Reviewer 2

Thank you for your thoughtful comments which have helped us to considerably improve the manuscript. We hope that the responses below, together with the modifications in the manuscript address all your concerns. We have responded to each of your comments (in bold) below.

1.NDVI based differencing approach is not new which authors also clarified in the method section. For novelty part, author have incorporated the cloud score, NDSI and temperature into the existing method. While I agree that the incorporation of NDSI and cloud score is necessary in snow covered areas (here, Gorkha), but in other areas such as in Haiti, this make things complicated. In previous studies, it has been found that vegetation recovery in earthquake affected areas take minimum 2 (Kashmir case) to more than 10 years (Wenchuan). Thus, cloud free composites of either Landsat8 or Sentinel2 images within the first or second year of event can easily be prepared in GEE platform, and should be used in cases other than Gorkha. This essentially makes two different algorithms, but I believe things will be less complicated.

We agree that: 1) NDVI differencing is not new; 2) there are some cases where snow is not a concern; 3) cloud free composites can easily be prepared in Google Earth Engine. However, we disagree that: a) it is less complicated to present two different algorithms; and b) our algorithm converges on a comparison of before and after cloud free composites. Finally we note that even this comparison of pre- and post-event composites would require decisions about stack lengths to generate each composite.

Therefore we argue that our findings are novel not only in presenting a new algorithm (albeit following closely from previous work (e.g. Behling et al., 2014; 2016; Marc et al., 2019; Scheip and Wegmann, 2021) in a way that is outlined in Sections 1 and 3.1); but also, in identifying 'behavioural' (i.e. well performing) parameter values for the key parameters that must be defined for this type of analysis; and in demonstrating that even an algorithm as simple as the one we present here can identify landslide affected pixels with comparable skill to manual mapping.

2.Further, to improve the performance of NDVI based difference approach in areas such as in Haiti, I would suggest authors to take minimum NDVI approach rather than average NDVI (i.e., minimum NDVI of the pixel of interest in the last, say, 5 years preceding the event). This approach will make sure that fresh barren surface caused by landslides have lower NDVI values than pre-event, and can be easily detectable and also helps in reduce the false positives.

This is a good suggestion. Developing a minimum NDVI based algorithm could be a fruitful avenue for future research. There is certainly a rational theoretical basis for such an algorithm and it would be interesting to compare the two approaches but doing so would involve developing a second algorithm. This would take the paper in an entirely different direction and (we feel) broaden its scope beyond that which is tractable for a single paper. Thus we do not pursue it here.

3. Although the authors have validated their method with manually delineated landslides, readers would like to know where the new approach stands when compared with other automated approaches such as HazeMapper (Scheip and Wegmann, 2020), supervised classification or machine learning techniques. These should be incorporated in discussion section.

We have added a new section (5.5 in the discussion) comparing ALADIN to these alternative approaches:

"5.5 Comparison to other automated detection methods

Automated detection of landslides typically relies on vegetation change detection and involves either generating indices of surface disturbance from which landslides can be manually identified (e.g. Scheip and Wegmann, 2020), or performing a supervised classification (e.g. Barlow et al., 2003; Behling et al., 2014; 2016; Prakash et al., 2020).

A recent example of automated surface disturbance detection, HazMapper (Scheip and Wegmann, 2020), uses similar image data (Landsat) and the same platform (Google Earth Engine) as ALADIN, but for a different purpose and using different functions to combine and transform the imagery. HazMapper is designed to generate a qualitative metric for surface change rather than a landslide-specific mapping tool. As a result, the approach does not mask snow-covered areas in case these are of interest for a user's particular application. The approach is simpler than that of ALADIN in that HazMapper calculates the NDVI difference only, rather than accounting for post-event NDVI, seasonal variability and noise in the NDVI signal for each pixel. It is currently only *applied to Landsat 7 onwards and only for individual sensors, rather than combining images from multiple Landsat sensors. This limits the events that can be examined to those occurring after 1999. However, results from HazMapper for the same study periods examined here show a good qualitative agreement with the ALADIN results. The similarity in approach, using stacks of Landsat imagery before and after a suspected trigger event, means that the two approaches will likely have many of the same strengths (e.g., the accurate georeferencing of Landsat imagery) and limitations (e.g., the coarse resolution of Landsat imagery and long wait times required to generate the post-event stack).*

Alternative approaches to landslide detection that involve supervised classification typically rely on machine learning (e.g. Prakash et al., 2020) or clustering methods (e.g. Barlow et al., 2003; Behling et al., 2014; 2016). These more complex approaches are compatible with the data and platforms that we use here. Although we have taken a simpler approach, the classification surfaces generated by ALADIN could be coupled with modern machine learning approaches to improve ALADIN's landslide detection skill. However, our results also highlight an important potential limitation to the use of supervised learning for landslide detection in general. Given the very severe disagreement between manually-mapped landslide inventories, any supervised learning method will have a very high risk of propagating gross errors into the classifier unless the training inventory is precisely co-located with the imagery used by the classifier. ALADIN could help improve existing supervised classification efforts by providing additional well-referenced landslide inventories, or by correcting existing ones."

4. Among all the inventories applied in this study, the Wang et al, 2019 (Hokkaido case) is the most recent one, and is mapped from 3 m Planet imageries. There is one more inventory available for Hokkaido case (see Dou et al. 2020) which was mapped from aerial images (less than 1 m). I would like to see this results in table 3.

We could continue to add inventories indefinitely but chose to stop at five study sites. We feel that this is sufficient. Hokkaido complicates the analysis because it opens the possibility of a Sentinel based analysis. This would require re-calibration and a more complete introduction to the properties of the Sentinel satellite and we feel that this is out of scope for the current study. We have used datasets from USGS sciencebase throughout our quantitative analysis in order to ensure consistent and traceable analysis. We include the Hokkaido dataset as an exception for illustrative purposes but do not use it in our quantitative analysis.

5.More explanation is needed on how the ALDI pixel are converted to landslide objects. I can see that in Kashmir, Aisen and Wenchuan cases, the large landslides identified by ALDI are more than manual methods (Fig. 7). Comment on this.

We now explain that the continuous ALADIN index is thresholded to generate the same FPR as the comparison inventory:

"For manually-mapped inventories this information is generally captured automatically since landslides are mapped as discrete objects rather than on a pixel-by-pixel basis. However, automated classifiers like ALADIN require additional steps to convert a continuous pixel-based classification surface to a set of landslide objects. First, we generate a binary prediction of landslide presence or absence by thresholding the ALADIN classification surface to match the manually-mapped FPR, as described above."

We then explain that landslide pixels identified by ALADIN are converted to landslide objects based on a connected components clustering:

"Second, we convert the binary landslide map to a set of landslide objects by identifying connected components at the 30 m resolution of the Landsat imagery (Haralick and Shapiro, 1992). This connected components clustering is one of the simplest of many possible clustering algorithms."

We have also added reference to specific examples in Fig. 7 in the results section to describe and explain the increased frequency of large landslides in the ALADIN-based distributions:

"However, the ALADIN-based distributions, are clearly different from those derived from manual mapping, they lack: 1) the roll-over at small areas (in all cases, Figure 7a-e); 2) the positive curvature to the right tail (particularly clear for Haiti, Figure 7d); and 3) the roll-off at very large areas (resulting in oversampling of landslides >10^5 m^2 for Wenchuan, Figure 7c)."

and

"These differences can be explained in terms of amalgamation and censoring. Amalgamation of multiple neighbouring landslides increases the frequency of large landslides, fattening the right tail (Marc and Hovius, 2015); and in some cases considerably increasing the size of the largest landslide (e.g. Aisen and Wenchuan, Figure 7b-c)."

Minor comments

1. **Xu et al 2014 inventory is having serious problem. It would be better if authors used the Fan et al., 2018 inventory for Wenchuan.**

We have used datasets available at USGS sciencebase throughout our quantitative analysis in order to ensure consistent and traceable analysis. We are not aware of an open access version of the Fan et al., 2018 inventory. However, we note that: 1) the Xu et al. (2014) inventory was collected as part of a study published in a peer reviewed journal; 2) we have not found clear evidence in the literature that the Xu et al. (2014) inventory is more seriously problematic than other inventories for Wenchuan in particular or such that it is not reflective of co-seismic inventories in general; and 3) we find that Xu et al. (2014) is consistent with ALADIN than Li et al. (2014) in some parts of the study area while the converse is true in other parts. Thus we conclude that Xu et al. (2014) is likely to contain mapping errors and that the same is true of Li et al. (2014) but that these errors are a fair reflection of the current state of the art in co-seismic landslide mapping.

2. Figure quality should be improved, A grey background or another color would hel to distinguish the ALDI 0 from positive ALDI values.

Thank you, this is useful. We have now sought to add a light green background to distinguish ALADIN 0 from other values. We chose to avoid grey because we already use grey to indicate areas that were masked during manual mapping. We have also sought to improve the quality of the maps by: 1) removing the grey transparencies in the full study area maps to make the ALADIN pattern easier to see at that scale; and 2) simplifying the mask layer in Figure 2 so that all masks are shown in grey to enable introduction of an additional colour to represent ALADIN=0.

3. Delete the Xu et al inventory from Fig 8. The offset in Xu et al inventory make the visualization difficult.

In Fig 8 the Xu et al. (2014) inventory is actually more consistent with ALADIN (and thus we argue more likely to be correct) than Li et al. (2014). Therefore it is difficult to argue for a removal of the Xu et al data without independent evidence of issues with the dataset. We have argued above that both Xu et al. (2014) and Li et al. (2014) inventories should be retained in our analysis and if that is the case then we feel that both should be displayed in Fig 8. Indeed, the striking offset between the two manually mapped inventories is one of the major points that we hope the reader will take from this figure.

4.Title can be a more relaxed one than current one. Automated mapping still have problem of delineating source and accumulation zones. Further, in automated method, separating two landslide in adjacent slopes (special case – Hokkaido) is difficult.

This is a point of agreement between both reviewers and while we suspect that this may reflect the dominance of the manual mapping paradigm in this field we have sought to relax the title, removing the claim that automated detection out-performs manual mapping, to reflect their concerns.

"Automated determination of landslide locations after large trigger events: advantages and disadvantages compared to manual mapping"

We have also added to the discussion section to expand on the limitations of ALADIN relative to manual mapping (Section 5.3) and to clarify the domains in which one out performs the other (Section 5.4).

5. Shorten the paper a little for better reading

We have shortened our description of the methods which was identified by R1 as a section with scope for shortening. We have also thoroughly edited the paper for clarity and readability. However, unfortunately we suspect the paper has expanded very slightly overall as a result of revisions to address comments from both reviewers. We are happy to consider further reductions in length, especially if the reviewer can provide more specific guidance.