Dear reviewer,

Thanks a lot for your great efforts to read through this paper and give very valuable comments. Here we have addressed the comments from you and the detailed description is attached in this document.

Best regards,

Qian Zhu, Xiaodong Qin, Dongyang Zhou, Tiantian Yang, Xinyi Song

Point 1: Model calibration considering parts of discharge time series is not a new idea

Response 1: Thank you very much for your comment. Input data, model and calibration strategy can affect the accuracy of flood events simulation and prediction. To our best knowledge, the sensitivity of models with different structures, such as lumped hydrological model, semi-distributed/distributed hydrological model, and data-driven model, to the spatio-temporal resolutions of precipitation has not been investigated. In this study, we investigated the impacts of temporal and spatial resolutions of precipitation on flood events simulation over a large-scale catchment, and we accomplished the study with the applicability of HBV, SWAT, DHSVM and LSTM forced by high spatio-temporal resolution gauge-based and satellite-based precipitation products.

Point 2: Lines 20: It is not clear what you mean by "flood event." Also, I am not comfortable with the term "to match continuous streamflow." May be you can write "to match the entire streamflow time series."

Response 2: Thank you very much for your comments. We have modified the relevant description of flood event in Lines 20: “Two calibration strategies are carried out, one of which targets at matching the flood, and the other one is the conventional strategy to match the entire streamflow time series.”

Point 3: How did you select the flood events

Response 3: Thank you for your question. ***In 2.2 Data description***, we have explained how we choose flood events: “Fig. 2 shows the time series of the hourly streamflow and corresponding gauge-based precipitation between 2015 and 2017, where eleven historical flood events are selected with flood peak exceeding the threshold of 8,600 m^3/s in this study.”

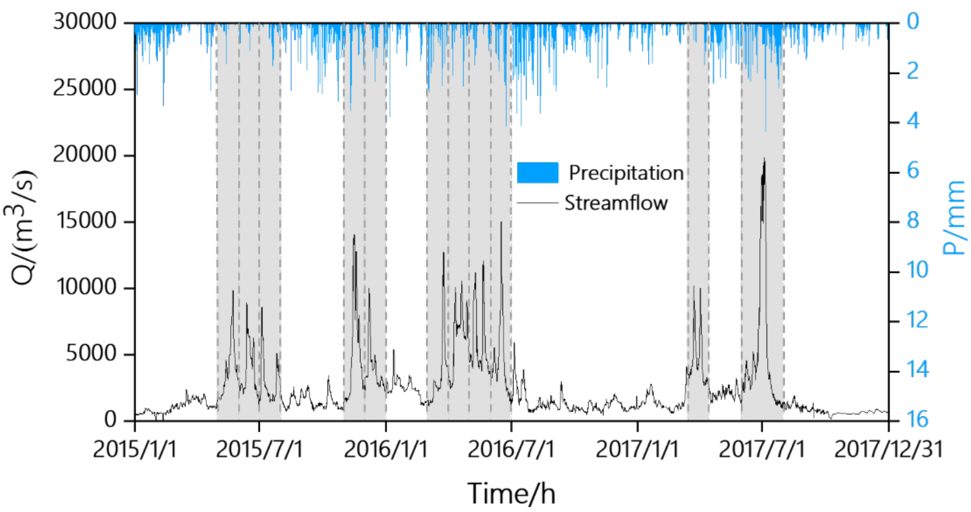

Fig. 2. Time series of observed hourly streamflow in Xiangtan station and basin-average precipitation from CMA, with eleven selected flood events covered by shaded areas.

Point 4: Line 295: Mean NSE may not be a reliable indicator. You should consider median, 75th and 25th percentile NSE.I see 75th NSE falling in case of CMA. The authors need to discuss it.

Response 4:

Thank you for your suggestion. Since our target is to explore the impacts of different calibration strategies on flood events simulation, mean NSE is used in our study for it is more suitable for flood events as many previous studies proved (Yu et al. 2018, Kao et al. 2020). Meanwhile, the mean and median NSE have the same pattern in our study, the mean and median NSE of calibration strategy II are better than that of calibration strategy I as a whole, which is illustrated in Fig. 6, for HBV, the mean NSE values of CMA, IMERG-E, IMERG-L, IMERG-F increase from 0.78, 0.54, 0.54, 0.72 with calibration strategy I to 0.79, 0.62, 0.67, 0.75 with calibration strategy II, the median NSE increase from 0.78, 0.67, 0.79, 0.68 with calibration strategy I to 0.80, 0.78, 0.83, 0.79 with calibration strategy II.

Point 5: NSEs in Figure 6: I don't see any consistent pattern. The results are not discussed properly.

Response 5:

Thank you for your question, and sorry for the misunderstanding. In order to discuss the results more thoroughly, results and discussion are presented in two separate sessions. The mean and median NSE of calibration strategy II are better than that of calibration strategy I as a whole, which is illustrated in Fig. 6. For HBV, the mean NSE values of CMA, IMERG-E, IMERG-L, IMERG-F increase from 0.78, 0.54, 0.54, 0.72 with calibration strategy I to 0.79, 0.62, 0.67, 0.75 with calibration strategy II, the median NSE increase from 0.78, 0.67, 0.68, 0.79 with calibration strategy I to 0.80, 0.78, 0.79, 0.83 with calibration strategy II. For SWAT, the NSE values in the validation period of IMERG-E, IMERG-L, IMERG-F show a significant increase from 0.70, 0.58, 0.63 with the strategy I to 0.75, 0.78, 0.73 with the strategy II, the median NSE increase from 0.67, 0.53, 0.51 with the strategy I to 0.70, 0.67, 0.63 with the strategy II. For the LSTM, the NSE values of flood events simulation also show higher mean values and smaller uncertainty based on the strategy II for all precipitation products, the flood events simulation based on IMERG-F shows the most significant improvement with the mean NSE value increasing from 0.59 with the strategy I to 0.75 with the strategy II, the median NSE value increase from 0.62 to 0.77.

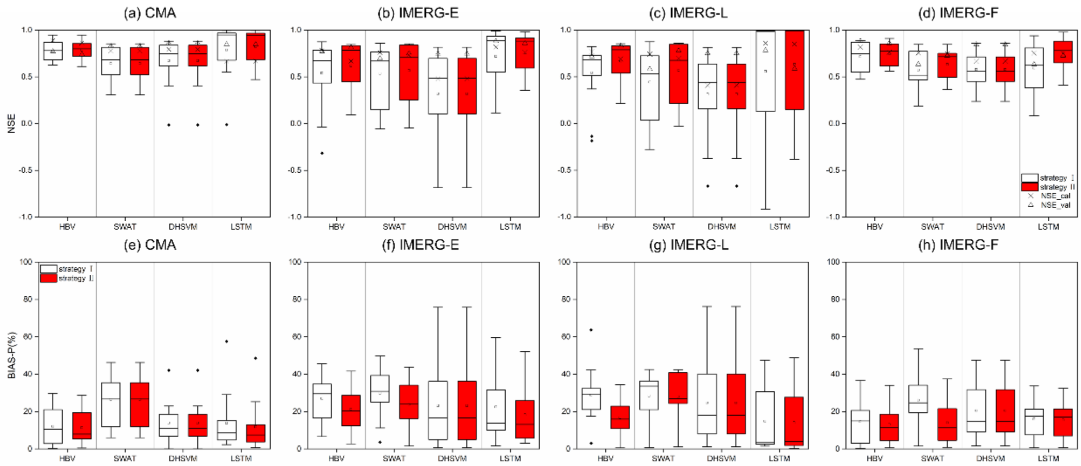

Fig. 6. The NSE and BIAS-P of flood events simulation forced by (a, e) CMA, (b, f) IMERG-E, (c, g) IMERG-L and (d, h) IMERG-F using two calibration strategies (White box is based on calibration strategy I; red box is based on calibration strategy II). The box plots show the 25th, 50th, and 75th percentiles, and the mean value is given and shown by a square. The cross represents the NSE of simulated streamflow during calibration, and the triangle represents the NSE of simulated streamflow during validation.

Point 6: NSEs in Figure 7: Again, I do not see a consistent pattern.

Response 6: Thank you for your comment. In the original manuscript, we have discussed why there is not a consistent pattern for NSEs, and we think the impacts of spatial resolution on flood events simulation behave differently among different models and precipitation sources. The discussion part is as follows:

Page 19-20 Line 450-474 'For the study area, under 0.25° spatial resolution, the CMA obtains the best flood events simulation based on SWAT and LSTM. The impact of spatial resolution on the capture of precipitation variability during flood event periods can propagate to the flood events simulation. Best results are obtained under 0.25° spatial resolution, the possible reason can be that finer spatial resolution (0.1°) increases the uncertainty of precipitation sets, nevertheless coarser spatial resolution (0.5°)

decreases the sufficiency of datasets.

The SWAT and DHSVM model driven by IMERG performs similarly under different spatial resolutions, which is consistent with previous research results (Lobligeois et al. 2014, Huang et al. 2019), where insignificant improvement was reported with higher spatial resolution of observed rainfall. It probably dues to the large catchment area and only the outlet station is used for calibration. Liang et al. (2004) found a critical resolution (1/8° for the VIC model) for a watershed with 1,233 km2, beyond which the spatial resolution shows limited impact on model performance. For our study area (82,375 km2), when the spatial resolution of precipitation changed from 0.1° to 0.5°, small variety is shown in the performance of flood events simulation, which indicates the critical resolution may be larger for large watershed.

For data-driven model, IMERG-E and IMERG-F show better performance under 0.1° spatial resolution in the LSTM-based simulation, which indicates that a higher spatial resolution (larger data set) can improve the performance of flood events simulation. Similar conclusion is drawn from previous study conducted by Sun et al. (2017), which also found that deep learning model performs better with larger datasets. In addition, the simulation with IMERG-L at 0.1° spatial resolution is not satisfactory, which may be related to the choice of hyperparameters and the limited data. However, after upscaling, the performance of LSTM in flood events simulation is greatly improved when the IMERG-L data is applied with 0.25° spatial resolution, which implies that scale transformation can be regarded as an approach of data enhancement in hydrological simulation based on deep learning.'

Reference:

Huang, Y., Bárdossy, A. and Zhang, K. 2019. Sensitivity of hydrological models to temporal and spatial resolutions of rainfall data. Hydrology and Earth System Sciences, 23(6), 2647-2663.

Lobligeois, F., et al. 2014. When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation using 3620 flood events. Hydrology and Earth System Sciences, 18(2), 575-594.

Liang, X., Guo, J. and Leung, L. R. 2004. Assessment of the effects of spatial resolutions on daily water flux simulations. Journal of Hydrology, 298(1-4), 287-310.

Sun, C., et al. 2017. Revisiting Unreasonable Effectiveness of Data in Deep Learning Era. 2017 Ieee International Conference on Computer Vision (Iccv), 843-852.

Point 7: Results and discussions should be put together. It is difficult to follow discussion when results are not immediately available.

Response 7: Thank you very much for your comments. We are very sorry for the difficulty in reading. As we mentioned above, we separate the results and discussion parts for the reason that we can discuss the results more thoroughly. In order to make it easy to follow, we have pointed out where to find the corresponding results, for example, "Compared with the conventional method choosing the fit parameter set based on entire streamflow time series (Calibration Strategy I), selecting the parameter set that results in the best flood events simulation (Calibration Strategy II) shows better performance on flood event simulation (Fig. 6)." Hope for your understanding.