We appreciate Referee #2's careful review and insightful suggestions. The followings are our response to comments and revision.

Using a dynamic hydrologic model with varying parameters with the recovery of a burned watershed in California and radar rainfall estimates, the authors investigated changes in rainfall intensity-duration thresholds inducing flash floods in the watershed as it recovers from fire. The study is of practical value for informing post-fired flash floods. My comments are as follows.

1. The title does not accurately summarize what has been done in the study as there is little information about how the spatiotemporal distributions of rainfall affect the intensity-duration thresholds inducing flash floods or their changes with recovery.

R: We changed the title to "Temporal changes in rainfall intensity-duration thresholds for post-wildfire flash floods in Southern California." The spatial analysis was an initial intent of the study, but was not as interesting as expected and therefore not a focus of this paper.

2. Given the small sample of rainfall events (34 in total), should sampling uncertainty (thus robustness) of the estimated intensity-duration thresholds be evaluated. One simple method might be to bootstrap these events or values of I(D, j), with which to obtain a set of plausible estimates of intensity-duration thresholds and evaluate robustness of the reported results.

R: Good suggestion! We bootstrapped original samples obtaining 50 replications. With the resampled events, we estimated 95% confidence intervals for rainfall intensity-duration thresholds (Fig.3-4 and Fig 6-7).

3. There are several prescribed quantities, such as a value of 15mm/h for extracting moderate-to-high rainfall events and a value of 2m^3/s for defining effective discharge. I did not see discussions about the rationale of these choices, nor their potential influences on the identified intensity-duration thresholds inducing flash floods.

R: The 15 mm/h threshold was chosen so that we could extract storms that have the potential to generate floods. This threshold provided a balance between having too many rainfall events to process and having a sufficient number of storms to perform the model and ID threshold analyses. We explained this value in Section 3.1 as follows,

"This threshold generally corresponds with a 1-year average recurrence interval storm event in the study area (NOAA Atlas 14). This value falls between the California-Nevada River Forecast Center's flash flood guidance for unburned areas in the region (~22-25 mm/h; CNRFC 2021) and regional thresholds for post-wildfire debris flows in this region at a point (12.7 mm/h, Cannon et al. 2008; Staley et al. 2013). This threshold allows us to focus on storms that have a high potential to generate floods, while keeping the number of storms to a manageable level for data processing."

Low flow is not our focus in this study. Flow depths associated with Q of less than 2 m^3/s are very small and any potential impacts from such flow would be negligible.

4. The whole third paragraph on page 6 is spent on describing the dominating spatial patterns of rainfall events in the study watershed. Nevertheless, they seem to be forgotten after that. How these spatial patterns affect the identified intensity-duration thresholds and their changes? See my first comment.

R: Spatial patterns associated with different storm types and their effects on rainfall thresholds were an initial interest of ours, but the results show thresholds are similar across different storm types. For example the I(15, 80) threshold for year 1 for all storm types is roughly 20 mm/hr. Details are attached in Figure S4-5. However, we do still think that it is beneficial to describe the different rainfall patterns to illustrate that our modeling analyses are forced with rainfall data that reflects the different types of systems that impact this area.

5. Lines 174-176, given such many ignored factors that may influence the reliability of derived radar rainfall estimates, how "the realistic spatial and temporal patterns of rainfall" can be guaranteed?

R: While the rainfall intensities at each grid cell and time step will vary somewhat across Z-R relationships, the spatial and temporal characteristics of the storm will be retained. As the radar data are observed rather than simulated, the spatial and temporal patterns offer the best representation of "reality" available over the area. The comment on radar measurement uncertainty is meant to provide general guidance for things to consider when using radar data, but this area exhibits quality radar coverage, as noted now in the first line of section 3.1.

"Weather radar coverage is adequate for estimating rainfall over the study area (NOAA 2021), and radars have been operational since the mid-1990s. This allows us to utilize observed data to capture temporal and spatial characteristics of storms impacting the study area, a region of complex terrain."

6. Line 179, as the used Z-R relationships are not calibrated for the study watershed, the derived rainfall events are not "realistic storms". Replace "realistic" with "plausible".

R: Modified.

7. Page 8, a more concrete interpretation to I(D, j) can substantially improve readability. Possibly, adding somewhere "I(D, j) indicates that 100(1-j)% of the watershed experiences rainfall of duration D with intensity of I or larger."

R: We agree, thank you for this suggestion. We added this interpretation in Section 3.2 as follows:

"A threshold defined by I_D^j would denote a threshold where (100-j)% of the watershed experiences rainfall of duration D with an intensity of I or greater."

We also described the j^th percentile in the Discussion and Conclusion sections,

"In other words, a good indicator of the potential for a flash flood is the presence of intense pulses of rainfall over durations of 15-60 minutes that cover at least 15%-25% of the watershed (Figure 9)."

"The optimal threshold for predicting the occurrence of a flash flood in our study areas is the 75^th-85^th percentile of peak rainfall intensity averaged over 15-60 minutes, i.e., I_{30}^{75}-I_{30}^{85}. In other words, a flash flood tends to be produced when rainfall intensity over 15%-25% of the watershed area exceeds a critical value."

8. Line 383, the statement is not convincing

R: It has been rephrased as, "While the magnitude of rainfall thresholds estimated here may only work for similar, recently burned watersheds within the San Gabriel Mountains, this work provides a general methodology for exploring a reliable predictors of post-fire flash floods for other watersheds and settings."

9. Lines 394-397, inaccurate statements. Are all existing rainfall generators unable to produce physically realistic rainfall fields? I think this question has been addressed in many published studies.

R: We agree that this statement is misleading and needs revision. We have updated the text in the last paragraph of section 5.1 to increase clarity.

"Several limitations are present in this work. First, we assess a small number of storm events (34) in the area as we are limited by the length of radar and gage records as well as and the number of events that impact the indicator rain gages, though applying the five Z-R relationships provides us with 170 rainfall realizations to assess. We prefer the use of observed rainfall data (radar and gauges) over simulated products, such as output from a rainfall generator (e.g., Zhao et al., 2019; Evin et al., 2018), as the radar is able to capture the spatial and temporal patterns of rainfall intensity in the study area's complex terrain. Though rainfall generators have advanced to represent some synoptic-to-mesoscale features, such as frontal and convective precipitation (e.g., Zhao et al. 2019), they are fundamentally designed to represent statistical characteristics of rainfall in places with limited observations (Wilks and Wilby 1999) and cannot be relied upon to replicate small scale storm characteristics in complex terrain (e.g., Camera et al. 2016). Future work could compare results from this hydrologic modeling experiment with observed versus simulated rainfall. Second, the challenges of Z-R relationships to convert reflectivity to precipitation also presents challenges in accurately representing precipitation values. This can be addressed in future work through studies to constrain Z-R relationships for storms producing intense rainfall in this region and through the deployment or installation of high-resolution gap-filling radars (e.g., Johnson et al., 2019)."

Additional modifications were made in the first paragraph of section 3.1 that help explain our preference for observed over simulated data.

"Weather radar coverage is adequate for estimating rainfall over the study area (NOAA 2021), and radars have been operational since the mid-1990s. This allows us to utilize observed data to capture temporal and spatial characteristics of storms impacting the study area, a region of complex terrain."