Authors' response to comments

We would like to thank the two reviewers for their comments, which have improved the manuscript greatly. Here we provide a complete response to comments from both reviewers, along with the revised manuscript in which all changes are indicated.

Reviewer 1

Comment 1

Line 5 "triggering events" You can specify the type of events

Manuscript Change: We have changed this to "*triggering earthquakes*" (Line 5, revised manuscript)

Comment 2

Line 6 "ARIA", specify the acronym

Response: We are renaming this method based on comments from reviewer 2 so this will be changed to "co-event coherence loss method"

Manuscript change: "*the ARIA method performs best*" at line 6 of original manuscript has been changed to "*co-event coherence loss (CECL) is the best performing method*" at line 6 of the revised manuscript.

Comment 3

Line 10 "useful landslide density", I would say landslide mapping

Response: We used landslide density so that readers do not expect individual mapped landslides, which is the most common output from mapping landslides using optical satellite imagery.

Manuscript change: none

Comment 4

Line 22 "could have been much longer had the earthquake occurred during Nepal's monsoon season", check this

Response: This statement is supported by Robinson et al. 2019, which is referenced earlier in the same sentence

Manuscript Change: none

Comment 5

Line 30 "reliability of global empirical rainfall-triggered landslide susceptibility maps", the scale itself is a limitation

Response: The spatial scale of the empirical models may be a disadvantage for some applications, but immediately after the earthquake, low resolution information over a large spatial scale is useful for giving an overview of severely impacted areas to emergency response coordinators (Discussed in Williams et al. *NHESS*. 2018). Although our analysis is carried out at a slightly higher resolution than these empirical models (200 x 220 m pixels as opposed to 1 km2, which is used by Nowicki-Jessee), the improvement in resolution in this study is fairly limited.

Manuscript Change: Text added at line 26 of revised manuscript: "*The outputs of these models have a comparatively low spatial resolution (around 1 km2 in the case of Nowicki Jessee et al. 2018), but can be used to provide an overview of the most severely impacted areas*"

Comment 6

Line 31 "style of landsliding,", I would say landslide type

Manuscript Change: "*style of landsliding*" at line 31 of original manuscript changed to "*type of landslides*" in line 33 of revised manuscript

Comment 7

Line 41 there are thousands of published paper in this field. I think the reference list can be extended, see some recent examples: _Dai, K., Li, Z., Tomás, R., Liu, G., Yu, B., Wang, X., ... & Stockamp, J. (2016). Monitoring activity at the Daguangbao mega-landslide (China) using Sentinel-1 TOPS time series interferometry. Remote Sensing of Environment, 186, 501-513. _Solari, L., Del Soldato, M., Raspini, F., Barra, A., Bianchini, S., Confuorto, P., ... & Crosetto, M. (2020). Review of Satellite Interferometry for Landslide Detection in Italy. Remote Sensing, 12(8), 1351. _Aslan, G., Foumelis, M., Raucoules, D., De Michele, M., Bernardie, S., & Cakir, Z. (2020). Landslide Mapping and Monitoring Using Persistent Scatterer Interferometry (PSI) Technique in the French Alps. Remote Sensing, 12(8), 1305. _Reyes-Carmona, C., Barra, A., Galve, J. P., Monserrat, O., Pérez-Peña, J. V., Mateos, R. M., ... & Azañón, J. M. (2020). Sentinel-1 DInSAR for Monitoring Active Landslides in Critical Infrastructures: The Case of the Rules Reservoir (Southern Spain). Remote Sensing, 12(5), 809. _Hu, X., Bürgmann, R., Lu, Z., Handwerger, A. L., Wang, T., & Miao, R. (2019). Mobility, thickness, and hydraulic diffusivity of the slowˇARˇmoving Monroe landslide in California revealed by LˇARˇ band satellite radar interferometry. Journal of Geophysical Research: Solid Earth, 124(7), 7504-7518.

Response: Thank you for drawing our attention to these materials, these references have now been added

Manuscript Change: References at line 41 (original manuscript) "*e.g. Boni et al, 2018; Handwerger et al., 2019*" changed to "*e.g. Aslan et al., 2020; Bonì et al., 2018; Dai et al., 2016; Handwerger et al., 2019; Hu et al., 2019; Reyes-Carmona et al., 2020; Solari et al., 2020)*" at line 43 of the revised manuscript. Aslan et al., (2020); Dai et al., 2016; Hu et al., 2019; Reyes-Carmona et al., 2020; Solari et al., 2020 have been added to the bibliography.

Comment 8

Line 68 "method"–> technique

Manuscript Change: "*method*" at line 68 of original manuscript changed to "*technique*" at line 71 of the revised manuscript.

Comment 9

Line 77 any reference for the Boxcar method?

Response: This is the most common method of calculating coherence, we have altered the text to make this clearer and added some examples of texts that use this method.

Manuscript Change: Line 77 of original manuscript "*The ensemble of pixels used in the calculation can be selected either using a 'boxcar' method, in which a small rectangular ensemble of pixels immediately adjacent to and centred on the target pixel are used…*" changed to "*The ensemble is chosen so that the pixels used in the calculation are expected to be similar. In a "boxcar" method, it is assumed that pixels immediately adjacent to and centred on the target pixel are similar to it (e.g.*
Hanssen, 2001; Yun et al., 2015)." at line 81 of the revised manuscript. A reference for Hansen, 2001 has been added to the bibliography.

Comment 10

Line 131 "landslides were triggered by the earthquake", the occurrence of a typhoon the day before is certainly a trigger as well.

Response: We agree that the typhoon the day before the earthquake will have caused widespread pore-pressure increase across the study area and that this may have played an important role in the initiation of some of the observed landslides. However, the landslides have been widely attributed to the earthquake as the primary trigger, with debate over the role of rainfall being focussed on the extent to which it pre-conditioned the slopes for failure (e.g. Osanai et al., 2019; Wang et al., 2019; Yamagishi and Yamazaki, 2018; Zhang et al., 2019). We are interested in this event because of the close sequencing of intense rain and shaking rather than as a test of our ability to distinguish the different triggers. We have removed reference to the earthquake as the trigger since this is unnecessary and replaced this with information on how the landslide data were collected.

Manuscript Change: "*with the advantage that because the landslides were triggered by the earthquake, we know more precisely when they occurred*." at line 130 of the original manuscript changed to "*with the advantage that because the typhoon and earthquake occurred one day apart, and aerial imagery of the triggered landslides was acquired immediately afterwards, we know more precisely when the landslides occurred (Yamagishi and Yamazaki, 2018)*." at line 143 of the revised manuscript. A reference for Yamagishi and Yamazaki (2018) has been added to the bibliography.

Comment 11

L140 and followings, what about the type of landslides that have been triggered?

Response: Thank you for identifying this omission, we will add information on landslide type.

Manuscript Change: Text added at line 127 of the revised manuscript. "*Finally, the type of landslides triggered by our four case study earthquakes were typical of landslides triggered by earthquakes (Keefer, 1984). The majority of ground failures in the four earthquakes were slides. In Nepal, ground failures were primarily a mixture of slides and falls, with the exception of a large debris avalanche in the Langtang Valley. For all four earthquakes, failure surfaces were at shallow depths in most cases with a small number of exceptions (Collins and Jibson, 2015; Ferrario, 2019; Yamagishi and Yamazaki, 2018)*." References for Keefer (1984) and Yamagishi and Yamazaki (2018) have been added to the bibliography.

Comment 12

Fig.1 is the inset in (c) Lombok 2? You should specify this in the caption

Response: Thank you for identifying this omission, we will add this to the figure caption

Manuscript Change: The Figure caption for Figure 1 has been changed from "*Figure 1. (a, b, c) Landslides triggered by the 2015 Gorkha, Nepal earthquake, the 2018 Hokkaido, Japan earthquake, and the 2018 Lombok, Indonesia earthquakes, respectively, plotted as the areal density of landsliding based on 1 km2 cells (Roback et al., 2018; Zhang et al., 2019; Ferrario, 2019). ALOS-2 SAR acquisitions are shown in green and Sentinel-1 in blue. (d, e, f) Acquisition dates and track numbers of SAR imagery used in this study. Earthquakes are shown as red vertical lines, black symbols show pre-event image acquisition dates, and red symbols show post-event dates.*" to "*Figure 1. (a, b, c) Landslides triggered by the 2015 Gorkha, Nepal earthquake, the 2018 Hokkaido,*

Japan earthquake, and the 2018 Lombok, Indonesia earthquakes, respectively, plotted as the areal density of landsliding based on 1 km2 cells (Roback et al., 2018; Zhang et al., 2019; Ferrario, 2019). ALOS-2 SAR acquisitions are shown in green and Sentinel-1 in blue. In (c), a second earthquake, referred to as "Lombok-2" is inset. (d, e, f) Acquisition dates and track numbers of SAR imagery used in this study. Earthquakes are shown as red vertical lines, black symbols show pre-event image acquisition dates, and red symbols show post-event dates."

Comment 13

L185 check the double "by a landslide"

Manuscript Change: Second "*by a landslide*" removed at line 206 of the revised manuscript.

Comment: L379 and one day with images from ROSE-L

Response: Thank you, I was not aware of this satellite and have added a reference to this mission

Manuscript Change: "*In the future, however, the availability of L-band SAR is likely to increase with the planned NiSAR satellite constellation, which will acquire data continuously with a 12 day repeat time over all landmasses globally (Sharma, 2019).*" at line 377 of original manuscript changed to "*In the future, however, the availability of L-band SAR is likely to increase with the planned NISAR and ROSE-L satellite constellations. NISAR, a joint NASA-ISRO mission planned to launch in 2022, will acquire L-band data continuously with a 12-day repeat time over all landmasses globally (Sharma, 2019), while the ESA ROSE-L satellite, whose launch date is planned in 2026, will have 6-day global repeat coverage, and 3-day in Europe (Pierdicca et al., 2019).*" at line 414 of the revised manuscript. A reference for Pierdicca et al (2019) has been added to the bibliography.

Comment 14

L398 do you have an estimate of the number/percentage of false positives due to the buildings? The same comment for the number of false positives due to wind or snow.

Response: It is difficult to quantify the number of false positives resulting from a specific source as we do not threshold the classifiers in our analysis. However, we have tried to assess the impact of buildings by repeating the ROC analysis with urban areas masked based on the ESA CCI landcover map. We found very little change in ROC AUC (<0.02), as these land cover types make up a relatively small part of the study area.

Quantifying false positives in snow-covered areas is difficult, since these are also likely to be underrepresented in the landslide inventory generated from optical satellite imagery (Roback et al. 2018). Furthermore, the dependence of snow cover on season, elevation and climate means that any measure we obtain of false positives due to snow is unlikely to be applicable in future events.

Quantification of false positives due to wind damage is also difficult as there are no reliable validation data available for this. From visual inspection, it appeared that the signal was strongest for areas mapped as "evergreen needleleaf forest" but masking these areas did not noticeably improve ROC AUC (<0.02).

It is difficult to predict how these false positives might affect the applicability of SAR data to future events, as this will depend on the extent of urban development (in the case of building damage), the climate, elevation and season (in the case of snow) and on the occurrence of storms (in the case of wind damage). Therefore, while we note that these are potential causes of error, we do not think it is useful to try to quantify these further.

Manuscript change: none

Comment 15

You could mask the rivers. Since the single landslide is not the target of your approach, the detection of a single landslide dam should not be possible and the river-mask duable.

Response: We agree this would be possible, but although we cannot map individual landslide dams, points where rivers pass through areas of high landslide density can be used to identify locations where a landslide dam is more probable (e.g. Robinson et al*. BSSA*, 2018) and therefore which are worthy of further investigation. Therefore, although we acknowledge that a landslide signal from SAR data is less reliable in a river than elsewhere, masking them would lead to a more confusing product where landslides are not allowed to intersect with rivers. Instead we suggest that rivers should be drawn over the coherence products. In response to the comments in this review and the following comment made by the other reviewer "*Line 425 - You mention rivers giving false positives for the Bx-S method. Would it be possible to identify rivers by applying the Bx-S method to a pre-seismic coherence image (using the same pixel siblings) and then use that as a mask on your co-seismic landslide image? You could propose this if so.",* we will add the following sentences to the manuscript:

Manuscript Change: "*A variety of methods could be used to identify and remove rivers from our analysis (including using a pre-event Bx-S surface). However, since areas where landslides and rivers intersect are particularly hazardous due to the potential for landslide dams and associated flash flooding, we did not mask rivers in this study. We suggest that any product based on SAR coherence supplied to emergency response*

coordinators should have rivers overlaid. This would both mask false positives due to rivers and allow identification of locations where rivers pass through areas of intense landsliding." added at line 469 of revised manuscript.

Reviewer 2

Comment 1

Quantification of different methods:

I would like to see further discussion of the use of the ROC curve to evaluate these classification methods. My understanding is that ROC curves are generally best deployed when the dataset is balanced between positive and negative examples (i.e. roughly equal numbers of landslide and non-landslide pixels), however I wonder if that is the case here, or if there are many more non-landslide pixels than landslide pixels in general?

If it's the case that the dataset is imbalanced, with more non-landslide pixels, then the ROC curve might not be the best way to assess how well your method is performing. **For example, if I take a balanced dataset then add a large number of negative examples, with their classification scores drawn from the same distribution at the existing negative examples, the false positive rate (FPR=FP/(FP+TN)) and true positive rate(TPR=TP/(TP+FN)) remain the same for a given threshold, and the ROC curve doesn't change**. However I now have many more false positives at a given threshold, meaning the precision of my classifier (=TP/(TP+FP)) will decrease (i.e. a smaller fraction of my positive classifications will be true positives). In some circumstances it could be the case that only a small fraction of my samples that are classified as positive are actually true positives, even as my TPR and FPR appear good.

Additionally, as I expand the region spanned by the SAR data, it's possible that I include more and more pixels that will have lower noise levels, e.g. as they're further away from the earthquake and so have less of the building damage, surface rupture, liquefaction etc. that can lead to false positives for landslide classification. This would lead to having a larger number of true negatives, and so an improved false positive rate, thus an improved ROC curve, but would have minimal effect on the precision, which doesn't consider true negatives.

In addition to the currently presentation, you should consider presenting a precision-recall curve, or mention why the the ROC curve is preferred over the precision-recall curve. It would also be good to have you mention the fraction of each region that is covered with landslide and non-landslide pixels.

Response: We thank the reviewer for the suggestion of using precision recall (P-R) instead of ROC. There are pros and cons to both methods and we have now checked that our conclusions are not affected by our decision to use ROC rather than P-R curves. This is because our conclusions depend on relative ranking of methods using these curves rather than absolute AUC values, and these rankings are insensitive to the choice between ROC and P-R. However, we choose to retain ROC based results within the paper. They are more appropriate than P-R for this study for two reasons

• First, true negatives are not used at all in calculation of P-R curves. Therefore P-R is only concerned with correct prediction of the smaller positive class (i.e. landslides for this study) and implicitly suggests that we are not interested in correct prediction of the larger negative class (non-landslides). Whilst this will be appropriate for many imbalanced classification problems, it is not appropriate for landslide mapping, where it is important to accurately identify where landslides have not occurred as well as where they have occurred. This is important for correct allocation of resources in emergency response.

• Second, the example above from the reviewer (highlighted in bold) highlights that the ROC curve is essentially insensitive (or at least very weakly sensitive) to the size of the area of negative samples (i.e. the area of the data with no landslides). This is in contrast to P-R curves, which are strongly sensitive to the size of the area of the dataset that has no landslides. In our case, this is exactly why ROC is more appropriate than P-R and is a strong argument in favour of using ROC. The proportion of landslide pixels in each dataset varies significantly between the four events, and is dependent both on landslide density within the landslide affected area (which is event-specific) and on the area of data that is processed, i.e. how much SAR data is processed beyond the region of intense landsliding (which is an arbitrary processing decision, and could depend on the frame size, which can vary across different satellites). We wish to be able to compare performance of our classification methods across these events in a fair manner, and also do not want our performance metric to depend on arbitrary processing choices. The baseline of a precision-recall curve is dependent on the proportion of landslide pixels in the dataset, and this varies significantly between the four events (and even between different sensors, due to differences in frame size). Therefore, P-R AUC values vary a lot between events, which distracts from the main result. A classifier with no ability will consistently have an ROC AUC of 0.5 for all events, which makes this metric more suitable when carrying out systematic analysis across several events and sensors as we have done here.

We have, however, tested using P-R instead to check that the relative *rankings* of the methods don't change for each SAR track, even if the absolute values do change and preclude useful cross-comparison. The results are very similar, and we include these results in an appendix. If the methods are ranked for each SAR track (to avoid landslide density effects on P-R AUC), the ranking for ROC AUC and P-R AUC is identical in 46/59 of cases for the results in Figure 2 and only differ by more than 1 rank in 6 cases. The conclusions we draw from our ROC analysis are the same as those that we would draw from P-R analysis. As the fraction of landslide pixels in the validation dataset for each track represents the baseline P-R AUC, this information is included in this appendix. We have also added a short section to the main text, referring to the appendix and clarifying why ROC is more appropriate for this study.

Manuscript Change: text added at line 285 of revised manuscript: *"On all SAR tracks, there are many more non-landslide than landslide pixels. It has been suggested that for such imbalanced data, precision-recall curves can better represent classification ability than ROC AUC (Saito and Rehmsmeier, 2015). Here, we chose to use ROC analysis since precision-recall curves do not allow comparison between datasets with different proportions of landslide and non-landslide pixels and therefore between different earthquakes and SAR tracks. However, when considering the relative performance of classifiers for each track independently, we found the same conclusions could be drawn from precision-recall curves as from ROC curves. A recreation of Figure 2a using precision-recall rather than ROC AUC values can be found in Supplementary Information."* A reference for Saito and Rehmseier (2015) has been added to the bibliography.

Comment 2

Histogram matching:

The use of histogram matching between the pre- and co-seismic coherence images could do with some further explanation so the reader understands the motivation and assumptions. My understanding is that this adjustment assumes that only a small number of pixels are anomalous (i.e. contain landslides)ã˘Aˇ˘T otherwise you would end up removing the signal you were looking for. Furthermore, I think there is an assumption that the coherence of the pixels in the image that's adjusted have all been affected in the same way. For example, if only the southern part of the image had been covered in snow between the final pre-seismic and first post-seismic SAR images, then adjusting the entire coseismic coherence image based on a simple matching of histograms would not be the correct approach, however if the second coherence image just had a longer temporal baseline then the extra temporal decorrelation might be removable by histogram matching.

Response: Yes that is correct. We will clarify this in the text.

Manuscript Change: *"This process accounts for different levels of bulk temporal decorrelation in the pre-event and co-event interferograms. It assumes only a small fraction of the pixels are affected by landslides so that the landslide signal is not removed from the co-event interferogram."* added at line 184 of revised manuscript.

Comment 3

Optimum thresholds:

It might be helpful for the discussion to go into more detail on what the optimum threshold for flagging landslides would be for each method, as this is what would be required before use by a first responder. Currently the presentation of your results is threshold free (i.e. in terms of the ROC curve), apart from when you make plots. The discussion around line 300-305 explains that you choose a threshold for plotting in order to be consistent with the number of points plotted in the landslide area plot in Fig 4, but it's not clear if this would be a useful threshold for response. The comparison you chose to present in figure 4 was also slightly confusing to me; in line 236 you say that you're only classifying pixels as 'landslide' pixels if they have >25% area of landslides, however for the plot in figure 4 you plot pixels that have >1% area of landslides, and then choose the classification threshold based on the number of pixels that have >1% landslide area. Is there a reason for the different threshold of landslide area? Would it not be better to choose the classification threshold from some desired trade off between true positive and false positive rates on the ROC curve (or a trade off between precision and recall)?

Response:

This comment relates to three different thresholds: 1) the minimum landslide area density below which landslide density is not displayed in Figures 3 and 4; 2) the minimum classifier value below which the classification surface is not displayed in Figures 3 and 4; and 3) the minimum landslide area density below which a pixel is not considered 'landslide affected' in ROC and P-R performance evaluation. We have chosen to approach these differently as they have different purposes.

The first two are 'display thresholds', and as the reviewer points out this would not necessarily be a useful threshold for response, in fact, establishing the threshold used in (2) would not be possible since the landslide density would be unknown. However, these 'display thresholds' only affect the look of the figures rather than our findings, and are selected simply to graphically illustrate our results in an appropriate way. In the 'ground

truth' maps of observed landslide density (Figures 3g and 4b, d, f) we chose to plot landslide areal density and to mask pixels with landslide area density <1%. This threshold of 1% was chosen because it captures the majority of the landslide-affected pixels. We chose the thresholds for the coherence-based surfaces such that the masked area in each would be approximately equal to that in the observed landslide density map. We agree, we have not explored this choice of threshold here and whether it would be suitable for emergency response. Since the figure shows ALOS-2 data and the time between image acquisitions for these data varies between events, which affects the coherence, we did not expect to be able to extract a threshold that could be applied in future events at this stage. Choosing the classification threshold based on a trade-off between true positive rate and false positive rate, or precision and recall would result in differences between events in the figures that might not fairly represent how well they were modelled. We have added text to Section 5.3 (Future Work) in the Discussion to make this clear.

The final threshold is required to coarsen the resolution of the landslide observations to match that of the classifiers. In this case, we chose to set the threshold higher than the 1% threshold used in Figures 3 and 4 and considered only cells with >25% landslide area to be 'landslide affected' for the purpose of ROC and precision-recall analysis. Pixels containing more landslides are more important to identify, for example because the chance that a road within the pixel has been blocked by a landslide is higher. But if the threshold is set too high, not enough pixels are classed as 'landslide' to be able to carry out meaningful ROC or precision-recall analysis, especially in Lombok where the landslide density was lowest, and many regions of landsliding will be missed by the classifier . The choice of 25% is therefore somewhat arbitrary but represents a trade-off between reliability of the ROC and P-R analysis and sensitivity to pixels less affected by landslides.

We have also tested at other thresholds (1%, 10% and 50%). These results are now included as an additional figure in Supplementary Information and referred to in the main text. In general, more severely impacted pixels have a stronger signal in the coherence surfaces and so are easier to detect. Therefore, increasing this threshold (e.g. to 50%) produces a higher ROC AUC. The difference in ROC AUC between setting this threshold at 1% and 25% can be up to 0.18, but in most cases was less than 0.1. Again, this choice makes little difference to the relative ranking of the methods and thus does not influence our conclusions.

Manuscript Change:

"We explore the effect that this choice of a 25% threshold to define a landslide pixel has on our results in the supplementary material, but varying this threshold between 1% and 50% was found to have little effect on the relative performance of different classifiers." added at line 264 of the revised manuscript.

"Finally, here we assessed classifier performance using ROC analysis, which does not require a threshold to be applied to the classification surface. However, if SAR methods are to be applied to future events for emergency response, it will be necessary to set a threshold between 'landslide' and 'non-landslide'. In this study the time between image acquisitions varied significantly between events, making it unlikely that a threshold could be selected that would work well for both Sentinel-1 and ALOS-2 across all events. However, this may be possible in the future when more events have been studied and SAR data with more regular acquisitions are available. Further work is therefore needed to establish such thresholds, which will be determined according to the requirements of emergency responders and their relative tolerance for false positives and false negatives" added at line 510 of the revised manuscript.

Comment 4

'ARIA' method naming:

The use of the descriptor 'ARIA' method isn't standard to my knowledge, and may be misleading. JPL's ARIA Project does employ this method, although it is also working on many alternative approaches. Furthermore, the method, or similar variants of it, has been used by many groups (e.g. Fieldling et al (2005)) and is not referred to as the 'ARIA' method in these publications. One term I have seen applied is 'Coherence Change Detection' (e.g. Washaya et al. 2018), however this may not be the most useful, as all of your methods are change detection using coherence. Perhaps 'coherence loss' would be a better term?

Citations for the paragraph above:

- Fielding, E. J. et al. (2005) 'Surface ruptures and building damage of the 2003Bam, Iran, earthquake mapped by satellite synthetic aperture radar interferometric correlation', Journal of Geophysical Research: Solid Earth, 110(3), pp. 1–15. doi:10.1029/2004JB003299.

- Washaya, P., Balz, T. and Mohamadi, B. (2018) 'Coherence Change-Detection with Sentinel-1 for Natural and Anthropogenic Disaster Monitoring in Urban Areas', Remote Sensing, 10(7), p. 1026. doi: 10.3390/rs10071026.

Response: Thank you for drawing our attention to these papers, we will add these to the text and include a more complete description of damage-detection methods based on coherence loss. To match the naming of the method we refer to as "post-event coherence increase" (PECI), we will alter the text and refer to the "ARIA" method as "co-event coherence loss" (CECL)

Manuscript changes: the acronym *"ARIA"* has been altered to *"CECL"* throughout the manuscript, and the method renamed to "Co-event coherence loss".

"This method was developed by the NASA Advanced Rapid Imaging and Analysis (ARIA) project for use in urban damage mapping… " at line 164 of the original manuscript has been changed to *"The coherence loss between a pre-event interferogram and a co-event interferogram can be used to detect physical changes to the ground surface associated with an earthquake, such as surface rupture, building damage and landslides (Fielding et al., 2005; Washaya et al., 2018; Yun et al., 2015). This method has been applied by the NASA Advanced Rapid Imaging and Analysis (ARIA) project for use in urban damage mapping…"* at line 179 of the revised manuscript. References for Washaya et al. (2018) and Yun et al. (2015) have been added to the bibliography.

A reference to Fielding et al. (2005) was added at line 46 of the revised manuscript.

Comment 5

Line 73 - ï z£'The signal-to-noise ratio of each pixel in an InSAR image is described by its coherence'. This statement feels incomplete to meã̆Aˇ Tin your work the 'signal' is the decorrelation of the pixel, and in circumstances where the phase is the signal (e.g. deformation time series), the decorrelation is only one contributor to the noise (e.g. can also have atmospheric and ionospheric noise which doesn't affect the coherence).

Response: By 'signal to noise', we meant the signal quality of the interferogram when it is used for ground deformation studies, not the signal in coherence studies, and were referring to high-frequency noise, not the longer wavelength nuisance signals that arise from ionospheric or tropospheric phase changes. We have reworded this statement to make it clearer.

Manuscript Change: *"The signal-to-noise ratio of each pixel in an InSAR image is described by its coherence. The coherence of a pixel can be estimated from the similarity between two SAR images in amplitude and phase difference, for a small ensemble of pixels n (Eq. 1, Just and Bamler, 1994):"* at line 72 of the original manuscript has been changed to *"When using an interferogram to map ground deformation, it is important that the signals recorded at a given location in the two SAR images are correlated, as decorrelation will result in high-frequency noise. In order to assess this and to identify noisy pixels, the coherence is estimated for every pixel from the similarity in the two SAR images in amplitude and phase difference, for a small ensemble of n pixels (Eq. 1, Just and Bamler, 1994):"* at line 75 of the revised manuscript.

Comment 6

Line 77 - I think some mention of the assumptions of the boxcar and sibling methods would be useful. My understanding is that the box-car method assumes that the surrounding pixels are statistically similar, and the sibling method identifies pixels that are similar through time (based on the amplitude) and assumes that these remain similar for subsequent acquisitions. I note that the sibling method is briefly explained on line 212, but more information in the introduction I feel would be helpful to the reader.

Response: Yes this is correct, we have added more information here for clarity

Manuscript Change: *"The ensemble of pixels used in the calculation can be selected either using a 'boxcar' method, in which a small rectangular ensemble of pixels immediately adjacent to and centred on the target pixel are used, or a 'sibling' method (e.g. Spaans and Hooper, 2016) where the ensemble of pixels is selected from within a wider window"* at line 77 of the original manuscript changed to *"The ensemble is chosen so that the pixels used in the calculation are expected to be similar. In a "boxcar" method, it is assumed that pixels immediately adjacent to and centred on the target pixel are similar to it (e.g. Hanssen, 2001; Yun et al., 2015). In a "sibling" method an assessment is carried out for every pixel to identify pixels that are statistically similar to it. For example, the sibling method of Spaans and Hooper (2016) identifies pixels that have similar amplitude behaviour through time."* at line 81 of the revised manuscript.

Comment 7

perhaps include a definition of perpendicular baseline, and mention that orbital controls on modern satellites have rendered this a much smaller issue?

Response: Thank you, we have added this to the text.

Manuscript Change: *"When the perpendicular baseline of the SAR image pair used to form an interferogram is sufficiently small, this spatial component will be small compared to any temporal decorrelation (Zebker and Villasenor, 1992)"* at line 87 of the original manuscript changed to *"This decorrelation is particularly sensitive to the SAR image pair's perpendicular baseline (the distance between the locations at which the satellite acquired the two SAR images measured perpendicular to both the flight and look directions). When the perpendicular baseline of the image pair used to form an interferogram is sufficiently small, this spatial component will be small compared to any temporal decorrelation (Zebker and Villasenor, 1992). For modern satellites, this will be the case most of the time."* at line 93 of the revised manuscript.

Comment 8

Line 100 - add citation on the effect of vegetation on coherence at different wavelengths?

Response: We have cited Zebker and Villasenor (1992) here, who compare decorrelation due to movement of scatterers at L-band and C-band wavelengths.

Manuscript Change: Reference to Zebker and Villasenor added at line 108 of the revised manuscript

Comment 9

Line 106 'as it is impossible to combine data from both tracks' - it would be helpful to clarify what is meant by 'combine' (i.e. you can't calculate the coherence between different tracks)

Manuscript Change: "*As it is impossible to combine data from both tracks*" at line 106 of the original manuscript has been changed to "*As it is impossible to calculate a combined coherence surface using data from two tracks*" at line 113 of the revised manuscript

Comment 10

Line 119 - what is meant by 'high resolution' data for ALOS-2? Might be helpful to mention the acquisition mode for clarity?

Response: We have specified in the text that these data were acquired in "stripmap" mode at a resolution of 3 – 10 m. The resolutions of each track can be found in Table 1.

Manuscript Change: "*(acquired in stripmap mode at a resolution of 3 - 10 m*)" added at line 127 of the revised manuscript.

Comment 11

Line 165 - need to clarify that the user must choose a coherence loss threshold for flagging damage, the current sentence structure implies that all pixels that have any coherence decrease at all are flagged as damaged

Response Here we apply this method to landslides and do not apply a threshold for the ROC analysis, but in the original Yun et al. 2015 paper, they use 0 as the threshold, so that any pixel whose coherence decreases is flagged as damaged. We have clarified this in the text.

Manuscript Change: "*The pre-event surface is then subtracted from the co-event surface (Yun et al., 2015).*" at line 167 of the original manuscript changed to "*The pre-event surface is then subtracted from the co-event surface and pixels whose coherence has decreased are flagged as damaged (Yun et al., 2015*)." at line 187 of the revised manuscript

Comment 12

Line 179 - I think it's worth reminding the reader here that the sibling pixels have been specifically identified to be behaving similarly before the earthquake by looking at the amplitude time series, and it's on that basis that they're expected to behave similarly

Response Yes, that's a good suggestion, we have added this to the text.

Manuscript Change: "*When a sibling-based method such as that of Spaans and Hooper (2016) is used to estimate coherence, the coherence of a pixel becomes dependent on 'siblings' that are not immediately adjacent to it but that are expected to behave similarly.*" at line 177 of the original manuscript changed to "*When a sibling-based method is used to estimate coherence, the coherence of a pixel becomes dependent on "siblings" that are not immediately adjacent to it but that are expected to behave similarly. In the method of Spaans and Hooper (2016) used here, an ensemble of siblings is selected for every pixel that have similar amplitude behaviour in a time series of pre-event imagery*" at line 197 of revised manuscript.

Comment 11

Line 185 - post-event coherence - make it clear that this coherence is calculated using the boxcar method (which I assume it is?)

Response: This was stated at line 176, but we have altered the text to ensure it is clear that the boxcar coherence is used in sections 3.3.1, 3.3.3, 3.3.4, and 3.3.5.

Manuscript Change "*A "boxcar" coherence estimate is used for CECL, PECl, ΔC_sum and ΔC_max.*" added at line 176 of the revised manuscript.

Comment 12

Line 233 - 'If over 95% of an aggregate pixel is made up of masked pixels, the aggregate pixel was masked' - how do you choose this number? Are you results sensitive to this choice? What fraction of pixels end up being masked out?

Response: The effect of this choice on ROC AUC is very low (< 0.02). However, altering this can significantly alter the number of pixels masked, for example on track 19 in Nepal, altering this threshold from 95% to 5% decreases the number of aggregate pixels used in the analysis from 246,375 to 148,829. The effect was much weaker in Hokkaido and Lombok, where fewer pixels were masked due to distortion on steep slopes. For example, in Hokkaido on track 68, altering the threshold from 95% to 5% decreases the number of aggregate

pixels from 19,754 to 19,235. We chose to set the threshold high to avoid masking any landslides unnecessarily and to therefore maintain good spatial coverage for our classification surface.

Manuscript change*: "This high threshold of 95% was chosen to minimise the loss of spatial coverage due to the masks. Varying the threshold between 95% and 5% had little difference in terms of the number of pixels used in the analysis in Hokkaido and Lombok (<5%), but in Nepal, where more pixels were masked due to distortion on steep slopes, decreasing the threshold to 5% resulted in a loss of coverage of around 40% on S085a. Altering this threshold made very little difference to the results presented in Section 4.1.*
." added at line 254 of the revised manuscript.

Comment 13

Line 234 - ï z£'In this study, we did not attempt to map SAR classification surface values directly to landslide areal density values, as this has not been attempted in previous studies and may not be possible' - would be good to get more info about why this is/ isn't possible, and a citation for the previous work. Is this comment based on the work of Burrows et al. (2019)?

Response: All of the studies done so far applying SAR data to landslide detection have used a binary validation landslide dataset and have not attempted to recreate landslide areal density (Aimaiti et al. 2019; Burrows et al. 2019; Ge et al. 2020; Jung and Yun 2019; Masato et al. 2020; Yun et al. 2015). Although it may be possible to extract landslide areal density from SAR data, this is uncertain as differences in viewing geometry, land cover type and (particularly with the ALOS-2 data used here) differences in the temporal baselines of the coherence maps may mean that there is too much variation between events to establish rules on this. This is now clarified in the text.

Manuscript change: "*In this study, we did not attempt to map SAR classification surface values directly to landslide areal density values, as this has not been attempted in previous studies and may not be possible*" at line 234 of the original manuscript changed to *"In this study, we did not attempt to map SAR classification surface values directly to landslide areal density values, as this has not been attempted in previous studies (e.g. Aimaiti et al., 2019; Burrows et al., 2019; Jung and Yun, 2019; Yun et al., 2015) and may not be possible due to differences in viewing geometry, land cover and, particularly with the ALOS-2 data used here, differences in temporal baseline between events."* at line 259 of the revised manuscript.

Comment 14

Line 236 - how did you select the 25% threshold? How much does this choice affect your results? How large is this area, and what fraction of the total landslide area is missed by virtue of the fact that it falls below this threshold? Would be helpful to have these questions discussed in the text, particularly if this choice has a large affect on AUC values

Response: The choice of this threshold has been explained in our response to Comment 3, Reviewer 2**.** The effect of this choice is now demonstrated in Supplementary material where we varied this threshold between 1% and 50%. In general, setting the threshold high results in slightly higher ROC AUC because the signal of an aggregate pixel containing more landslide pixels is stronger but the relative performance of different classifiers remains the same.

For a 200 x 220 m pixel, 25% corresponds to an area of 11,000 m2 out of the total 44,000 m2 area of each aggregate pixel. Applying this threshold results in around 27% of the total landslide area being classified as 'non-landslide' in Hokkaido, around 61% in Nepal and in Lombok, where the average landslide size is comparably small, around 86% for Lombok-1 and 85% for Lombok-2. When the threshold is dropped to 10%, the total excluded landslide area decreases to 6% in Hokkaido, 26% in Nepal and 55% and 53% in Lombok-1 and -2 respectively. Information on the proportion of excluded pixels with the threshold set at 1%, 10%, 25% and 50% is now included along with the supplementary information supplied in response to Comment 3, Reviewer 2

Manuscript Change See comment 3, reviewer 2

Comment 15

Line 277 - 'ï z£the 6-day Sentinel-1 acquisition window' - this is a little unclear, to my mind 'window' implies that data was acquired at some point in that time range. Rather than 'window' could say 'acquisition frequency'?

Response We have changed this to repeat time to avoid confusion.

Manuscript change: "*Acquisition window*" at line 277 of the original manuscript changed to *"acquisition repeat time"* at line 313 of the revised manuscript.

Comment 16

Line 287 - for clarity it may be worth reminding the reader that the L-band image is from the ALOS 2 satellite

Manuscript Change: *"from ALOS-2"* added at line 323 of the revised manuscript.

Comment 17

Line 324 - a brief discussion of what is meant by 'higher quality siblings' would be helpful (I imagine that they're more statistically similar?)

Response: If siblings are selected using SAR images acquired more recently before the earthquake, then there is less time for pixels to have been altered by changes to the ground surface, and their behaviour is more likely to remain similar in the co-seismic image (assuming there are no landslides)

Manuscript change

"As Sentinel-1 imagery is acquired every 12 days, more images were available for this calculation and were acquired over a shorter time period (Fig. 1). Therefore it may be the case that the siblings selected by RapidSAR when using Sentinel-1 imagery were of a higher quality than when using ALOS-2, giving a more reliable sibling coherence estimate." in the original manuscript changed to *"As Sentinel-1 imagery is acquired every 12 days, more images were available for this calculation and were acquired over a shorter time period. This allows less time for pixels to be altered by changes to the ground surface, meaning that, for non-landslide pixels, a pixel and its siblings are likely to be more similar. In this way, the siblings selected by RapidSAR for Sentinel-1 imagery may have been of a higher quality than those for ALOS-2, giving a more reliable coherence estimate."* at line 357 of the revised manuscript.

Comment 18

Line 425 - You mention rivers giving false positives for the Bx-S method. Would it be possible to identify rivers by applying the Bx-S method to a pre-seismic coherence image (using the same pixel siblings) and then use that as a mask on your co-seismic landslide image? You could propose this if so.

Response: Yes this might be possible and could be a good way to remove these false positives. This has been covered in our response to Reviewer 1, Comment 15

Manuscript Change: See Reviewer 1, Comment 15

Comment 19

Line 469 - ï z£ 'The ARIA method is the best performing method when only one L-band image is available' - does this mean 'one post-seismic L-band SAR image'? I think this could be clarified, as I could read this to mean that only on L-band SAR image is available in total.

Response Thank you for pointing this out, we have changed this for clarity

Manuscript Change: *"only one L-band image"* at line 469 of the original manuscript changed *to "only one post-event L-band image"* at line 521 of the revised manuscript.

Comment 20

Figure 6 - how is the threshold for these plots chosen? Could you clarify in the caption?

Response: The 10% of pixels most likely to be landslides are plotted. This has been added to the figure caption

Technical Corrections

4 technical corrections were suggested by the reviewer, all of which have been implemented in the manuscript.

1. Technical Corrections: For words in quotation marks apostrophes are being used throughout, rather than opening and closing quotation marks. Please adjust.
2. Line 185 - double use of 'by a landslide' in this sentence.
3. Line 235 and 443 - add a comma after 'Thus' (consistent with 'Thus' on line 10)
4. Table 1 - 'x' has inconsistent formatting or is missing in some cases

A Systematic Exploration of Satellite Radar Coherence Methods for Rapid Landslide Detection

Katy Burrows[1], Richard J. Walters[1], David Milledge[2], and Alexander L. Densmore[3]

[1]COMET, Department of Earth Sciences, Durham University
[2]School of Engineering, Newcastle University
[2]Department of Geography, Durham University

Correspondence: Katy Burrows (katy.a.burrows@durham.ac.uk)

Abstract. Emergency responders require information on the distribution of triggered landslides within two weeks of an earthquake or storm. Useable satellite radar imagery is acquired within days of any such event worldwide. Recently, several landslide detection methods that use these data have been developed, but testing of these methods has been limited in each case to a single event and satellite sensor. Here we systematically test five methods using ALOS-2 and Sentinel-1 data across four triggering

5 earthquakes. The best performing method was dependent on the satellite sensor. For three of our four case study events, an initial ALOS-2 image was acquired within 2 weeks, and with these data, co-event coherence loss (CECL) is the best performing method. Using a single post-event Sentinel-1 image, the best-performing method was the boxcar-sibling method. We also present three new methods which incorporate a second post-event image. While the waiting time for this second post-event image is disadvantageous for emergency response, these methods perform more consistently and on average 10% better across event and sensor type than the boxcar-sibling and CECL methods. Thus, our results demonstrate that useful landslide density

10 information can be generated on the timescale of emergency response, and allow us to make recommendations on the best method based on the availability and latency of post-event radar data.

Copyright statement. TEXT

1 Introduction

15 Information on the spatial distribution of earthquake- or rainfall-triggered landslides needs to be generated as quickly as possible in order to be useful for emergency response efforts, ideally within two weeks of an event (Inter-Agency Standing Committee, 2015; Williams et al., 2018). This information is commonly generated from analysis of optical satellite imagery (e.g., Bessette-Kirton et al., 2019; Kargel et al., 2016). However, relying solely on optical satellite imagery in landslide assessment is problematic, as the mapping process can be significantly delayed by cloud cover (Robinson et al., 2019). For example,

20 following the 2015 M_w 7.8 Gorkha, Nepal earthquake, almost no cloud-free optical imagery was acquired over the region of most intense landsliding for a full week following the earthquake (Williams et al., 2018), and Robinson et al. (2019) have

demonstrated that this delay could have been much longer had the earthquake occurred during Nepal's monsoon season, with some areas unlikely to be successfully imaged at all between June and September.

When optical imagery is not available, empirical models based on factors such as the topographic slope and measurements or predictions of earthquake-induced shaking or rainfall data are used to predict the likely location and intensity of triggered landsliding (e.g. Kirschbaum and Stanley, 2018; Kritikos et al., 2015; Nowicki Jessee et al., 2018). The outputs of these models have a comparatively low spatial resolution (around 1 km^2 in the case of Nowicki Jessee et al. (2018)), but can be used to provide an overview of the most severely impacted areas. While these models can be generated within hours of an earthquake or rainfall event, they do not always perform well. For example, the USGS ground failure model of Nowicki Jessee et al. (2018) under-predicted the landsliding triggered by the M_w 6.9 2018 Lombok, Indonesia earthquake sequence (Ferrario, 2019) and multiple empirical models over-predicted the spatial extent of landslides triggered by the 2016 M_w 7.8 Kaikoura, New Zealand earthquake (Allstadt et al., 2018). The reliability of global empirical rainfall-triggered landslide susceptibility maps such as that of Kirschbaum and Stanley (2018) is dependent on the type of landslides, on the rainfall dataset used and on the intensity of the triggering rainfall event (Jia et al., 2020; Kirschbaum and Stanley, 2018). Similarly, the accuracy of empirical models for earthquakes is strongly dependent on the quality of the shaking data or model. Earthquake-induced shaking information is typically updated several times following an event, and the details of the shaking dataset that is used can have a strong effect on the modelled landslide spatial distribution and impacts (Allstadt et al., 2018; Robinson et al., 2017). For example, empirical models generated immediately following the Gorkha earthquake failed to capture the spatial pattern of triggered landslides (Robinson et al., 2017). This spatial information is critically important for use in emergency response coordination, so this limitation is a significant disadvantage when applying such empirical models.

Synthetic aperture radar (SAR) satellite imagery presents a means of generating landslide information in all weather conditions as radar is able to penetrate cloud. For landslide studies, SAR is most commonly used to measure the downslope velocity of slow-moving landslides (e.g. Aslan et al., 2020; Bonì et al., 2018; Dai et al., 2016; Handwerger et al., 2019; Hu et al., 2019; Reyes-Carmona et al., 2020; Solari et al., 2020). However, SAR can also be used to detect modifications to the Earth's surface and it has been demonstrated that radar methods can be used to automatically detect wind damage to forests (Rüetschi et al., 2019), flooding (Martinis et al., 2015) and urban damage following earthquakes, typhoons and wildfires (Fielding et al., 2005; NASA, 2018; Yun et al., 2015).

Recently, there have been several attempts to develop similar SAR-based change detection methods for rapid landslide mapping, based on SAR amplitude (Konishi and Suga, 2018, 2019; Mondini et al., 2019), coherence (Burrows et al., 2019; Olen and Bookhagen, 2018; Yun et al., 2015), or some combination of these (Aimaiti et al., 2019; Jung and Yun, 2019), or based on polarimetric SAR methods (e.g. Yamaguchi et al., 2019). However, with the exception of Mondini et al. (2019) who used a global selection of landslides, these studies are generally tested on a single landslide event and use a single radar sensor. For example, Aimaiti et al. (2019); Konishi and Suga (2019); Jung and Yun (2019); Yamaguchi et al. (2019) tested their methods using ALOS-2 imagery of the 2018 Hokkaido earthquake. If such methods are to be applied in future events, wider testing is needed. This would allow us to establish whether different methods work equally well for different events, and to determine the best method to use with data from a given SAR sensor and within a given time window.

To address this need, we carried out a systematic statistical comparison of the performance of five radar-based methods of landslide detection, using imagery acquired by the Sentinel-1 and ALOS-2 PALSAR-2 sensors spanning four case study landslide-triggering earthquakes. We chose to test on earthquakes rather than rainfall events for two reasons. First, it can be assumed that the majority of landslides occurred concurrently with or very shortly after the shaking, and this information on landslide timing simplifies the validation of the methods. Second, radar imagery is more likely to be acquired immediately after an earthquake as part of emergency tasking of satellite acquisitions, as these data are commonly used to measure earthquake-related ground deformation. We tested on four large (M_w > 6.6) events: the 2015 Gorkha, Nepal earthquake, the 2018 Hokkaido, Japan earthquake, and two earthquakes of the 2018 Lombok, Indonesia sequence (Fig. 1). All of these events triggered thousands of landslides, which have been mapped using optical satellite imagery (Ferrario, 2019; Roback et al., 2018; Zhang et al., 2019). We assessed the ability of each method and radar dataset to predict these validation data, and demonstrate the wide applicability of SAR coherence methods to landslide detection, making recommendations for which method is most suitable depending on the type of SAR data that is available and the timing of data acquisition.

2 Satellite radar coherence for change detection

A SAR system works by illuminating the Earth's surface with microwave radiation and measuring the amplitude and phase of the returned signal. In the interferometric SAR (InSAR) technique, the difference in phase between two images acquired over the same area at different times can be used to map the change in distance between the ground and the satellite. SAR amplitude is the strength of the backscattered signal and is partially dependent on the material at the ground surface: its orientation relative to the satellite, its roughness, and its dielectric properties.

When using an interferogram to map ground deformation, it is important that the signals recorded at a given location in the two SAR images are correlated, as decorrelation will result in high-frequency noise. In order to assess this and to identify noisy pixels, the coherence γ is estimated for every pixel from the similarity in the two SAR images in amplitude and phase difference, for a small ensemble of n pixels (Eq. 1, Just and Bamler, 1994):

$$\gamma = \frac{\frac{1}{n}\sum_{i=1}^{n} A_i \cdot \overline{B_i}}{\sqrt{\frac{1}{n}\left(\sum_{i=1}^{n} A_i \cdot \overline{A_i} \sum_{i=1}^{n} B_i \cdot \overline{B_i}\right)}} \tag{1}$$

A_i and B_i are complex representations of the phase and amplitude of each pixel i within the ensemble, with the complex conjugate shown by the overline. The ensemble is chosen so that the pixels used in the calculation are expected to be similar. In a "boxcar" method, it is assumed that pixels immediately adjacent to and centred on the target pixel are similar to it (e.g. Hanssen, 2001; Yun et al., 2015). In a "sibling" method an assessment is carried out for every pixel to identify pixels that are statistically similar to it. For example, the sibling method of Spaans and Hooper (2016) identifies pixels that have similar amplitude behaviour through time.

Coherence can be decomposed into 3 components, with the total coherence dependent on their product (Eq. 2, Zebker and Villasenor, 1992):

$$\gamma_{total} = \gamma_{temporal} \cdot \gamma_{spatial} \cdot \gamma_{thermal} \tag{2}$$

Here we are interested in temporal coherence $\gamma_{temporal}$, as decorrelation of this component reflects changes in the physical properties of the Earth's surface between image acquisitions. The spatial coherence, $\gamma_{spatial}$, is dependent on the geometric properties of the satellite acquiring the image and the ground surface. Decorrelation of the spatial component of coherence $\gamma_{spatial}$ is the result of small changes in satellite viewing geometry between acquisitions and can be stronger in areas of steep topography, as it is dependent on incidence angle. This decorrelation is particularly sensitive to the SAR image pair's perpendicular baseline (the distance between the locations at which the satellite acquired the two SAR images measured perpendicular to both the flight and look directions). When the perpendicular baseline of the image pair used to form an interferogram is sufficiently small, this spatial component will be small compared to any temporal decorrelation (Zebker and Villasenor, 1992). For modern satellites, this will be the case most of the time. We removed distorted pixels, which were likely to be more strongly affected by decorrelation of $\gamma_{spatial}$, from our analysis in Section 3.4. Decorrelation of the thermal coherence $\gamma_{thermal}$ was assumed to be insignificant, following Zebker and Villasenor (1992).

3 Data and methods

3.1 SAR data

In this study, radar imagery was used from two satellite systems. Sentinel-1 uses C-band radar (wavelength ~ 5.6 cm), whilst ALOS-2 PALSAR-2 uses lower frequency L-band radar (wavelength ~ 24 cm). The difference in wavelength between the two systems means that, in forested areas, L-band radar penetrates further into the canopy than C-band. The shorter wavelength of C-band radar means it is sensitive to surface modifications on a smaller spatial scale. For example, in a forest, C-band radar data may detect change in the location or orientation of leaves, while L-band is sensitive instead to changes in the location and orientation of branches. L-band InSAR is often more useful in vegetated areas, as its deeper penetration allows it to retain higher coherence in the absence of major vegetation changes (Zebker and Villasenor, 1992).

Radar systems acquire data at an oblique angle to the vertical on near-polar ascending and descending orbital tracks (referred to here as a and d), which are acquired on different dates. The satellite look direction is perpendicular to the orbit direction. The data used in this study were acquired at an angle of between 31.4° and 43.8° to vertical. This oblique acquisition angle means that on a given track, some hillsides will be more favourably oriented to the sensor than others, and so information from ascending and descending tracks can be combined to obtain more complete coverage of an event. As it is impossible to calculate a combined coherence surface using data from two tracks, and in some cases imagery will only be acquired on one track within two weeks of an event, here we considered each track separately. We will refer to tracks according to their sensor, track number, and orbit direction, for example, A018d (ALOS-2, track 18, descending orbit).

4

3.2 Case studies

We used four case study events: the 2015 M_w 7.8 Gorkha, Nepal earthquake (Fig. 1a, d); the M_w 6.6 2018 Hokkaido, Japan earthquake (Fig. 1b, e); and two M_w 6.8 and 6.9 earthquakes from the 2018 Lombok, Indonesia sequence (Fig. 1c, f). These four events have several traits in common which made them suitable for this study. First, they were all large earthquakes, triggering thousands of landslides, making them of interest from an emergency response perspective. This also meant that the earthquakes and associated landslides had previously been investigated, and inventories of triggered landslides had been compiled from optical satellite imagery, enabling direct testing of the radar methods against these independent datasets (Ferrario, 2019; Roback et al., 2018; Zhang et al., 2019). Second, while the vegetation types differ between the three regions, the presence of dense vegetation across each region meant that we could expect landslide and non-landslide pixels to have a similar first-order signal in the radar data across the events. Third, ascending and descending track Sentinel-1 data and at least one track of high resolution ALOS-2 data (acquired in stripmap mode at a resolution of 3 - 10 m) were available for all case study areas. Finally, the type of landslides triggered by our four case study earthquakes were typical of landslides triggered by earthquakes (Keefer, 1984). The majority of ground failures in the four earthquakes were slides. In Nepal, ground failures were primarily a mixture of slides and falls, with the exception of a large debris avalanche in the Langtang Valley. For all four earthquakes, failure surfaces were at shallow depths in most cases with a small number of exceptions (Collins and Jibson, 2015; Ferrario, 2019; Yamagishi and Yamazaki, 2018).

The M_w 7.8 2015 Gorkha, Nepal earthquake (Fig. 1a) occurred on 25 April 2015 and triggered around 25,000 landslides over an area hundreds of kilometres wide, as mapped by Roback et al. (2018). Sentinel-1 imagery was acquired on tracks S019d and S085a. ALOS-2 data on track A157a were divided into subtracks with acquisitions on different dates, shown as separate polygons on Fig. 1a, which will be referred to as East (E), Central (C) and West (W).

The M_w 6.6 2018 Hokkaido, Japan earthquake (Fig. 1b) occurred on 5 September 2018. Two inventories have been published for this event: one containing 7,837 landslides (Wang et al., 2019) and one containing 5,265 (Zhang et al., 2019). Neither provided information on the mapping extent so we assumed that this could be approximated by the convex hull of the data locations. As the inventory of Zhang et al. (2019) has the largest convex hull, we used this inventory for validation of the radar methods. Both descending and ascending ALOS-2 data were available for this event, with a higher spatial and temporal frequency than for the other events. The earthquake occurred the day after Typhoon Jebi passed over Hokkaido, and so this case study was also an opportunity to test landslide detection methods following a rainfall event, with the advantage that because the typhoon and earthquake occurred one day apart, and aerial imagery of the triggered landslides was acquired immediately afterwards, we know more precisely when the landslides occurred (Yamagishi and Yamazaki, 2018). This is important if SAR methods are to be used in the future for mapping storm-triggered landslides because factors such as wind damage and the water content of the soil are known to affect SAR coherence and amplitude (Rüetschi et al., 2019; Scott et al., 2017).

The 2018 Lombok earthquake sequence comprised 4 earthquakes with $M_w > 6$: M_w 6.4 on 28 July; M_w 6.8 on 5 August; and M_w 6.3 and 6.9 on 19 August. Ferrario (2019) generated two landslide inventories for this sequence. Although cloud-free imagery was not available across the whole affected area following the earthquake on 28 July, no landslides were visible in the

areas that could be mapped. Ferrario (2019) thus presented their first inventory of 4,823 landslides triggered following the 5 August earthquake, and a second inventory of 9,319 landslides which had been triggered by the end of the sequence. We refer to the 5 August inventory as Lombok-1 (Fig. 1c) and the 19 August inventory as Lombok-2 (Fig. 1c, inset).

There are several key ways in which the events differed. First, the triggered landslides had very different spatial patterns, with the Gorkha earthquake triggering landslides across an area spanning hundreds of km, while the Lombok and Hokkaido earthquakes triggered landslides across only a few km. It can be seen in Fig. 1 that much denser landsliding was triggered by the Hokkaido earthquake than by the other events. Second, the sizes and shapes of the landslides were very different between events: the Lombok earthquakes triggered a large number of small landslides with a median landslide area of 460 m^2 for Lombok-1 and 580 m^2 for Lombok-2, compared to equivalent median areas of 4,350 m^2 for Hokkaido and 1,070 m^2 for Nepal (measured from the inventories of Ferrario, 2019; Roback et al., 2018; Zhang et al., 2019). Third, the events occurred under different weather conditions: the Hokkaido earthquake followed months of heavy rain and occurred one day after Typhoon Jebi, while the Gorkha and Lombok earthquakes occurred during the dry season. Finally, the topographic relief in the three case study areas varied significantly. In Nepal the majority of landslides occurred on slopes of over 40° (Roback et al., 2018). For the Hokkaido event, the proportion of slopes > 40° was very low and so the majority of landslides occurred on much shallower slopes (Wang et al., 2019), with Lombok lying between these two extremes. This is highly relevant to the application of SAR coherence methods to landslide detection. Steep slopes can lead to distortion of the radar image, and coherence is also dependent on the geometry of the hillslope and radar sensor. Therefore, we might expect that, as hillslopes in Hokkaido and Lombok are shallower than in Nepal, landslide detection using SAR may be more successful in these areas. These differences between the four events made them ideal for testing the wider applicability of SAR-coherence-based landslide detection methods in vegetated areas. If a method is to be widely applied in the future, we need to be confident that its performance is consistent across differing events and settings.

3.3 Landslide detection methods

We tested two existing methods: the co-event coherence loss (CECL) method of Yun et al. (2015) and the boxcar-sibling method of Burrows et al. (2019). Each of these existing methods uses a single post-event SAR image. We also present three new methods that incorporated a second post-event image: the post-event coherence increase (PECI); the sum of the coherence increase and decrease (ΔC_sum); and the maximum of coherence increase or decrease (ΔC_max). A "boxcar" coherence estimate is used for CECL, PECI, ΔC_sum and ΔC_max.

3.3.1 Co-event coherence loss (CECL)

The coherence loss between a pre-event interferogram and a co-event interferogram can be used to detect physical changes to the ground surface associated with an earthquake, such as surface rupture, building damage and landslides (Fielding et al., 2005; Washaya et al., 2018; Yun et al., 2015). This method has been applied by the NASA Advanced Rapid Imaging and Analysis (ARIA) project for use in urban damage mapping and identifies "damaged" pixels as those where coherence has decreased in the co-event map relative to the pre-event map. First the pre-event coherence map is adjusted so that it has the same

coherence frequency distribution as the co-event coherence map using exact histogram matching (Coltuc et al., 2006). This process accounts for different levels of bulk temporal decorrelation in the pre-event and co-event interferograms. It assumes only a small fraction of the pixels are affected by landslides so that the landslide signal is not removed from the co-event interferogram. The pre-event surface is then subtracted from the co-event surface and pixels whose coherence has decreased are flagged as damaged (Yun et al., 2015). Although this method was developed for detecting damage to buildings, Yun et al. (2015) tested it on the 2015 Gorkha earthquake and noted that landslides in the Langtang Valley corresponded spatially to areas of coherence decrease in the ARIA surface. Coherence decrease between pre-event and co-event interferograms has since been used as an input in the landslide detection methods of Aimaiti et al. (2019) and Jung and Yun (2019) applied to the 2018 Hokkaido earthquake.

3.3.2 The boxcar-sibling method (Bx-S)

Rather than relying on the coherence change through time, the Bx-S method of Burrows et al. (2019) uses the difference between two alternative co-event spatial methods of coherence calculation as a landslide classification surface. The four other methods tested here all use a traditional "boxcar" coherence, in which the coherence of a pixel is estimated from the similarity in phase change of the pixels immediately adjacent to it. When a sibling-based method is used to estimate coherence, the coherence of a pixel becomes dependent on "siblings" that are not immediately adjacent to it but that are expected to behave similarly. In the method of Spaans and Hooper (2016) used here, an ensemble of siblings is selected for every pixel that have similar amplitude behaviour in a time series of pre-event imagery. For a landslide pixel, this means that its coherence is calculated from a more dispersed ensemble of pixels than with a traditional boxcar coherence estimate and so proportionally less of the ensemble will also lie within the landslide. A sibling-based coherence surface is therefore relatively insensitive to landslides, and landslides can be identified as those whose co-event boxcar coherence is lower than their co-event sibling coherence (see Burrows et al. (2019) for detail).

3.3.3 Post-event coherence increase (PECI)

The coherence decrease caused in a co-event interferogram by a landslide is a temporary effect, assuming that the landslide stops moving following the earthquake. Therefore in a post-event interferogram (calculated from two post-event images) coherence should be higher for landslide pixels than in a co-event interferogram. As landslides expose bare rock or soil, which is likely to have higher coherence than vegetated areas, this co-event to post-event increase may actually be larger than the pre-event to co-event decrease used in the CECL method to measure landslides, making the signal easier to detect. This is particularly the case when using C-band SAR, which experiences more decorrelation than L-band in vegetated regions. Applying the same histogram-matching step as in the CECL method, we propose this co-event to post-event coherence increase as a new potential landslide detection method.

3.3.4 Sum of coherence increase and decrease (ΔC_sum)

215 As landslides are expected to exhibit both a decrease in co-event coherence and an increase in post-event coherence, we summed the absolute magnitudes of these changes to form a new landslide classification surface. As in the CECL method, the pre-event and post-event coherence maps were histogram-matched to the co-event coherence map, removing bulk changes in coherence between the three maps. This method is equivalent to CECL + PECI.

3.3.5 Maximum of coherence increase or decrease (ΔC_max)

220 For every pixel, we took whichever was largest of the pre-event to co-event coherence loss (CECL) and the co-event to post-event increase (PECI). This method is similar to the ΔC_sum method but uses whichever has the strongest signal. The relative signal strength of the CECL or PECI methods will vary by location, for example according to the vegetation type. This method takes whichever of the two methods has a stronger signal for any given location.

3.4 Data processing

225 SAR data were processed using GAMMA, with the LiCSAR processing software used for Sentinel-1 (Li et al., 2016). The data were multilooked by a factor of 5 in range and 1 in azimuth (Sentinel-1) or by 5 in both range and azimuth (ALOS-2) to improve the signal to noise ratio. See Table 1 for information on the data resolution and pixel size at various stages of the processing. For geometric coregistration, we used the 1-arcsecond Shuttle Radar Topography Mission (SRTM) digital elevation model (Farr et al., 2007).

230 The boxcar coherence estimate used in all methods was calculated using a 3×3 pixel moving window (Table 1). The sibling-based coherence estimate used for the Bx-S method was calculated using the RapidSAR algorithm of Spaans and Hooper (2016). Siblings were calculated based on all pre-event images shown in Fig. 1 (a minimum of 6 images). For every pixel, between 15 and 50 siblings were identified within an 81×81 pixel window based on their amplitude and amplitude variability (window sizes in Table 1). Unfortunately, insufficient pre-event ALOS-2 data were available to carry out this calculation in the 235 case of the 2015 Gorkha, earthquake and so the Bx-S method could not be tested for this case.

In most cases, we used the co-event, pre-event, and post-event coherences with the shortest temporal baseline; however for the inventory provided by Ferrario (2019) for the landslides triggered by the Lombok earthquake on 19 August, we used a co-event image that spans this earthquake and the 5 August earthquake, as the landslide inventory contains landslides triggered by both events. After calculating each landslide classification surface from the methods described in Section 3.3, we used 240 the GAMMA software to convert the surfaces to a geographic coordinate system. This process, involving reprojection and interpolation, results in uniform pixel size (20 m $\times$ 22 m). We then normalised the values of each surface by the theoretical maximum and minimum value that could be obtained from each method, resulting in a set of classification surfaces with values between 0 and 1, with 1 most likely to be a landslide.

Before statistical testing, we removed distorted pixels. The oblique angle at which SAR imagery is acquired meant that 245 some pixels were distorted by topography and were badly imaged by the SAR system, experiencing shadow, foreshortening,

or layover (Franceschetti et al., 1994) and decorrelation of $\gamma_{spatial}$ (Section 3.1). Following Burrows et al. (2019), we masked these according to the area in geographic coordinates that contributes to the pixel in radar geometry, removing pixels with an area of 0 and those where this area was over six times larger than their multilooked pixel spacing in radar coordinates (see Table 1).

250   We carried out statistical testing of the results at two resolutions: first, at the initial resolution of the processed radar data in geographical coordinates (20 m × 22 m), with the vector landslide inventories of Ferrario (2019); Roback et al. (2018); Zhang et al. (2019) rasterized at this resolution; and second, at an aggregated resolution of 200 m × 220 m, calculated by amalgamating 10 × 10 grids of the 20 m × 22 m pixels. Aggregated classifier pixels were given the mean value of the unmasked pixels in the 10 × 10 grid. If over 95% of an aggregate pixel was made up of masked pixels, the aggregate pixel was masked. This

255 high threshold of 95% was chosen to minimise the loss of spatial coverage due to the masks. Varying the threshold between 95% and 5% had little difference in terms of the number of pixels used in the analysis in Hokkaido and Lombok (<5%), but in Nepal, where more pixels were masked due to distortion on steep slopes, decreasing the threshold to 5% resulted in a loss of coverage of around 40% on S085a. Altering this threshold made very little difference to the results presented in Section 4.1.

    In this study, we did not attempt to map SAR classification surface values directly to landslide areal density values, as this

260 has not been attempted in previous studies (e.g. Aimaiti et al., 2019; Burrows et al., 2019; Jung and Yun, 2019; Yun et al., 2015) and may not be possible due to differences in viewing geometry, land cover and, particularly with the ALOS-2 data used here, differences in temporal baseline between events. Thus, a binary ground truth was preferable, which we generated by assigning aggregate pixels as "landslide" if they were composed of over 25% landslide by area according to the rasterized landslide inventories we used for verification. We explore the effect that this choice of a 25% threshold to define a landslide

265 pixel has on our results in the supplementary material, but varying this threshold between 1% and 50% was found to have little effect on the relative performance of the different classifiers.

    The aggregation process was done for several reasons. First, the boxcar coherence estimation has the effect of blurring neighbouring pixels, so that the minimum size of an object we could resolve was dependent on the size of the 3 × 3 boxcar (see Table 1). In all cases this was larger than the 20 m × 22 m pixel spacing in geographic coordinates. Second, it has already

270 been demonstrated that SAR coherence-based methods perform better at lower resolutions (Burrows et al., 2019). Third, when comparing landslide inventories, for example those of Zhang et al. (2019) and Wang et al. (2019) for the Hokkaido earthquake, there is often some variation in the exact shape and location of individual polygons. We expect this effect to be decreased when inventories are downsampled. We therefore chose a resolution that was larger than the boxcar window but that resulted in enough mapped landslide and non-landslide aggregate pixels for analysis. This is similar to the resolution of other landslide

275 products generated for emergency response (e.g. Bessette-Kirton et al., 2019; Nowicki Jessee et al., 2018)

3.5   ROC analysis

We used the receiver operating characteristic (ROC) area under the curve (AUC) to evaluate and compare the classification ability of each surface. For each classification surface, a threshold was set for which no pixels were classified as landslides and then incrementally decreased until it reached a value where all pixels were classified as landslides. At each incremental

280 threshold, the false positive rate (the ratio of false positives to real non-landslide pixels) was plotted against the true positive rate (the ratio of true positives to real landslide pixels) to form an ROC curve. The area under this curve is equal to the probability that if a landslide pixel and a non-landslide pixel were randomly selected from the dataset, the classifier would rank them correctly (Hanley and McNeil, 1982). For a randomly generated surface with no classification ability, the AUC = 0.5. For a perfect classifier, the AUC = 1.0.

285 On all SAR tracks, there are many more non-landslide than landslide pixels. It has been suggested that for such imbalanced data, precision-recall curves can better represent classification ability than ROC AUC (Saito and Rehmsmeier, 2015). Here, we chose to use ROC analysis since precision-recall curves do not allow comparison between datasets with different proportions of landslide and non-landslide pixels and therefore between different earthquakes and SAR tracks. However, when considering the relative performance of classifiers for each track independently, we found the same conclusions could be drawn from

290 precision-recall curves as from ROC curves. A recreation of Figure 2a using precision-recall rather than ROC AUC values can be found in Supplementary Information.

4   Results

4.1   Results at 200 m × 220 m resolution

Figure 2 shows the ROC AUC values for each classification method described in Section 3.3 and each track of radar data shown

295 in Fig. 1. The cells are coloured so that the best performing classification surfaces are shown in green and the weakest in red. A classifier that performs well for all events and both sensors should therefore appear as a green row. The eastern and western tracks of S157a and data for Lombok-2 on track A129a have been omitted, as the waiting times for the first post-event image were very long (77, 63 and 139 days respectively). This long time window resulted in widespread co-event coherence loss that adversely affected classifier performance. Such a long waiting time would make it extremely unlikely that these data could be

300 used in emergency response and the poor performance is unlikely to be representative of classifier behaviour when using more timely imagery.

The methods are grouped by the number of post-event images that are required, and the waiting time in days for this imagery for each event and SAR track is also given. As the primary scientific use for post-event ALOS-2 imagery is to form a co-event interferogram, only one post-event image was acquired immediately following the Lombok-1 and Nepal earthquakes, with the

305 waiting time for the second post-event image being considerably longer. We still include these results as the waiting time for L-band radar data is likely to decrease in the future with the planned NISAR satellite constellation, which is expected to launch in 2022, acquiring data globally with a 12 day repeat time (Sharma, 2019).

If we consider only data acquired within the two week emergency response window (Inter-Agency Standing Committee, 2015), the highest AUC in Hokkaido was initially 0.58 on day 0 using the Bx-S method on S046d, rising to 0.89 after 1 day

310 with data from track A018d and the CECL method. In Nepal, the first radar data were acquired on day 4 on track S019d, from which the highest AUC of 0.74 was obtained using the Bx-S method. On day 7, the first ALOS-2 image was acquired on A157a, and more accurate information could have been generated using CECL (AUC 0.81), although this ALOS-2 scene

covers a smaller area than the Sentinel-1 data (Fig. 1). In Lombok, the 6-day Sentinel-1 acquisition repeat time meant that a large volume of data was available within two weeks of these events. The best data and method for Lombok-1 evolved as follows: CECL, S032d (day 0, AUC 0.56); CECL or Bx-S, S156a (day 3, AUC 0.64); PECI, S032d (day 6, AUC 0.69); ΔC_max, S156a (day 9, AUC 0.72); CECL, A129a (day 13, AUC 0.88). The corresponding classification surfaces for this evolution are plotted in Fig. 3. For Lombok-2, the first Sentinel-1 image was acquired on day 1, track S156a, and had an AUC of 0.67 using the Bx-S method, which could be slightly improved upon using the second post-event image that was acquired on day 7, using PECI (AUC 0.68). No ALOS-2 data was available within 2 weeks of this earthquake. Therefore in most cases, initially only Sentinel-1 data are available and the best option is the Bx-S method. This can then be improved upon when the first ALOS-2 image becomes available using CECL. For Lombok-1 and Lombok-2, where two post-event Sentinel-1 images became available before the first ALOS-2 image, incorporating these data also improved on the accuracy of the result.

With a single L-band post-event image from ALOS-2, CECL was the best-performing landslide classification surface in all cases. An improvement was seen when an additional post-event image was acquired, and methods requiring this image were used, for the case of Lombok and Nepal, but not for Hokkaido. When a single C-band post-event image was used, the Bx-S method outperformed CECL for Hokkaido, Gorkha and Lombok-2 and had a similar performance for Lombok-1. The addition of a second post-event image and adoption of methods that used this image showed an improvement in Hokkaido and Lombok-1, but not in Nepal. However, when considering Fig. 2a as a whole, methods that used a second post-event image were both better performing and more consistent across event and sensor type. When looking across all three events, the best option in terms of AUC was to use the ΔC_sum method with L-band imagery. When grouped by event and radar look direction and ranked according to AUC, Δ C_sum using ALOS-2 was ranked 3rd out of 10 classification surfaces for descending imagery over Hokkaido, 1/9 for ascending imagery over Nepal and 2/10 for ascending imagery over Lombok-1. A comparison cannot be made for ascending track data in Hokkaido, as A116a and S085a had different look directions (westwards and eastwards respectively).

Figure 4 shows the landslide classification surfaces calculated using the ΔC_sum method and ALOS-2 data of each event, alongside the aggregated validation landslide data. In Figure 4b, d, f, cells made up of <1% landslide by area are masked. In order to recreate this in the radar surface, it was necessary to threshold and plot only pixels that were most likely to be landslides based on their classifier value. Here we applied a threshold such that the number of pixels plotted in panels (a, c, e) is the same as the number in panels (b, d, f), respectively. These threshold values were similar, but not identical, and we expect that more case study sites would be required to determine a more general threshold for application in future events. However, in each case, the spatial pattern of landsliding in (b, d, f) was recreated using the radar surfaces (a, c, e), which would allow the worst affected areas to be identified for emergency response even without strict definition of this threshold.

4.2   Variation across event and sensor type

Our results show significant variation across event and sensor type. L-band radar outperformed C-band in the majority of cases, but for some methods, we observed additional effects causing differences in performance.

For methods relying on a co-event vs. pre-event coherence decrease (i.e. CECL, ΔC_sum, ΔC_max), landslides associated with Lombok-2 were more difficult to predict than Lombok-1 (Fig. 2). This is likely to be due to the co-event acquisition time window, which was 6 days for the first inventory, but had to be increased to 18 days on track S156a and 24 days on S032d in order to span both earthquakes. This will have caused temporal decorrelation of vegetated non-landslide pixels, resulting in a smaller difference between landslide and non-landslide co-event coherence and making it more difficult for the classifier to distinguish between these. The same effect was seen in Nepal, where 12 day interferograms were used for S019d, but 24 day pre-event and co-event interferograms were used for S085a.

Generally, the Bx-S method was more consistent across sensor type than the CECL method. The CECL method performed better than the Bx-S method with ALOS-2 data, but worse with Sentinel-1. There are several possible reasons for this. First, the longer wavelength of L-band SAR meant that it was able to maintain a higher coherence in the pre-event interferogram, so that the coherence difference for a landslide pixel in a pre-event and co-event interferogram was larger, resulting in better performance from CECL. Second, in the Bx-S method, siblings are identified using pre-event imagery. As Sentinel-1 imagery is acquired every 12 days, more images were available for this calculation and were acquired over a shorter time period. This allows less time for pixels to be altered by changes to the ground surface, meaning that, for non-landslide pixels, a pixel and its siblings are likely to be more similar. In this way, the siblings selected by RapidSAR for Sentinel-1 imagery may have been of a higher quality than those for ALOS-2, giving a more reliable coherence estimate.

4.3 Results at 20 m × 22 m resolution

While the main aim of this study was not to map individual landslides, we also show results at a 20 m x 22 m scale (the resolution of the classification surfaces in geographic coordinates), with pixels that are sufficiently small to resolve individual landslides. Our results show that SAR methods were less successful at this resolution than when downsampled, with AUC on average 16 % lower using Sentinel-1 data and 11 % lower using ALOS-2 data. Using Sentinel-1 data, AUC values were low, ranging from 0.49 (CECL, Lombok-2) to 0.61 (PECI, Hokkaido). From this we conclude that mapping individual landslides with Sentinel-1 data was not possible using the methods tested here, which supports similar preliminary findings by Burrows et al. (2019).

Classification surfaces using ALOS-2 are more promising, with AUC up to 0.80 (ΔC_max, Lombok-1). As in Section 4.1, ΔC_sum performed best, having the highest AUC on tracks A116a and A157a, and an AUC of only 0.01 less than the best performing classification surface on tracks A018d and A129a. Figure 6 shows small regions of this surface overlain with landslide polygons for each event. While the landslides generally coincide spatially with high classifier values, it is clear that it would not have been possible in most cases to map the landslide polygons using the radar data. Some large landslides (Fig. 6b, d) contain pixels with high classifier values. However, in (d), there are also areas of mapped landslide that do not have high classifier values, and in (a) and (c) the landslides are too small. Therefore we conclude that, while radar coherence methods show promise in individual landslide detection, it was not possible using the methods tested here, and there is likely to be a minimum detectable landslide size.

5   Discussion

380   We have demonstrated that SAR data is widely applicable to landslide detection in vegetated areas within the timeframe of the emergency response effort. For example, within the two week limit suggested by Inter-Agency Standing Committee (2015) and Williams et al. (2018), it would have been possible to generate triggered landslide density information using the CECL method with ALOS-2 data with an ROC AUC of 0.81 in Nepal, 0.89 in Hokkaido and 0.88 in Lombok-1 (Fig. 2). In this section, we first consider the applications of the SAR methods to future events, and then turn to potential sources of errors. Finally we

385   discuss future work that would allow wider application of SAR methods to landslide detection.

5.1   Application

5.1.1   Landslide density estimation

We found that radar methods were better suited to the production of landslide density maps than to the identification of individual landslides. The aggregated resolution of 200 m x 220 m that we used here was not high enough to identify individual

390   landslides. However it was higher than the resolution of most empirical landslide susceptibility models designed for rapid response (Allstadt et al., 2018; Nowicki Jessee et al., 2018) and landslide maps generated from optical satellite imagery for use in aid efforts by Bessette-Kirton et al. (2019) following Hurricane Maria, Puerto Rico in 2017, and by Williams et al. (2018) following the 2015 Gorkha earthquake. SAR data seem best suited to producing products at this spatial scale. For example, Bessette-Kirton et al. (2019) produced a grid of 2 x 2 km pixels and assigned them as "high landslide density" (>25

395   landslides), "low landslide density" (1-25 landslides) or "no landslides". This was published a month after the hurricane using optical imagery acquired between 6 and 18 days after the event. As radar data are available within a few days of an earthquake, equivalent products could easily be generated from radar within this timescale.

Landslide density maps can be combined with data on population density in order to estimate exposure, as in Nowicki Jessee et al. (2018). They can be used alongside maps of roads to identify transport routes that are likely to be blocked, and maps of

400   river networks in order to identify areas where a landslide may have temporarily dammed a river, posing a risk of flash flooding when the dam collapses (Robinson et al., 2018). Information of this kind may guide aerial assessments of landslides such as that by Collins and Jibson (2015), which was carried out 32-36 days after the 2015 Gorkha earthquake to identify possible landslide dams.

5.1.2   Recommendations on data and methods

405   We showed in Section 4 that, in the majority of cases, landslide classification surfaces generated using L-band data outperformed those generated from C-band data. Currently, the main source of L-band data is the ALOS-2 satellite system. This system has a 14 day repeat time and one of its objectives is disaster response. Therefore, in most cases, a post-seismic image will be made available within two weeks of an earthquake, with the main purpose of producing a co-event interferogram that can be used to measure ground displacement and infer fault geometry. However, as this interferogram requires only one post-

410 event image, the waiting time for a second image may be several months, as was the case following the Nepal and Lombok earthquakes. Therefore in an emergency response situation, it is likely that only one L-band image will be available within the required timeline, and so CECL would be the best available classification surface. L-band data are also less likely to be acquired following a rainfall event, even if this event triggers many landslides, as ground displacement maps and therefore interferograms are not needed. In the future, however, the availability of L-band SAR is likely to increase with the planned

415 NISAR and ROSE-L satellite constellations. NISAR, a joint NASA-ISRO mission planned to launch in 2022, will acquire L-band data continuously with a 12-day repeat time over all landmasses globally (Sharma, 2019), while the ESA ROSE-L satellite, whose launch date is planned in 2026, will have 6-day global repeat coverage, and 3-day in Europe (Pierdicca et al., 2019).

While Sentinel-1 data yield generally lower AUC values (Figs. 2a, 5), the system acquires data continuously over global

420 tectonic belts with a 12 day repeat time, with all data made freely available. Image acquisitions on ascending and descending tracks are offset in time so that the waiting time for a single post-event track is always less than 6-12 days. Over Europe, Sentinel-1 data are acquired at twice this frequency, meaning that the first post-event image should be available within 3 days. As the data are regularly acquired, they will be available for rainfall events as well as earthquakes. In some cases, we may have two post-event Sentinel-1 images before the first ALOS-2 image becomes available. It is more difficult to make a definitive

425 recommendation for the best method to use in this situation, as performance varies significantly between events. With one post-event image, the Bx-S method is the best performing classifier with Sentinel-1, and in Nepal remains the best classifier even after additional post-event Sentinel-1 images have been acquired. With two post-event images, the best performing classifier in 2 out of 8 cases is either ΔC_sum or ΔC_max. However in cases where the presence of vegetation means that pre-event coherence and non-landslide co-event coherence are very low, the best option is to use PECI (e.g., Lombok-1 S032d, Lombok-

430 2, Hokkaido). Inspection of the pre-event coherence should therefore be carried out first to select which method to use.

5.2 Sources of incorrect classifications

5.2.1 Building damage

As the CECL method of Yun et al. (2015) was originally designed to detect urban damage, it is unsurprising that damaged buildings cause false positives in SAR coherence-based methods for landslide detection. Large scale signals, such as the

435 town of Atsuma in Hokkaido, could be removed using a land cover map, but masking landslides in built-up areas is clearly disadvantageous as this is where they would do the most damage. Additionally, buildings located outside of large towns may not be included in such maps, but damage to these would still result in false positives. These false positives will occur using all the methods in this study. They are particularly likely to have negatively affected the AUC values for A116a in Hokkaido, as this track had a larger proportion of non-landslide pixels lying in built up areas.

14

5.2.2 Wind damage

In Hokkaido, we observed some large false-positive patches in CECL, PECI, ΔC_sum and ΔC_max classification surfaces using ALOS-2 data and PECI, ΔC_sum and ΔC_max surfaces using Sentinel-1. These correspond spatially to areas of ever-green needleleaf forest in the Japanese Aerospace Exploration Agency (JAXA) high resolution land use and land cover map of Japan (JAXA, 2018). Areas with this forest cover, which we have mapped using Sentinel-2 imagery, are outlined in white on Fig. 6b.

We hypothesise that this type of vegetation may have been damaged by wind during Typhoon Jebi, which passed over the area the day before the earthquake, causing a coherence decrease that is similar to that caused by landsliding. This type of forest has a comparatively high coherence in both pre-event and post-event L-band interferograms, giving it a strong signal in CECL, PECI, ΔC_sum and ΔC_max. This effect may also explain why the low coherence area extends beyond the area affected by landslides (an effect visible in our data, and also observed by Fujiwara et al. (2019)). Unfortunately, none of the SAR data in this study were acquired between the typhoon and the earthquake, which would have allowed separation of the two events. However, these forest patches have no signal in amplitude-based methods, which has been observed in the study of Fransson et al. (2010) of wind-damaged forest at this resolution. This lack of an amplitude signal may allow these false positives to be removed through a combination of amplitude and coherence-based classification surfaces. Additionally, we observed that these patches of forest are generally much larger than the landslides and so could perhaps be removed if object- rather than pixel-based classification were used.

5.2.3 Snow

Snow is a potential source of error when using SAR in landslide detection. There is a strong difference in SAR backscattering properties between wet snow, dry snow and no snow (Koskinen et al., 1997). As such, decreased coherence can be caused by snow melt, drift or fall between image acquisitions. Examination of Sentinel-2 imagery shows there was no snow cover at the time of the Hokkaido earthquake, and snow is not likely in Lombok due to the high temperatures. We do not have exact snow cover data for Nepal at the time of the earthquake, but we expect that in April approximately 1/5 of the country's total area would have been covered in snow (ICIMOD, 2013). Therefore at high altitudes, it is reasonable to assume that some false positives may have been caused by snow, particularly on track A129a due to the long time interval between image acquisitions.

5.2.4 Rivers

The Bx-S method is effectively a spatial filter on co-event coherence, removing signals that cover a large area. Therefore small, low coherence objects will be identified as landslides. This includes rivers, as demonstrated by Spaans and Hooper (2016). As the low coherence caused by a river is not temporary, this should have less of an effect on the other coherence methods tested here. However, changes in position or water level are likely to result in false positives in all coherence-based classifiers. A variety of methods could be used to identify and remove rivers from our analysis (including using a pre-event Bx-S surface). However, since areas where landslides and rivers intersect are particularly hazardous due to the potential for landslide dams and

475   5.2.5   Landslide density

When using the Bx-S method in Hokkaido, we found low AUC values with both L-band and C-band SAR. Visual comparison of the boxcar and sibling coherence surfaces shows that the area affected by landslides had a low coherence in both, rather than the expected case where the boxcar coherence is lower than the sibling coherence. We suggest this may be due to the intensity of the damage caused by this event, both in terms of the landslides, which had a much higher density than for Nepal
480   or Lombok, and possibly also in terms of vegetation damage caused by Typhoon Jebi. Since the Bx-S method relies on siblings of a landslide pixel lying outside the landslide, it may not work well in the case where many landslides are close together. This is similar to the case found by Burrows et al. (2019), where the Bx-S method was unable to identify large landslides if the sibling search window was not sufficiently large.

5.3   Future work

485   We have demonstrated that the classification ability of a method can vary significantly based on the data used, the nature of the triggering event, and the resolution of the analysis. For example, the Bx-S method performed significantly better in the case of Nepal than Hokkaido, and CECL was up to 32% more successful when using ALOS-2 compared to Sentinel-1 data with the same event and look direction. Thus, it is clear that any future SAR-based classification method must be widely tested on multiple events, and it cannot be assumed that a method performing well with one SAR dataset will be equally successful using
490   data from a different sensor.

While we tested ascending and descending tracks separately, a more complete picture would be obtained through combining these two tracks, as hillslopes that are unfavourably oriented to the satellite in one track are likely to be more favourably oriented to the other track. A complete map would be easier to interpret for emergency responders than two maps, each of which is missing landslides on unfavourably oriented slopes. This is particularly important since landslides are most likely on
495   steep slopes where these orientation effects are likely to be most severe.

Although we selected earthquakes to use as case studies, SAR methods could be equally useful in the case of rainfall-triggered landslides, where cloud cover may also cause delays to mapping using optical satellite imagery. We have shown that SAR methods performed well in Hokkaido, where the co-event coherence maps spanned both an earthquake and a typhoon. This demonstrates that the alterations in SAR coherence and amplitude that can result from changes in soil moisture content
500   are not prohibitive to the application of SAR methods to rainfall triggered landslides, although some false positives may have been caused by wind damage.

The Lombok case studies demonstrate the importance of knowing the timing of the landslides when applying current methods. When moving from Lombok-1, in which all landslides were assumed to have been triggered by a single earthquake, to Lombok-2, which contained landslides triggered by a series of earthquakes over a two-week period, it was necessary to in-

505 crease the timespan of the Sentinel-1 co-event interferograms from 6 days to 24 days on track S032d and 18 days on track S156a, which significantly reduced the AUC of the classification methods. Therefore, if we wish to apply radar methods more widely to landslide triggering events that are dispersed through time, such as long rainfall events (e.g., those associated with the monsoon) or earthquake sequences, more work will be required on the characterisation of the signals of landslides in radar imagery through time.

510   Finally, here we assessed classifier performance using ROC analysis, which does not require a threshold to be applied to the classification surface. However, if SAR methods are to be applied to future events for emergency response, it will be necessary to set a threshold between 'landslide' and 'non-landslide'. In this study, the time between image acquisitions varied significantly between events, making it unlikely that a threshold could be selected that would work well for both Sentinel-1 and ALOS-2 across all events. However, this may be possible in the future when more events have been studied and SAR data with

515 more regular acquisitions are available. Further work is therefore needed to establish such thresholds, which will be determined according to the requirements of emergency responders and their relative tolerance for false positives and false negatives.

6   Conclusions

We have demonstrated that it is possible to generate landslide density information from SAR coherence within a typical two-week emergency response time-frame and at a useful spatial resolution. The best performing method is dependent on the

520 wavelength of the available imagery and the number of post-event images available during the emergency response. The CECL method is the best performing method when only one post-event L-band image is available (ROC AUC = 0.7-0.89), and this method could be applied within two weeks for three out of our four case study events. However, Sentinel-1 data were available earlier than ALOS-2 for all of the events studied here, with the first image available within 4 days of each earthquake. Using the first post-event Sentinel-1 image acquired, the best method is the Bx-S method with AUC between 0.58-0.74. Methods that

525 use a second post-event image improve overall accuracy by an average of 10 % and are more consistently reliable across event and sensor type. These approaches could be valuable for landslide mapping when latency is less important, but the additional waiting time for the second post-event image is a disadvantage in an emergency response situation.

535

References

540   Aimaiti, Y., Liu, W., Yamazaki, F., and Maruyama, Y.: Earthquake-Induced Landslide Mapping for the 2018 Hokkaido Eastern Iburi Earthquake Using PALSAR-2 Data, Remote Sensing, 11, 2351, 2019.

Allstadt, K. E., Jibson, R. W., Thompson, E. M., Massey, C. I., Wald, D. J., Godt, J. W., and Rengers, F. K.: Improving Near-Real-Time Coseismic Landslide Models: Lessons Learned from the 2016 Kaikōura, New Zealand, EarthquakeImproving Near-Real-Time Coseismic Landslide Models, Bulletin of the Seismological Society of America, 108, 1649–1664, 2018.

545   Aslan, G., Foumelis, M., Raucoules, D., De Michele, M., Bernardie, S., and Cakir, Z.: Landslide Mapping and Monitoring Using Persistent Scatterer Interferometry (PSI) Technique in the French Alps, Remote Sensing, 12, 1305, 2020.

Bessette-Kirton, E. K., Cerovski-Darriau, C., Schulz, W. H., Coe, J. A., Kean, J. W., Godt, J. W., Thomas, M. A., and Hughes, K. S.: Landslides Triggered by Hurricane Maria: Assessment of an Extreme Event in Puerto Rico, GSA Today, 29, 2019.

Bonì, R., Bordoni, M., Colombo, A., Lanteri, L., and Meisina, C.: Landslide state of activity maps by combining multi-temporal A-DInSAR

550   (LAMBDA), Remote sensing of environment, 217, 172–190, 2018.

Burrows, K., Walters, R. J., Milledge, D., Spaans, K., and Densmore, A. L.: A New Method for Large-Scale Landslide Classification from Satellite Radar, Remote Sensing, 11, 237, 2019.

Collins, B. D. and Jibson, R. W.: Assessment of existing and potential landslide hazards resulting from the April 25, 2015 Gorkha, Nepal earthquake sequence, Tech. rep., US Geological Survey, 2015.

555   Coltuc, D., Bolon, P., and Chassery, J.-M.: Exact histogram specification, IEEE Transactions on Image Processing, 15, 1143–1152, 2006.

Dai, K., Li, Z., Tomás, R., Liu, G., Yu, B., Wang, X., Cheng, H., Chen, J., and Stockamp, J.: Monitoring activity at the Daguangbao megalandslide (China) using Sentinel-1 TOPS time series interferometry, Remote Sensing of Environment, 186, 501–513, 2016.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., et al.: The shuttle radar topography mission, Reviews of geophysics, 45, 2007.

560   Ferrario, M.: Landslides triggered by multiple earthquakes: insights from the 2018 Lombok (Indonesia) events, Natural Hazards, pp. 1–18, 2019.

Fielding, E. J., Talebian, M., Rosen, P. A., Nazari, H., Jackson, J. A., Ghorashi, M., and Walker, R.: Surface ruptures and building damage of the 2003 Bam, Iran, earthquake mapped by satellite synthetic aperture radar interferometric correlation, Journal of Geophysical Research: Solid Earth, 110, 2005.

565   Franceschetti, G., Marino, R., Migliaccio, M., and Riccio, D.: SAR simulation of three-dimensional scenes, in: SAR Data Processing for Remote Sensing, vol. 2316, pp. 192–202, International Society for Optics and Photonics, 1994.

Fransson, J. E., Pantze, A., Eriksson, L. E., Soja, M. J., and Santoro, M.: Mapping of wind-thrown forests using satellite SAR images, in: 2010 IEEE International Geoscience and Remote Sensing Symposium, pp. 1242–1245, IEEE, 2010.

Fujiwara, S., Nakano, T., Morishita, Y., Kobayashi, T., Yarai, H., Une, H., and Hayashi, K.: Detection and interpretation of local surface

570   deformation from the 2018 Hokkaido Eastern Iburi Earthquake using ALOS-2 SAR data, Earth, Planets and Space, 71, 64, 2019.

Handwerger, A. L., Fielding, E. J., Huang, M.-H., Bennett, G. L., Liang, C., and Schulz, W. H.: Widespread initiation, reactivation, and acceleration of landslides in the northern California Coast Ranges due to extreme rainfall, Journal of Geophysical Research: Earth Surface, 2019.

Hanley, J. A. and McNeil, B. J.: The meaning and use of the area under a receiver operating characteristic (ROC) curve., Radiology, 143,

575   29–36, 1982.

Hanssen, R. F.: Radar interferometry: data interpretation and error analysis, vol. 2, Springer Science & Business Media, 2001.

Hu, X., Bürgmann, R., Lu, Z., Handwerger, A. L., Wang, T., and Miao, R.: Mobility, thickness, and hydraulic diffusivity of the slow-moving Monroe landslide in California revealed by L-band satellite radar interferometry, Journal of Geophysical Research: Solid Earth, 124, 7504–7518, 2019.

580 ICIMOD: Snow cover statistic - Nepal, http://www.icimod.org/?q=10216, 2013.

Inter-Agency Standing Committee: Multi-Sector Initial Rapid Asessment Guidance, https://www.humanitarianresponse.info/en/programme-cycle/space/document/multi-sector-initial-rapid-assessment-guidance-revision-july-2015, accessed on 16 October 2018, 2015.

JAXA: High-Resolution Land Use and Land Cover 30m resolution map of Japan [2014-2016], https://www.eorc.jaxa.jp/ALOS/en/lulc/lulc_

585   index.html, version 18.03, 2018.

Jia, G., Tang, Q., and Xu, X.: Evaluating the performances of satellite-based rainfall data for global rainfall-induced landslide warnings, Landslides, 17, 283–299, 2020.

Jung, J. and Yun, S.-H.: A Hybrid Damage Detection Approach Based on Multi-Temporal Coherence and Amplitude Analysis for Disaster Response, in: IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, pp. 9330–9333, IEEE, 2019.

590 Just, D. and Bamler, R.: Phase statistics of interferograms with applications to synthetic aperture radar, Applied optics, 33, 4361–4368, 1994.

Kargel, J. S., Leonard, G. J., Shugar, D. H., Haritashya, U. K., Bevington, A., Fielding, E., Fujita, K., Geertsema, M., Miles, E., Steiner, J., et al.: Geomorphic and geologic controls of geohazards induced by Nepal's 2015 Gorkha earthquake, Science, 351, aac8353, 2016.

Keefer, D. K.: Landslides caused by earthquakes, Geological Society of America Bulletin, 95, 406–421, 1984.

Kirschbaum, D. and Stanley, T.: Satellite-based assessment of rainfall-triggered landslide hazard for situational awareness, Earth's future, 6,

595   505–523, 2018.

Konishi, T. and Suga, Y.: Landslide detection using COSMO-SkyMed images: a case study of a landslide event on Kii Peninsula, Japan, European Journal of Remote Sensing, 51, 205–221, 2018.

Konishi, T. and Suga, Y.: Landslide detection with ALOS-2/PALSAR-2 data using convolutional neural networks: a case study of 2018 Hokkaido Eastern Iburi earthquake, in: Proc. of SPIE Vol, vol. 11154, pp. 111 540H–1, 2019.

600 Koskinen, J. T., Pulliainen, J. T., and Hallikainen, M. T.: The use of ERS-1 SAR data in snow melt monitoring, IEEE Transactions on geoscience and remote sensing, 35, 601–610, 1997.

Kritikos, T., Robinson, T. R., and Davies, T. R.: Regional coseismic landslide hazard assessment without historical landslide inventories: A new approach, Journal of Geophysical Research: Earth Surface, 120, 711–729, 2015.

Li, Z., Wright, T., Hooper, A., Crippa, P., Gonzalez, P., Walters, R., Elliott, J., Ebmeier, S., Hatton, E., and Parsons, B.: Towards InSAR

605   everywhere, all the time, with Sentinel-1., International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, 41, 2016.

Martinis, S., Kersten, J., and Twele, A.: A fully automated TerraSAR-X based flood service, ISPRS Journal of Photogrammetry and Remote Sensing, 104, 203–212, 2015.

Mondini, A. C., Santangelo, M., Rocchetti, M., Rossetto, E., Manconi, A., and Monserrat, O.: Sentinel-1 SAR Amplitude Imagery for Rapid

610   Landslide Detection, Remote Sensing, 11, 760, 2019.

NASA: Advanced Rapid Imaging and Analysis (ARIA) Project for Natural Hazards, https://aria.jpl.nasa.gov/, accessed on 27 June 2018, 2018.

Nowicki Jessee, M., Hamburger, M., Allstadt, K., Wald, D. J., Robeson, S., Tanyas, H., Hearne, M., and Thompson, E.: A Global Empirical Model for Near-Real-Time Assessment of Seismically Induced Landslides, Journal of Geophysical Research: Earth Surface, 123, 1835–1859, 2018.

Olen, S. and Bookhagen, B.: Mapping damage-affected areas after natural hazard events using Sentinel-1 coherence time series, Remote Sensing, 10, 1272, 2018.

Pierdicca, N., Davidson, M., Chini, M., Dierking, W., Djavidnia, S., Haarpaintner, J., Hajduch, G., Laurin, G. V., Lavalle, M., López-Martínez, C., et al.: The Copernicus L-band SAR mission ROSE-L (Radar Observing System for Europe)(Conference Presentation), in: Active and Passive Microwave Remote Sensing for Environmental Monitoring III, vol. 11154, p. 111540E, International Society for Optics and Photonics, 2019.

Reyes-Carmona, C., Barra, A., Galve, J. P., Monserrat, O., Pérez-Peña, J. V., Mateos, R. M., Notti, D., Ruano, P., Millares, A., López-Vinielles, J., et al.: Sentinel-1 DInSAR for Monitoring Active Landslides in Critical Infrastructures: The Case of the Rules Reservoir (Southern Spain), Remote Sensing, 12, 809, 2020.

Roback, K., Clark, M. K., West, A. J., Zekkos, D., Li, G., Gallen, S. F., Chamlagain, D., and Godt, J. W.: The size, distribution, and mobility of landslides caused by the 2015 Mw7. 8 Gorkha earthquake, Nepal, Geomorphology, 301, 121–138, 2018.

Robinson, T. R., Rosser, N. J., Densmore, A. L., Williams, J. G., Kincey, M. E., Benjamin, J., and Bell, H. J.: Rapid post-earthquake modelling of coseismic landsliding intensity and distribution for emergency response decision support., Natural hazards and earth system sciences., 17, 1521–1540, 2017.

Robinson, T. R., Rosser, N. J., Davies, T. R., Wilson, T. M., and Orchiston, C.: Near-Real-Time Modeling of Landslide Impacts to Inform Rapid Response: An Example from the 2016 Kaikōura, New Zealand, EarthquakeNear-Real-Time Modeling of Landslide Impacts to Inform Rapid Response, Bulletin of the Seismological Society of America, 108, 1665–1682, 2018.

Robinson, T. R., Rosser, N., and Walters, R. J.: The Spatial and Temporal Influence of Cloud Cover on Satellite-Based Emergency Mapping of Earthquake Disasters, Scientific Reports, 9, 1–9, 2019.

Rüetschi, M., Small, D., and Waser, L. T.: Rapid detection of windthrows using Sentinel-1 C-band SAR data, Remote Sensing, 11, 115, 2019.

Saito, T. and Rehmsmeier, M.: The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets, PloS one, 10, e0118 432, 2015.

Scott, C., Lohman, R., and Jordan, T.: InSAR constraints on soil moisture evolution after the March 2015 extreme precipitation event in Chile, Scientific reports, 7, 4903, 2017.

Sharma, P.: Updates in Commissioning Timeline for NASA-ISRO Synthetic Aperture Radar (NISAR), in: 2019 IEEE Aerospace Conference, pp. 1–12, IEEE, 2019.

Solari, L., Del Soldato, M., Raspini, F., Barra, A., Bianchini, S., Confuorto, P., Casagli, N., and Crosetto, M.: Review of Satellite Interferometry for Landslide Detection in Italy, Remote Sensing, 12, 1351, 2020.

Spaans, K. and Hooper, A.: InSAR processing for volcano monitoring and other near-real time applications, Journal of Geophysical Research: Solid Earth, 121, 2947–2960, 2016.

Wang, F., Fan, X., Yunus, A. P., Subramanian, S. S., Alonso-Rodriguez, A., Dai, L., Xu, Q., and Huang, R.: Coseismic landslides triggered by the 2018 Hokkaido, Japan (M w 6.6), earthquake: spatial distribution, controlling factors, and possible failure mechanism, Landslides, 16, 1551–1566, 2019.

Washaya, P., Balz, T., and Mohamadi, B.: Coherence change-detection with sentinel-1 for natural and anthropogenic disaster monitoring in urban areas, Remote Sensing, 10, 1026, 2018.

Wessel, P. and Smith, W. H.: New, improved version of Generic Mapping Tools released, Eos, Transactions American Geophysical Union, 79, 579–579, 1998.

Williams, J. G., Rosser, N. J., Kincey, M. E., Benjamin, J., Oven, K. J., Densmore, A. L., Milledge, D. G., Robinson, T. R., Jordan, C. A., and Dijkstra, T. A.: Satellite-based emergency mapping using optical imagery: experience and reflections from the 2015 Nepal earthquakes., Natural hazards and earth system sciences., 18, 185–205, 2018.

Yamagishi, H. and Yamazaki, F.: Landslides by the 2018 Hokkaido Iburi-Tobu Earthquake on September 6, Landslides, 15, 2521–2524, 2018.

Yamaguchi, Y., Umemura, M., Kanai, D., Miyazaki, K., and Yamada, H.: ALOS-2 polarimetric SAR observation of Hokkaido-Iburi-Tobu earthquake 2018, IEICE Communications Express, 8, 26–31, 2019.

Yun, S.-H., Hudnut, K., Owen, S., Webb, F., Simons, M., Sacco, P., Gurrola, E., Manipon, G., Liang, C., Fielding, E., et al.: Rapid Damage Mapping for the 2015 M w 7.8 Gorkha Earthquake Using Synthetic Aperture Radar Data from COSMO–SkyMed and ALOS-2 Satellites, Seismological Research Letters, 86, 1549–1556, 2015.

Zebker, H. A. and Villasenor, J.: Decorrelation in interferometric radar echoes, IEEE Transactions on geoscience and remote sensing, 30, 950–959, 1992.

Zhang, S., Li, R., Wang, F., and Iio, A.: Characteristics of landslides triggered by the 2018 Hokkaido Eastern Iburi earthquake, Northern Japan, Landslides, pp. 1–18, 2019.
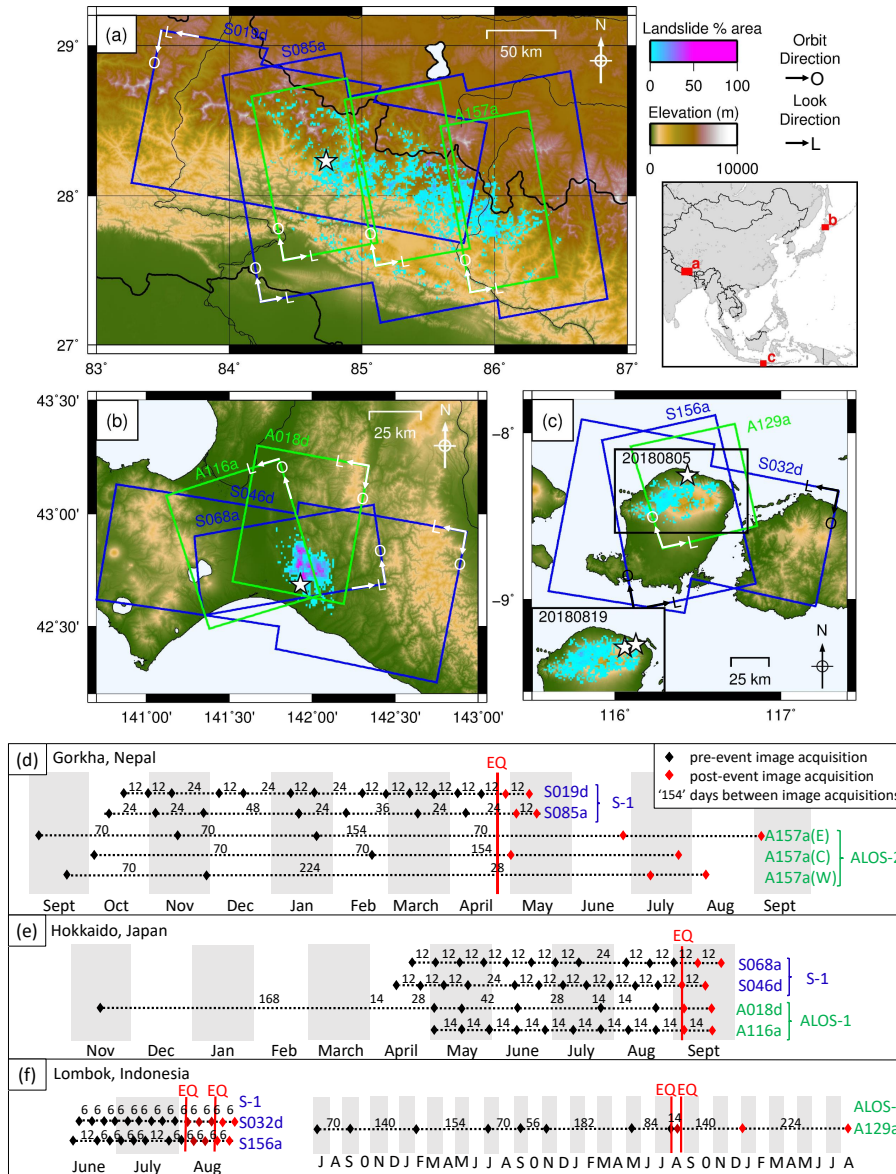
Figure 1. (a, b, c) Landslides triggered by the 2015 Gorkha, Nepal earthquake, the 2018 Hokkaido, Japan earthquake, and the 2018 Lombok, Indonesia earthquakes, respectively, plotted as the areal density of landsliding based on 1 km^2 cells (Roback et al., 2018; Zhang et al., 2019; Ferrario, 2019). ALOS-2 SAR acquisitions are shown in green and Sentinel-1 in blue. In (c), a second earthquake, referred to as "Lombok-2" is inset. (d, e, f) Acquisition dates and track numbers of SAR imagery used in this study. Earthquakes are shown as red vertical lines, black symbols show pre-event image acquisition dates, and red symbols show post-event dates.
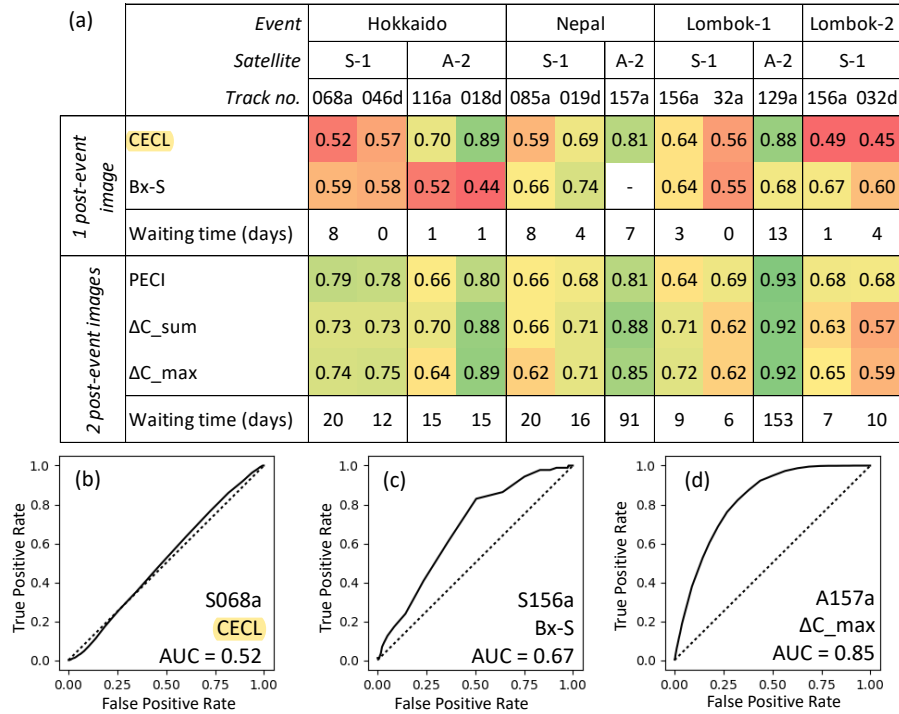
23

(a)

	Event	Hokkaido				Nepal			Lombok-1			Lombok-2	
	Satellite	S-1		A-2		S-1		A-2	S-1		A-2	S-1	
	Track no.	068a	046d	116a	018d	085a	019d	157a	156a	32a	129a	156a	032d
1 post-event image	CECL	0.52	0.57	0.70	0.89	0.59	0.69	0.81	0.64	0.56	0.88	0.49	0.45
	Bx-S	0.59	0.58	0.52	0.44	0.66	0.74	-	0.64	0.55	0.68	0.67	0.60
	Waiting time (days)	8	0	1	1	8	4	7	3	0	13	1	4
2 post-event images	PECI	0.79	0.78	0.66	0.80	0.66	0.68	0.81	0.64	0.69	0.93	0.68	0.68
	ΔC_sum	0.73	0.73	0.70	0.88	0.66	0.71	0.88	0.71	0.62	0.92	0.63	0.57
	ΔC_max	0.74	0.75	0.64	0.89	0.62	0.71	0.85	0.72	0.62	0.92	0.65	0.59
	Waiting time (days)	20	12	15	15	20	16	91	9	6	153	7	10

Figure 2. (a) AUC values for each classifier described in Section 2 at a resolution of 200 m × 220 m. For Hokkaido, we use the inventory of Zhang et al. (2019). Insufficient pre-event data were available on A157a to calculate the Bx-S classification surface. Colours range from red (worst performing AUC < 0.55) to green (best performing, AUC >0.80). (b, c, d) Examples of ROC curves for (b) the CECL method on track S156a, Hokkaido; (c) the Bx-S method on track S156a, Lombok-2; (d) ΔC_max on track A157a, Nepal.
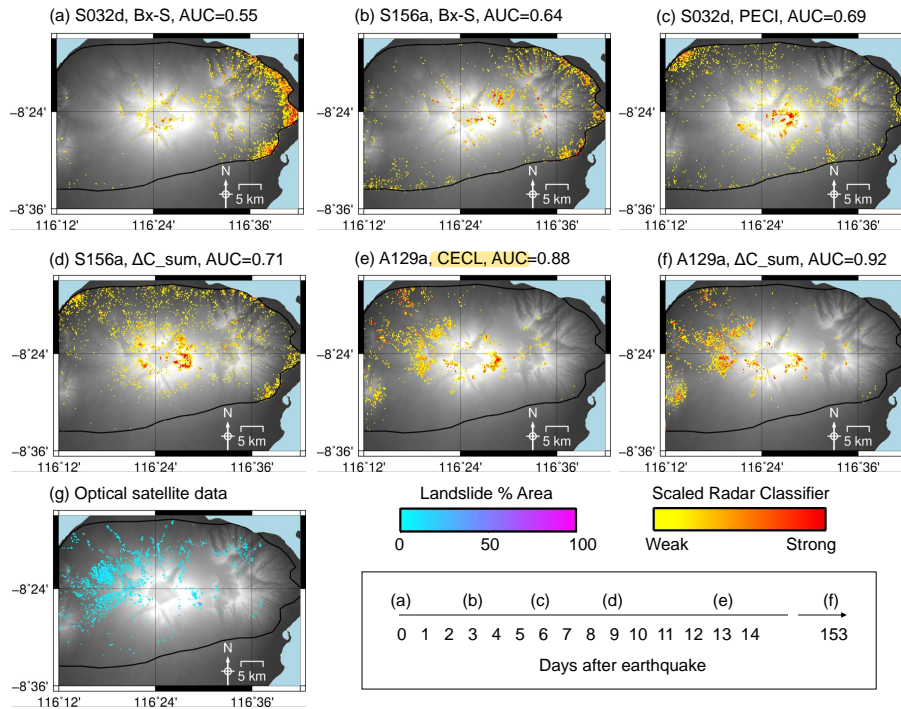
Figure 3. (a-f) Time series of classification surfaces in the order that SAR data were acquired on tracks S032d, S156a and A129a following the 5 August 2018 Lombok earthquake, using the methods we recommend in Section 4.1.(g) Observed Landslide areal density for 200 m x 220 m pixels (calculated from Ferrario, 2019).
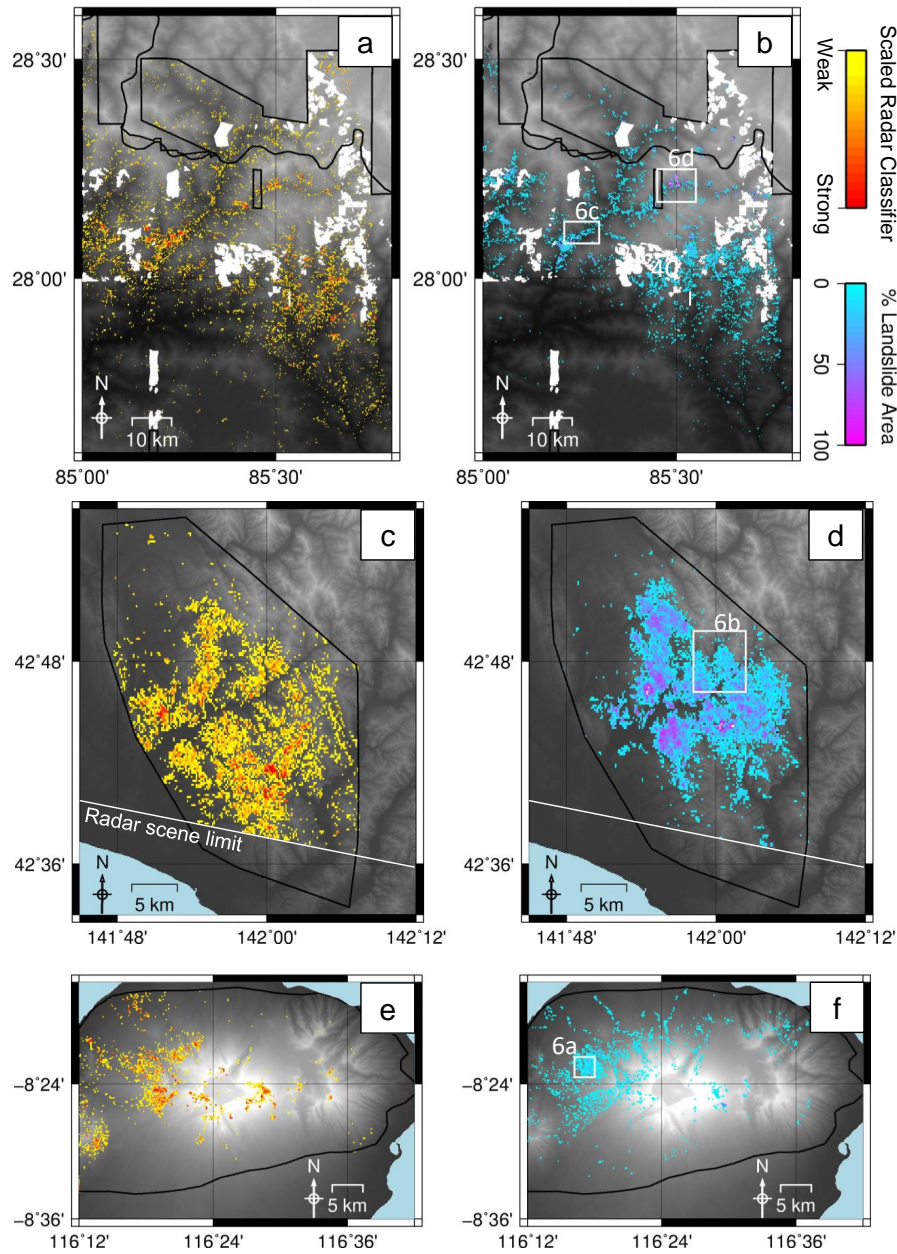
Figure 4. SAR-based landslide classification surfaces for the Gorkha (a), Hokkaido (c) and Lombok-1 (e) earthquakes calculated with ALOS-2 radar data using the ΔC_sum method at a 200 m × 220 m resolution. (b, d, f) Observed landslide density calculated as the percentage of each 200 m × 220 m cell covered by landsliding for each event according to the inventories of Ferrario (2019), Roback et al. (2018), Zhang et al. (2019) respectively. Cells where landslide density was 0 were masked. For the ΔC_sum surface, a threshold value of the classifier was selected such that the number of cells plotted in panels (a, c, e) for each event was the same as the number plotted in panels (b, d, f)

		Hokkaido				Nepal			Lombok			Lombok 2		Event
		S-1		A-2		S-1		A-2	S-1		A-2	S-1		Satellite
		68A	46D	116A	18D	85A	19D	157A	156A	32D	129A	156A	32D	Track number
1 post-event image	CECL	0.51	0.51	0.62	0.72	0.52	0.58	0.68	0.53	0.50	0.76	0.50	0.49	
	Bx-S	0.52	0.52	0.52	0.51	0.56	0.59	-	0.54	0.52	0.69	0.53	0.52	
	Waiting time	8	0	1	1	8	4	7	3	0	13	1	4	
2 post-event images	PECI	0.60	0.61	0.60	0.65	0.54	0.55	0.67	0.53	0.54	0.76	0.53	0.55	
	ΔC_sum	0.57	0.58	0.63	0.71	0.54	0.57	0.72	0.54	0.52	0.79	0.52	0.52	
	ΔC_max	0.58	0.59	0.60	0.72	0.53	0.57	0.70	0.54	0.53	0.80	0.52	0.54	
	Waiting time	20	12	15	15	20	16	91	9	6	153	7	10	

Figure 5. AUC values for each classifier described in Section 2 at a resolution of 20 m × 22 m. For Hokkaido, we use the inventory of Zhang et al. (2019). Insufficent pre-event data were available on A157a to calculate the Bx-S classification surface. Colours range from red (worst performing AUC < 0.55) to green (best performing, AUC >0.80).
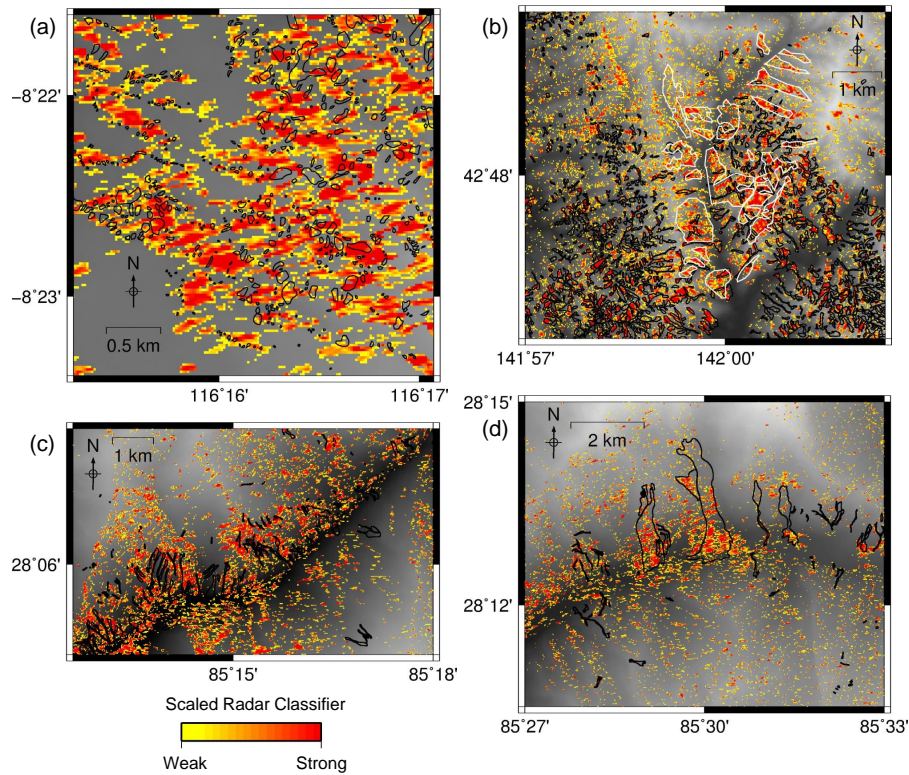
Figure 6. ΔC_sum classification surface calculated using ALOS-2 data displayed at 20 m x 22 m resolution with mapped landslide polygons in black. The 10% of pixels most likely to be landslide pixels are coloured from yellow (least likely) to red (most likely) (a) Lombok: track A129a for the 5 August 2018 event (Ferrario, 2019). (b) Hokkaido: track A018d, landslide polygons from Zhang et al. (2019). White polygons show the locations of forested areas mapped using Sentinel-2 data. (c,d) Nepal: track A157a(C), landslide polygons from Roback et al. (2018).

Table 1. The resolution and pixel spacing of the data at different stages throughout the processing for Sentinel-1, ALOS-2 on tracks A018d and A116a (Hokkaido), track A157a (Nepal) and track A129a (Lombok). Resolutions are given in Range × Azimuth coordinate system

	Resolution (m)	Radar pixel spacing (m)	Multilooked radar pixel spacing (m)	Boxcar window size (m)	Sibling search window size (m)
All Sentinel-1	20 × 22	2.3 × 14.0	11.6 × 14.0	34.8 × 42.0	940 × 1134
ALOS-2 018d	3 × 3	1.4 × 2.1	7.2 × 10.6	21.6 × 31.8	583 × 859
ALOS-2 116a	3 × 3	1.4 × 1.9	7.2 × 9.3	21.6 × 27.9	583 × 753
ALOS-2 157a	10 × 10	4.3 × 3.8	21.5 × 19	64.5 × 57.0	1742 × 1539
ALOS-2 129a	10 × 10	4.3 × 3.3	21.5 × 16.3	64.5 × 48.9	1742 × 1320