Review of Bernard et al. 2020, "*Beyond 2D inventories : synoptic 3D landslide volume calculation from repeat LiDAR data*" for ESURF by David Milledge

Summary

This is a really exciting paper that takes a careful and robust approach to surface differencing in order to identify topographic changes associated with co- and post-seismic processes. In this respect the paper provides a novel and useful contribution both in developing a methodology for identifying the changes and in documenting the changes themselves. It is also (if correct) a potentially very important paper! It argues that the current form of landslide size distributions are a result of observational error for all but the right tail. Given the importance of these findings it is essential that the paper is very clear about its landslide detection process and the implications of each processing step for landslide detection (in terms of location, size and shape). Fundamentally I remain unconvinced at the end of the paper that the findings on landslide scaling and size distribution reflect those of real landslides in the study area.

The approach to surface differencing is rigorous, the writeup is clear (subject to a few minor comments that could be easily resolved). It is the step from change detection to landslide detection that I find problematic. Landslides are detected as connected patches of surface difference above a threshold. The validity of this detection method is not tested against any independent observations. I believe that the detection process is: 1) sensitive to topographic errors; and 2) prone to amalgamate some landslides and break up others. The resultant landslide inventory is then used to make claims about the size distribution and scaling properties of landslides, both of which are extremely sensitive to the errors detailed above. The conclusions of the paper are then built around these later size distribution and scaling findings, which I don't think the data are currently capable of supporting.

To me, the two main missing elements of the manuscript are: 1) a very clear definition of the range of processes and landforms that the authors would include within the category of 'landslide' (and therefore what set of processes their inventory represents); and 2) a detailed comparison of the landslide inventory generated here against independent observations within the study area, these are likely to include optical imagery but would ideally also include field investigation. These two elements are essential if the authors are to support their conclusions on landslide scaling and landslide size-frequency distributions.

The estimates of total landslide volume and of net mass loss from the study area are important contributions on their own. However, I think that more work is needed to post-process that volume estimate to account for counter-factual observations (such as deposition areas with no upslope erosion). It would also be very interesting to investigate the work associated with the event in terms of elevation reduction for the landslide mass.

Finally, examination of the properties of the landslides that have been identified would be valuable both because the dataset should enable interesting insight and because disagreement between findings from this dataset and more traditional inventories may highlight uncertainties or errors not only in the traditional approaches but also in this new approach.

Major Comments

MC1) The paper needs to more clearly define: 1) landslides (i.e. what the inventory includes) and 2) what your inventory can and cannot be used for.

Early on you introduce the idea that there are different landslide processes (L43, "process specific") but you don't follow this logic through into your results. Instead, your analysis may contain an implicit definition of landslides as all processes responsible for surface change that cannot be attributed to fluvial processes (L217-9). You certainly need to make this definition of landslides explicit in your introduction.

The introduction needs a much clearer explanation for what you expect the inventory to be useful for. If it is for understanding landslide mechanics then it is essential that you make an effort to distinguish individual landslides on a mechanistic basis (see MC3 on amalgamation). If that is not an expected application of the inventory you should say so, otherwise there is a real risk it will be misapplied.

Appropriate uses of the inventory (e.g. volume estimation or landslide mechanics) depend not only on its purpose but also on entry criteria into, and distinctions within, the inventory. Non-fluvial surface change that might result from earthquake shaking includes: tree-throw, ravel, rockfalls and slides, slow earthflows, rapid soil slides and debris flows. The processes responsible for these surface changes differ from one another to varying degrees. If there is no distinction between them this precludes the inventory's use in analysis of landslide mechanics and therefore prediction. It allows comment on correlation e.g. of volumes and areas of change, but makes it extremely difficult to make any inference about causation. It also opens the work to the criticism that the bulk statistics mask important differences in behaviour between processes. For example: the differing size distributions for rockfalls (where others have reported no detectable rollover) and landslides (where there usually is).

MC2) Elevation errors need to be better quantified and more thoroughly discussed

The manuscript needs a more thorough treatment of errors in the topographic data dealing with both: 1) the properties of the elevation errors that you have identified (e.g. spatial pattern, wavelength, covariation with landscape properties); and 2) the possible sources of error.

RE error properties: The amplitude and correlation length of elevation uncertainty from different sources and how they interact to generate a 2D elevation error field with a particular amplitude and wavelength is really important for this particular application, where landslides are identified by thresholding then segmenting differences. The error appears to have a fairly long wavelength in many areas (tens to hundreds of metres). It also appears to have some aspect dependence. The spatial correlation of these errors is important because you assume uniform isotropic registration errors.

RE error sources: It isn't clear to me what you mean by imperfect alignment, nor ICP related errors (L204-5). Identifying errors on Fig 3 and hypothesising the sources from which they derive are useful but need to be discussed in the manuscript as well. You recognise the presence of “internal flight line height mismatch” and indicate that it results in “large scale low amplitude topographic change” (L418-22). They should be introduced earlier in the article with a more complete explanation of what they are and how you found them.

Error consequences: It would be useful to say something about the implications of the topographic errors. I can see two implications: **First**, incorrectly assuming uniform isotropic errors will result in a confidence interval for identifying significant geomorphic change that is too strict on some slopes (e.g. some aspects) and not sufficiently so on others. This in turn will lead to false negative change detection on some slopes and false positives on others. **Second**, change detection false positives will result in false identification of landslide objects or false representation of their geometry. False negatives are equally problematic since they could result in not only changes to landslide geometry but also cluster breakup (biasing the size distribution). These problems are illustrated in Fig 3 where a number of error patches would be identified as landslides by the detection algorithm if these areas were not assumed to be stable. If such false positives exist here it is they likely also exist elsewhere. It is essential that you quantify their impact on your findings.

Suggested additional analysis: You could apply your landslide detection method to only the pre-defined ‘stable areas’ and generate landslide geometries. These geometries and their scaling relationships might indicate the impact of error on your findings (particularly: number and area density, pdfs, and scaling relationships of artefact landslides). If the results are similar to your findings for ‘non-stable areas’ it would be very difficult to argue that the data support your claims with any certainty (the same results could have been generated purely from topographic errors in the absence of any landslides).

MC3) You recognise that segmentation results in amalgamation but don’t quantify its extent or impact

Severe amalgamation can result from automated segmentation of a thresholded classifier. In the landslide maps that you show here (e.g. Fig 2), amalgamation appears a severe problem for the largest landslides. The argument that it can be solved by tuning Dm (as you suggest on L426-7) is unconvincing since two separate landslides can be within millimetres of one another but have different failure mechanisms. You later say that you “cannot resolve the amalgamation” (L511) and that it “is still a potential issue” (L558). I would argue that it is not potentially but certainly an issue. Your figures show that amalgamation is present (perhaps pervasive) in your inventory but its extent or impact is not quantified. The total volume is insensitive to amalgamation but your landslide pdfs and scaling relationships are not.

Amalgamation makes it difficult to use the inventory to understand landslide mechanics and therefore susceptibility or hazard. If the inventory is to be useful in understanding landslide mechanics then it is essential that you make an effort to distinguish individual landslides on a mechanistic basis. In the extreme case, an individual landslide where subsets of the failed material moved in opposite directions (e.g. Fig 2) is clearly problematic for the mechanics of that failure.

Amalgamation introduces bias towards large landslides in size distributions. Power law exponents and even the appropriateness of any power law can be sensitive to landslide amalgamation. Observations such as that on L569-70 that “largest and deepest landslides deviate significantly from this trend” become questionable since these are the landslides most likely to be the result of amalgamation.

Suggested additional analysis: If you are to retain your findings on landslide scaling and size distributions, then you must examine the potential impact of amalgamation on these findings. There are methods available that quantify amalgamation in landslide mapping. You should investigate these. You should preferably also address this amalgamation, either by removing affected landslides or ideally, by decomposing amalgamated landslides into their individual failures.

MC4) Topographic errors propagate through segmentation to introduce a bias towards small landslides. Existing experiments to quantify this bias are insufficient.

You say on L559-60 that “our approach has the benefits of more systematically capturing small landslides than traditional approaches”. However, this is one of my main problems with the paper given potential propagation of topographic errors through thresholding and segmentation. You show (in SI) that: 1) in the absence of any real topographic change your detection algorithm generates artefact landslides of 1-20 m^2 purely due to spatially uncorrelated topographic noise; 2) this noise generates many more small than large landslides; 3) in this experiment artefacts >20 m^2 were extremely rare. However, this does not demonstrate that predicted landslide size is insensitive to longer wavelength topographic errors (known to be present in the data); nor even to short wavelength noise in the presence of longer wavelength surface differences (e.g. real landslides).

First, even without any real topographic change (i.e. no real landslides), the size distribution of erroneous landslide-like clusters will depend on the spatial correlation length of the difference errors, which in turn depends on the correlation lengths of the errors in the surfaces being differenced. Longer error wavelengths will enable the generation of larger error clusters. Your figures show that topographic errors are clearly not uncorrelated and you recognise this yourselves (L418-22); nor do the errors appear to have a single characteristic wavelength. This is a hard problem but one that you must deal with if you are to convince the reader that the landslide inventory you have generated is not hopelessly biased by these landslide-like artefacts. **Second,** the problem is not only that clusters of erroneous negative surface difference due to roughness (or other errors) can create artefacts that appear to be landslides, but also that clusters of erroneous positive surface difference are collocated with real topographic change (e.g. due to a landslide); these can interfere negatively with real changes reducing the surface difference below the threshold for detection and breaking a single landslide into multiple patches.

Oversampling of small landslides is important because it undermines your most surprising and high impact claim: that rollover reported in previous inventories is due to under detection (L573-4). I am not currently convinced by this claim because you do not exclude the possibility that the lack of a rollover is solely due to detection errors. You need to quantify these detection biases before you can make these claims.

Suggested additional analysis: A “more advanced segmentation” (L431-3) may be out of scope for this paper. However, an indication of the impact of the simple segmentation on object based classification skill is an essential requirement of this paper if it is to retain the current approach to identifying discrete landslides. Two possible avenues could be followed to provide such an indication. **First,** your analysis of topographic changes on the stable surfaces (Fig 3) would allow you to perform the same analysis that you have performed in the SI but using the pre- and post-EQ surfaces for the stable zones identified in Fig 3. This would enable you to identify the size distribution of artefacts that can be generated from topographic errors with a spatial correlation length closer to that for the unstable parts of the study area. This though still does not account for the possibility that the landslide erosion signal itself is altered by the noise (e.g. by disconnecting clusters). **Second,** you could collect a landslide check dataset using independent observations. This might take the form of an entirely independent inventory but should certainly also involve cross-checking to confirm the existence and characteristics (e.g. area, shape, depth) of your predictions.

MC5) Findings that differ from previous work

There are a number of unusual findings that are worthy of comment because they are some of your most interesting and potentially important findings. It is essential though that each is carefully examined and that the critique that it might have arisen due to methodological errors is dealt with head on.

First, it is not unusual to identify more sources than deposits due to amalgamation of landslide deposits. However, it is unusual for deposit areas to be smaller than source areas (L281-2), and for deposit depths to be thicker than source depths (L265-70 and L460). These result should be compared with results from previous studies.

Second, the estimate that Massey et al. (2020) "potentially missed around 169,000" landslides (L491-2), i.e. 92% of the landslides is surprising because of its magnitude. However, to be convincing you need to demonstrate that the landslides that you have detected genuinely are landslides.

Third, you identify areas of deposition where there is no upslope erosion (L256 and Fig 4). I don't think these can be real deposition zones but instead must be a consequence of incorrect landslide detection. Their spatial extent and depth distribution would be a useful indication of the precision of the technique.

Fourth, it is extremely unusual that locations classed as vegetated in post-event optical imagery but identified as a landslide by another technique are considered by the authors to be genuine landslides (L347-8). Instead, the presence of vegetation at the location strongly suggests a false positive.

Fifth, you find no statistically significant difference between total landslide erosion and deposition (L459), does that mean that there was no significant mass loss due to landsliding from the study area? If so, the earthquake's role has been almost exclusively to break up the material and redistribute it within the study area. This seems to differ from the findings of previous studies on the influence of coseismic landslides on earthquake mass balance. It would be very interesting to see the elevation change of the mass (perhaps as elevation pdfs for scars and deposits). This might be a valuable contribution on the instantaneous impact of coseismic landslides on mass balance.

MC6) Areas should be calculated surface parallel

I agree that variation in surface normal orientation could introduce bias in the volume estimate (L238-9) and that volume should be calculated normal to the plane on which landslide area is calculated. However, it isn't clear to me that a horizontal plane is most appropriate for area calculation. This will severely underestimate landslide area on the very steep slopes typical of rockfall initiation. Surface parallel area calculations would be more appropriate and orthogonal measurements for volume calculation would still be possible. I don't find the argument based on retaining consistency with other studies (L279-80) convincing. Your inventory appears to contain a significant number of rockfalls and their area and volume are often calculated in a vertical rather than horizontal reference frame. If you do not alter the reference frame you should certainly report the distribution of surface slopes of your landslides and comment on the effect of vertically projected area given this distribution.

MC7) Reactivation is actually under-detection on bare or sparsely vegetated slopes

I am not convinced that the set of landslides that you are describing are really re-activated landslides (L344-5) but I'm also not convinced that reactivation is really the relevant issue. I think you demonstrate that you are able to detect landslides in vegetation sparse or vegetation free areas that have generally suffered serious underreporting. Some of these areas may be bare due to previous landsliding, others may not but the key point is that inventories have underestimated landslide density due to underdetection in these areas.

However, if you do want to focus specifically on reactivation then you need to clearly define reactivated landslides and detail how you classify them as such. For example: Does a landslide need to reoccur within the footprint of an existing landslide to be a reactivation or is retrogression included within reactivation? Within what time window following the first landslide must the second occur for it to be considered reactivation?

MC8) Landslide object detection needs to be tested against an independent dataset

The findings in this paper depend critically on the skill with which the proposed method can classify landslide scars and deposits. Thus it is essential that the paper reports testing results that quantify this object based classification skill. At present, "orthophotos were used to visually validate" the classifier (L115-6) but without reporting results of this analysis. I think it is essential that you explicitly explain your sampling and mapping strategy for landslide detection from orthophotos in the methods. You should then include a section in the results where you compare your orthophoto based mapping to the surface differencing approach. However, it is not enough to simply say the orthophoto mapping did not identify landslides that were identified by the surface differencing. You should then go to a carefully chosen (e.g. stratified random) subset of the landslides detected by each method that were not detected by the other (i.e. surface differencing but not ortho-photo mapping and vice versa) to establish as far as possible which of the two methods was in error and why. While finding and mapping thousands of landslides might be time-consuming (L255), confirming their existence and characteristics (e.g. area, shape, depth) would not.

Minor Comments

L32: You need to define landslides early in the paper, it will influence interpretation of your later findings.

L34: “spatial distribution, total volume…”? explaining why these are useful would be helpful because this sets the motivation for everything that follows. Landslide mapping has different requirements depending on its purpose. The purposes you choose here set the metrics against which your own mapping should be evaluated.

L35: “associated direct and secondary hazards” What is the connection between mapping, the first three elements of the list and these later two? It would help to make this explicit. As above, this relates to the purpose of the mapping.

L43: “process specific”: here you recognise the importance of different landslide processes. This is an important point that needs to be reflected in your own analysis (see MC1).

L110&113: It isn't clear what statistic is being referred to here as a metric for vertical accuracy. RMSE? If so it would be better to quote mean error and SDE, giving an indication of the contributions of precision and bias.

L114: “using the classification provided”: More detail is needed on the method used to classify ground points.

L115-6: “orthophotos were used to visually validate”: The results of this validation are missing from the paper.

L135: “core point”: What is the sensitivity of the results to varying core point spacing? The orientation and spacing of the grid of core points should also be described and justified either here or in section 3.2.

L136: “first dataset”: Which is the first dataset? Are your results sensitive to this choice, both in terms of depth area scaling and disagreement in 1) change surfaces and 2) maps of significant positive and negative change?

L137-8: This is not clear: do you mean that you calculate the centroid of each point cloud in 3D then take the magnitude of the 3D vector that connects these two points; or that (for each point cloud) you take the arithmetic mean of differences from the reference plane (defined by D) in a direction normal to that plane? In either case you are performing a spatial averaging at length scale d/2 assuming a uniform kernel. First, is it problematic to perform averaging over length scales larger than the core point spacing? Second does it make sense to assume equal weight in the average with plane parallel radial distance from the core point, or should some form of inverse distance weighted average be used? I would have assumed a weighted average was more appropriate but it would be useful for you to explain why an unweighted mean is more appropriate.

L139: “if not intercept is found…”: This is not clear to me. Do you mean 'if the cylinder does not intersect any points in the second surface? Why would this happen? Does this only occur at the boundaries of the point clouds? How do these intersection failures influence the surface differencing and how do you report them in your later analysis?

L140: “provides uncertainty”: what is the basis / justification for the uncertainty estimate taking this particular form? It looks familiar as it has some similarities to a confidence interval but also some differences. This threshold is important to explain and justify in detail because it is used to threshold discrete landslides in the following analysis. Why threshold at 95% confidence? What is the impact on your findings (total volumes and scaling relationships) of thresholding at a difference CI (e.g. 99 or 90% confidence)?

L145: “detrended roughness”: I think it would be useful to add that these are standard deviations if this is true.

L145: “reg” is quantified using the standard deviation of differences between the surfaces. I think it would be helpful here and elsewhere to use similar notation for the registration error to the other errors being examined here. Why are the local terms converted to standard errors but reg is left as a standard deviation? Finally, the length scale over which reg was calculated would seem to be important here.

L164: “at least 5 data points”: Does removing these points from your analysis alter your results on pdfs and scaling?

L165: “not deemed interesting”: I don't think that this is the right phrase, can you rephrase? What was the impact on your results of applying d=10 m throughout?

L173: “may result in…”: How do you identify when this has happened? What is the objective that you are optimising?

L187: “larger than 1 m”: perhaps give a range, >1 could be very large but really it is on the order of 1-2 m.

L187: “rasterizing the dataset”: at what resolution?

L196: “standard deviation…” These stable areas would be an excellent test of your landslide detection method, indicating the scaling relationships, size distributions, and total volumes generated by artefacts alone.

L199-200: “The registration error…”: this definition should come earlier. It is important for interpreting eqn 2.

L200-1: “manually selected”: This doesn't seem consistent with L191. Was a threshold of change selected manually then all areas with changes smaller than threshold included? Or, did you take a subset of pixels from the areas with changes smaller than the threshold so that only large patches were considered?

L202-4: “The standard deviation”: This is great. Clear and useful!

L215: 15-20% it might be useful to show the location of these points on one of your maps.

L217-9: “all geomorphic processes…”: This assumes that all non-fluvial changes are landslides (see MC1).

L221-234: “Landslide source and deposit segmentation”: This process is much more clearly articulated in the SI. When I finished this paragraph I was totally confused but when I read the SI it became clear. The SI text on this analysis is really important and needs to be included within the main manuscript.

L225: “compact”: I don't understand what compact means in this context

L225: More detail on how this method works is essential here. This step is a key component of your method and it is very important that the algorithm is given the same clear treatment that others have been in previous sections. I guess that Dm sets the distance between points within which two points are considered connected. This is a tuning parameter and it is not clear to me what objective you are seeking to optimise during tuning.

L226: “amalgamation effect and the over-segmentation”: These two effects and your method for quantifying them need to be explained. I don't think this can be included only in SI.

L228: “small artefacts”. Do you mean that there are a small number or that they are small in magnitude?

L229: Why would it be the smallest landslides that are particularly affected by “these artefacts”? I had assumed that these were due to things like inclusion of points on either side of ridges.

L230: “minimum number of points”: You haven't explained what this parameter controls. Is it the minimum number of connected components required for a patch to be retained (i.e. minimum landslide size)?

L231: I must have misunderstood Np, it does not follow that all detected changes will be artefacts if Np simply removes landslides smaller than some minimum size. On reading the SI I now understand, this is because you aren't comparing before and after point clouds. You explain this well in the SI, but need to bring some of that text in here.

L231-2: Reference to pre- and post- earthquake data here is confusing unless read in conjunction with the SI.

L233: “artefacts”: What are artefacts? How do you identify them and why do they become smaller/less frequent/less important for larger areas? What does it mean that they are limited (e.g. they make up less than x% of the landslide objects)? The SI text on this analysis is really important point that needs to be included within the main manuscript.

L238-9: I agree that volume should be calculated normal to the plane on which landslide area is calculated. But disagree that a horizontal reference frame is most appropriate. This will underestimate landslide area (severely on the very steep slopes typical of rockfall initiation). Surface parallel calculations seem more appropriate (see MC6).

L245-7: “specific to each landslide”: This potentially presents a nice opportunity to check your method by comparing source and deposit volumes for individual landslides over your study area.

L252-3: “Most of the detected changes on hillslopes correspond to…” How do you know this? Is it based on your definition that all non-fluvial change is due to mass movement? Or, are you drawing on additional information to interpret the change map? Ideally you would have some independent check data.

L254: “previously unstable bare rock”: It is useful to know that many are located on bare rock but how do you know that it was previously unstable and what do you mean by previously (this implies a timescale)?

L255: “Their large number illustrates how difficult”: Why would their large number make them difficult to manually extract? Perhaps time-consuming rather than difficult. An inventory with 1431 sources is a very modest sample size by the standards of modern landslide inventories.

L256: “deposit areas” Some of these areas do not have any upslope erosion. How do you explain this? See MC5

L259: “stable areas”: This argument seems somewhat circular since stable areas were identified from surface differencing rather than independent information.

L259: What do you mean by “artificial changes” here?

L286: It is not clear what “substantial” means in this context.

L291: A rollover is considered characteristic of landslide, but not of rockfall, distributions I think. As I understand it your inventory includes both. It would be useful to comment on this amalgamation of processes within a single inventory and what it implies for interpretation of the landslide distribution. See MC1.

L292-3: “if we reduce the minimum landslide size to…” I don't think that you can make this statement if you have excluded landslides smaller than 20 m^2 for good reason. Your methods text makes a compelling argument that surface differencing errors will generate landslide-like artefacts that will distort the distribution and thus should be censored.

L305-7: This is connected to my earlier concern about planimetric areas and suggests a need either to address the problem as you suggest here or to include a more detailed explanation for when the problem presents itself, what fraction of the landslides in your inventory might suffer from it. See MC6.

L308: Why frame your analysis in terms of volume-area relationships rather than depth-area? Since volume is the product of depth and area the x and y variables are correlated by definition. Others e.g. Larsen have examined depth-area scaling and the translation from one to the other is straightforward. But using depth de-trends the y-axis making it easier to see the scatter in the relationship, which is a major advantage.

L315: It is not surprising that your depth area scaling relationship is gentler than that of Larsen since you censor core points with difference < 0.33, making it impossible for small landslides to also be shallow.

L325: An analysis of the uncertainty associated with total volume estimates would be very valuable. Your study is unusual in not only being able to resolve scar volumes but also deposit volumes, thus you can calculate a net volume removed from the hillslopes. This value is of considerable interest and comparing it to the estimates that have been made elsewhere using different approaches seems worthwhile. You should however deal with your own uncertainties as well and within those uncertainties should be included those areas of deposition that have no landslide source upslope.

L344-5: A clear definition of reactivated landslides is needed here. See MC7.

L347: In many areas "bare rock areas" surfaces do not necessarily indicate recent landsliding (depending on how recent is defined). They may in this landscape though, so it would be useful to say so and evidence the claim.

L349: "following classical approaches": This is not a convincing justification for the method. It argues for the method's validity because others have previously adopted it without any further indication of why it might be valid here.

Section 5.1.1: This section is excellent, this alone is a major contribution!

L384: I don't think you do explain the subsampling exercise in section 3.2 it is currently in the SI as I understand it.

L431-3: "more advanced segmentation": an indication of the classification skill of the simple segmentation of object based classification skill is an essential requirement of this paper.

L460: "deposits form more concentrated and thicker patches": This result is surprising and should be compared with expectations from other studies on landslide runout.

L461-2: I would have expected "very shallow rockfalls" to result in even shallower rockfall deposits because they are initially small in volume and spread out over a wide area during deposition. Reference to rockfalls here increases my concern around use of horizontal (i.e. planform) area calculations.

L465: It is not clear what you mean by "landslide analysis". I suggest cutting this phrase, I don't think it is necessary.

L466: Whether the right tail is a "power law" or not is debated. See Medwedeff et al. (2020) among others.

L473-4: Power law scaling can be sensitive to amalgamation errors such as those in your inventory. As a result I struggle to know how much confidence I can place in these values. See MC3.

L488-90: "rollover … is likely caused by an under detection of small landslides": This is a very important claim with significant implications for our understanding of landslide mechanics. But it depends on both: 1) your findings being robust to error in segmentation; and 2) the inventories that you are discussing having comparable definitions of landslides. Given that the landslide detection method has not been tested and that there are reasons to expect that the method introduces considerable bias in the size distribution I am not convinced that the claim is true. The lack of a rollover in your dataset might instead be due to detection errors.

L499-500: "most of our inventory is relevant to shallow landsliding": shallow landsliding is usually defined as landslides initiating above the soil-bedrock interface, and thus distinct from rockfall. Can you make this distinction in your inventory?

L511: "cannot resolve the amalgamation…": The problem of amalgamation needs a more detailed treatment. Total volume is insensitive to amalgamation but landslide size distributions and scaling relationships are not. See MC3.

L535: "a much lower detection level than optical methods": This phrase is unclear, can you rephrase?

L551: "95% confidence of 0.34 m": This is quoted here as spatially invariant but the method captures spatial variability. The value here doesn't account for the impact of local roughness on the confidence intervals.

L558: "Amalgamation in 3D is still a potential issue": Amalgamation needs a more detailed treatment, given its potential impact on your findings it cannot simply be flagged as a potential issue.

L561: “reactivation” needs to be much more clearly defined. However, your argument doesn't really depend on reactivation but on landslide detection in vegetation sparse or vegetation free areas. I suggest restating in these terms.

L566: “V-A relationship”: depth-area should be examined as well as or instead of V-A because it removes spurious correlation with area. All three geometric properties are likely to be very severely affected by detection errors.

L569-70: “largest and deepest landslides deviate significantly”: these slides are also most likely to suffer amalgamation.

Figures

Figs: 3,4&6: Though they are very well presented I find the 3D plots (Fig. 3, 4 and 6) very difficult to interpret. I feel strongly that all these Figures should be presented in 2D map view instead of as oblique images.

Fig 4: The colourmap in Fig4a does not match description in the caption or in the text. It appears to show a single colour for erosion and another for deposition. It also appears to apply a filter to areas with non-significant change. Areas of significant change are shown in Fig4b so the mask is not necessary here. By masking non-significant changes you lose the opportunity to compare the magnitude of change before significance tested. The colourbars in both Fig4a and c are a good length but contain little usable information. As I see it there is no change in colour for erosion/deposition >1 m in Fig 4a and 5 m in Fig 4c. In addition, the scaling is asymmetric in 4c making it difficult to compare erosion and deposition.

Fig 5: The 3D minimum volume line needs explaining in the caption. I would have expected this line to be oriented parallel to the depth contours since minimum volume for a given area is set by minimum detectable depth. If so the minimum detectable volume might explain the sharp lower boundary to the volume area point cloud.

Fig 6: The inventory of Massey is useful additional information but scar outlines would be more useful than controids.

Fig 6b: It would be useful to identify the largest four or five landslides in the right hand panel of Fig 6b either by giving them individual outlines or different colours this would give an indication of the degree of amalgamation.