**RESPONSE LETTER INSTRUCTIONS**

1. After receiving a review from the journal, prepare a Review Response letter based on the format provided below. Every question and comment should be turned into a bulleted list before answering them point by point.
2. Submit 3 things:
   1. Manuscript with all changes in track-change mode (name\_trackchange.docx)
   2. Manuscript with all changes after accepting all changes (name\_clean.docx)
   3. Response letter (reviewresponse.docx)

**Review Report**

**Manuscript Title**: On the Impact of Rainfall Spatial Variability, Geomorphology, and Climatology on Flash Flood Severity

**Authors: Manabendra Saharia, Pierre-Emmanuel Kirstetter, Humberto Vergara, Jonathan J.Gourley, Isabelle Emmanuel, Hervé Andrieu**

**Manuscript No.**: 2020WR029124

Dear Editor, thank you for giving us the opportunity to submit a revised draft of the manuscript. We appreciate the time and effort that you and the reviewers dedicated to providing feedback on our manuscript and are grateful for the insightful comments on and valuable improvements to our paper. We have incorporated most of the suggestions made by the reviewers. Those changes are highlighted within the manuscript. Please see below, in blue, for a point-by-point response to the reviewers’ comments and concerns. All page numbers refer to the revised manuscript file with tracked changes.

**Reviewers' Comments to the Authors**

**Review 1**

This paper uses a vast, continental-scale observational dataset to examines the impact of catchment and precipitation properties, among which precipitation spatial variability, on the "flashiness" (as introduced in Saharia et al., 2016) of flash flood events. The study is relevant to the readership or WRR, and the use of a large observational dataset adds considerable information over the literature addressing this issue which is typically based on hydrological modeling results. I think the study deserves publication. However, the adopted methodology is not clearly described, and the text needs improvement in presentation and terminology. Overall, I think this is a relevant contribution which suffers from sub-optimal presentation.   
Please see below here my specific comments.

Thank you for the positive feedback on our manuscript. We agree that the presentation of the work can be improved. Please see our point-by-point responses below.

**Major comments**

1. Title/terminology

The paper stands on the equivalence "flash flood severity" = "flashiness". The typical definitions of severity are generally based on the flash flood peak discharge and, in some cases, on some measure of its rarity (e.g. the yearly exceedance probability or the recurrence interval). Conversely, the definition of "flashiness" includes both peak discharge and time-to-peak. This might cause quite some misunderstanding in the readers of the paper, both in terms of wrong expectances and of difficulty in comparing results with previous studies. This is even more true since some of the analyses specifically focus on static catchment attributes, with clear intent of understanding time to peak over peak discharge recurrence interval.   
I think the word "flashiness" should replace severity at least in the title. This difference should then be highlighted more clearly from the beginning of the text (now it appears at line 168), and the differences with respect to previous studies in which only peak discharge was examined should be made clearer. At the same time, I think the paper addresses much more than "rainfall spatial variability", as one can see from the bullet-points in the conclusions. In the end there is almost no conclusion about rainfall! Perhaps, the paper would benefit from a further update to the title/abstract to better deliver this.

Thank for this comment. It is true that typically flash flood severity is defined in terms of flash flood peak discharge or recurrence interval. However, such frequency of flood peak-based approaches often fail to identify regions that may not flood on a frequent basis, but when they do flood, it can be catastrophic. Moreover, the peak is only one aspect of a flood, and the severity of a flood should depend both on the timing and peak of a flood. To address this, we defined flood severity in terms of flashiness to characterize a hydrologic response to input rainfall. The flashiness metric gives the rate of rise of the hydrograph during flooding conditions and thus captures both the magnitude and timing aspects, with higher values corresponding to more severe floods (Saharia et al. 2017). Especially, it identifies basins responses with a large-magnitude discharge in a short period of time. It promotes a more process-based interpretation of flash flood severity in the community. This is a significant departure from widely used frequentist approaches that use the annual likelihood of flash flooding. Thus, we would like to keep that part of the title.

To avoid any misunderstanding, flashiness is introduced in the abstract in L32-36: “this study overcomes limitations of prior works based on limited case studies or simulations to characterize the catchment response in terms of flashiness. The objective is to develop a robust understanding of how rainfall spatial variability influences flash flood severity and to assess its contribution relative to basin physiography and climatology”.

This difference is now clearly highlighted in the introduction section in L129-135: “Flashiness is introduced in Saharia et al. (2017) to represent the severity of floods. It gives the rate of rise of the hydrograph during flooding conditions and thus captures both the magnitude and timing aspects with higher values corresponding to more severe floods (Eq. 2 in Saharia et al. 2017). Especially, it identifies basins responses with a large-magnitude discharge in a short period of time. Flashiness provides a process-based interpretation of flash flood severity that differs from classical frequentist approaches based on peak discharge”.

The comment about the paper representing more than rainfall spatial variability is well-taken. The title of the paper has been changed to “On the Impact of Rainfall Spatial Variability, Geomorphology, and Climatology on Flash Flood Severity” to better reflect this. The abstract of the paper has been updated. We have also expanded the conclusion section with the following additions in L.593-L605:

“3. The contribution of rainfall spatial variability on flashiness was found on par with geomorphology and climatology. There is an inflection point in the relationship between the mean of the accumulated precipitation and flashiness. This is due to the physiography of the basin dampening the effect of lower rainfall values, while higher rainfall overwhelms other factors and primarily contributes to flashiness.

4. The partial term plots reveal that increased flashiness is associated with higher basin slope, higher first-order channel frequency, and higher occurrence of high rainfall values (i.e., high relative standard deviation of the accumulated precipitation). Dispersed precipitation with respect to the flow paths and lower relative standard deviation of the rainfall weighted flow distance are associated with lower flashiness values. These findings, based on unprecedented spatial and temporal representativeness, highlight the importance of accounting for rainfall spatial variability in hydrologic modeling”

Reference:

Saharia, M., Kirstetter, P.-E., Vergara, H., Gourley, J. J., Hong, Y., & Giroud, M. (2017). Mapping Flash Flood Severity in the United States. Journal of Hydrometeorology, 18(2), 397–411. <https://doi.org/10.1175/JHM-D-16-0082.1>

1. Introduction

Overall, the introduction needs to be deeply reorganized/rewritten to solve the issue highlighted at point 1, to avoid repetition of concepts, and to be more focused on the important points. More attention should be devoted to the novel aspects introduced by this study, namely the use of observational over modeling dataset, the focus on flashiness, the inclusion in the analysis of static catchment properties, etc. Some of these aspects are already mentioned in the intro, but in a scattered way, without guiding the reader to the point.

Thank you for this comment. The entire introduction section has been reworked for clarity and to highlight the novel aspects of the study – observational over modeling, flashiness, inclusion of geophysical and climatological parameters etc. The discussion on rainfall spatial variability has also been reorganized to provide greater clarity.

1. Methods
2. how are the flash floods selected? At line 221 there is a mention to 45,000 km2 basins, which seems quite large for the typical definitions of flash floods

The NWS definition of flash floods is based on a flow of water above a pre-determined flood level and beginning six hours of the causative event. A few studies over Europe such as Gaume et al. (2009) and Marchi et al. (2010) have used basin thresholds of 500 km2 and 1000 km2 respectively. However, the definition of flash flooding through these basin scale thresholds is limited for several situations, e.g., as that the effective basin area can be quite small for a localized convective storm near the basin outlet, which can produce a rapid response for a relatively large catchment. This paper builds on the definition of flood severity proposed in Saharia et al. (2017), which is based on the flood response as the difference between the peak discharge and action stage discharge normalized by the flooding rise time and basin area, as visualized in the figure below. An empirical cumulative distribution function (ecdf) was then used to scale the values between 0 and 1.

Diagram

Description automatically generated

Figure 1: Graphical representation of event-level flashiness. Reproduced from Saharia et al. (2017)

References:

Gaume, E., and Coauthors, 2009: A compilation of data on European flash floods. J. Hydrol., 367, 70–78, doi:10.1016/ j.jhydrol.2008.12.028.

Marchi, L., M. Borga, E. Preciso, and E. Gaume, 2010: Characterisation of selected extreme flash floods in Europe and implications for flood risk management. J. Hydrol., 394, 118–133, doi:10.1016/ j.jhydrol.2010.07.017.

Hydrologic Services Program, 2012: National Weather Service manual 10-950. National Weather Service, 5 pp. [Available online at <http://www.nws.noaa.gov/directives/sym/pd01009050curr.pdf>.]

Saharia, M., Kirstetter, P.-E., Vergara, H., Gourley, J. J., Hong, Y., & Giroud, M. (2017). Mapping Flash Flood Severity in the United States. Journal of Hydrometeorology, 18(2), 397–411. <https://doi.org/10.1175/JHM-D-16-0082.1>

1. it is not clear if lines 223-244 are methods done in this study or already in the united flash flood database;

Thank you for the comment. The method of augmenting the dataset with landscape properties has been specifically done for this study. It is not a part of the original Unified Flash Flood Database which can be found here: <https://inside.nssl.noaa.gov/flash/database/>. A sentence has been added that provides a clarification.

L.209-210: “The geomorphologic parameters for delineated catchments were extracted from these grids using custom libraries.”

1. same lines: it is not clear to me how and why the DEM (or the flow direction?) is downscaled to 1-km, please try to be more precise in the description.

Thank you for the comment. The DEMs were simply resampled (averaged) to our desired resolution, after which the flow direction and flow accumulation were derived through a series of iterative quality control measures to ensure the streams follow the correct flow paths. The DEM was kept at 1-km in order to match the resolution of the MRMS precipitation forcing dataset. We kept everything at a consistent resolution for ease of computation and analysis.

# L.204-209 edit: “The National Hydrography Dataset (NHD; http://nhd.usgs.gov/) was used to resample (averaging) the 30-m DEM to a 1-km grid to ensure compatibility between DEM-based flow accumulations and the actual river network across the Continental United States (CONUS). The DEM was kept at 1-km in order to match the resolution of the MRMS precipitation forcing dataset. Everything was kept at a consistent resolution for ease of computation and analysis.”

1. it is not clear to me how a quantity defined for individual floods ("flashiness") could provide generic basin information ("likelihood of a basin to produce...", line 293). This is quite a crucial concept in this study, and needs to be better described.

Thank you for this comment that improves the clarity of the manuscript. The “flashiness” variable is introduced in Saharia. et al. (2017) as a measure of flood severity. It is defined as the difference between the peak discharge and action stage discharge normalized by the flooding rise time and basin area as given (see Eq. 1 in Saharia. et al. 2017). The flashiness information is available at two levels: event and basin. The basin flashiness (i.e. the median flashiness value computed from all flooding events observed at a given basin) represents the climatological flash flood severity of the basin, and it is used to describe the likelihood of a basin to produce a flashy response in Saharia et al. (2017). Further details are available in Section 3 of Saharia et al. (2017).

The event-level flashiness, given by Eq. (2) in Saharia et al. (2017), is used in this study to characterize the hydrologic response of a catchment to rainfall forcing in each event. It gives the rate of rise of the hydrograph during flooding conditions and thus captures both the magnitude and timing aspects with higher values corresponding to more severe floods. Especially, it identifies basins responses with a large-magnitude discharge in a short period of time.

It is now clearly stated in the introduction section in L129-L135: “Flashiness is introduced in Saharia et al. (2017) to represent the severity of floods. It gives the rate of rise of the hydrograph during flooding conditions and thus captures both the magnitude and timing aspects with higher values corresponding to more severe floods (Eq. 2 in Saharia et al. 2017). Especially, it identifies basins responses with a large-magnitude discharge in a short period of time. Flashiness provides a process-based interpretation of flash flood severity that differs from classical frequentist approaches based on peak discharge”.

The sentence was also reformulated for more clarity in L.264-269: “The metric is scaled between 0 and 1 with higher values corresponding to more severe floods (Eq. 2 in Saharia et al., 2017). The relative amounts and timings at the outlet depend on the spatial distribution of rainfall forcing, runoff, and transfer in channels. The event-based flashiness characterizes the basin response for individual floods, especially in terms of to producing a large-magnitude discharge in a short period of time in response to a heavy rainfall”.

Reference

Saharia, M., Kirstetter, P.-E., Vergara, H., Gourley, J. J., Hong, Y., & Giroud, M. (2017). Mapping Flash Flood Severity in the United States. Journal of Hydrometeorology, 18(2), 397–411. <https://doi.org/10.1175/JHM-D-16-0082.1>

1. equations for the Delta\_1, Delta\_2, HG, VG would be helpful to complement the eqs now at page 15;

Thank you for the comment. The equations for have now been included added in L.324 and 326. Equations for the Horizontal and Vertical Gaps are also added. These metrics are based on distributions, the definitions of which have been included:

L.337-360 addition:“Further, two more indices proposed by Emmanuel et al. (2015) based on the width function (Kirby, 1976; Rigon et al., 2016), denoted *w*, and on the rainfall weighted width function, denoted *wp*, were also computed for all flooding events. The rainfall weighted width function is defined by:

Where d is the distance to the outlet along the flow line, is the rainfall averaged over all pixels at distance d, and the average rainfall over the basin. Here, *w* and *wp* can be regarded as probability density functions whose cumulative distributions functions are denoted by *W* and *Wp*, respectively. The first index, Vertical Gap (VG), is defined as the absolute value of the maximum vertical difference between *W* and *Wp*. This criterion is calculated by the Kolmogorov-Smirnov test which compares two cumulative distributions. VG values close to zero indicate a rainfall distribution with low spatial variability, while higher values of VG indicate greater concentration of the rainfall over a small part of the catchment. The value of VG is obtained for a quantile value *Wp*(*dp,VG*) and a flow distance *dp,VG* of *Wp*. The distance *dVG* associated with the same quantile value is different for *W*. The second index, Horizontal Gap (HG), is defined as the absolute difference of these two distances normalized by the maximum flow distance. It can be understood as a measure of the deviation between the rainfall weighted flow distances and the flow distances, expressed in distance units. It represents a spatially homogeneous rainfall or concentrated close to the catchment centroid position for values close to zero, while values lower (or greater) than 0 indicate a rainfall distribution downstream (or upstream).

1. more details about HG VG should be provided, from what one can get from the text they seem alternatives to Delta\_1 Delta\_2; do they provide additional/complementary info? Not clear

Please see the response to the previous question. The details of HG/VG have now been added in L.337-360and they provide additional insights into the rainfall structure. (Delta\_1, Delta\_2) and (VG, HG) both compare the width function of the watershed to the rainfall width function which summarizes the rain pattern on the watershed. (Delta\_1, Delta\_2) compares the two first moments of these pdfs while (VG, HG) are based on the criterion used by the classical Kolmogorof-Smirnov test to compare the distributions. They provide additional insights into the rainfall structure.

1. the fact that the precipitation-related quantities are computed up to the time of the peak (now line 358) should come earlier in the section. Alternatively, less misleading details should be given earlier (e.g. line 329 "accumulated rainfall field" makes one think about total event accumulation)

Thank you for this comment. The section of rainfall accumulation has now been moved as the second paragraph of the section for clarity (now, L.299-306). With these changes, the reader should now have greater clarity on accumulated rainfall field.

1. Section 5 was very difficult to follow. I suggest trying to reorganize it in a clearer way

Thank you for this comment. Section 5 text has been reworded to provide greater clarity.

1. Figure 4: what is "relative standard deviation"? Is it some kind of coefficient of variation?

The relative standard deviation (RSD) is equivalent to the coefficient of variation. It is often more convenient than the standard deviation as it is expressed in percentage. It is obtained by multiplying the standard deviation by 100 and dividing it by the mean.

RSD = (100 \* S)/Xbar, where S is the standard deviation and Xbar is the mean.

1. Figures 4, 5: how are these figures supposed to handle the different density in the x-axis?

Thank you for the comment. Depending on the scale of the variable, linear or logarithmic scales are used in the x-axis.

1. The GAMLSS model.
2. I probably missed something, but I cannot understand why a model specifically focusing on location, scale, shape (three parameters, three degrees of freedom) is used to compute only two (mean, standard deviation - as said in lines 509-517). Overall, it seems to me there is a lot of talking about the advantages of including the shape and I was quite disappointed to see that in the end only two parameters were used. Is the complexity of GAMLSS required when only 2 parameters are required? Could a simpler method be enough/more accurate?

It is true that the GAMLSS modelling framework allows for the fit of distributions with various number of parameters. For variables that take on values in the range ]0, 1[ like flashiness (see section 2.1), the most appropriate option is the beta distribution (<http://www.gamlss.com/wp-content/uploads/2013/01/gamlss-manual.pdf>). We agree that more distribution parameters would provide additional insight on the relation between flashiness and the predictors (precipitation variability, geomorphology, and climatology). However, additional parameters also introduce more degrees of freedom that can lead to overfitting at the expense of model interpretability. We believe that a parsimonious model is required to extract robust relations between the predictand (i.e., flashiness), and the predictors (see section 6.2). One distribution parameter is theoretically sufficient to extract these relations. Yet in case of heteroscedasticity or conditional asymmetry in the data distribution (see Fig. 4 and 5), it is appropriate to model the spread of the distribution (scale) in addition to the location. Therefore, we believe that we strike a balance between fitting performance and model interpretability with using 2 parameters.

Simpler methods like the Generalized Additive Models (Hastie & Tibshirani, 1990) or Generalized Linear Models (McCullagh et al., 1989) don't provide the same flexibility in terms of distribution or link functions.

L.473-476 addition: “GAMLSS is an extension over the traditional frameworks and offers higher flexibility as the response variable, which can be continuous, discrete, or mixed, can follow a general distribution function instead of being restricted to the exponential family”.

1. which distributions were tested in order to choose Beta? Why?

Other distributions were tested, that can fit variables taking on values in the range ]0, 1[: normal, lognormal, gamma, Gumbel, reverse gaussian, reverse Gumbel, etc. The goodness-of-fit was checked with the Akaike information criteria (AIC), the Gaussianity and independence of residuals with their first four moments, Filliben correlation coefficient, and quantile–quantile plots (Stasinopoulos and Rigby, 2007). The best fitting results were obtained with the beta distribution.

L.494-498 addition: “Several distributions were tested (e.g., normal, lognormal, Gumbel) and their goodness-of-fit was checked following Stasinopoulos and Rigby (2007). The beta distribution was identified as the most appropriate distribution for modeling the dependence of flashiness on various geomorphological, climatological, and rainfall variables by using the AIC and by checking the normality and independence of residuals”.

1. do you have any physical interpretation for the choice of Beta? Do you envision situations in which a different distribution is needed? Would the results be different?

We don’t have any physical interpretation for the choice of Beta. As the flashiness is defined as a scaled variable in this study, the beta distribution is appropriate. In a situation where the flashiness should be unscaled, e.g. for a more direct physical interpretation of the flashiness values in a forecasting context, a different distribution appropriate for variables taking on values in the range ]0, +∞[ would be required. Of course, the influence of individual predictors would be equivalent, and we expect the conclusions drawn from the analysis in section 6.2 to be identical.

1. I suggest providing more details on the use of penalized splines. Why is this operation needed?

A GAMLSS model involves link functions that relate predictors to the distribution parameters to model conditional distributions. Each parameter (μ, σ) of the beta distribution is a function of the predictors that describe precipitation variability, geomorphology, and climatology using penalized splines. Penalized splines are used to fit the non-linear relations between each predictor and the beta distribution parameters (μ, σ).

L.506-508 addition: “To relate explanatory variables and the beta distribution parameters, penalized splines are used as link functions for fitting trends for each parameter as they offer more flexibility in modeling complex nonlinear relationships”.

1. what is the "systematic part" (line 530)? Is it the parameter \mu?

Thanks for the comment. The systematic part is the parameter μ that represents the conditional expectation, conditioned on the predictors (precipitation variability, geomorphology, and climatology). It is now indicated as follows in L.518-519 : the systematic part (i.e., parameter μ of the beta distribution).

**Minor comments**

* line 57-59 need to be substantiated with references

It has now been edited to substantiate with an extensive literature review. The introduction has also been reworked substantially.

L.67-79 edits: “The extent to which spatial heterogeneity of rainfall impacts catchment response and its influence in comparison to basin physiography and climatology remains an open research topic. The literature has not yielded a consensus, which has implications for understanding flood processes and adequately representing them in distributed hydrologic models for improved forecasting of floods. While studies typically based on hydrological modeling results have concluded that the spatial variability of rainfall exerts a significant impact on the hydrograph (Anquetin et al., 2010; Kim et al., 2008; Looper & Vieux, 2012; Mei et al., 2014; Sangati et al., 2009; J. A. Smith et al., 2007; Vieux et al., 2009; Zoccatelli et al., 2010), others have indicated limited influence of rainfall spatial variability (Adams et al., 2012; Brath et al., 2004; Cole & Moore, 2008; Nicótina et al., 2008; M. B. Smith et al., 2004). Pokhrel and Gupta (2011) have reported that the influence of rainfall spatial variability on the hydrologic response can be greatly diminished by the damping effect of routing while Lobligeois et al. (2014) concluded that there is a regional dependence.”

* line 71: how is storm size defined?

We don’t measure storm size. Instead, we rely on the variables describing rainfall spatial variability to represent the storm.

* lines 133 and following: this concept was already presented earlier, I suggest reorganizing

Thank you for this comment. The sentence regarding Smith et al. (2004) has been deleted and merged with the previous section where the concept has already been discussed. The two limitations in previous studies of a) limited sample size, and b) observed v/s modeled are now addressed in 2 separate paragraphs, that we hope lead to greater clarity.

* line 150: there are more up-to-date references on this topic

Thank you for this comment. Considering the revised introduction section, these references are no longer necessary.

* Section 2.2: it would be helpful to have some quantitative information on the quality of MRMS

Thank you for this comment. There are quite a few studies that attest to the quality of MRMS. According to Zhang et al. (2020), the change to the synthetic algorithm that uses an attenuation-based estimator in rain has yielded an increase in the mean bias ratio in heavy rain from 0.81 to 0.95. You can find more results and details here: Zhang, J., L. Tang, S. Cocks, P. Zhang, A. Ryzhkov, K. Howard, C. Langston, and B. Kaney, 2020: A dual-polarization radar synthetic QPE for operations. J. Hydrometeor., 21, 2507–2521, <https://doi.org/10.1175/JHM-D-19-0194.1>.

L.238-239 addition: “Recent advances in MRMS precipitation estimation using multi-sensor approaches and attenuation-based estimators in rain has yielded improved accuracy (Zhang et al. 2020).”

* Lines 353-355: this seems like introduction

Thank you for this comment. This line has now been edited and moved to the introduction section.

* Figures: it is not clear why some figures show precipitation in mm and some in m

Thank you for this comment. The plots are now all in m.

* Acknowledgments: now it is not clear if the used precipitation data is MRMS or Nexrad (cannot access the link)

The link for MRMS data has now been updated to : http://edc.occ-data.org/nexrad/mosaic/

**Review 2**

This is a great study providing new insights on the relative significance of precipitation spatial variability to the modeling of flooding. I only have a few suggestions that aim at improving the discussion and presentation of results.

* Introduction - There is a study in US that tested the precipitation moments and flow distances based on a long record of data, which worthies mentioning and comparing results to their findings.   
  Mei, Y., E. N. Anagnostou, D. Stampoulis, E. I. Nikolopoulos, M. Borga, H. J. Vegara, Rainfall organization control on the flood response of mild-slope basins, 2014: Journal of Hydrology, Volume 510, 14 March 2014, Pages 565-577

Thank you for this comment. This paper that has now been referenced in L.73.

* Line 287 - is there a typo here? Saharia et al. {Citation}

Thank you for the comment. The typo has now been corrected.

* Section 3 - Discussion connecting the various spatial variability indices to the mechanistic processes that define flashiness.

Thank you for the comment. We discuss this aspect extensively in the following sections.

* Line 369 - is this correct (should this be precipitation event)? "start of the flooding event (t1)"

Thank you for this comment. This a typo and it has now been corrected to “start of the rainfall event (t1)”.

* Case study - some discussion on the case study is needed. How are these parameters relate to the flashiness of the basin. Overall, up to this point, the readers would benefit from a better connection/transition from the description of the spatial indices to the interpretation in terms of basin's flashiness and the presented case study. Maybe the flashiness section should go before the case study?

Thank you for this comment. The introduction section has been significantly reworked to provide more context as to why flashiness is related to rainfall spatial variability, climatology, and geomorphology indices. Section 5 of the paper has also been edited to provide more interpretation. The explanation of the flashiness/flash flood severity (Section 2.1) precedes the spatial variability indices (2.2), which is followed by the case study to illustrate both. This should now provide a logical flow to the reader.

* 6.2 this is an important section. The reader can benefit from more details and mathematical formulations describing how the partial predictions of flashiness were determined for the different variables.

Thank you for this comment.

L.526-530 addition: “GAMLSS is an additive model and allows to analyze the influence of individual predictors on the response variable (i.e., flashiness). A partial term plot is a powerful diagnostic tool for disaggregating competing influences. It shows the marginal effect one predictor has on the predicted outcome by averaging all other predictors and plotting the response as a function of the predictor of interest.”

**Review 3**

The authors examine the dependence of flash flood severity on the spatial variability of rainfall. Analyses build on the interesting and important prior work developing a "Flash Flood Database", a flash flood severity index and the MRMS rainfall data set. Analyses of spatial variability of rainfall focus on the rainfall weighted flow distance and normalized dispersion metrics introduced in Smith et al. 2002 and 2005a (see also Zoccatelli et al. 2011). The goal of the paper - developing a "robust understanding of how rainfall spatial variability influences flash flood severity... relative to basin physiography" is important and the "large-sample" approach is valuable and likely to provide interesting insights on the problem. The biggest weakness of the paper, in its current form, is that conclusions drawn from the analyses do not match the lofty goals of the study. The first item in the listing of "findings" notes that a "Large number of variables were used to model event flashiness". The second item points to the "systematic part of the ... model ... yielded a correlation of 0.84...meaning there is adequate skill ... to explain dependencies". The final finding gets more at the crux of the problem - "contribution of rainfall spatial variability on flashiness was found on par with geomorphology and climatology", but it sheds little new light on the problem. The authors need to dig in more forcefully to develop compelling conclusions concerning rainfall spatial variability and flash flooding. The authors note in the Introduction, that the "literature review of this topic reveals that our understanding of rainfall spatial variability on flooding under a wide variety of rainfall, physiographic and antecedent conditions remains limited". There has, however, been substantial research on the topic that provides useful directions for sharpening the analyses developed by the reviewer. I've focused below on research by my colleagues, but also point to other groups that have addressed aspects of the problems pursued by the authors.

We thank the reviewer for his comments. We agree on the need on our part to provide a better exposition of our results and re-organization in certain parts of our paper. The conclusions section has been expanded.   
  
**Comments:**

1. Literature Review – Rainfall-weighted flow distance metrics are used in Smith et al. 2002 to demonstrate the role of spatial (and temporal) variability of rainfall in controlling flood peak magnitudes and response times. Drainage network structure and antecedent soil moisture are also important elements considered in the study (for additional developments, see Cristiano et al. 2019, ten Veldhuis et al. 2017, Zhou et al. 2017 and Yang et al. 2017, along with references and citations). Normalized dispersion and rainfall weight flow distance metrics are used in Smith et al. 2005a to explain how storms with sharply contrasting spatial rainfall distributions can have strikingly similar flood response, i.e. why GIUH-like procedures often work well for flood modeling.

Woods and Sivapalan (1999) proposed an analytical method to express the variability of catchment-averaged storm runoff rates in terms of space and time variability of hydrological inputs. They represent the spatial variability of rainfall excess at the scale of a rain event by introducing the average rainfall excess with flow distance.

In order to analyze the role of spatial rainfall variability in the analysis of flood formation, Smith et al. (2002) consider the rainfall-weighted centroid distance to the basin outlet that varies with time, who’s mean value characterizes the rainfall pattern at the event scale Smith et al (2005).

L. 107-112 addition: “Woods and Sivapalan (1999) proposed an analytical method to express the variability of catchment-averaged storm runoff rates in terms of space and time variability of hydrological inputs. Smith et al. (2002, 2005) considered the rainfall-weighted centroid distance to the basin outlet that varies with time, whose mean value characterizes the rainfall pattern at the event scale. They represent the spatial variability of rainfall excess at the scale of a rain event by introducing the average rainfall excess with flow distance”.

1. The authors also specifically note that "we seek to investigate the severity of a flood in this paper, a hitherto unexplored aspect of the hydrograph in the context of rainfall spatial variability". This is in fact the central theme of Smith et al. 1996 and the subsequent studies in Giannoni et al. and Landel et al. The details of spatial variability of rainfall are central to the magnitude of one of the most extreme floods in the eastern US, the 27 June 1995 Rapidan flood. Rainfall analyses in Smith et al. 1996 provided one of the first demonstrations that the WSR-88D could provide useful, high-resolution information on the most extreme of floods. Spatial variability of rainfall and extreme floods are also central themes of Smith et al. 2018 and 2019. The 2019 paper provides end-member representations of rainfall spatial variability and flood response.   
   The examples above are taken from the material with which I'm most familiar. There are a number of other groups that are tackling similar problems (I'm most familiar with efforts in the US, EU, UK, Israel and China). A closer look at previous work would help to sharpen the analyses presented in the present study.

Thank for this comment. We discuss previous contributions on the impact of rainfall variability on floods in L.67-L115.

Flash flood severity is usually defined in terms of flash flood peak discharge or recurrence interval. We defined flashiness or flash flood severity in Saharia et al. (2017) as the potential for a flashy response to input rainfall. This definition includes both the peak and timing of a flood, and it is a significant departure from widely used frequentist approaches. It identifies hydrological responses with a large-magnitude discharge in a short period of time, and it is not dependent on annual likelihood of flash flooding. While rainfall spatial variability has been studied before, this has never been done in the context of flashiness. The Smith el al. (1996) paper discusses storm motion and evolution and how it is guided by spatial characteristics of the watershed.

References:

Saharia, M., Kirstetter, P.-E., Vergara, H., Gourley, J. J., Hong, Y., & Giroud, M. (2017). Mapping Flash Flood Severity in the United States. Journal of Hydrometeorology, 18(2), 397–411. https://doi.org/10.1175/JHM-D-16-0082.1

1. Urban Flash Floods –

There is no specific treatment or discussion of urban flash flooding despite the fact that loss of life and property damage concentrate around urban regions. In examining basin properties that control flash flooding, omission of impervious cover and other variables characterizing urban surfaces and drainage systems is a serious omission. The USGS Gages II data base includes these basin characteristics, along with others that could enhance the land surface response side of analyses. Directly addressing issues that are unique to urban flash flooding would improve the paper.

Thank you for this comment. Our study is limited by the availability of USGS gauges and corresponding NWS flood stage cutoffs, which form the basis of the flood database used in this study. Further, to separate out the relationship between rainfall spatial variability and watershed characteristics, we have concentrated only on unregulated basins. A study on regulated basins requires considering anthropogenic factors that are out of the scope of the analysis. We acknowledge the potential usefulness of the Gages II dataset in augmenting our understanding. We will consider this for a future study.

Our objective has also been to treat the severity of flooding as a continuum using “Flashiness” as a metric. We acknowledge the importance of including as many useful variables as possible to characterize the land surface response. Table 2 in the paper provides a list of the final variables that were included in the model. Note that imperviousness is a variable included in our dataset but has not been identified as important as other variables. We wrote custom scripts to calculate a large number of geomorphic variables from a variety of sources to do this, which has been published as a tool for wider usage [1].

References

[1] Shen, X., Vergara, H. J., Nikolopoulos, E. I., Anagnostou, E. N., Hong, Y., Hao, Z., et al. (2016). GDBC: A tool for generating global-scale distributed basin morphometry. Environmental Modelling & Software, 83, 212–223. https://doi.org/10.1016/j.envsoft.2016.05.012

1. Flash Flood Severity –

I don't have a good sense for how well the flash flood severity index corresponds with other measures of severity. The August 27, 2006 event in the Blue River (Figure 2) seems like a modest flood with a large flash flood severity index. It is an annual peak in the Blue River, but on the small end. The NCEI storm events data base includes one flash flood episode in Kansas City, but not in this watershed. The big one is August 22, 2017 - peak is almost 7 times larger than 2006 peak and from the form of the hydrograph it should max out flash flood severity. It could make for a better illustration of methods and provide "case study" insights to rainfall variability and flash flooding. Have you examined how often "complex" hydrographs create odd values of rise time and flash flood severity? During my time at NWS, I was tangentially involved in the world of determining "flood stages" and bankfull stages. My impression was that local considerations (often a particular damage location) drove the determination of flood stages and that bankfull determinations were not carried out in the manner that a USGS geomorphologist would follow. Prior papers have documented some of these issues, but it would be useful to provide the reader with your best assessment of what the most serious problems are with the "action" stage determinations.

Thank you for this comment. While severity of a flood is often associated with its peak value, this paper deals with flashiness, which combines both the peak and timing aspect of floods.

L. 129-135: “Flashiness is introduced in Saharia et al. (2017) to represent the severity of floods. It gives the rate of rise of the hydrograph during flooding conditions and thus captures both the magnitude and timing aspects with higher values corresponding to more severe floods (Eq. 2 in Saharia et al. 2017). Especially, it identifies basins responses with a large-magnitude discharge in a short period of time. Flashiness provides a process-based interpretation of flash flood severity that differs from classical frequentist approaches based on peak discharge”.

The case study included in this paper has a flashiness value of 0.85, which identifies it as a flash flood. In our previous study, we had identified a flashiness cutoff of 0.75 for differentiating floods and flash floods by comparing our values with National Weather Service (NWS) Flash Flood Reports (Figure reproduced below from Saharia et al., 2017). As such, we believe this is a relevant case study in the context of our paper.

Chart, line chart

Description automatically generated

Figure 2: PDF of flashiness for flash flood (black) and flood (grey). Reproduced from Saharia et al. (2017)

The unified flash flood database is an even-based database that informs the timing and peak discharge of floods. As stated in the manuscript in L.136-138, this study explores the first-order dependencies of flashiness on event-level rainfall spatial variability. Complex hydrographs are out of the scope of the analysis and can be specifically analyzed in a future study.

We are reliant on NWS for the action stage values which are used for a wide variety of operational work in US. While we are aware of uncertainties with regards to USGS flood stage, this discussion is beyond our scope.

Reference:

Saharia, M., Kirstetter, P.-E., Vergara, H., Gourley, J. J., Hong, Y., & Giroud, M. (2017). Mapping Flash Flood Severity in the United States. Journal of Hydrometeorology, 18(2), 397–411. https://doi.org/10.1175/JHM-D-16-0082.1  
  
**Minor Points -**

1. Page 3. First paragraph of Introduction extends for more than a page. Kicking off the paper with a more compact and compelling formulation of core ideas would enhance the paper.

Thank you for this comment. The first paragraph and the introduction have been significantly reworded and made more concise to provide a compact understanding and logical flow.

1. Line 131. "The study tested the spatial rainfall variability indices described in Zoccatelli et al. (2010)". Here and throughout, note that the rainfall weighted flow distance and dispersion metrics were introduced in Smith et al. 2002 and 2005a.

Thank you for this comment.

L107-L112 addition: “Woods and Sivapalan (1999) proposed an analytical method to express the variability of catchment-averaged storm runoff rates in terms of space and time variability of hydrological inputs. Smith et al. (2002, 2005) considered the rainfall-weighted centroid distance to the basin outlet that varies with time, whose mean value characterizes the rainfall pattern at the event scale. They represent the spatial variability of rainfall excess at the scale of a rain event by introducing the average rainfall excess with flow distance.”

1. Lines 220 - 221 "The maximum basin area in this study is approximately 45,000 km2 ". What is the rationale for such a large basin in a flash flood study.

Thank you for this comment. The NWS definition of flash floods is based on a flow of water above a pre-determined flood level and beginning six hours of the causative event. A few studies over Europe such as Gaume et al. (2009) and Marchi et al. (2010) have used basin thresholds of 500 km2 and 1000 km2 respectively. However, the definition of flash flooding through these basin scale thresholds is limited for several situations, e.g., as that the effective basin area can be quite small for a localized convective storm near the basin outlet, which can produce a rapid response for a relatively large catchment. This paper builds on the definition of flood severity proposed in Saharia et al. (2017), which is based on the flood response as the difference between the peak discharge and action stage discharge normalized by the flooding rise time and basin area.

1. Lines 255 - 256. "The MRMS project, initiated by the NOAA NSSL, has revolutionized the way precipitation is estimated by producing a seamless high-resolution dataset that updates in the order of five minutes without human intervention". Useful and I'm a fan, but "revolutionized" is too strong.

Thank you for this comment. The word “revolutionized” has been replaced with “significantly advanced” (L. 229).

1. Lines 296 - 299. "The database was subjected to extensive post-processing based on radar beam height and snow percent of total precipitation in a basin to reduce input uncertainties in modeling results. Firstly, all events that fall in basins with mean radar beam height of greater than 2 kilometers above the ground level were discarded." Is there a specific rationale for the 2 km choice?

Thank you for this comment. The beam height threshold is used to reduce uncertainties with QPEs. In general, the higher the radar beam is in the cloud, the greater the uncertainties are in estimating surface precipitation. We focus the dataset to regions with trustworthy data, so we have used a beam height threshold to minimize uncertainties (e.g., Gourley et al. 2017). With using 2 kms it is expected to robustly evaluate the impact of rainfall spatial variability and minimize issues with radar rainfall forcing.

Reference

Gourley, J. J., Z. Flamig, H. Vergara, P. Kirstetter, R. Clark III, E. Argyle, A. Arthur, S. Martinaitis, G. Terti, J. Erlingis, Y. Hong, and K. Howard, 2017: The Flooded Locations And Simulated Hydrographs (FLASH) project: improving the tools for flash flood monitoring and prediction across the United States, *Bull. Amer. Meteor. Soc.*, **98**, 361-372. <http://dx.doi.org/10.1175/BAMS-D-15-00247.1>

1. Figure 2. Wrong caption - Leaf River...

Thank you for this comment. It has now been corrected.