Response to reviewer 3

We thank the reviewer for taking the time to review our paper. The following is a line by line response to all comments made on our manuscript (reviewer comments given in bold type).

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 doublechecking 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.

We will revise the manuscript to make it clearer throughout that our methods cannot establish timings for all the landslides in an inventory (and in fact, will only provide timings for ~20%).
The quote here is taken from the abstract. In response to this and to similar comments made by reviewer 2, we will change the end of the abstract.
Previous text: *"our methods allow 20% of landslides to be timed with an accuracy of 80%. This will allow multi-temporal landslide inventories to be generated for long rainfall events such as the Indian summer monsoon, which triggers large numbers of landslides every year and has until now been limited to annual-scale analysis."*
New text: *" our methods allow 20-30% of landslides to be timed with an accuracy of 80%. Application of our methods could provide an insight on landslide timings throughout events such as the Indian summer monsoon, which triggers large numbers of landslides every year and has until now been limited to annual-scale analysis*" .
Similarly in the conclusions section, we will change "*These methods will allow us to generate multi-temporal landslide inventories for long rainfall events, unlocking comparisons between rainfall data, hydrological models and triggered landsliding*." to "*These methods will provide information on the timings of some individual landslides and allow spatio-temporal clusters of landslides to be associated with peaks in rainfall during long rainfall events, unlocking comparisons between rainfall data, hydrological models and triggered landsliding."*

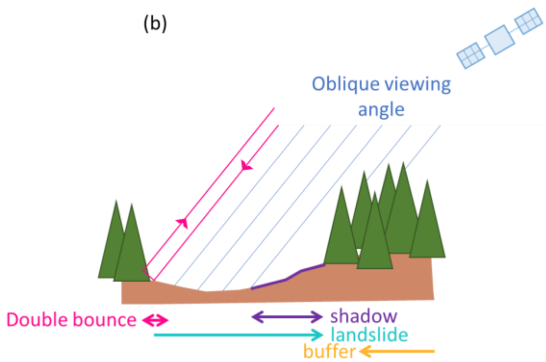
This removes the words "multi-temporal landslide inventories", which were misleading since we cannot provide a complete inventory where all the landslides have timings assigned.
We will make also make sure this is clear in the discussions / conclusions and throughout the manuscript.
We will also make sure it is clear throughout that we are not assigning specific dates, but instead time windows of (in most cases) 12 days to each landslide.

We have also made some adjustments to the methods, which somewhat improve the number of landslides for which we can assign a timing from 20% to 30% in favourable cases. This does not change the overall message of the manuscript, since we keep the same 80% accuracy and still do not assign a timing for the majority of the landslides in an inventory, but it represents an improvement to the method. All following replies and alterations to the text therefore incorporate these new sensitivity levels.

(i) A new method based on bright spots observed at the edges of landslide polygons

As described in Section 2.4.3 of our original manuscript, shadows are cast by trees at the edges of landslide scars due to the imaging geometry of the SAR sensor. (See figure below, altered from Figure 2b of the original manuscript). On the opposite side of the scar, we may observe a bright patch in the SAR amplitude images due to double bounce scattering between the exposed soil and trees on the far side of the landslide scar, and the focussing of the microwave energy into a small area (Villard and Borderies, 2007). Similarly to the shadow method, we compare the pre-event and post-event SAR time series and identify pixels which have experienced a strong increase in amplitude (we found a threshold of >5dB to perform best in this case). Incorporating this method means that, increases the final number of landslide that are assigned a timing by 3% in Hiroshima, 2% in Zimbabwe and 7% in Trishuli.

A description of this method will be added to Section 2.4 (SAR amplitude techniques for landslide timing) as " 2.4.4 Method 4: Geometric bright spots"



(ii) Increase in size of the landslide polygons

As described at lines 185-195 of our original manuscript, there may be a spatial mismatch between the optically-derived landslide polygons and the SAR imagery. Previously, we increased the size of landslide polygons by 10m (1 SAR pixel) for the geometric shadows method to try to account for this. However, this effect may also affect the results from other methods, and decrease the number of landslides for which we are able to assign a time window.

Therefore, for landslides that are not assigned a date by at least 2/4 of our methods, we now increase the size of the landslide polygon using a 20 m buffer for all methods and repeat, with the aim of trying to assign landslide timings to some of the landslides that experience this spatial mismatch. This improves the number of landslides we are able to detect in each case study event by 4% in Hiroshima and 5% in Zimbabwe and Trishuli, and will be added to the next version of the manuscript.

Since this will now be described in its own subsection in the methods section, text referring to this at lines 185-195 and 317-326 of our original manuscript will be removed.

There will definitely be a biases in terms of identified slides,
The biases towards which slides can be assigned a timing is related to section 3.3 "Factors effecting the performance of each method." For example, it is clear that larger landslides are more likely to be assigned a timing than smaller landslides using our methods (Fig. 5 d-f).
In the manuscript, we considered how this effected where our methods could be applied (for example, they will not work well for inventories of small landslides or in arid environments), but we did not consider how well the timings of the 30% of landslides we assign a timing to using SAR methods will represent the full inventory. In fact, this 30% will be biased to contain a somewhat higher proportion of large landslides and landslides in more heavily vegetated areas than the original inventory. We will add this point to the revised version manuscript

New text at line 252 of original manuscript: *“For future applications, this helps to determine the environments where our methods can be expected to work well. It also provides an insight on potential biases in terms of the subset of a landslide inventory that is assigned timings by our methods.”*

a vast majority of sildes will be missed or - worse - labelled incorrectly,

Our methods should not incorrectly label a large percentage of slides. We expect that if we apply our methods to an inventory of rainfall-triggered landslides, we will obtain an inventory in which ~70% have no timing information, ~24% are correctly timed and ~6% are incorrectly timed. We will be careful to make this clearer in the abstract, results and conclusions sections of the revised manuscript.

and things might look dire when thinking beyond the scope of this study, e.g. if no polygons are availabe.

If polygons are not available for an event, it will not be possible to apply our methods in their current form – they are not designed to be applied to events for which we do not have a pre-existing landslide inventory. We will ensure this is clear in the revised version of the manuscript.

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

The new manuscript will be given a new title that better reflects the results in the paper. For example *“Using Sentinel-1 radar amplitude time series to constrain the timing of individual landslides: a step towards understanding the controls on monsoon-triggered landsliding”*

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.

The manuscript was submitted to the special issue since, while only 1 of our 3 landslide inventories was located in the Himalayas, the methods we develop are expected to be particularly useful in Nepal due to high levels of landsliding and cloud cover during the monsoon season in that country.

We will emphasise this in the revised introduction

previous text (lines 31-34 of original manuscript): *“the Indian Summer Monsoon (June – September) triggers hundreds of landslides every year in the Nepal Himalaya and cloud-free optical satellite imagery is unlikely to be available throughout this period (Robinson et al. 2019). This limits analysis of these landslides to the annual scale (e.g. Marc et al. 2019a; Jones et al. 2021).*

new text: *“the Indian Summer Monsoon (June – September) triggers hundreds of landslides every year in the Nepal Himalaya and cloud-free optical satellite imagery is unlikely to be available throughout this period (Robinson et al. 2019). This limits analysis of these landslides to the seasonal scale and prevents association of individual landslides or landslide clusters with specific peaks in rainfall (e.g. Marc et al. 2019a; Jones et al. 2021).”*

This is also better reflected in the proposed new manuscript title.

Specific Comments

Abstract: suggest to remove “thousands of” as this is somewhat unspecific without a time unit and potentially misleading

“thousands of” will be removed in the revised version of the abstract.

"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.
We do not really consider small landslides in this study as these are difficult to detect with Sentinel-1 SAR, which has a relatively coarse resolution (once multi-looked to reduce noise) and our methods are shown in Figure 5 of the original manuscript to perform better on larger landslides.
Also, optical satellite images are also subject to cloud cover so, while they allow smaller landslides to be mapped, they do not address the problem we are tackling in this manuscript.
We will change “*typically*” here to “*often*”

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.
We will restructure this section so that we first describe the inventories and then describe how they were filtered to remove small landslides. We will also change “*10 x 10 m SAR pixels*” to “*Sentinel-1 GRD pixels, of size 10 x 10 m*”
line 74: Planetdove: Do you mean PlanetScope DOVEs?
Yes, we should have used “PlanetScope” here, not Planetdove. Thank you for correcting this.
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.
This was a mistake. At line 119 of the previous version of the manuscript, we refer to “*(e.g. Tozang landslide, Mondini et al, 2021, Fig 4)”.* “*Fig 4”* there was not intended to refer to Figure 4 of our paper, but to Figure 4 in the paper of Mondini et al. 2021. This was not clear, and in the next version of the manuscript, we will reduce this reference to *“(Mondini et al. 2021)”* only.

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.

Yes, this was also raised by reviewer 2, who suggested a change of terminology from e.g. “H083A” to “H_asc” to describe the tracks. We will make this change in the next version of the manuscript, which should also resolve this comment.

"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?

Yes it is interesting since other studies (e.g. Handwerger et al. 2022) use VH and find it to be the most successful option. We attach a version of table 2 in the original manuscript which contains the same results from VH in Zimbabwe and Hiroshima (VH data are not available for the Nepal case study, which occurred soon following the launch of Sentinel-1 before dual-pol data began to be consistently acquired – this would be a further disadvantage to using VH).

	Hiroshima		Zimbabwe		Trishuli		Buri Gandaki	Bhote Kosi			Hiroshima (VH)		Zimbabwe (VH)	
Track	H090D	H083A	Z079D	Z072A	Tr019D	Tr085A	BG019D	BK121D			H090D	H083A	Z079D	Z072A
Total landslides	543		383		650		922	1554			543		383	
non-masked	543	540	383	383	485	474	592	894			543	540	383	383
ls-b inc	137 (26%)	92 (36%)	90 (50%)	66 (41%)	106 (38%)	107 (35%)	152 (36%)	313 (36%)			94 (12%)	97 (29%)	40 (20%)	26 (12%)
ls-b dec	126 (38%)	205 (57%)	182 (23%)	155 (34%)	156 (37%)	143 (22%)	113 (27%)	310 (32%)			182 (25%)	251 (63%)	262 (28%)	266 (36%)
pix var	160 (45%)	181 (59%)	134 (50%)	83 (47%)	141 (42%)	125 (44%)	141 (30%)	261 (43%)			155 (32%)	222 (56%)	152 (16%)	110 (27%)
shadow	79 (49%)	122 (75%)	43 (55%)	58 (72%)	45 (80%)	50 (80%)	17 (88%)	52 (87%)			144 (43%)	227 (74%)	125 (39%)	140 (47%)
combined >2m	51 (75%)	99 (88%)	48 (65%)	39 (85%)	45 (76%)	33 (88%)	33 (64%)	86 (63%)			71 (54%)	166 (80%)	88 (33%)	87 (60%)
combined >3m	10 (80%)	22 (100%)	5 (80%)	2 (50%)	8 (100%)	3 (100%)	1 (100%)	5 (100%)			9 (89%)	38 (97%)	17 (35%)	9 (89%)
asc & desc	122 (80%)		81 (73%)		70 (81%)		-	-			196 (69%)		142 (44%)	
asc & desc 2m,1t	79 (73%)		64 (72%)		52 (77%)		-	-			139 (65%)		102 (23%)	
asc & desc 3m	43 (91%)		17 (76%)		18 (94%)		-	-			57 (81%)		40 (75%)	
baseline (1/n)	7%	17%	10%	7%	8%	8%	8%	14%			7%	17%	10%	7%

As you can see, we have a generally lower accuracy for VH data, especially in Zimbabwe.

Since this may be interesting for future studies, we can include it as a supplement, which we will reference at line 110 of the original manuscript.

In general, it is expected that VH should be more sensitive to changes in volume scattering (here, scattering within the forest canopy). Therefore, since the removal of vegetation due to a landslide should reduce this component, we would expect VH to decrease for a landslide polygon. Therefore, of the methods analysed here, we might expect VH to perform better than VV for the landslide-background decrease method (ls-b dec in the above table), which is in fact the case for H083A, Z079D and Z072A (but not H090D). However, since VH performs worse than VV for the other methods, it is better to use VV.
This table was calculated prior to the addition of our new method based on bright patches at the edges of landslide polygons and so this method is not included in the table. However, the double-bounce scattering component is generally very weak in cross-polarised SAR images, therefore we do not expect that this method would have been successful using VH.

The copernicus DEM would have been a more recent DEM version, also available globally at a resolution of 30 m.

We chose to use the SRTM DEM here since it is already available as a dataset in Google Earth Engine, making it easier to integrate into the slope correction module for future users.

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
Previous text: "*we used the Google Earth Engine Sentinel-1 ground range detected (GRD) data set. These data are preprocessed following the workflow of Filipponi et al. (2019) to obtain the sigma0 backscatter coefficient at a resolution of 20 x 22 m and a pixel size of 10 x 10 m*."
New text
"*we used the Google Earth Engine Sentinel-1 ground range detected (GRD) data set. These data are preprocessed following the workflow described in Filipponi et al. (2019) to obtain the backscatter coefficient sigma0 at a resolution of 20 x 22 m in radar coordinates. The data are then resampled onto a 10 m grid in geographic coordinates.*"
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?

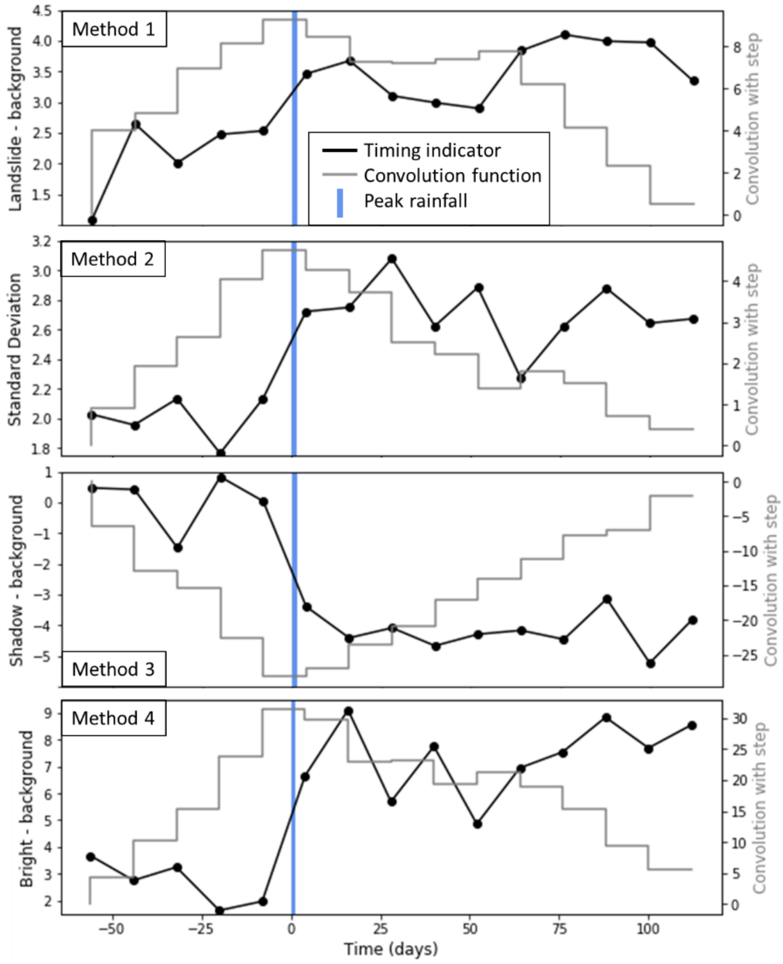


Figure 1 Example time series for each method described in Section 2.4 for a single landslide from the Hiroshima data set using SAR data from Sentinel-1 track 019D. The blue bar shows the duration of the peak rainfall associated with this event (6-7 July 2018).

Yes, this is a good point. In response to this and to similar comments from reviewer 2, we have prepared 3 panels (above) which will be incorporated into Figure 2. This makes the step change clearer (Note we have also changed to a different landslide polygon due to changes requested by reviewer 2 – these time series are thus not comparable to those in Figure 2 of the original manuscript).

These panels also show as a grey line the convolution between the method and a step function to make it clearer how a step change results in a peak or trough in the convolution function.

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?
This line will be removed here and we will describe the combination of methods later in Section 3.1

2.5 Step change identification: Up until here I was able to follow the text with rereading 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.

This Section will be rewritten in response to this and to comments made by reviewer 2 in order to make it easier to follow. Additionally, the revision of Fig. 2, which now displays the time series and convolution with the step function, should help to clarify the text.

l. 216: reporting a baseline as reference is a good idea for putting the achieved results in context.
Thank you.

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.
This was our mistake – Table-2 does not show specificity and we will alter the table caption to better describe the results.

The confusion matrix in Table 1 of the original manuscript was designed to assess how the size of the peak in the convolution function can be used to predict whether or not an assigned date is correct and to allow the calculation of the F1-score.

After the thresholds for each method have been selected and we move to combining predictions from multiple methods and tracks, there is no way to divide the landslides into TP, FP, FN, TN because we have three categories: landslides that are assigned the correct timing, landslides that are assigned an incorrect timing and landslides that are not assigned any timing.

In response to this, and to comments made by reviewer 2, we will change to showing *correct / assigned* rather than *assigned (% of assigned that are correct)* for each method in Table 2. By providing the raw numbers, any statistic can be calculated by the reader if they require it.

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.
The problem is not that a vast majority of landslides will be assigned an incorrect time window, but instead that the majority of landslides will not be assigned any time window at all. The subpart of the inventory that has timings assigned will be biased towards larger landslides and those in more vegetated areas. We will discuss this in the revised version of the manuscript. (See response to earlier comments.)

Of those landslides that are assigned a time interval, we expect that 80% of the time, this interval should be correct. Therefore if, in a new dataset, we observe a spatio-temporal cluster of landslides, we can assume the timing of this cluster is correct, since it is very unlikely that all the timed landslides in this cluster would be assigned the same incorrect date.

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.

The GEE and python codes will be shared as a supplement to the next version of the manuscript.

Technical comments

Please double check Equation (1), the dot and "area" are somewhat floating around there.
Yes, we made a mistake with the equation here in LaTeX here, the dot is in the wrong place. This will be fixed in the next version of the manuscript

Figure 2(a): green text on green background is hardly readable.
We will adjust the colours on this figure to make it easier to read

Figure 2(c): y-axis label is unreadable.
We are replacing this panel and (d) with the panels shown in our response to your comment above on line 161. We will ensure the y-axis is legible in these new panels

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 rightjustified for better readability.
Thank you for these suggestions, we will make these adjustments to the table to improve it's 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.
In response to this and to comments made on this table by reviewer 2, we will change this so that each column contains *correct / assigned* rather than *assigned (% of assigned that are correct)*.

Overall, I suggest to use a more consistent plotting style (including readable colorscales) throughout the manuscript.

We will ensure this is the case in the revised manuscript.