Response to reviewers

RC1

This work explores the relationships between wildfires and landslide susceptibility in various regions of the world. The results outline the complexity of these relationships and have permitted to derive some conclusions on the smaller amounts of precipitation needed for landslide triggering in burned areas and on the seasonal shift in landslides occurrence. I am reporting below some suggestions for paper revision.

We address the reviewer's concerns below:

How were the study regions selected? Since the availability of data on both vegetation fires and landslides is fundamental in the choice of the study areas, one could ask why other regions where such data are available, for instance, Europe and Australia, were not considered.

We thank the reviewer for this question. The data from Europe and Australia were excluded because only a very small percentage of the landslides in these regions could be identified as recently burned. We add the following text to clarify this aspect of the study region selection:

> Regions were determined using the AGglomerative NESting (AGNES) hierarchical clustering algorithm (Kaufman and Rousseeuw, 2009) considering the latitude and longitude of the landslides, and clusters were subsequently combined, split, or eliminated on the basis of sample sizes as described below. First, the cluster tree was truncated at 30 clusters, after which all the clusters with fewer than 100 data points or less than 5% burned sites were eliminated. ***Notably, two commonly studied regions for landslides - Europe and Australia (e.g. Van Den Eekhaut, 2020; Nyman, 2011) – were eliminated due to a lack of verifiable post-wildfire landslides available in the GLC.*** Cases where two nearby regions with lower numbers of landslides, for example, Central America and Caribbean/Venezuela, were joined manually. Finally, the largest region, encompassing Western US and Canada, was split into three sub-regions based on an additional identical clustering process over this sub-domain. The final regions are shown in Fig. 1panel (a). The Pacific Northwest of North America was included even though the percentage of burned sites is lower than threshold, but at 4.4% it was nearly double the highest percentage among the eliminated regions 2.25% in the Eastern US). Some landslides were not included in any of the final regions.

> These events were not, however, eliminated from any analysis of all landslides.

Nyman, P., Sheridan, G. J., Smith, H. G., & Lane, P. N. (2011). Evidence of debris flow occurrence after wildfire in upland catchments of south-east Australia. *Geomorphology*, *125*(3), 383–401. https://doi.org/10.1016/j.geomorph.2010.10.016
Van Den Eeckhaut, M., & Hervás, J. (2020). State of the art of national landslide databases in Europe and their potential for assessing landslide susceptibility, hazard and risk. *Geomorphology*, *139–140*, 545–558. https://doi.org/10.1016/j.geomorph.2011.12.006

Although this paper deals with rainfall-triggered landslides, other factors that influence the occurrence of landslides - e.g. earthquakes - could be mentioned, even if only to clarify that these factors are not relevant in the study regions and the considered years.

We thank the reviewer for this observation. The following section clarifies that landslide sites were excluded if they were marked as related to other factors such as earthquakes or snowmelt:

> To reduce errors resulting from including a variety of types of rainfall-triggered landslides within the same dataset, the selected landslides were limited to those categorized by a `landslide trigger' value of `rain,' `downpour,' `flooding,' or `continuous rain.' *Landslides with a second trigger such as an earthquake were eliminated.* Snowmelt-driven landslides were also not included because the impact of precipitation is delayed in those cases -- an analysis of the snow record in California/Nevada revealed only a single event with enough antecedent snow to suggest it could have been mislabeled.

Although it focuses on a specific issue and a particular type of mass movement, the work by Riley et al. (2013) on the frequency-magnitude relationships of debris flows could be mentioned in the introduction and/or in the discussion as it compares fire-related and non-fire related debris flows at the global scale. *Riley KL, Bendick R, Hyde KD, Gabet EJ. 2013. Frequency-magnitude distribution of debris flows compiled from global data, and comparison with post-fire debris flows in the western US. Geomorphology, 191: 18–128.* https://doi.org/10.1016/j.geomorph.2013.03.008*.*

We thank the reviewer for this suggestion, and will include this reference in the introduction:

A study by Riley et al. (2013) comparing post-wildfire and non-fire-related debris flows on a global scale found that the volumes of the post-wildfire debris flows tended to be smaller. This finding suggests an increase in debris flow hazard and frequency after wildfires.

While it is important to acknowledge the problems in the quality of data, the possible occurrence of "many false positive burned landslides" mentioned in the discussion (page 20, lines 414-416) could partly undermine the results of this study. Saying that a validation of which landslides were truly post-wildfire is outside the scope of the study is a rather weak way to cope with this issue. The authors could try to better clarify which datasets are affected by these problems and delimit the extent and severity of these errors.

We thank the reviewer for this comment. We include additional analysis and discussion of the issue of false positive burned areas as follows. Firstly, we have computed the percentages of burned sites among landslides with differing location accuracies to quantify the potential error. Secondly, we propose to include additional analysis of the results, splitting the landslides into 'high accuracy' and 'low accuracy groups. The following will be added to the methods section to explain this analysis:

To explore the effects of variability in location accuracy and landslide type within the GLC, validation analyses were performed to quantify the extent of errors due to these factors. Firstly, the percentages of burned sites in each region were computed for each location accuracy. Subsequently, the results of the Mann-Whitney hypothesis tests comparing pre-landslide precipitation percentiles were duplicated splitting the data in the high- and low-accuracy groups (<=1 km and > 1 km respectively). The number of days with significantly significant differences in precipitation percentile in the 14 days prior to the landslide and 7 days are computed in each group.

The following additional figure and accompanying text will be included to address this issue (the figure number 3a is a placeholder so as not to confuse it with existing figures):

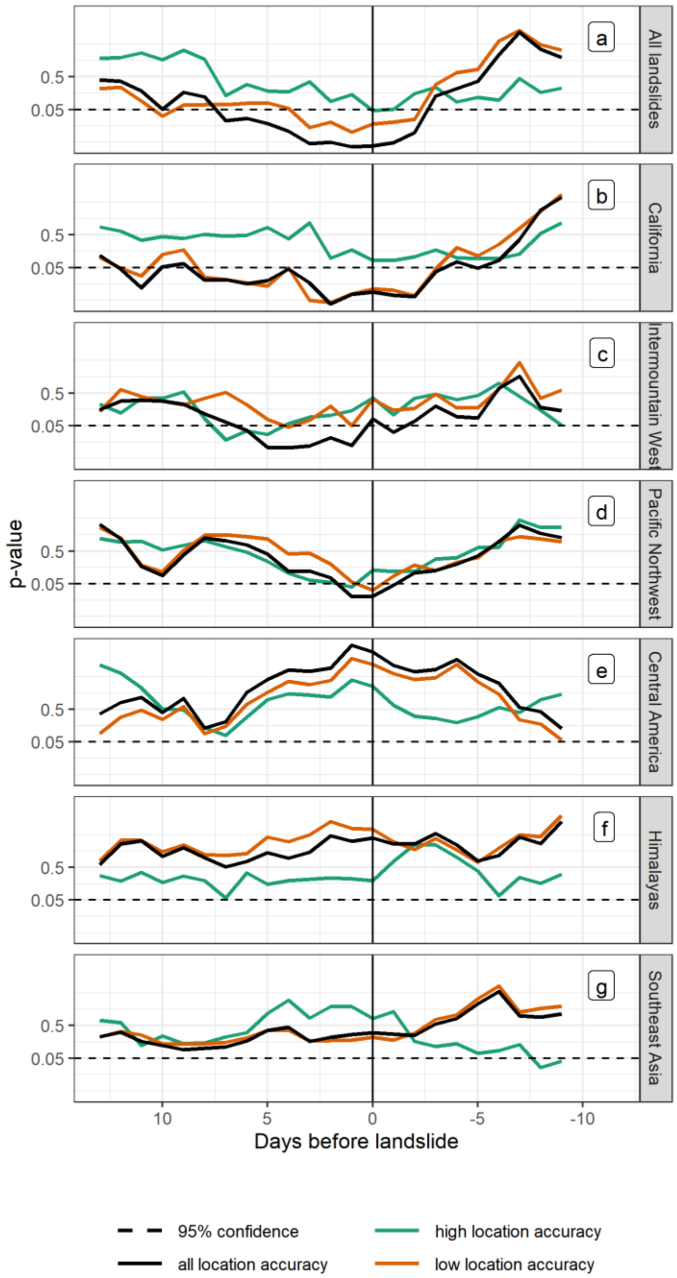

Figure 3b: p-values for Mann-Whitney hypothesis tests comparing precipitation percentiles at burned and unburned sites. The thick black line shows the p-values for all landslides, while green and orange lines show high (1 km or less) and low (greater than 1 km) location accuracies. A horizontal black line shows the p=0.05 significance threshold, while a vertical black line indicates the day of the landslide.

Figure 3b shows p-values for Mann-Whitney hypothesis tests comparing precipitation percentiles for burned and unburned groups for high and low location accuracy groups *of landslides. High accuracy indicates less than 1 km. Several regions, such as California (Fig. 3b panel (b)) show substantial differences between the high-accuracy and low-accuracy p-values. Sample sizes of burned locations among the exact locations are low, ranging from 2 to 34 in each region, with overall only 3.7% of high-accuracy landslides classified as burned (below the threshold used to exclude regions from this study). The low percentage of burned sites may partially account for high p-values among the high-accuracy group. An additional important consideration is the likelihood of a greater number of false positive burned sites among the low-accuracy group. Notably, the percentage of identified burned sites using this method increases with the location accuracy radius – globally 12.5% of low-accuracy landslides were identified as burned in contrast with only 3.7% of high-accuracy landslides.*

Finally, we will expand the discussion:

Low landslide location accuracy and lower number of burned landslides may have also contributed to the lack of conclusive results in the Pacific Northwest, Southeast Asia and Central America. The regions outside the US and Canada tended to have less accurate landslide locations. *Furthermore, less accurate locations were also more likely to be marked as burned, with a threefold increase in the percentage of landslides identified as burned between high- and low-accuracy groups.* This occurs because larger landslide radii were more likely to contain burned area by chance alone, and hence become `false positive' post-wildfire landslides, i.e.~landslides that occurred nearby but not coincident to a burned area. This idea is supported by the lower cumulative burned fractions within the regions outside the US and Canada (see Fig. 1 panels (c) and (d)). *Though landslide accuracy in the GLC is an approximate measure, introducing the possibility of false negative unburned sites, false positive post-wildfire landslides nonetheless represent an important potential source of uncertainty in this analysis. These uncertainties introduce the possibility that some of differences in triggering precipitation percentiles between burned and unburned sites may be related to unique qualities of fire-prone areas rather than fire itself.* The degree to which fires and landslides are statistically linked also contributes to the rate of false positives. Some regions may have many false positive burned landslides because there was a larger percentage of low accuracy locations, or alternatively because there was no significant increase in the probability that a landslide would occur in a burned location. Such a low posterior landslide probability given that a fire has occurred would tend to greatly increase the number of false positive burned areas by decreasing the probability that a landslide occurred in the burned section of the landslide radius, thus negating the effects of larger landslide buffers. Future studies using visible and other satellite imagery to pinpoint landslide locations and dates could help further clarify the post-wildfire posterior landslide probability by essentially eliminating the location error.

Caption of Fig.3: it could be specified that the grey belt corresponds to the day of landslide occurrence.

The caption of Fig. 3 will read:

> Seven-day precipitation percentile in the lead-up to landslides for all landslides in (a) and for the six individual regions labeled (b)--(g), whether classified as part of one of the regions or not. *The day of the landslide is indicated with a vertical grey column.* Days where a significant difference was found between the burned and unburned groups are indicated in bold coloring (Mann--Whitney hypothesis test, $p > 0.05$).

The caption of Fig. 4 is very long and not easy to follow: I wish to suggest moving part of it to the text of the manuscript.

The following modifications have been made to the caption and text related to Fig. 4:

> Figure 4: *p-values of Mann--Whitney hypothesis tests comparing landslide-triggering precipitation relative to 100 bootstrapped samples (n~100 for each sample) drawn from a 38-year precipitation record from the landslide locations. The y-axes are shown with a probit transform to expand the section of the axis where p-values are below 0.05 (significant at 95% confidence, shown as a dashed black line). The y-axis has also been inverted so that larger differences in precipitation (lower p-values) are higher on the y-axis for consistency with the percentile plots in Fig. 3. In panels (h)-(u), an example of the kernel density estimate (kde) for day-of-landslide precipitation in black separated by burned and unburned groups is compared with kdes of all bootstrapped samples in orange (burned group) or purple (unburned group).*
>
> Figure 4 highlights the increase in precipitation in the days before a landslide relative to historical amounts for that location and time of year, i.e., relative to climatology, offering a robust assessment of the landslide precipitation departure. The Mann--Whitney p-values comparing the precipitation record on each day to each of the (~100) samples are shown

in \ref{fig:bootstrap} panels (a)--(g). *Landslide events have been split into burned and unburned groups (shown in orange and purple respectively) for six regions and for all landslides in the study. Bootstrapped samples were drawn from the same DOY and locations as the landslides but from a randomly selected year. In panels (a)-(g), box plots of p--values represent the degree to which the landslide-triggering precipitation differed from climatological precipitation with lower values indicating a larger difference between the two precipitation distributions.*

Examples of the kernel density estimates of each bootstrap sample as compared to the precipitation on the day of the landslide are shown in Fig. 4 panels (h)--(u) to better illustrate the comparisons made by the hypothesis tests in panels (a)--(g). Each orange or purple curve was tested against the black curve to obtain the boxplots of p-values at 0 days before the landslide.

RC2

This manuscript pulls in several interesting global datasets to try to add more data and a global perspective to the existing literature on wildfire and landslides. Currently, there are a few relatively large challenges for the manuscript that lead to a lack of clarity, generally. I will point out several of these challenges and potential solutions that might help the authors to refine their description to enhance clarity and ultimately usability of the results.

The first challenge is that the authors do not differentiate between landslides and debris flows following wildfire. This is problematic because there is a very large body of work that exists on post-wildfire debris flows, and a smaller, but important body of work on post-fire landsliding. I would highly encourage the authors to make this distinction using terminology such as the Varnes 1978 classification. The reason this is important is because the mechanisms that generate these different types of mass movement are very different and occur at very different times following wildfire. For example, post-wildfire debris flows typically happen in the first year after a fire and they are generated by distributed overland flow that coalesces into channels and mobilizes sediment (see for example McGuire 2017 and references therein). By contrast, shallow landsliding often happens decades after fire due to soil saturation and loss of root cohesion (e.g. Jackson and Roering, 2009 and references therein). These mechanisms are nearly polar opposite, in that the first is generated by very low infiltration after fire, the second is generated during a condition of very high infiltration after fire. Lumping

these two types of mass movement together makes it extremely confusing for readers to put your precipitation analysis into the proper context. Even though debris flows and shallow landslides both move rock and sediment and involve some water, most of the erosion by debris flows happens in channels whereas most of the erosion from shallow landslides happens on hillslopes. This is sort of like saying that bread and dog biscuits are similar because they involve grain and baking, but functionally, they are extremely different. Consequently, if you could clarify what types of mass movement you are focusing on, that would go a very long way to improving the current manuscript.

We appreciate this concern. The largest category of landslides included in the NASA GLC is labeled 'landslide', and includes mass movements of all types. We acknowledge that the lack of differentiation as to the types of mass movements is of concern with this data source, and will include the following additional analysis to highlight this issue:

Subsequently, the results of the Wilcox tests comparing pre-landslide precipitation percentiles are duplicated splitting the data in the high- and low-accuracy groups (<=1 km and > 1 km respectively). The number of days with significantly significant differences in precipitation percentile in the 14 days prior to the landslide and 7 days are computed in each group. Finally, a similar analysis compared debris flows (labeled as 'debris flow' or 'mudslide' in the GLC) and other types of mass movements.

In addition, we include an additional figure and analysis as described comparing the day-of-landslide precipitation percentile from the undifferentiated 'landslide' group with landslide specifically labeled as 'mudslide' or 'debris flow':

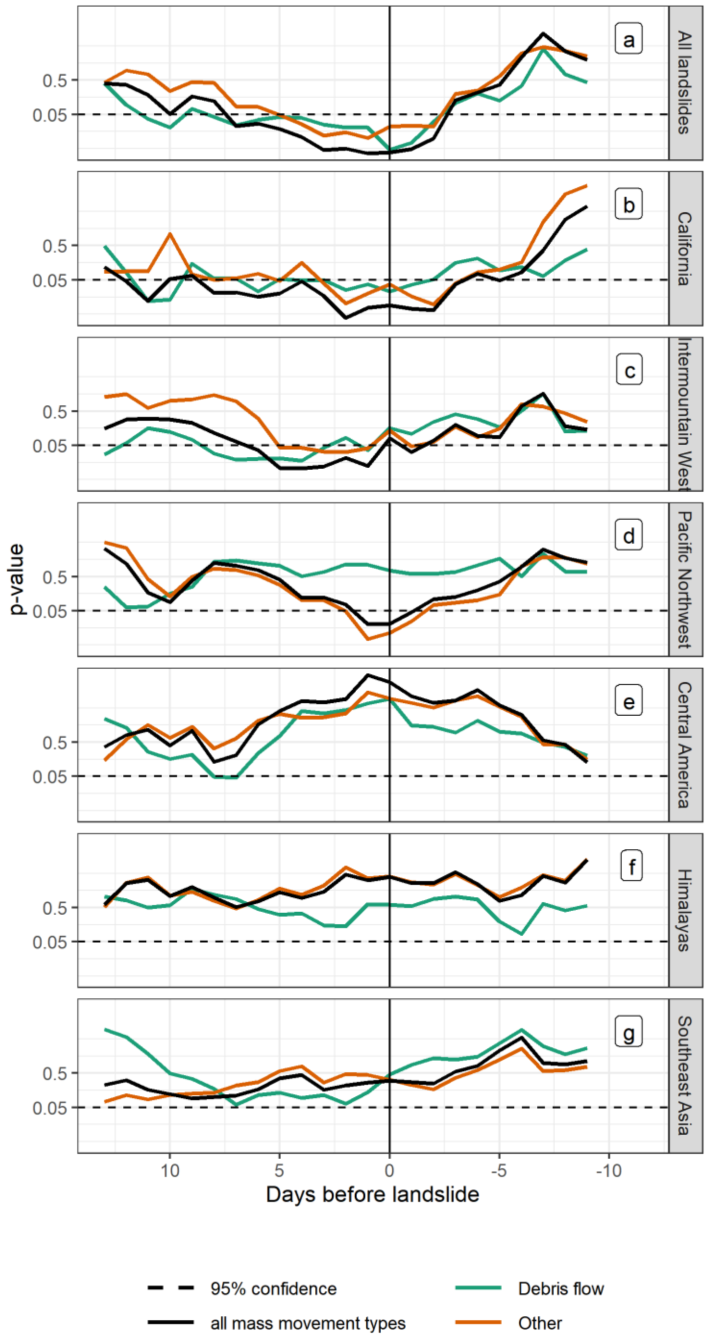

Figure 3a: p-values of Mann-Whitney tests comparing landslide-triggering precipitation percentiles at burned and unburned sites. The black line shows results for all landslides, while debris flows and other mass movements are shown in green and orange respectively. A horizontal black line shows a 95% confidence level for the hypothesis test, and a vertical black line indicates the day of the landslides

Figure 3a shows the p-values of Mann-Whitney hypothesis tests, similarly to those performed for Fig. 3. The results in Fig. 3a are split into categories by landslide type, with 'debris flow' and 'mudslide' landslide types labeled as debris flows and all other types labeled as other. With the exception of the Pacific Northwest (Fig. 3a panel (d)), the landslide type has limited impact on the number of days with significant differences ($p < 0.05$) in precipitation in the 14 days prior to the landslide in regions with any such significant differences. For example, in California (Fig. 3a panel (b)), nine days have a statistically significant difference for both groups. In the Intermountain West eight days have a statistically significant difference for debris flows while similarly six days have a statistically significant difference for other types of mass movements.

The second challenge is the imprecision in the spatial location of your landslide database. Currently you are using a 10km buffer to see if there are burned areas near the landslide. In the case of shallow landslides, that can be extremely small (on the order of 10-100 m in cross-hillslope width if you are talking about true landslides and not debris flows). A buffer of 10km will often be much larger than a wildfire perimeter therefore it would be very easy to accidently confuse an unburned landslide with a burn area, resulting in spurious conclusions. Moreover, in many studies that focus on true landslides after fire, the rainstorms that trigger slides in burn areas also trigger slides in unburned areas (See for example: Meyer et al., 2001). I suggest that you carve out a small case-study to convince readers that you have a handle on the location or can quantify the uncertainty. If you can use a subset of the data with very well known locations and show the applicability at a known location with post-fire landsliding I think this will help people to trust the generalizations you make.

We appreciate this concern. Additional analysis will be included to explore the magnitude of the uncertainty introduced by location errors, as described in the methods section:

To explore the effects of variability in location accuracy and landslide type within the GLC, validation analyses were performed to quantify the extent of errors due to these factors. Firstly, the percentages of burned sites in each region were computed for each location accuracy. Subsequently, the results of the Mann-Whitney hypothesis tests comparing pre-landslide precipitation percentiles were duplicated splitting the data in the high- and low-accuracy groups (<=1 km and > 1 km respectively). The number of days with significantly significant differences in precipitation percentile in the 14 days prior to the landslide and 7 days are computed in each group.

The following additional figure and accompanying text will also be included to address this issue (the figure number 3a is a placeholder so as not to confuse it with existing figures):

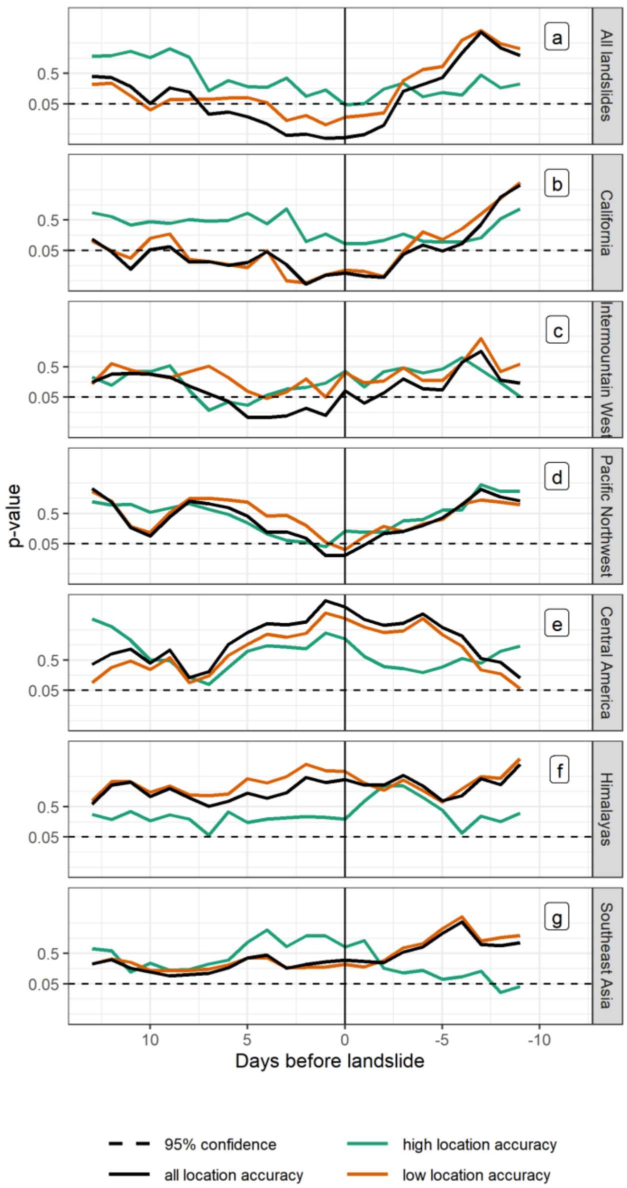

Figure 3b: p-values for Mann-Whitney hypothesis tests comparing precipitation percentiles at burned and unburned. The thick black line shows the p-values for all landslides, while green and orange lines show high (1 km or less) and low (greater than 1 km) location accuracies. A horizontal black line shows the p=0.05 significance threshold, while a vertical black line indicates the day of the landslide.

Figure 3b shows p-values for Mann-Whitney hypothesis tests comparing precipitation percentiles for burned and unburned groups for high and low location accuracy groups of landslides. High accuracy indicates less than 1 km. Several regions, such as California (Fig. 3b panel (b)) show substantial differences between the high-accuracy and low-accuracy p-values. Sample sizes of burned locations among the exact locations are low, ranging from 2 to 34 in each region, with overall only 3.7% of high-accuracy landslides classified as burned (below the threshold used to exclude regions from this study). The low percentage of burned sites may partially account for high p-values among the high-accuracy group. An additional important consideration is the likelihood of a greater number of false positive burned sites among the low-accuracy group. Notably, the percentage of identified burned sites using this method increases with the location accuracy radius – globally 12.5% of low-accuracy landslides were identified as burned in contrast with only 3.7% of high-accuracy landslides.

Finally, we will expand the discussion:

Low landslide location accuracy and lower number of burned landslides may have also contributed to the lack of conclusive results in the Pacific Northwest, Southeast Asia and Central America. The regions outside the US and Canada tended to have less accurate landslide locations. *Furthermore, less accurate locations were also more likely to be marked as burned, with a threefold increase in the percentage of landslides identified as burned between high- and low-accuracy groups.* This occurs because larger landslide radii were more likely to contain burned area by chance alone, and hence become `false positive' post-wildfire landslides, i.e.~landslides that occurred nearby but not coincident to a burned area. This idea is supported by the lower cumulative burned fractions within the regions outside the US and Canada (see Fig. 1 panels (c) and (d)). *Though landslide accuracy in the GLC is an approximate measure, introducing the possibility of false negative unburned sites, false positive post-wildfire landslides nonetheless represent an important potential source of uncertainty in this analysis. These uncertainties introduce the possibility that some of differences in triggering precipitation percentiles between burned and unburned sites may be related to unique qualities of fire-prone areas rather than fire itself.* The degree to which fires and landslides are statistically linked also contributes to the rate of false positives. Some regions may have many false positive burned landslides because there was a larger percentage of low accuracy locations, or alternatively because there was no significant increase in the probability that a landslide would occur in a burned location. Such a low posterior landslide probability given that a fire has occurred would tend to greatly increase the number of false positive burned areas by decreasing the probability that a landslide occurred in the burned section of the landslide radius, thus negating the effects of larger landslide buffers. Future studies using visible and other satellite imagery to pinpoint landslide locations and dates could help further clarify the post-wildfire posterior landslide probability by essentially eliminating the location error.

The third challenge is timing of the landslide database that you are using with respect to the wildfire. The issue of timing cross-cuts the first challenge. We know that, in general, shallow landslides happen several years after a wildfire and post-fire debris flows happen very soon after a wildfire, but you show the timing of the landsliding in any of your plots so it is very hard to analyze the how precipitation forcing should work based on the differences in those landslides with respect to time since fire. Consequently, explicitly analyzing time since fire will go a long way to helping readers to understand how to interpret your data.

We thank the reviewer for this suggestion. We have compiled an additional figure that takes into account the delay between fire and landslide (less than or greater than 1 year):

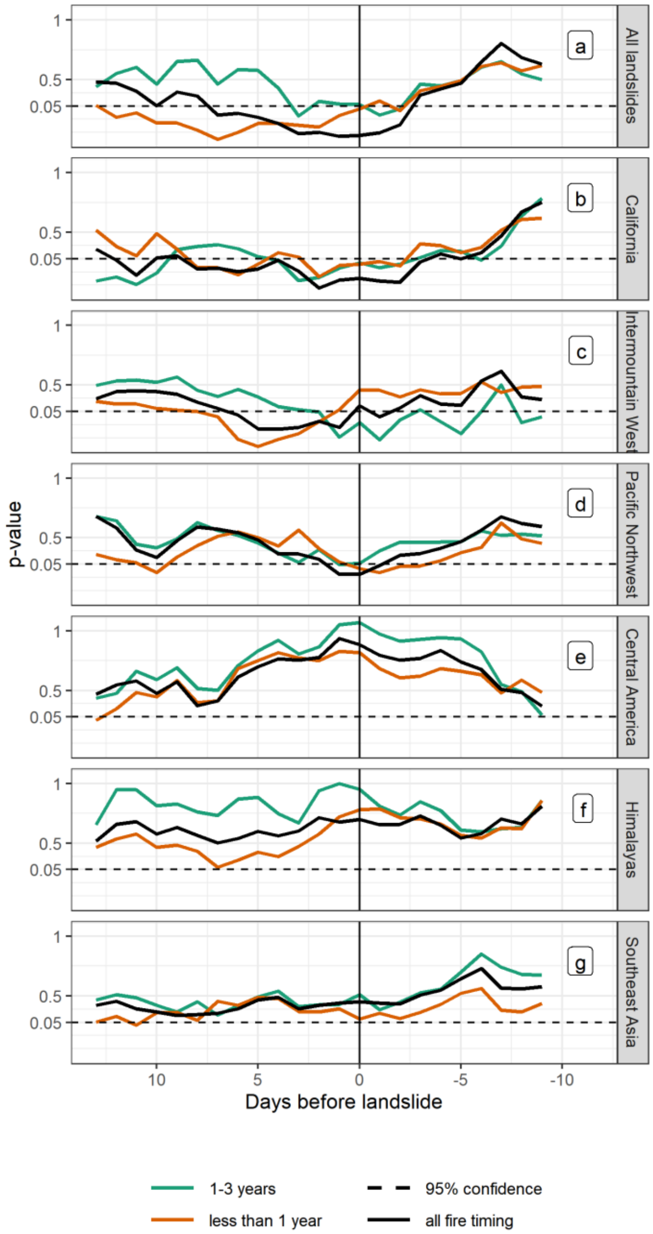

Figure 5a: *p-values for Mann-Whitney hypothesis tests comparing precipitation percentiles at burned and unburned sites. The thick black line shows the p-values for all landslides, while orange and green lines show landslides occurring within one year of a wildfire and between one and three year of a wildfire respectively. A horizontal black line shows the p=0.05 significance threshold, while a vertical black line indicates the day of the landslide.*

The following text will be inserted in the Results section to describe this figure (note that the figure number is a placeholder to avoid confusion with existing figures):

Figure 5a shows the p-values of Mann-Whitney tests comparing precipitation percentiles of groups of mass movements with different timing relative to wildfire with precipitation percentiles of mass movements at unburned sites. Landslides at burned sites were divided into two groups: within one year after a wildfire, landslide between one and three years after a wildfire. In California and the Pacific Northwest of the US (Figure 5a panels (b) and (d)), the p-values are similar among the two timing groups. By contrast, in the Intermountain West of the US (Figure 5a panel (c)), the lower precipitation percentiles at burned sites are only statistically significant at the time of the for landslides occurring 1-3 years after a wildfire. However, precipitation is significantly lower in the 'less than one year' group in the seven-to-three days before the landslide. In Central America, the Himalayas, and Southeast Asia (Figure 5a panels (e), (f), and (g)), differences between burned and unburned sites are not statistically significant for either group.

The following text will be inserted into the Discussion:

The timing of landslides relative to wildfire may also influence the magnitude of triggering storms. While in some regions, such as California and the Pacific Northwest, timing does not have a major impact on precipitation percentile differences, the Intermountain West of the US displays two distinct behaviors depending on the timing of landslides relative to wildfire. In the year immediately after a fire, the precipitation percentile is lower than for landslides at unburned locations in the seven-to-three days before the landslide, before rising to match precipitation percentile at unburned locations (see Figure 5a panel (c)). This pattern matches the result from Figure 6 panel (c) in which post-wildfire landslides in this region appear to manifest as a large storm preceded by a period of infrequent precipitation. In contrast, timing appears to make little difference to the precipitation percentile in other regions

A final general comment is that some of the precipitation analysis is very vague for readers unfamiliar with the type of data you are using. For example, you often refer to changes in percentiles, but often it isn't clear what the precipitation is a percentile of? Is it the percentile of the max 7 day rainfall, the max rainfall in a 38 year record, or something else. More detail in explaining your methods would really help readers. Similar comment for the figures. Many of the figures are missing axis labels or labeled tick marks like the inset figures in Figure 1, Figure 4 h-u, and Figure 5.

We thank the reviewer for this suggestion and have clarified the methods section below:

First, the seven-day running total precipitation depth percentile for the 30 days surrounding the day of the year and across the total 38-year record (see Sect. 2.4) was used as a proxy for landslide susceptibility.

And further details in Sect. 2.4:

Precipitation data were further processed to facilitate the comparison of landslide-triggering events across a variety of seasons and climates. The precipitation values were normalized for both location and time of year by computing a 30-day rolling percentile of the 7-day running precipitation values based on 38 years of historical precipitation climatology from 1981–2019 for each location. The percentile was computed from all the precipitation values from up to 15 days before or after the day of the year (DOY) on which the landslide occurred, and from all years in the record. This statistic controls for geographic and seasonal differences across landslide events by producing a normalized precipitation distribution that remains uniform for location and time of year. As a result, anomalous precipitation events are highlighted, facilitating the comparison of landslide triggers across locations and seasons.

In addition, revised Figures 1, 4, and 5 are included as supplements

Below I will mention several line specific comments.

24: odd ref to Shakesy and Moody here as neither of those papers deal with landslides.

We thank the reviewer for this observation, and will change the references here to (Cannon, 2010; Staley, 2018)

Cannon, S. H., Gartner, J. E., Rupert, M. G., Michael, J. A., Rea, A. H., & Parrett, C. (2010). Predicting the probability and volume of postwildfire debris flows in the intermountain western United States. *Bulletin*, *122*(1-2), 127-144.

Staley, D. M., Tillery, A. C., Kean, J. W., McGuire, L. A., Pauling, H. E., Rengers, F. K., & Smith, J. B. (2018). Estimating post-fire debris-flow hazards prior to wildfire using a statistical analysis of historical distributions of fire severity from remote sensing data. *International journal of wildland fire*, *27*(9), 595-608.

25: Do Kirshbaum and Stanley reference wildfire?

We thank the reviewer for this observation, and will revise the text as follows:

Mass movement hazards in general may also depend on dynamic factors such as soil moisture, meteorology and the length of time since the most recent fire (Kirschbaum and Stanley, 2018; McGuire et al., 2021; DeGraff et al. 2015)

DeGRAFF, J. V., Cannon, S. H., & Gartner, J. E. (2015). The timing of susceptibility to post-fire debris flows in the Western United States. Environmental & Engineering Geoscience, 21(4), 277-292.

McGuire, L. A., Rengers, F. K., Oakley, N., Kean, J. W., Staley, D. M., Tang, H., ... & Youberg, A. M. (2021). Time Since Burning and Rainfall Characteristics Impact Post-Fire Debris-Flow Initiation and Magnitude. Environmental & Engineering Geoscience, 27(1), 43-56.

56: Ebel 2012 said that ash holds much more water, not that it reduces infiltration

This sentence will be revised as follows:

A layer of post-fire ash caused by fire can also increase storage potential depending upon the thickness and hydraulic conductivity of the layer (Ebel et al., 2012)

67: There are many more up to date references you should add along with Cannon and Gartner, 2005. See refs in Moody et al., 2013; Santi and Rengers, 2020.

70: I'd add references to Pelletier and Orem, 2014

82: You are referencing papers about post-fire debris flows here, which are very different than landslides. 104: The Donnellan paper is about debris flows. To my knowledge there have not yet been landslides reported for the Thomas fire.

We thank the reviewer for the above two observations. Where possible we have used the same vocabulary as the GLC, in which 'landslide' refers to all types of rainfall-triggered mass movements. However, this terminology is misleading. We propose to replace the term 'landslide' with 'mass movement,' or 'debris flow' where specified throughout the manuscript to reduce confusion.

139: Please provide a more detailed definition of the precipitation depth percentile. 224: A 30-day rolling percentile of what?

We appreciate the above two concerns, and will provided additional details about the percentile calculation below:

First, the seven-day running total precipitation depth percentile for the 30 days surrounding the day of the year and across the total 38-year record (see Sect. 2.4) was used as a proxy for landslide susceptibility.

And further details in Sect. 2.4:

Precipitation data were further processed to facilitate the comparison of landslide-triggering events across a variety of sea-sons and climates. The precipitation values were normalized for both location and time of year by computing a 30-day rolling percentile of the 7-day running precipitation values based on 38 years of historical precipitation climatology from 1981–2019 for each location. The percentile was computed from all the precipitation values from up to 15 days before or after the day of the year (DOY) on which the landslide occurred, and from all years in the record. This statistic controls for geographic and seasonal differences across landslide events by producing a normalized precipitation distribution that remains uniform for location and time of year. As a result, anomalous precipitation events are highlighted, facilitating the comparison of landslide triggers across locations and seasons.

236: Again the median percentile of what?

The text will be modified as follows, as will any other locations where the percentile is not clarified:

The null hypothesis of the Mann–Whitney test was that the median *precipitation* percentile of the burned sites is greater than or equal to the median precipitation percentile of the unburned sites.

156: It isn't clear how you define those categories (e.g. what defines "rain" versus "downpour")

We have clarified the text below:

In order to reduce errors resulting from including a variety of types of rainfall-triggered landslides within the same dataset, the selected landslides were limited to those *labeled in the GLC* with a 'landslide trigger' value of 'rain,' 'downpour,' 'flooding,' or 'continuous rain.'

179: Previous studies say that debris flow susceptibility increases within six months of a fire, but landsliding can take many years to occur. See Benda and Dunne, 1997

Gartner et al. (2014) found that the increase in debris flow probability in a watershed due to wildfire is greatest immediately after wildfire, but can last a total of 2-5 years. *Other studies suggest that the overall mass movement hazard evolves over* *time in a more complex manner, with debris flow hazards increasing for the year after the fire followed by an increase in the frequency of shallow landslides as tree roots decay in subsequent years (Rengers et al., 2020; Benda and Dunne, 1997).*

Benda, L., & Dunne, T. (1997). Stochastic forcing of sediment supply to channel networks from landsliding and debris flow. *Water Resources Research*, *33*(12), 2849-2863.

Rengers, F. K., McGuire, L. A., Oakley, N. S., Kean, J. W., Staley, D. M., & Tang, H. (2020). Landslides after wildfire: Initiation, magnitude, and mobility. *Landslides*, *17*(11), 2631-2641.

194: You should acknowledge that severe wildfire is most common in semi arid regions. Humid regions can have fires, but the severity is limited and very few fires from humid regions result in landslides or debris flows because they don't reach very high burn severity.

We appreciate this observation, and will include the following reference to this effect:

These five studies model the probability of landslides following fire using logistic regressions to demonstrate that both burn severity (Staley et al., 2016) and burn extent within a watershed (Cannon et al. 2010) are associated with increased debris flow likelihood. Notably, burn severity and extent are both increased by drought and other low antecedent soil moisture (Westerling et al., 2003), and thus we expect to find more post-wildfire debris flows in dry climates.

Westerling, A. L., Gershunov, A., Brown, T. J., Cayan, D. R., & Dettinger, M. D. (2003). Climate and wildfire in the western United States. *Bulletin of the American Meteorological Society*, *84*(5), 595-604.

196: Would CHIRPS even pick up a storm like the NCFR that hit Montecito, CA in January 2018?

We appreciate this point, and will elaborate on the choice of the CHIRPS dataset:

Time series of precipitation at the landslide sites were obtained from the CHIRPS precipitation dataset (Funk et al., 2015). *CHIRPS is a gauge-corrected global precipitation database derived from satellite-based cloud temperature measurements.The CHIRPS dataset was chosen because of its global coverage and relatively long climatological record (1981-present). Though the ~5.5 km resolution of CHIRPS may present challenges in capturing high-intensity storms that sometimes trigger landslides (Hong et al., 2006), Gupta et al. (2019) found that CHIRPS performed well in detecting extreme precipitation across India. Furthermore, this resolution matches the 5 km resolution of the plurality of records in the GLC.* Precipitation was averaged for each landslide location within the radius of the provided location accuracy. Additional pre-processing steps described below were performed to distinguish anomalously high precipitation events from potential seasonal shifts and climatic differences across sites

215: Please provide a more detailed description of both CHIRPS and Daymet.

266: Note the wide literature that wildfire is more likely during droughts.

440: Be more specific in the length of time you are referring to when you say "a dry spell followed by a sharp uptick in precipitation" Are you talking about decadal drought, a few weeks, ?

The text will be clarified as indicated below:

In contrast, in the Intermountain West burned landslide locations appear to be characterized by a *dry spell of at least 20 days* followed by a sharp uptick in precipitation, suggesting that burned and dry soil may be the most vulnerable to extreme erosion in that region.

446: Since you don't differentiate between debris flows and landslides, it is entirely unclear how to assess your conclusion that you think landslides are caused by isolated intense thunderstorms on dry soil. Wall et al., 2020 offers a really nice overview of literature in the Pacific Northwest about true post-fire landslides (not debris flows). Note that the authors referenced therein often saw landsliding after very wet periods many years after wildfire.

Figure 3: Not sure what you mean by "bold coloring" in the caption. What makes a color bold? Also there are 6 symbols in the legend, I only see four symbols in the plot.

The caption of Figure 3 will be changed as follows:

Days where a significant difference was found between the burned and unburned groups are indicated in darker colors.

Figure 5: I am very confused on what the y-axis is supposed to represent here, it is very hard to understand what this plot is showing with the current description.

The caption of Figure 5 will be changed as follows:

DOY of landslides, DOY of fires, and the length of time in between fire and mass movement by region. *Each horizontal line represents one event, arranged on the y-axis in order of the delay between wildfire and mass movement.* Black dots on the right show the day of the year the landslide occurred, and horizontal lines represent the duration of time elapsed in between the fire and the landslide. Lines are colored by the season of the fire and are ordered by the day of the fire relative to the landslide. The black lines, or rug, at the top of each panel as well as the colored rug on the left duplicate the day-of-year of the fires to highlight seasonal patterns.

Figure 6: In the legend are the first two lines supposed to be dashed? Also in the text can you explain what exactly you are doing with the kernel density. I'm unclear on the anslysis.

We will amend the caption of Fig. 6 to explain the legend:

Precipitation frequency anomaly relative to the long-term mean aligned by the landslide date. In panels (a)(g), frequency is shown both daily and smoothed with a 90-day moving average to highlight shifts. *Daily precipitation frequency is represented as thin lines in orange and purple (burned and unburned groups) while the 90-day average is a thicker line.* The long-term mean has been removed from all the frequency curves. Landslides are in burned and unburned groups for each region separately and for all landslides. In panels (h)--(n), the kernel density estimate of landslides by the time of year is shown for both the burned and unburned groups in a radial plot.

In addition, we propose to clarify the kernel density analysis:

Figure 6 shows differences in seasonality between burned and unburned landslide seasonality on the right and the results of the precipitation frequency analysis on the left. *The kernel density estimates on the right show changes in the seasons (e.g. Fall or Winter) in which landslides at burned and unburned sites occurred. By contrast, the analysis on the left shows when landslides in each group tended to occur relative to the times of year with greater precipitation frequency.* While all regions except for Central America…

Benda L, Dunne T (1997) Stochastic forcing of sediment supply to channel networks from landsliding and debris flow. Water Resour Res 33:2849–2863

Jackson, M. and Roering, J.J., 2009. Post-fire geomorphic response in steep, forested landscapes: Oregon Coast Range, USA. Quaternary Science Reviews, 28(11-12), pp.1131-1146.

Meyer GA, Pierce JL, Wood SH, Jull AT (2001) Fire, storms, and erosional events in the Idaho Batholith. Hydrol Process 15:3025–3038

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 grainâ□□byâ□□grain sediment bulking or en masse failure to blame?. Geophysical Research Letters, 44(14), pp.7310-7319.

Pelletier, J. D., and C. A. Orem (2014), How do sediment yields from post-wildfire debris-laden flows depend on terrain slope, soil burn severity class, and drainage basin area? Insights from airborne-LiDAR change detection, Earth Surface Processes and Land- forms, 39(13), 1822–1832, doi:https://doi.org/10.1002/esp.3570.

Wall, S.A., Roering, J.J. and Rengers, F.K., 2020. Runoff-initiated post-fire debris flow Western Cascades, Oregon. Landslides, 17(7), pp.1649-1661.

Varnes, D.J., 1978. Slope movement types and processes. Special report, 176, pp.11-33.