Nat. Hazards Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/nhess-2020-168-RC2, 2020

Natural Hazards and Earth System Sciences
Discussions

Interactive comment on "A Systematic Exploration of Satellite Radar Coherence Methods for Rapid Landslide Detection" *by* Katy Burrows et al.

Anonymous Referee #2

Received and published: 3 August 2020

General comments:

The authors present a quantitative comparison of multiple different methods of mapping landslides using the coherence of synthetic aperture radar images. The quantitative comparison between different methods, with tests in different regions and examination of the local features that can affect results, is particularly valuable given the ongoing work on using SAR for disaster response. The work is well written and figures are of a good standard.

There are some places where I think the assumptions underlying certain choices, both for the landslide mapping methods and the quantification of the results could be better explained; I have provide further comments on these below. There are also numerous places where I think small tweaks to the wording and slightly more explanations would

be helpful to the reader.

Technical comments:

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.

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)âĂŤotherwise 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.

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)?

'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 2003 Bam, 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.

Line by Line comments:

Line 73 - ïż£'The signal-to-noise ratio of each pixel in an InSAR image is described by its coherence'. This statement feels incomplete to meâĂŤin 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).

Line 77 - I think some mention of the assumptions of the box-car 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.

Line 87 - perhaps include a definition of perpendicular baseline, and mention that orbital controls on modern satellites have rendered this a much smaller issue?

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

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)

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

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

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

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

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?

Line 234 - ïż£'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)?

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.

Line 277 - 'ïż£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'?

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

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

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.

Line 469 - ïż£'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. Figure 6 - how is the threshold for these plots chosen? Could you clarify in the caption?

Technical Corrections:

For words in quotation marks apostrophes are being used throughout, rather than opening and closing quotation marks. Please adjust.

Line 185 - double use of 'by a landslide' in this sentence.

Line 235 and 443 - add a comma after 'Thus' (consistent with 'Thus' on line 10)

Table 1 - 'x' has inconsistent formatting or is missing in some cases

---

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2020-168, 2020.