Improved Dating of Landslides in Zimbabwe by Combining Satellite Multispectral and Synthetic Aperture Radar Observations: Response to Reviewers

We thank the reviewers for the detailed and constructive comments on this manuscript. We have considered these comments and have revised the manuscript accordingly. Below we provide the individual comments and responses, along with manuscript changes. Review comments are provided in grey boxes, responses in **blue bold**, and manuscript changes in normal black text.

Reviewer 1 Comments

REVIEWER 1: COMMENT 1

I am concerned that there could be a problem in the methodology, and the only real way to tell is by adding another case study event. The problem is this: the authors are using SAR time series to detect a landslide based on the change in scattering properties caused by the removal of vegetation. This change should be permanent (or at least last for as long as the landslide scar remains unvegetated). However, in Figure 6 and lines 187-193, the authors do not see any sustained change in the SAR time series, they only see a peak in the amplitude. This peak in the amplitude could reflect the landslide timing (but in that case, why does the change not last longer?) or it could reflect the increase in soil moisture associated with the rainfall event. For this landslide inventory, the temporary soil moisture increase, and the landslide timing coincide, so there is no way to tell. For this reason, I strongly recommend to also test the method using earthquake-triggered landslides where there is no change in soil moisture (and make sure there is no significant precipitation event at the same time, so not Hokkaido 2018, Turkiye 2023, Haiti 2021 etc.). This will allow you to ensure that you are detecting the landslide itself and not the change in soil moisture. Without carrying out this extra validation, I don't think you have a reliable test of your method.

...

RESPONSE

We thank the reviewer for this comment, which has prompted us to further investigate whether the SAR response identified by our method could be attributed to changes in soil moisture associated with rainfall rather than to the landslide itself.

Firstly, following the other reviewer comments and clarification of the calibration dataset, the SAR input used within the method has been changed from VV alone to a combined VV + VH dataset. This resulted in a small improvement in performance

within the calibration dataset. VV and VH individually produced similar results (approximately 92% accuracy), whereas the combined VV + VH dataset achieved 93.5% accuracy. The use of the combined VV + VH dataset is discussed further in Reviewer 2, Comment 9.

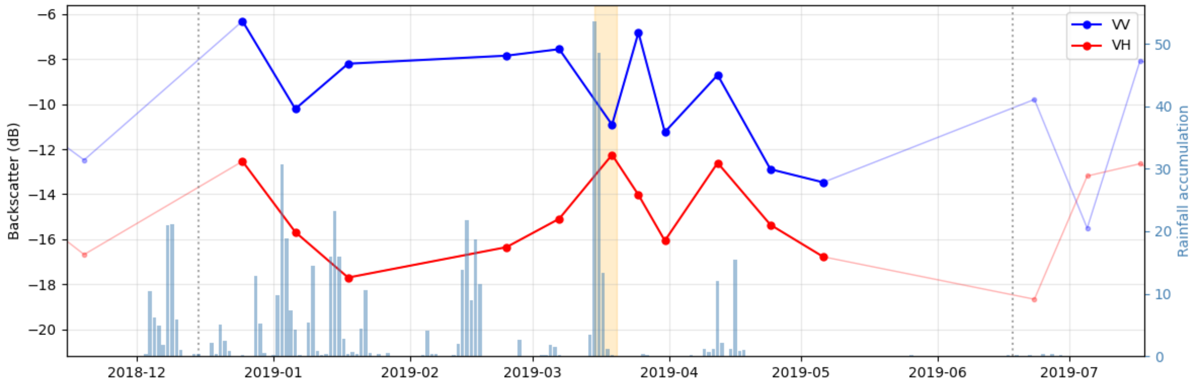
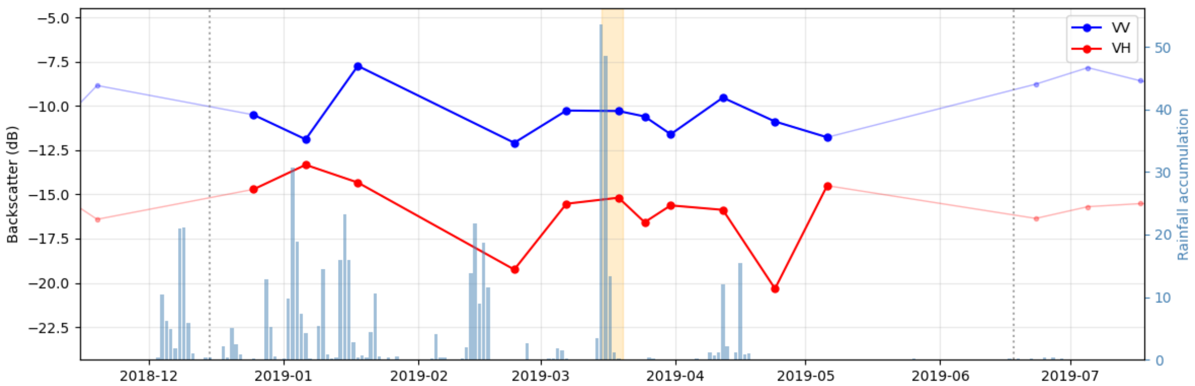


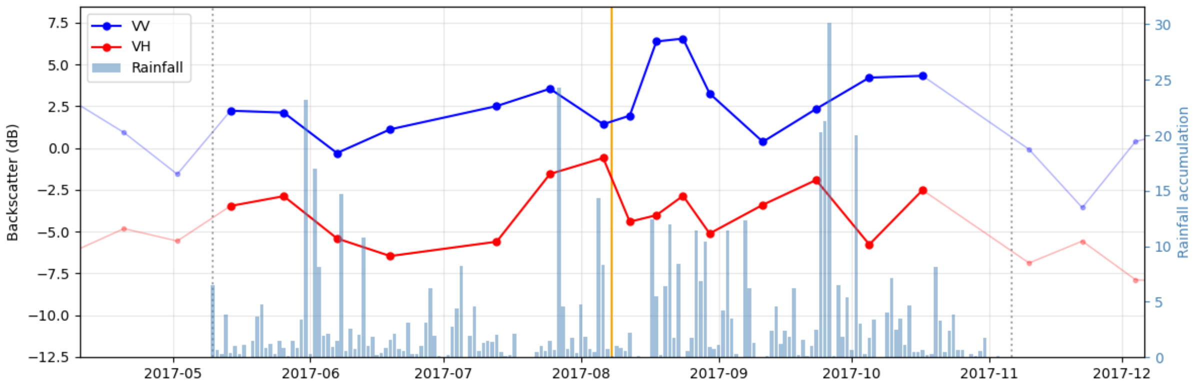
Figure 1 shows an example of a correctly dated landslide and the corresponding SAR VV and VH time series, alongside the rainfall time series for the region. A pronounced rainfall event occurs during the period in which the landslide is dated. However, several earlier rainfall events are also visible in the time series without producing a comparable response in both SAR channels. This suggests that the response identified by the method is not simply a direct response to every rainfall-driven soil moisture increase. We acknowledge, however, that because the rainfall event and landslide timing coincide in this example, this case alone cannot distinguish conclusively between a landslide-related change in scattering properties and a rainfall-induced soil moisture response.



To investigate this further, Figure 2 shows the corresponding time series for a point within the same region that is not associated with a detected landslide. The rainfall recorded at the two locations is comparable in both magnitude and timing, whereas the non-landslide point does not show a comparable response in the SAR time series during the period of the landslide identified in Figure 1. This provides

additional evidence that the pronounced response at the landslide location cannot be explained solely by the regional rainfall event. The contrasting response between the landslide and non-landslide locations suggests that local surface changes at the landslide location contribute to the SAR signal.

Importantly, we agree with the reviewer that these rainfall-associated examples cannot completely exclude soil moisture as a contributing factor, particularly because the landslide and rainfall occur at approximately the same time. We therefore additionally investigated an earthquake-triggered landslide, for which the triggering mechanism is independent of rainfall. Figure 3 shows the SAR VV and VH time series for the six months surrounding the earthquake-triggered landslide in Jiuzhaigou County, Sichuan Province, China, on 8 August 2017.



The SAR response around the earthquake-triggered landslide is similar to that observed in Figure 1, with a pronounced change in the SAR time series around the landslide date despite the absence of a comparable rainfall anomaly at that time. In contrast, rainfall events elsewhere in the six-month period do not consistently produce the same response in both VV and VH. For example, the rainfall increase in September is accompanied by a response in VH, but does not produce a comparable simultaneous response in both VV and VH. This distinction is important because the method uses the combined response of VV and VH rather than relying on a single-polarisation response. The observation that rainfall events do not consistently produce the same response across both polarisations provides additional evidence that the detected signal is not simply a response to changes from dry to wet conditions.

We agree that removal of vegetation and exposure of bare soil would be expected to produce a persistent change in surface properties, whereas soil moisture produces a predominantly transient response. We also thank the reviewer for directing us to Cabré et al. (2020), who distinguish between transient changes in InSAR coherence associated with variations in soil moisture and more persistent changes associated with modification of the surface through erosion and sedimentation. Our results do not show a simple, permanent offset in backscatter

following every landslide. Consequently, we do not interpret the transient SAR peak itself as definitive evidence of permanent vegetation removal. Rather, our method identifies an anomalous change in the SAR scattering properties that provides information on the timing of the landslide. The additional earthquake-triggered case study demonstrates that such an anomaly can occur in the absence of a rainfall triggering, reducing the likelihood that the method is simply assigning landslide timing according to the strongest soil-moisture fluctuation.

We thank the reviewer for highlighting this distinction. We agree that soil moisture is an important confounding factor for SAR-based landslide detection, and have therefore added the earthquake-triggered case study and clarified the interpretation and limitations of the SAR response in the revised manuscript.

MANUSCRIPT CHANGES

[Edited Line 181]

Unlike phase-based interferometry, which measures precise displacement, amplitude changes indicate changes in radar scattering properties that can occur following surface disturbance, including landsliding.

[Added Line 188]

In this study, the SAR response is therefore treated as an anomalous change in scattering properties that provides information on landslide timing, rather than as evidence of a permanent change in surface properties. This transient response may reflect changes in radar scattering associated with the landslide, although SAR backscatter is also sensitive to changes in surface soil moisture. Therefore, the observed response is interpreted primarily as an indicator of the timing of the landslide rather than as evidence of a persistent post-landslide change in surface properties.

REVIEWER 1: COMMENT 2-4

Figure 4. It is not clear what the different symbols represent in this figure. Why are there two separate sets of unfilled blue circles (one at 100% and the other increasing from 0 to 50)? What are the circles and crosses? By diameter, do you mean the width of the box? It is a bit confusing to interchangeably use “box” and “buffer”. Also, in the caption, you should specify 1x1, 3x3 etc. “pixels”.

RESPONSE

Thank you for this comment. The figure caption for Figure 4 has been altered to address this unclear or missing information.

MANUSCRIPT CHANGES

[Added Figure 4 Caption]

Figure 4. Percentage of the landslide within the AOI (blue) and the percentage of the AOI in the landslide (orange). Each buffer is represented by the mean, interquartile range (bar). The minimum and maximum percentages within the dataset are shown as blue unfilled circles and orange crosses respectively. The diagram to the right shows one landslide example, with the buffer for 1x1, 3x3, 5x5, and 9x9 pixels drawn.

REVIEWER 1: COMMENT 5

Lines 137-138 Is the size of this buffer something that can easily be changed in the code? This could allow you to adjust if you know you are looking at an area with a lot of large landslides. I assume that making it larger from the start would make the code take longer

RESPONSE

This can be changed within the dataset. It has been moved from the automatically set section to the manual code section before Step 1. It has been clarified that this does not need changing, but it is an option if desired.

MANUSCRIPT CHANGES

[Added Line 143]

This buffer value is adjustable within the code workflow, which may benefit datasets that contain larger landslides on average, however 100m has been deemed optimal for this case study that shows a range of landslide geometries.

REVIEWER 1: COMMENT 6

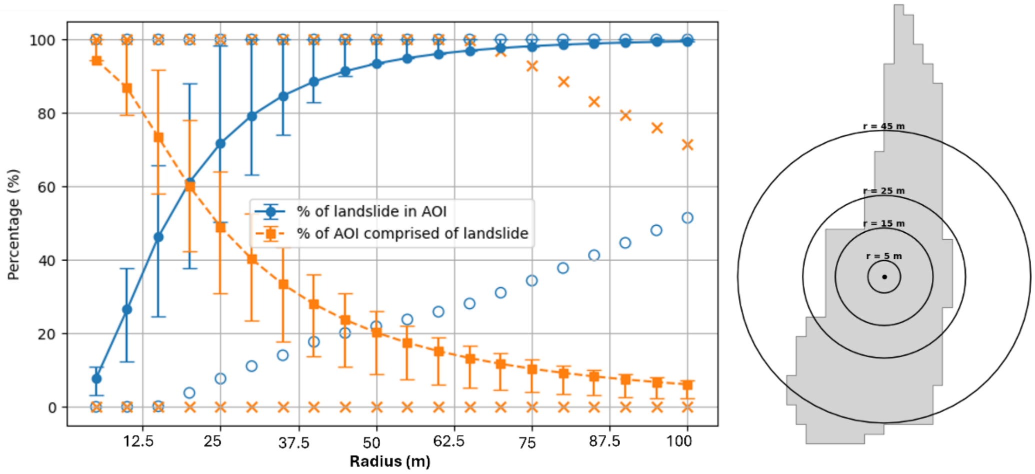
Line 139 “However, Fig. 4 shows and average of less than 10% of this 100 m radius (200 m diameter) is comprised of the landslides, with most of the dataset falling between 0-10%” But in Figure 4, it looks to me as if an average of 20% of the AOI is comprised of landslide, not 10%

RESPONSE

The original Figure 4 x-axis was labelled as diameter, which caused the values to appear twice as large as intended. This has been corrected to radius for consistency with the text and to clarify that the 100 m buffer refers to a 100 m radius.

MANUSCRIPT CHANGES

[Replaced Figure 4 x-axis label with radius values]



[Text Line 139]

However, Fig. 4 shows an average of less than 10% of this 100 m radius is comprised of the landslides, with most of the dataset falling between 0-10%.

REVIEWER 1: COMMENT 7

Line 203 “NDVI is expected to result in a drastic drop” this phrasing is a little odd, I suggest “NDVI is expected to show a drastic drop”.

RESPONSE

Thank you for this clarification. This wording has been altered in the text.

MANUSCRIPT CHANGES

[Changed Line 203]

NDVI is expected to show a drastic drop when a landslide occurs.

REVIEWER 1: COMMENT 8

Line 223, I think you mean an improvement in the precision, but a decrease in the sensitivity rather than the accuracy. Is that correct?

RESPONSE

Thank you, this section has been reworded. The term “accuracy” has been defined within the manuscript, following another reviewer comment, to refer specifically to the percentage of the dataset correctly dated within the event window. To avoid

ambiguity, the revised text describes the change directly as a decrease in the percentage of results correctly dated within the event window.

MANUSCRIPT CHANGES

[Changed Line 217]

Figure 8B and Table 4 show the results arising when combining the dates given by Step 1 and 2, where we see an improvement in temporal uncertainty (12.8 days to 11 days average) but a decrease in the percentage of the results correctly dated within the event window (68.5%).

[Second Results paragraph collapsed into first for redundancy]

REVIEWER 1: COMMENT 9

Line 230 “the results in this study utilise the descending signal only” why is this? It’s true that more landslides were triggered on eastern facing slopes in this event (as seen in Fig. 9), but there are a reasonable number also on west-facing slopes that might be better detected with the ascending track. Couldn’t using both increase the sensitivity?

RESPONSE

We did test ascending and combined ascending/descending data, but neither improved sensitivity; descending-only performed better. Figure 9 shows the distribution of all landslides detected by a) descending, c) ascending and e) both Sentinel-1 SAR datasets. The method returned worse results for all ascending slope angles, as shown in Figure 9c below. The combined track of both the ascending and descending data (Figure 9e) also shows worse results than those within the descending track.

We concluded that this is most likely due to the offset footprints of the ascending fields of view, whilst the descending track images all align with each other (see Appendix Figure A3).

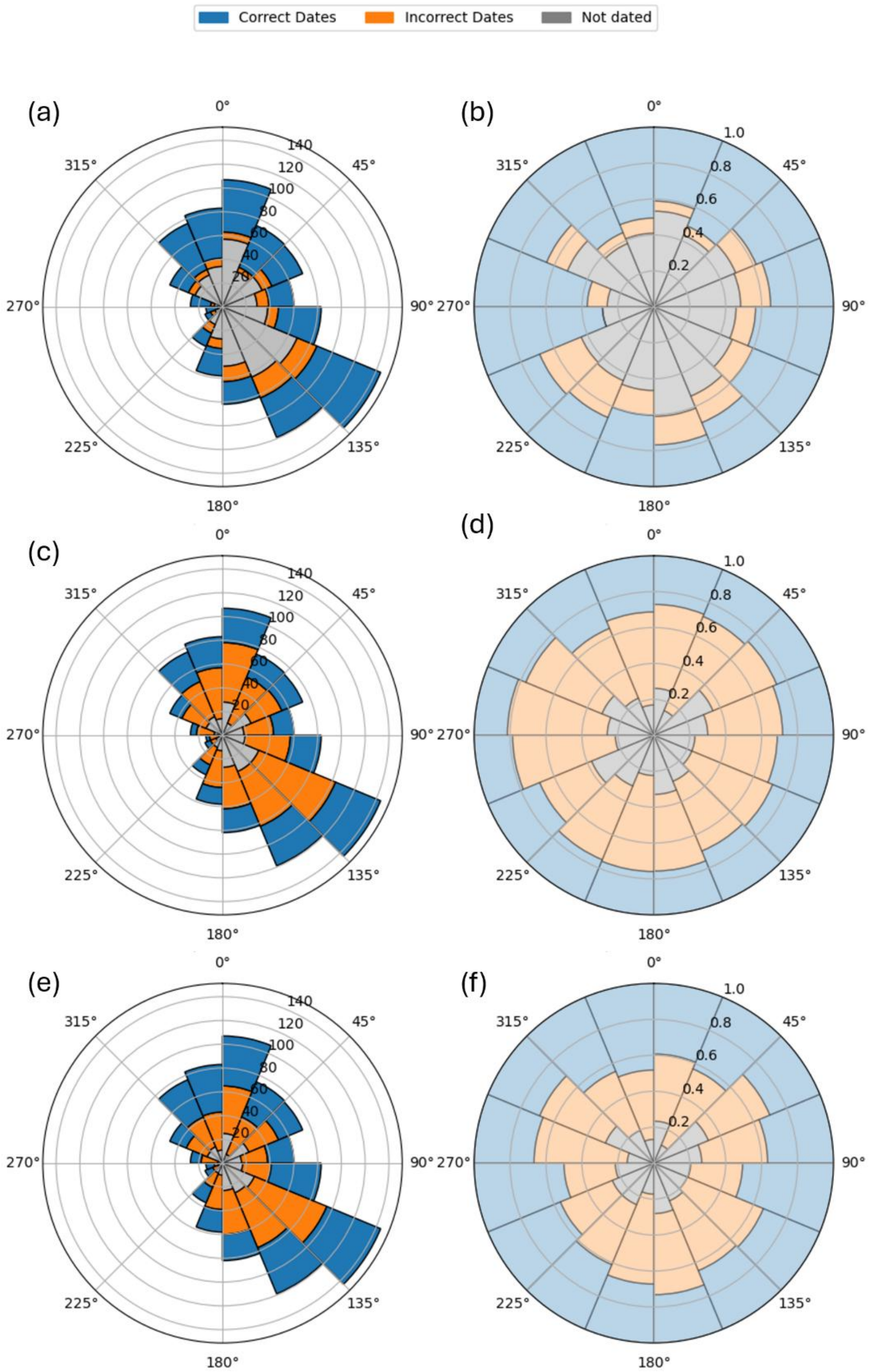


Figure 9 Caption: Distribution of correctly (blue), incorrectly (orange), and not dated (grey) landslides with respect to slope aspect derived from the SRTM 1 Arc-Second Global digital elevation model NASA JPL (2013). The results from the descending (a and b), ascending (c and d) and both (e and f) SAR data are shown. a), c), and e) show the absolute distributions in respect to aspect, whilst b), d) and f) show the normalised distribution across each aspect.

This is expanded upon within the discussion paragraph started on line 274.

REVIEWER 1: COMMENT 10

Line 238 “Burrows et al. additionally focusses on landslide delineation” This is not really the case. They use landslide polygons, but their method does not delineate these polygons, they were taken from pre-existing studies. Their method provides dating only. For me the complementarity of their method to yours is that while theirs was designed to be applied in monsoon-affected regions (and in a later paper, they go on to use it in Nepal), yours works better when reasonably frequent multi spectral images are available (as you describe later on in lines 315-320)

RESPONSE

Thank you for this clarification. The wording has been modified to correct this inaccuracy and clarify the distinction between the two approaches.

MANUSCRIPT CHANGES

[Edited Line 238]

It is important to note that Burrows et al. 2022 additionally is designed to be applied on monsoon affected regions, whereas the present study is better when more frequent multi-spectral images are present.

REVIEWER 1: COMMENT 11

Line 315-320 It’s true that this is a limitation, but in monsoonal climates, the 6-month window in which a landslide occurred is almost always known a-priori from the manual mapping. For example there are several inventories of monsoon-triggered landslides in Nepal mapped on an annual basis (see Harvey et al. 2025 for a summary https://doi.org/10.1007/s11069-024-07013-1) , if you assume all landslides occurred during the rainy season between April-November, could you use this information in place of the constraint you currently place using NDVI?

RESPONSE

This is a useful point not considered within the original report. If this information is known, this method could use this pre-existing information.

There would be limitations for events that happen at the very beginning and end of the period depending on the frequency of Sentinel-1 data, as the events may be missed out, or be contained within the first few data points, which are unlikely to be accurately detected.

MANUSCRIPT CHANGES

[Added Line 320]

Inventories that provide seasonal information, such as that discussed in Harvey et al. 2024 for Nepal, could use the yearly monsoon dates as an approximate replacement for the ±90 day time-period to bypass Step 1. However, there is limitation on using the monsoon season for events that would occur on the peripheries of this period, as these are likely to be dated less effectively if the Sentinel-1 data is infrequent.

REVIEWER 1: COMMENT 12

Section 5.1 Another limitation is that the method only works (or at least has only been tested) in vegetated environments

RESPONSE

Thank you for highlighting this limitation. This has been added to Section 5.1, clarifying that the method is dependent on vegetation cover and may not detect landslides in areas with little to no vegetation.

MANUSCRIPT CHANGES

[Added Line 321]

Furthermore, this method will be unable to detect landslides in areas with little to no vegetation cover, due to both NDVI and BSI depending on the removal of vegetation. Landslides in areas of low vegetation, particularly on mountainous slopes, may therefore not be detected by this method

---

REVIEWER 1: COMMENT 13

Line 102: no need to capitalise Earth here, since you refer to the material, not the planet Line 139 “and” should be “an”.

RESPONSE

Thank you for this comment. The capitalisation of “Earth” and the spelling error “and”/“an” have been corrected in the manuscript.

MANUSCRIPT CHANGES

This capitalisation and misspelling have been fixed in the manuscript.

<u>Reviewer Comments 2</u>

REVIEWER 2: COMMENT 1

1. The parameter optimisation appears to be performed on the validation dataset?

Several methodological choices appear to have been selected using the same Zimbabwe inventory that is later used to evaluate performance. These include a 100 m background buffer, preferred Ruptures penalties, hierarchical ranking of change-point models, and a ±90-day SAR window, filtering based on physically plausible NDVI/BSI changes. Although these choices are individually reasonable, they collectively raise concerns regarding overfitting. The manuscript would be considerably strengthened if the authors clearly separated parameter calibration, parameter validation, and final model evaluation. If an independent validation dataset is unavailable, the authors should, at a minimum, explicitly discuss this limitation and quantify the sensitivity of the results to these parameter choices.

RESPONSE

Thank you for this comment, as this was not included in the original version and the subset used to fit the method was not removed from the original results given.

Within the original decision making of this method, a smaller subset of the dataset was used to make these decisions. This is seen within the new Figure A1 which is expanded upon within the following comment.

The results from these 100 points have now been removed from the results of this method to ensure no overlap between the calibration and the evaluation values, as this was overlooked in the original version.

The filtering after Step 1 was carried out due to physical understandings, not on the actual results. This was decided because landslides should always result in a decrease of NDVI and increase in BSI, due to the removal of vegetation. Therefore, the only points that are filtered after Step 1 are those that result in an increase in the NDVI or decrease in BSI after the detected date.

MANUSCRIPT CHANGES

[Added Line 94]

The decisions made for fixed parameters within this method were chosen using a 100-point subset of the dataset, taken systematically through the dataset to ensure area variability. These points were removed from the calibration shown in Table 3 to reduce the risk of overfitting.

[Edited Table 3]

Stage	Metric	Value
Step 1	Total Points Assessed	1219
	Points Dated (NDVI only)	184
	Points Dated (BSI only)	41
	Points Dated (both)	673
	Total Points Dated	898
	In March 2019	683 (76.1%)
	Overlapping 15-20 March 2019	355 (39.5%)
Step 2	No. available for Step 2	898
	Total Points Dated by Step 2	827
	In March 2019	728 (88.0%)
	Overlapping 15-20 March 2019	711 (86.0%)
Overlap	Overlapping 15-20 March 2019	568 (68.7%)

[These results have been altered throughout the text.]

REVIEWER 2: COMMENT 2

2. Sensitivity analysis may require more empirical testing

Many important parameters appear to have been fixed after empirical testing, but the manuscript only briefly explains why. For example, why exactly 100 m?, why ±90 days rather than ±60 or ±120?, why L2 penalty = 1 for SAR?, and why this specific hierarchy of penalties? Although Figures 5 and 7 partially justify these decisions, a more systematic sensitivity analysis would greatly improve confidence in the methodology. Ideally, the supplementary material could include performance curves showing how accuracy, precision, and percentage of dated landslides vary across parameter space.

RESPONSE

The parameters were empirically calibrated using a subset of 100 landslides from the full dataset of 1,319 landslides. The figure shows the relationship between the percentage of landslides dated and the percentage of landslides dated correctly (overlapping 15-20 March 2019) for each parameter tested. These analyses were carried out keeping all other variables consistent, showing the distributions of accuracy for each parameter. Figure 1a and c show the best performing parameters are a 90-day SAR window for Step 2, and a 100m buffer range for Step 1.

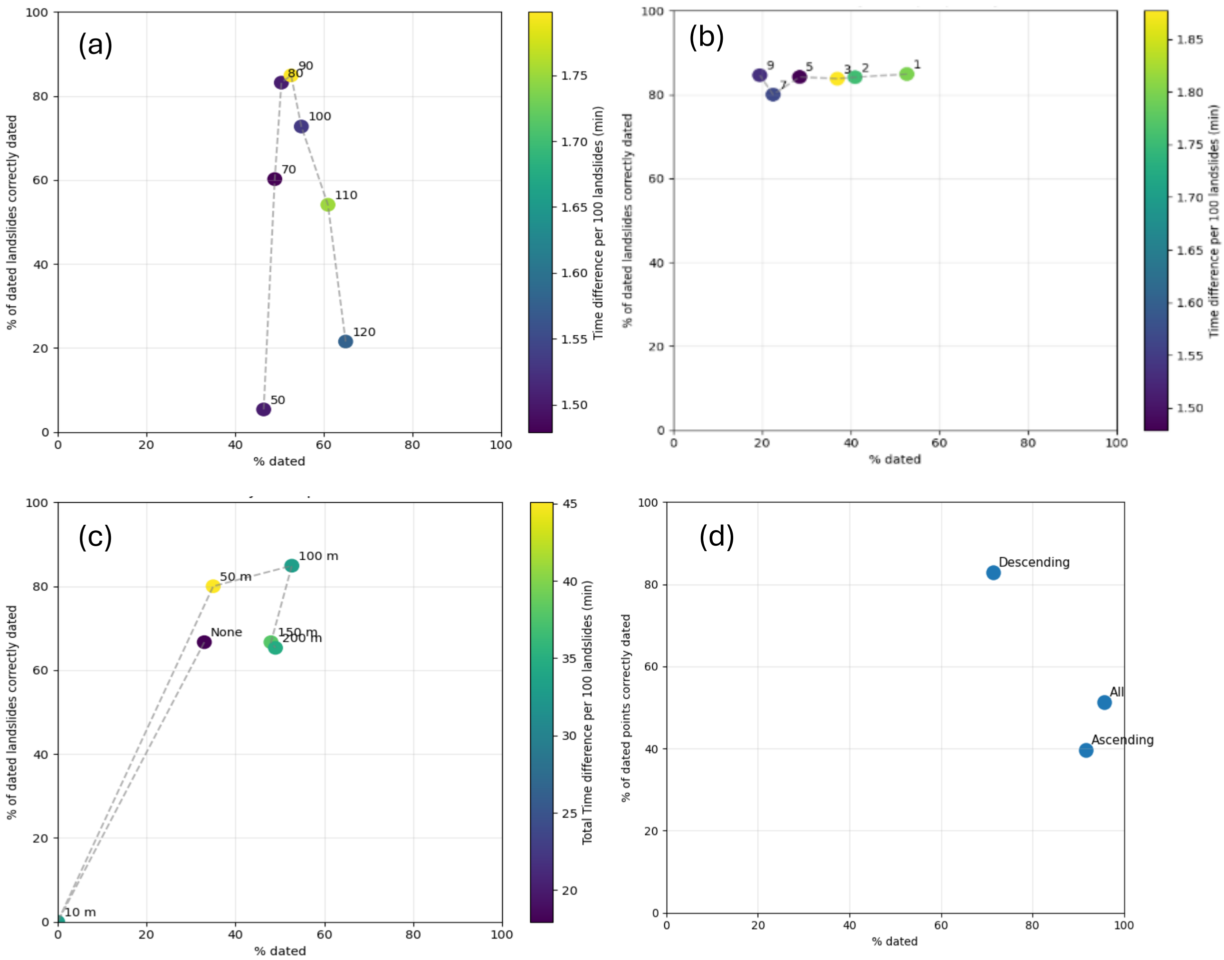


Figure A1: Relationships between the number of landslides dated and the number of those dated landslides dated correctly for a) Step 2 date window, b) Step 2 Ruptures Penalty, c) Step 1 buffer distance, d) ascending/descending Sentinel-1 track option. Sensitivity analysis conducted on a calibration subset of 100 landslides from the full dataset of 1,319 landslides. These analyses were carried out keeping all other variables consistent (Step 2 window = 90, Step 2 method and penalty = L2 penalty 9, Step 2 data = VV descending, Step 1 buffer = 100m).

Figure 1b shows that while there is little impact in the accuracy of the results using different penalties in Step 2, using a lower penalty results in a larger number of landslides being dated. Therefore, we chose a penalty value of 1 for Step 2.

The hierarchy of penalties used within Step 1 was established from the preliminary analysis presented in Figure 5. Higher penalties provide more selective detection,

but reduce the number of landslides dated. The hierarchy therefore progresses from more selective to less selective configurations to balance the reliability and coverage of Step 1 detections. The rationale and selected hierarchy are already described in the manuscript.

Temporal uncertainty was not included as a sensitivity parameter because this is determined by the acquisition frequency of the available Sentinel-1 imagery rather than by the methodology itself. The method can only assign a landslide to the interval between available observations, meaning that the possible temporal uncertainties are fixed by the dataset.

The descending track was used as it shows a significant improvement of the results of the ascending track for SAR data in Step 2. It is mentioned later in this manuscript that this is likely due to the available data in the region and is likely case study specific.

MANUSCRIPT CHANGES

[Added Figure to Appendix]

[Added Line 129]

Using a fixed subset of 100 points, Figure A1A shows the impact of different buffer sizes on the number of points dated, and the percentage of those points that are dated within the event window. A 100~m buffer was optimal for this subset and was therefore used within the wider dataset.

[Added Line 155]

These specific configurations were selected based on the empirical results shown in Figure 5 to span a range of detection stringency, from L2 penalty 9, which provides high selectivity but low coverage, to L2 penalty 0.5, which provides substantially greater coverage but lower accuracy.

[Added Line 179]

Figure A1D highlights that for the 100-landslide fixed sample, the descending track resulted in a much higher number of events being correctly dated within the event window, so this was used. The type of data that is used however is changeable within the code and should be determined in each case.

[Added Line 193]

This is also shown in Figure A1B which shows that a 90-day window has the highest number of correctly dated events within the event window for the fixed 100 points used to calibrate the method.

[Added Line 198]

Figure A1C shows the impact of different penalty levels on the 100-point subset for Step 2, showing that there is little impact on the number of events correctly dated. However a lower penalty results in a higher number of events being dated, so a penalty of 1 was used.

REVIEWER 2: COMMENT 3

3. It would be good for the readers to have a definition of the evaluation metrics used in this study.

The manuscript uses terms such as accuracy, precision, correctly dated, overlap, and percentage dated throughout. However, these definitions are not always intuitive. For example, 84.6% "accuracy" is defined as the percentage of dated landslides for which the estimated temporal window overlaps the known event window. This differs substantially from the conventional statistical definition of classification accuracy. I recommend including a dedicated subsection that defines each evaluation metric mathematically. Similarly, "precision" is used to describe temporal uncertainty rather than predictive precision, which may confuse readers.

RESPONSE

We thank the reviewer for highlighting the potential ambiguity in the terminology used to describe the evaluation metrics. The term 'precision' has been replaced in the manuscript with the term 'temporal uncertainty'. We have added a dedicated "Terminology and Evaluation Metrics" section to clearly define the terms used throughout the manuscript and provide their mathematical formulations.

MANUSCRIPT CHANGES

[Added Line 49]

Terminology and Evaluation Metrics

This section will provide a description of the terminology used within the paper for clarity.

Percentage Dated: This is the number of the total landslide dataset dated by the end of the method. Within the paper, unless stated otherwise, this will refer to the number of landslides dated by the end of Step 2. Some instances will refer to the percentage dated by the end of Step 1 alone.

$$Percentage\ Dated = \frac{The\ number\ of\ events\ dated}{The\ total\ number\ of\ events} x100$$

Overlap: Each landslide is given a temporal window, corresponding to the Sentinel-1 images between which the event has been detected. Furthermore, Emberson et al.

2022 has dated the landslides within this report between 15-20 March 2019. Overlap in this report refers to when these two events overlap.

Correctly dated: A landslide has been deemed correctly dated if the temporal window given overlaps the window provided by Emberson et al. 2022.

Accuracy: Accuracy refers to the percentage of the dated landslides that are found to overlap the window of interest and be correctly dated.

$$Accuracy\ (\%) = \frac{The\ number\ of\ events\ correctly\ dated}{The\ number\ of\ events\ dated} x100$$

Temporal Uncertainty: The temporal uncertainty refers to the length of the determined window for each landslide. The length of time between the two Sentinel-1 images between which the landslide has been detected.

$$Temporal\ Uncertainty = Date\ of\ Second\ image - Date\ of\ first\ image$$

The terms listed above are used throughout the manuscript and will be referring to the definitions listed above when used.

REVIEWER 2: COMMENT 4

4. The authors shall consider performing an uncertainty analysis.

Every detected landslide date is ultimately treated as deterministic. However, uncertainty exists at multiple stages, viz., Sentinel revisit interval, cloud masking, rupture change-point selection, SAR smoothing, and parameter selection. The manuscript would benefit from a discussion of how these uncertainties propagate through the workflow. For example, are there confidence scores associated with individual detections? Could multiple candidate change points be retained rather than selecting only the first one?

RESPONSE

Thank you for this comment. We agree that uncertainty occurs at multiple stages of the workflow. A qualitative uncertainty analysis has therefore been added to identify the major sources of uncertainty, their potential effects on the results, and where possible, how these could be assessed or mitigated.

Source	Uncertainty	Effect on Results
Sentinel-1 and 2 revisit intervals	Observations are not continuous.	Landslide can only be dated to the temporal resolution of available observations. This is more limiting within Step 2,

		as Step 1 is used as an approximate dating tool for Step 2.
Cloud masking	Images may be removed even when an observation would otherwise be useful. Or the image may not be removed, resulting in artificially low NDVI values.	Creates gaps or artificial brightness and can move the detected date away from the actual event. In extreme cases, this could result in the event date falling outside of the window provided in Step 2.
Change-point detection	Currently only the first point is considered, and therefore artificial early detection may disguise genuine detection later in the timeseries.	False/multiple candidate changes may be identified. There is little way of determining if these are true or not. The penalty and method in Step 1 could indicate level of assuredness.
SAR analysis	SAR observations are noisy and temporally variable. Infrequent revisit time can mean that the 6-month window of interest is ineffective at detecting landslides. Furthermore, falsely identified Step 1 results can be dated in Step 2 when there is no landslide present in the data.	Candidate date may shift or weak changes may be missed or falsified. Only one penalty is used so there is no confidence in this step.
SAR smoothing	Smoothing suppresses noise but can also shift/blur a transition.	Introduces temporal uncertainty.
Parameter selection	Buffer, penalty, SAR window, etc. influence sensitivity.	Different parameters can produce different detected dates.

This approach does introduce a deterministic element, as stated by the reviewer, which is a limitation of the method. The method could be altered to retain multiple candidate change points rather than only the first detected change. This could

increase the likelihood that the true landslide date is detected; but would also introduce complexity into the workflow. For example, three potential dates identified during Step 1 would then lead to multiple workflow runs for Step 2. This would increase computational requirements and, more importantly, would result in multiple potential dates for a single landslide without providing a means of determining which represents the true event. A more robust implementation of multiple candidate dates would therefore require the method to treat the outputs as a distribution of possible event dates, rather than as multiple equally valid detections. For this reason, the current method retains only the first candidate date, while the code could be adapted to retain multiple candidates where required. Work to expand upon this existing method could add this dimension.

MANUSCRIPT CHANGES

[Added Line 314]

Uncertainty is introduced at several stages of the workflow. The temporal precision of detected dates is constrained by the revisit intervals of Sentinel-1 and Sentinel-2 and, for optical data, by cloud masking. Missing observations can shift the apparent timing of a change or prevent detection altogether. Also, changepoint detection is sensitive to parameter selection and the choice of penalty values, which may influence the timing and number of detected changes. Furthermore, SAR backscatter exhibits substantial temporal variability, and smoothing operations used to reduce noise may introduce additional temporal uncertainty. Finally, the current implementation retains only the first detected change point, resulting in a deterministic estimate of event timing. This does not assign a formal confidence score to individual detections, although the selected Ruptures method and penalty values may provide an indication of the strength of a detected change. Future work could retain multiple date results and represent event timing as a probability distribution rather than a single date estimate.

REVIEWER 2: COMMENT 5

5. Discussion of computational efficiency could be expanded

The manuscript reports that approximately 100 landslides can be processed in 45–60 minutes. This implies roughly 10–13 hours for the full inventory. Since one of the principal advantages of the proposed workflow is accessibility, readers would benefit from additional discussion regarding computational bottlenecks, opportunities for parallelisation, use of GEE map/reduce operations, and scalability to inventories containing tens of thousands of landslides.

RESPONSE

The principal computational bottleneck occurs in Step 1, where residuals are calculated to remove seasonal variation prior to change-point detection. Although the underlying image processing and spectral-index calculations are performed using Google Earth Engine (GEE) server-side operations, the time-series information is retrieved for individual landslides through repeated client-server requests. These requests and the subsequent conversion of the results into pandas Dataframes for change-point analysis introduce substantial processing time.

There are opportunities for further optimisation. In particular, the time-series extraction could be batch processed rather than retrieved individually, reducing repeated client-server communication. The change-point analysis is also independently performed for each landslide and is therefore well suited to parallel processing, allowing multiple landslides to be analysed together. Similarly, greater use of GEE map/reduce operations could allow more of the time-series preparation and reduction to be performed server-side before the results are transferred to the Colab environment. However, there is a 5000-element limit for collection queries, and for extended timeseries, processing multiple points at once returns an error. The current implementation therefore uses individual point extraction to provide a predictable and robust workflow across different inventories.

The workflow was deliberately designed to remain self-contained within Google Colab, with GEE providing the remote image-processing infrastructure and Python handling the analysis. Moving a greater proportion of the workflow into GEE could improve scalability but would require a more substantial restructuring of the current implementation. We therefore consider the current approach a balance between computational efficiency and accessibility.

The method is currently suitable for inventories containing several thousand landslides, although processing time increases with inventory size. Inventories containing tens of thousands of landslides would result in substantial processing times using the current implementation and would benefit from batch extraction,

parallel processing, or a more GEE-native implementation. We have expanded the discussion of these computational limitations and potential future improvements in the revised manuscript.

MANUSCRIPT CHANGES

[Line 302 emended]

The current implementation processes approximately 100 landslides in 30 minutes using a 13th Gen Intel Core i7 processor with 16 GB RAM, corresponding to approximately six and a half hours for the complete 1,319-landslide dataset.

[Added Line 305]

The calculation of residuals to remove seasonal variation prior to event detection is a significant computational bottleneck in Step 1, particularly because the required time-series data are retrieved from Google Earth Engine through repeated client-server requests. Further efficiency improvements could be achieved through batch extraction of time-series data and parallel processing of individual landslides, as the change-point analysis for each landslide is independent. However, it is important to note that there are element limits when retrieving information from GEE, and therefore so longer timeseries, batch processing in the current method format is not possible for more than a few individual points. Greater use of Google Earth Engine map/reduce operations could also reduce the volume of data transferred to the Colab environment.

[Added Line 309]

The current method can process inventories containing several thousand landslides. Inventories containing tens of thousands of landslides would require substantial computation times and may benefit from more advanced GEE-based batch-processing approaches. Therefore, while the proposed workflow offers a practical and accessible solution for regional-scale landslide inventories, computational efficiency remains a limitation for very large datasets and time-sensitive applications.

REVIEWER 2: COMMENT 6

6. Comparison with previous studies shall be more balanced

Table 1 compares numerous studies using different datasets, landslide types, spatial resolutions, inventory quality, vegetation conditions, and validation strategies. Consequently, direct comparison of percentages dated or reported accuracy may not always be meaningful. The discussion should acknowledge these differences more explicitly to avoid implying superiority where experimental conditions differ substantially.

RESPONSE

We thank the reviewer for highlighting this point. We agree that the performance metrics reported in Table 1 cannot be considered directly comparable because the studies differ substantially in landslide size, vegetation conditions, inventory characteristics, satellite datasets, spatial information available, sample size, and validation strategies. We have therefore revised the manuscript to make clear that Table 1 is intended to provide methodological context rather than a direct ranking of reported accuracies.

MANUSCRIPT CHANGES

[Added Line 33]

Table 1 summarises previous approaches that have used satellite observations to retrospectively constrain landslide timing when the approximate landslide location is known. The studies differ substantially in their datasets, landslide characteristics, satellite observations, spatial information available, and validation approaches. Consequently, the reported performance metrics should not be considered directly comparable measures of method accuracy. Instead, the table is intended to provide context for the methodological design of this study and to highlight differences in the information required to apply each approach.

[Edited Line 41]

In contrast, the method presented here is designed around a minimal spatial input: a landslide point location. This is intended to increase the range of existing inventories to which retrospective dating can potentially be applied, rather than to provide a directly superior alternative to approaches requiring more detailed landslide information.

[Edited Line 234]

This proportion is comparable to, and in some cases higher than, the proportion of landslides dated in several previous studies (Table 1); however, these values should not be interpreted as directly comparable measures of method performance because previous studies differ substantially in landslide size, vegetation conditions, inventory characteristics, satellite data, and validation strategies. The relatively large number of landslides considered here nevertheless demonstrates the potential for the proposed point-based approach to be applied across larger inventories.

[Edited Line 253] [Whole paragraph]

Previous studies have generally evaluated their methods on substantially smaller landslide populations (7–66 landslides in several studies; Table \ref{tab: other studies}). Direct comparison of reported accuracy is therefore difficult. For example, Barbera et al. 2025 restrict their analysis to landslides encompassing at least ten 10 m x 10 m pixels and occurring in well-vegetated areas. Similar spatial, vegetation or landslide-size constraints are present in several of the studies summarised in Table 1. These restrictions may improve the suitability of the selected landslides for satellite-based detection but also mean that the resulting performance metrics describe a different target population from that considered here. The present method was deliberately evaluated without a minimum landslide-size or polygon-geometry requirement, allowing 1319-point locations to be assessed. This demonstrates the potential scalability and broader applicability of the point-based approach, although further testing across different inventories, landslide sizes and vegetation conditions is required to establish how these factors influence dating performance.

REVIEWER 2: COMMENT 7

The manuscript repeatedly states that the workflow "requires only point locations." Strictly speaking, this is true for the execution of the workflow, but the optimisation and validation were performed using polygon inventories converted to centroids. This distinction should be stated more explicitly.

RESPONSE

We thank the reviewer for highlighting this distinction. The workflow itself requires only point locations; however, the landslide inventory used for parameter calibration and evaluation was originally provided as polygon geometries, which were converted to centroid points before analysis. We have clarified this distinction in the revised manuscript and explicitly noted it as a limitation.

MANUSCRIPT CHANGES

[Added Line 66]

All validations within this study have been carried out using these centroid point locations.

[Added Line 334]

A further limitation of this method is that the reference landslide inventory consists of polygon geometries that were converted to centroid point locations. Although this provides the point locations required by the workflow, other datasets may provide point locations that differ substantially from polygon-derived centroids, such as failure points located at the tip of the scar, which may be less effectively detected by this method. Consequently, the performance demonstrated here may not directly reflect the ability of the method to detect landslides represented by independently derived point locations. Future validation using datasets containing independently recorded point locations would help to assess the robustness of the method under these conditions.

REVIEWER 2: COMMENT 8

The discussion could better explain why NDVI and BSI were selected instead of other vegetation or disturbance indices (e.g., NBR, EVI, NDMI). Readers may wonder whether these alternatives were evaluated.

RESPONSE

NDVI and BSI were selected following preliminary testing of several commonly used optical indices for landslide detection. Early versions of the methodology evaluated NDVI, BSI, and NBR. However, the results obtained from BSI and NBR showed little difference in dating performance across the assessed landslides. As a result, NBR was removed to reduce redundancy and improve computational efficiency. The final combination of NDVI and BSI was used because the two indices capture different aspects of landslide-induced surface change. NDVI is commonly used in landslide detection due to its sensitivity to vegetation loss, while BSI responds to the exposure of bare soil and rock following slope failure.

Other indices, such as NDMI and EVI, were not evaluated during the development of this methodology. NDVI was included as a baseline because of its extensive use and demonstrated effectiveness in landslide detection within the literature. EVI was not considered because it requires coefficients designed to correct for atmospheric and canopy effects, which would introduce additional assumptions and sources of uncertainty.

Future work could investigate whether alternative indices such as NDMI or EVI provide additional discriminatory power for landslide dating in different environmental settings.

MANUSCRIPT CHANGES

[Added Line 114]

Other indices were not assessed during the development of the methodology. NDVI and BSI were selected because they capture different aspects of landslide-induced surface change, with NDVI responding to vegetation loss and BSI responding to the exposure of bare soil and rock. Early assessment of the methodology included Normalised Burn Ratio (NBR), but little variation was observed between NBR and BSI in the resulting landslide dates. NBR was therefore removed to reduce redundancy. Other indices, such as the Enhanced Vegetation Index (EVI), were not considered because they require coefficients designed to correct for atmospheric and canopy effects, which would introduce additional assumptions and sources of uncertainty. Alternative indices, such as Normalised Difference Moisture Index (NDMI) and EVI, could be investigated in future work to assess whether they provide additional information for landslide dating.

REVIEWER 2: COMMENT 9

The rationale for using only VV polarisation deserves slightly more explanation. Although Figure 7 shows better performance than VH, quantitative statistics comparing the two channels would improve transparency.

RESPONSE

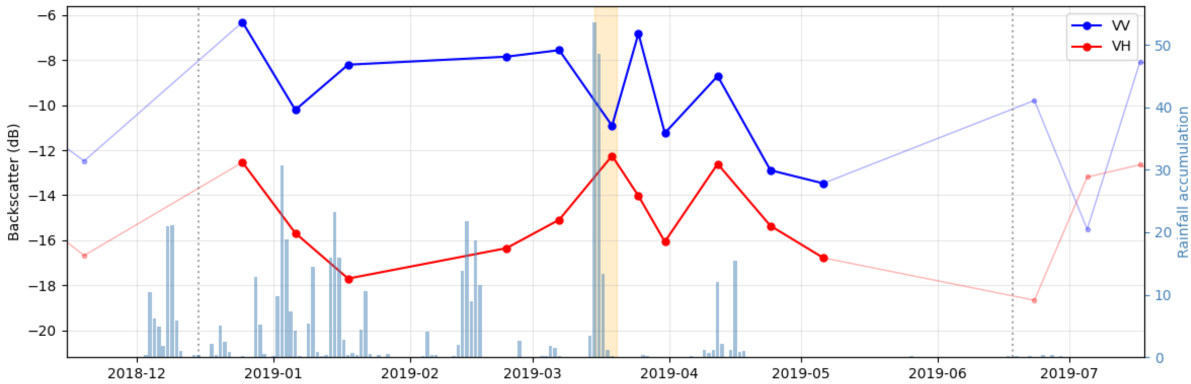


FIGURE: 6-month timeseries of VV (blue) and VH (red) SAR data from Sentinel-1 in Google Colab. This figure shows the daily rainfall distribution on the right y axis, with the event window highlighted in orange.

We thank the reviewer for highlighting the need for a quantitative comparison between the VV and VH polarisations. We have revisited the calibration procedure and identified an inconsistency in the original implementation: although the manuscript described calibration using 100 points, the original analysis used the first 100 points in the dataset rather than the specified calibration set. We have corrected this so that the same 100 calibration points are now used consistently across the analyses, aligning with the earlier comments from the reviewer.

Following this correction, the performance of VV and VH is very similar. Using the 100 calibration points, VV identified 50 events, of which 46 were correctly dated (92.0%), while VH identified 51 events, of which 47 were correctly dated (92.2%). We therefore find no meaningful difference in detection performance between the two individual polarisations.

We have also evaluated a combined VV+VH input, given that the two polarisations provide complementary information on radar scattering and are commonly considered together for landslide detection. This identified 62 events, of which 58 were correctly dated (93.5%). The combined input therefore provides a small improvement in the proportion of correctly dated events compared with either VV or VH individually.

These results have led us to revise the rationale for the use of SAR data in the manuscript. The result has been altered to use the VV + VH result, as suggested by

the calibration points. All three have been mentioned, and the comparison shown in the figure below provided.

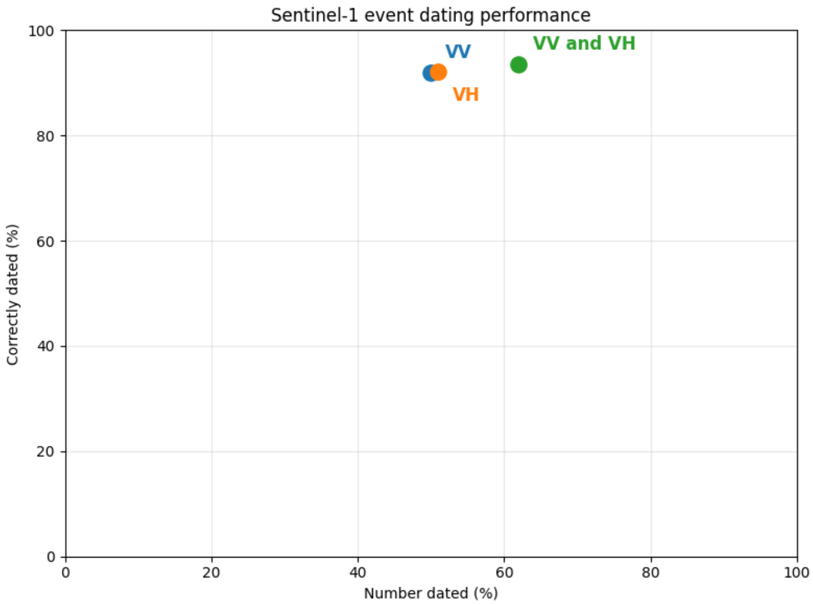


FIGURE: Relationships between the number of landslides dated and the number of those dated landslides dated correctly for VV, VH and VV + VH as a multivariable input. Sensitivity analysis conducted on a calibration subset of 100 landslides from the full dataset of 1,319 landslides.

MANUSCRIPT CHANGES

[VV and VH corrected throughout the text where VV alone has previously been written.]

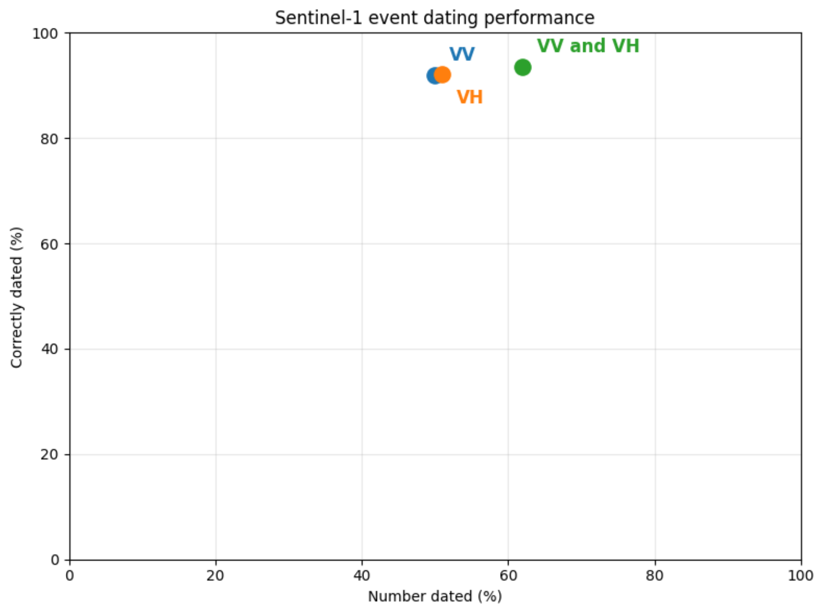
[Edited Line 73]

We use Sentinel-1 C-band vertical-vertical (VV) and vertical-horizontal (VH) backscatter polarised Synthetic Aperture Radar (SAR) data. Both backscatters were considered alone, but due to the slight improvement in the detection seen in VV and VH together, the multi dimension option was chosen for this method (see Figure A2).

[Edited Line 179]

SAR VV backscatter is sensitive to surface properties such as roughness and soil moisture, while VH is sensitive to vegetation structure and volume scattering (Burrows et al., 2022; Torres et al., 2021). Both polarisations have been successfully applied to landslide detection, with studies demonstrating that their complementary responses can improve identification of landslide-related changes (Santangelo et al., 2022; Nava et al., 2026). This study retains VV and VH as complementary dimensions of a temporal SAR signal and analyses them jointly using multivariate change-point detection. Calibration experiments showed that VV and VH individually achieved similar dating accuracies, while their combined use increased the number of correctly dated landslides. Therefore, both polarisations were retained in the final workflow.

[Figure below added to appendix]



[Figure 6 Edited to contain VH time series]

REVIEWER 2: COMMENT 10

Figure 5 is somewhat difficult to interpret because multiple penalty values and methods are plotted together. A clearer legend or separate panels would improve readability.

RESPONSE

This figure has been edited for clarity. It has been separated into two panels to distinguish between L1 and L2.

MANUSCRIPT CHANGES

[Replaced Figure 5]

[Edited Values within the manuscript to reflect these values after editing to same 100 systematically chosen points as figures from Reviewer 2 Comment 2]

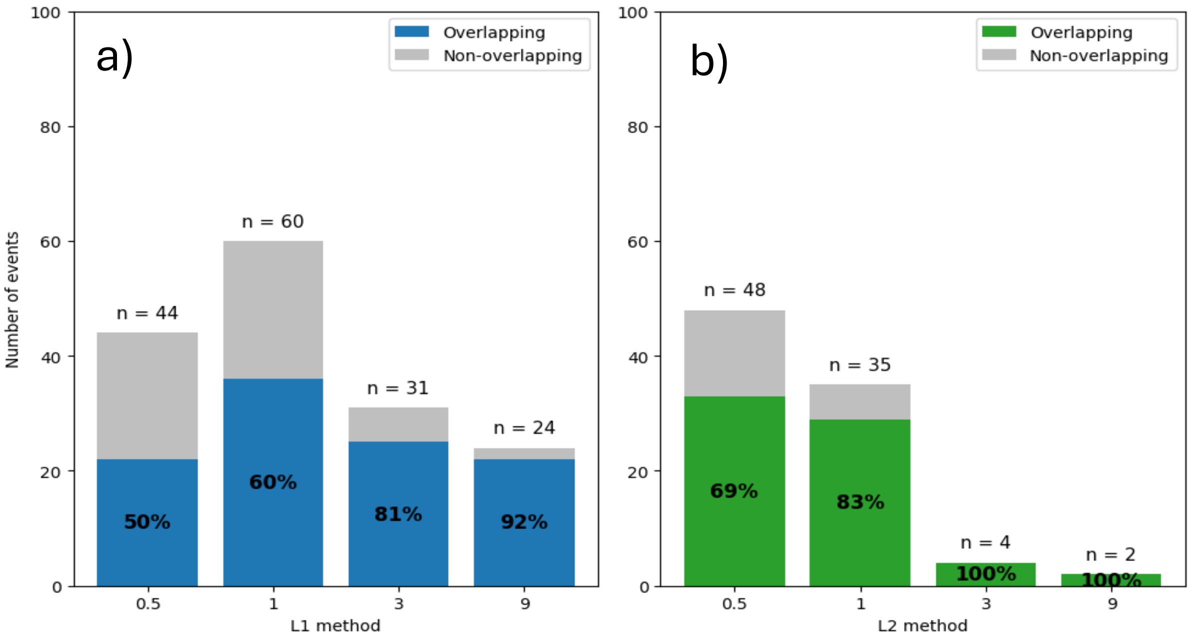


Figure 5. The number of points dated (grey) for each method, and the number of those dated correctly (spanning 15-20 March 2019) for A) Ruptures method L1 (blue) and b) L2 (green). This uses the established method for the first 150 indices in the dataset. This shows the total number of each penalty dated, and the percentage of that dated correctly.

REVIEWER 2: COMMENT 11

Several figures display only one representative landslide. It would be valuable to include additional examples illustrating successful and failed detections, as well as noisy optical and SAR time series. These examples would help readers understand the robustness of the workflow.

RESPONSE

Additional example landslides from this case study have been included to provide examples with greater noise and less clear time series, as well as examples that resulted in incorrect or unsuccessful dating. This has been carried out for Figures 3 and 6.

MANUSCRIPT CHANGES

[Replaced Figure 3, 6]

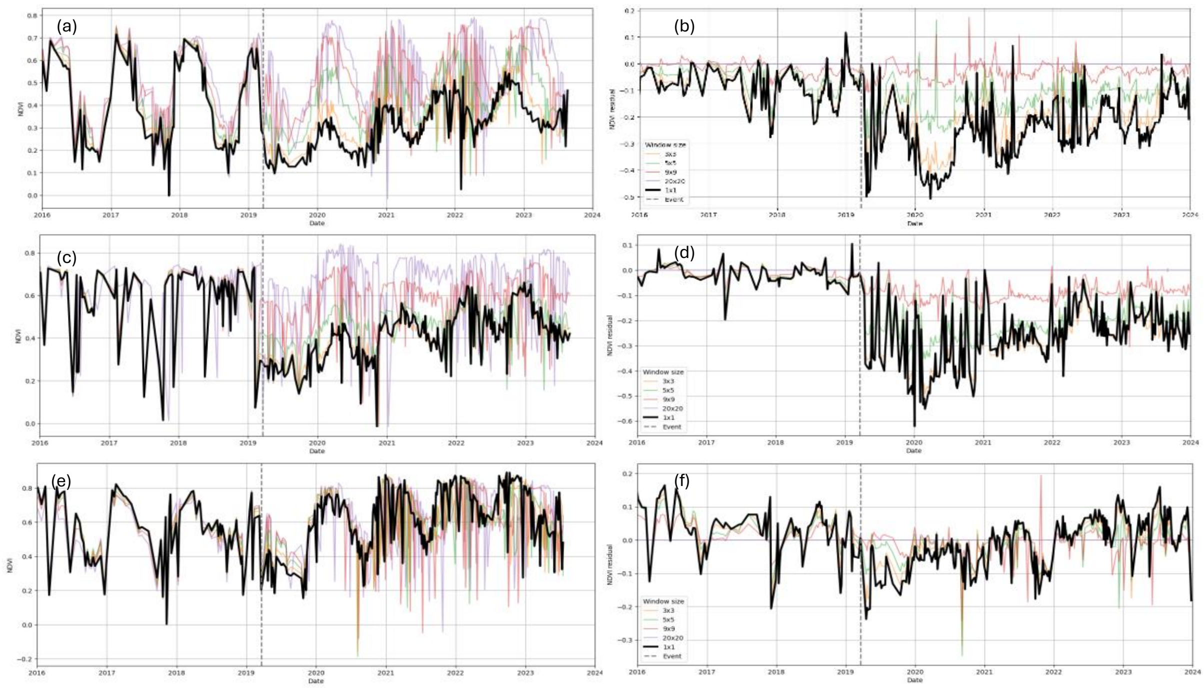


Figure 3. NDVI time series for a landslide within the Zimbabwe dataset showing the (a), (c), (e) NDVI for three given points and (b), (d), (f) the corresponding NDVI for that point minus the NDVI mean for the surrounding 100 m radius. Black shows the results for a single pixel 1x1 (10mx10m), orange = 3x3 pixels, green = 5x5 pixels, red = 9x9 pixels, and purple = 20x20 pixels. The event is shown by the dashed line at 2019-03-15.

[Added Line 137]

Figure 3C shows a significantly weaker seasonal variability prior to the landslides but produces a strong change during the landslide event. Both Figure 3A and C (and therefore B and D) result in correctly detected landslide dates, with Figure 3E and F providing an example that was not dated.

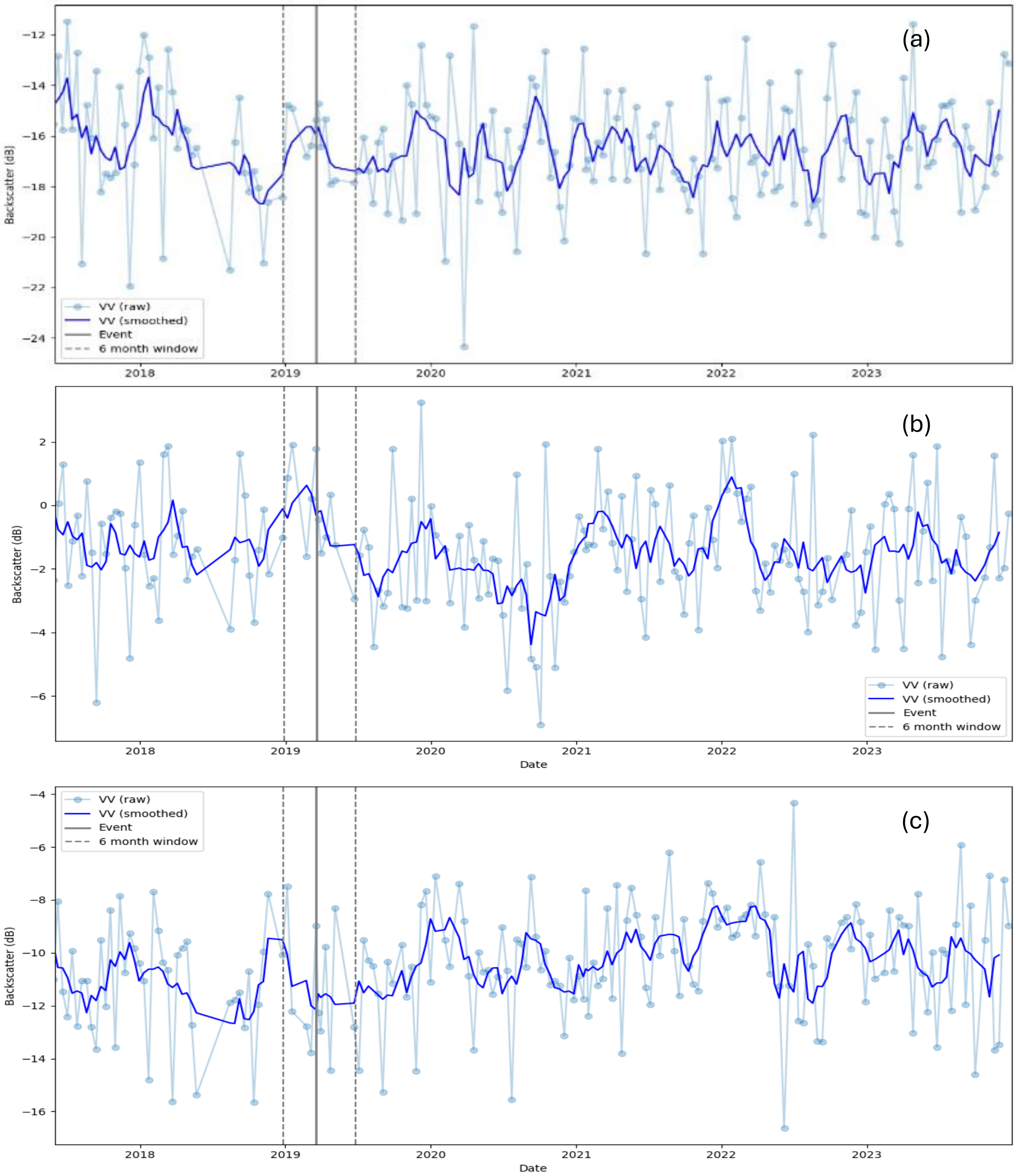


Figure 6: Complete Time Series of VV SAR backscatter. This shows both the point values and the smoothed (3 consecutive point) time series. The landslide event is shown on 20 March 2019 (solid vertical line), with the +90 day window falling between the dashed lines. (a), (b) and (c) are three examples of landslide SAR timeseries. Both (a) and (b) resulted in correctly dated landslides, and (c) has been incorrectly dated.

[Added Line 186]

Figure 6a and b both represent examples that were dated correctly, with Figure 6c showing an example that was not dated correctly.

REVIEWER 2: COMMENT 12

The manuscript frequently uses the phrase "dated correctly." Since the true date is only constrained to a temporal window (15–20 March), wording such as "correctly assigned to the known event window" would be more precise.

RESPONSE

Throughout the document "dated correctly" has been altered in many places to read as "dated correctly within the event window"/ " dated within the event window" or a similar variation, depending on the context. This has helped to clarify that there is a date window rather than a specified known date of the event.

MANUSCRIPT CHANGES

Changes made throughout the manuscript to clarify this point.

REVIEWER 2: COMMENT 13

The authors mention false positives using randomly selected non-landslide points. This is an interesting experiment and could be expanded slightly. For example, what land-cover classes generated the most false positives? Could these be filtered automatically?

RESPONSE

The distribution of correctly, incorrectly and not dated results are shown in the figure. This shows the area of interest encompassed predominantly of tree cover, grassland, bare/sparse vegetation and shrubland. Tree cover shows the best dating skill, with 87.9% being dated within the window of the event. No major land-cover type showed a clear reduction in dating performance within this study area. Built-up regions and Cropland provide insufficient sample sizes to comment on the reliability of this method, although it is noted none of the four landslides that within built-up areas are dated correctly.

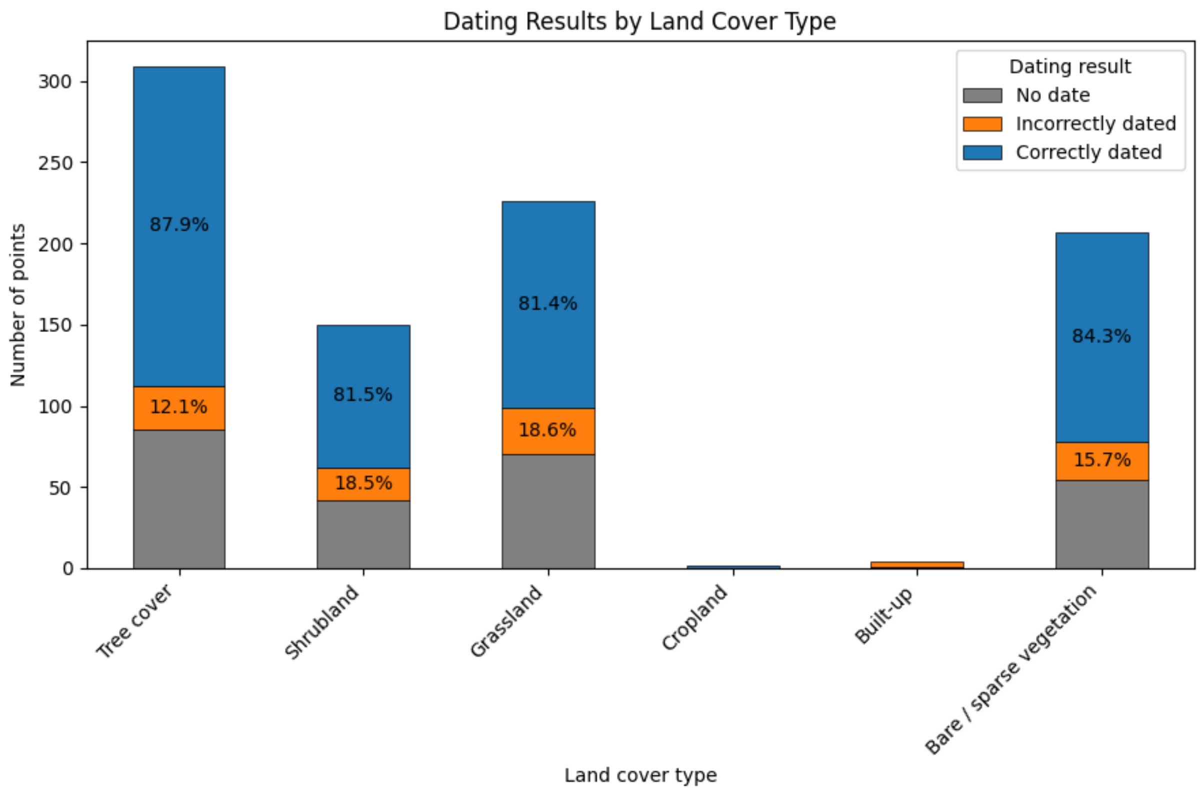


Figure: Land use of each landslide, showing the landslides that were note dated (grey), were dated incorrectly (orange) and correctly (blue).

MANUSCRIPT CHANGES

[Added Line 189]

The distribution of landcover was also explored within these results to assess whether certain regions suggested significant impacts on the reliability of the results. The area is comprised predominantly of tree cover, shrubland, grasslands and bare/sparse vegetation. Of these four regions, tree cover shows the best results, with 87.9% of the dated landslides being dated across the desired event window. However, the other three categories show little variation, all falling within 3% of the overall dataset accuracy. Therefore, no landcover classification included within this case study was shown to require filtering. The sample size for built-up areas was too small to make significant claims, however all four of the landslides within these areas were dating incorrectly.

REVIEWER 2: COMMENT 14

The Conclusions occasionally repeat numerical results already discussed extensively. This section could instead focus more on the broader implications and future applications of the workflow.

RESPONSE

The numerical values within the conclusion have been reduced to the key values.

A paragraph on potential future work or expansions of this method has been provided.

MANUSCRIPT CHANGES

[Edited Conclusion to reduce repetition of details]

[Added Line 364]

This methodology demonstrates the potential of Ruptures changepoint detection in determining long-term optical observations for rapid changes in the landslides. Similar approaches could be applied to the retrospective identification of events such as wildfires, floods, landslides or rapid anthropogenic change in different regions with a range of indices.

REVIEWER 2: COMMENT 15-16

There are occasional inconsistencies in capitalisation (e.g., "Step 1", "step 1"). Some figure captions are overly detailed and could be shortened. Ensure all abbreviations (AOI, BSI, VV, VH, L1, L2) are defined at first appearance. Carefully proofread for minor grammatical issues and repeated wording.

RESPONSE

Thank you for this comment, which has led to an edits in the manuscript.

MANUSCRIPT CHANGES

Capitalisation of Step has been fixed throughout the document. Abbreviations have been defined at their first appearance within the text. The document has been proofread for grammar issues and repeats.

Community Reviewer

COMMUNITY REVIEWER: COMMENT 1

First, the work by Dille et al. (2026). This study investigates the same event as your paper, i.e. the landslides triggered by cyclone Idai. This work, that was published after your manuscript has already been submitted for review, provides a substantially more comprehensive landslide inventory. Nearly 12,000 landslide source areas (predominantly shallow) are mapped as polygons, together with a comparable number of runout areas. Overall, this inventory contains approximately ten times more mapped features than the inventory of Emberson et al (2022), which forms the basis of your analysis. I believe these differences, which may raise questions about the completeness and representativeness of the Emberson et al. (2022) inventory, could be acknowledged and briefly discussed in your manuscript.

RESPONSE

We thank Olivier for spending time to provide comments for this paper, and for highlighting the recent inventory produced by Dille et al. (2026). This study was published after the initial submission of our manuscript and therefore was not available during the development of the methodology. We agree that the substantially larger number of mapped landslides may have implications for the completeness and representativeness of the Emberson et al. (2022) inventory used in this study. We have therefore added a discussion acknowledging the newer inventory and its potential implications for future applications and validation of the proposed workflow.

MANUSCRIPT CHANGES

[Added Line 334 before Reviewer 2 comment 7]

Furthermore, since the completion of this study, Dille et al. (2026) have published a substantially larger inventory for this region, containing approximately 12,000 sources. This is more extensive than Emberson et al. (2022), which highlights the further issue of inventory completeness. Future work could apply the workflow presented here to the Dille et al. (2026) inventory to assess the transferability of the method and its performance across a more comprehensive set of landslides.

COMMUNITY REVIEWER: COMMENT 2

Second, the work by Deijns et al. (2024). This study presents a semi-automatic approach for landslide detection based on times series of eight spectral indexes, including NDVI. The approach of Deijns et al. (2024) was calibrated and validated using rainfall triggered landslides in Africa as well and relies on Sentinel-2 imagery. In addition to mapping landslides, the approach enables the timing of landslide occurrence to the determined. The resulting inventory comprises nearly 4,000 mapped landslides and may therefore provide useful context for your discussion on landslide dating methodologies.

RESPONSE

We thank the reviewer for highlighting Deijns et al. (2024). This study has now been added to Table 1, which summarises existing landslide dating methodologies and provides additional context for comparison with the approach presented here.

MANUSCRIPT CHANGES

[Added to Table 1]

Study	Data Soruce	Data Type	Landslide Data	Geometry Input	Satellite Data Length	Resolution	Accuracy	Temporal Uncertainty
Deijns et al. 2024	Sentinel-2	Multi-spectral	3900 landslides	Geometries	Varied	10mx10m	NA	2-4 weeks