Author's Response

Manuscript Number: hess-2020-56

Authors: Jianzhuang Pang, Huilan Zhang, Quanxi Xu, Yujie Wang, Yunqi Wang, Ouyang Zhang, Jiaxin Hao

Title: Hydrological evaluation of open-access precipitation data using SWAT at multiple temporal and spatial scales

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 bellowing point-to-point responses to the main concerns and minor comments.

Response to the reviews

Response to Comments from Reviewer #1

General Comments: *This paper uses two open-access precipitation products (CHIRPS and CPC) and a dataset from rain gauges to drive the SWAT hydrological model in the Jiang river watershed in China. All three precipitation datasets are shown to produce generally similar hydrological model performances, with the calibrated parameterisation reducing the effect of the identified differences in the precipitation datasets through changing hydrological processes. This is a potentially useful paper for the hydrological modelling community in that it highlights that an acceptable hydrological performance according to the commonly-used Moriasi et al (2007) criteria for the Nash Sutcliffe Efficiency metric does not mean that the hydrological processes are correctly simulated – but that it can indicate that the calibration process has merely been successful in altering the catchment hydrological processes to compensate for inadequacies in the input data.*

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 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 $P \geqslant 1$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 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 were articulated in sections of **Abstract**, **Discussion** and **Conclusions of the revised manuscript**. Supplements relating to the first aspect are marked blue, while those relating to the second aspect are marked red.

Original version of abstract:

P1-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 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:

P19-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 was discussed, and the multi-objective calibration and parameterization was 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, 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 7. Spatial variation of annual precipitation at sub-basin scale for (a) Gauge (b) CHIRPS and (c) CPC.

Original version of conclusion section:

P21-L461-470 – "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.

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 modeling 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 reanalyzed data. Furthermore, fully consideration on performing the hydrological evaluation from both spatial and temporal scales is also key for the future development of OPPs."

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 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.

2. Given that the authors are simulating a 159,000km^2 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 km^2, 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 In-stitute of Aquatic Science and Technology, Eawag, Dübendorf, Switzerland.

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

Original version of abstract:

P1-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 performed slightly better in both model simulation and fairly better in uncertainty analysis than that of CPC.

Revised version of materials and methods section:

At the end of Sect.2.4.2 (P12-L249), 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.

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:

P15-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). Further, among all the three models, 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. The underestimation of the peak flows during flood seasons, would be the main reason of the slightly worse performance of the two OPPs inputs.

Figure 9. Observed and simulated discharges at the outlet of JRW at monthly scale using precipitation inputs of Gauge, CHIRPS and CPC, respectively.

Original version of result section 3.3.2:

P16-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. 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 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. The performances in describing the peak flows were not good for all 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….

Figure 11. Observed and simulated discharges at the outlet of JRW at daily scale using precipitation inputs of Gauge, CHIRPS and CPC, respectively.

Original version of conclusions section:

P20-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 monthly scale presented slightly better performance than that of daily 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 uncertainty level and higher spatial accuracy than that of CPC, thus it can be a fairly good option 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…

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) was 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 16. 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.

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

Parameters	Initial range	Gauge Monthly	Daily	CHIRPS Monthly	Daily	CPC Monthly	Daily
a__SOL_K().sol	−10/10	1.988/10	-0.706/10	-0.471/7.681	-0.396/10	5.264/10	-2.106/10
v__ESCO.hru	0/1	0.879/1	0.405/1	0.775/1	0.355/1	0.914/1	0.462/1
v__ALPHA_BF.gw	0/1	0.401/0.963	0.299/0.896	0.055/0.677	0.183/0.728	0.216/0.901	0.415/1

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

Time scale	Datasets	Statistics	SURQ	LATQ	GW_Q	GW_Q_D	ET	Summation
Monthly	Gauge	Average amount/mm	4500.00	2977.22	299.07	60.61	9076.60	16913.50
		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	16978.78
		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	15972.40
		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	17003.48
		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	18699.11
		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	15978.90
		Percentage/%	15.60%	14.41%	10.70%	0.55%	58.73%	

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 In-stitute 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.

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

Original version of discussion section:

P18-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, 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 discussion 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. 15 & 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 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.

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.

P1-L17 – "All three products satisfactorily reproduce the stream discharges at the JRW outlet with better performance than the Gauge model." was 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.

P5-L123 – "the annual average evapotranspiration ranges from 800 to 1000 mm." was 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:

P10-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 was 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:"

P10-L217 – "E_a = evapotranspiration" was corrected as "ET = actual evapotranspiration".

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

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

The procedure of deriving LUCC 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 LUCC 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%.

P6-L135-137 – "The data included six primary classifications—cultivated land, woodland, grassland, water area, construction land, and unused land, as well as 25 secondary classifications. The cultivated land's classification accuracy was 85 %, and other data classification accuracies reached 75 %."

The sentence was revised as "The data included six primary classifications—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.

P7-L156 – "The most recent gridded format CHIRPS product (V2.0 datasets) was completed and released in February 2015."

The sentence was revised as "The first gridded format CHIRPS product was released in February 2015, which has first recorded in 1981 and continues to be updated."

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.

P9-L193 – "The STD_n values range from 0 to 1, and the optimal value is 0. The *STD* value is mathematically expressed as follows:"

The sentence should be corrected as: "The STD_n values range from 0 to ∞, and the optimal value is 1. The STD_n value is mathematically expressed as follows:"

The units of *STD* and *PBIAS* was revised throughout the manuscript as follows:

P8-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 $\overline{Q}$ and $\overline{s}$ are the mean values of the measured and estimated values (or simulated values), respectively."
The sentence was revised as "Where n is the number of the time series; Q_i and S_i are measured values and estimated values, respectively; and $\overline{Q}$ and $\overline{s}$ 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)."

P9-L191 – "Standard deviation (*STD*) represents the discretization degree of the datasets."
The sentence was revised as "Standard deviation (*STD*) represents the discretization degree of the precipitation datasets (mm)."

P12-L248 – "*PBIAS* describes the OPPs' systematic bias. *PBIAS* ranges from 0 to +∞, and the optimal value is 0."
The sentence was revised as "*PBIAS* describes the OPPs' systematic bias (%). *PBIAS* ranges from 0 to + ∞ %, and the optimal value is 0 %."

P12-L266 – "The *STD* values for Gauge-CHIRPS and Gauge-CPC are 1.06 and 0.94, respectively."
The sentence was revised as "The STD_n values for Gauge-CHIRPS and Gauge-CPC are 1.06 and 0.94, respectively."

P13-L268-270 – "Nevertheless, *PBIAS* values of Gauge-CHIRPS and Gauge-CPC were 9.58 and -6.70, respectively"
The sentence was revised as "Nevertheless, *PBIAS* values of Gauge-CHIRPS and Gauge-CPC were 9.58 % and -6.70 %, respectively"

P16-L343-344 – "The *PBIAS* values for Gauge-CHIRPS and Gauge-CPC are 5.85 and -5.38, respectively."
The sentence was 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:
P17-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 was 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). (excerpts)

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

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

Parameters	Initial range	Gauge		CHIRPS		CPC	
		Monthly	Daily	Monthly	Daily	Monthly	Daily
v__PLAPS.sub	−1000/1000	0.012/0.067	0.061/0.183	0.079/0.135	0.068/0.205	0.017/0.078	-0.014/0.095
v__ALPHA_BF.gw	0/1	0.401/0.963	0.299/0.896	0.055/0.677	0.183/0.728	0.216/0.901	0.415/1
v__ALPHA_BNK.rte	0/1	0.492/0.863	0.444/1	0.201/0.696	0.467/1	0.564/1	0.307/0.92

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, "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 low proletarian flow and thus WYLD." was corrected as:

P18-L392-393 – "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.

P18-L405 – "According to the SWAT model's water balance equation (Eq. 9), …" was corrected as "According to the SWAT model's water balance equation (Eq. 7), …"

10. L560 – "streamflow photograph"? hydrograph?

Authors' Response: thank you very much for pointing out this typo.

P20-L452 – "The differences in the statistics at the monthly and daily scale correspondingly affected the streamflow photograph, …" was corrected as "However, the differences of precipitation inputs in the statistics at the monthly and daily scale correspondingly affected the streamflow hydrograph, …"

Response to Comments from Reviewer #2

General comment: *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 spatial resolutions of CHIRPS (0.05°) and CPC (0.5°) were higher than that of the geographic datasets, "some of the grid records are potentially missed, especially for the high - resolution CHIRPS products." Duan Z et al. (2019) proposed an area-weighted method to calculated precipitation for each subbasin, "Calculate the area-weighted average daily CHIRPS data (after disaggregated by 10 times (0.005°)) from all grids within the subbasin to represent the effective daily precipitation for each subbasin". This might be an alternative way to solve the data problem.

Authors' response: Greatly appreciate your suggestion. Following this advice, we recalculated the precipitation inputs for each sub-basin via "area-weighted" (AW) method, and compared the results with those derived by "Nearest Distance" (ND) adopted in this manuscript. Take the sub-basin No.411, which is located at Beibei hydrological station, as an example, the calculated precipitation inputs by the above two methods were depicted in Fig. S1. It's shown in Fig.S1 that no significant difference (P = 0.88) were detected by using these two methods. Compared with the ND method, the amount of rainfall obtained by AW method is slightly underestimated (*PBIAS* = -0.78%), especially when the rainfall intensity is between 50 mm and 100mm. Both methods slightly increase the uncertainty of precipitation inputs. Considering the effectiveness in producing peak discharges of streamflow in SWAT model, the ND method in the original manuscript is adopted. Actually, the "potentially missed" grid records mainly refers to CHIRPS products. More than 10000 grids within the JRW or nearby were adopted, which is much higher than that of the CPC (~165 grids) and rain gauges (20 stations). Considering that the "missed" grids could be covered by other grids within the same sub-basin due to the high resolution of CHIRPS.

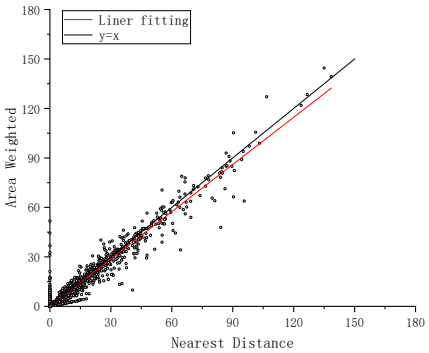


Figure S1. Scatter plot of the CHIRPS precipitation comparing between Nearest Distance method and Area Weighted method.

Therefore, to avoid misrepresentation of data accuracy, the manuscript will be revised as follows (**line 177** to **line 178)**:

Original version of abstract:

P8-L177-178 – "Using this method, some of the grid records are potentially missed, especially for the high-resolution CHIRPS products."

Revised version of abstract:

Using this method, the grid records of high-resolution CHIRPS products within the same sub-basin will be uniformly assigned the grid value closest to the centroid, which will offset the high resolution advantage of CHIRPS products.

2. Moriasi et al. (2007), cited by the author, used three indicators RSR (ratio of the root mean square error to the standard deviation of measured data), NSE, and PBIAS to establish a model to evaluate performance level, while the author used only two. Why not use all three metrics? Besides, since only the NSE index was graded into different evaluation levels, was the evaluation on model performance reliable without the evaluation grades from other two indicators?

Authors' response: Greatly appreciate the comment. After an extensive literature review on model evaluation system, we followed the suggestion and adopted all three indicators to evaluate the performance of the hydrologic models with different precipitation datasets. The models' performances were classified using the standard defined by Moriasi et al. (2007), which are shown in Table 2. And the evaluation results by using all three metrics at monthly and daily scales were depicted in Table S1 and Table S2, respectively.

Comparably, in the original manuscript, like most existing papers did (Zhu et al., 2015; Tuo et al., 2018; Duan et al., 2019), only *NSE*, *CC*, and *PBIAS* were adopted to evaluate the simulation performance of hydrologic models. But they did not give a classified evaluation of the simulation results. When we used only *NSE* to classify the simulation performance of the model, as presented in the original manuscript, the evaluation results are "Very good" for all three models at the monthly scale, but when we used three indicators to evaluate, CPC only reached the level of "Good" (Table S1.).

Table 2. General performance ratings statistics recommended by Moriasi et al. (2007).

Performance Rating	*RSR*	*NSE*	*PBIAS* (%)
Very good	0.00 < RSR ≤ 0.50	0.75 < NSE ≤ 1.00	PBIAS ≤ ±10
Good	0.50 < RSR ≤ 0.60	0.65 < NSE ≤ 0.75	±10 < PBIAS ≤ ±15
Satisfactory	0.60 < RSR ≤ 0.70	0.50 < NSE ≤ 0.65	±15 < PBIAS ≤ ±25
Unsatisfactory	RSR > 0.70	NSE < 0.50	PBIAS > ±25

Table S1. Evaluation results of monthly scale SWAT model

	Calibration			Validation		
	Gauge	CHIRPS	CPC	Gauge	CHIRPS	CPC
RSR	0.28	0.42	0.36	0.36	0.49	0.46
NSE	0.92	0.82	0.87	0.87	0.76	0.79
PBIAS (%)	7.9	2.3	10.8	1.2	-6.6	4.2
Evaluation	Very good	Very good	Good	Very good	Very good	Very good

Table S2. Evaluation results of daily scale SWAT model

	Calibration			Validation		
	Gauge	CHIRPS	CPC	Gauge	CHIRPS	CPC
RSR	0.52	0.67	0.67	0.53	0.69	0.63
NSE	0.73	0.55	0.55	0.72	0.53	0.6
PBIAS (%)	5.1	-1.7	19.9	0.2	-12.2	9.9
Evaluation	Good	Satisfactory	Satisfactory	Good	Satisfactory	Satisfactory

Duan, Z., Tuo, Y., Liu, J., Gao, H., Song, X., Zhang, Z., Yang, L., and Mekonnen, D. F.: Hydrological evaluation of open-access precipitation and air temperature datasets using SWAT in a poorly gauged basin in Ethiopia, J. Hydrol., 569, 612-626, https://doi.org/10.1016/j.jhydrol.2018.12.026, 2019.

Tuo, Y., Marcolini, G., Disse, M., and Chiogna, G.: 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, 2018.

Zhu, H., Li, Y., Liu, Z., Shi, X., Fu, B., and Xing, Z.: Using SWAT to simulate streamflow in Huifa River basin with ground and Fengyun precipitation data, J. Hydroinform., 17, 834–844, https://doi.org/10.2166/hydro.2015.104, 2015.

The above considerations was supplemented in sections of **Materials and methods** and **Result** of the revised manuscript.

Original version of materials and methods section 2.3:

P9-L195-196 – "(3) *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 +∞ mm, and an optimal value of 0 mm. The *RMSD* value is expressed as follows:"

$$RMSD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[(Si-\overline{S})-(Qi-\overline{Q})\right]^2}, \quad (4)$$

Revised version of materials and methods section 2.3:

(3) *RSR* value: observations standard deviation ratio (*RSR*) is an error index statistic between the OPPs and Gauge datasets. Root mean square error (*RMSE*) divided by *STD* values would derive the *RSR* value. *RSR has a* range from 0 to ∞ with 0 as the optimal value. The calculation equation is expressed as follows:

$$RSR = \frac{RMSE}{STD} = \frac{\sqrt{\sum_{i=1}^{n}(S_i - Q_i)^2}}{\sqrt{\sum_{i=1}^{n}\left(S_i-\overline{S}\right)^2}}, \quad (4)$$

Original version of materials and methods section 2.4.2:

P11-L244-247 – "The model performance was classified using the *NSE* values defined by Moriasi et al. (2007): unsatisfactory performance ($NSE \leq 0.50$), satisfactory performance ($0.50 < NSE \leq 0.65$), good performance ($0.65 < NSE \leq 0.75$) and very good performance ($0.75 < NSE \leq 1.00$). The three models' parameter range after 3000 iterations is shown in Table 2. *CC*, *NSE*, and Percentage bias (*PBIAS*) were used to evaluate the model simulation results."

Revised version of materials and methods section 2.4.2:

The model performance was classified using *RSR*, *NSE*, and percentage bias (*PBIAS*) values defined by Moriasi et al. (2007), which is shown in Table 2. The three models' parameter range after 3000 iterations is shown in Table 3. *CC*, *NSE*, *PBIAS and RSR* were used to evaluate the model simulation results.

Original version of result section:

P15-L325-328 – "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)."

Revised version of result section:

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).

3. Explain what "The correlation coefficients' spatial variation" is? The spatial correlation of the three precipitation datasets should be a value rather than a graph. Explain how Fig. 8 was calculated and obtained. Explain why distinguish average precipitation in daily and monthly scales?

Authors' response: Thanks a lot for the question. Basically, the term "correlation coefficients" (CC_{sub}) refers to the correlation between two OPPs and Gauge at either daily or monthly scales for each sub-basin. "The correlation coefficients' spatial variation" is the variation of CC_{cub} in different sub-basins. The calculation steps of CC_{sub} are as follows. Firstly, the spatial distribution diagrams of Gauge records and the two OPPs are calculated. The correlation coefficient for each sub-basin between the Gauge series and OPPs series is calculated by using the correlation coefficient method. Coefficients for the all 414 sub-basins will form a spatial distribution plot of the correlation coefficients. Fig. 8 aims to distinguish the correlation of rainfall amounts in different sub-basins, and to preliminarily judge the performance of OPPs in different sub-basins, especially at different time scales. Distinguish average precipitation at daily and monthly scales are important, since the difference in precipitation amount statistics may highly resulted in different modelling performance and inner mechanics of hydrologic processes. Actually, the *CC* values between spatial-aggregated CPC and Gauge records are higher than that of CHIRPS and Gauge at both scales, yet the *CC* values at the monthly scale are much higher than that at daily scale. Similar variation regularity was found in spatial distribution, which was described in Fig. 8. However, *CC* values between CPC and Gauge are smaller than that of CHIRPS and Gauge at a portion of sub-basins, which was explained at **line313** to **line315** in manuscript:

"Spatially, the higher CC values between the Gauge and CPC at the monthly scale are mainly distributed in areas with comparably low or high rainfall amounts, such as Wudu and Wangyuan. Yet, the CC value was less relevant in areas with moderate rainfall (e.g., Suining), compared with to that of the Gauge and CHIRPS."

The manuscript will be accordingly revised as follows (**line 311** to **line 315**). Additions to the spatial correlation *CC* are marked blue, and additions to the correlation coefficients' spatial variation are marked red:

Original version:

P14-L311-312 – "The correlation coefficients' spatial variation between the Gauge and OPPs at monthly and daily scales are illustrated in Fig. 8."

Revised version:

"The *CC*, *STDn*, and *RSR* values of precipitation spatial distribution between CHIRPS and Gauge are 0.89, 0.96, and 0.55, respectively, and 0.82, 0.87, 0.62 between CPC and Gauge, respectively. These statistics indicate that both CHIRPS and CPC estimates can describe the spatial distribution of precipitation in JRW, among which CHIRPS depicts better performance. The correlation coefficients between the Gauge and OPPs at monthly or daily scales for every sub-basin are illustrated in Fig. 8.

Minor revision comments:

1. Double-check: L11- "Jiang River". L114- "larges" should be "largest". L299- "varations" should be "variations".

Authors' Response: Thanks a lot for pointing out the typos, and we made revisions all through the manuscripts.

P1-L11-12 – "Herein, we implemented a comprehensive evaluation of three selected precipitation products in the Jiang River Watershed (JRW) located in southwest China." was corrected as "Herein, we implemented a comprehensive evaluation of three selected precipitation products in the Jialing River Watershed (JRW) located in southwest China."

P5-L114 – "The Jialing River is the primary tributary of the Yangtze River, with the larges drainage area of 159,812 km^2 and a total length of ~1345 kilometers." was corrected as "The Jialing River is the primary tributary of the Yangtze River, with the largest drainage area of 159,812 km^2 and a total length of ~1345 kilometers."

P14-L299 – "Spatial varations of the three products' long-term mean annual precipitation, …" was corrected as "Spatial variations of the three products' long-term mean annual precipitation, …"

2. Grammar errors. L98- The verb "have" should be "has". L92- The article "an" here should be "a". L310- "relative" should be "relatively".

Authors' Response: Thank you very much for pointing out the grammar errors, and we have corrected them as follows:

P4-L98 – "However, evidence for water balance component variations under the influence of different precipitation inputs have not been fully investigated." was corrected as "However, evidence for water balance component variations under the influence of different precipitation inputs has not been fully investigated."

P4-L92 – "Bai & Liu (2018) used an HIMS model …" was corrected as "Bai & Liu (2018) used a HIMS model…"

P14-L310 – "the overall precipitation values estimated by the CHIRPS are relative higher" was corrected as "the overall precipitation values estimated by the CHIRPS are relatively higher"

3. L342 & L360-361 unit of CC should be decimal rather than percentage.

Authors' Response: Thank you so much for the suggestion.

P16-L342 – "The spatial correlation between WYLD and precipitation for rainfall for the Gauge, CHIRPS, and CPC products reached 84.8 %, 84.3 %, and 90.84 %, respectively."
The sentence was corrected as "The spatial correlation between WYLD and precipitation for rainfall for the Gauge, CHIRPS, and CPC products reached 0.85, 0.84, and 0.91, respectively."

P16-L360-361 – "The *CC* values between the WYLD and precipitation for Gauge, CHIRPS, and CPC at the daily scale are 83.45 %, 84.41 % and 91.70 %, respectively, …"
The sentence was corrected as "The *CC* values between the WYLD and precipitation for Gauge, CHIRPS, and CPC at the daily scale are 0.83, 0.84 and 0.92, respectively, …"

4. L257-As IPCC reported, "Extreme rainfall" was defined as the 95th percentile of daily precipitation data. Therefore, Fig.3, shown as monthly rainfall box chart, failed to capture "extreme rainfall values"

Authors' Response: Thanks a lot for the comment and suggestion.

The term "extreme rainfall" here means that estimation of Gauge monthly rainfall in the rainy season (especially in July) is significantly higher than that of the other two OPPs. For better interpretation, the term will be revised as:

P12-L257 – "Fig. 3 shows that the extreme rainfall values captured by Gauge are higher than those of CHIRPS and CPC." was corrected as "Fig. 3 shows that the rainfall values in the rainy season (especially in July) captured by Gauge are higher than those of CHIRPS and CPC."

5. L327-Usually we use "validation" instead of "verification".

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

P15-L327 – “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, …”
The sentence was corrected as “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 validation periods, …”

P15-L336 – “CPC showed significant overestimation in 2017 and 2018 during the verification period.” was corrected as “CPC showed significant overestimation in 2017 and 2018 during the validation period.”

6. L451- “although they performed slightly better at the daily scale.” the model should perform better at the monthly scale?

Authors’ Response: We apologize for this error, and we have revised it as follows:
P20-L451 – “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.” was corrected as “All three precipitation inputs successfully forced the model to replicate the discharge records at the Beibei station at monthly and daily scale, and results at monthly scale presented slightly better performance than that of daily scale.”

Marked-up manuscript version

Hydrological evaluation of open-access precipitation data using SWAT at multiple temporal and spatial scales

Jianzhuang Pang[1], Huilan Zhang[1*], Quanxi Xu[2], Yujie Wang[1], Yunqi Wang[1], Ouyang Zhang[2], Jiaxin Hao[1]

[1]Three-gorges reservoir area (Chongqing) Forest Ecosystem Research Station, School of Soil and Water Conservation, Beijing Forestry University, Beijing 100083, China

[2] Bureau of Hydrology, Changjiang Water Resources Commission of the Ministry of the Water Resources, Wuhan 430010, China

Correspondence to: Huilan Zhang (zhanghl@bjfu.edu.cn)

Abstract. Temporal and spatial precipitation information is key to conducting effective hydrological process simulation and forecasting. Herein, we implemented a comprehensive evaluation of three selected precipitation products in the ~~Jiang~~ Jialing River Watershed (JRW) located in southwest China. A number of indices were used to statistically analyze the differences between two open-access precipitation products (OPPs), i.e. Climate Hazards Group Infra-Red Precipitation with Station (CHIRPS) and CPC-Global (CPC), and the rain gauge (Gauge). The three products were then categorized into sub-basins to drive SWAT simulations. The results show: (1) the three products are highly consistent in temporal variation on a monthly scale, yet distinct on a daily scale. CHIRPS is characterized by overestimation of light rain, underestimation of heavy rain, and a high probability of false alarm. CPC generally underestimates rainfall of all magnitudes; (2) ~~All three products~~Both OPPs satisfactorily reproduce the stream discharges at the JRW outlet with ~~better~~ slightly worse performance than the Gauge model. Model with CHIRPS as inputs performed slightly better in both model simulation and fairly better in uncertainty analysis than that of CPC. On a temporal scale, the OPPs are inferior with respect to capturing flood peak, yet superior at describing other hydrograph features, e.g. rising and falling processes and base~~-~~flow. On a spatial scale, CHIRPS offers the advantage of deriving smooth, distributed precipitation and runoff due to its high resolution; (3) The water balance components derived from SWAT models with equal simulated streamflow discharges are remarkably different between the three precipitation inputs. The precipitation spatial pattern results in an increasing surface flow trend from upstream to downstream. 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. ~~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 to temporal and spatial variation of water balance components, demonstrating the complexity in describing natural hydrologic processes.~~

1. Introduction

Precipitation has been established as the most significant meteorological parameter with respect to forcing and calibrating hydrologic models; as its spatial–temporal variability considerably influences hydrological behavior and water resource availability (Galván et al., 2014; Lobligeois et al., 2014; Roth & Lemann, 2016). Previous studies have demonstrated that reducing precipitation data uncertainty has a sizeable impact on stabilizing model parameterization and calibration (Mileham et al., 2008; Cornelissen et al., 2016; Remesan & Holman, 2015). However, accurate portrayal of authentic basin rainfall inputs' spatial–temporal variability has severe limitations (Bohnenstengel et al., 2011; Liu et al., 2017), and successfully acquiring such data from available resources has typically posed numerous challenges for hydrologic modeling (Long et al., 2016; Zambrano-Bigiarini et al., 2017).

Conventionally, hydrologists have regarded gauge measurements as actual rainfall (Zhu et al., 2015; Musie et al., 2019), and used point rainfall measurements from rain gauges to perform spatial interpolation and illustrate rainfall field in basin/sub-basin regions (Weiberlen & Benitez, 2018; Belete et al., 2019). Ideally, if rain gauges are positioned with reasonable density and uniform distribution, the spatial precipitation variation described by this method is still valid (Duan et al., 2016). However, in remote or developing regions, the meteorological stations are usually scarce and irregularly distributed, resulting in inconsistent and erroneous distributed rainfall field (Hwang et al., 2011; Peleg et al., 2013; Cecinati et al., 2018; Luo et al., 2019). In other cases, when data observation is accidentally missing, the data quality might be unreliable (Alijanian et al., 2017; Sun et al., 2018). These challenges make ground-based precipitation measurements subject to large uncertainties for driving and calculating hydrologic models.

Over the last few decades, open-source precipitation products (OPPs) have provided a promising alternative for detecting temporal and spatial precipitation variability (Qi et al., 2016; Jiang et al., 2018; Jiang et al., 2019). A variety of studies have demonstrated the accuracy differences among the various OPPs, as well as within the same OPP among different regions (Gao

et al., 2017 Zambrano-Bigiarini et al., 2017; Lu et al., 2019; Wu Y et al., 2019). Sun et al (2018) evaluated and compared the advantages and disadvantages of 29 OPPs with different spatial and temporal resolutions with respect to their ability to describe global precipitation. Unlike most OPPs with a spatial resolution of 0.25°~ 0.5°, the Climate Hazards Group Infra-Red Precipitation with Station (CHIRPS), a "satellite-gauge" type precipitation product, provides very fine spatial resolution (Funk et al., 2015), with 0.05° being equivalent to a resolution of one gauge station for every 30.25 km^2 area. This characteristic has facilitated widespread use and consistent admiration of CHIRPS in recent years. Duan Z et al. (2019) evaluated three precipitation products in Ethiopia and found that CHIRPS performed best among them; Lai et al. (2019) used PERSIANN-CDR and CHIRPS driven hydrological simulation in the Beijiang River basin of China, and determined that CHIRPS performed significantly better than the PERSIANN-CDR. The Climate Prediction Center Gauge-Based Analysis of Global Daily Precipitation (CPC-Global) is a unified precipitation analysis product from the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC) (Xie et al., 2007) that contains unified precipitation data collected from > 30,000 monitoring stations belonging to the WMO Global Telecommunication System, Cooperative Observer Network, and other national meteorological agencies. The product was created using the optimal interpolation objective analysis technique and the data is considered to be relatively accurate. Tian et al. (2010) used CPC as reference data to evaluate the applicability of GSMaP in the United States; and Beck et al. (2017~~b~~) used CPC to modify the MSWEP data product they built. Marked differences have been found in the accuracy and spatio-temporal patterns of different precipitation datasets, which highlights the critical importance of dataset selection for both scientific researchers and decision makers.

The hydrological model or rainfall-runoff model is an important tool for understanding hydrological processes and aids water resources operation decision-makers (Yilmaz et al., 2012; Yan et al., 2016; Wu J et al., 2019). Most frequently used hydrologic models have been shown to efficiently incorporate data from rain gauges, while open-access precipitation has also been continuously improved and adopted into different modules that evaluate its performance in simulating watershed runoff (Bhuiyan et al., 2019; Solakian et al., 2019). Among all the various existing hydrologic models, the Soil & Water Assessment Tool (SWAT) is widely employed by the scientific community and others interested in watershed hydrology research and management (Price et al., 2013; Wang et al., 2017; Li et al., 2018; Qiu et al., 2019). ~ 4000 peer-reviewed papers in reputable academic journals worldwide (SWAT literature database, from 1984 to 2020) have used SWAT modeling results to support

their scientific endeavors. Moreover, a new version (SWAT+) is currently in development that will provide a more flexible spatial representation of interactions and processes within a watershed (Volk et al., 2009; Gabriel et al., 2014; Ayana et al., 2015; Jin et al., 2018). As mentioned above, numerous researchers concur that designing accurate watershed models requires realistic depiction of temporal and spatial precipitation variability. Huang et al. (2019) used hourly, sub-daily, and diurnal precipitation data to simulate runoff from Baden-Württemberg state in Germany and found a positive correlation between higher rainfall temporal resolution and model performance. As such, hydrological processes simulated by the SWAT model, that were based on environmental data lacking accurate regional precipitation distribution figures, will unquestionably be faulty and unreliable. For example, Lobligeois et al. (2014) used rain-gauge measurements (2,500 stations within an area of 550,000 km^2), and Weather radar network data with a spatial resolution of 1km, to simulate runoff from France. Their results clearly showed that the higher resolution radar data significantly improved the simulation accuracy.

To date, the effect of combined ground-based and satellite-based precipitation estimates on streamflow simulation accuracy is not well understood, particularly when the data covers a variety of temporal and spatial resolutions. Hence, this study aims to elucidate these unknowns. More importantly, hydrologic models are expected to describe internal hydrologic processes and subsequently present a unique interpretation of the water balance components (Pellicer-Martínez et al., 2015; Tanner & Hughes, 2015; Wang et al., 2018); yet very limited studies have been conducted to investigate the effects of temporal and spatial resolution on hydrologic processes or water balance components. Thus, this fundamental issue must be addressed before hydrologic modeling with open-access precipitation datasets can be utilized at maximum capacity; as without a thorough understanding of the water cycle's inner processes, the hydrologic models may be highly misleading and facilitate inappropriate management decisions. Bai & Liu (2018) used a~~n~~ HIMS model to simulate the runoff driven by CHIRPS, CMORPH, PERSIANN-CDR, TMPA 3B42, and MSWEP at the source regions of the Yellow River and Yangtze River basins in the Tibetan Plateau. They reported that parameter calibration significantly counterbalanced the impact of diverse precipitation inputs on runoff modeling, resulting in substantial differences in evaporation and storage estimates. Their research helps enhance our understanding of how water balance components are impacted by precipitation data and hydrologic model parameters. However, evidence for water balance component variations under the influence of different precipitation inputs

ha~~ve~~s not been fully investigated. In this study, we aimed to verify the ability of CPC-global and CHIRPS to accurately simulate watershed runoff, and analyze how different OPP characteristics modify the hydrological processes simulation.

The Yangtze River is the largest river in China. Its upper segment is home to its primary tributary with the largest drainage area—the Jialing River, which exhibits spatial heterogeneity with respect to climate, geomorphologies, and land cover conditions. Over the past six decades, anthropogenic activity in conjunction with climate change have substantially reduced the drainage basin's streamflow (Meng et al., 2019), which significantly impacts the inflow condition of the three Gorge reservoir. Therefore, it is debatably imperative to elucidate how varying rainfall characteristics impact runoff and hydrological processes in the Jialing River Watershed (JRW), especially in spatio-temporal dimensions. Given the above considerations, herein, we attempt to: (1) statistically quantify the differences of ground-based and typical open-source precipitation datasets in the JRW, (2) evaluate the performances of different precipitation datasets in simulating the watershed streamflow using SWAT, and (3) investigate the potential behaviors of different precipitation dataset in describing hydrologic processes. All of the above objectives were analyzed on temporal and spatial scales. The goal of this study was to surpass an accuracy assessment of rainfall estimates, and evaluate the use of diverse precipitation data types as model operation forcing data and in hydrologic process portrayal.

2. Materials and methods

2.1. Study area

The Jialing River is the primary tributary of the Yangtze River, with the largest drainage area of 159,812 km^2 and a total length of ~1345 kilometers. The JRW is situated between 29°17′30″ N and 34°28′11″ N and 102°35′36″ E and 109°01′08″ E and geographically extends over the northern part of the transition zone under the eastern Tibet Plateau. The JRW's elevation difference is ~5000 m and the average gradient is ~ 2.05 ‰. Due to the sharply changing topographic gradients, the area features northwest highlands, northern mid-low mountains, middle-eastern hills, and southern plains (Fig. 1). The hydrometeorological conditions follow a similar spatial distribution pattern, i.e., relatively colder and drier in the north and warmer and wetter in the south (Meng et al., 2019). The long-term annual precipitation, based on records from 1956–2018,

ranges from ~ 900 to 1200 mm, a product of southwest China's warm, low latitude air. The rainfall is mainly concentrated from May to September, which accounts for about 60% of the annual precipitation. The annual average temperature ranges from 4.3 to 27.4℃, and the annual average actual evapotranspiration (ET) ranges from 800 to 1000 mm. The daylight duration is ~1890 h/year, with an annual wind speed of 0.7 ~ 1.8 m/s. Annual relative humidity ranges from 57 to 79 % (Herath et al., 2017). The controlled hydrologic Beibei station is gauged at the JRW outlet, and the long-term mean annual runoff is ~ 6.55 × 10^8 m^3 $year^{-1}$, according to the Changjiang Sediment Bulletin. The Digital Elevation Model (DEM), river network, and meteorological stations are shown in Fig.1.

2.2. Data sources

The data required for SWAT modeling and validation consisted of geographic information, meteorological, and hydrological datasets.

2.2.1 Geographic information dataset

The Geographic dataset included the DEM, land use and land cover data (LULC), and soil properties. SRTM 90 m resolution data was the DEM used in this study and was provided by the Geospatial Data Cloud (http://www.gscloud.cn/). LULC data was obtained by manual visual interpretation based on 2010 Landsat TM/ETM remote sensing images, which were preprocessed by Beijing Digital View Technology Co., Ltd, with a spatial resolution of 30 m. The data included six primary classifications—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 spots for repeated interpretation analysis, it is proved that ~~The~~ the cultivated land's classification accuracy was 85 %, and other data classification accuracies reached 75 %. The soil data included a soil type distribution map and soil attribute database. The soil type distribution map is a product of the second national land survey provided by the Nanjing Institute of Soil Research, Chinese Academy of Sciences. It depicted a spatial resolution of 1 km and used the FAO-90 soil classification system. Harmonized World Soil Database v1.2, which was provided by the Food and Agriculture Organization (FAO), was the soil attribute database used in this study and can be downloaded from http://www.fao.org/land-water/databases-and-

software/hwsd/zh/. Most soil attribute data can be obtained directly from HWSD v1.2, such as soil organic carbon content, soil profile maximum root depth, and soil concrete gradation, etc. Parameters that cannot be directly obtained from the HWSD, e.g., texture class, matrix bulk density, field capacity-wilting point, and saturation hydraulic coefficient, can be calculated from the acquired data using the Soil-Air-Water Field & Pond Hydrology model developed by Washington State University. All the above geographic information data were processed by ArcMap 10.2 to obtain 250 m spatial resolution data of the JRW, using the Beijing_1954_GK_Zone_18N Projection coordinate system and GCS_Beijiing_1954 Geographic coordinate system.

2.2.2 Meteorological and hydrological dataset

Daily observed discharges are documented from 1997–2018 at Beibei hydrological station. The daily meteorological records—including precipitation, temperature, relative humidity, sunshine hours, and wind velocity, were measured by 20 meteorological stations in and around the JRW, which were provided by the China meteorological data network (http://data.cma.cn/). The solar radiation data required for establishing the meteorological database was calculated using the sunshine hours (n), and the calculation method consisted of employing the solar radiation (Rs) index in the FAO-56 Penman-Monteith method.

The first gridded format CHIRPS product was released in February 2015, which has first recorded in 1981 and continues to be updated. ~~The most recent gridded format CHIRPS product (V2.0 datasets) was completed and 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. The data is available for download at http://chg.geog.ucsb.edu/data/chirps/. The CHIRPS product is composed of various types of precipitation products, including ground measurement, remote sensing, and reanalysis data—e.g. The Climate Hazards Group Precipitation Climatology (CHPclim) is built from monthly precipitation data supplied by the United Nations FAO, Global Historical Climate Network (CHCN), Cold Cloud Duration from NOAA, and TRMM 3B42 Version 7 from NASA, etc. Essentially, the above-mentioned data were synthesized into 5-day rainfall records, and then the rain gauge observations from multiple data sources were used to correct the deviation, which was further interpolated to daily scaled CHIRPS product. More detailed information about CHIRPS products is available in Funk et al. (2015).

CPC-Global, in gridded format, is the first generation product of NOAA's ongoing CPC unified precipitation project. This product offers daily precipitation estimates from 1998 to the present, at a spatial resolution of 0.5° over land. Daily precipitation data for CPC-Global can be downloaded from: http://ftp.cpc.ncep.noaa.gov/precip/CPC_UNI_PRCP/GAUGE_GLB/V1.0/. This dataset integrates all existing CPC information sources and employs the optimum-interpolation objective analysis technique to form a set of cohesive precipitation products with consistent quantity and enhanced quality. The data was collected from > 30,000 monitoring stations belonging to the WMO Global Telecommunication System, Cooperative Observer Network, and other national meteorological agencies (Xie et al., 2007). For the sake of brevity, the CPC-Global data is referred to as CPC in this article, and the corresponding SWAT model is denoted as the CPC model.

In the SWAT model, all meteorological data were categorized into sub-basins according to the "nearest distance" principle. As such, for the point-formatted gauge observations, a SWAT sub-basin will read the precipitation records from the weather station that is closest to its centroid. Similarly, for the grid-formatted estimates, i.e. CHIRPS and CPC, a SWAT sub-basin will read the precipitation observations from the grid that is closest to its centroid. Using this method, the grid records of high-resolution CHIRPS products within the same sub-basin will be uniformly assigned the grid value closest to the centroid, which will offset the high resolution advantage of CHIRPS products.~~Using this method, some of the grid records are potentially missed, especially for the high resolution CHIRPS products.~~ In order to incorporate the advantages of CHIRPS' spatial resolution and the SWAT model's effectiveness when using the other two products, we selected 400 sub-basins, so that the number of effective CHIRPS is ~ 20 times greater than that of the Gauge and CPC. Note that all of the precipitation statistics in this study are based on the sub-basin scale, which ensures that the precipitation data was correctly categorized in the SWAT model.

2.3. Statistical analysis method

The following indices were selected to statistically compare the three precipitation products:

(1) *CC* value. The correlation coefficient (*CC*) is a numerical measure (ranging from -1 to 1) of the linear statistical relationship between two variables, e.g., simulations and observations, with respect to strength and direction. The closer the

absolute value of *CC* to 1, the higher the correlation between simulation and observation. *CC* is mathematically expressed as follows:

$$CC = \frac{\sum_{i=1}^{n}(Q_i - \overline{Q})(S_i - \overline{S})}{\sqrt{\sum_{i=1}^{n}(Q_i - \overline{Q})^2}\sqrt{\sum_{i=1}^{n}(S_i - \overline{S})^2}}, \quad (1)$$

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 $\overline{Q}$ and $\overline{S}$ are the mean values of the measured and estimated values ~~(or simulated values)~~, respectively. The value may refer to either precipitation (mm) or streamflow discharge (m^3/s).

(2) *STD* value: Standard deviation (*STD*) represents the discretization degree of the datasets (mm). The *STD* of Gauge observations is used to normalize the OPPs' *STD*, and denoted as STD_n to compare the dispersion of OPPs relative to Gauge. The STD_n values range from 0 to ∞ ~~1~~, and the optimal value is ~~0~~1. The ~~*STD*~~ STD_n value ~~is~~ is mathematically expressed as follows:

$$STD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(S_i - \overline{S})^2}, \quad (2)$$

$$STD_n = STD_{OPP} / STD_{Gauge}, \quad (3)$$

Where STD_{OPP} and STD_{Gauge} are the *STDs* of OPPs and Gauge, respectively.

(3) *RSR* value: observations standard deviation ratio (*RSR*) is an error index statistic between the OPPs and Gauge datasets. Root mean square error (*RMSE*) divided by *STD* values would derive the *RSR* value. *RSR* has a range from 0 to ∞ with 0 as the optimal value. The calculation equation is expressed as follows: ~~(3) *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 +∞, and an optimal value of 0. The *RMSD* value is expressed as follows:~~

$$RSR = \frac{RMSE}{STD} = \frac{\sqrt{\sum_{i=1}^{n}(S_i - Q_i)^2}}{\sqrt{\sum_{i=1}^{n}(S_i - \overline{S})^2}}, \quad (4)$$

~~$$RMSD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[(Si - \overline{S}) - (Qi - \overline{Q})\right]^2}, \quad (4)$$~~

(4) *POD* value: Probability of detection (*POD*) is a ratio that reflects the number of times OPPs correctly detected the frequency of rainfall relative to the total number of rainfall events. *POD* has a range from 0 to 1, and an optimal value of 1. It is mathematically expressed as follows:

$$POD = \frac{t_H}{t_H + t_M}, \quad (5)$$

Where t is the number of qualified data pairs; H denotes when the Gauge and OPPs both detect a rainfall event; and M represents when Gauge detects a rainfall event, but the OPPs do not.

(5) *FAR* value: False-alarm rate (*FAR*) represents the frequency that precipitation is detected using OPPs, but not detected using Gauge. *FAR* has a range from 0 to 1, and an optimal value of 0. The *FAR* value is mathematically expressed as follows:

$$FAR = \frac{t_F}{t_H + t_F}, \quad (6)$$

Where F denotes that the OPPs detected a rainfall event, while the Gauge does not.

2.4. Hydrological model

2.4.1 SWAT description

The process-based SWAT model is an all-inclusive, temporally uninterrupted, and semi-distributed simulation that was developed by the Agricultural Research Service of the United States Department of Agriculture (Arnold et al., 1998; Arnold & Fohrer, 2005). The model's smallest simulation unit is the Hydrological Response Unit (HRU). Fields with a specific LULC and soil, which may be scattered throughout a sub-basin, are lumped together in one HRU. The model assumes that there is no interaction between HRUs in any one sub-basin. According to the water balance cycle, the HRU hydrologic process is first calculated and then summed to obtain the total hydrologic process of the sub-basin (Arnold et al., 1998; Arnold et al., 2012). 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:

$$SW_t = SW_0 + \sum_{i=1}^{t}(P - Q_{surf} - ET - W_{seep} - Q_{lat} - Q_{gw}), \quad (7)$$

Where, SW_t is the soil's water content at the end of period t (mm); SW_0 is the soil water content at the beginning of period t (mm); and t is the calculation period length. P = precipitation; Q_{surf} = surface runoff; ~~E_a~~ ET = actual evapotranspiration; W_{seep} = the amount of percolation and by-pass flow exiting the soil profile bottom; Q_{lat} = lateral flow, and Q_{gw} = base-flow, including return flow from the shallow aquifer (GW_Q) and flow out from the deep aquifer (GW_Q_D)—all on day i in (mm).

Surface runoff, lateral flow, and base-flow add up to Water Yield (WYLD). SWAT uses the Soil Conservation Services-Curve Number method (SCS-CN) to simulate surface runoff. Surface runoff is mathematically expressed as follows:

$$\frac{F}{S} = \frac{Q_{surf}}{P - I_a}, \tag{8}$$

Where I_a is the initial loss (mm)—i.e., precipitation loss before surface runoff; F is the final loss— i.e., precipitation loss after surface runoff is generated; S is the maximum possible retention in the basin at that time (mm), and is the upper limit of F.

The WYLD from the sub-basin forms in the connected channel network, then, through routing water processes, enters the downstream reach segment and repeats the water balance process, eventually converging at the drainage outlet. This is the watershed hydrological process simulated in SWAT that is shown in Fig. 2.

2.4.2 SWAT calibration and validation

In this study, the period of calibration and validation are set for 1999-2008 and 2009-2018, respectively. The two years before both the calibration and validation periods are delineated as the warmup phase for the purpose of initializing model state variables, e.g., soil moisture and groundwater concentration. The model is automatically calibrated using the Sequential Uncertainty Fitting algorithm version 2 (SUFI-2) in the SWAT- Calibration and Uncertainty Program (SWAT-CUP). This algorithm has been successfully applied in many related studies (Abbaspour et al., 2017; Shivhare et al., 2018; Tuo et al., 2018). The model parameters and the initial calibration range are shown in Table 1. The parameters were selected based on literature published by Arnold et al. (2012) and Tuo et al. (2016) and the official manual. In order to consider the impact of elevation on precipitation and the fact that precipitation in the OPPs is horizontal, we introduced the Precipitation Lapse Rate (PLAPS) parameter and divided each sub-basin into 10 Elevation Bands. Tuo et al (2016) demonstrated that this method is able to correct the rainfall error caused by ignoring elevation and effectively improve the model's simulation performance. Furthermore, the

models were calibrated following three iterations of 1000 times each. Following each iteration, the SWAT-CUP generated a fresh set of parameter ranges. This new set was used for the next iteration after considering the upper and lower bounds of the physical meaning. The above method was repeated for the three different precipitation data sets. The Nash efficiency coefficient (*NSE*) was used as the objective function to optimize the model calibration, and is mathematically expressed as follows:

$$NSE = \frac{\sum_{i=1}^{n}(S_i - Q_i)^2}{\sum_{i=1}^{n}(Q_i - \overline{Q})^2}, \quad (9)$$

The model performance was classified using *RSR*, *NSE*, and percentage bias (*PBIAS*) values defined by Moriasi et al. (2007), which is shown in Table 2. Parameter ranges of ~~T~~the three models~~'~~ ~~parameter range~~ after 3000 iterations ~~is~~are shown in Table 3. *CC*, *NSE*, *PBIAS* and *RSR* were used to evaluate the model simulation results. ~~The model performance was classified using the *NSE* values defined by Moriasi et al. (2007): unsatisfactory performance (*NSE* ≤ 0.50), satisfactory performance (0.50 < *NSE* ≤ 0.65), good performance (0.65 < *NSE* ≤ 0.75) and very good performance (0.75 < *NSE* ≤ 1.00). The three models' parameter range after 3000 iterations is shown in Table 2. *CC*, *NSE*, and Percentage bias (*PBIAS*) were used to evaluate the model simulation results.~~

PBIAS describes the OPPs' systematic bias (%). *PBIAS* ranges from 0 to +∞ %, and the optimal value is 0 %. The calculation equation is expressed as follows:

$$PBIAS = \frac{\sum_{i=1}^{n}(S_i - Q_i)}{\sum_{i=1}^{n} Q_i} \times 100\%, \quad (10)$$

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 *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.

3. Results

3.1. Evaluation of the OPPs on a temporal scale

3.1.1. Monthly scale

A comparison of the monthly precipitation time series (Gauge, CHIRPS, and CPC) across the watershed is shown in Fig. 3. Note that the time series in Fig. 3 represents the average value of the whole watershed and was calculated as follows: the original point or grid formatted rainfall records were first categorized into every SWAT model sub-basin, according to the nearest distance principle, and then spatially synthesized into one time series by sub-basin area using the weighted average method. Fig. 3 shows that the rainfall values in the rainy season (especially in July)~~the extreme rainfall values~~ captured by Gauge are higher than those of CHIRPS and CPC. Moreover, the *CC* values between CHIRPS and Gauge records, as well as CPC and Gauge records, are 0.97 and 0.98 ($P<0.01$, i.e., extremely significant positive correlation), respectively. The high *CC* values demonstrate the highly correlated linear relationship between the two OPPs and Gauge records on a monthly scale, indicating that both CHIRPS and CPC products are equally as effective at describing the monthly precipitation variation within the JRW as the Gauge records.

The box diagrams of the three precipitation records are shown in Fig. 4. Note that July is the largest contributor to the yearly precipitation, as well as the annual flood peak calculation. According to the July results, when compared with Gauge records, the CHIRPS product has a large median, small maximum, and large minimum, while the three CPC values are all smaller than the Gauge values. These characteristics will potentially lead to different hydrological modeling in flood peak simulation. The STD_n values for Gauge-CHIRPS and Gauge-CPC are 1.06 and 0.94, respectively. The ~~*RMSD*~~ *RSR* values for

Gauge-CHIRPS and Gauge-CPC are ~~15.80~~0.27 and ~~12.95~~0.22, respectively. These statistics indicate that both CHIRPS and CPC estimates are able to provide equally effective precipitation values compared with that of the Gauge records. Nevertheless, *PBIAS* values of Gauge-CHIRPS and Gauge-CPC were 9.58 % and -6.70 %, respectively, indicating the overestimation of CHIRPS products and underestimation of CPC products compared with that of the Gauge records. Specifically, overestimation of the CHIRPS products mainly occurs between April and September, which is the JRW rainy season; while during the dry season, i.e., October - March, the CHIRPS estimates are closely consistent with the Gauge records. In contrast, the CPC estimates parallel the Gauge records for the rainy season, yet rainfall is underestimated during dry season.

3.1.2. Daily scale

Intensity and frequency are the most critical parameters for characterizing rainfall features on a daily scale (Azarnivand et al., 2019; Wen et al., 2019). The scatter plots in Fig. 5 depict a precipitation intensity comparison between the OPPs and Gauge records, at a daily scale, at Beibei hydrological station (NO.411). Based on Fig. 5, the angle between the CHIRPS' 95 % line estimates and the horizontal axis is > 45 degrees (the 1:1 line), indicating that CHIRPS overestimates precipitation relative to the Gauge records. CPC estimates demonstrate the exact opposite. More specifically, the scatter distribution indicates that compared to the Gauge records, the CHIRPS' estimates tend to overestimate precipitation for light rains, and underestimate it for heavy rains. Meanwhile, the CPC products underestimate both light and heavy rains. Statistically, the *CC*, STD_n, and ~~*RMSD*~~*RSR* values between CHIRPS and the Gauge records are 0.53, 1.14, and ~~5.16~~1.04, respectively, and 0.64, 0.87, and ~~3.95~~0.80, respectively, between the CPC and Gauge products. At the daily scale, the OPPs and Gauge products showed an evident decrease in consistency when compared with the monthly scale. The above indicators demonstrate the deceased consistency of the OPPs' estimates and Gauge observations; while the CPC shows superior performance relative to the CHIRPS.

The three precipitation products' cumulative daily precipitation intensity frequencies are shown in Fig. 6. Note that on the right side of the figure, the 50 mm/day demarcation divides the horizontal axis into two sections of rainfall intensity, in order to depict the three products' frequency trends more clearly. Overall, the three products display a high probability of occurrence for precipitation intensity of 0.1~25 mm/day—87 %, 94 %, and 98 % for CHIRPS, Gauge, and CPC, respectively.

However, the probability for precipitation intensity > 100 mm/day is 99.70 %, 99.73 % and 99.99 %, respectively, indicating the potential upper limit extreme rainfall event value within this area. The CPC product fails to detect extreme rainstorm events.

Table 4 is the recognition capability evaluation of the two OPPs for rainfall intensity between 0.1 and 50 mm and > 50 mm events. The CPC and CHIRPS *POD* values for rainfall intensity between 0.1 and 50 mm are 83.53 % and 27.29 %, respectively, demonstrating that the CPC product has a strong ability to capture the onset of rainfall. Nevertheless, the CPC *POD* value for rainfall intensities≥ 50mm decreases to 9.42 %, indicating its poor ability to capture rainstorms, while that of the CHIRPS product is relatively higher, with a *POD* value of 18.12 %. Moreover, the *FAR* fractions for both OPPs are between 44 % and 66 %, demonstrating its lower ability to detect rainstorm values.

3.2. Evaluation of OPPs on a spatial scale

Spatial variations of the three products' long-term mean annual precipitation, for all of the partitioned sub-basins, are shown in Fig. 7. The three products' precipitation values exhibit an obvious upward trend from the JWR's upstream to downstream region. The precipitation's shifting pattern in space highly correlates with the topography variation (shown in Fig. 1), indicating that meteorological and hydrologic variables throughout the region are potentially influenced by and respond to catchment landscape modification. It should be noted that in Fig. 7(a), all the sub-basins are divided into several regions, each of which has the same rainfall value, while abrupt rainfall value changes occurred between adjacent regions. In Fig. 7(c), the transition between adjacent sub-basins is smoother than that observed with Gauge. In Fig. 7(b), the CHIRPS product shows the smoothest precipitation transition between the adjacent sub-basins, illustrating the advantages of the high resolution CHIRPS product. In this study, precipitation records from 20 rain gauge stations, 411 CHIRPS grids, and 76 CPC grids were categorized into the SWAT sub-basins (as mentioned in section 2.3), leading to differences in the continuity or smoothness of rainfall spatial distribution among the three products. Compared with the Gauge observations, the overall precipitation values estimated by the CHIRPS are relatively higher, while that of the CPC is relatively lower.

The *CC*, STD_n, and *RSR* values of precipitation spatial distribution between CHIRPS and Gauge are 0.89, 0.96, and 0.55, respectively, and 0.82, 0.87, 0.62 between CPC and Gauge, respectively. These statistics indicate that both CHIRPS and CPC estimates can describe the spatial distribution of precipitation in JRW, among which CHIRPS depicts better performance. The

correlation coefficients between the Gauge and OPPs at monthly or daily scales for every sub-basin are illustrated in Fig. 8. The correlation coefficients' spatial variation between the Gauge and OPPs at monthly and daily scales are illustrated in . Overall, the monthly scale *CC* values (with a rang of 0.7~1) are comparably larger than that of the daily scale values (with a rang of 0.5~0.7). Spatially, the higher *CC* values between the Gauge and CPC at the monthly scale are mainly distributed in areas with comparably low or high rainfall amounts, such as Wudu and Wangyuan. Yet, the *CC* value was less relevant in areas with moderate rainfall (e.g., Suining) relative to that of the Gauge and CHIRPS. However, at the daily scale, the correlation of Gauge and CPC is higher than that of Gauge and CHIRPS, except for a few individual sub-basins located in the east-south area.

3.3. Hydrological performance of different precipitation products in the SWAT model

3.3.1 Spatio-temporal performance at a monthly scale

OPPs ignores terrain differences when forcing the model, which may increase potentially systematic errors in hydrologic modeling (Tuo et al., 2016). Thus, in this work, elevation bands (see Sect. 2.4.1) were used to normalize precipitation at different elevations. The monthly observed runoff and simulated runoff subjected to this procedure, and used for the SWAT model during the calibration and validation periods, are presented in Fig. 9. The results show that the three precipitation inputs successfully stimulated the model to reproduce the discharge records at the Beibei station; the rising and falling simulated flood event processes are in good agreement with that of the observed ones. 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 validation 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) values, while CPC only reached the level of "Good" due to higher *PBIAS* (10.8 %) (Fig. 9). 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. Further, among all the three models, 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. The underestimation of the peak flows during flood seasons (June to August), would be the main reason of the slightly worse performance of the two OPPs inputs. The Gauge model demonstrates the best performance, which may reflect its strong ability to ascertain the peak rainfall during the flood seasons (Fig. 4). Note that in Fig. 4, the CHIRPS medians are larger than that of the Gauge, while the maxima are smaller, and the minima are larger during the flood seasons. These features facilitated the best performance for describing the base-flow and medium floods, like those in years 2003 and 2014. As a result, the CHIRPS model achieved the best simulation base–flow–; although it overestimated the precipitation with light rain intensity, and obviously overestimated the streamflow with discharge< 6000m^3/s, which also led to its final performance deviation. CPC showed significant overestimation in 2017 and 2018 during the validation~~verification~~ period. Although it approximated CHIRPS' estimated results, it clearly deviated from its previous tendency to underestimate precipitation during these two years.

In terms of simulated WYLD spatial variation at the sub-basin scale (as shown in Fig. 10), the consistency of the Gauge and CHIRPS models is slightly better than that of the CPC model, potentially demonstrating the advantage of the CHIRPS' high resolution in simulating precipitation. Furthermore, the WYLD distribution pattern is highly consistent with the corresponding precipitation distribution (Fig. 7). The spatial correlation between WYLD and precipitation for rainfall for the Gauge, CHIRPS, and CPC products reached 0.85, 0.84, and 0.91~~84.8 %, 84.3 %, and 90.84 %~~, respectively. Compared to the Gauge simulation, the CHIRPS overestimated and the CPC underestimated the WYLD. The *PBIAS* values for Gauge-CHIRPS and Gauge-CPC are 5.85 % and -5.38 %, respectively.

3.3.2. Spatial-temporal performance at a daily scale

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. 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 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. The performances in describing the peak flows ~~are~~ were 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. As shown in Fig~~s~~. 5 and Fig. 6, both the CHIRPS and CPC underestimate heavy rainfall intensities compared with the Gauge observations. Conversely, the CHIRPS model performs best in simulating the base-flow, since CHIRPS tend to capture higher values of light rainfalls than that of the Gauge and CPC.

~~Note that at a daily scale, the three evaluation parameters are significantly smaller than at the monthly scale. The primary reason is that the daily scale sample size is nearly 30 times larger than that of the monthly scale, so the cumulative systematic deviation led to poorer evaluation parameters. Therefore, in order to better compare the simulation results of the two scales, the streamflow discharge daily process was integrated into the monthly process, so that it had the same sample size as the monthly model. As shown in Fig. 12, the three types of precipitation models maintained their inherent advantages and attained equal or superior performance at the daily scale.~~ The WYLD spatial variation for all sub-basins at the daily scale are shown in ~~Fig. 13~~Fig. 12. The spatial variation of daily and monthly scale of the three precipitation datasets is basically the same. The *CC* values between the daily and monthly WYLD spatial variation for Gauge, CHIRPS, and CPC at the daily scale are 0.98, 0.99 and 0.97, respectively. Similar to the results from the monthly scale, the WYLD spatial pattern is highly correlated with that of precipitation. The *CC* values between the WYLD and precipitation for Gauge, CHIRPS, and CPC at the daily scale are 0.83~~.45 %~~, 0.84~~.41 %~~ and 0.92~~1.70 %~~, ~~83.45 %, 84.41 % and 91.70 %,~~ respectively, which are even higher than those of the monthly scale. The *CC*, STD_n, and ~~*RMSD*~~ *RSR* values between CHIRPS and Gauge are 0.92, 1.06, and ~~0.23~~0.46, respectively, and 0.81, 0.94, ~~0.33~~0.66 between CPC and Gauge, respectively.

4. Discussion

4.1. Comparison of precipitation products in terms of rainfall events with different magnitudes

In the aforementioned results, compared to Gauge product, the CHIRPS tends to overestimate the intensity and frequency of light rain, but underestimate heavy rain, which is consistent with the results reported by Gao et al. (2018). However, CPC tends to underestimate the intensity and overestimate the frequency of rainfall for light and heavy rain, although light rain is more underestimated. These results are consistent with those reported by Ajaaj et al. (2019). The differences in the capture of different magnitude rainfall intensities may potentially influence hydrologic process and forecasting. From Eq. (8), the basin's WYLD is directly proportional to the amount of precipitation. In other words, heavy rainfall tends to produce large amounts of streamflow. Duan J et al. (2019) conducted a slope experiment and found that there was a significant difference in the runoff coefficients between extreme rainfall events and normal rainfall events; the former produced much more runoff and sediment than the latter. 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~~early stage~~ conditions. After extreme flood events, antecedent~~early stage~~ conditions had no effect on rainfall-runoff processes, and rainfall significantly affected the streamflow discharge. Moreover, the evaluation index *NSE* performance is mainly determined by the peak streamflow. Thus, it is critical to identify the magnitudes of different rainfall events at both a temporal and spatial scale. As a consequence, we derived the temporal and spatial distributions of the rainfall events with different magnitudes. The spatial scale dimension was implemented by identifying the serial numbers of all sub-basins (~~Fig. 14~~Fig. 13); and the temporal dimension was fulfilled by detecting rainfall events of different magnitudes throughout the study period (~~Fig. 15~~Fig. 14).

Overall, ~~Fig.15~~ Fig.14 shows that the CPC tends to capture more light rainfall events with precipitation intensities between 0.1 and 50 mm/day (LR events), the CHIRPS identified more medium rainfall events with precipitation intensities between 50 and 100 mm/day (MR events), and both the Gauge and CHIRPS detected more heavy rainfall events with precipitation intensities larger than 100 mm/day (HR events). Accordingly, the total annual precipitation amounts of the three products are ranked as CHIRPS (956.4 mm) > Gauge (872.8 mm) > CPC (814.3 mm). Even with the advantage of detecting MR and HR events, the CHIRPS' ability to simulate flood events is inferior to that of the Gauge. Potential reasons may consist

of: 1) the HR events detected by CHIRPS are more scattered at a temporal scale, which disperses the flood peak value; and 2) the high frequency of the MR detected by CHIRPS resulted in parameter sets in the SWAT model that tended to derive a lower runoff coefficient, in order to avoid a large systematic bias in terms of *PBIAS*.

The CPC estimates, 60% of which are detected as LR, tend to be incapable of driving the SWAT model to capture small streamflow discharge, especially the ones equivalent to base-flow. Consequently, the CPC model *CC* values are relatively low. 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~~low proletarian flow and thus WYLD~~. Furthermore, the CHIRPS has a high probability of MR event false-alarm, which is consistent with the results reported by Zambrano-Bigiarini et al. (2017). Thus, a significant number of erroneous peaks exist in the CHIRPS, just like the temporal variation at a daily scale, which has a very low correlation with the Gauge. Erroneous precipitation peaks tend to produce erroneous streamflow peaks. Although SWAT can repair the peak position deviation to some extent, the *CC* is inevitably reduced.

4.2. Effect of OPPs difference on hydrological process simulation

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 modeling are also different. As such, it is critical that researchers and decision makers adequately understand the benefits and limitations of different precipitation products in modeling the hydrologic processes.

According to the SWAT model's water balance equation (Eq. ~~9~~7), WYLD equals the sum of Q_{surf}, Q_{lat} and Q_g, where Q_{gw} can be divided into flow out from a shallow aquifer (GW_Q) and flow out from a deep aquifer (GW_Q_D). If the soil water content W and percolation/bypass flow into the deep aquifer w_{seep} remains unchanged over a long time period, then the equation is modified to $P = Q_{surf} + ~~E_a~~ET + Q_{lat} + Q_{gw}$. Thus, we calculated the water balance component portions, Q_{surf}, Q_{lat}, Q_{gw}, and ~~E_a~~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. 15 & 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. ~~It is evident from and 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 highly possibly related in parameter adjustment. As shown in Table 3, 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. ~~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, resulting in significant differences in evaporation and storage estimates.

Second, rainfall characteristics also have a significant impact on hydrological processes in the watershed. A large number of studies show that rainfall intensity is a key player in the watershed's hydrological process (Zhou et al., 2017; Du et al., 2019; Zhang et al., 2019). Studies conducted by Redding & Devito (2010) showed that the occurrence of lateral flow is mainly determined by rainfall intensity. When the rainfall intensity is greater than the surface soil hydraulic conductivity, the rainfall mainly forms surface runoff. When rainfall intensity is between the soil surface and bedrock hydraulic conductivity, the rainfall mainly forms lateral flow. When rainfall intensity is less than the bedrock's hydraulic conductivity, the rainfall will infiltrate into the groundwater. In this study, the precipitation recorded by CHIRPS was mainly distributed between 25 and 100 mm/day, while that of CPC was mainly distributed between 0.1 and 25 mm/day. This may be the reason why CHIRPS overestimated the proportion of surface runoff and CPC overestimated the proportion of lateral flow, compared with that of the Gauge model. 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, hydrologic components exhibited distinct behaviors 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.

5. Conclusions

The sparsity and unevenness of ground-based precipitation observations pose great challenges to the establishment of hydrological models. In this study, the Gauge, CHIRPS, and CPC were evaluated by statistically comparing the different

precipitation products as well as their performance at driving the hydrological model. Specifically, the potential behaviors of different precipitation datasets in describing precipitation magnitudes and hydrologic processes in terms of water balance components are further discussed. The main conclusions are summarized as follows：

1. The three precipitation datasets exhibited similar temporal records at a monthly scale, and the Gauge measures were more capable of capturing maxima than those of the OPPs. During rainy seasons, the CHIRPS median, maxima, and minima were larger, smaller, and larger, respectively, than that of the Gauge records, while all three CPC statistical values were smaller than that of the Gauge. At a daily scale, the CHIRPS tends to overestimate light rains and underestimate heavy rains, while all the CPC rainfall intensities were underestimated. Spatially, precipitation in all sub-basins increases from the upstream to the downstream region, and the CHIRPS derives the most smoothly distributed precipitation pattern.

2. All three precipitation inputs successfully forced the model to replicate the discharge records at the Beibei station at ~~a~~ monthly and daily scale, and results at monthly scale presented slightly better performance than that of daily scale.~~although they performed slightly better at the daily scale.~~ However, the differences of precipitation inputs in the statistics at ~~the~~ monthly and daily scales correspondingly affected the streamflow hydrograph, e.g. flood peak, base flow, and the rising and falling processes.~~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.~~ Overall, the CHIRPS dataset performs better in hydrological evaluation because of its lower uncertainty level and higher spatial accuracy than that of CPC, thus it can be a fairly good ~~choice~~ option 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 3) were quite different. In other words, evaluating only the streamflow simulation accuracy of the precipitation products will conceal the differences between these precipitation products, which is primarily because that hydrological models are able to offset the influences of precipitation inputs on streamflow simulations using parameter calibration and validation.

3. The calibrated parameters are adjusted to alter the hydrologic mechanics in terms of water balance components. Thus, they effectively fill the potential gaps in the WYLD that may be introduced by the varying precipitation amounts and intensities detected by different precipitation products. 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. Multi-objective calibration would be recommended for hydrological modelers in parameter calibration and optimization, especially for large and spatial resolved watersheds. Lastly, the spatial precipitation pattern also significant impacted the spatial 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 modeling 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 reanalyzed data. ~~Furthermore, f~~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.

Data availability

Sources of the geospatial and climate forcing data used to configure the SWAT model have been described in the "2.2. Data sources". The simulations of streamflow data are shown in the figures of the paper and are available through contacting the authors.

Author contributions

HZ conceptualized the work. HZ and JP designed the core structure and collected the data required. JP conducted the data analyses and model simulation and drafted the manuscript. HZ revised the manuscript. QX collected data and contributed to key analyses and discussions. YJW, YQW and OZ contributed to the discussion of the results, and JH assisted carrying out model configuration and data analyses.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

This study was supported by the Fundamental Research Funds for the Central Universities (NO.2016ZCQ06 and NO.2015ZCQ-SB-01), the Joint Fund of State key Lab of Hydroscience and Institute of Internet of Waters Tsinghua-Ningxia Yinchuan (sklhse-2019-Iow08), the National Natural Science Foundation of China (51309006, 41790434), and the National Major Hydraulic Engineering Construction Funds “Research Program on Key Sediment Problems of the Three Gorges Project” (12610100000018J129-01). We gratefully acknowledge the Beijing Municipal Education Commission for their financial support through Innovative Transdisciplinary Program "Ecological Restoration Engineering". We also sincerely thank anonymous reviewers and editor for their constructive comments and suggestions.

References

Abbaspour, K., Vaghefi, S., and Srinivasan, R.: 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, 2017.

Ajaaj, A. A., Mishra, A. K., and Khan, A. A.: Evaluation of Satellite and Gauge-Based Precipitation Products through Hydrologic Simulation in Tigris River Basin under Data-Scarce Environment, J. Hydrol. Eng., 24, 05018033, https://doi.org/10.1061/(asce)he.1943-5584.0001737, 2019.

Alijanian, M., Rakhshandehroo, G. R., Mishra, A. K., and Dehghani, M.: Evaluation of satellite rainfall climatology using CMORPH, PERSIANN-CDR, PERSIANN, TRMM, MSWEP over Iran, Int. J. Climatol., 37, 4896-4914, https://doi.org/ 10.1002/joc.5131, 2017.

Arnold, J. G., Srinivasan, R., Muttiah, R. S., and Williams, J. R.: Large area hudrologic modeling and assessment part I: model development1, J. Am. Water. Resour. As., 34, 73-89, https://doi.org/ 10.1111/j.1752-1688.1998.tb05961.x, 1988.

Arnold, J. G. and Fohrer, N.: SWAT2000—Current capabilities and research opportunities in applied watershed modelling, Hydrol. Process., 19, 563-572, https://doi.org/ 10.1002/hyp.5611, 2005.

Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., Santhi, C., Harmel, R. D., van Griensven, A., Liew, M. W. V., and Jha, M. K.: SWAT: Model Use, Calibration, and Validation, T. ASABE., 55, 1491–1508, https://doi.org/10.13031/2013.42256, 2012.

Ayana, E. K., Worqlul, A. W., and Steenhuis, T. S.: Evaluation of stream water quality data generated from MODIS images in modeling total suspended solid emission to a freshwater lake, Sci. Total. Environ., 523, 170-177, https://doi.org/10.1016/j.scitotenv.2015.03.132, 2015.

Azarnivand, A., Camporese, M., Alaghmand, S., and Daly, E.: Simulated response of an intermittent stream to rainfall frequency patterns, Hydrol. Process., 13610, 1-18, https://doi.org/10.1002/hyp.13610, 2019.

Bai, P. and Liu, X.: Evaluation of Five Satellite-Based Precipitation Products in Two Gauge-Scarce Basins on the Tibetan Plateau, Remote Sensing, 10, 1316, https://doi.org/10.3390/rs10081316, 2018.

~~Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A. I. J. M., Weedon, G. P., Brocca, L., Pappenberger, F., Huffman, G. J., and Wood, E. F.: Global scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling, Hydrol. Earth. Syst. Sc., 21, 6201-6217, https://doi.org/ 10.5194/hess-21-6201-2017, 2017a.~~

Beck, H. E., van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., and de Roo, A.: MSWEP: 3-hourly 0.25° global gridded precipitation (1979~2015) by merging gauge, satellite, and reanalysis data, Hydrol. Earth. Syst. Sc., 21, 589–615, https://doi.org/10.5194/hess-21-589-2017, 2017~~b~~.

Belete, M., Deng, J., Wang, K., Zhou, M., Zhu, E., Shifaw, E., and Bayissa, Y.: Evaluation of satellite rainfall products for modeling water yield over the source region of Blue Nile Basin, Sci. Total. Environ., 134834, https://doi.org/10.1016/j.scitotenv.2019.134834, 2019.

Bhuiyan, M. A. E., Nikolopoulos, E. I., Anagnostou, E. N., Polcher, J., Albergel, C., Dutra, E., Fink, G., Torre, A. M., and Munier, S.: Assessment of precipitation error propagation in multi-model global water resource reanalysis, Hydrol. Earth. Syst. Sc., 23, 1973-1994, https://doi.org/ 10.5194/hess-23-1973-2019, 2019.

Bohnenstengel, S. I., Schlunzen, K. H., and Beyrich, F.: Representativity of in situ precipitation measurements - A case study

for the LITFASS area in North-Eastern Germany, J. Hydrol., 400, 387-395, https://doi.org/ 10.1016/j.jhydrol.2011.01.052, 2011.

Cecinati, F., Moreno-Ródenas, A. M., Rico-Ramirez, M. A., ten Veldhuis, M. C., and Langeveld, J. G.: Considering Rain Gauge Uncertainty Using Kriging for Uncertain Data, Atmosphere, 9, 446, https://doi.org/10.3390/atmos9110446, 2018.

Cornelissen, T., Diekkruger, B., and Bogena, H. R.: Using High-Resolution Data to Test Parameter Sensitivity of the Distributed Hydrological Model HydroGeoSphere, Water, 8, 22, https://doi.org/ 10.3390/w8050202, 2016.

Du, J., Niu, J., Gao, Z., Chen, X., Zhang, L., Li, X., and Zhu, Z.: Effects of rainfall intensity and slope on interception and precipitation partitioning by forest litter layer, CATENA, 172, 711–718, https://doi.org/ 10.1016/j.catena.2018.09.036, 2019.

Duan, J., Liu, Y. J., Yang, J., Tang, C. J., and Shi, Z. H.: Role of groundcover management in controlling soil erosion under extreme rainfall in citrus orchards of southern China, J. Hydrol., 124290, https://doi.org/10.1016/j.jhydrol.2019.124290, 2019.

Duan, Z., Liu, J., Tuo, Y., Chiogna, G., and Disse, M.: Evaluation of eight high spatial resolution gridded precipitation products in Adige Basin (Italy) at multiple temporal and spatial scales, Sci. Total. Environ., 573, 1536-1553, http://doi.org/10.1016/j.scitotenv.2016.08.213, 2016.

Duan, Z., Tuo, Y., Liu, J., Gao, H., Song, X., Zhang, Z., Yang, L., and Mekonnen, D. F.: Hydrological evaluation of open-access precipitation and air temperature datasets using SWAT in a poorly gauged basin in Ethiopia, J. Hydrol., 569, 612-626, https://doi.org/10.1016/j.jhydrol.2018.12.026, 2019.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, S., Rowland, J., Harrison, L., Hoell, A., and Michaelsen, J.: The climate hazards infrared precipitation with stations - a new environmental record for monitoring extremes, Sci. Data., 2, 150066, https://doi.org/ 10.1038/sdata.2015.66, 2015.

Gabriel, M., Knightes, C., Dennis, R., and Cooter, E.: Potential Impact of Clean Air Act Regulations on Nitrogen Fate and Transport in the Neuse River Basin: a Modeling Investigation Using CMAQ and SWAT, Environ. Model. Assess., 19, 451-465, https://doi.org/10.1007/s10666-014-9410-x, 2014.

Galván, L., Olías, M., Izquierdo, T., Cerón, J. C., and Villarán, R. F.: Rainfall estimation in SWAT: An alternative method to

simulate orographic precipitation, J. Hydrol., 509, 257-265, https://doi.org/ 10.1016/j.jhydrol.2013.11.044, 2014.

Gao, Z., Long, D., Tang, G., Zeng, C., Huang, J., and Hong, Y.: Assessing the potential of satellite-based precipitation estimates for flood frequency analysis in ungauged or poorly gauged tributaries of China's Yangtze River basin, J. Hydrol., 550, 478–496, https://doi.org/10.1016/j.jhydrol.2017.05.025, 2017.

Gao, F., Zhang, Y., Ren, X., Yao, Y., Hao, Z., and Cai, W.: Evaluation of CHIRPS and its application for drought monitoring over the Haihe River Basin, China, Nat. Hazards., 92, 155–172, https://doi.org/10.1007/s11069-018-3196-0, 2018.

Herath, I. K., Ye, X., Wang, J., and Bouraima, A. K.: Spatial and temporal variability of reference evapotranspiration and influenced meteorological factors in the Jialing River Basin, China, Theor. Appl. Climatol., 131, 1417–1428, https://doi.org/10.1007/s00704-017-2062-4, 2017.

Huang, Y., Bárdossy, A., and Zhang, K.: Sensitivity of hydrological models to temporal and spatial resolutions of rainfall data, Hydrol. Earth. Syst. Sc., 23, 2647-2663, https://doi.org/ 10.5194/hess-23-2647-2019, 2019.

Hwang, Y., Clark, M. P., and Rajagopalan, B.: Use of daily precipitation uncertainties in streamflow simulation and forecast, Stoch. Env. Res. Risk. A., 25, 957–972, https://doi.org/10.1007/s00477-011-0460-1, 2011.

Jin, X., He, C., Zhang, L., and Zhang, B.: A Modified Groundwater Module in SWAT for Improved Streamflow Simulation in a Large, Arid Endorheic River Watershed in Northwest China, Chinese. Geogr. Sci., 28, 47-60, https://doi.org/ 10.1007/s11769-018-0931-0, 2018.

Jiang, S., Ren, L., Xu, C., Yong, B., Yuan, F., Liu, Y., Yang, X., and Zeng, X.: Statistical and hydrological evaluation of the latest Integrated Multi-satellite Retrievals for GPM (IMERG) over a midlatitude humid basin in South China, Atmos. Res., 214, 418-429, https://doi.org/ 10.1016/j.atmosres.2018.08.021, 2018.

Jiang, L. and Bauer-Gottwein, P.: How do GPM IMERG precipitation estimates perform as hydrological model forcing? Evaluation for 300 catchments across Mainland China, J. Hydrol., 572, 486-500, https://doi.org/ 10.1016/j.jhydrol.2019.03.042, 2019.

Lai, C., Zhong, R., Wang, Z., Wu, X., Chen, X., Wang, P., and Lian, Y.: Monitoring hydrological drought using long-term satellite-based precipitation data, Sci. Total. Environ., 649, 1198-1208, https://doi.org/ 10.1016/j.scitotenv.2018.08.245, 2019.

Li, D., Christakos, G., Ding, X., and Wu, J.: Adequacy of TRMM satellite rainfall data in driving the SWAT modeling of Tiao xi catchmen, J. Hydrol., 556, 1139-1152, https://doi.org/ 10.1016/j.jhydrol.2017.01.006, 2018.

Liu, J. B., Kummerow, C. D., and Elsaesser, G. S.: Identifying and analysing uncertainty structures in the TRMM microwave imager precipitation product over tropical ocean basins, Int. J. Remote. Sens., 38, 23-42, https://doi.org/ 10.1080/01431161.2016.1259676, 2017.

Lobligeois, F., Andréassian, V., Perrin, C., Tabary, P., and Loumagne, C.: When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation using 3620 flood events, Hydrol. Earth. Syst. Sc., 18, 575-594, https://doi.org/ 10.5194/hess-18-575-2014, 2014.

Long, Y. P., Zhang, Y. N., and Ma, Q. M.: A Merging Framework for Rainfall Estimation at High Spatiotemporal Resolution for Distributed Hydrological Modeling in a Data-Scarce Area, Remote Sensing, 8, 7, https://doi.org/ 10.3390/rs8070599, 2016.

Lu, Y. J., Jiang, S. H., Ren, L. L., Zhang, L. Q., Wang, M. H., Liu, R. L., and Wei, L.Y.: Spatial and Temporal Variability in Precipitation Concentration over Mainland China, 1961–2017, Water, 11, 881, https://doi.org/10.3390/w11050881, 2019.

Luo, X., Wu, W., He, D., Li, Y., and Ji, X.: Hydrological Simulation Using TRMM and CHIRPS Precipitation Estimates in the Lower Lancang-Mekong River Basin, Chinese. Geogr. Sci., 29, 13–25, https://doi.org/10.1007/s11769-019-1014-6, 2019.

Meng, C. C., Zhang, H. L. Wang, Y. J., Wang, Y. Q., Li, J., and Li, M.: Contribution Analysis of the Spatial-Temporal Changes in Streamflow in a Typical Elevation Transitional Watershed of Southwest China over the Past Six Decades, Forests, 10, 495, https://doi.org/10.3390/f10060495, 2019.

Mileham, L., Taylor, R., Thompson, J., Todd M., and Tindimugaya, C.: Impact of rainfall distribution on the parameterisation of a soil-moisture balance model of groundwater recharge in equatorial Africa, J. Hydrol., 359, 46-58, https://doi.org/10.1016/j.jhydrol.2008.06.007, 2008.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L.: Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations, T. ASABE., 50, 885–900, https://doi.org/10.13031/2013.23153, 2007.

Musie, M., Sen, S., and Srivastava, P.: Comparison and evaluation of gridded precipitation datasets for streamflow simulation

in data scarce watersheds of Ethiopia, J. Hydrol., 579, 124168, https://doi.org/ 10.1016/j.jhydrol.2019.124168, 2019.

Peleg, N., Ben-Asher, M., and Morin, E.: Radar subpixel-scale rainfall variability and uncertainty: lessons learned from observations of a dense rain-gauge network, Hydrol. Earth. Syst. Sc., 17, 2195–2208, https://doi.org/10.5194/hess-17-2195-2013, 2013.

Pellicer-Martínez, F., González-Soto, I., and Martínez-Paz, J. M.: Analysis of incorporating groundwater exchanges in hydrological models, Hydrol. Process., 29, 4361–4366, https://doi.org/10.1002/hyp.10586, 2015.

Price, K., Purucker, S. T., Kraemer, S. R., Babendreier, J. E., and Knightes, C. D.: Comparison of radar and gauge precipitation data in watershed models across varying spatial and temporal scales, Hydrol. Process., 28, 3505–3520, https://doi.org/10.1002/hyp.9890, 2013.

Qi, W., Zhang, C., Fu, G., Sweetapple, C., and Zhou, H.: Evaluation of global fine-resolution precipitation products and their uncertainty quantification in ensemble discharge simulations, Hydrol. Earth. Syst. Sc., 20, 903–920, https://doi.org/10.5194/hess-20-903-2016, 2016.

Qiu, J., Yang, Q., Zhang, X., Huang, M., Adam, J. C., and Malek, K.: Implications of water management representations for watershed hydrologic modeling in the Yakima River basin, Hydrol. Earth. Syst. Sc., 23, 35–49, https://doi.org/10.5194/hess-23-35-2019, 2019.

Redding, T. and Devito, K.: Mechanisms and pathways of lateral flow on aspen-forested, Luvisolic soils, Western Boreal Plains, Alberta, Canada, Hydrol. Process., 24, 2995–3010, https://doi.org/10.1002/hyp.7710, 2010.

Remesan, R. and Holman, I. P.: Effect of baseline meteorological data selection on hydrological modelling of climate change scenarios, J. Hydrol., 528, 631-642, https://doi.org/10.1016/j.jhydrol.2015.06.026, 2015.

Roth, V. and Lemann, T.: 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, 2016.

Shivhare, N., Dikshit, P. K. S., and Dwivedi, S. B.: A Comparison of SWAT Model Calibration Techniques for Hydrological Modeling in the Ganga River Watershed, Engineering, https://doi.org/10.1016/j.eng.2018.08.012, 2018.

Solakian, J., Maggioni, V., Lodhi, A., and Godrej, A.: Investigating the use of satellite-based precipitation products for

monitoring water quality in the Occoquan Watershed, Journal of Hydrology: Regional Studies, 26, 100630, https://doi.org/10.1016/j.ejrh.2019.100630, 2019.

Solano-Rivera, V., Geris, J., Granados-Bolaños, S., Brenes-Cambronero, L., Artavia-Rodríguez, G., Sánchez-Murillo, R., and Birkel, C.: Exploring extreme rainfall impacts on flow and turbidity dynamics in a steep, pristine and tropical volcanic catchment, CATENA, 182, 104118, https://doi.org/10.1016/j.catena.2019.104118, 2019.

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., and Hsu, K-L.: A Review of Global Precipitation Data Sets: Data Sources, Estimation, and Intercomparisons, Rev. Geophys., 56, 79-107, https://doi.org/ 10.1002/2017RG000574, 2018.

Tanner, J. L. and Hughes, D. A.: Surface water–groundwater interactions in catchment scale water resources assessments—understanding and hypothesis testing with a hydrological model, Hydrolog. Sci. J., 60, 1880-1895, https://doi.org/10.1080/02626667.2015.1052453, 2015.

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

Tian, Y., Peters-Lidard, C. D., Adler, R. F., Kubota, T., and Ushio, T.: Evaluation of GSMaP Precipitation Estimates over the Contiguous United States, J. Hydrometeorol., 11, 566–574, https://doi.org/10.1175/2009jhm1190.1, 2010.

Tuo, Y., Duan, Z., Disse, M., and Chiogna, G.: Evaluation of precipitation input for SWAT modeling in Alpine catchment A case study in the Adige river basin, Sci. Total. Environ., 573, 66-82, https://doi.org/ 10.1016/j.scitotenv.2016.08.034, 2016.

Tuo, Y., Marcolini, G., Disse, M., and Chiogna, G.: 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, 2018.

Volk, M., Liersch, S., and Schmidt, G.: Towards the implementation of the European water framework directive lessons learned from water quality simulations in an agricultural watershed, Land Use Policy, 26, 580-588, https://doi.org/ 10.1016/j.landusepol.2008.08.005, 2009.

Wang, H., Sun, F., Xia, J., and Liu, W.: Impact of LUCC on streamflow based on the SWAT model over the Wei River basin on the Loess Plateau in China, Hydrol. Earth. Syst. Sc., 21, 1929-1945, https://doi.org/ 10.5194/hess-21-1929-2017, 2017.

Wang, L., Wang, Z., Yu, J., Zhang, Y., and Dang, S.: Hydrological Process Simulation of Inland River Watershed: A Case Study

of the Heihe River Basin with Multiple Hydrological Models, Water, 10, 421, https://doi.org/10.3390/w10040421, 2018.

Weiberlen, F. O. and Benitez J. B.: Assessment of satellite-based precipitation estimates over Paraguay, Acta. Geophysica., 66, 369–379, https://doi.org/10.1007/s11600-018-0146-x, 2018.

Wen, T., Xiong, L., Jiang, C., Hu, J., and Liu, Z.: Effects of Climate Variability and Human Activities on Suspended Sediment Load in the Ganjiang River Basin, China, J. Hydrol. Eng., 24, 05019029, https://doi.org/10.1061/(asce)he.1943-5584.0001859, 2019.

Wu, Y., Zhang, Z., Huang, Y., Jin, Q., Chen, X., and Chang, J.: Evaluation of the GPM IMERG v5 and TRMM 3B42 v7 Precipitation Products in the Yangtze River Basin, China, Water, 11, 1459, https://doi.org/10.3390/w11071459, 2019.

Wu, J., Chen, X., Yu, Z., Yao, H., Li, W., and Zhang, D.: Assessing the impact of human regulations on hydrological drought development and recovery based on a "simulated-observed" comparison of the SWAT model, J. Hydrol., 123990, https://doi.org/10.1016/j.jhydrol.2019.123990, 2019.

Xie, P., Chen, M., Yang, S., Yatagai, A., Hayasaka, T., Fukushima, Y., and Liu, C.: A Gauge-Based Analysis of Daily Precipitation over East Asia, J. Hydrometeorol., 8, 607–626, https://doi.org/10.1175/jhm583.1, 2007.

Yan, R., Gao, J., and Huang, J.: WALRUS-paddy model for simulating the hydrological processes of lowland polders with paddy fields and pumping stations, Agr. Water. Manage., 169, 148–161, https://doi.org/10.1016/j.agwat.2016.02.018, 2016.

Yilmaz, A. G., Imteaz, M. A., and Ogwuda, O.: Accuracy of HEC-HMS and LBRM Models in Simulating Snow Runoffs in Upper Euphrates Basin, J. Hydrol. Eng., 17, 342–347, https://doi.org/10.1061/(asce)he.1943-5584.0000442, 2012.

Zambrano-Bigiarini, M., Nauditt, A., Birkel, C., Verbist, K., and Ribbe, L.: Temporal and spatial evaluation of satellite-based rainfall estimate across the complex topographical and climatic gradients of Chile, Hydrol. Earth. Syst. Sc., 21, 1295-1320, https://doi.org/10.5194/hess-21-1295-2017, 2017.

Zhang, D. J., Lin, Q. Y., Chen, X. W., and Chai, T.: Improved Curve Number Estimation in SWAT by Reflecting the Effect of Rainfall Intensity on Runoff Generation, Water, 11, 163, https://doi.org/ 10.3390/w11010163, 2019.

Zhang, H. L., Meng, C. C., Wang, Y. Q., Wang, Y. J., and Li, M.: Comprehensive evaluation of the effects of climate change and land use and land cover change variables on runoff and sediment discharge. Sci. Total. Environ., 702, 134401,

https://doi.org/10.1016/j.scitotenv.2019.134401, 2020.

Zhou, Z., Ouyang, Y., Li, Y., Qiu, Z., and Moran, M.: Estimating impact of rainfall change on hydrological processes in Jianfengling rainforest watershed, China using BASINS-HSPF-CAT modeling system, Ecol. Esng., 105, 87–94, https://doi.org/10.1016/j.ecoleng.2017.04.051, 2017.

Zhu, H., Li, Y., Liu, Z., Shi, X., Fu, B., and Xing, Z.: Using SWAT to simulate streamflow in Huifa River basin with ground and Fengyun precipitation data, J. Hydroinform., 17, 834–844, https://doi.org/10.2166/hydro.2015.104, 2015.

760

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
a__SOL_K().sol	Saturated hydraulic conductivity [mm/h]	−10/10	Soil layer specific
r__SOL_BD().sol	Moist bulk density [g/cm^3]	−0.5/0.5	Soil layer specific
a__CN2.mgt	SCS streamflow curve number	−20/20	HRU specific
v__ESCO.hru	Soil evaporation compensation factor	0/1	0.95
a__HRU_SLP.hru	Average slope steepness [m/m]	−0.2/0.4	HRU specific
a__SLSUBBSN.hru	Average slope length [m]	−9/130	HRU specific
v__CH_K2.rte	Effective hydraulic conductivity [mm/h]	0/400	0
v__CH_N2.rte	Mannings n value for main channel	0/0.3	0.014
a__GWQMN.gw	Threshold depth of water in the shallow aquifer required for return flow to occur [mm]	−500/500	1000
a__REVAPMN.gw	Threshold depth of water in the shallow aquifer for "revap" to occur [mm]	−500/500	750
v__ALPHA_BF.gw	Baseflow alpha factor [days^{-1}]	0/1	0.048
v__GW_REVAP.gw	Groundwater "revap" coefficient	0.02/0.2	0.02
v__GW_DELAY.gw	Groundwater delay [days]	0/300	31
v__ALPHA_BNK.rte	Baseflow alpha factor for bank storage	0/1	0

Table 2. General performance ratings statistics recommended by Moriasi et al. (2007).

Performance Rating	*RSR*	*NSE*	*PBIAS* (%)
Very good	$0.00 < RSR \leq 0.50$	$0.75 < NSE \leq 1.00$	$PBIAS \leq \pm 10$
Good	$0.50 < RSR \leq 0.60$	$0.65 < NSE \leq 0.75$	$\pm 10 < PBIAS \leq \pm 15$
Satisfactory	$0.60 < RSR \leq 0.70$	$0.50 < NSE \leq 0.65$	$\pm 15 < PBIAS \leq \pm 25$
Unsatisfactory	$RSR > 0.70$	$NSE < 0.50$	$PBIAS > \pm 25$

765

Table ~~2~~3: Optimal parameters calibrated for all three models.

Parameters	Initial range	Gauge Monthly	Daily	CHIRPS Monthly	Daily	CPC Monthly	Daily
v__PLAPS.sub	−1000/1000	0.012/0.067	0.061/0.183	0.079/0.135	0.068/0.205	0.017/0.078	-0.014/0.095
a__SOL_K().sol	−10/10	1.988/10	-0.706/10	-0.471/7.681	-0.396/10	5.264/10	-2.106/10
r__SOL_BD().sol	−0.5/0.5	0.036/0.5	-0.111/0.5	-0.130/0.5	-0.126/ 0.5	0.262/0.5	-0.04/0.5
a__CN2.mgt	−20/20	-16.141/17.309	-1.371/20	12.825/20	-1.491/20	-4.092/20	-1.992/20
v__ESCO.hru	0/1	0.879/1	0.405/1	0.775/1	0.355/1	0.914/1	0.462/1
a__HRU_SLP.hru	−0.2/0.4	0.261/0.4	0.013/0.4	0.157/0.280	-0.2/0.116	0.181/0.4	0.049/0.4
a__SLSUBBSN.hru	−9/130	-9/40.518	-9/75.760	68.303/108.959	23.139/94.386	-9/19.244	-9/74.023
v__CH_K2.rte	0/400	0/101.266	0/252.317	0/113.457	56.486/285.514	16.056/326.448	0/220.314
v__CH_N2.rte	0/0.3	0.019/0.188	0/0.173	0.091/0.183	0/0.187	0.138/0.233	0.091/0.272
a__GWQMN.gw	−500/500	-500/-241.312	-76.285/500	-104.708/178.914	-500/21.287	-500/-235.592	-500/118.785
a__REVAPMN.gw	−500/500	-500/-98.63	-500/125.78	-429.291/69.66	-232.285/303.28	-189.739/295.43	-432.284/189.28
v__ALPHA_BF.gw	0/1	0.401/0.963	0.299/0.896	0.055/0.677	0.183/0.728	0.216/0.901	0.415/1
v__GW_REVAP.gw	0.02/0.2	0.038/0.141	0.02/0.123	0.127/0.188	0.02/0.124	0.037/0.141	0.077/0.192
v__GW_DELAY.gw	0/300	37.681/215.552	118.714/300	96.336/188.942	81.664/245.036	0/74.581	0/182.936
v__ALPHA_BNK.rte	0/1	0.492/0.863	0.444/1	0.201/0.696	0.467/1	0.564/1	0.307/0.92

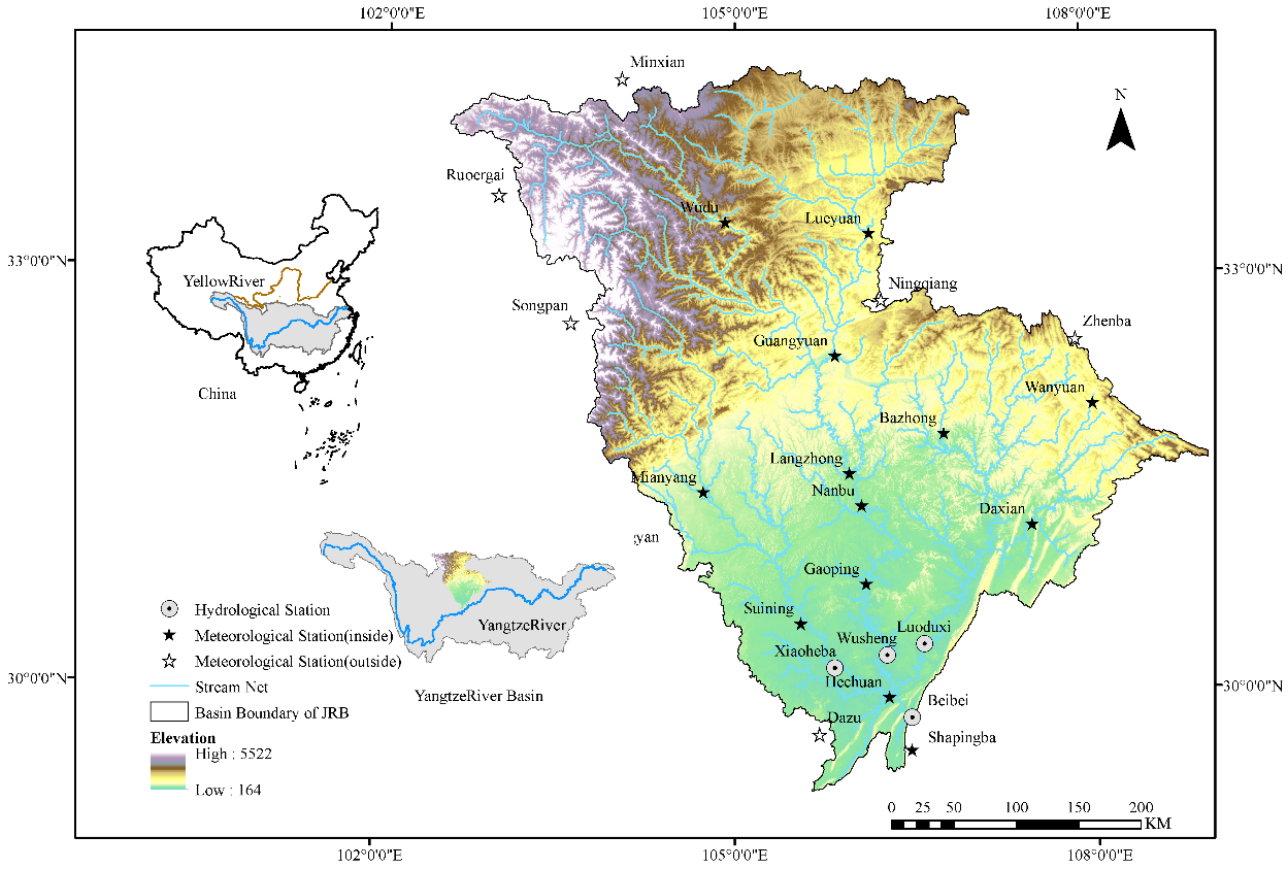
Table ~~3~~4: *POD* and *FAR* values for different rainfall intensities.

	> 0.1mm		≥ 50mm	
	POD	*FAR*	*POD*	*FAR*
CHIRPS	27.29%	54.12%	18.12%	65.56%
CPC	83.53%	46.76%	9.42%	44.71%

Table 45: Summarization of water balance components of the three models for the whole JRW.

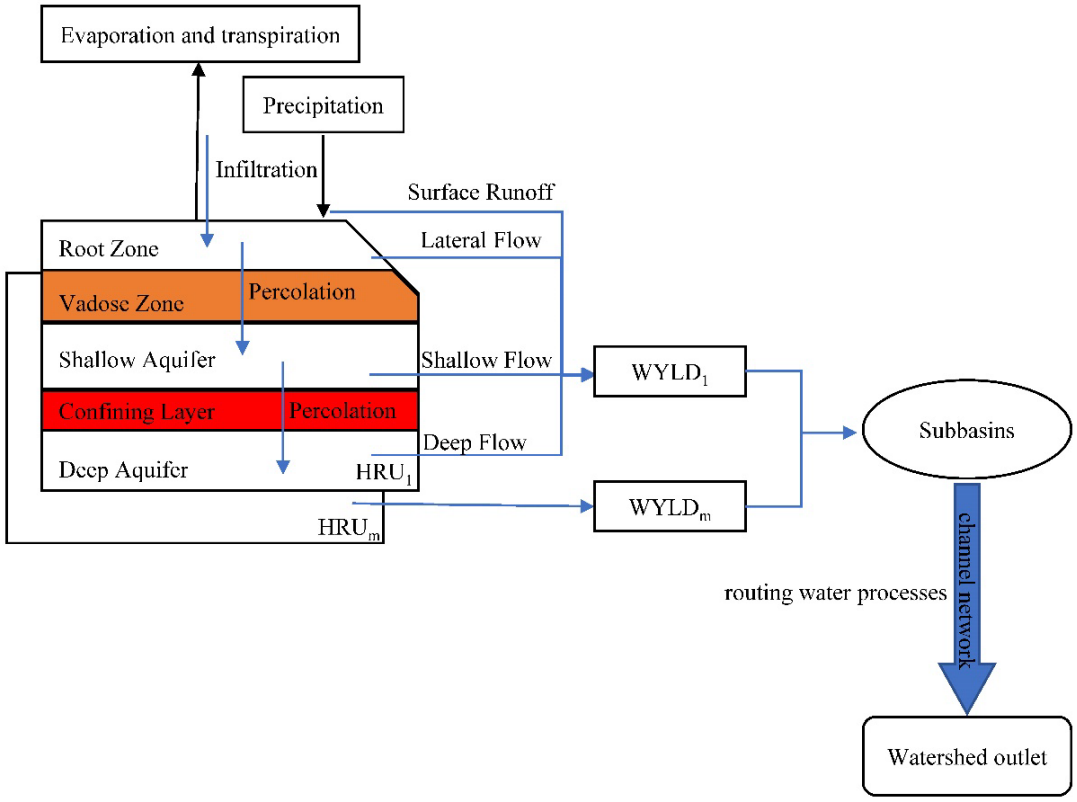
Datasets	Statistics	SURQ	LATQ	GW_Q	GW_Q_D	ET	Summation
Gauge	Average	3832.20	2471.80	274.05	54.48	8153.94	14786.47
	Percentage	25.92%	16.72%	1.85%	0.37%	55.14%	
CHIRPS	Average	5188.42	629.06	809.19	120.64	8161.22	14908.52
	Percentage	34.80%	4.22%	5.43%	0.81%	54.74%	
CPC	Average	910.37	4707.44	547.85	28.32	7806.84	14000.82
	Percentage	6.50%	33.62%	3.91%	0.20%	55.76%	

Time scale	Datasets	Statistics	SURQ	LATQ	GW_Q	GW_Q_D	ET	Summation
Monthly	Gauge	Average amount/mm	4500.00	2977.22	299.07	60.61	9076.60	16913.50
		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	16978.78
		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	15972.40
		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	17003.48
		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	18699.11
		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	15978.90
		Percentage/%	15.60%	14.41%	10.70%	0.55%	58.73%	



780

Figure 1. Sketch map of the Jialing River Basin with meteorological stations.



785 **Figure 2.** Schematic representation of the modified SWAT model structure.

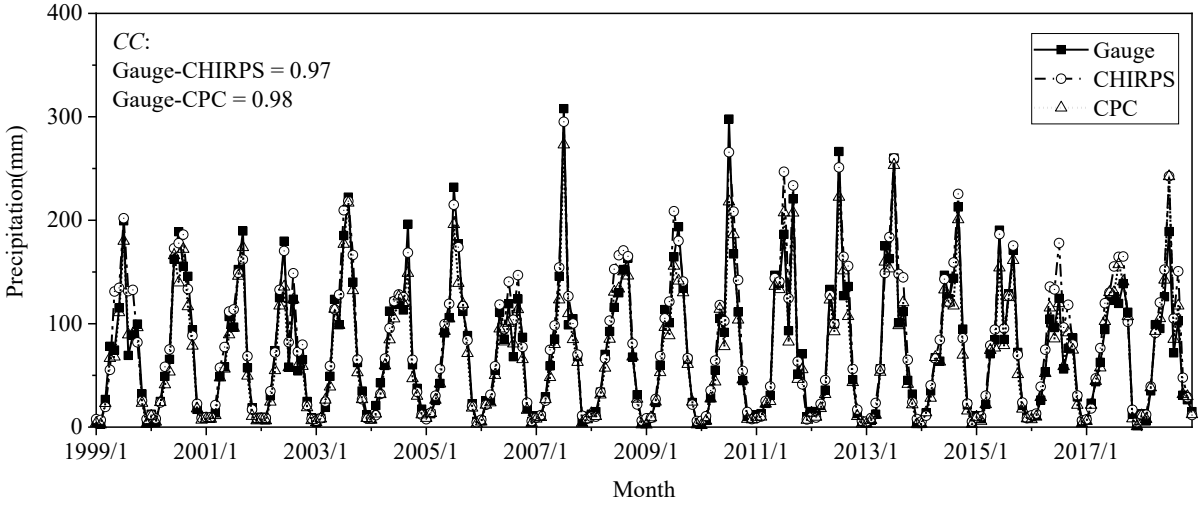
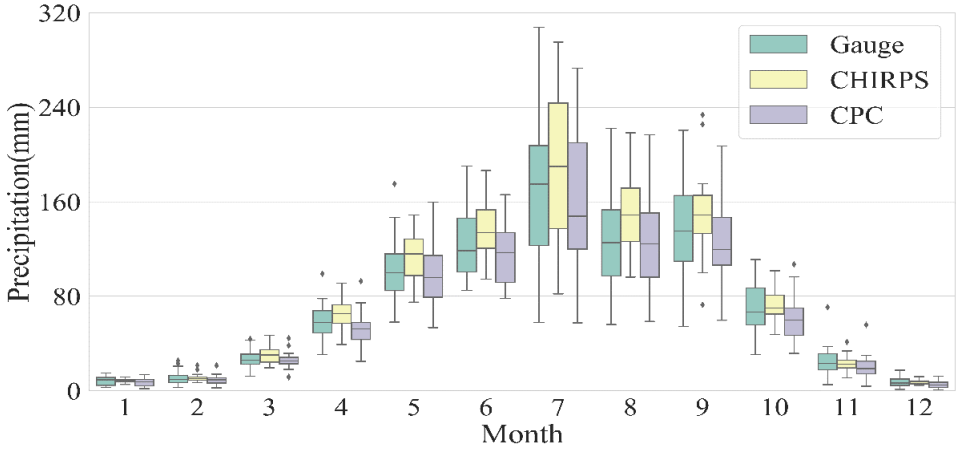


Figure 3. Time series of three different precipitation records at monthly scale in JRW (the *CC* values of 0.97 and 0.98 indicating extremely significant positive correlation (P<0.01, where P stands for the probability of being rejected when there was a significant difference)).



795

Figure 4. Box Diagrams of three different precipitation records at monthly scale in JRW.

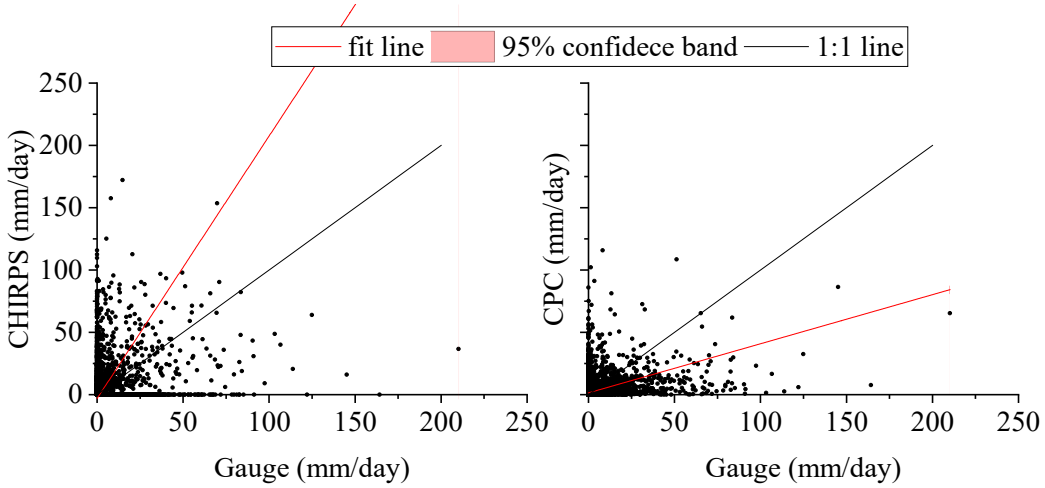
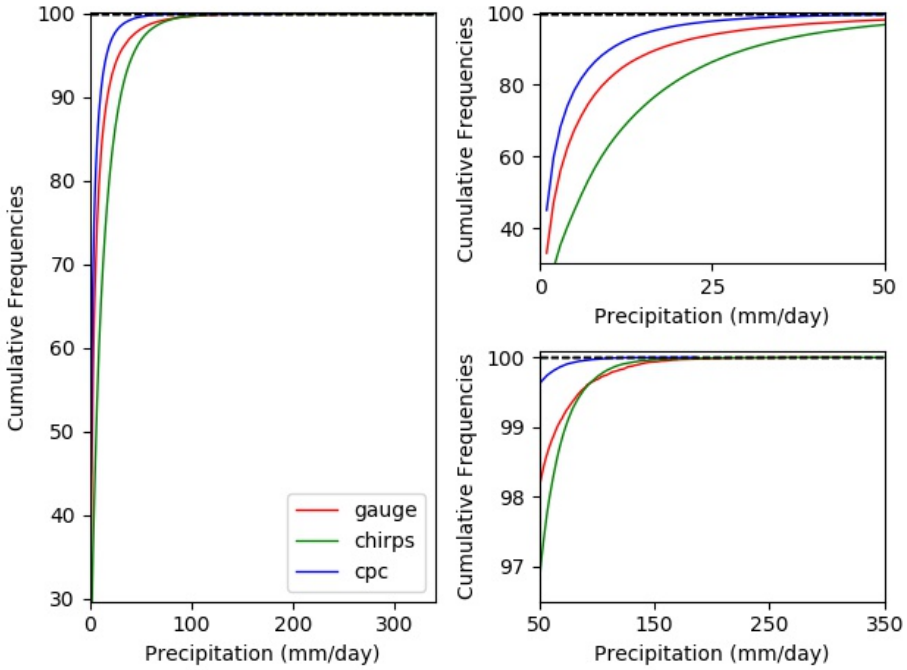


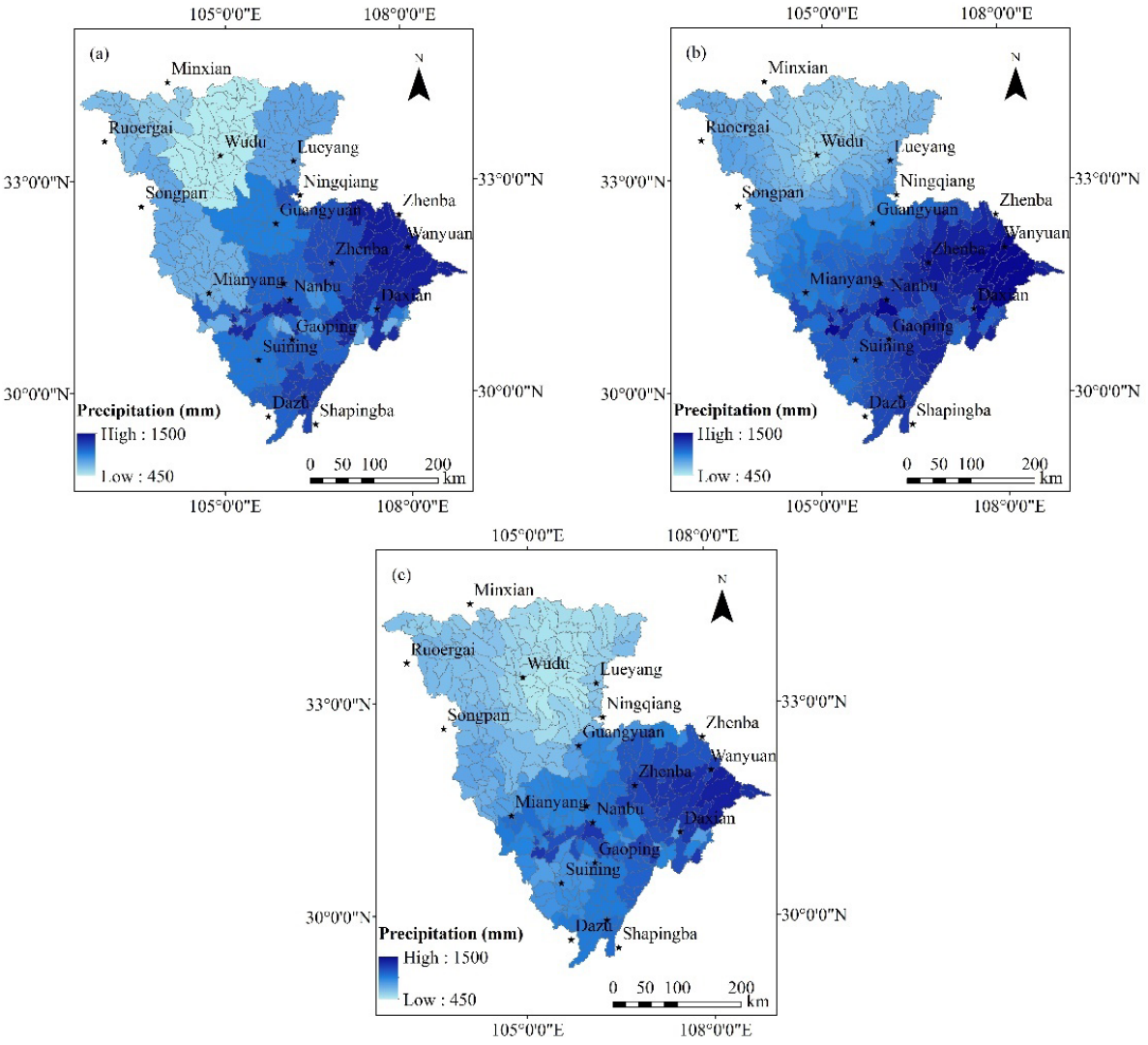
Figure 5. Scatter plot of the OPPs records comparing with Gauge records at daily scale: (a) comparison of CHIRPS and Gauge; (b) comparisons of CPC and Gauge.



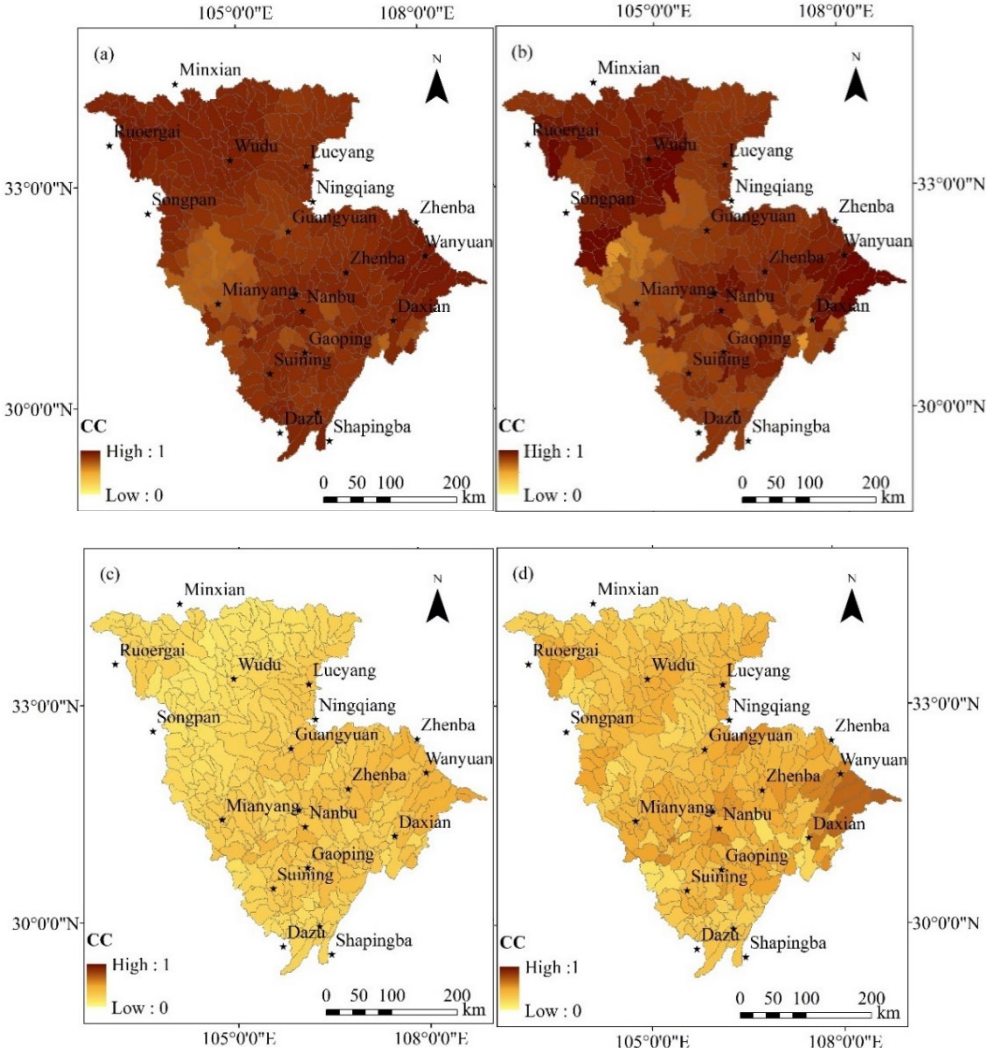
805

Figure 6. Cumulative Frequencies of daily precipitation intensity for the three precipitation products (Gauge, CHIRPS, CPC) in JRW: (a) distribution of all precipitation values; (b) distribution of precipitation values that are<100 mm; (c) distribution of precipitation values that are ≥100 mm.

810

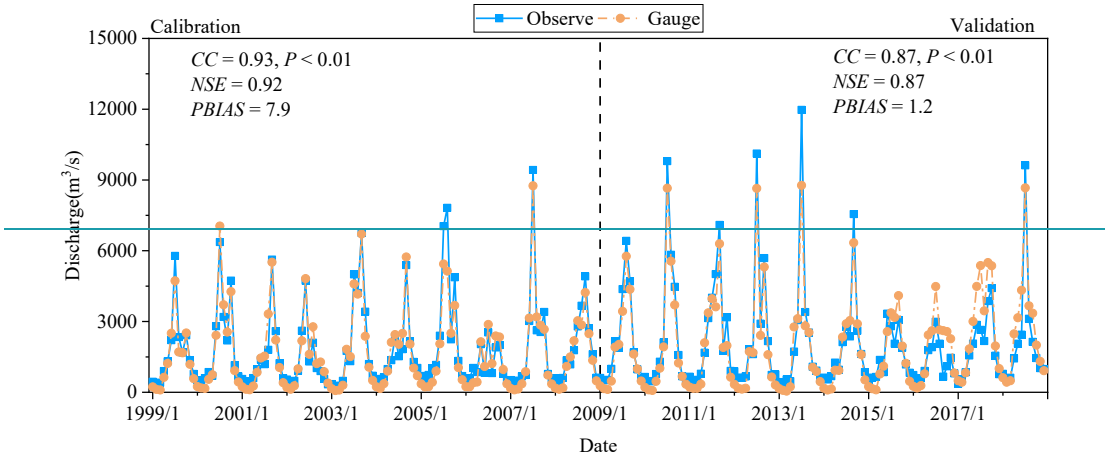


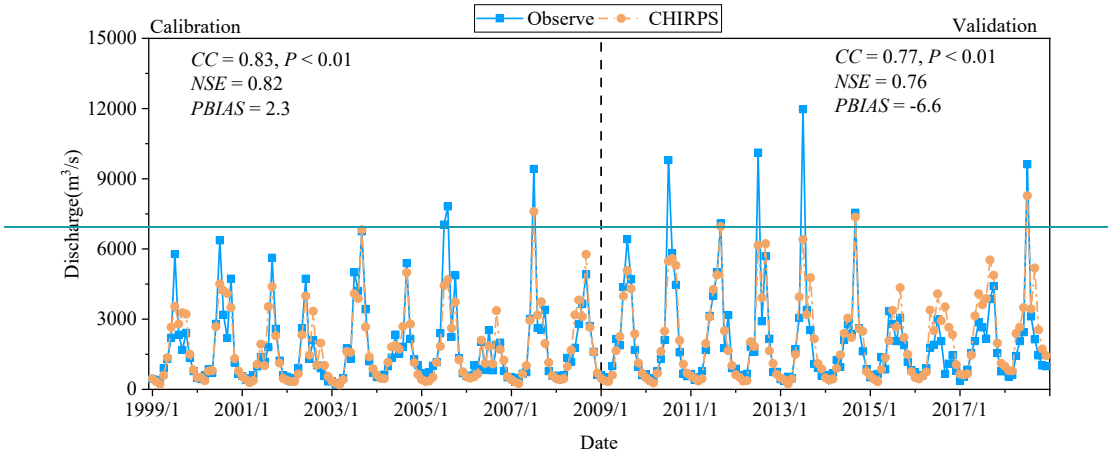
815 **Figure 7.** Spatial variation of annual precipitation at sub-basin scale for (a) Gauge (b) CHIRPS and (c) CPC.



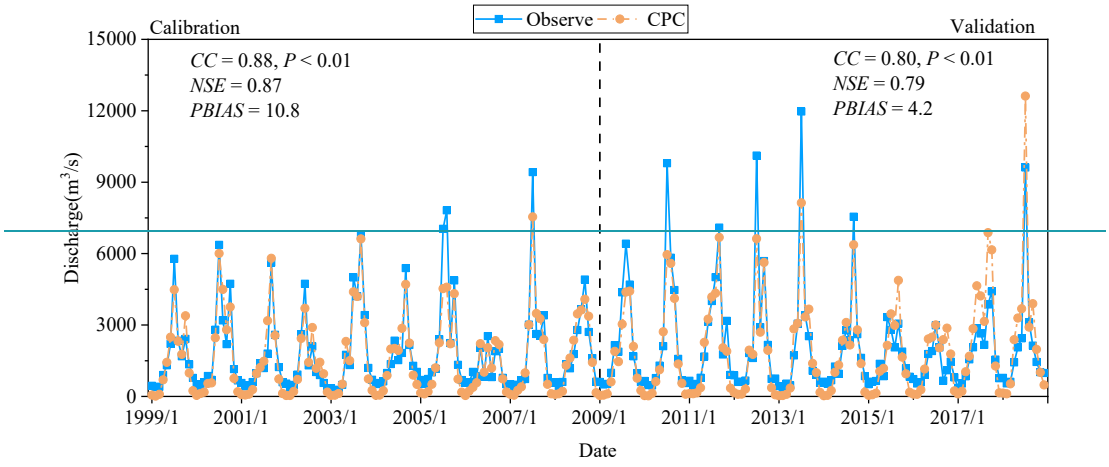
820

Figure 8. Spatial variation of *CC* values of the precipitation between (a) Gauge and CHIRPS, (b) Gauge and CPC at monthly scale and (c) Gauge and CHIRPS, (d) Gauge and CPC at daily scale.


Observe
Gauge
Calibration
Validation
CC = 0.93, P < 0.01
NSE = 0.92
PBIAS = 7.9
CC = 0.87, P < 0.01
NSE = 0.87
PBIAS = 1.2
Discharge(m3/s)
Date


Observe
CHIRPS
Calibration
Validation
CC = 0.83, P < 0.01
NSE = 0.82
PBIAS = 2.3
CC = 0.77, P < 0.01
NSE = 0.76
PBIAS = -6.6
Discharge(m3/s)
Date

825


Observe
CPC
Calibration
Validation
CC = 0.88, P < 0.01
NSE = 0.87
PBIAS = 10.8
CC = 0.80, P < 0.01
NSE = 0.79
PBIAS = 4.2
Discharge(m3/s)
15000
12000
9000
6000
3000
0
1999/1
2001/1
2003/1
2005/1
2007/1
2009/1
2011/1
2013/1
2015/1
2017/1
Date

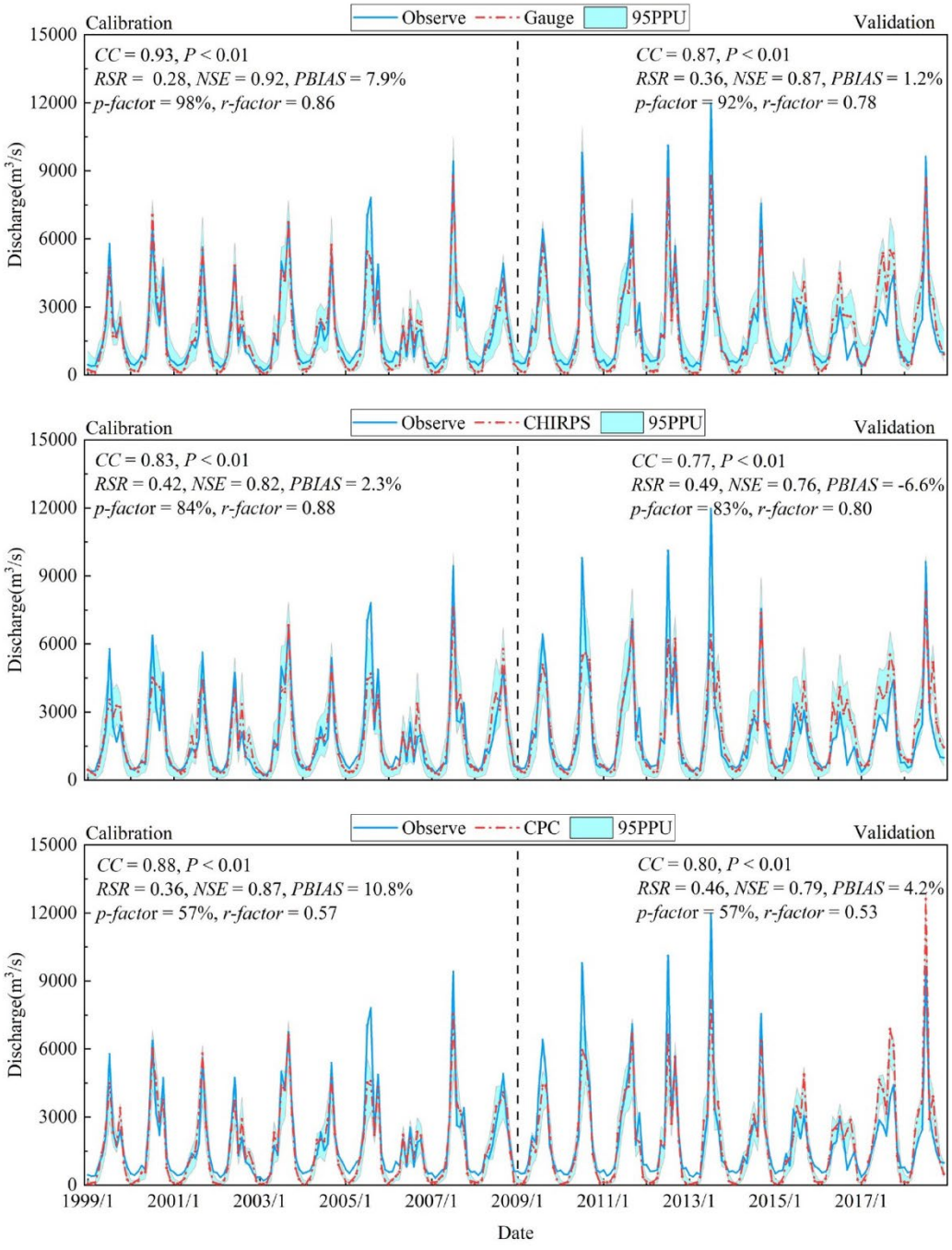


Figure 9. Observed and simulated discharges at the outlet of JRW at monthly scale using precipitation inputs of Gauge, CHIRPS and CPC, respectively.

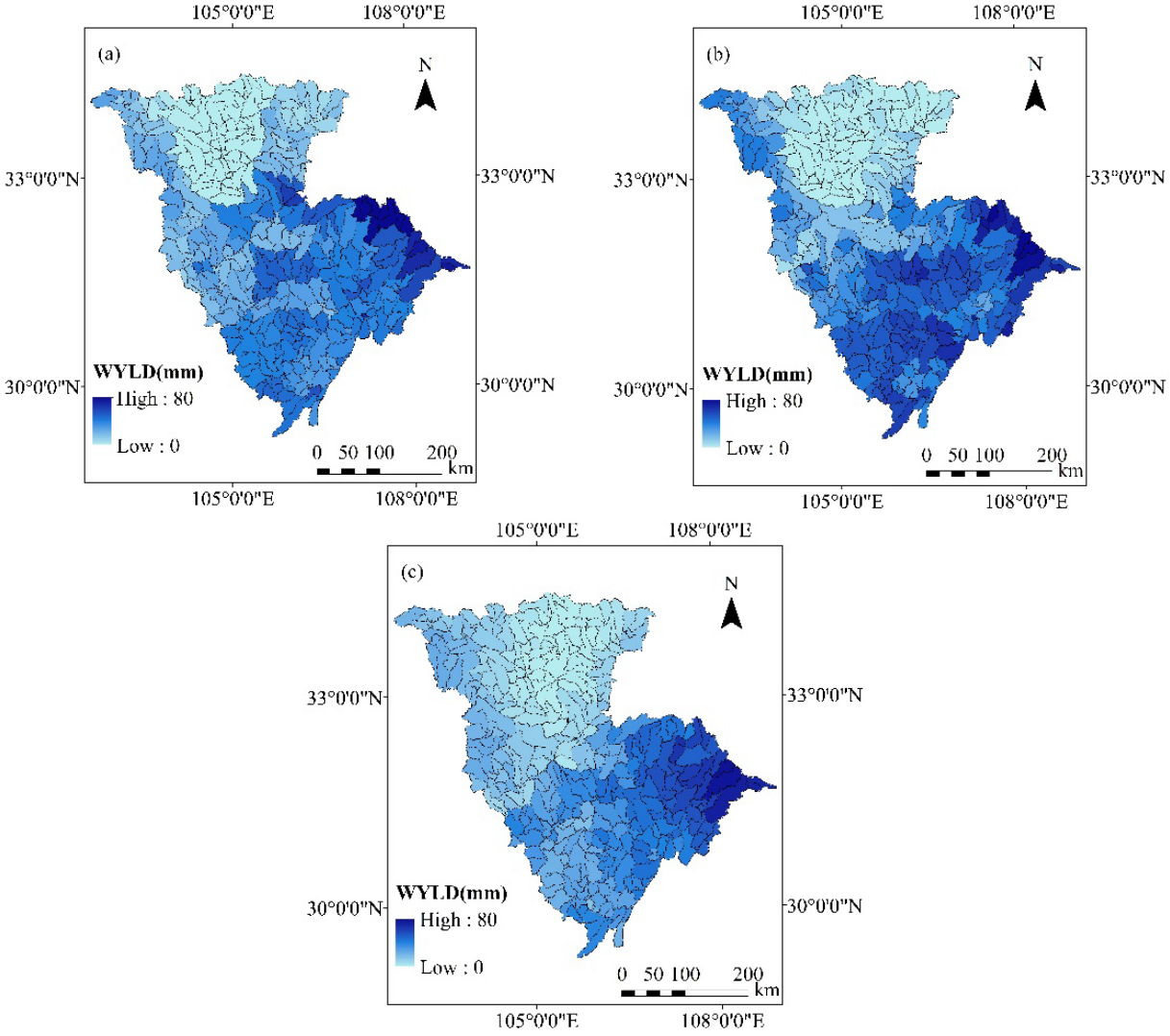
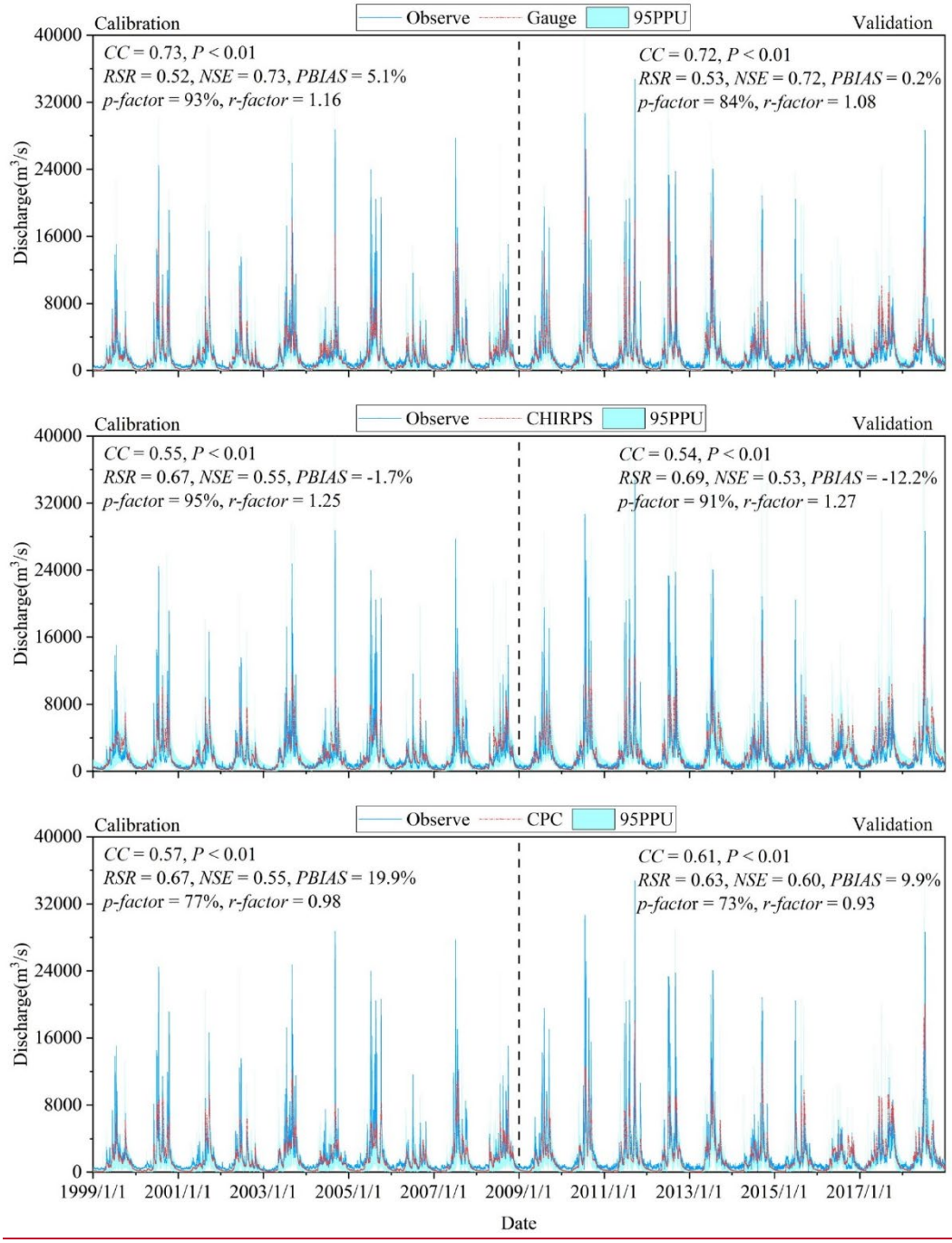
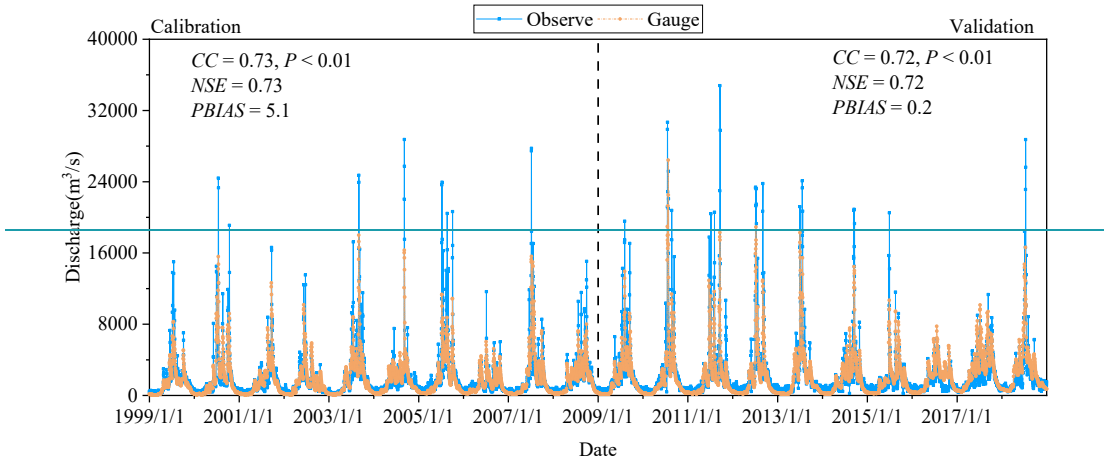
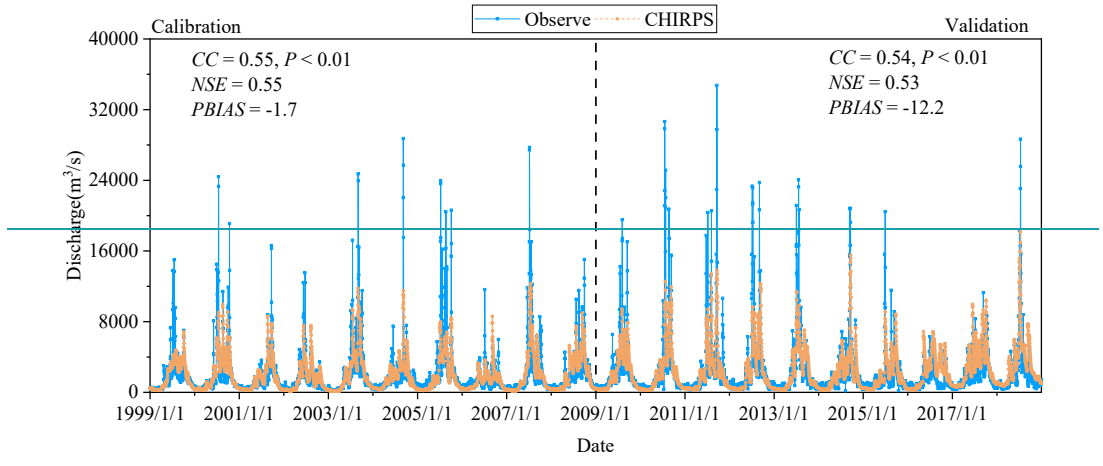
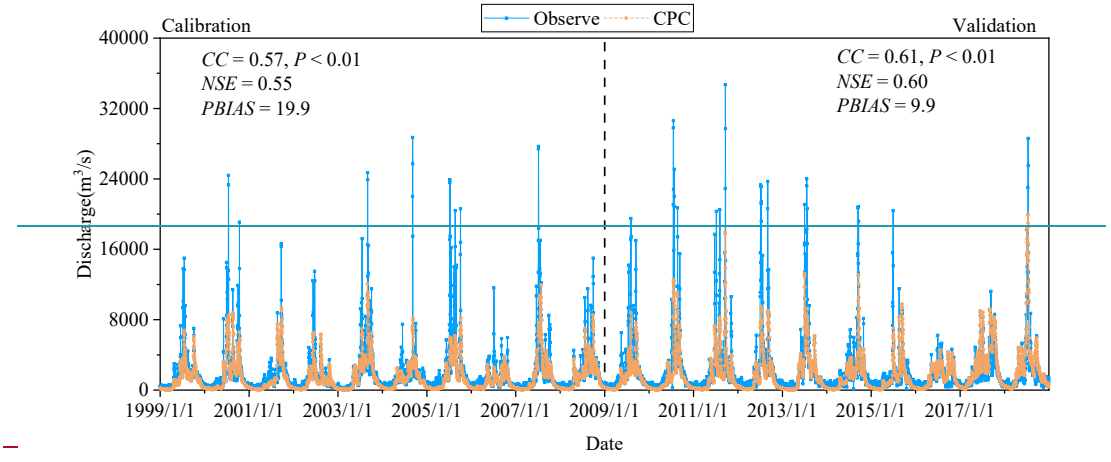


Figure 10. Spatial variation of water yield at monthly scale for all sub-basins calculated with precipitation inputs of (a) Gauge (b) CHIRPS and (c)CPC.


Calibration
Validation
Observe
Gauge
95PPU
CC = 0.73, P < 0.01
RSR = 0.52, NSE = 0.73, PBIAS = 5.1%
p-factor = 93%, r-factor = 1.16
CC = 0.72, P < 0.01
RSR = 0.53, NSE = 0.72, PBIAS = 0.2%
p-factor = 84%, r-factor = 1.08
Observe
CHIRPS
95PPU
CC = 0.55, P < 0.01
RSR = 0.67, NSE = 0.55, PBIAS = -1.7%
p-factor = 95%, r-factor = 1.25
CC = 0.54, P < 0.01
RSR = 0.69, NSE = 0.53, PBIAS = -12.2%
p-factor = 91%, r-factor = 1.27
Observe
CPC
95PPU
CC = 0.57, P < 0.01
RSR = 0.67, NSE = 0.55, PBIAS = 19.9%
p-factor = 77%, r-factor = 0.98
CC = 0.61, P < 0.01
RSR = 0.63, NSE = 0.60, PBIAS = 9.9%
p-factor = 73%, r-factor = 0.93
Discharge(m3/s)
0
8000
16000
24000
32000
40000
1999/1/1
2001/1/1
2003/1/1
2005/1/1
2007/1/1
2009/1/1
2011/1/1
2013/1/1
2015/1/1
2017/1/1
Date

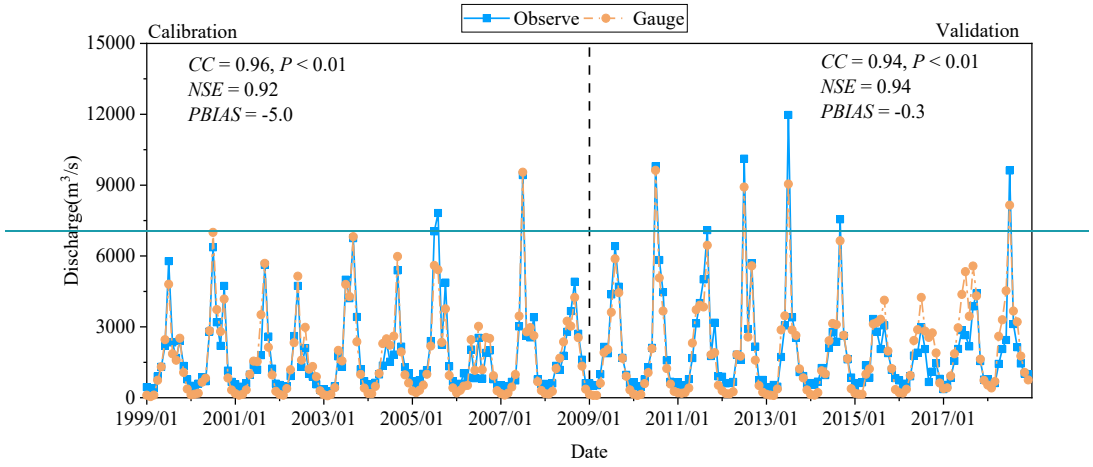

Observe
Gauge
Calibration
Validation
CC = 0.73, P < 0.01
NSE = 0.73
PBIAS = 5.1
CC = 0.72, P < 0.01
NSE = 0.72
PBIAS = 0.2
Discharge(m3/s)
40000
32000
24000
16000
8000
0
1999/1/1
2001/1/1
2003/1/1
2005/1/1
2007/1/1
2009/1/1
2011/1/1
2013/1/1
2015/1/1
2017/1/1
Date

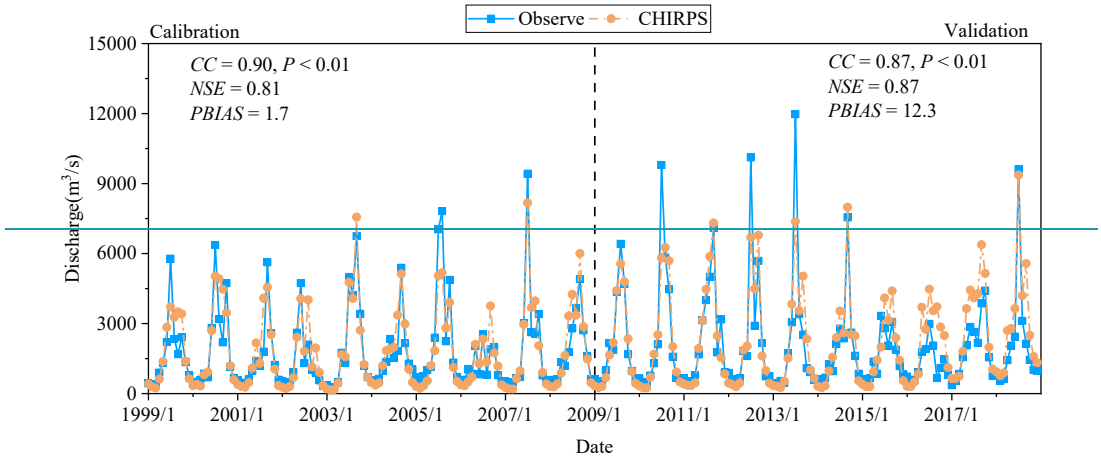

Observe
CHIRPS
Calibration
Validation
CC = 0.55, P < 0.01
NSE = 0.55
PBIAS = -1.7
CC = 0.54, P < 0.01
NSE = 0.53
PBIAS = -12.2
Discharge(m3/s)
40000
32000
24000
16000
8000
0
1999/1/1
2001/1/1
2003/1/1
2005/1/1
2007/1/1
2009/1/1
2011/1/1
2013/1/1
2015/1/1
2017/1/1
Date


Observe
CPC
Calibration
Validation
CC = 0.57, P < 0.01
NSE = 0.55
PBIAS = 19.9
CC = 0.61, P < 0.01
NSE = 0.60
PBIAS = 9.9
Discharge(m3/s)
40000
32000
24000
16000
8000
0
1999/1/1
2001/1/1
2003/1/1
2005/1/1
2007/1/1
2009/1/1
2011/1/1
2013/1/1
2015/1/1
2017/1/1
Date

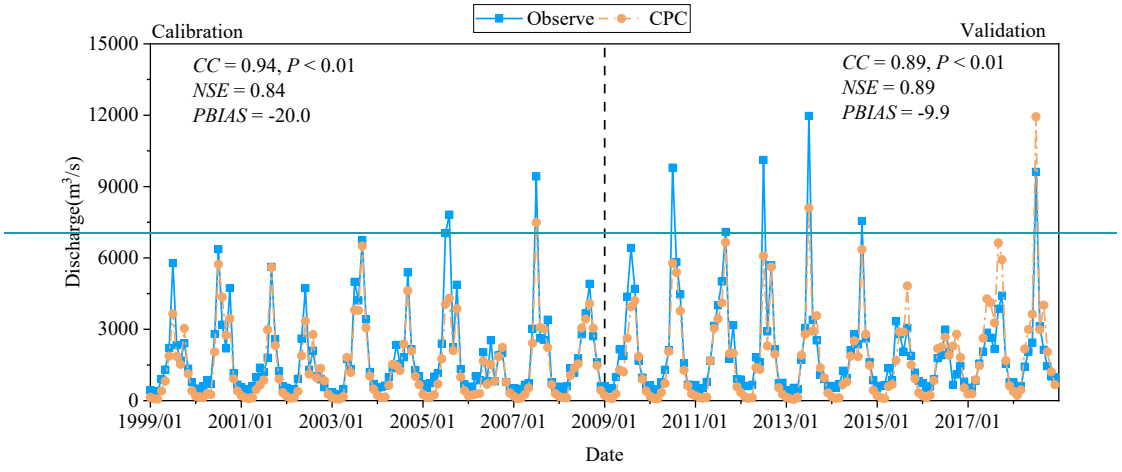
840

Figure 11. Observed and simulated discharges at the outlet of JRW at daily scale using precipitation inputs of Gauge, CHIRPS and CPC, respectively.

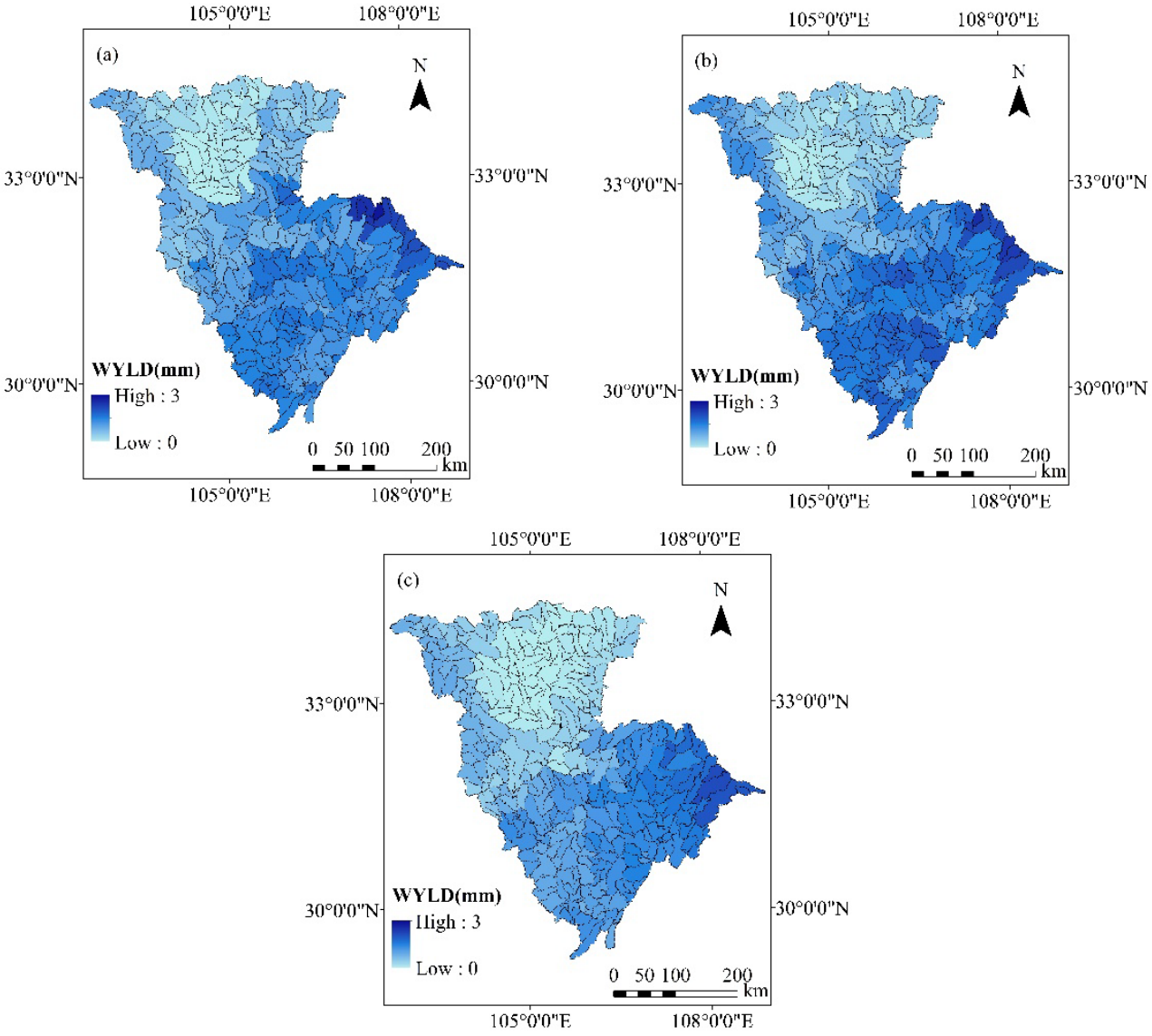

Observe
Gauge
Calibration
Validation
CC = 0.96, P < 0.01
NSE = 0.92
PBIAS = -5.0
CC = 0.94, P < 0.01
NSE = 0.94
PBIAS = -0.3
Discharge(m3/s)
Date


Observe
CHIRPS
Calibration
Validation
CC = 0.90, P < 0.01
NSE = 0.81
PBIAS = 1.7
CC = 0.87, P < 0.01
NSE = 0.87
PBIAS = 12.3
Discharge(m3/s)
Date

845


Observe
CPC
Calibration
Validation
CC = 0.94, P < 0.01
NSE = 0.84
PBIAS = -20.0
CC = 0.89, P < 0.01
NSE = 0.89
PBIAS = -9.9
Discharge(m3/s)
Date

~~**Figure 12.** Observed and simulated discharges at the outlet of JRW calculating at daily scale and presenting at monthly scale using precipitation inputs of Gauge, CHIRPS and CPC, respectively.~~



850

Figure ~~13~~12. Spatial variation of water yield at daily scale for all sub-basins calculated with precipitation inputs of (a) Gauge (b) CHIRPS and (c) CPC.

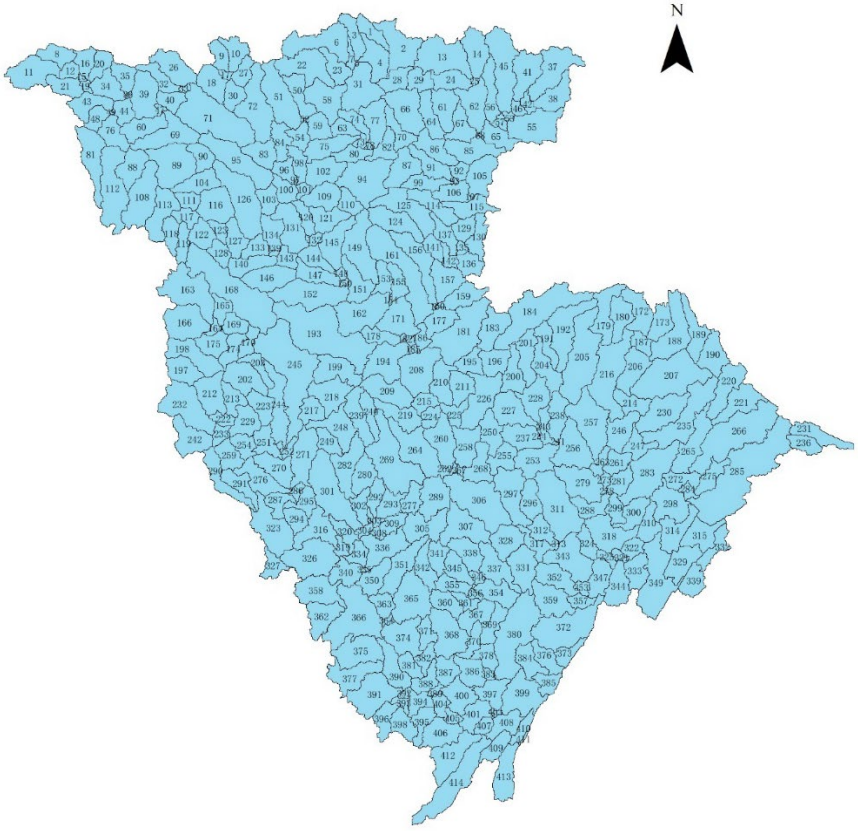


Figure ~~14~~13. Spatial distribution of sub-basins in SWAT, named by Numbers.

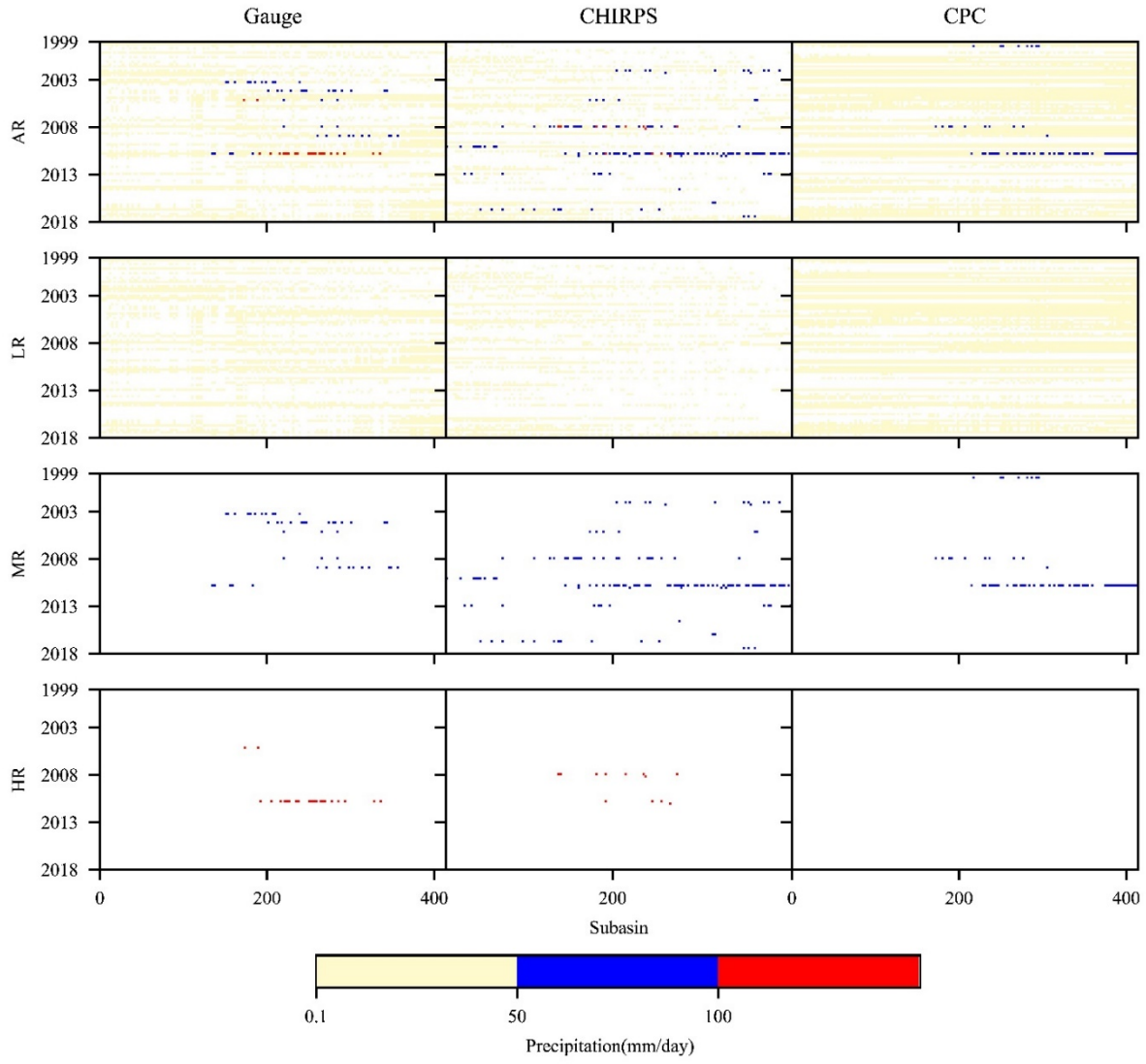


Figure ~~15~~14. Full records of flood events occurred throughout the study period and all sub-basins detected by three precipitation products, where the AR, LR, MR, HR stand for precipitation of all rainfall intensities, intensity between 0.1 and 50 mm/day, intensity between 50 and 100 mm/day, intensity more than 100 mm/day, respectively.

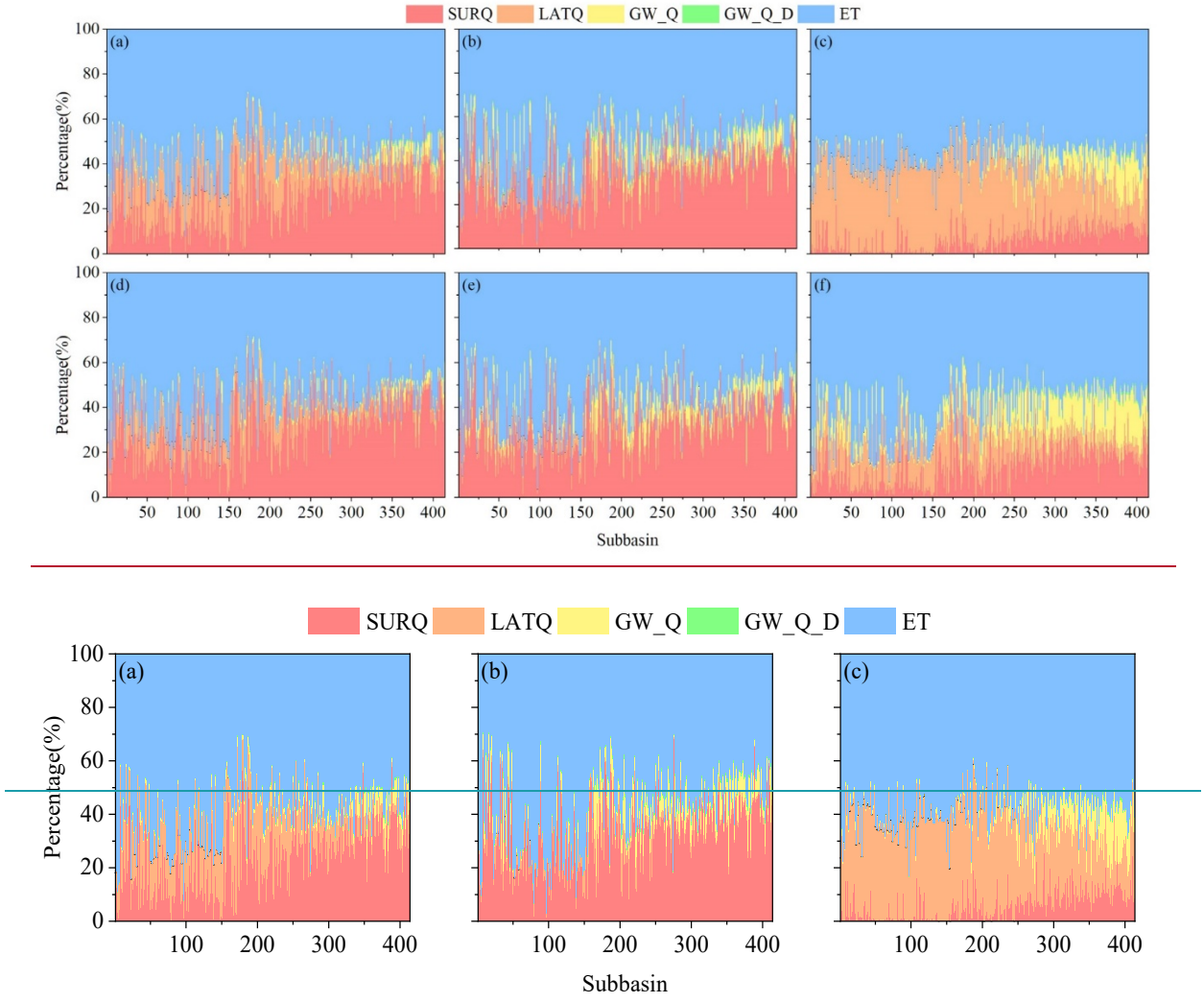


Figure ~~16~~15. Water balance components for all sub-basins derived from SWAT models using precipitation inputs of (a) Gauge (b) CHIRPS and (c) CPC at monthly scale and (d) Gauge (e) CHIRPS and (f) CPC at daily scale (where SURQ represents surface runoff Q_{surf}; LATQ represents lateral flow Q_{lat}; 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*. ~~Water balance components for all sub-basins derived from SWAT models using precipitation inputs of (a) Gauge (b) CHIRPS and (CPC) (where SURQ represents surface runoff Q_{surf}; LATQ represents lateral flow Q_{lat}; GW_Q is the base flow from the shallow aquifer; GW_Q_D is the base flow from the deep aquifer, and the sum of GW_Q and GW_Q_D~~

~~equals to Q_{gw}; ET represents evapotranspiration E_a.~~