Nat. Hazards Earth Syst. Sci. Discuss., author comment AC1
https://doi.org/10.5194/nhess-2021-111-AC1, 2021

Reply on RC1

Elsa S. Culler et al.

Author comment on "A data-driven evaluation of post-fire landslide susceptibility" by Elsa S. Culler et al., Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2021-111-AC1, 2021

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 will 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 propose to include additional analysis and discussion of the issue of false positive burned areas. 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 supplemental additional figure and accompanying text will be included to address this issue (the figure number 3b is a placeholder so as not to confuse it with existing figures):

Caption:

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.

Results

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 a major potential source of error 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 caption*: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).*

Text:

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.

Please also note the supplement to this comment:
https://nhess.copernicus.org/preprints/nhess-2021-111/nhess-2021-111-AC1-supplement.zip