Response to Anonymous Referee #2

Comments/text of reviewer posted in **bold** *and italics*; the authors's answeres start with "***Response***:"; the sentences in the revised version is in blue.

Anonymous Referee #2

Referee comment on "The impact of spatiotemporal structure of rainfall on flood frequency over a small urban watershed: an approach coupling stochastic storm transposition and hydrologic modeling" by Zhengzheng Zhou et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2021-161-RC2, 2021

This paper addresses the issue of the impact of small scale spatio-temporal rainfall variability on the hydrologic response of small urbanized catchment. The topic is relevant for the community and HESS journal. The paper is well written, and rather straightforward to read. However I believe that major modifications are needed before a potential publication in HESS. I have three major comments :

- Results are often not so clear with contrasting trends. In order to clarify the outcome of the study, I think that statistical significance of results should be quantified more systematically. This could help to clarify some results.

Response: Thank you for your suggestions. The significance level of rainfall frequency is done in the study of Zhou *et al.* (2019). In this study, we use the results of rainfall frequency and then simulate the flood discharges based on the rainfall results. The upper and lower level of flood simulation can be found in Figure 5 in the revised version.

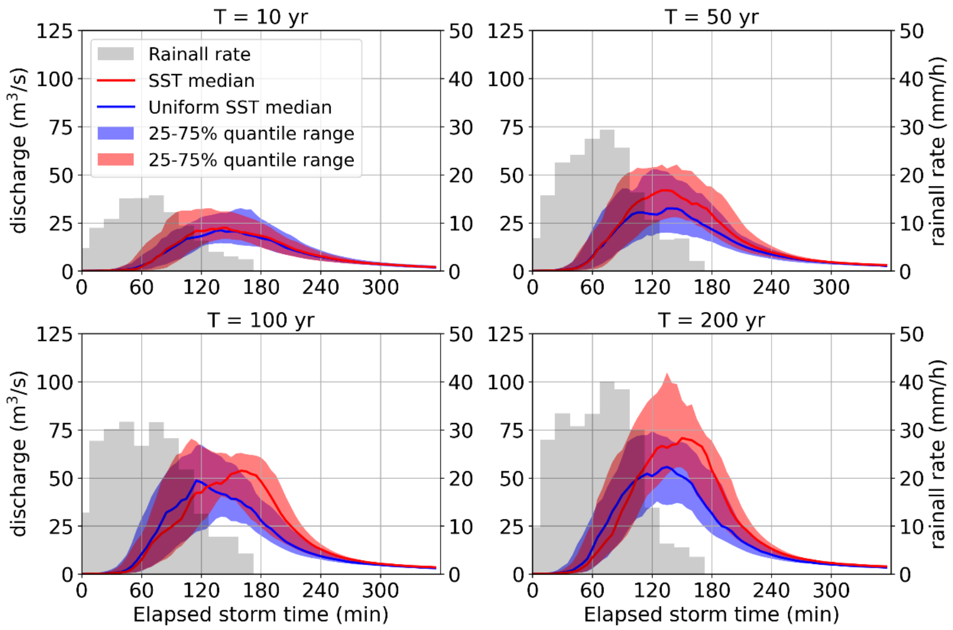

Figure 5. Time series of simulated hydrographs for Franklintown based on the 3-h design storms from 10-yr to 200-yr return periods with spatially uniform (blue) and spatially distributed (red) rainfall. The grey bar indicates the median value of basin-averaged rainfall rate.

- Physical interpretation of the results should also be systematically provided, with quantification of the various effect implemented (notably the percentage of imperviousness of the various catchments). This would enable to enlarge the potential application of the paper's results to other locations.

Response: Generally, the differences in hydrologic responses among urban basins cannot be fully accounted for by differences in impervious cover (Zhou *et al.*, 2017). For example, we compared the flood peaks at DR1, DR2 and DR5 which have similar basin scales (1.32, 1.92 and 2.05 km^2, respectively). DR1 with higher imperviousness and detention controlled area than DR2, has higher flood peaks than DR2. However, DR5 with lower imperviousness and detention controlled area has lower peaks than DR1 and DR2. A common agreement is that imperviousness tends to increase flood peaks while detention infrastructure tends to decrease peaks. Thus, the comparison among the three small-scale basins show that difficulties remain in attributing specific changes in urban flood peak distributions to specific urbanization characteristics. We have addressed the complexities in urban flood responses among the DR watershed, which can provide implementations for other urban watersheds. The role of urban features in flood responses is beyond the scope of this paper. Thank you.

- It is not clear at what spatial scale are the rainfall return periods estimated ? This could have a strong impact on the results.

Response: The rainfall frequency is estimated for the entire DR watershed with a spatial scale of 14.3 km^2.

Detailed comments :

2) Data and methods

2.1) Study region and data

- Fig. 1.b : could you please clarify the configuration of the catchment, and notably the path of the various rivers through the catchment, which do not seem to converge toward the outlet of the catchment ? Please also update the colours because the map is difficult to read.

Response: Figure 1 has been modified. The outline of streams are clarified in Fig.1b. The colors are changed to present the main types of land use land cover. For the details of the configuration of watershed used in GSSHA model, the readers are directed to Smith *et al.* (2015).

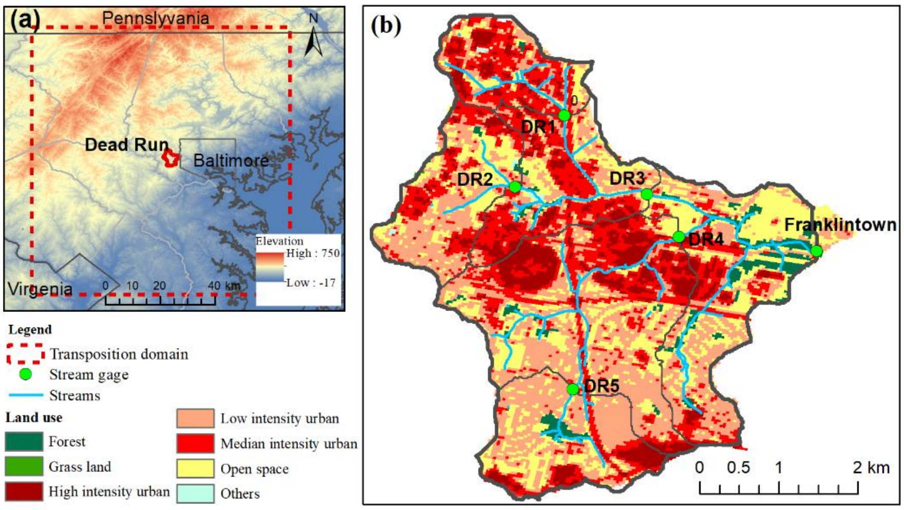

Figure 1. Overview of Dead Run study region including (a) location of DR, elevation, and transposition domain of SST; (b) land use land cover and stream gages. The red outline and grey outline in (a) indicates the boundary of DR watershed and Baltimore City, respectively.

- Table 1 : please clarify what is "developed land" and "controlled area".

Response: The note is added in the end of Table 1 in the revised version as follows.

"Note:

a. Developed lands include "Developed, open space" (>20% impervious surface), "Developed, low intensity" (20%-49% impervious surface), "Developed, medium intensity" (50%-79% impervious surface), and "Developed, high intensity" (80% or more impervious surface). Data source: USGS 2012 National Land Cover Dataset (NLCD).

b. Dentention controlled areab refers to the area controlled by detention infrastructure."

- l. 94-98 : sentence weirdly written, please rephrase

Response: We have rephrased the paragraph and added a general introduction of bias correction for radar data. In the revised version, it is written as follows. "High resolution (15-min temporal resolution, 1-km^2 spatial resolution) 34radar rainfall fields for the 2000-2015 period were derived from volume scan reflectivity fields from the Sterling, Virginia WSR-88D (Weather Surveillance Radar-1988 Doppler) radar. The two-dimentional radar rainfall fields are then developed from the reflectivity fields using the Hydro-NEXRAD algorithms (Krajewski *et al.*, 2011) which have been used in rainfall and hydrological studies (Smith *et al.*, 2007; Lin *et al.*, 2010; Smith *et al.*, 2013; Wright *et al.*, 2014; Zhou *et al.*, 2017). The Hydro-NEXRAD algorithms includes quality control algorithms, *Z-R* conversion of reflectivity to rainfall rate, time integration, and spatial mapping algorithms (Seo *et al.*, 2011). To improve the rainfall estimates, a multiplicative mean-field bias correction (Smith and Krajewski, 1991; Wright *et al.*, 2012) is applied on a daily basis using a network of 54 rain gauges in and around the Baltimore County. The bias computation takes the form $\mathrm{B}_i = \frac{\sum_{S_i} G_{ij}}{\sum_{S_i} R_{ij}}$. Where G_{ij} is the rainfall accumulation for gage j on day i, R_{ij} is the daily rainfall accumulation for the co-located radar pixel accumulation on day i, and S_i is the index of the rain gage stations for which both the rain gage and the radar report positive rainfall accumulations for day i. Each 15-min radar rainfall field from day i is then multiplied by B_i."

2.2) GSSHA hydrological model

- Please explain how the interactions between surface flow and stormwater system is handled, because it is crucial in highly impervious areas.

Response: The 2-D overland flow grid empties into 1-D stream channels represented by cross sections that contain both the stream channel and the floodplain. Surface water is accumulated until the specified retention depth of the cell is exceeded. The overland flow is then routed in two orthogonal directions using Manning's equation with the diffusion wave form of the de St. Venant equations to estimate friction slope. When the overland flow reaches a model grid cell that contains a channel node, the flow is passed into the channel and routed using a 1-D technique through the stream network.

Culverts were measured in the field and represented in-stream as cross sections. Storm pipes were represented through cross sections with a rounded half-circle bottom and walls reaching up vertically. The circle's diameter was set to the pipe diameter. Detention basins were represented within the channel with cross sections extracted from the 1-m lidar topographic data. Water was backed up within the detention basins using rating curves at the detention basin outlets. The rating curves were based on a simple orifice equation, calculated for a fully submerged outlet pipe with a size determined by the downstream pipe size. Over all streams, culverts, pipes, and detention basins 989 cross sections were included within the model for an average spacing of 72.8 m between distinct cross sections.

2.3) SST procedure

- Over which area is rainfall computed to estimate the return periods ?

Response: The rainfall is estimated for the entire DR watershed (14.3 km^2) with spatiotemporal structure.

- l. 153 : "to generate estimates for return periods up to 500 years". I believe an assessment of the corresponding uncertainties should be provided.

Response: In the study of *Zhou et al.* 2019, 1000 realizations of 500-yr series are generated, and the median values of 1000 realizations are used to generate estimates for return periods up to 500 years. The 90^{th} and 10^{th} quantiles of 1000 realizations are set as the lower and upper bound of SST estimates (Figure R1). The uncertainties of rainfall estimations are demonstrated in detail in *Zhou et al.* 2019.

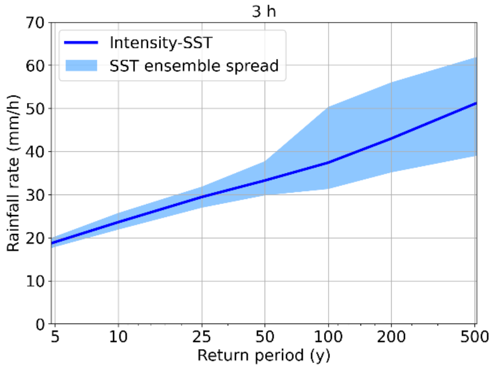

Figure R1. The SST- based intensity-duration-frequency curves.

2.4) Characteristics of rainfall and hydrologic response

- Eq 1 : given it is a basin average at time t, I guess that the integral should not be over time (T), but over the spatial domain.

Response: Eq.1 has been revised as $\mathrm{M(t)} = \int_A R(t,x)dx$. Thanks.

- Eq. 3 : if M(t) is a rain rate, I guess the time step should appear in the computation of the cumulative depth. More generally, please indicate units of all the quantities used.

Response: The units of the quantities are labeled as "Peak basin-average rainfall rate (mm/h)" and "storm total rainfall depth (mm)". Thanks.

- l. 182-183 : please clarify what you are calling a "unimodal distribution for rainfall".

Response: The unimodal distribution indicates that there is spatially one peak over the entire watershed. It is clarified in the revised version.

- Are all the indicators really used ? I do not have the feeling that they are really all used and that the high number is just creating a bit of confusion. I would either really exploit all of them or reduce their number.

Response: The high-resolution radar rainfall data provides us a good opportunity to examine the details of rainfall structure. We then extract the indicators include temporal characteristics of rainfall (rainfall peak, rainfall rate and total rainfall) and spatial characteristics of rainfall (spatial coverage, rainfall location and rainfall-weighted flow distance). By using these indicators, the rainfall structure can be described comprehensively. And then coupled with a hydrological model, we can capture the rainfall-flood relation well. Through the examination of all the indicators, the main features that impact flood responses can be found.

Therefore, we prefer to keep all the indicators in the content. They are all mentioned in the results section. A complimentary demonstration is shown in Figure A3, showing the relationship between spatiotemporal rainfall structure and flood responses. The indicators that are highly related to flood responses are highlighted and addressed.

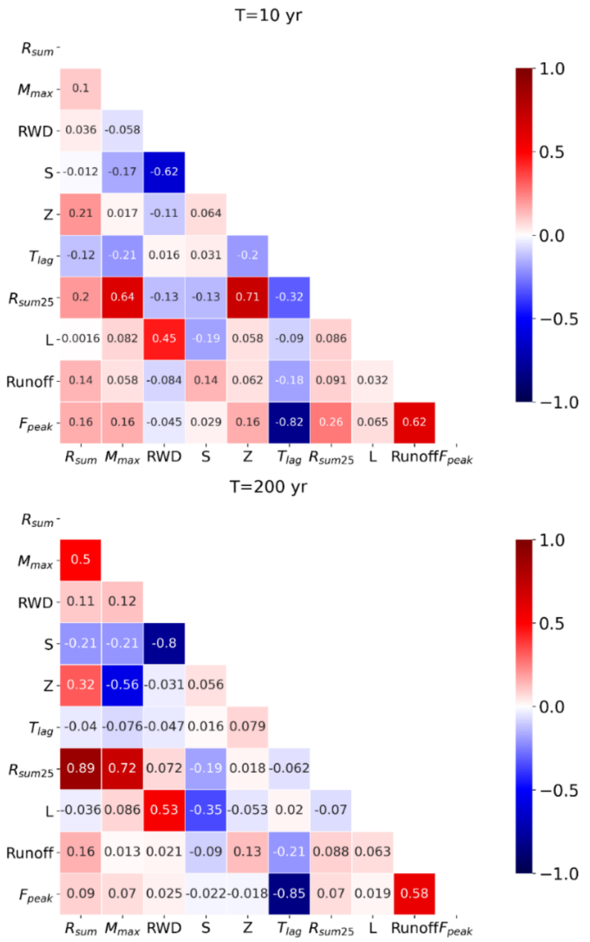

Figure A3: Correlation between space-time rainfall structure and flood responses at Franklintown under 10-yr and 200-yr return periods.

3) Result and discussion

3.1) Model validation

- I believe that an example of hydrographs should be included in the main document.

<u>Response</u>: Thank you for the suggestion. The hydrograph of the 14 August 2011 storm event is added as a representative simulation (Figure 3 in the revised version).

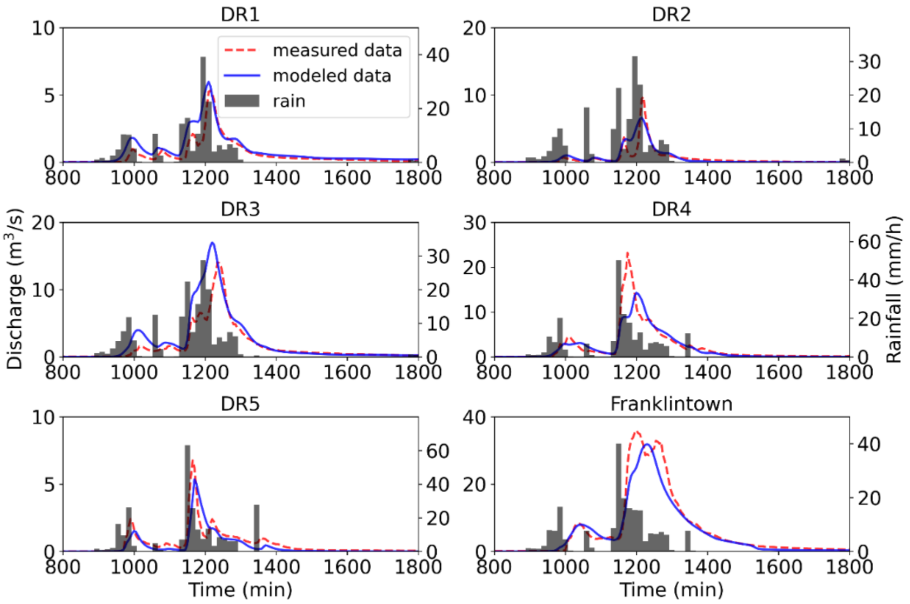

Figure 3. Hydrographs and rainfall for the the 14 August 2011 storm event. Time refers to minutes from the start of the model simulation.

- The differences observed in Fig. 2.a are quite significant and should be more discussed.

<u>Response</u>: Thank you for the suggestion. We have revised the paragraph to make it clearer. The flood magnitude at the downstream Franklintown gage is well captured with the median peak discharge difference of -14%. The largest median peak discharge difference of 57% is at DR1. The reason is likely that it has a large area of land which was not represented fully on county storm sewer maps (Smith *et al.*, 2015). A similar situation is found in Smith *et al.* (2015), and their research shows the model can be effective in producing hydrological responses. As mentioned in Section 2.1, the rating curves used to compute the discharge data at DR3 and DR4 are not provided from USGS. It may increase the error in the measurements and modeling results. Overall, for a large collection of flood events with various rainfall characteristics and peak discharges ranging from 70 m^3/s to 253 m^3/s, the model performs well to reproduce peak discharges, especially at the basin outlet.

The discussion of error from discharge is added in the revised version as follows. "It should be noted that the error in simulated response may be attributable to measurement errors tied to stage discharge curves and to conversions of radar reflectivity to rainfall rate, as well as to the features that were simplified within the model, such as initial soil moisture and some aspects of the storm drain network (Smith *et al.*, 2015). For example, it has been documented that the average error of discharge between USGS direct measurements and stage-discharge curves for Franklintown is 17.4% between 2008 and 2010 (Lindner and Miller, 2012). As mentioned in Section 2.1, the rating cruve used to compute the discharge data at DR3 and DR4 is not provided from USGS. It may increase the error in the measurements and modeling results. For the rainfall data set used in this study, the difference of the storm total rainfall between a rain gage and the bias-corrected radar rainfall data for the pixel of that gage is compared. The median difference for all gages over the 21 storms is 22.6%(Smith *et al.*, 2015)."

- Also, limiting the validation of the model to peak discharge and time of peak do not seem sufficient. Adding indicators using the whole hydrographs (such as the imperfect Nash-Sutcliffe efficiency and not only the skewness) would add some relevancy to the validation.

Response: Thank you for your suggestion. NSE is added in the revised version as follows. "The median Nash-Sutcliffe Efficiency (NSE) for the 21 events at Franklintwon is 0.77 (Fig. 2c). The best NSE at Franklintown is 0.97 indicating that the match between model and measured data was nearly exact. For the subwatersheds, the best median NSE is at DR-4 with a value of 0.74, while the least median NSE is at DR-1 with a value of 0.21. The results show that the main tendency of flood response is captured by the model."

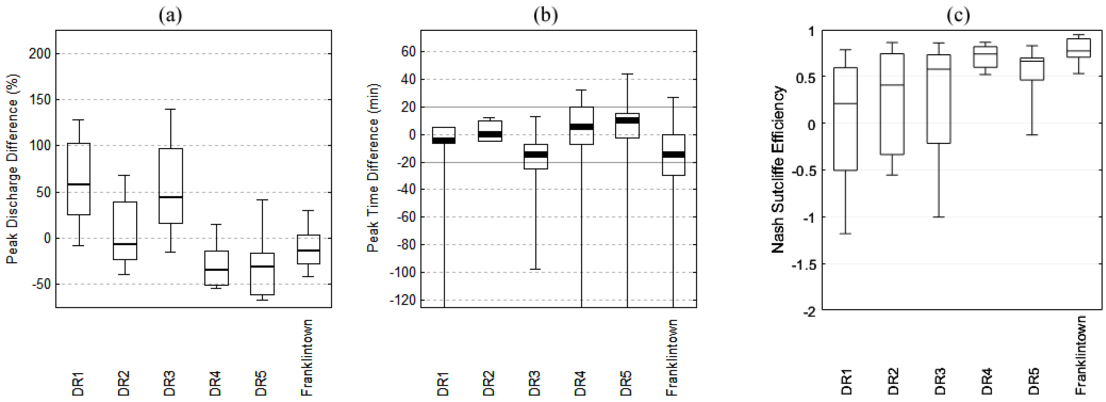

Figure 2. Comparison of (a) flood peak discharges, (b) response times and (c) NSE for 21 historical rainfall events.

- Fig 3.a : clarify how normalization is implemented. Also more explanation/interpretation of the multimodal behaviour for flood should be added. Is it due to location of rainfall ? Is the same behaviour also observed for the other stations.

Response: The equation for normalization is added in the revised version as follows. "The normalization is the ratio of values minus the minimum to the maximum minus minimum." The multimodal behavior for flood is caused by the mixed impact of rainfall process and the urban drainage system. We cannot attribute the behavior simply to rainfall nor the urban drainage system. The following results show that flood peak is more correlated with spatial rainfall features, implying that the multimodal distribution of flood peaks is more likely associated with spatial rainfall distributions. We have discussed in the revised version as follows. "The following results will show that flood peak is more correlated with spatial rainfall features, implying that the multimodal distribution of flood peaks is more likely associated with the spatial distribution of rainfall."

- l. 242-243 : please provide some explanations / interpretations to this retrieved behaviour.

Response: The interpretation is added in the revised version as follows. "For the 100-yr return period, the upper spread shows a tendency toward dual peaks, which cannot be revealed from conventional design flood practices. Since in the conventional rainfall flood frequency approach, the design storm is temporally idealized as a unimodal peak process. By using theses design storm, the flood response is generally simulated as a unimodal peak process. The above results imply the uncertainty and insufficiency of flood frequency analysis in the conventional method. For the 200-yr return period, the hydrograph is peakiest with a large upper spread."

- Fig. 5 and associated comments : the fact that no clear trend with basin size is found is somehow surprising. Differences in land use are mentioned and should be further explored and quantified I

think. Also, was a sensitivity analysis carried out on the choice of rainfall events ?

Response: Generally, basin scale plays an important role in determining the distribution of flood magnitudes. We discussed the difference in land use and linked the difference with the difference in flood peaks. However, the impacts are subtle, and the data are not robust enough to do more than suggest these implementations.

We did not manually select rainfall events to drive the hydrologic model. Based on the SST framework, the rainfall events are selected and transposed randomly. Thus, for each return period, the rainfall results are not from one or several specific rainfall events. A sensitivity analysis is not needed.

3.3) Rainfall-Flood relationship

- l. 284 : a correlation of 0.16 is very low, so I am not sure the wording "somewhat correlated is appropriate"

Response: The sentence is revised as "the flood peak is slightly correlated with…"

- Fig 7 and associated comments : please provide more insights about the "random forest regression model" for the non-specialist reader. Also, a lot of rainfall features while floods are quantified simply by flood peak. How significant are the differences between 10yr and 200 yr results ? Would a different selection of initial rainfall events lead to different results?

Response: The general introduction of RF model is discussed in the revised version "Random forests (RF) is an ensemble learning method (Breiman, 2001) that aggregates results from multiple models to achieve better accuracy. RF is one of the most widely-used method for regression and classification. Moreover, it is relatively easy to train and tests. In this study, rainfall spacetime structure characteristics are used as RF model features. The flood peak is set as the model target. The relationship between rainfall structure and flood peak is then explored under the RF-based regression method.".

The magnitude of flood peak is the main target in flood frequency analysis. Furthermore, in this study, the response time is less impacted in the small urbanized watershed. We hence focus on the flood peak.

According to the values of feature importance, the importance of the total rainfall decreased by 2% compared with the 10-yr return period. The largest difference is for the fractional coverage of storm core which increases by 5% from 10-yr to 200-yr return period. Both rainfall total and rainfall peak decrease from small return periods to large return periods; on the contrary for the dispersion of RWD (S) and fractional coverage of storm core (Z). Though it appears that the difference is moderate, but for a such small watershed, the tendency of the change of spatiotemporal rainfall feature importance is noteworthy. We have mentioned it in the revised version as follows. "For extreme storms, the maximum discharge is more closely linked to the spatial structure of rainfall, which is consistent with the results in (Peleg *et al.*, 2017; Zhu *et al.*, 2018). Though it appears that the difference is moderate, but for a such small watershed, the tendency of the change of spatiotemperol rainfall feature importance is noteworthy."

It is possible that the different rainfall events will lead to different results. But unlike many previous studies which used a limited number of rainfall events, in this study, we selected a storm catalog with 200 events with various spatiotemporal structures. And their transposed results are then further computed as the rainfall frequency analysis. For each return period, the 300 transposed events contain various rainfall structures which can provide more general and reasonable results.

- l. 306-309 : It is not obvious to me how the general conclusion is obtained from the Fig. 8. Please clarify ? Notably, the figure does not provide answers on the rapidness of flood response if I

understood it well …

Response: The Figure 8 and the discussion are removed from the manuscript. Thanks.

- l. 311-318 : I believe that the paragraph is actually quite interesting, and that more interpretation should be done. However, the spatial scale used to determine the rainfall return period should be clarified and is likely to play a significant role, notably in relation with the size of the studied catchment. For which of the catchments was Fig. 9 obtained ? Are different trends found for other catchments ?

Response: The rainfall results are for the entire watershed (14.3 km^2) and the flood frequency results are also for the same watershed. The limitation of this analysis is that the total number of realizations is only 30 with return periods up to 100 years. With more run of the model, it is possible to obtain some functional relationship between flood return periods and rainfall return periods, which could be done in our future work.

As shown in the following figure, similar results are found for the sub-basins (flood return periods at sub-basins vs. rainfall return periods at the DR outlet).

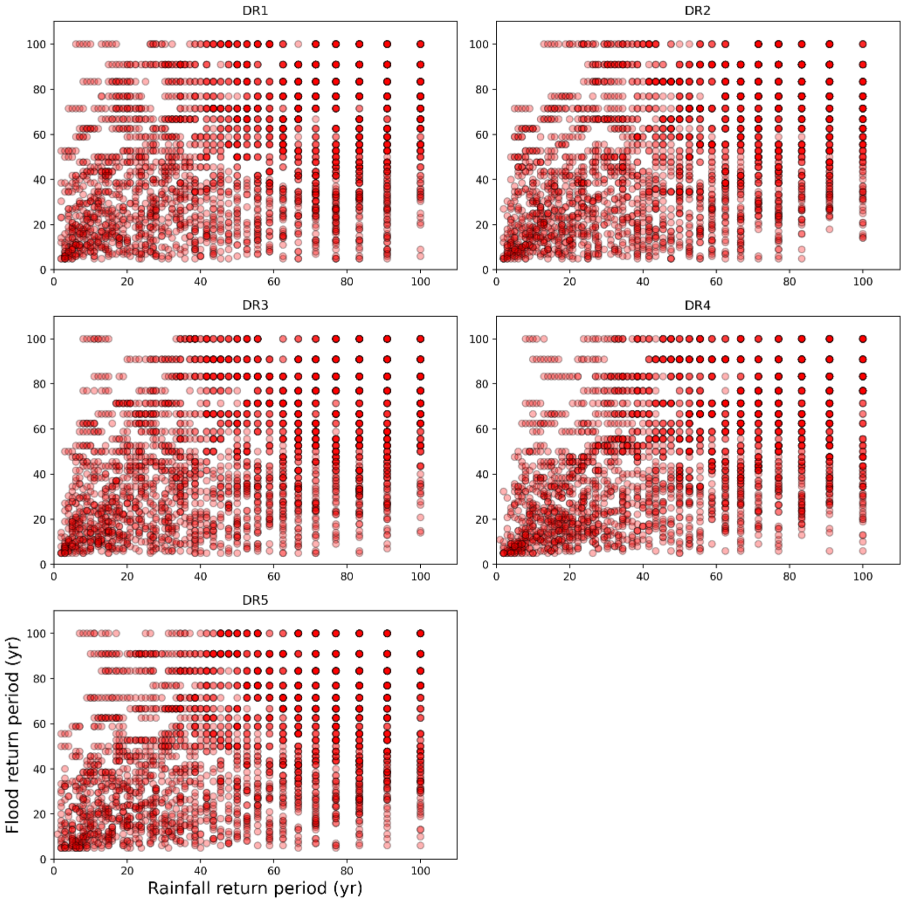

Figure R3. Scatterplot of return periods for DR-scale rainfall and subbasin-scale peak discharge.

Reference

Breiman, L. (2001), Random Forests, *Machine Learning*, 45(1), 5-32. 10.1023/a:1010933404324

Krajewski, W. F., A. Kruger, J. A. Smith, R. Lawrence, C. Gunyon, R. Goska, B.-C. Seo, P.

Domaszczynski, M. L. Baeck, and M. K. Ramamurthy (2011), Towards better utilization of NEXRAD data in hydrology: an overview of Hydro-NEXRAD, *Journal of hydroinformatics*, 13(2), 255-266. doi:10.2166/hydro.2010.056

Lin, N., J. A. Smith, G. Villarini, T. P. Marchok, and M. L. Baeck (2010), Modeling extreme rainfall, winds, and surge from Hurricane Isabel (2003), *Weather and forecasting*, 25(5), 1342-1361. doi:10.1175/2010WAF2222349.1

Lindner, G. A., and A. J. Miller (2012), Numerical Modeling of Stage-Discharge Relationships in Urban Streams, *Journal of Hydrologic Engineering*, 17(4), 590-596. doi:10.1061/(ASCE)HE.1943-5584.0000459

Peleg, N., F. Blumensaat, P. Molnar, S. Fatichi, and P. Burlando (2017), Partitioning the impacts of spatial and climatological rainfall variability in urban drainage modeling, *Hydrology and Earth System Sciences*, 21(3), 1559. doi:10.5194/hess-21-1559-2017

Seo, B.-C., W. F. Krajewski, A. Kruger, P. Domaszczynski, J. A. Smith, and M. Steiner (2011), Radar-rainfall estimation algorithms of Hydro-NEXRAD, *Journal of Hydroinformatics*, 13(2), 277-291. doi:10.2166/hydro.2010.003

Smith, B., J. Smith, M. Baeck, and A. Miller (2015), Exploring storage and runoff generation processes for urban flooding through a physically based watershed model, *Water Resources Research*, 51(3), 1552-1569. doi:10.1002/2014WR016085

Smith, B. K., J. A. Smith, M. L. Baeck, G. Villarini, and D. B. Wright (2013), Spectrum of storm event hydrologic response in urban watersheds, *Water Resources Research*, 49(5), 2649-2663. doi:10.1002/wrcr.20223

Smith, J. A., and W. F. Krajewski (1991), Estimation of the mean field bias of radar rainfall estimates, *Journal of Applied Meteorology*, 30(4), 397-412.

Smith, J. A., M. L. Baeck, K. L. Meierdiercks, A. J. Miller, and W. F. Krajewski (2007), Radar rainfall estimation for flash flood forecasting in small urban watersheds, *Advances in Water Resources*, 30(10), 2087-2097. doi:10.1016/j.advwatres.2006.09.007

Wright, D. B., J. A. Smith, G. Villarini, and M. L. Baeck (2012), Hydroclimatology of flash flooding in Atlanta, *Water Resources Research*, 48(4). https://doi.org/10.1029/2011wr011371

Wright, D. B., J. A. Smith, G. Villarini, and M. L. Baeck (2014), Long-term high-resolution radar rainfall fields for urban hydrology, *Journal of the American Water Resources Association*, 50(3), 713-734. doi:10.1111/jawr.12139

Zhou, Z., J. A. Smith, D. B. Wright, M. L. Baeck, and S. Liu (2019), Storm catalog-based analysis of rainfall heterogeneity and frequency in a complex terrain, *Water Resources Research*, 55(3), 1871-1889. doi:10.1029/2018WR023567

Zhou, Z., J. A. Smith, L. Yang, M. L. Baeck, M. Chaney, M.-C. Ten Veldhuis, H. Deng, and S. Liu (2017), The complexities of urban flood response: Flood frequency analyses for the Charlotte Metropolitan Region, *Water Resources Research*, 53(8), 7401-7425. doi:10.1002/2016WR019997

Zhu, Z., D. B. Wright, and G. Yu (2018), The Impact of Rainfall Space-Time Structure in Flood Frequency Analysis, *Water Resources Research*, 54(11), 8983-8998. doi:10.1029/2018wr023550