Author’s response

Dear editor and reviewers,

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

Response to editor

Point 1: The comments of referees are very valid. The authors have shown that they appreciate these and have a plan for the paper revision. I would suggest to give special attention to the general comments of Referee 1, more clearly highlighting the novelty of this work.

Response 1: Thank you very much for your comment. We have carefully responded to the comments of Referee 1 and the novelty of this work is clearly highlighted in the introduction part, which are listed as follows:

Page 3-4 Lines 85-94: ‘But rare studies have been conducted to probe the effects of spatio-temporal of satellite-based precipitation on flood simulation, not to mention its impact on flood simulation with models based on DL methods (*e.g.*, LSTM). What’s more important, 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. Therefore, three widely used and typical physically based models (lumped HBV model, semi-distributed SWAT model, and distributed DHSVM model), and one data-driven model (LSTM) which shows good performance in hydrological simulation, are employed to probe the impacts of spatio-temporal resolutions of precipitation on flood events simulation.'

Page 4 Lines 105-108: 'However, studies about event-based calibration are still quite limited, particularly for LSTM. Therefore, in this study, we conduct different calibration strategies aimed at obtaining the best possible flood events simulation.'

Response to community comment

Point 1: Meaningful study! could the authors distinguish the sensitivities of these different models to the spatio-temporal resolutions of precipitation? and explain the reasons?

Response 1: Thank you for your question. In our study, the hydrological models are more sensitive than the machine learning model on the whole, but the sensitivity of the model is related to the precipitation input.

As illustrated in Fig.7 and Fig.8, when the spatiotemporal resolution of CMA changes, DHSVM is the most sensitive one, the mean NSE of flood events simulated with which declines from 0.68 to 0.45 when the spatial resolution of precipitation changes from 0.1° to 0.5°. Perhaps, in the case of CMA-driven DHSVM, the impact of spatial resolution on the capture of precipitation variability during flood event periods can propagate to the flood events simulation.

But when the spatiotemporal resolution of IMERG changes, the SWAT and DHSVM model perform similarly under different spatial resolutions, which is consistent with previous research (Lobligeois et al. 2014, Huang et al. 2019), where insignificant improvement was reported with higher spatial resolution of observed rainfall in a large catchment area. 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 km^2, beyond which the spatial resolution shows limited impact on model performance. For our study area (82,375 km^2), when the spatial resolution of precipitation changes from 0.1° to 0.5°, a small variation is shown in the performance of flood events simulation, which indicates the critical resolution may be larger for large watersheds. For HBV, it is not sensitive to changes in temporal resolutions because its simple hydrological model structure.

For LSTM, even though its sensitivity to the precipitation is lower than that of hydrological models, a higher resolution shows better performance. A similar conclusion is drawn from a previous study conducted by Sun et al. (2017), which found that deep learning model performs better with larger datasets.

References:

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.

Response to Referee 1

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

Response 1: Thank you very much for your comment. We agree that model calibration considering parts of discharge time series is not a new idea for hydrological model. As we clarified in the introduction part, “However, studies about event-based calibration are still quite limited, particularly for LSTM. Therefore, in this study, we conduct different calibration strategies aimed at obtaining the best possible flood events simulation.” Furthermore, besides calibration strategy, input data and model structure are the two main factors which affect the accuracy of flood events simulation and prediction, which are actually our primary focus. 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 application 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 comment. We have modified the relevant description of flood event in Lines 19-22: “Two calibration strategies are carried out, one of which targets at matching the flood events with peak discharge exceeding 8600 m^3/s between January 2015 and December 2017, 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."

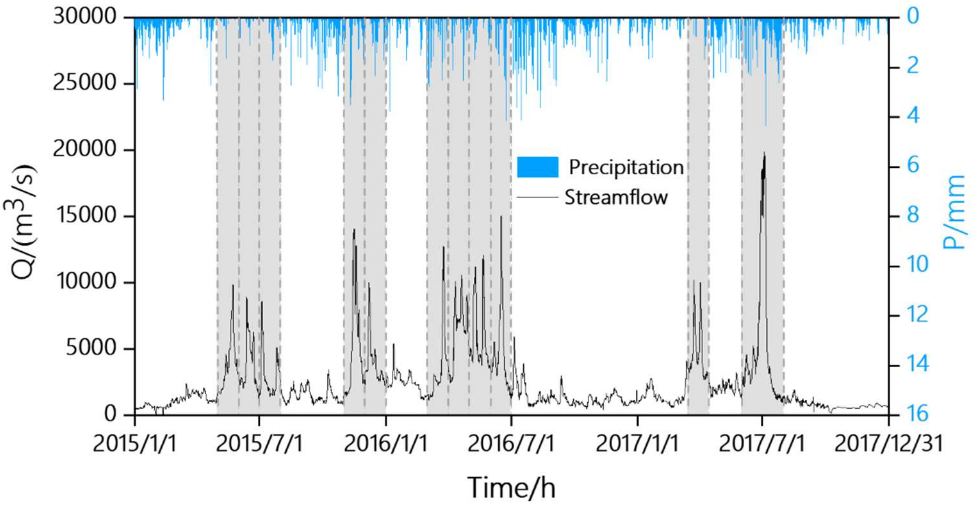


Fig. 2. Time series of observed hourly streamflow in Xiangtan station and basin-average precipitation from CMA, with eleven selected flood events covered by shaded areas.

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, while 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.

As you said, the 75th percentile of NSE decreases in case of CMA. Upon checking the values, we found that it falls from 0.865 with calibration strategy I to 0.855 with calibration strategy II, indicating a very slight difference. Additionally, the other evaluation index, BIAS-P, shows better performance for calibration strategy II compared to calibration strategy I. Therefore, since it is targeted to compare the two calibration strategies, as a whole, we can summarize that calibration strategy II is better than calibration strategy I.

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.

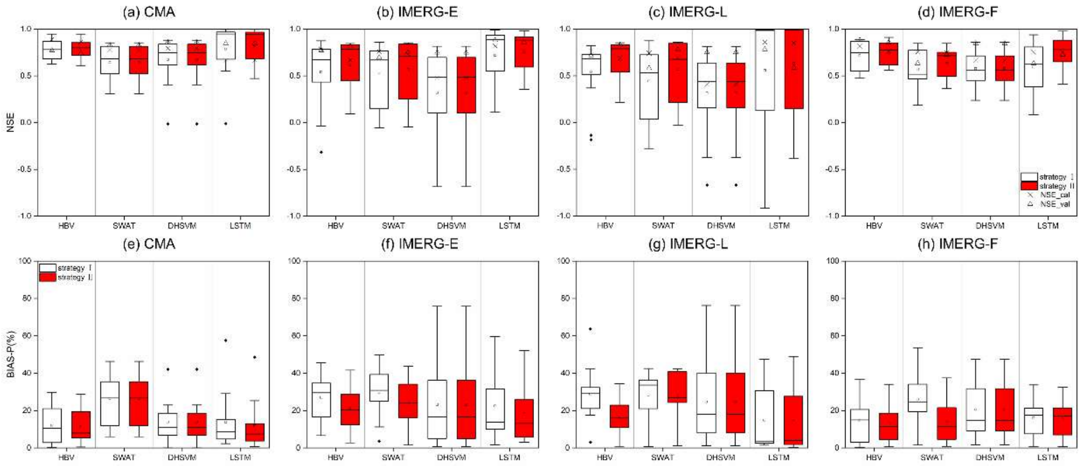


Fig. 6. The NSE and BIAS-P of flood events simulation forced by (a, e) CMA, (b, f) IMERG-E, (c, g) IMERG-L and (d, h) IMERG-F using two calibration strategies (White box is based on calibration strategy I; red box is based on calibration strategy II). The box plots show the 25th, 50th, and 75th percentiles,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

Please refer to “5.1 Comparison of two different calibration strategies” for the corresponding discussion. Thanks.

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

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

Page 18-19 Line 426-450 ‘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. The 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 perform similarly under different spatial resolutions, which is consistent with previous research (Lobligeois et al. 2014, Huang et al. 2019), where insignificant improvement was reported with higher spatial resolution of observed rainfall in a large catchment area. 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 km^2, beyond which the spatial resolution shows limited impact on model performance. For our study area (82,375 km^2), when the spatial resolution of precipitation changes from 0.1° to 0.5°, a small variation is shown in the performance of flood events simulation, which indicates the critical resolution may be larger for a 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, namely a larger dataset, 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 a 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.’

References:

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: We sincerely apologize for any difficulties you may have experienced while reading. As we outlined previously, we elected to separate the results and discussion sections, enabling us to delve into a more comprehensive examination of our findings. To enhance clarity and facilitate comprehension, we have highlighted where to locate the relevant results, for instance, "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.

Response to Referee 2

Point 1: Line 66: “the study”-> “they”.

Response 1: Thank you for your suggestion. This sentence has been re-edited:

Page 3 Lines 65-68: ‘Su et al. (2020) assessed the IMERG products at multiple spatial and temporal resolutions by upscaling, and they summarized that degrading the spatio-temporal resolution improves the accuracy of IMERG products.’

Point 2: Line 124-127: The structure of these two sentences is suggested to be revised. The conjunction "so" in the beginning of the second sentence may be unclear.

Response 2: Thank you for your suggestion. This sentence has been re-edited:

Page 5 Lines 126-129: ‘Concentrated storm events during the flood season cause frequent floods throughout the basin. Since the Xiang River basin is the most densely populated and economically developed area in Hunan Province (Zhu et al. 2020a), it is critical to accurately simulate and predict flood events in the region for effective flood risk management.’

Point 3: Line 139: “(hereafter CMA)” needs to be put behind “China Meteorological Administration”.

Response 3: Thank you for your suggestion. This sentence has been re-edited:

Page 6 Lines 141: ‘A precipitation product released by China Meteorological Administration (hereafter CMA),’

Point 4: Line 185: the reference “(AghaKouchak et al. 2013)” should be located behind the “HBV model”.

Response 4: Thank you for your suggestion. This sentence has been re-edited:

Page 8 Lines 186-187: 'A lumped version of HBV model (AghaKouchak et al. 2013) is used in this study,'

Point 5: Line 230-233: please pay attention to the format of the variables, such as xt, and t.

Response 5: Thank you for your question. I am sorry for our carelessness; the format of the variables has been corrected:

Page 10 Lines 233-236: 'The inputs for the complete sequence $x = [x_1, ..., x_n]$, where x_t is a vector containing the input features of time t, and the dimension of the x_t corresponds to the number of grids of the precipitation data. The outputs for the complete sequence $y = [y_1, ..., y_n]$, where y_t is the streamflow of time t.'

Point 6: Line 268-269: please explain how the eleven historical flood events are selected.

Response 6: 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."

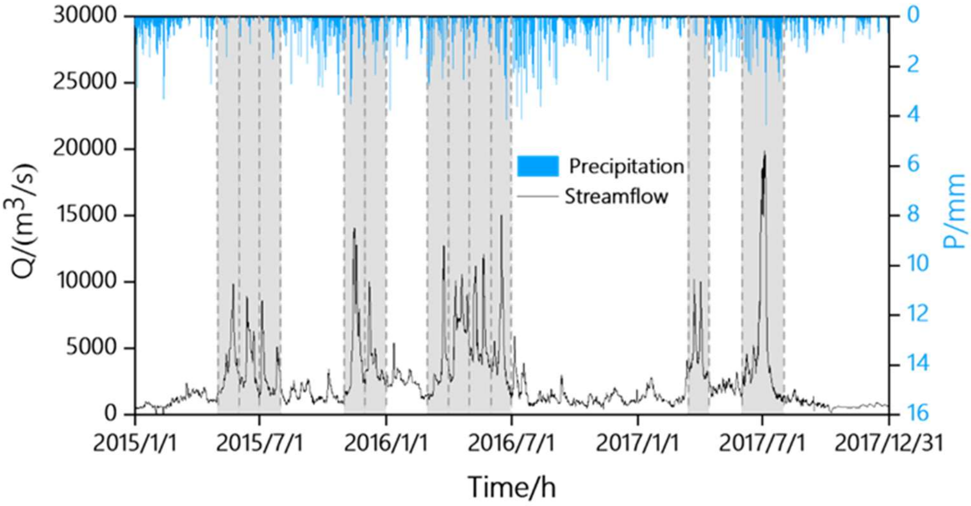


Fig.2. Time series of observed hourly streamflow in Xiangtan station and basin-average precipitation from CMA, with eleven selected flood events covered by shaded areas.

Point 7: Line 338: "as the resolution get coarser"-> as the resolution is coarser or as the resolution gets coarser.

Response 7: Thank you for your suggestion. This sentence has been re-edited:

Page 14 Lines 340-341: 'The performance of IMERG-F gets worse as the resolution is coarser,'

Point 8: Line 350-352: "However, the uncertainty of NSE, KGE and BIAS-P values of flood events simulated with IMERG is decreasing as the spatial resolution." As the spatial resolution what? finer or coarser?

Response 8: Thank you for your question. We are very sorry for the difficulty in reading. This sentence has been re-edited:

Page 15 Lines 354-355: 'However, the uncertainty of NSE, KGE and BIAS-P values of flood events simulated with IMERG decreases as the spatial resolution is finer.'

Point 9: Line 365: in most instances -> in most cases.

Response 9: Thank you for your suggestion. This sentence has been re-edited:

Page 16 Lines 368: 'The mean NSEs of LSTM are higher than 0.7 in most cases,'

Point 10: Line 407-408: "the same results" means the results are exactly the same, does that what the authors indicate? Otherwise, the same results -> the comparable/similar results or the results are almost the same.

Response 10: Thank you for your suggestion. This sentence has been re-edited:

Page 18 Lines 410-412: 'However, the CMA shows the similar results under two different calibration strategies in SWAT-based flood events simulation.'

Point 11: Line 417-418: the calibration strategy II is an effective way for training the LSTM model to obtain the best flood events simulation results -> the calibration strategy II is an effective way to train the LSTM model to obtain the best flood events simulation.

Response 11: Thank you for your suggestion. This sentence has been re-edited:

Page 18 Lines 419-421: 'When comparing the two calibration strategies, the calibration strategy II is an effective way to train the LSTM model to obtain the best flood events simulation.'

Point 12: Line 430: performs -> perform.

Response 12: Thank you for your suggestion. This sentence has been re-edited:

Page 19 Lines 433: 'The SWAT and DHSVM model driven by IMERG perform similarly under different spatial resolutions,'

Point 13: Line 431: please delete the "results". And please check the whole manuscript for this issue.

Response 13: Thank you for your suggestion. This sentence has been re-edited:

Page 19 Lines 434: 'which is consistent with previous research (Lobligeois et al. 2014, Huang et al. 2019),'

And we checked the whole manuscript for this issue as you suggested. Thanks.

Point 14: Line 440: larger data set -> larger dataset. Isn't the "Fig. 9" shall be colored red to be consistent with other figures?

Response 14: Thank you for your suggestions. This sentence has been re-edited:

Page 19 Lines 443-444: 'which indicates that a higher spatial resolution, namely a larger dataset, can improve the performance of flood events simulation.'

We have changed the color of Fig. 9 to make it consistent with other figures.:

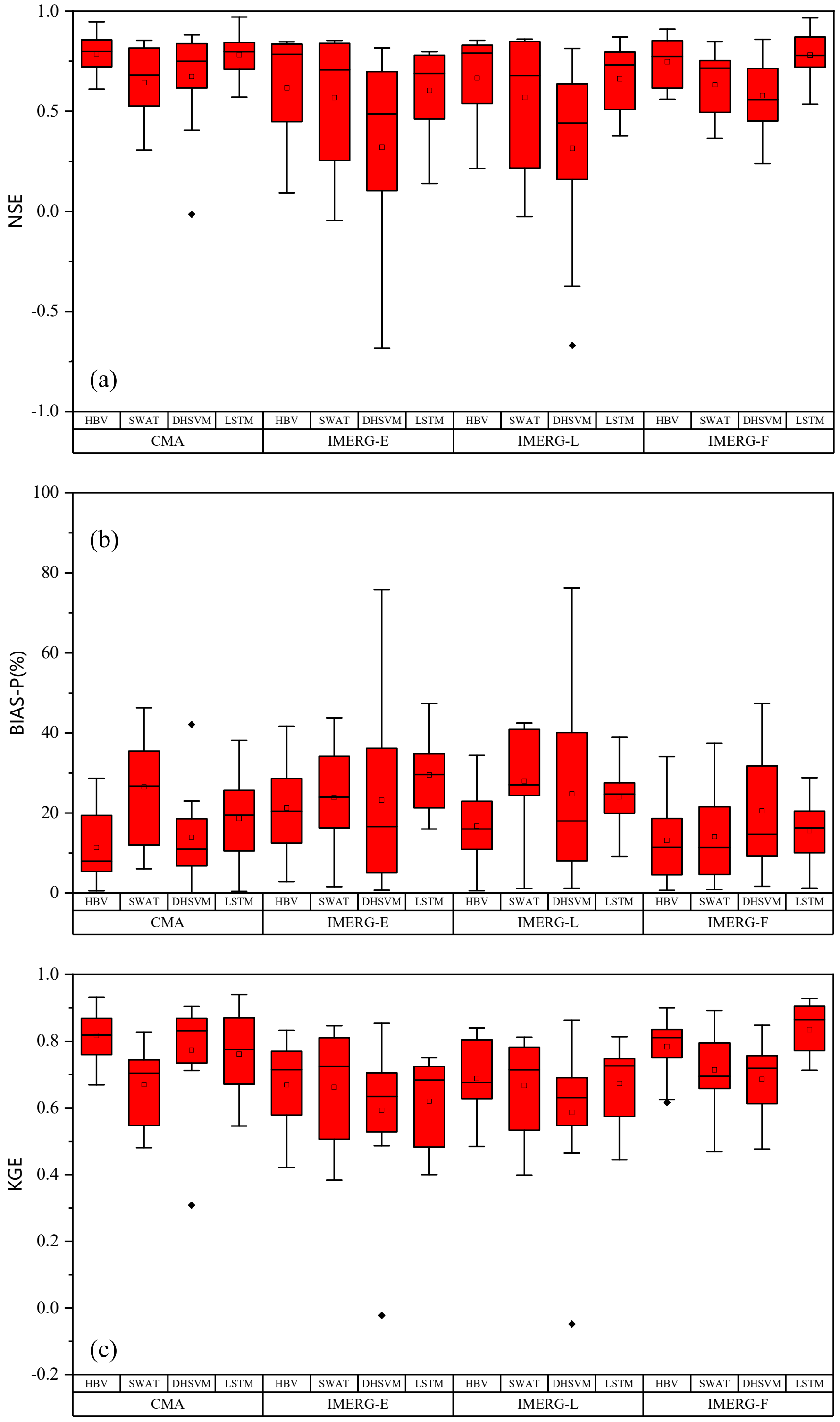


Fig. 9. The (a) NSE, (b) BIAS-P and (c) KGE of flood events simulation forced by CMA, IMERG-E, IMERG-L and IMERG-F using calibration strategies II. The box plots show the 25th, 50th, and 75th percentiles, and the mean value is given and shown by a square.

Point 15: The colors used in Fig.10 are not so easy to distinguish.

Response 15: Thank you for your question. We are very sorry for the difficulty in reading. We have changed the color of Fig.10 to make it easier to distinguish:

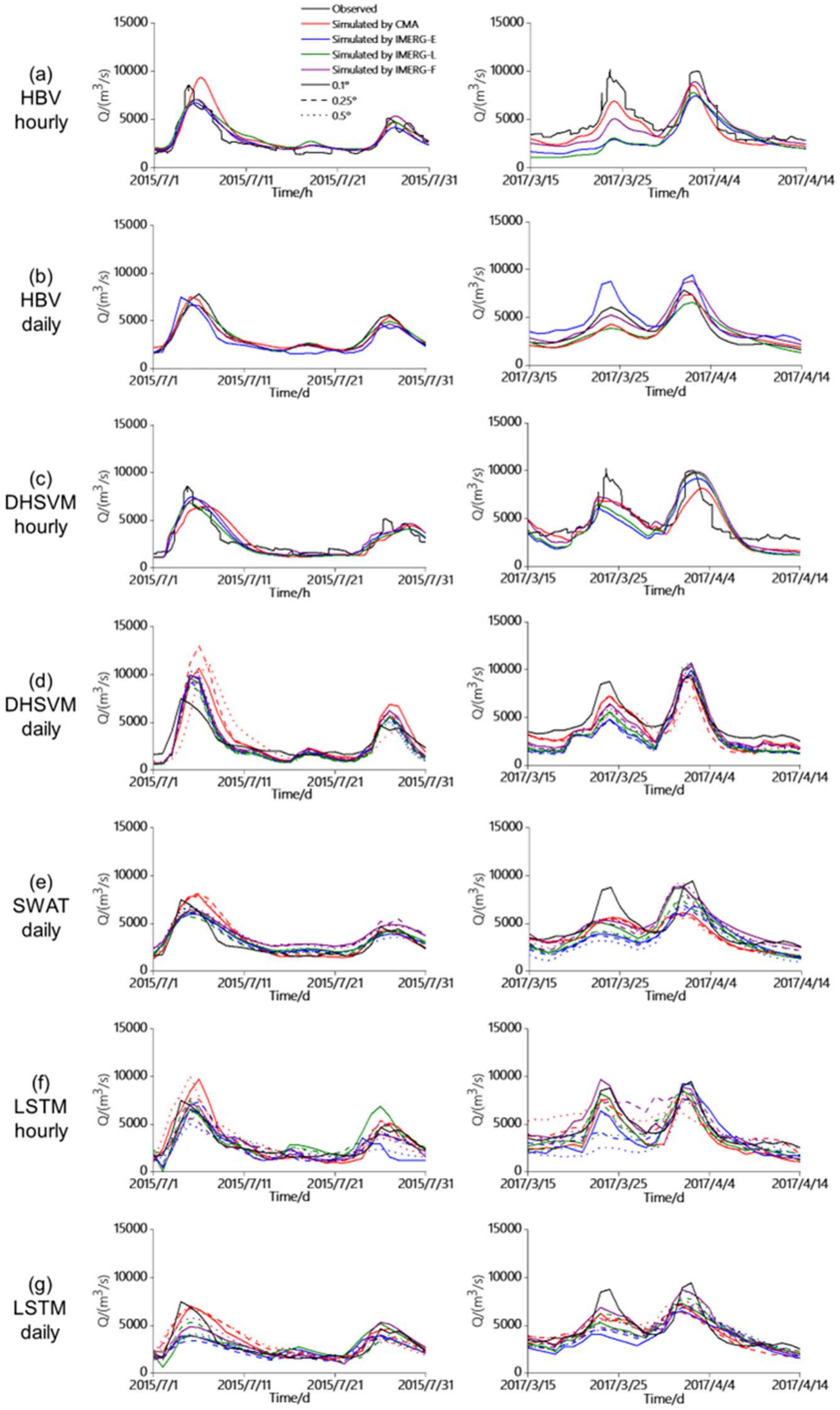
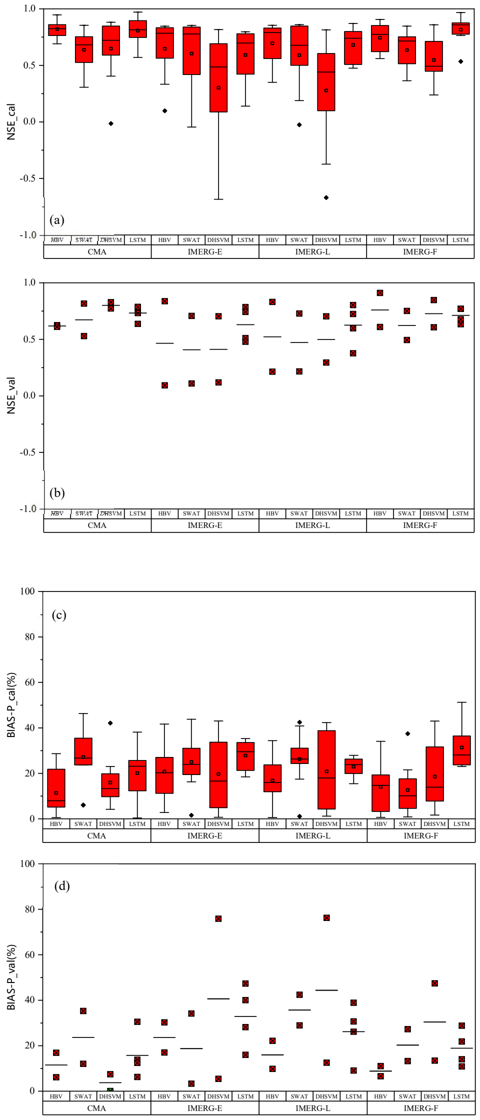


Fig. 10. Comparison of HBV, SWAT, DHSVM, and LSTM based flood events simulation from July 1st, 2015 to July 31th, 2015, and from March 15th, 2017 to April 14th, 2017 forced by CMA, IMERG-E, IMERG-L, and IMERG-F with different spatio-temporal resolutions.

Point 16: Same issue of Appendix C, and please refer to the comment #15

Response 16: Thank you for your question. We are very sorry for the difficulty in reading. We have changed the color of Appendix C to make it consistent with other figures.:



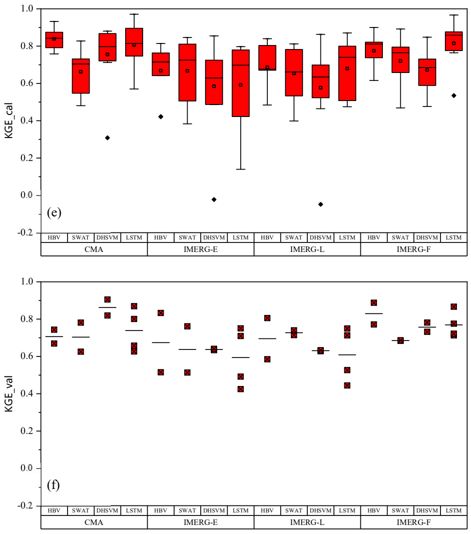


Fig. C0. Same as Fig. 9, but the results in calibration and validation periods are separated