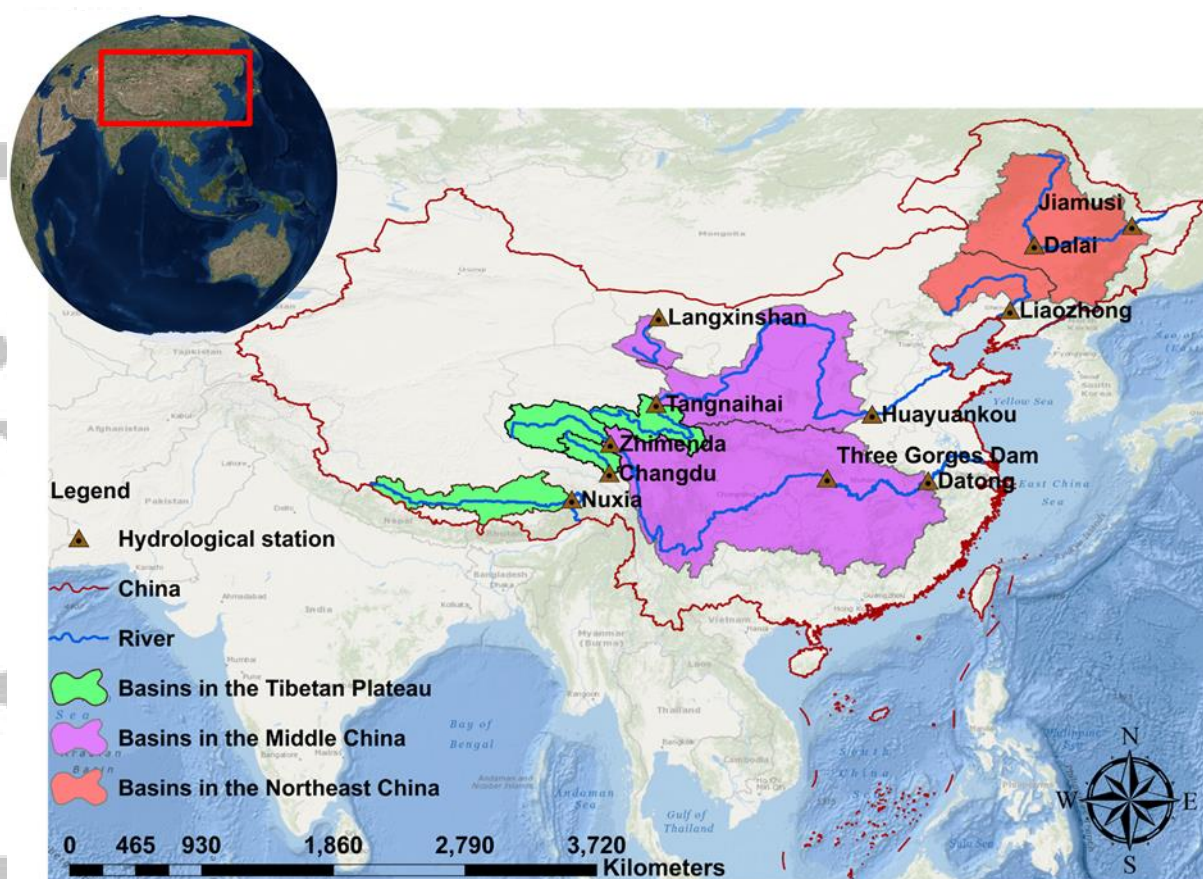
Fig. 1 The regions studied.
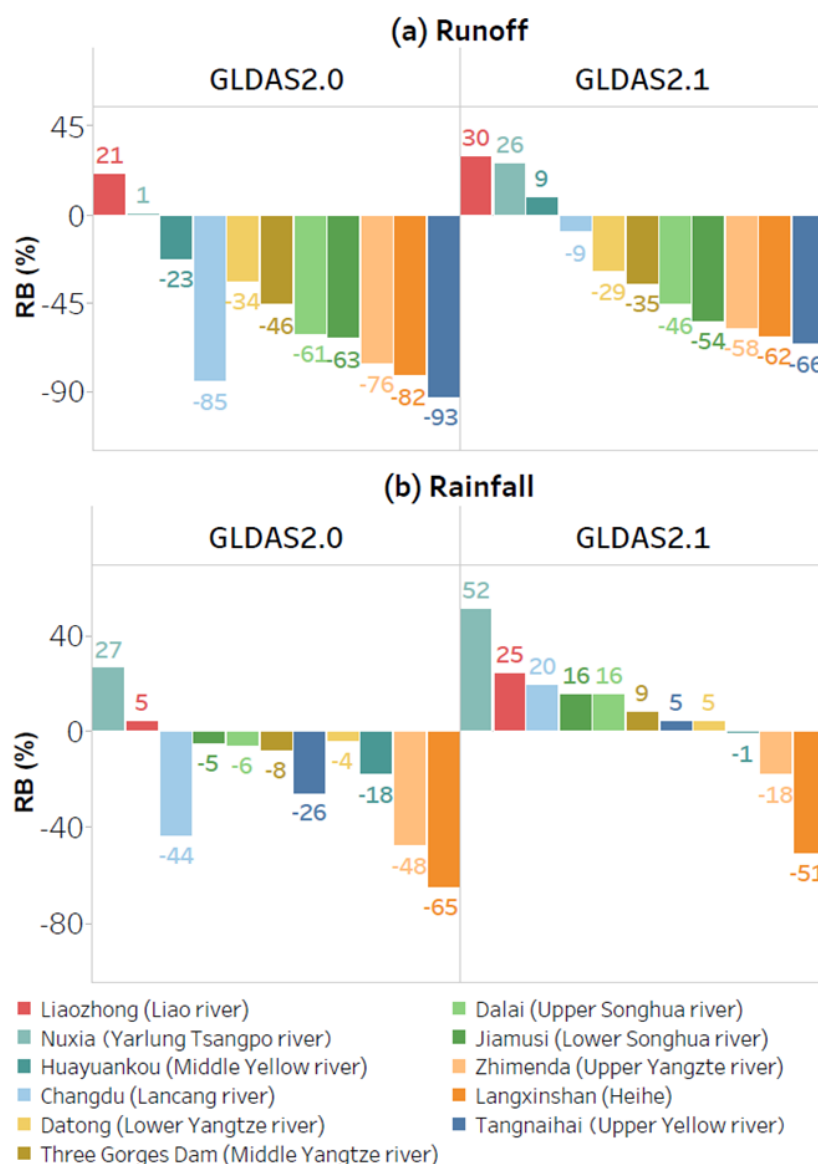
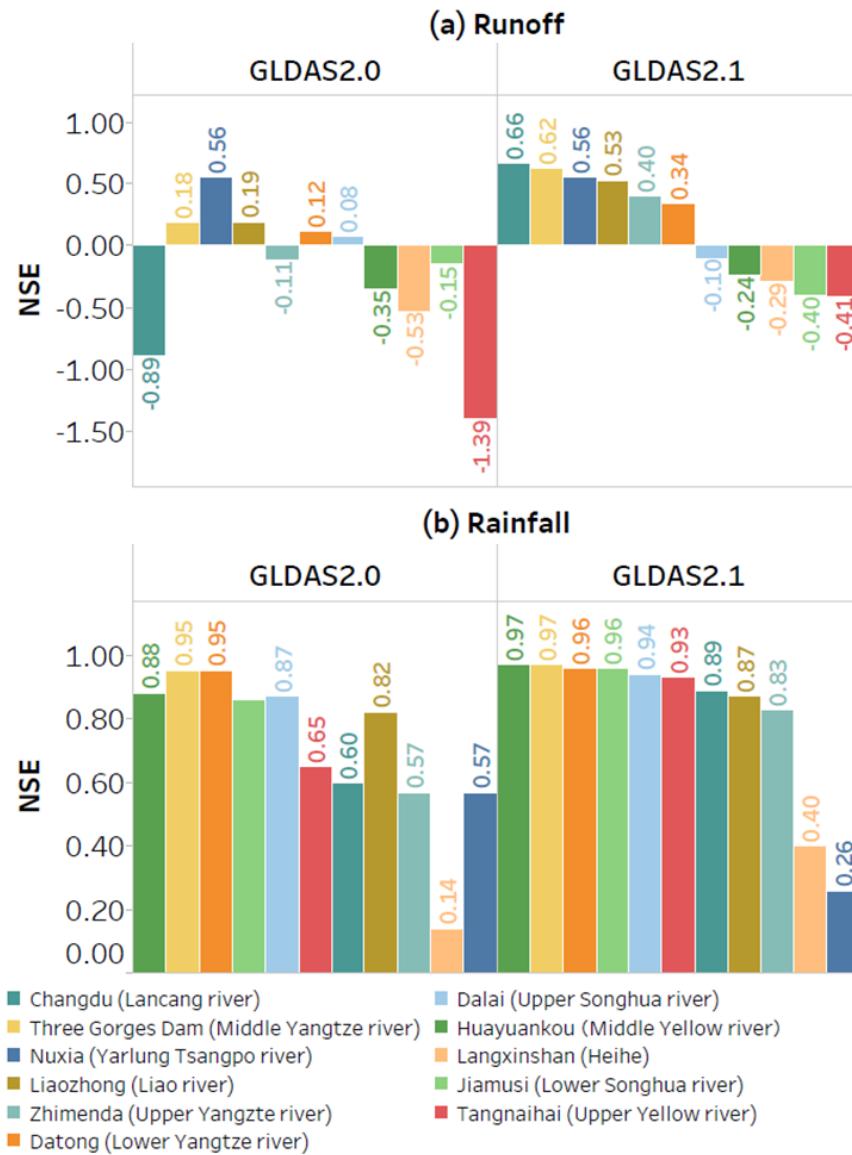Fig. 2 Relative Bias (RB) (%) of runoff and rainfall.

Fig. 3 Nash-Sutcliffe Efficiency (NSE) of runoff and rainfall on a monthly scale.
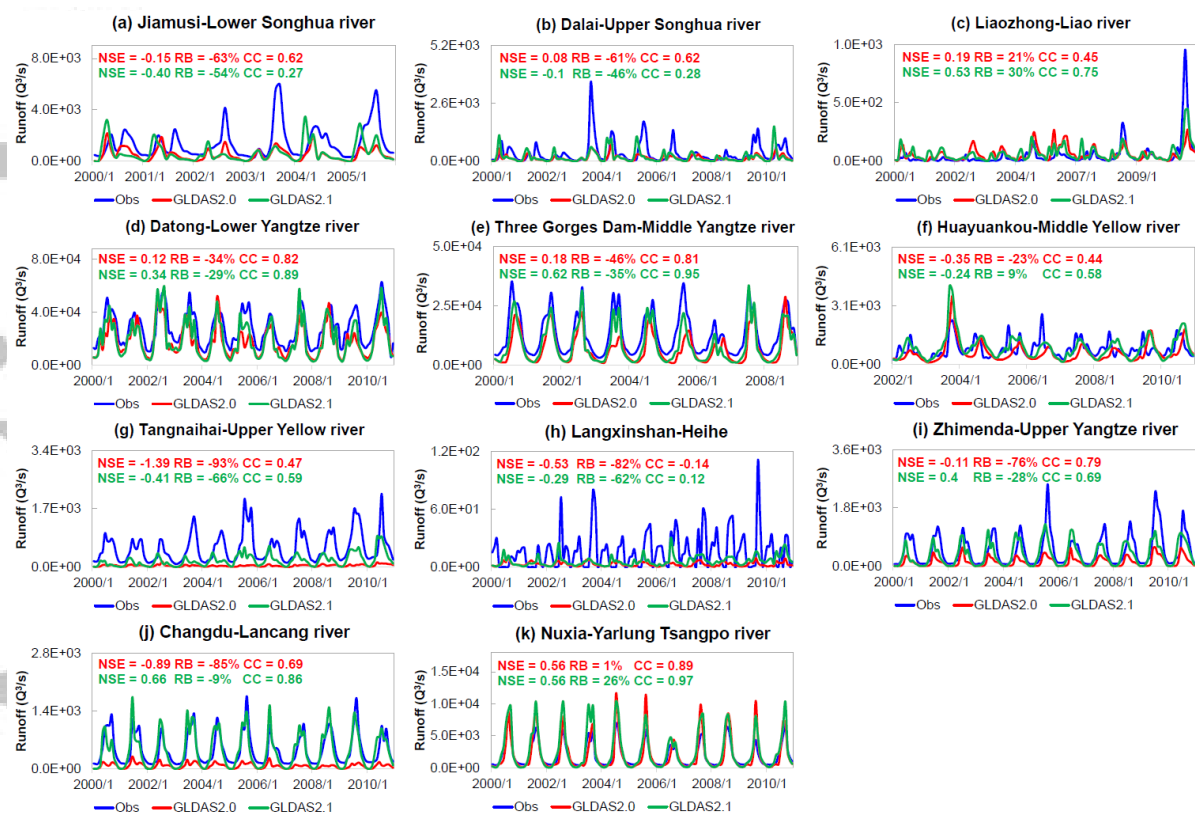
Fig. 4 Comparison of monthly runoff among observation, GLDAS2.0 simulation and GLDAS2.1 simulation. NSE = Nash-Sutcliffe Efficiency; RB = Relative Bias; CC= Correlation Coefficient.
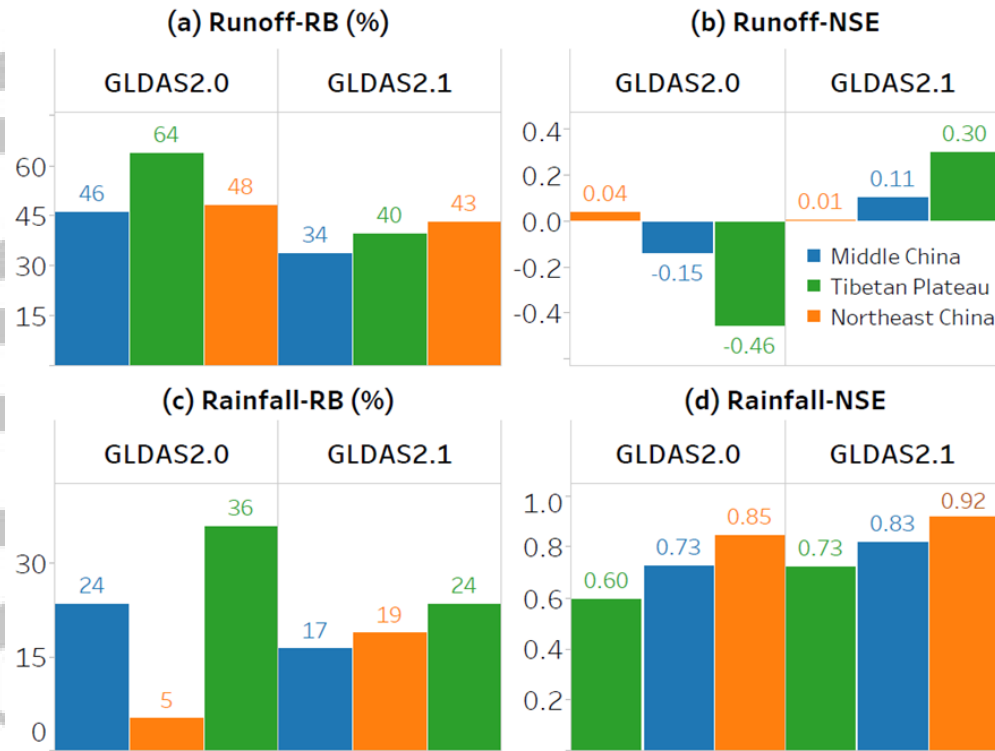
Fig. 5 Regional average absolute Relative Bias (RB) and Nash-Sutcliffe Efficiency (NSE) for runoff and rainfall.