Natural Hazards and Earth System Sciences Discussions

Nat. Hazards Earth Syst. Sci. Discuss., referee comment RC3
https://doi.org/10.5194/nhess-2022-21-RC3, 2022

Comment on nhess-2022-21

Anonymous Referee #3

Referee comment on "Using Sentinel-1 radar amplitude time series to constrain the timings of individual landslides: a step towards understanding the controls on monsoon-triggered landsliding" by Katy Burrows et al., Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2022-21-RC3, 2022

General comments

This manuscript presents a method for estimating the time-window of landslide occurrence based on Sentinel-1 GRD data. The method is tested against inventories of landslides with known timestamp. The underlying research question is definitely an interesting one and may also be of importance for e.g. finding timestamps to known polygons or for double-checking timestamps reported in landslide inventories. I do commend the authors as clearly a lot of work was put into the analysis.

Having worked with Sentinel-1 data for similar endeavors, the reported goodness-of-fit of results is well in line with my previous experience and expectations. However, the fact that only 1/5 of all landslides can be detected at all with reasonable accuracy indicates one of the following two conclusions for me:

- the level of maturity of the analyses is still rather low (this is not limited to the study at hand, but rather a general statement);
- SAR data is only suitable to a limited extent for the task at hand, or rather for the detection of sudden gravitational natural hazard events in general.

Overall, I think that this is an interesting contribution that is worth publishing subject to major revision. Generally, I suggest to report the results more as a potential contribution towards using Sentinel-1 data for estimating time-windows of event occurrence. Some parts of the manuscript read as if a well-working method is presented that works generically for identifying timings of landslides. However, this is still very much work in progress. For instance, I don't think that *"This will allow multi-temporal landslide inventories to be generated for long rainfall events such as the Indian summer monsoon"*

in a comprehensive manner. There will definitely be a biases in terms of identified slides, a vast majority of sildes will be missed or - worse - labelled incorrectly, and things might look dire when thinking beyond the scope of this study, e.g. if no polygons are availabe. Rather, I suggest to present the status quo and clearly highlight the limitations and highlight needs for further research based on the findings of this study. Also, the title should be clarified to indicate that time-intervals are identified rather than exact time stamps in terms of exact dates.

On a sidenote, I was slightly confused when I saw that the special issue title concerns the "Himalayan region", and study areas in this manuscript include Hiroshima and Zimbabwe. Since the Tr, BG and BK case studies are located in Nepal I think that's fine. The authors might consider adding some indication of the case study areas in the title, as I think it is actually very nice consider inventories from three different locations.

Specific comments

- Abstract: suggest to remove "thousands of", as this is somewhat unspecific without a time unit and potentially misleading.
- "Landslide locations are typically mapped using optical satellite imagery". There are many more methods that are "typically" used for such purposes, including ALS and orthophotos. The authors even mention this in section 2.1 ("drone and aerial imagery"). This might not be the case for all regions around the world as this clearly depends on the country under consideration, but the regions of interest have not yet been specified up to this point. VHR optical satellite imagery is expensive, while the spatial resolution of free data (e.g. Sentinel-2) is often too coarse to detect small slides. Free VHR data might be available e.g. through Google Earth, but not at the temporal resolution required to pinpoint the time windows to periods of some days.
- Section 2.1: I think the structure of this section can be improved. For instance:
  - l.64f: "We used three published polygon inventories of landslides whose timings are known a-priori to test and develop landslide timing methods." Since the authors continue "We filtered each inventory to remove ..." I was wondering whether there was a reference for these data sets? This point is re-established two lines later with a reference to Emberson et al. (2021), leading to some interruption of the flow from a reader's perspective.
  - l.66: "10 × 10 m SAR pixels" are mentioned. So I assume at this point that S-1 GRD data was used. Yet, the data source is unclear at this point. Also, why 20?
  - I suggest to keep methodological considerations (e.g. filtering slides < 2000m², minimum number of SAR pixels, etc) separate from the initial inventory description.
- line 74: Planetdove: Do you mean PlanetScope DOVEs?
- Please double check figure references in the body text. Fig. 1 - specifically, only Fig.1(d) - is referenced the first time on line 135. Fig. 4 is the first figure to be referenced in the text.
- It took me a while till I figured out the meaning of the terminology you used for the orbit IDs (e.g. "H083A"). Please specify more clearly that this is a combination of study area, orbit number and orbit direction. It might be helpful to include a "Hiroshima" Label in Figure 1(a), similar to 1(b) and 1(c).
- l. 110: "We tested both of these polarisations, but found VV to perform better than VH so present only the results for VV." This is an interesting finding. How was this evaluated?

- l. 123: "... geographic coordinates at a resolution of 20 x 22 m and a pixel size of 10 x 10 m". I suggest to specify this further, this statement might be confusing to an audience from the broader field of natural hazards research not familiar with (SAR) satellite data.
- l. 125: The copernicus DEM would have been a more recent DEM version, also available globally at a resolution of 30 m.
- l. 161/Figure 2: "A step change in the difference between the median landslide amplitude and the median background amplitude is then used as an indicator of landslide timing." It might be beneficial to plot this difference?
- l. 164: "When combining methods, we found that using ..." Since the other two methods have not yet been described it might be better to move such statements towards the end of your methods section, when all three methods have been properly introduced?
- 2.5 Step change identification: Up until here I was able to follow the text with re-reading some parts several times again. Here I really had to pause and ponder upon backtracking multiple times to be able to understand what is described here. Please be more concise here on how all the aforementioned methods are combined exactly, how the step function is set up and why. Some sort of graphical depiction of the workflow would probably help a lot to foster overall understanding of the whole processing pipeline.
- l. 216: reporting a baseline as reference is a good idea for putting the achieved results in context.
- Specificity is reported in table 2, F1-score is reported in Fig. 3. Providing confusion matrices of all results in the appendix might be interesting as a more detailed reference of results.
- Overall, appropriate performance metrics and their interpretation is of key importance. In fact, when thinking about the implications of the method presented here, this is crucial. If no validation data are available (e.g. when this method is applied to a new data set), a vast majority of identified dates (more precise: time windows) will be incorrect. This needs to be discussed.
- Publishing the code (e.g. on GitLab/GitHub) would be welcome for the final manuscript, but also of interest from a reviewer's perspective. If there are concerns with respect to sharing code before the publication is accepted, there are surely opportunities for embargos.

Technical comments

- Please double check Equation (1), the dot and "area" are somewhat floating around there.
- Figure 2(a): green text on green background is hardly readable.
- Figure 2(c): y-axis label is unreadable.
- Table formatting in Table 2 is off (e.g. first line - alignment of "Total landslides"). The "Asc & Desc" columns are also aligned in a confusing way. Numbers should be right-justified for better readability.
- Table 2: I suggest to split the information in the columns, and avoid combining multiple units (number and percentage, i.e. specificity) in one cell.
- Overall, I suggest to use a more consistent plotting style (including readable colorscales) throughout the manuscript.