Response to Referee document for NHESS manuscript 2021-345 by Li et al.

We thank the referee for their insightful comments. We have made constructive changes in response to each of these comments and believe that the manuscript is substantially improved as a result.

This Response to the Referee document provides a complete documentation of the changes that are responsive to each Referee Comment. The document is designed so that the changes that we have made in response to each comment can be immediately read and understood. Referee Comments are shown in black, while Author Responses are shown in blue. Quoted text from the revised manuscript is shown in red. For clarity, we group our responses to the referee's thematically similar General and Specific comments into individual numbered responses below.

Referee #3:

This paper presents the hydrologic analysis in post-fire debris flow by employing WRF-Hydro. The research extends an important hydrologic modeling tool to an important topic, post-fire hydrologic hazards. The paper is mainly an application of WRF-Hydro over a large area including burned scars in CA. It integrates multiple techniques and datasets, such as hydrologic simulations, radar data, and debris flow identification using RS, and so on, to improve hydrologic analysis. Also, the revised manuscript has big improvement. It definitely merits publication after consideration of these comments.

Response: We thank the referee for their careful review and positive assessment of our manuscript. In the below, we address each of the referee's comments. We hope that the changes have helped improve the clarity of the manuscript and increased the insights gained from the results. We thank the referee for judging this manuscript suitable for publication, and hope that our revision efforts further that sentiment.

1. Peak Flow Comments

General Comments: One problem with this research is that the debris flow susceptibility (or likelihood) is indicated using water volume. Runoff-generated debris flow occurs when flow strengths exceed the threshold, authors also noticed and mentioned in the manuscript. The key to debris flow occurrence is flow strengths, such peak flow, maximum depth, maximum velocity and so on, rather than volume. Of course, generated volume is closely related to peak flow but also controlled by duration.

Specific Comments:
Line 563-565. I suspect the accumulated volume is a reasonable metrics for runoff-generated debris flow assessment. Initiation of runoff-generated debris flow is a threshold behavior, which means the peak flow or unitless peak flow must be larger than certain threshold, or it is unlikely to happen. You may want to look at this paper and references therein:
https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019GL083623?msclkid=28d5425bbaed11ecaab0029541ed7fcc

Response #1: We thank the referee for their feedback regarding the use of accumulated volume in our susceptibility analysis, however we are confused by the contradictory nature of the above two comments. In the first comment the referee indicates that the use of accumulated volume may be a "problem", while in the second comment they indicate their suspicion that "the accumulated volume is a reasonable metric." Despite the contradictory nature of these comments, in the following, we assume the referee is seeking greater clarity on the use of accumulated volume versus peak discharge in debris flow susceptibility assessments.

It is important to note that in our manuscript we do not attempt to simulate debris flow initiation or the dynamics of debris flows. In this manuscript, we restrict our analysis to the assessment of debris flow susceptibility by locating areas with elevated environmental conditions conducive to debris flow occurrence (Brabb 1985; Guzzetti et al., 2005). We agree with the referee that if we were to attempt to simulate the dynamics of individual debris flows (e.g., initiation, entrainment, runout, etc.) rather than assess susceptibility, a different modeling framework would be needed [e.g., the individual catchment-level framework used by Rengers et al. (2016) and McGuire et al. (2017)]. Indeed, to resolve debris flow initiation requires spatiotemporal resolutions of meters and seconds that are not conducive to regional analyses.

To clarify the goal of this study, we revised the sentence in lines 96 – 98, which now reads (the edits we made is underscored): In other words, debris flow susceptibility neither simulates debris flow dynamics such as initiation nor estimates debris flow size or considers the timing or frequency of the debris flow occurrence. Rather, it focuses on locating areas prone to debris flows considering local environmental factors (Brabb 1985; Guzzetti et al., 2005).

We are in agreement with the referee that accumulated volume is closely related to peak discharge. Indeed, in our susceptibility analysis, we find similar susceptibility results regardless of variable choice. However, we concede that in the mechanistic debris flow world, peak discharge is likely a more valuable metric to quantify. As such, we have substantially revised our manuscript to now assess debris flow susceptibility using peak discharge output from WRF-Hydro. In addition, we have altered all related text, figures, and tables accordingly. Below, we provide the new peak flow versions of Figures 9, 10, and 11 and Table 3 for your reference.

In our revision, we've moved all accumulated discharge volume figures and tables to Appendix B (Figs. B7–9 and Table B5). For the convenience of comparison, we also reproduce accumulated discharge volume figures below (Fig. B7). For this domain under these meteorological conditions we generally find that similar conclusions regarding susceptibility can be drawn using either peak discharge or accumulated discharge volume, i.e., Rat Creek had medium susceptibility, Mill Creek had high susceptibility, and Big Creek and Nacimiento had very high susceptibility in the burn scar simulation, and catchments with catchment-area normalized peak discharge correspond well with the post-event debris flow identification (Figs. 9b,e & B7b,e). However, it is in the difference plots that the value of a peak discharge-based analysis is apparent. That is, streams and catchments with elevated susceptibility are more evident in the peak discharge maps (Figs. 9c,f & B7c,f). We thank the referee for this suggestion and hope that our substantial changes better highlight our methodological capabilities and results.

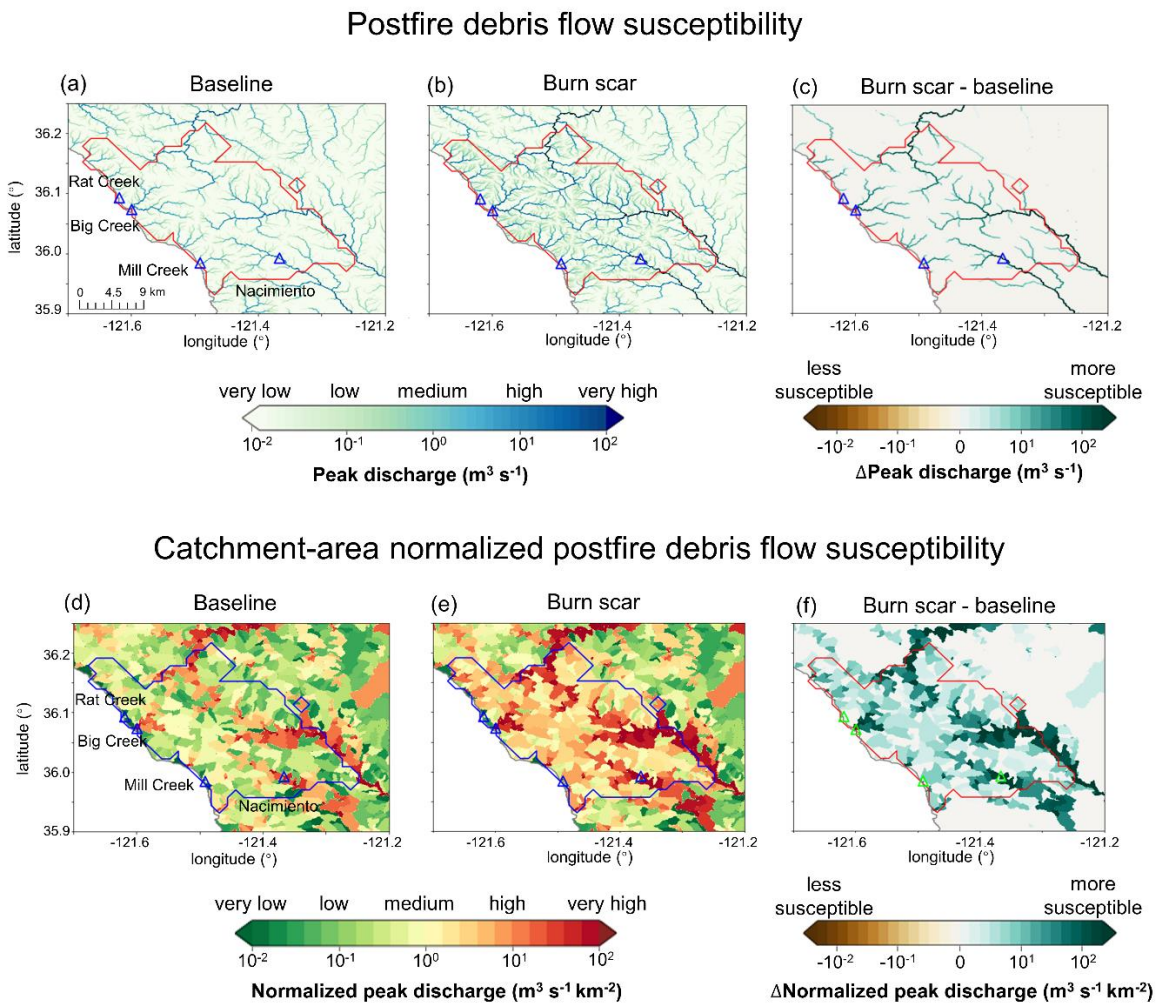

Fig. 9| Peak discharge-based postfire debris flow susceptibility. Peak discharge at individual stream level for the (a) baseline, (b) burn scar, and (c) difference between burn scar and baseline simulations from January 27th 00:00 to 28th 12:00 ($m^3 s^{-1}$). (d)–(f) Normalized peak discharge by catchment area at catchment level ($m^3 s^{-1} km^{-2}$; shading). For each catchment, the peak discharge is the maximum discharge rate at the catchment outlet from January 27th 00:00 to 28th 12:00 divided by catchment area. Triangles stand for debris flow deposition locations and are annotated in (a) and (d). We conduct similar analyses using accumulated discharge volume in Fig. B7 in Appendix B.

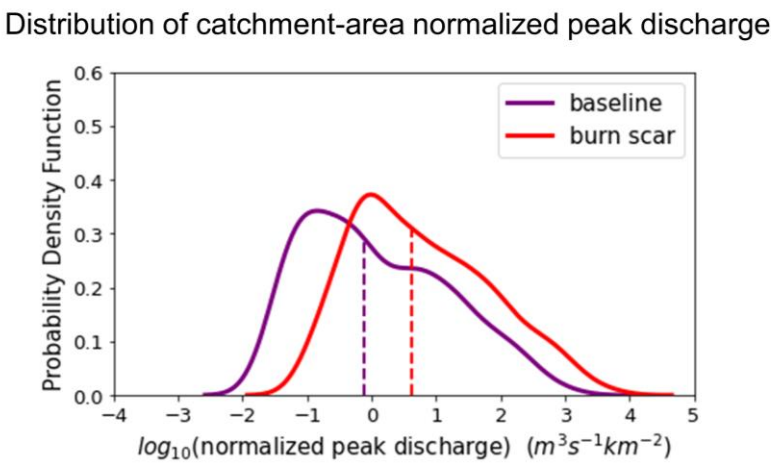

Fig. 10| Distributions of peak discharge at the outlet of the 404 catchments normalized by upstream catchment areas within Dolan burn scar in the baseline simulation (purple line) and in the burn scar simulation (red line). Dashed vertical lines indicate median values. We conduct similar analyses using accumulated discharge volume in Fig. B8 in Appendix B.

Table 3

Statistics of catchment area-normalized peak discharge in baseline and burn scar simulations

	mean	**std**	**5P**	**25P**	**50P**	**75P**	**95P**
Baseline simulation ($m^3\ s^{-1}\ km^{-2}$)	25.88	±95.71	0.04	0.14	0.76	8.21	129.54
Burn scar simulation ($m^3\ s^{-1}\ km^{-2}$)	110.80	±423.82	0.19	0.84	4.16	36.21	603.15
Relative percent change	328%	343%	375%	500%	447%	341%	366%

Table 3| Statistics, including the mean, standard deviation (std), 5P, 25P, 50P, 75P, and 95P, of the catchment-area normalized peak discharge for all the 404 basins within the Dolan burn scar in

the baseline and burn scar simulation and their relative percent changes. We conduct similar analyses using accumulated discharge volume in Table B5 in Appendix B.

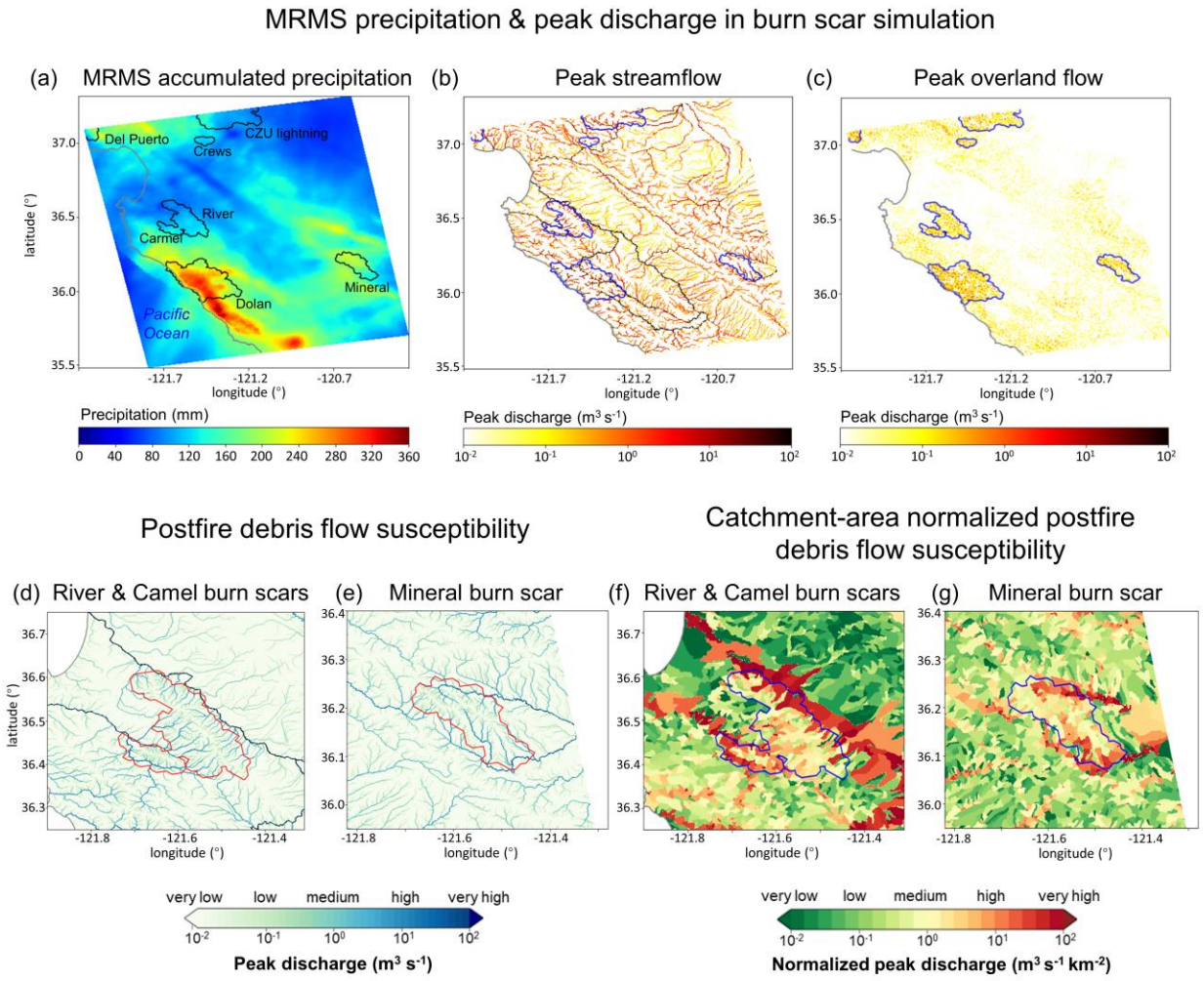

Fig. 11| MRMS accumulated precipitation and peak discharge informed regional debris flow susceptibility. (a) MRMS accumulated precipitation during January 27th 00:00 to 29th 23:00 over the model domain (mm; shading). Names of burn scars are labeled in black. (b) Peak streamflow ($m^3 s^{-1}$; yellow-to-red shading) and (c) peak overland flow from 27th 00:00 to 28th 12:00 over the model domain ($m^3 s^{-1}$; yellow-to-red shading). (d)–(e) Stream-level postfire debris flow susceptibility as Fig. 9b but for River and Camel burn scars. (f)–(g) Catchment-area normalized debris flow susceptibility as Fig. 9e but for River and Camel burn scars. Wildfire perimeters of

2020 wildfire season are outlined in black in (a), in blue in (b), (c), (f), and (g), and in red in (d) and (e). The coastline of California is depicted in grey.

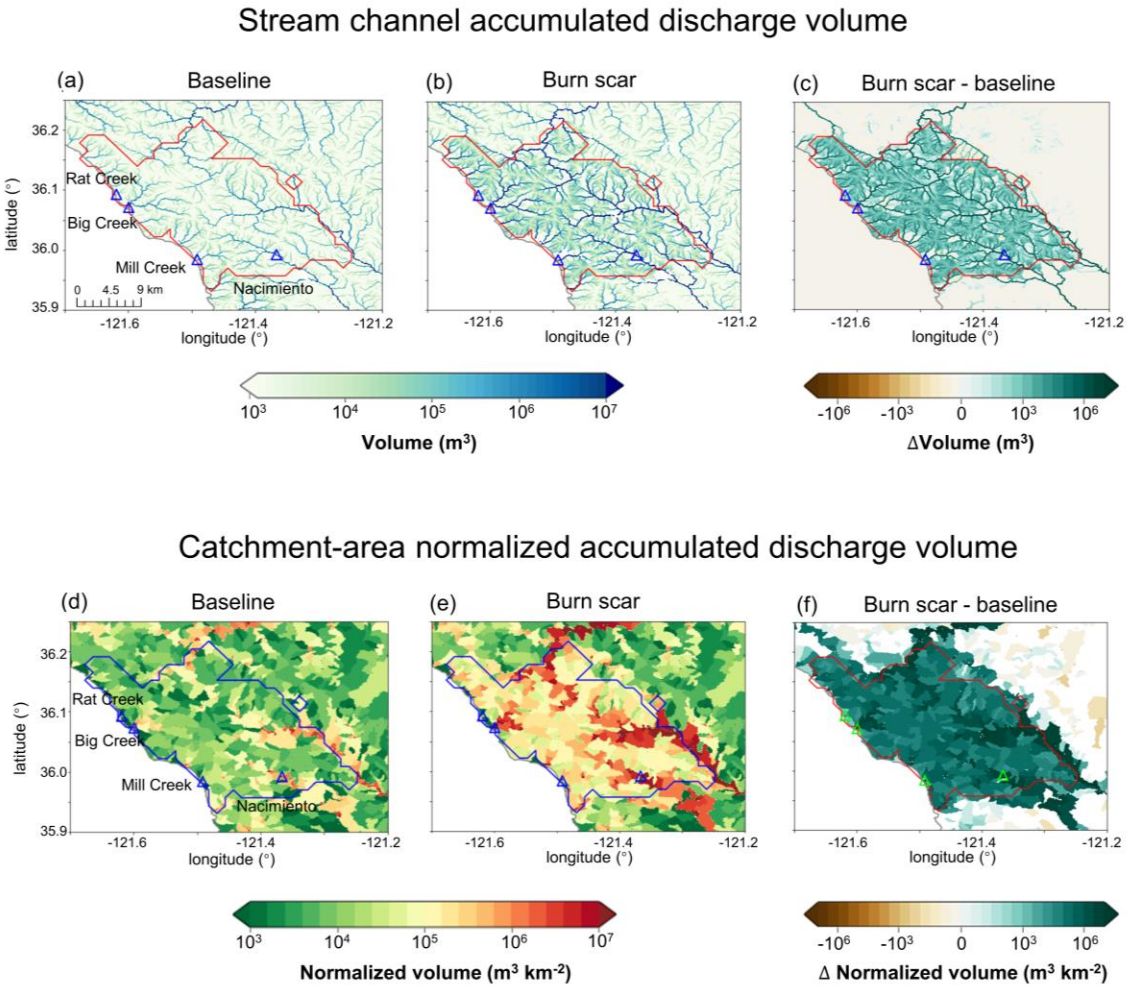

Fig. B7| Accumulated discharge volume at individual stream level for the (a) baseline, (b) burn scar, and (c) difference between burn scar and baseline simulations (m^3). Total discharge volume is accumulated from January 27th 00:00 to 28th 12:00. (d)–(f) Normalized discharge volume by catchment area at catchment level (m^3 km^{-2}; shading; Santi & Morandi, 2013). For each catchment, the discharge volume is accumulated at the catchment outlet from January 27th 00:00 to 28th 12:00 divided by catchment area. Triangles stand for debris flow deposition locations and are annotated in (a) and (d).

In addition to Figure and Table changes, we have altered the text of Section 5.3, which now reads:

5.3 Debris flow susceptibility assessment for the Dolan burn scar

Since high magnitude runoff is often the cause and precursor of runoff-generated debris flows in burned areas (Cannon et al., 2003, 2008; Rengers et al., 2016), we use peak discharge of overland flow and streamflow to assess runoff-generated debris flow susceptibility under pre-fire (i.e., baseline; Fig. 9a&d) and postfire (i.e., burn scar simulation; Fig. 9b&e) conditions [we conduct similar analyses using accumulated discharge volume in Figs. B7–9 and Table B5 in Appendix B]. We assess changes at both stream and catchment levels and use the difference between burn scar and baseline simulations to assess added debris flow susceptibility (Fig. 9c&f). Consistent with the increasing erosive and entrainment power associated with increasing discharge, our debris flow susceptibility increases as the peak discharge increases. To reduce the effects of catchment size on the peak discharge-based susceptibility levels, we normalize a catchment's discharge by the area of the catchment (Leopold et al., 1964; McCormick et al., 2009; Fig. 9d–f). Non-normalized catchment susceptibility maps are also provided (Fig. B10).

In the pre-fire baseline simulation, the AR-induced precipitation produces lower debris flow susceptibility over most of the domain, but elevated susceptibility along stream channels (Fig. 9a). We note no substantial differences between areas in or out of the burn scar. In the burn scar simulation, debris flow susceptibility levels increase across the Dolan burn scar and along channels outside but downstream of the burn scar (Fig. 9b–c). The peak discharge near Rat Creek, Big Creek, Mill Creek, and Nacimiento more than triples (Table 2 & Fig. 9a–c). Within the burn scar, susceptibility along major stream channels, such as the Nacimiento River and San Antonio River increase. Outside the burn scar, susceptibility levels along river channels downstream of the burn scar, such as the Arroyo Seco River, also increase (Fig. 9c).

At the catchment level, debris flow susceptibility is assessed using peak discharge normalized by catchment areas at the outlet of each catchment between January 27th 00:00 to 28th 12:00 (Fig. 9d–f). The catchment-area normalized peak discharge is classified into five categories based on equal intervals on log10 scale. The susceptibility categorization follows: "very low" (~10^{-2} m^3 s^{-1} km^{-2}), "low" (~10^{-1} m^3 s^{-1} km^{-2}), "medium" (~10^{0} m^3 s^{-1} km^{-2}), "high" (~10^{1} m^3 s^{-1} km^{-2}), and "very high" (~10^{2} m^3 s^{-1} km^{-2}). In the baseline simulation, majority of catchments are subject to low or very low debris flow susceptibility with normalized peak discharge less than 1 m^3 s^{-1} km^{-2} (Fig. 9d). In the burn scar simulation, about half of the catchments within the Dolan burn scar have medium susceptibility or above, and about 1/4 of basins are subject to high to very high debris flow susceptibility (Fig. 9e and Table 3). The additional debris flow susceptibility brought about by the inclusion of wildfire burn scar characteristics is substantial (Fig. 9f).

To summarize changes in debris flow susceptibility as a result of including burn scar characteristics in WRF-Hydro simulations, we create distributions of pre-fire baseline and burn scar catchment-area normalized peak discharge from the 404 catchments located within the Dolan burn scar perimeter (Fig. 10). After incorporating burn scar characteristics, the full distribution shifts to the right, indicating increased susceptibility levels – a shift considered robust by a Student's t-test (*p* value: 5.3E-23). A quantitative assessment of this shift indicates that both the mean and the standard deviation of catchment area normalized peak discharge increase by more

than 300% (Table 3). We also assess shifts at a range of distribution percentiles: 5P: 375%, 25P: 500%, 50P: 447%, 75P: 341%, and 95P: 366% (Table 3). In the burn scar simulation, more than half of catchments have normalized peak discharge > 10^0 m^3 s^{-1} km^{-2} (i.e., medium susceptibility) and about 1/4 of catchments have normalized peak discharge > 10^1 m^3 s^{-1} km^{-2} (i.e., high susceptibility) – values that correspond to the 70P and 90P of the baseline simulation, respectively. Disproportionate shifting of the distribution suggests that debris flow susceptibility increases non-linearly under simulated burn scar conditions.

Our catchment-area normalized peak discharge-based susceptibility assessment also indicates that the catchments containing Mill Creek, Big Creek, and Nacimiento have high or very high susceptibility (Fig. 9d–f), consistent with our (limited) debris flow observations. Other areas with elevated susceptibility include catchments containing the Arroyo Seco and San Antonio Rivers. Beyond the burn scar perimeter, effects of fire expand to adjacent and downstream catchments, and some drainage basins along the Arroyo Seco and Nacimiento Rivers are simulated to have very high susceptibility, i.e., normalized peak discharge volumes in excess of 10^2 m^3 s^{-1} km^{-2} (Fig. 9e&f).

2. Spatial and Temporal Resolution Comments

General Comments:
A related problem is the spatial resolution of computational cell. The 100m grid is used in the study. It is okay to simulate the overland flow generation and hydrograph at the outlets, but it provides few information about flow dynamics at fine resolution, which is the key to observe the debris flow generation and occurrence.

Specific Comments:
Line 151. Spatial resolution of 100m is problematic for capturing debris flow behavior. This comment applies to the same issue throughout the text (e.g., lines 273, 332, etc.).
Line 376. Justify 1-hour results can represent the flow peaks.

Response #2: We thank the reviewer for this question and an opportunity for us to clarify (a) the objectives of our paper and (b) the spatial and temporal resolution of our different modeling components:

(a) As mentioned in Response #1, it is important to note that in our manuscript we do not attempt to simulate debris flow initiation or the dynamics of debris flows. In this manuscript, we restrict our analysis to the assessment of debris flow susceptibility by locating areas with elevated environmental conditions conducive to debris flow occurrence. We agree with the referee that if we were to attempt to simulate the dynamics of individual debris flows (e.g., initiation, entrainment, runout, etc.) rather than assess susceptibility, a different modeling framework would be needed [e.g., the individual catchment-level framework used by Rengers et al. (2016) and McGuire et al. (2017)]. Indeed, to resolve debris flow initiation requires spatiotemporal resolutions that are not conducive to regional analyses.

(b) To clarify the spatial resolutions used in our study, here we describe the various components of our modeling scheme. Channelized streamflow is simulated by the channel routing module of WRF-Hydro at fine spatial resolutions that range from 1.5 to 100 m depending on the USGS DEM-indicated stream order. This information can be found in Figure B3 and Table B1 in our manuscript. Since our model domain is quite large (~35,000 km^2), streamflow is then output on the 100-m grid due to limited storage space. However, discharge dynamics are solved at a resolution consistent with the stream order.

As for temporal resolution, the MRMS precipitation forcing is hourly but our WRF-Hydro terrain routing and channel routing modules compute overland flow and channelized streamflow every 6 seconds. However, we output simulated data at an hourly resolution due to limited storage space.

To clarify our methodology, we added a sentence to lines 288 – 293, lines 328 – 331, and lines 356 – 361 in our revised manuscript (the added sentence is underscored). Lines 288 – 293 now read: The channel routing module then calculates channelized flows assuming a trapezoidal channel shape (Fig. B2). Parameters related to the trapezoidal channel, such as channel bottom width (B_w), Manning's roughness coefficient (n), and channel side slope (z) are functions of channel stream order (Fig. B3 and Table B1). Channelized streamflow is computed at spatial resolutions ranging from 1.5 m to 100 m depending on the channel stream order (Table B1). Computed streamflow is then output on the 100-m grid.

Lines 328 – 331 now read (the added sentence is underscored): Noah-MP passes excess water to the terrain routing module, which simulates overland flow using a 2-dimensional fully-unsteady, explicit, finite-difference diffusive wave equation adapted from Julien et al. (1995) and Ogden (1997). In this application, overland flow is computed at each 6 second time step and is archived hourly at 100-m spatial resolution.

Lines 356 – 361 now read (the added sentence is underscored): If overland flow intersects grid cells identified as channel grids (2nd Strahler stream order and above; pre-defined by the hydrologically conditioned USGS 30-m DEM), the channel routing module routes the water as channelized streamflow using a 1-dimensional, explicit, variable time-stepping diffusive wave formulation. In this work, the channel routing module calculates streamflow at 6-s temporal resolution and spatial resolutions ranging from 1.5 m to 100 m depending on the channel stream order (Fig. B3 and Table B1).

3. MRMS Data

General Comments: The third problem is the performance of MRMS data in the study area. I believe MRMS data is the best option author may have, but its poor quality in mountainous area always unable catch the intensive storms. I'd like to see how authors consider this issue.

Response #3: We thank the referee for pointing out the contingencies with using MRMS data, and we agree that given all available options, it is the "best choice." MRMS data has relatively high

spatial (1-km) and temporal resolutions (hourly) compared to other available datasets. In addition, it covers a large spatial domain (i.e., CONUS) and thus it is valuable for regional susceptibility assessments. We acknowledge the uncertainties in the MRMS precipitation data on lines 701 – 706 and in Appendix A. We note that MRMS Gauge-Corrected and MRMS Mountain Mapper Precipitation datasets may be superior products and are therefore preferred in mountainous areas, however these datasets are not yet available for our study period (Jan 1–31, 2021) (https://mtarchive.geol.iastate.edu/2021/01/31/mrms/ncep/; retrieved May 2022). In addition, it is important to note that from an operational forecast application perspective, predicted precipitation products will suffer similar (and likely worse) limitations to that of MRMS. It is therefore important to work with the best data resources available. We agree with the referee that the best available data set is MRMS.

A discussion on this topic can be found in the manuscript on lines 701 – 706: However, this also means the accuracy of WRF-Hydro predictions depends on the accuracy of precipitation forcing, and in our hindcast application, MRMS precipitation data (Appendix A). Accordingly, our WRF-Hydro-based assessment could benefit from precipitation products mosaiced from various sources to constrain precipitation-based uncertainties (e.g., gauge-corrected and/or Mountain Mapper MRMS), although the long processing time of these datasets inhibits timely post-event assessments.

We have also added a sentence in Appendix A to highlight the caveats of using MRMS in mountainous areas. Appendix A now reads (the changes we made are underscored):

Appendix A

Text A1. Multi-Radar/Multi-Sensor System (MRMS) radar-only precipitation estimate and uncertainty

MRMS is a precipitation product that covers the contiguous United States (CONUS) on 1-km grids. It combines precipitation estimates from sensors and observational networks (Zhang et al., 2011, 2014, 2016), and is produced at the National Centers for Environmental Prediction (NCEP) and distributed to National Weather Service forecast offices and other agencies. Input datasets used to produce MRMS include the U.S. Weather Surveillance Radar-1988 Doppler (WSR-88D) network and Canadian radar network, Parameter-elevation Regressions on Independent Slopes Model (PRISM; Daly et al., 1994, 2017), Hydrometeorological Automated Data System (HADS) gauge data with quality control (Qi et al., 2016), and outputs from numerical weather prediction models. There are four different MRMS quantitative precipitation estimates (QPE) products incorporating different input data or combinations: radar only, gauge only, gauge-adjusted radar, and Mountain Mapper. One limitation of using MRMS radar only precipitation data is that radars struggle to capture rainfall in mountainous areas due to orographic beam blocking (Anagnostou et al. 2010; Germann et al. 2007). Gauge-corrected and Mountain Mapper MRMS are thought to be superior products in mountainous terrains and therefore are preferred. However, for our study period (i.e., January 1–31, 2021), the gauge-corrected and Mountain Mapper MRMS are not available (as of May 2022).

We acknowledge that precipitation data has uncertainties. Use of different precipitation products may help to constrain uncertainties. A study comparing different gridded precipitation datasets including satellite-based precipitation data, gauge dataset, and multi-sensor products

revealed large uncertainties in precipitation intensity (Bytheway et al., 2020). However, comparing different precipitation datasets to characterize uncertainties is beyond the scope of this study. MRMS provides gridded precipitation at high temporal (hourly) and spatial (1-km) resolutions, making it a useful tool to demonstrate the utility of WRF-Hydro in post-wildfire debris flow susceptibility assessments.

4. Overland Flow Augmentation

General Comments: A highlight of this study is output overland flow by modifying the source code. The overland flow generation is represented by Noah-MP. The related variables are *q_sfcflx_x* and *q_sfcflx_y*. and combined with *qqsfc* and *Noah_distr_routing* to calculate the amount (forgive me if I remember wrong). It is very valuable to point out the modification. I believe the modification of the source code will be published with this work. It will benefit WRF-Hydro community!

Response #4: We agree and thank the reviewer characterizing our contribution as very valuable. We hope so! We provide a link to our modified code in the *Code availability statement* at the end of the manuscript (lines 1018 – 1020). At this link we provide the modified source code and instructions on how to use the modified Fortran files. In our revision, we've added a sentence to point readers to the *Code availability statement* on lines 349 – 351.

Lines 349 – 351 now read (the sentence we added is underscored): One key advance made in this work is that we modified WRF-Hydro source code to output overland flow (see the Code availability statement for the modified source code).

5. Atmospheric Rivers

General Comments: Another problem is "atmospheric river" (AR) mentioned many times, and it seems authors emphasize the AR is the major reason of heavy rainstorm of post-fire debris flow. Generally, it is fine. But AR does not the directly produce rainfall and AR is a very large scale atmospheric pattern occurred many places. AR-triggered certain synoptic system, such as NCFR carrying plenty of moisture provides opportunities of heavy rainfall.

Response #5: We thank the referee for highlighting the confusion produced by our manuscript regarding atmospheric rivers. According to the glossary of the American Meteorological Society (https://glossary.ametsoc.org/wiki/Atmospheric_river), an atmospheric river (AR) is defined as "a long, narrow, and transient corridor of strong horizontal water vapor transport" that "frequently leads to heavy precipitation where they are forced upward—for example, by mountains or by ascent in the warm conveyor belt.".

We agree that ARs are large synoptic systems, which partly motivates our use of a regional hydrologic model to study debris flow susceptibility, that is, an individual landfalling AR will alter debris flow likelihood over a large region (the area of an AR is typically on the order of 10^5 km^2), rather than within an individual catchment. Indeed, ARs are reported to produce 30–50% of the annual precipitation and 60%–100% of the extreme precipitation along the U.S. west coast (Collow et al., 2020; Eldardiry et al., 2019; Hecht & Cordeira, 2017). Given their importance to

hydrological conditions over our modeling domain, we have added text to lines 75 – 88 that defines ARs and contextualizes their importance:

Lines 75 – 88 now read: On the U.S. west coast, atmospheric rivers (ARs) are the dominant synoptic weather systems responsible for producing postfire debris flows (Barth et al., 2017; Oakley et al., 2017, 2018; Young et al., 2017). ARs are long filament-like bands of elevated water vapor within the lower troposphere that often form over ocean basins. They are responsible for over 90% of poleward water vapor transport (Zhu & Newell, 1998) and often result in heavy precipitation upon landfall, particularly with orographic uplift (Ralph et al., 2004; Neiman et al., 2008). It is reported that 30–50% of annual precipitation and 60%–100% of extreme precipitation along the U.S. west coast is the result of ARs (Collow et al., 2020; Eldardiry et al., 2019; Hecht & Cordeira, 2017). In California, anthropogenic climate change is projected to increase AR intensity (Huang et al., 2020a, 2020b), increase the intensity and frequency of wet-season precipitation (Polade et al., 2017; Swain et al., 2018), increase wildfire potential (Brown et al., 2020; Swain 2021), and extend the wildfire season (Goss et al., 2020). As such, the occurrence and intensity of postfire debris flows are likely to increase as the effects of anthropogenic climate change persist (Cannon & DeGraff, 2009; Kean & Staley, 2021; Oakley 2021).

6. Calibration

General Comments: A similar problem is only three stream gauges are used to calibrate the model for such a large area. One reason might be lack of natural flow record. You may want to use natural flow data, such as https://rivers.codefornature.org/, https://pubs.er.usgs.gov/publication/70046617 Temporal resolution may be different but still provide valuable information.

Specific Comments:

Line 369. Justify 3 gauges used for calibration is sound. Or you may consider to use natural flow data:
https://rivers.codefornature.org/, https://pubs.er.usgs.gov/publication/70046617, or https://pubs.er.usgs.gov/publication/70046617

Response #6: We thank the reviewer for providing extra data sources. Indeed, records of natural flows are scarce in California. We chose the three USGS stream gages because they are located downstream of Dolan burn scar which is the focus of the case study.

For reference, lines 385 – 393 of our manuscript read: Due to the Mediterranean climate of California, many USGS stream gages experience low or no flow during the dry season. In addition, many gages are under manual regulation to mitigate wet-season flood risks and better distribute water resources. As such, it can be challenging to obtain natural streamflow observations for model calibration. Here, three USGS stream gages [i.e., Arroyo Seco NR Greenfield, CA (ID 11151870), Arroyo Seco NR Soledad, CA (ID 11152000), and Arroyo Seco BL Reliz C NR Soledad, CA (ID 11152050)] (Fig. 1a) on streams that have measurable flows during our study period and are free of human regulation are used. These gages are located downstream of the Dolan burn scar and hence are useful in calibrating the parameters associated with burn scar effects.

We investigated the two links provided by the reviewer. The first suggested dataset (https://rivers.codefornature.org/) provides machine-learning estimates of natural flows at monthly

resolution. Their methodology is provided at this link: https://rivers.codefornature.org/#/science. In line with the referee's opinion that coarse temporal resolutions are not ideal for debris flow assessments, we do not think monthly flow records are helpful in calibrating the model to assess debris flow susceptibility.

The second dataset the reviewer suggested is the USGS GAGESII data (https://pubs.er.usgs.gov/publication/70046617). However, this is not a streamflow time series data. Instead, it provides geospatial data and classifications for the USGS stream gages including shapefiles of the point locations, basins, and streamlines of the USGS stream gages. After carefully comparing the point locations of GAGESII with the locations of USGS stream gages which we are already using (https://maps.waterdata.usgs.gov/mapper/?state=ca), we found that they closely match. As a result, the above two datasets do not provide additional useable information.

Detailed Comments

Line 153-154. Other studies did use water-only models but those modeling works were done by using high resolution DEM, such as 1-m lidar data. The flow dynamics at fine resolution describes the initiation of debris flow (e.g., Rengers et al., 2016; Mcguire et al., 2017)

Rengers, F.K., McGuire, L.A., Kean, J.W., Staley, D.M. and Hobley, D.E.J., 2016. Model simulations of flood and debris flow timing in steep catchments after wildfire. Water Resources Research, 52(8), pp.6041-6061.

McGuire, L. A., Rengers, F. K., Kean, J. W., and Staley, D. M. (2017), Debris flow initiation by runoff in a recently burned basin: Is grain-by-grain sediment bulking or en masse failure to blame?, Geophys. Res. Lett., 44, 7310– 7319, doi:10.1002/2017GL074243.

Response #7: We agree with the reviewer that these studies simulated debris flow initiation at a high resolution. However, we are not attempting to do what these studies have done. We are assessing debris flow susceptibility over a regional domain. We have balanced the computational costs between fine spatiotemporal resolution and large spatial extent. Our model domain covers more than 10,000 catchments and 35,000 km^2, while the above mentioned mechanistic studies are focused on individual catchments [the study area is 0.01 km^2 in McGuire et al. (2017) and <2 km^2 in Rengers et al. (2016)].

To avoid the confusion regarding our intent, we deleted the word "behavior" from the following sentence:

The sentence in lines 164 – 168 of our revised manuscript now reads: Previous efforts, employing shallow water equations, diffusive, kinematic, and diffusive-kinematic wave models, have demonstrated that water-only models can provide critical insights on runoff-driven debris flows (Arattano & Savage, 1994; Arattano & Franzi, 2010; Di Cristo et al., 2021), even in burned watersheds (Rengers et al., 2016; McGuire & Youberg, 2020).

And given the lack of clarity in our original manuscript, we have modified the text in the Introduction Section that discusses these studies.

Lines 130 – 141 now read: Studies that have investigated postfire hydrologic responses using physics-based models have largely focused on mechanistic studies such as short-term responses at high spatiotemporal resolutions (Rengers et al., 2016; McGuire et al., 2016, 2017) or long-term runoff responses at coarse temporal resolutions (McMichael & Hope, 2007; Rulli & Rosso, 2007) in individual catchments. For example, process-based models have employed shallow water equations to better understand the triggering (McGuire et al., 2017; Tang et al., 2019a, 2019b) and sediment transport mechanisms (McGuire et al., 2016) of postfire debris flows as well as the timing of postfire debris flows (Rengers et al., 2016). The numerical models employed by these studies are used to simulate debris flow dynamics rather than assess susceptibility over regional domains, as such they focus on individual catchments (with drainage areas of ~1 km2) with very high spatiotemporal resolutions (Rengers et al., 2016; McGuire et al., 2016, 2017; Tang et al., 2019a, 2019b).

Line 179. Not clear. You may want to point out the time window of rainfall intensity, such as 15-min or 30-min.

Response #8: To clarify, we have modified this sentence to indicate that the precipitation data we used to calculate the 24 mm hr^{-1} is hourly MRMS.

The sentence now reads: On January 27–29, 2021, an atmospheric river (AR) made landfall on the Big Sur coast, bringing more than 300 mm of rainfall to California's Coast Ranges (Fig. 2), with a peak rainfall rate of 24 mm h^{-1} [calculated with Multi-Radar/Multi-Sensor System (MRMS) precipitation; Zhang et al., 2011, 2014, 2016].

Line 192. Add spatial resolution in this section.

Response #9: The spatial resolution of Sentinel-2 optical data that is used to calculate rdNDVI is 10 m. The SAR-change is calculated from the Sentinel-1 satellites which also have a spatial resolution of 10 m.

In our revised manuscript, we added the spatial resolution to lines 215 – 217. It now reads (the changes are underscored): …where NIR is the near-infrared response and Red is the visible red response. rdNDVI was calculated from 10-m Sentinel-2 satellite data using the HazMapper v1.0 Google Earth Engine application (Scheip & Wegmann, 2021).

And lines 224 – 226: Lastly, we searched for debris flows (and other ground surface deformation) by examining SAR backscatter change with data acquired by the 10-m Copernicus Sentinel-1 (S1) satellites [see full description in Handwerger et al. (2022)].

Lines 223-225. You may add a figure to compare the RS-based debris flow and field observation.

Response #10: Field observations were conducted by our co-author Dr. Noah Finnegan. However, these field excursions were not formal analyses, but rather were meant to confirm that the remotely sensed debris flow events indeed occurred. We are aware of an exhaustive field-based analysis in

the Dolan burn scar (Cavagnaro et al., 2021) and when that effort is complete, a comparison with the RS-based observations would be ideal.

Line 262. Simplifying the model description? WRF-Hydro has been used extensively.

Response #11: We agree that WRF-Hydro has been used extensively in simulating streamflow, but to our knowledge it has only rarely been employed in the landslide debris flow community. As it is a new tool in the debris flow, postfire hydrology, and natural hazards communities, we prefer to maintain our model description for the benefit of these newer audiences.

Line 453,455. Change unit of Ks to mm/hr.
Response #12: We prefer to keep the unit of Ks to be m/s because m/s is the unit that the Noah-MP LSM and WRF-Hydro utilize (Chen & Dudhia, 2001; see Table B3).

Line 505-506. Great metrics you used here in the context of hazards and their impacts.
Response #13: We thank the referee for their kind words.

Line 669-672. I agree the uniform precip introduce bias of hazards distribution over watersheds. But I highly suspect the assessments based on the data and method in this study, such as the computation grid size, MRMS quality in the study area, parameters on burned scars, and metrics (water volume rather than peak flow) used for debris flow assessment.

Response #14: It is unclear to us what the referee is asking here. We thank the referee for acknowledging their agreement that the use of uniform precipitation is one of the limitations of the USGS methods.

Line 730-736. You got the idea but confused why the volume of water is used for debris flow assessment.
Response #15: We thank the referee for identifying this confusing sentence. In this manuscript, we are not attempting simulate debris flow triggering processes. This sentence was intended to explain a future direction that could potentially advance our susceptibility assessment to better characterize debris flow occurrence likelihoods. To clear the confusion, we revised the sentence and the paragraph in our revised manuscript.

Lines 751 – 762 now read: A second capability in need of development is the use of WRF-Hydro to identify debris flow triggering time and location by employing a domain-specific rainfall ID threshold trained with historic landslide inventory and triggering rainfall events (Tognacca et al., 2000; Gregoretti & Dalla Fontana, 2007, 2008) or a newly developed dimensionless discharge and Shields stress threshold (Tang et al., 2019a; McGuire & Youberg, 2020). While in this study we do not attempt to simulate debris flow dynamics such as triggering, we note that WRF-Hydro is capable of simulating overland flow and streamflow at higher spatiotemporal resolutions [on scales that are similar to other debris flow mechanistic studies such as Rengers et al. (2016), McGuire et al. (2016, 2017), and Tang et al. (2019a, 2019b)]. Therefore, WRF-Hydro's capability to simulate the triggering processes of runoff-generated debris flows is potentially only limited by the spatiotemporal resolution of precipitation forcing and computing resources.

Figure 4: change rainrate unit to mm/hr, which is more common for post-fire hydrologic study. You may tune font size in all figures.

Response #16: We have changed the rainrate unit to mm/hr and tuned the font size for all figures in our revised manuscript.

References:

Anagnostou MN, Kalogiros J, Anagnostou EN, Tarolli M, Papadopoulos A, Borga MJJoh (2010) Performance evaluation of high-resolution rainfall estimation by X-band dual-polarization radar for flash flood applications in mountainous basins 394:4-16

Arattano, M., & Franzi, L. (2010). On the application of kinematic models to simulate the diffusive processes of debris flows. Nat. Hazards Earth Syst. Sci., 10(8), 1689-1695. doi:10.5194/nhess-10-1689-2010

Arattano, M., & Savage, W.Z. (1994). Modelling debris flows as kinematic waves. Bulletin of the International Association of Engineering Geology 49, 3–13. https://doi.org/10.1007/BF02594995

Barth NA, Villarini G, Nayak MA, White K (2017) Mixed populations and annual flood frequency estimates in the western United States: The role of atmospheric rivers Water Resour Res 53:257-269

Brabb E.E. (1985) Innovative approaches to landslide hazard and risk mapping. In: International Landslide Symposium Proceedings, Toronto, Canada. pp 17-22

Brown, E.K., Wang, J., & Feng, Y. (2020). U.S. wildfire potential: a historical view and future projection using high-resolution climate data. Environmental Research Letters. 16, 034060

Bytheway, J. L., Hughes, M., Mahoney, K., & Cifelli, R. (2020). On the Uncertainty of High-Resolution Hourly Quantitative Precipitation Estimates in California, Journal of Hydrometeorology, 21(5), 865-879. Retrieved Oct 25, 2021, from https://journals.ametsoc.org/view/journals/hydr/21/5/jhm-d-19-0160.1.xml

Cannon, S. H., Gartner, J. E., Parrett, C., & Parise, M. (2003). Wildfire-related debris-flow generation through episodic progressive sediment-bulking processes, western USA. Debris-Flow Hazards Mitigation: Mechanics, Prediction, and Assessment. Millpress, Rotterdam, pp. 71-82.

Cannon, S. H., Gartner, J., Wilson, R., Bowers, J., & Laber, J. (2008). Storm Rainfall Conditions for Floods and Debris Flows from Recently Burned Basins in Southwestern Colorado and Southern California. Geomorphology, 96, 250-269. Doi:10.1016/j.geomorph.2007.03.019

Cannon, S. H., & DeGraff, J. (2009). The increasing wildfire and postfire debris flow threat in western USA, and implications for consequences of climate change. In Landslides–disaster risk reduction (pp. 177-190): Springer.

Cavagnaro, D. et al. (2021) Variability in hydrologic response to rainfall across a burn scar: observations from the Dolan Fire, California. AGU abstract. https://agu.confex.com/agu/fm21/meetingapp.cgi/Paper/921613

Collow ABM, Mersiovsky H, Bosilovich MG (2020) Large-Scale Influences on Atmospheric River–Induced Extreme Precipitation Events along the Coast of Washington State J Hydrometeor 21:2139-2156 doi:10.1175/JHM-D-19-0272.1

Chen, F., & Dudhia, J. (2001). Coupling an Advanced Land Surface–Hydrology Model with the Penn State–NCAR MM5 Modeling System. Part I: Model Implementation and Sensitivity, Monthly Weather Review, 129(4), 569-585. Retrieved Oct 24, 2021, from https://journals.ametsoc.org/view/journals/mwre/129/4/1520-0493_2001_129_0569_caalsh_2.0.co_2.xml

Daly, C., R. P. Neilson, & D. L. Phillips. (1994). A statistical-topographic model for mapping climatological precipitation over mountainous terrain. J. Appl. Meteor., 33, 140–1

Daly, C., M. E. Slater, J. A. Roberti, S. H. Laseter, & L. W. Swift Jr. (2017). High-resolution precipitation mapping in a mountainous watershed; Ground truth for evaluating uncertainty in a national precipitation dataset. Int. J. Climatol., 37, 124–137, https://doi.org/10.1002/joc.4986.

Di Cristo, C., Iervolino, M., Moramarco, T., & Vacca, A. (2021). Applicability of Diffusive model for mud-flows: An unsteady analysis. Journal of Hydrology, 600, 126512. Doi:https://doi.org/10.1016/j.jhydrol.2021.126512

Eldardiry H, Mahmood A, Chen X, Hossain F, Nijssen B, Lettenmaier DP (2019) Atmospheric River–Induced Precipitation and Snowpack during the Western United States Cold Season J Hydrometeor 20:613-630 doi:10.1175/JHM-D-18-0228.1

Germann U, Galli G, Boscacci M, Bolliger M (2007) Radar Precipitation Measurement in a Mountainous Region Q J Roy Meteorol Soc 132:1669-1692 doi:10.1256/qj.05.190

Goss M, Swain DL, Abatzoglou JT, Sarhadi A, Kolden CA, Williams AP, Diffenbaugh NS (2020) Climate change is increasing the likelihood of extreme autumn wildfire conditions across California Environmental Research Letters 15:094016 doi:10.1088/1748-9326/ab83a7

Gregoretti, C., & Dalla Fontana, G. (2007). Rainfall threshold for the initiation of debris flows by channel-bed failure in the Dolomites. Debris-Flow Hazards Mitigation: Mechanics, Prediction, and Assessment, 11-22.

Gregoretti, C., & Dalla Fontana, G. (2008). The triggering of debris flow due to channel-bed failure in some alpine headwater basins of the Dolomites: analyses of critical runoff. Hydrological Processes, 22(13), 2248-2263. doi:https://doi.org/10.1002/hyp.6821

Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M., & Ardizzone, F. J. G. (2005) Probabilistic landslide hazard assessment at the basin scale 72:272-299

Handwerger AL, Huang MH, Jones SY, Amatya P, Kerner HR, Kirschbaum DB (2022) Generating landslide density heatmaps for rapid detection using open-access satellite radar data in Google Earth Engine Nat Hazards Earth Syst Sci 22:753-773 doi:10.5194/nhess-22-753-2022

Hecht, C. W., & Cordeira, J. M. (2017). Characterizing the influence of atmospheric river orientation and intensity on precipitation distributions over North Coastal California. Geophysical Research Letters, 44(17), 9048-9058.

Huang X., Stevenson, S., Hall, A.D. (2020a) Future warming and intensification of precipitation extremes: A “double whammy” leading to increasing flood risk in California. Geophysical Research Letters, 47:e2020GL088679

Huang X., Swain D.L., Hall A.D. (2020b) Future precipitation increase from very high resolution ensemble downscaling of extreme atmospheric river storms in California. Science Advances. 6:eaba1323 doi:10.1126/sciadv.aba1323

Ice, G. G., Neary, D. G., & Adams, P. W. (2004). Effects of Wildfire on Soils and Watershed Processes. Journal of Forestry, 102(6), 16-20. doi:10.1093/jof/102.6.16

Julien, P. Y., Saghafian, B., & Ogden, F. L. (1995). RASTER-BASED HYDROLOGIC MODELING OF SPATIALLY-VARIED SURFACE RUNOFF1. JAWRA Journal of the American Water Resources Association, 31(3), 523-536. doi:https://doi.org/10.1111/j.1752-1688.1995.tb04039.x

Kean, J. W., & Staley, D. M. (2021). Forecasting the Frequency and Magnitude of Postfire Debris Flows Across Southern California. Earth's Future, 9(3), e2020EF001735. doi:https://doi.org/10.1029/2020EF001735

Leopold, L. B., M. G. Wolman, and J.P. Miller, Fluvial Processes in Geomorphology, W. H. Freeman, New York, 1964

McCormick, B. C., Eshleman, K. N., Griffith, J. L., & Townsend, P. A. (2009). Detection of flooding responses at the river basin scale enhanced by land use change, Water Resour. Res., 45, W08401, doi:10.1029/2008WR007594.

McGuire, L. A., Rengers, F. K., Kean, J. W., and Staley, D. M. (2017), Debris flow initiation by runoff in a recently burned basin: Is grain-by-grain sediment bulking or en masse failure to blame?, Geophys. Res. Lett., 44, 7310– 7319, doi:10.1002/2017GL074243.

McGuire, L. A., & Youberg, A. M. (2020). What drives spatial variability in rainfall intensity-duration thresholds for post-wildfire debris flows? Insights from the 2018 Buzzard Fire, NM, USA. Landslides, 17(10), 2385-2399. doi:10.1007/s10346-020-01470-y

McMichael, C. E., & Hope, A. S. (2007). Predicting streamflow response to fire-induced landcover change: implications of parameter uncertainty in the MIKE SHE model. J Environ Manage, 84(3), 245-256. doi:10.1016/j.jenvman.2006.06.003

Neiman, P. J., Ralph, F. M., Wick, G. A., Lundquist, J. D., & Dettinger, M. D. (2008). Meteorological Characteristics and Overland Precipitation Impacts of Atmospheric Rivers Affecting the West Coast of North America Based on Eight Years of SSM/I Satellite Observations, Journal of Hydrometeorology, 9(1), 22-47. Retrieved May 13, 2022, from https://journals.ametsoc.org/view/journals/hydr/9/1/2007jhm855_1.

Oakley N.S., Lancaster J.T., Kaplan M.L., Ralph F.M. (2017) Synoptic conditions associated with cool season post-fire debris flows in the Transverse Ranges of southern California Nat Hazards 88:327-354 doi:10.1007/s11069-017-2867-6

Oakley N.S., Cannon F., Munroe R., Lancaster J.T., Gomberg D., Ralph F.M. (2018) Brief Communication: Meteorological and climatological conditions associated with the 9 January 2018 post-fire debris flows in Montecito and Carpinteria, California, USA Nat Hazards Earth Syst Sci 18:3037-3043 doi:10.5194/nhess-18-3037-2018

Oakley, N.S. (2021). A warming climate adds complexity to postfire hydrologic hazard planning. Earth's Future, 9, e2021EF002149. https://doi.org/10.1029/2021EF002149

Ogden, F.L. (1997). CASC2D reference manual. Department of Civil & Environmental Engineering, University of Connecticut, Storrs.

Polade, S.D., Gershunov, A., Cayan, D.R, Dettinger, M.D., & Pierce, D.W. (2017) Precipitation in a warming world: Assessing projected hydro-climate changes in California and other Mediterranean climate regions Scientific Reports 7:10783 doi:10.1038/s41598-017-11285-y

Qi, Y., Martinaitis, S., Zhang, J., & Cocks, S. (2016). A real-time automated quality control of hourly rain gauge data based on multiple sensors in MRMS system. J. Hydrometeor., 17, 1675–1691, https://doi.org/10.1175/JHM-D-15-0188.1.

Ralph, F. M., Neiman, P. J., & Wick, G. A. (2004) Satellite and CALJET aircraft observations of atmospheric rivers over the eastern North Pacific Ocean during the winter of 1997/98. Mon. Wea. Rev., 132, 1721–1745, https://doi.org/10.1175/1520-0493(2004)132<1721:SACAOO>2.0.CO;2.

Rengers, F.K., McGuire, L.A., Kean, J.W., Staley, D.M. and Hobley, D.E.J., 2016. Model simulations of flood and debris flow timing in steep catchments after wildfire. Water Resources Research, 52(8), pp.6041-6061.

Rulli, M.C., & Rosso, R. (2007). Hydrologic response of upland catchments to wildfires. Advances in Water Resources, 30(10), 2072-2086. doi:https://doi.org/10.1016/j.advwatres.2006.10.012

Scheip, C.M., & Wegmann, K.W. (2021). HazMapper: A global open-source natural hazard mapping application in Google Earth Engine. Natural Hazards and Earth System Sciences, 21(5), 1495-1511.

Tang, H., McGuire, L. A., Rengers, F. K., Kean, J. W., Staley, D. M., & Smith, J. B. (2019a). Developing and testing physically based triggering thresholds for runoff-generated debris flows. Geophysical Research Letters, 46, 8830– 8839. https://doi.org/10.1029/2019GL083623

Tang, H., McGuire, L. A., Rengers, F. K., Kean, J. W., Staley, D. M., & Smith, J. B. (2019b). Evolution of debris-flow initiation mechanisms and sediment sources during a sequence of postwildfire rainstorms. Journal of Geophysical Research: Earth Surface, 124, 1572– 1595. https://doi.org/10.1029/2018JF004837

Tognacca, C., Bezzola, G.R., & Minor, H.E. (2000). Threshold criterion for debris-flow initiation due to channel-bed failure. Debris-flow hazards mitigation: Mechanics, prediction and assessment (pp. 89-97).

Young AM, Skelly KT, Cordeira JM (2017) High-impact hydrologic events and atmospheric rivers in California: An investigation using the NCEI Storm Events Database Geophys Res Lett 44:3393-3401 doi:10.1002/2017GL073077

Zhang, J. et al. (2011). National Mosaic and Multi-Sensor QPE (NMQ) system: Description, results, and future plans. Bull. Amer. Meteor. Soc., 92, 1321–1338, https://doi.org/10.1175/2011BAMS-D-11-00047.1.

Zhang, J., Qi, Y., Langston, C., Kaney, B., & Howard, K. (2014). A real-time algorithm for merging radar QPEs with rain gauge observations and orographic precipitation climatology. J. Hydrometeor., 15, 1794–1809, https://doi.org/10.1175/JHM-D-13-0163.1.

Zhang, J. et al. (2016). Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. Bull. Amer. Meteor. Soc., 97, 621–638, https://doi.org/10.1175/BAMS-D-14-00174.1.

Zhu, Y., & Newell, R. E. (1998). A Proposed Algorithm for Moisture Fluxes from Atmospheric Rivers, Monthly Weather Review, 126(3), 725-735. Retrieved May 14, 2022, from

https://journals.ametsoc.org/view/journals/mwre/126/3/1520-0493_1998_126_0725_apafmf_2.0.co_2.xml