

Interactive comment on “Hydrological evaluation of open-access precipitation data using SWAT at multiple temporal and spatial scales” by Jianzhuang Pang et al.

Jianzhuang Pang et al.

zhanghl@bjfu.edu.cn

Received and published: 26 May 2020

Response to Referee Comments:Reviewer#1

First of all, we sincerely appreciate the comments on our manuscript. We both thank the positive remarks and the specific concerns, which provided great encouragement and specific guidance to the authors to improve the manuscript. Please see the following point-to-point responses to the main concerns and minor comments.

C1

General comments:

This paper presents a hydrological evaluation of two open-access precipitation products (CHIRPS and CPC) compared with rain gauge dataset, at multiple temporal and spatial scales. The content of this research is of great interest to readers of watershed hydrology, remote sensing, and satellite meteorology, since it provided valuable suggestions for researchers in these fields, especially for hydrologic modelers. It is demonstrated by the authors that, even with obvious statistical differences, performances of the three selected precipitation datasets in simulating water yield are parallel. Comparably, inconsistency were found when OPPs and rain gauge data were used to simulate hydrological components, e.g. Surface runoff, lateral flow, and base flow. Inner mechanism was highlighted from both spatial and temporal scales. Overall, this manuscript is quite well written and presented. Minor revision comments below aim to improve the quality of the manuscript.

Main concerns:

1. The paper fails to articulate the implications of its finding (that hydrological models can give very similar model performance, with differing process behaviour, with precipitation datasets with quite different characteristics) in either the Conclusions or the Abstract. For example, Remesan and Holman (2015) study cited by the authors showed that such ‘similar’ calibrated/validated models, when subsequently run using perturbed inputs (e.g. climate change scenario), can lead to different magnitudes and directions of hydrological change due to their differing parameterization. The authors should consider how their findings can guide modelers in the use of these different precipitation datasets for the hydrological modelling of the current and future climate.

Authors’ response: Greatly appreciate the comment and suggestion. As stated in Remesan and Holman’s (2015) study, “with similar historical model performance, model construction with different baseline meteorological data choices significantly

C2

condition the magnitude and direction of simulated hydrological impacts of climate change”, the current study has reached “similar” conclusions: “with similar performances in simulating river runoff, different types of precipitation data digested in hydrologic modeling tends to counterbalance their identified differences by differing parameterization and leads to different directions of hydrologic processes”. Considering that this research focuses on precipitation condition under current climate, it could generally provide implications to hydrological modelers of current and future climate from following two aspects:

1) From perspective of precipitation estimation: CHIRPS has a higher spatial resolution (with 0.05° being equivalent to a resolution of one gauge station for every 30.25 km^2 area) and a stronger ability to recognize heavy rain and extreme rainfall (Fig.4 – Fig.6 in the manuscript). These features would facilitate the widespread use of CHIRPS in future climate analyses. Take extreme climate analyses for example, it is reported that the frequency of extreme rainfall events in China has been significantly increased in past decades and this tendency will continue increasing in future climate change (Mou et al., 2020; Xi et al., 2018). With this background in future climate change, CHIRPS would provide high potential in future extreme rainfall event analyses with high spatial resolution. Actually, CHIRPS has been applied to identify extreme rainfall events by indicators of nP (Number of days with $P \geq 1\text{mm}$), PRCPTOT (Annual total precipitation), and R95pad (Total precipitation when $P > 95$ percentile of all days), etc. (Cavalcante et al., 2020). In contrast, the CPC’s strong ability to identify light rain represents a unique advantage in extreme drought-related research.

2) From perspective of hydrologic modeling: overall, the three precipitation types derive almost equivalent and acceptable hydrological performance according to Moriasi et al’s criteria (2007), while CHIRPS presented better performance in uncertainty analyses. Although the river runoff values simulated by the three models are basically

C3

consistent, there are significant differences among other hydrological components, such as surface runoff, lateral flow, and base flow. CHIRPS tends to derive more surface flow due to the higher precipitation detection, while CPC tends to yield more lateral flow due to the lower precipitation detection. As such, CHIRPS would suit broader applications in flood prediction of the future climate due to its ability in extreme precipitation identification and surface flow simulation. More importantly, multiple-objective calibration based on multiple hydrological components are recommended to improve SWAT modeling in large and spatial resolved watershed.

Cavalcante, R. B. L., Ferreira, D. B. da S., Pontes, P. R. M., Tedeschi, R. G., da Costa, C. P. W., & de Souza, E. B. (2020). Evaluation of extreme rainfall indices from CHIRPS precipitation estimates over the Brazilian Amazonia. *Atmos. Res.*, 238, 104879. doi:10.1016/j.atmosres.2020.104879.

Mou, S., Shi, P., Qu, S., Feng, Y., Chen, C., & Dong, F. (2020). Projected regional responses of precipitation extremes and their joint probabilistic behaviors to climate change in the upper and middle reaches of Huaihe River Basin, China. *Atmos. Res.*, 104942. doi:10.1016/j.atmosres.2020.104942.

Xi, Y., Miao, C., Wu, J., Duan, Q., Lei, X., & Li, H. (2018). Spatiotemporal Changes in Extreme Temperature and Precipitation Events in the Three-Rivers Headwater Region, China. *J. Geophys. Res-Atmos.*, 123, 5827–5844. doi:10.1029/2017jd028226.

The above considerations will be articulated in sections of **Abstract**, **Discussion** and **Conclusions** of the revised manuscript.

Original version of abstract:

L23-27 – “The results of this study demonstrate that evaluating precipitation products using only streamflow simulation accuracy will conceal the dissimilarities between these products. Hydrological models alter hydrologic mechanisms by adjusting calibrated parameters. Specifically, different precipitation detection methods lead

C4

to temporal and spatial variation of water balance components, demonstrating the complexity in describing natural hydrologic processes.”

Revised version of abstract:

The results of this study demonstrate that with similar performances in simulating watershed runoff, the three precipitation datasets tend to conceal the identified dissimilarities through hydrological model parameter calibration, which leads to different directions of hydrologic processes. As such, multiple-objective calibration is recommended for large and spatial resolved watershed in future work. The main findings of this research suggest that the features of OPPs facilitate the widespread use of CHIRPS in extreme flood events and CPC in extreme drought analyses in future climate.

Original version of discussion section:

L435-437 – “. . . Moreover, precipitation in the watershed’s upstream area tended to infiltrate into the land surface due to the lower precipitation detection (see Fig. 7); yet when the river flow converged in the watershed’s downstream area, the surface flow increased due to the larger detected precipitation values.”

Revised version of discussion section:

In Sect. 4.2, the features of the two OPPs in detecting precipitation and hydrologic components modelling will be discussed, and the multi-objective calibration and parameterization will be added.

Moreover, precipitation in the watershed’s upstream area tended to infiltrate into the land surface due to the lower precipitation detection (see Fig. 7); yet when the river flow converged in the watershed’s downstream area, the surface flow increased due to the larger detected precipitation values. The results of these findings demonstrated that although the river runoff simulated by the three models are basically consistent,

C5

hydrologic components exhibited distinct behaviours due to the different features in precipitation detection. CHIRPS has a stronger ability to recognize heavy rain and tends to produce more surface runoff, while CPC’s strong ability to identify light rain produces more lateral flow. As such, multi-objective calibration approach would be recommended for flood prediction in future climate. Tuo et al. (2018) use water yield (WYLD), snow water equivalent (SWE), combining WYLD and SWE as objectives to for parameter calibration and optimization in the SWAT model, and verified the effectiveness of the multi-object procedure.

Figure 1. Spatial variation of annual precipitation at sub-basin scale for (a) Gauge (b) CHIRPS and (c) CPC. (Fig.7 in the manuscript)

Original version of conclusion section:

L461-465 – “In particular, according to parameter adjustment, the three products’ precipitation detection features resulted in significantly different water balance component portions, i.e., the overestimation of MR by CHIRPS resulted in a larger portion of surface flow, while the underestimation of all rainfall by CPC reduced a larger portion of lateral flow. Lastly, the spatial precipitation pattern also significant impacted the spatial distribution of the water balance components from upstream to downstream.”

Revised version of conclusion section:

In particular, according to parameter adjustment, the three products’ precipitation detection features resulted in significantly different water balance component portions, i.e., the overestimation of MR by CHIRPS resulted in a larger portion of surface flow, while the underestimation of all rainfall by CPC resulted in a larger portion of lateral flow. Multi-objective calibration would be recommended for hydrological modellers in parameter calibration and optimization, especially for large and spatial resolved watersheds. Lastly, the spatial precipitation pattern also significantly impacted the spatial

C6

distribution of the water balance components from upstream to downstream.

Although the OPPs have advantages and limitations with respect to the accuracy of precipitation estimates at different spatial and temporal scales, as well as in hydrological modelling and describing hydrologic mechanics, they demonstrate good potential in our case study within the JRW. As such, the OPPs should merge the advantages of satellite, ground observations, as well as the reanalysed data. Fully consideration on performing the hydrological evaluation from both spatial and temporal scales is also key for the future development of OPPs. Furthermore, CHIRPS is advantaged in extreme rainfall detection and thus good as flood prediction, while CPC would be more potentially used in extreme drought analysis in future climate analyses and hydrologic modelling.

C7

2. Given that the authors are simulating a 159,000km² catchment using a single flow gauge for calibration / validation, there is huge equifinality in their results. Given that they used the SUFI-2 / SWATCUP, I would have expected some assessment and discussion of the uncertainty in their model results.

Authors' response: Greatly appreciate the comment. Assessment and discussions on the uncertainty of model results are quite important issues in hydrologic modelling (Abbaspour, 2015). In our study, the model calibration / validation use a single hydrologic station, with a monitored area of more than 159000 km², which would induced inevitable system or random deviation by parameter calibration. Therefore, as the comment suggested, uncertainty analyses on model results should be processed and discussed.

Abbaspour, K. C. (2015) SWAT-CUP 2012: SWAT Calibration and Uncertainty Programs - A User Manual. Tech. rep., Swiss Federal Institute of Aquatic Science and Technology, Eawag, Dübendorf, Switzerland.

With the considerations above, assessment and discussion on the uncertainty of model results will be added in the revised manuscript, and the modification will be specified as follows:

Original version of abstract:

L17-18 – “**All three products** satisfactorily reproduce the stream discharges at the JRW outlet with better performance than the Gauge model.”

Revised version of abstract:

Both OPPs satisfactorily reproduce the stream discharges at the JRW outlet with slightly worse performance than the Gauge model. Model with CHIRPS as inputs

C8

performed slightly better in both model simulation and uncertainty analysis than that of CPC.

Revised version of methodology section:

At the end of Sect.2.4.2, we added a description of the SWAT-CUP-based uncertainty analysis method:

The quality of model input data and the parameterization process increase the uncertainty risk associated with the model results, which has been identified in the application of SWAT (Thavhana et al., 2018; Tuo et al., 2018; Zhang et al., 2020). There are two factors, *p-factor* and *r-factor*, which are used for uncertainty analysis in SUFI-2 algorithm of SWAT CUP. *p-factor* refers to the percentage of the measured data distributed within the 95% prediction uncertainty (95PPU) band of the model results (%), and the *r-factor* graphically means the average thickness of the 95PPU band divided by Standard Deviation (*STD*) of the measured records (Abbaspour, 2017). Theoretically, *p-factor* ranges from 0 to 100% and takes 100% as the optimal value, and *r-factor* ranges from 0 to ∞ and takes 0 as the optimal value. It should be noted that the increase in the *p-factor* comes at the expense of the increase in the *r-factor*. It was stated in the study of Roth & Lemann (2016) that combined values of *p-factor* > 70% and *r-factor* < 1.5 are preferably uncertainty range, which is also referred to in this paper.

Abbaspour, K. C., Vaghefi, S., and Srinivasan, R. (2017) A Guideline for Successful Calibration and Uncertainty Analysis for Soil and Water Assessment: A Review of Papers from the 2016 International SWAT Conference, Water, 10, 6, <https://doi.org/10.3390/w10010006>.

Roth, V. and Lemann, T. (2016) Comparing CFSR and conventional weather data for discharge and soil loss modelling with SWAT in small catchments in the Ethiopian Highlands, Hydrol. Earth. Syst. Sc., 20, 921-934, <https://doi.org/10.5194/hess-20-921-2016>.

C9

Thavhana, M. P., Savage, M. J., and Moeletsi, M. E. (2018) SWAT model uncertainty analysis, calibration and validation for runoff simulation in the Luvuvhu River catchment, South Africa. Physics and Chemistry of the Earth, 105, 115–124, <https://doi.org/10.1016/j.pce.2018.03.012>.

Tuo, Y., Marcolini, G., Disse, M., and Chiogna, G. (2018) A multi-objective approach to improve SWAT model calibration in alpine catchments, J. Hydrol., 559, 347-360, <https://doi.org/10.1016/j.jhydrol.2018.02.055>.

Zhang, H. L., Meng, C. C., Wang, Y. Q., Wang, Y. J., and Li, M. (2020) Comprehensive evaluation of the effects of climate change and land use and land cover change variables on runoff and sediment discharge. Science of the total environment, 702, 134401, <https://doi.org/10.1016/j.scitotenv.2019.134401>.

Original version of result section 3.3.1:

L325-330 – “Based on the model performance classification scheme designed by Moriasi et al. (2007), all three models, each using a different precipitation product, achieved “very good” performance for both the calibration and verification periods, although the Gauge model attained the highest *CC* (0.93 for calibration and 0.87 for validation) and *NSE* (0.92 and 0.87). Compared with the model using Gauge input, the models using the two OPPs tended to underestimate the peak flows that occur mainly during flood seasons (June to August), which is the main reason behind the lower *NSE* values. . .”

Revised version of result section 3.3.1:

Based on the model performance classification scheme designed by Moriasi et al. (2007), Gauge and CHIRPS achieved “very good” performance for both the calibration and verification periods, although the Gauge model attained the highest *NSE* (0.92 for calibration and 0.87 for validation) values and lowest *RSR* (0.28 and 0.36) value, while CPC only reached the level of “Good” due to higher *PBIAS* (10.8) (Fig.9). The underestimation of the peak flows during flood seasons, would be the main reason of the lower *NSE* values of the two OPPs inputs. Further, among all the three models,

C10

the model with Gauge inputs performed best in uncertainty analyses (p -factor = 98%, r -factor = 0.86 for calibration and p -factor = 92%, r -factor = 0.78 for validation), which is followed by the model using CHIRPS as input (p -factor = 84%, r -factor = 0.88 and p -factor = 83%, r -factor = 0.80). Using CPC datasets as precipitation inputs resulted in the highest degree of uncertainty level (p -factor = 57%, r -factor = 0.57 and p -factor = 57%, r -factor = 0.53), which fails to reach a preferable level.

Figure 2. Observed and simulated discharges at the outlet of JRW at monthly scale using precipitation inputs of Gauge, CHIRPS and CPC, respectively. (Fig.9 in the manuscript)

Original version of result section 3.3.2:

L346-350 – “As shown in Fig. 11, the three precipitation inputs also successfully forced the model to replicate the discharge records at the Beibei station at a daily scale, with performance evaluations of “good,” “satisfactory,” and “satisfactory” for Gauge, CHIRPS, and CPC models, respectively. The performances in describing the peak flows are not very good for all of the three products, among which, the Gauge model performs best. The peak flows are usually caused by extreme precipitation events, like rainfall events with an intensity > 80 mm/day.”

Revised version of result section 3.3.2:

As shown in Fig. 11, the three precipitation inputs successfully forced the model to replicate the discharge records at the Beibei station at daily scale, with performance evaluations of “good,” “satisfactory,” and “satisfactory” for Gauge, CHIRPS, and CPC models, respectively. The performances in describing the peak flows were not good for all three products, among which, the Gauge model performs best. **Different from the monthly scale, the CHIRPS-driven daily scale model showed lowest uncertainty level among the three precipitation datasets. The p -factor of Gauge, CHIRPS and**

C11

CPC were 93%, 95%, and 77% for calibration and 84%, 91%, and 73% for validation, respectively, and r -factor were 1.16, 1.25, and 0.98 for calibration and 1.08, 1.27, and 0.93 for validation, respectively. Overall, the uncertainties of daily scale models with all three precipitation datasets as inputs were significantly lower than those of monthly scale, and the CPC-driven monthly model success to reach a preferable level.

Figure 3. Observed and simulated discharges at the outlet of JRW at daily scale using precipitation inputs of Gauge, CHIRPS and CPC, respectively. (Fig.11 in the manuscript)

Original version of conclusions section:

L450-455 – “2. All three precipitation inputs successfully forced the model to replicate the discharge records at the Beibei station at a monthly and daily scale, although they performed slightly better at the daily scale. The differences in the statistics at the monthly and daily scale correspondingly affected the streamflow photographs, e.g. flood peak, base flow, and the rising and falling processes. The three models’ spatial WYLD distributions are highly correlated to that of the precipitation records. While there were equivalent performances in simulating streamflow hydrographs, it should be noted that the calibrated parameters in all three models (Gauge, CHIRPS, and CPC models at monthly and daily scales, see Table 2) were quite different. . .”

Revised version of conclusions section:

2. All three precipitation inputs successfully forced the model to replicate the discharge records at the Beibei station, and results at daily scale presented slightly better performance than that of monthly scale. However, the differences of precipitation inputs in the statistics at the monthly and daily scale correspondingly affected the streamflow photographs, e.g. flood peak, base flow, and the rising and falling processes. **Overall, the CHIRPS dataset performs better in hydrological evaluation because of its lower**

C12

uncertainty level and higher spatial accuracy than that of CPC, thus it can be a good choice for researchers who are interested in this study area. The three models' spatial WYLD distributions are highly correlated to that of the precipitation records. While there were equivalent performances in simulating streamflow hydrographs, it should be noted that the calibrated parameters in all three models (Gauge, CHIRPS, and CPC models at monthly and daily scales, see Table 2) were quite different. . .

C13

3. *The paper provides three sets of SWAT output analyses – monthly, daily and daily aggregated to monthly. However, SWAT is a daily model so the monthly SWAT outputs are themselves an internal aggregation of its daily outputs; so the presentation and description of the daily aggregated to monthly outputs (L439-448 and Figures 12 and 13) are meaningless and should be removed.*

Authors' Response: Thanks a lot for this comment and advice. The presentation and description of the daily aggregated to monthly outputs (L439-448 and Figures 12 and 13) will be removed in the revised manuscript.

As one of the major objectives of this manuscript was to evaluate the performances of different precipitation datasets in simulating the watershed streamflow using SWAT on different temporal scales, the authors ran the SWAT models at monthly and daily scales, respectively. Essentially, SWAT is a daily model that monthly outputs can be derived by aggregating its daily outputs. For researchers, who are not able to collect daily streamflow records, may be more interested in the performance at monthly scale. With this consideration, the authors presented two sets of SWAT output analyses, i.e. daily and monthly, and further look into the corresponding water balance components (Fig.4 & Table 5) adjusting by calibrated parameters (Table 2). In the previous manuscript, proportions of water balance components at monthly scale were compared and analyzed. *In the revised manuscript, water balance components calculated at daily scale should also be presented and compared with results of monthly-scaled models.*

Figure 4. Water balance components for all sub-basins derived from SWAT models using precipitation inputs of (a) Gauge (b) CHIRPS and (c) CPC at monthly scale and (d) Gauge (e) CHIRPS and (f) CPC at daily scale (where SURQ represents surface runoff Q_{surf} ; LATQ represents lateral flow Q_{lat} ; GW_Q is the baseflow from the shallow aquifer; GW_Q_D is the baseflow from the deep aquifer, and the sum of GW_Q and GW_Q_D equals to Q_{gw} ; ET represents actual evapotranspiration ET). (Fig.16 in the manuscript)

C14

Table 2: Optimal parameters calibrated for all three models. (excerpts)

Parameters	Initial range	Gauge		CHIRPS		CPC
		Monthly	Daily	Monthly	Daily	Monthly
a_SOL_K().sol	-10/10	1.988/10	-0.706/10	-0.471/7.681	-0.396/10	5.264/10
v_ESCO.hru	0/1	0.879/1	0.405/1	0.775/1	0.355/1	0.914/1
v_ALPHA_BF.gw	0/1	0.401/0.963	0.299/0.896	0.055/0.677	0.183/0.728	0.216/0.90

Table 5: Summarization of annual average water balance components of the three models for the whole

Time scale	Datasets	Statistics	SURQ	LATQ	GW_Q	GW_Q_D	ET	S
Monthly	Gauge	Average amount/mm	4500.00	2977.22	299.07	60.61	9076.60	1
		Percentage/%	26.61%	17.60%	1.77%	0.36%	53.66%	
	CHIRPS	Average amount/mm	6068.35	773.24	949.56	140.79	9046.83	1
		Percentage/%	35.74%	4.55%	5.59%	0.83%	53.28%	
	CPC	Average amount/mm	1087.19	5577.20	583.45	30.15	8694.40	1
		Percentage/%	6.81%	34.92%	3.65%	0.19%	54.43%	
Daily	Gauge	Average amount/mm	5544.88	1856.00	244.94	48.29	9309.37	1
		Percentage/%	32.61%	10.92%	1.44%	0.28%	54.75%	
	CHIRPS	Average amount/mm	6202.63	834.78	1167.37	59.75	10434.58	1
		Percentage/%	33.17%	4.46%	6.24%	0.32%	55.80%	
	CPC	Average amount/mm	2493.11	2302.28	1709.95	88.66	9384.90	1
		Percentage/%	15.60%	14.41%	10.70%	0.55%	58.73%	

C15

Results showed that:

(1) Either at daily scale or monthly scale, all three models achieved acceptable and similar simulation performance for comparisons of both time series and spatial distributions. However, the parameter systems are completely different at two temporal scales (Table 2). The non-uniqueness of parameters has been proved a persistent drawback of SWAT (Abbaspour et al., 2004; Abbaspour, 2015; Zhang et al., 2015).

And we had explained this drawback at **line399** to **line404** in manuscript:

“In general, simulated and observed streamflow hydrographs, using OPPs and Gauge inputs, can successfully match at both monthly and daily scales. However, consistency between simulated and observed streamflow does not guarantee identical hydrologic processes. For example, the SWAT model calibrated parameters are not the same for all precipitation inputs, meaning that the hydrologic mechanics during SWAT modelling are also different. As such, it is critical that researchers and decision makers adequately understand the benefits and limitations of different precipitation products in modelling the hydrologic processes.”

(2) With differing parameterizations, different precipitation inputs tend to derive completely different hydrological component amounts at different time scales (Fig. 16 & Table 5).

Abbaspour, K. C., Johnson, C. A., & van Genuchten, M. T. (2004). Estimating Uncertain Flow and Transport Parameters Using a Sequential Uncertainty Fitting Procedure. *Vadose Zone Journal*, 3(4), 1340. doi:10.2136/vzj2004.1340.

Abbaspour, K. C. (2015) SWAT-CUP 2012: SWAT Calibration and Uncertainty Programs - A User Manual. Tech. rep., Swiss Federal Institute of Aquatic Science and Technology, Eawag, Dübendorf, Switzerland.

Zhang, J., Li, Q., Guo, B., & Gong, H. (2015). The comparative study of multi-site uncertainty evaluation method based on SWAT model. *Hydrological Processes*, 29(13), 2994–3009. doi:10.1002/hyp.10380.

C16

With the considerations above, discussion on the model parameters and water balance components will be added in the revised manuscript, and the modification will be specified as follows:

Original version of conclusions section:

L408-425 – “Thus, we calculated the water balance component portions, Q_{surf} , Q_{lat} , Q_{gw} , and E_a , for all the JRW sub-basins. It is evident from Fig.16 and Table 4 that the total portions of water balance components differ among the three precipitation products. However, they do share some similarities in that the evapotranspiration (ET) portions of all three products are above 50 %, resulting in a watershed runoff production coefficient of 0.45. Furthermore, the main Gauge model components are SURQ and LATQ, which account for 25.92 % and 16.72 %, respectively; the main CHIRPS component is SURQ, which accounts for 34.80 %, and the main CPC component is LATQ, which accounts for 33.62 %. Spatially, the surface flow portion increases from upstream to downstream.

The above water balance component regularities are primarily the result of two causes. First, the differences in the above hydrological component proportions are mainly controlled by the model parameters. For example, ESCO is a soil evaporation compensation factor that directly affects maximum evaporation from soil; the smaller the value, the larger the maximum evaporation. The SWAT model indirectly increases WYLD by using higher ESCO and thus decreases the ET value. In this study, the ESCO values for Gauge, CHIRPS, and CPC range from 0.879 - 1, 0.775 - 1, and 0.914 - 1, respectively. Furthermore, the total ET values during the study period were 8153.94, 8161.22, and 7806.84 mm, respectively. Apparently, the CPC model reduced its corresponding ET by using a higher ESCO parameter, so that the lack of precipitation inputs would be offset by less evaporation. This result is consistent with that reported by Bai & Liu (2018), who conducted a study at the source regions of the Yellow River and Yangtze River basins in the Tibetan Plateau. They further concluded that the impact of different precipitation inputs on runoff simulation is largely offset by parameter calibration,

C17

resulting in significant differences in evaporation and storage estimates.”

It should be noted that the average values of water balance components for the whole watershed were calculated by sub-basin area weighting method, i.e. the portion of the sub-basin area was assigned as the weight coefficient of the sub-basin's water balance values.

Revised version of conclusions section:

Thus, we calculated the water balance component portions, Q_{surf} , Q_{lat} , Q_{gw} , and ET , for all the JRW sub-basins. With differing parameterizations, different precipitation inputs tend to derive completely different hydrological component amounts at different time scales (Fig. 16 & Table 5). At monthly scale, all three models, with Gauge, CHIRPS and CPC as inputs, have similar ET portions, which account for above 54%. The major components of Gauge model are SURQ and LATQ, accounting for 25.92 % and 16.72 %, respectively, the major component of CHIRPS model is SURQ, which accounts for 34.80 %, and the primary component of CPC model is LATQ, which accounts for 33.62 %. However, at daily scale, SURQ of Gauge model increased largely, reaching a proportion 32.61%, while LATQ decreased to 10.92%; LATQ of CPC model decreased and SURQ and ET increased, accounting for 14.41%, 15.60% and 58.73%, respectively; water balance components proportions of CHIRPS model slightly changed.

The above water balance component regularities are primarily the result of two causes. First, the differences in the above hydrological component proportions are highly possibly related in parameter adjustment. As shown in table 2, the SURQ of Gauge and CPC models were significantly increased due to the decrease of the parameter SOL_K, which stands for saturated hydraulic conductivity. The decrease of the parameter ESCO in CPC model led to the increase of ET ratio, which influenced soil evaporation compensation. The variation of parameter ALPHA_BF, which is baseflow recession constant, caused the GW_Q components of the three models to vary in the same direction. This result is consistent with that reported by Bai & Liu (2018), who conducted a study at the source regions of the Yellow River and Yangtze River basins

C18

in the Tibetan Plateau. They further concluded that the impact of different precipitation inputs on runoff simulation is largely offset by parameter calibration, resulting in significant differences in evaporation and storage estimates.

C19

Minor revision comments:

1. L19 – change “All three products” to “Both OPPs” as the text is comparing to the gauge model.

Authors’ Response: Thanks a lot for pointing out this issue.

The sentence will be corrected as “Both OPPs satisfactorily reproduce the stream discharges at the JRW outlet with slightly worse performance than the Gauge model, . . .”

2. L153 – is the evapotranspiration “actual”, “potential” or “reference”?

Authors’ Response: It’s the actual evapotranspiration, and the sentence will be revised as “the annual average actual evapotranspiration (*ET*) ranges from 800 to 1000 mm.”

The descriptions related to evapotranspiration all through the manuscript have been corrected in the revised manuscript:

L214 – “Water balance, including precipitation, surface runoff, **evapotranspiration**, infiltration, lateral and base flow, and percolation to shallow and deep aquifers, is mathematically expressed as follows:”

The sentence will be corrected as “Water balance, including precipitation, surface runoff, **actual evapotranspiration**, infiltration, lateral and base flow, and percolation to shallow and deep aquifers, is mathematically expressed as follows:”

C20

L217 – “ E_a = evapotranspiration” will be corrected as “ ET = actual evapotranspiration”.

L410-411 – “However, they do share some similarities in that the **evapotranspiration (ET)** portions of all three products are above 50 %, resulting in a watershed runoff production coefficient of 0.45.” **The sentence will be corrected as** “However, they do share some similarities in that the **actual evapotranspiration (ET)** portions of all three products are above 50 %, resulting in a watershed runoff production coefficient of 0.45.”

L418-422 – “The SWAT model indirectly increases WYLD by using higher ESCO and thus decreases the **ET** value. In this study, the ESCO values for Gauge, CHIRPS, and CPC range from 0.879 - 1, 0.775 - 1, and 0.914 - 1, respectively. Furthermore, the total **ET** values during the study period were 8153.94, 8161.22, and 7806.84 mm, respectively. Apparently, the CPC model reduced its corresponding **ET** by using a higher ESCO parameter, so that the lack of precipitation inputs would be offset by less evaporation.”

The sentence will be corrected as “The SWAT model indirectly increases WYLD by using higher ESCO and thus decreases the **ET** value. In this study, the ESCO values for Gauge, CHIRPS, and CPC range from 0.879 - 1, 0.775 - 1, and 0.914 - 1, respectively. Furthermore, the total **ET** values during the study period were 8153.94, 8161.22, and 7806.84 mm, respectively. Apparently, the CPC model reduced its corresponding **ET** by using a higher ESCO parameter, so that the lack of precipitation inputs would be offset by less **ET** .”

3. L169-170 – how has the classification accuracy been determined, given that it was based on “manual visual interpretation”?

C21

Authors’ Response: Thank you very much for this question.

The procedure of deriving LUC types based on 2010 Landsat TM/ETM remote sensing images are as follows: The geometric shape, colour feature, texture feature and spatial distribution of ground objects were analysed and extracted according to the image spectral features. The remote sensing image interpretation marks were established based on the field measurement data and the reference map. Six primary classifications were recognized- cultivated land, woodland, grassland, water area, construction land, and unused land. The quality of the LUC product was checked by combining field survey and random sampling dynamic map spot for repeated interpretation analysis. Generally, the quality inspection result is that the classification accuracy of cultivated land data is 85%, and that of other data can reach more than 75%.

The manuscript will be revised as “The data included six primary classification- cultivated land, woodland, grassland, water area, construction land, and unused land, as well as 25 secondary classifications. **After checking the quality of data products by combining field survey and random sampling dynamic map spot for repeated interpretation analysis, it is proved that the cultivated land’s classification accuracy was 85 %, and other data classification accuracies reached 75 %.**”.

4. L194 – how does a dataset (CHIRPS v2.0) released in 2015 provide data to the “present”?

Authors’ Response: Thank you very much for this question.

Actually, the CHIRPS v2.0 dataset has been continuously updated since it was released in 2015, and we are sorry for the misinterpretation.

C22

The manuscript will be revised as “The first gridded format CHIRPS product was released in 1981 to present and the most recent one (V2.0 datasets) was released in February 2015. The dataset spans from 1981 to the present and provides daily precipitation data with a spatial resolution of 0.05° in a pseudo global coverage of 50° N - 50° S.”

5. L237 – looking at equation (3), isn't the optimal value of $STD_n = 1$ e.g. identical STD_s ? And why should STD_n values range from 0-1 which implies STD gauge can never be $< STD$ opp? General $-RMSE$, STD and $PBIAS$ have units – please use them throughout

Authors' Response: we are sorry to make this mistake for our neglect, which should be corrected as: “The STD_n values range from 0 to ∞ , and the optimal value is 1.”

The units of $RMSE$, STD and $PBIAS$ will be revised throughout the manuscript as follows:

L189-190 – “Where n is the number of the time series; Q_i and S_i are measured values and estimated values (or simulated values), respectively; and \bar{Q} and \bar{s} are the mean values of the measured and estimated values (or simulated values), respectively.”

The sentence will be revised as “Where n is the number of the time series; Q_i and S_i are measured values and estimated values, respectively; and \bar{Q} and \bar{s} are the mean values of the measured and estimated values, respectively. The value may refer to either precipitation (mm) or streamflow discharge (m^3/s).”

C23

L191 – “Standard deviation (STD) represents the discretization degree of the datasets.” The sentence will be revised as “Standard deviation (STD) represents the discretization degree of the precipitation datasets (mm).”

L195-196 – “ $RMSD$ value: Root mean square deviation ($RMSD$) is used to demonstrate the error between the OPPs and Gauge datasets. $RMSD$ has a range from 0 to $+\infty$, and an optimal value of 0.” The sentence will be revised as “ $RMSD$ value: Root mean square deviation ($RMSD$) is used to demonstrate the error between the OPPs and Gauge datasets (mm). $RMSD$ has a range from 0 to $+\infty$ mm, with an optimal value of 0 mm.”

L248 – “ $PBIAS$ describes the OPPs' systematic bias. $PBIAS$ ranges from 0 to $+\infty$, and the optimal value is 0.” The sentence will be revised as “ $PBIAS$ describes the OPPs' systematic bias (%). $PBIAS$ ranges from 0 to $+\infty$ %, and the optimal value is 0 %.”

L266-267 – “The $RMSD$ values for Gauge-CHIRPS and Gauge-CPC are 15.80 and 12.95, respectively.” The sentence will be revised as “The $RMSD$ values for Gauge-CHIRPS and Gauge-CPC are 15.80 mm and 12.95 mm, respectively.”

L281-283 – “Statistically, the CC , STD , and $RMSD$ values between CHIRPS and the Gauge records are 0.53, 1.14, and 5.16, respectively, and 0.64, 0.87, and 3.95, respectively, between the CPC and Gauge products.”

The sentence will be revised as “Statistically, the CC , STD_n , and $RMSD$ values between CHIRPS and the Gauge records are 0.53 mm, 1.14 mm, and 5.16 mm, respectively, and 0.64 mm, 0.87 mm, and 3.95 mm, respectively, between the CPC and Gauge products.”

C24

L361-362 – “The *CC*, *STD_n*, and *RMSD* values between CHIRPS and Gauge are 0.92, 1.06, and 0.23, respectively, and 0.81, 0.94, 0.33 between CPC and Gauge, respectively.” [The sentence will be revised as](#) “The *CC*, *STD_n*, and *RMSD* values between CHIRPS and Gauge are 0.92 **mm**, 1.06 **mm**, and 0.23 **mm**, respectively, and 0.81 **mm**, 0.94 **mm**, 0.33 **mm** between CPC and Gauge, respectively.”

L266 – “The ***STD*** values for Gauge-CHIRPS and Gauge-CPC are 1.06 and 0.94, respectively.” [The sentence will be revised as](#) “The ***STD_n*** values for Gauge-CHIRPS and Gauge-CPC are 1.06 and 0.94, respectively.”

L268-270 – “Nevertheless, *PBIAS* values of Gauge-CHIRPS and Gauge-CPC were 9.58 and -6.70, respectively”
[The sentence will be revised as](#) “Nevertheless, *PBIAS* values of Gauge-CHIRPS and Gauge-CPC were 9.58 % and -6.70 %, respectively”

L343-344 – “The *PBIAS* values for Gauge-CHIRPS and Gauge-CPC are 5.85 and -5.38, respectively.” [The sentence will be revised as](#) “The *PBIAS* values for Gauge-CHIRPS and Gauge-CPC are 5.85 % and -5.38 %, respectively.”

6. L463 – “antecedent” is the more usual term for “early-stage”.

Authors’ Response: Thank you very much for your advice, and we have revised this term into antecedent all through the manuscript:

C25

L373-375 – “Solano-Rivera et al. (2019) experimented in the San Lorencito headwater catchment and found that the rainfall-runoff dynamics before extreme events were mainly related to **early-stage** conditions. After extreme flood events, **early-stage** conditions had no effect on rainfall-runoff processes, and rainfall significantly affected the streamflow discharge.”

[The sentence will be changed as](#) “Solano-Rivera et al. (2019) experimented in the San Lorencito headwater catchment and found that the rainfall-runoff dynamics before extreme events were mainly related to **antecedent** conditions. After extreme flood events, **antecedent** conditions had no effect on rainfall-runoff processes, and rainfall significantly affected the streamflow discharge.”

7. L483 – there are no ALPHA-BF parameter ranges given in Table 1 and 2 to substantiate this. The values of ALHPA_BF and GWRECH_DP should be added to the tables.

Authors’ Response: Thanks a lot for pointing out this error, and ALHPA_BF [has been added to Table1 and Table2](#). Since the parameter RCHRG_DP is not a sensitive one, so it was not included in the calibration process.

Table 1: Hydrological parameters considered for sensitivity analysis (“a_”, “v_” and r_” means an absolute increase, a replacement, and a relative change to the initial parameter values, respectively).

Parameters	Description	Range	Default
v__PLAPS.sub	Precipitation lapse rate[mm]	-1000/1000	0
v__ALPHA_BF.gw	Baseflow alpha factor [days⁻¹]	0/1	0.048
v__ALPHA_BNK.rte	Baseflow alpha factor for bank storage	0/1	0

C26

Table 2: Optimal parameters calibrated for all three models. (excerpts)

Parameters	Initial range	Gauge		CHIRPS		CPC
		Monthly	Daily	Monthly	Daily	Monthly
v__PLAPS.sub	-1000/1000	0.012/0.067	0.061/0.183	0.079/0.135	0.068/0.205	0.017/0.0
v__ALPHA_BF.gw	0/1	0.401/0.963	0.299/0.896	0.055/0.677	0.183/0.728	0.216/0.9
v__ALPHA_BNK.rte	0/1	0.492/0.863	0.444/1	0.201/0.696	0.467/1	0.564/1

8. L486 – what is “proletarian” flow?

Authors’ Response: Thank you so much for pointing out this typo.

The authors tended to articulate that "For CPC dataset, the high proportion of LR events will lead to severe rainfall losses in the initial- and post- loss processes, resulting in very limited surface water yield. As such, [the sentence will be corrected as:](#)

“A potential reason for this phenomenon may be that the rainfall during LR events tends to be easily lost in the initial- and post-loss processes, resulting in very limited or even no WYLD.”

9. L500 – equation 7

Authors’ Response:thank you very much for pointing out this error, [and it will be corrected in the revised manuscript.](#)

10. L560 – “streamflow photograph”? hydrograph?

Authors’ Response:thank you very much for pointing out this typo, [and it will be corrected in the revised manuscript.](#)

C27

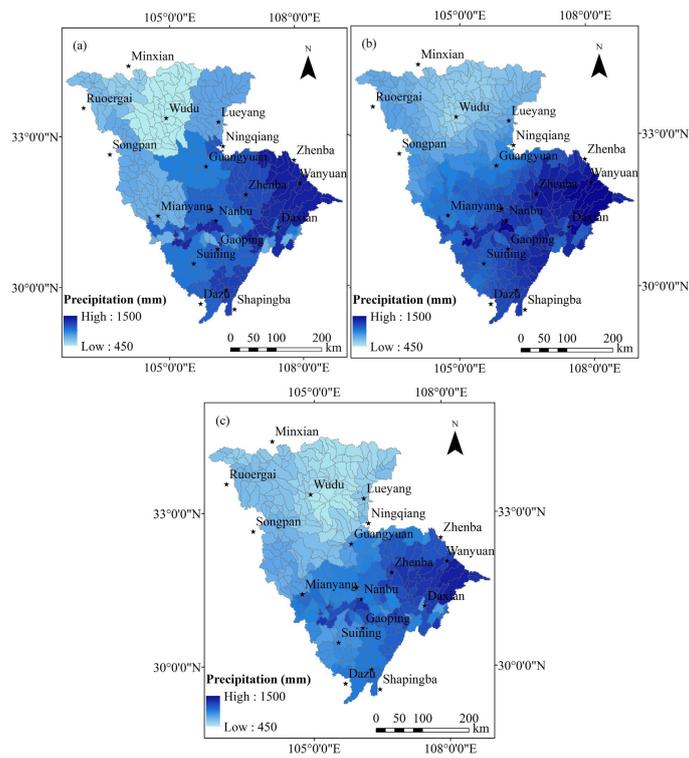


Fig. 1. Spatial variation of annual precipitation at sub-basin scale for (a) Gauge (b) CHIRPS and (c) CPC.

C29

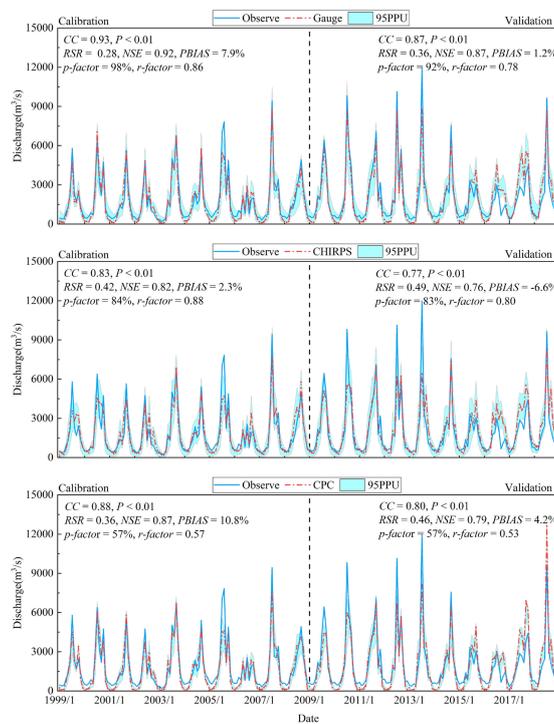


Fig. 2. Observed and simulated discharges at the outlet of JRW at monthly scale using precipitation inputs of Gauge, CHIRPS and CPC, respectively.

C30

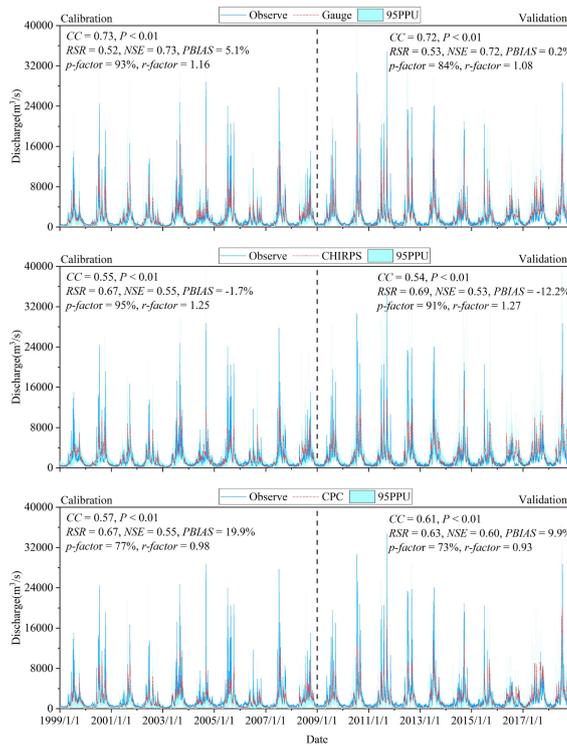


Fig. 3. Observed and simulated discharges at the outlet of JRW at daily scale using precipitation inputs of Gauge, CHIRPS and CPC, respectively.

C31

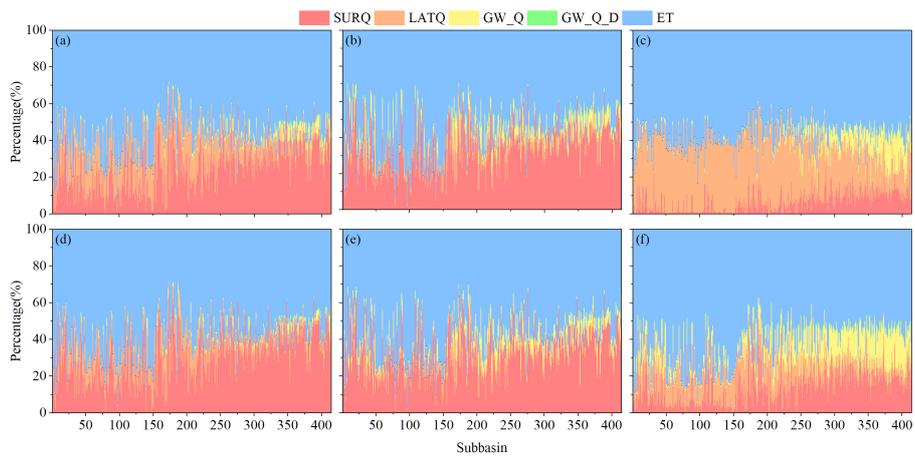


Fig. 4. Water balance components for all sub-basins derived from SWAT models using precipitation inputs of (a) Gauge (b) CHIRPS and (c) CPC at monthly scale and (d) Gauge (e) CHIRPS and (f) CPC at daily scale

C32