Second review of Bernard et al. 2020 by David Milledge.

This is an excellent paper that makes at least two significant contributions. The first is methodological, detailing a method with which to robustly determine topographic change over large areas that include steep slopes. The second is substantive, demonstrating that the size distributions (volume and area) and geometric scaling relationships for landslides differ from those previously found for landslides. The authors have considerably improved the manuscript since my last review. However, there are a number of major and minor comments from my previous review that have yet to be fully addressed. Below I have retained only these comments and explained (in bold) why I feel that they remain unaddressed. I have also made minor comments on the new draft of the manuscript by commenting on the PDF.

The workflow that you have introduced has great potential to improve the quality of landslide inventories. The paper is a significant and rigorous contribution because it: 1) introduces the workflow for a suitable case study, 2) shows that the workflow improves on alternative 3D methods and can detect landslides not detectable in 2D methods; 3) highlights common errors in 2D methods that have been proposed but rarely demonstrated, and 4) demonstrates for the first time that you can calculate a volumetric budget for landslide derived topographic change without the need for volume-area scaling relationships which are known to (and which you show) introduce considerable uncertainty. It also opens up a discussion about what constitutes a landslide and how this differs from the things that we currently map, whether in 2D or 3D.

I agree with you when you say in your response that *"despite its limitations, it currently represents the state of the art in terms of 3D landslide detection and landslide inventory creation"*. **This is an excellent methodological contribution and I have only very minor comments on presentation of the paper in relation to this aspect of the work. As a methodological contribution I agree when you say that:** *"Scientific research is incremental, and we fully expect that our workflow will be improved in coming years by others, as it was the case for 2D landslide inventories.* **In relation to this, it is excellent that you:** *"provide all the elements (code, data) for other researchers to apply the workflow and reproduce our results, or apply the workflow to their data."* **The reason I am so demanding of the checks you apply to the 3D method relative to the 2D method is that the flaws in the 2D method are relatively well rehearsed in the literature but you are presenting the 3D method as a new (and better) technique. To do so you must demonstrate that this is the case.**

However, I strongly disagree that you *"clearly demonstrate, at least in our study case, that the rollover in the pdf(A) observed in our 2D inventory but not in the 3D one, is due to a size-dependent under detection in 2D."* **I disagree because I think that non-trivial errors remain in: 1) the detection of small landslides, where some false positives remain due to spatially correlated errors; and 2) defining the boundaries of landslides, where automated segmentation continues to result in amalgamation. I know that I come to this with a bias: I have interpreted my own field observations as indicating a rollover in landslide size and have developed theoretical explanations for that rollover. So I am probably resistant to the idea that landslide size distributions lack a rollover. I'm trying to avoid this bias but may not manage it.**

You could easily address my outstanding major concerns by softening your claims. For example, you say in your response: *"its application in comparison with a 2D landslide inventory shows that landslide under-detection in image based inventories is extremely prevalent in our study area"*. **I broadly agree but I would say: 1) that the comparison "suggests" rather than** *"shows"* **because you cannot identify which inventory is in error only argue which is more probably the source of the error; and 2) that the under-detection is "present" rather than** *"extremely prevalent"* **because you can argue that it is very likely that some of the size-dependent bias between inventories is extremely likely to be due to 2D under-detection but some is also extremely likely to be due to 3D over-detection and you cannot currently identify their relative share.**

Two of your key conclusions (stated in the abstract) are that the manually mapped 2D inventory *"severely underestimates total area and volume"* [L20] and that there is *"a systematic size-dependent under-detection in the 2D inventory"* [L24]. However, both these statements are underpinned by an assumption that the 3D sources are correct (i.e. the ground truth) such that differences between them and the 2D inventory are attributable to error in the 2D inventory. This assertion needs to be justified in the paper but it is not at present. Instead you consistently assert and assume that in cases where the two inventories differ it is the 3D inventory that is correct.

Another key conclusion of the paper is that the 2D size distribution has a rollover whereas the 3D distribution does not and that this is due to missed landslides in the 2D inventory. However, it is not clear that this is a fair comparison. The rollover is detected in the manual mapping on the basis of a reduced frequency of landslides in the smallest class 13-20 m^2, relative to the class 20-31 m^2 (which is the modal class). This smallest class is below the lower limit of detection for the 3D method. If you enforced a single consistent lower size limit for your analysis and censored all landslides smaller than this limit for both datasets then I don't think you would conclude that the manual mapping displayed a rollover. Note, that the x-axis values for the size distributions in Figure 8 are lower bin limits not central values this is potentially confusing and should be adjusted.

I think you make one further important finding that you could highlight in the abstract: you demonstrate the variety of types of topographic change that occur in response to an earthquake and show that existing 2D landslide mapping captures only a small part of that range. Your results prompt questions about what constitutes a landslide within these landscapes and how we should delimit them. This is particularly important for size distributions because the way that you define both your term landslide (to say what is in or out of the class) and the boundaries of your landslide in space on the basis of post failure observations will differ depending on the motivation for examining them. For example, your point on L420 highlights the complexity of mapping post-earthquake topographic change and relating it to processes. Should subsidence / retrogressive slumping upslope of a catastrophic landslide be included within the same source zone? Is this one landslide or two? The processes and perhaps even the timing of movement are quite different. But it is a very important point that these movements will not be captured in conventional inventories though there is widespread recognition of the processes you discuss based on field reconnaissance.

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).

We added the following definition of what we consider as "landslide" in the introduction of the manuscript: "We use the generic term of "landslide" to define the spatially coherent changes detected by our method on hillslopes that result in at least several decimeter erosion (i.e.,scars or sources) or deposition". The discussion now features an entire paragraph (L596) addressing the various type of landslides that can be detected by our approach. The aim of this paper is not to better understand landslide mechanics as we cannot confidently identify the different landslide processes we detect. We are mainly interested by the estimation of co-seismic volume and to overcome issues such as under-detection and amalgamation on volume estimation. The introduction now clearly integrates these two problematics. We also believe that the new filtering approach that we introduce, which results in 3 time less landslide sources compared to the initial MS, results in a much more robust inventory.

RE: definitions. Simply defining landslides as decimetre scale change not due to fluvial processes is accurate but will be a very unusual definition to the reader. You can help them to see that your definition of a landslide is consistent with theirs by adding a little more explanation and I think that would be very worthwhile.

Crozier suggests that: "*The three most widely used classifications involving landslides (Sharpe, 1938; Varnes, 1958 and 1978; Hutchinson 1988) separate 'mass movements' (Fairbridge, 1968) into two categories: subsidence (which is the vertical sinking of material-see entry on Land Subsidence) and those movements that occur on slopes. These' slope movements' are then usually divided firstly into 'landslides,' as defined above, and secondly into the slower, more widespread and ill-defined movements such as 'creep,'' sagging,' and 'rebound."* The landslide definition that he refers to is: "*the downward or outward movement of a mass of slope-forming material under the influence of gravity, occurring on discrete boundaries and taking place initially without the aid of water as a transportational agent."*

Crozier M.J. (1999) Landslides. In: Environmental Geology. Encyclopedia of Earth Science. Springer, Dordrecht.

I think you can make the case that most of the change that you detect can be classified as landslides following the definitions of Sharpe (1938), Varnes (1958, 1978), Hutchinson (1988), and Crozier (1999). But you need to make that case. If you explain the timescale over which the change occurs and the spatial limits of detection that you will ultimately impose then you can argue that everything that you detect should fall within the class of landslide. It would help your later discussion if you gave a summary of what that might include (e.g. slides where the failed material is entirely removed from the source zone and those where movements that are small relative to the length of their failure surface). It would also help to explain which non-fluvial mass movement processes are not detected, particularly: tree-throw, ravel and other forms of creep (because the movements are either too small, too localised, or too slow).

It would perhaps also be worth saying that this definition differs from those commonly (implicitly) applied in 2D landslide inventories derived from satellite imagery since these rarely (or incompletely) capture slides where material is displaced by only a fraction of the failure surface. These inventories are censored by their ability to detect change from image properties and thus rarely capture rockfall source zones. The same censoring results in under-sampling of small landslides, landslides in bare or sparse vegetation and landslides obscured by forest canopy because these can't be confidently identified.

<u>RE: expected uses.</u> Clarifying the focus on co-seismic landslide volume estimation is useful as is the section that you have added on the different processes represented in your inventory.

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.

The reviewer suggested many areas to explore that are extremely interesting, but which, for some of them, would constitute an entire paper by themselves, in particular when it comes to the analysis of error properties suggested by the reviewer.

I am not suggesting that all these areas need to be exhaustively explored but that the inferences that you draw from them must be stated with appropriate confidence for the uncertainty in the data that underpins them. Mass balance, which was your primary objective is largely insensitive to the errors that I highlight here. Landslide size distributions and scaling relationships are potentially very sensitive to these errors.

We also aim at developing a generic workflow applicable to a variety of cases for which users may not necessarily perform extensive error properties analysis.

Your contribution in developing a workflow is extremely valuable and is not compromised by the continued presence of these errors. However, prospective users will also use this paper as a model for what the workflow can be used to calculate. Your discussion of errors and their implications is therefore important because it will influence not only how people interpret your results but also the capability and limitations of the workflow.

Hence, to improve the paper we have worked on two aspects: (i) improving and better identifying errors in our dataset, (ii) defining a new confidence metrics (the SNR) for each landslide source or deposit to filter out landslide with low confidence.

Both these aspects have very considerably improved both the workflow and its application in this paper. However, you have not addressed my original concern about spatially correlated error in this comment.

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 110 *registration errors.*

I do not see where this comment is addressed in the response. As far as I can tell: 1) there are spatially correlated errors in your difference surface; 2) these errors will not be captured by the SSDS analysis; 3) correlation of errors implies that if one core point has errors large enough to exceed LoD then there is a non-trivial probability that one of its neighbours will also have errors that exceed LoD; and 4) the distribution of erroneous patches will be strongly right skewed (i.e. smaller patches more probable than larger patches). I would be keen to know whether you agree. On this basis, I think you must quantify the impact of spatially correlated errors on your size distribution if you are to argue that your measured size distribution is the 'true' distribution or even that it is more correct than the 2D distribution.

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.

Errors in our dataset: entirely revisiting our data, we identified two sources of errors: (1) Remaining LiDAR point cloud misclassification in forest areas, inducing local topographic errors, and (2) imperfect flight line alignments from the pre-EQ data, inducing topographic errors of longer wavelength. To address the first issue, we first removed as much misclassified points as possible by interpolating a surface and remove outlier points (see detail in supplementary material and section 2 in the paper). To address the second, we estimated residual errors of each flight lines due to imperfect flight line alignments composing the pre-EQ point cloud and defined the registration error *reg* based on the maximum residual error of the flight line misalignments (section 3.3.1. and S3 in supplements). Compared to the previous version of the MS, the *reg* is now 3 cm larger (20 cm vs 17 cm). We also show that only 1% of points are detected as significant change in the stable area, validating our choice of $LoD_{95\%}$,. While our *reg* is considered uniform over the study area, the $LoD_{95\%}$ (eq.2) also take into account the local

point cloud density and roughness which are correlated to the presence of vegetation. The $LoD_{95\%}$ is thus spatially variable. In addition, we are aware that, ideally, a spatially variable model for point cloud error and registration would be preferable for each survey and combined into a more accurate and complete form of *LoD* than what the M3C2 approach currently offers. However, in the absence of the position and attitude information of the sensor (e.g., Smoothed Best Estimate of Trajectory file) and raw LiDAR data - rarely available on LiDAR data repositories -, or of dense ground control which is hard to get in mountainous environment, it is currently impossible. We now discuss this in the discussion (section 5.1.2).

This is a useful explanation but the decision to assume that registration error is spatially uniform still needs justifying in the text in a way that addresses the concern that long-wavelength errors might combine with short wavelength errors to generate patches of erroneous change in some places and break up patches of true change in others.

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.

Filtering landslides with low confidence with the SNR: To limit the false detection due to these errors, we also defined a signal-to-noise ratio threshold to efficiently remove suspicious landslides (section 3.3.4 and 4.1). This index is based on the mean ratio between the 3D-M3C2 distance and the $LoD_{95\%}$ for each landslide. We provide a way to evaluate the optimal value of SNR by comparing the number of landslides in the database to a case with no change (i.e., two versions of the same data but with different sampling, a test that we now call Same Data Different Sampling test (section 3.2). This SNR filtering removes a very large number of landslide source and deposits (fig. 5b), and in particular long wavelength low amplitude changes that occurred due flight line misalignment in the pre-EQ data, as well as many small landslides in forested region where point density is very low. We provide a systematic analysis of the impact of SNR (and *reg*) on pdf (A) (fig. 11), pdf (V) (fig. S 13) and V-A relationship (fig. S14).

This text does not directly address my concern above. SNR filtering is a good addition to the workflow and it does appear to successfully remove a very large number of changes that were erroneously identified as landslides before. The only test of whether this filtering is sufficient is the SSDS (Same surface different sampling, SSDS and SDDS used interchangeably in both the paper and the response). If I have understood it correctly, this is a very weak test of the workflow because it examines only the impact of random errors and ignores spatially correlated errors in the surfaces.

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).*

Did you undertake this analysis? It seems straightforward to do but I don't see any response to this comment in your response. You do not use the mapped distribution as a test for your distribution and I understand why. However, you don't currently offer any independent test of your distribution.

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.

We agree that the segmentation approach we use certainly do not allow to solve the amalgamation problem, and this is highlighted in many parts of the MS. However, the problem of amalgamation is inherently subjective and plagues all inventories.

I agree, that segmentation is a subjective problem but you have made it reproducible by removing the subjectivity. The problem is that the best reproducible (i.e. automated) segmentations still perform poorly (with respect to the segmentation that a human mapper would choose).

Our 3D data reveals a level of complexity, and a density of amalgamated landslides which makes the definition of a single landslide in relation to an ideal rupture mechanism extremely difficult.

This is a really important point and could be a key contribution of the paper. Your results show that it's complicated. Far more complicated than we capture in conventional 2D inventories.

Even the segmentation of the 2D inventory proved to be extremely complex and is entirely not reproducible. Hence, we favour a reproducible approach, even if currently limited, that can be applied exhaustively to much larger datasets, than a non-reproducible one (2D manual mapping) that we now demonstrate misses a very large number of landslides and incorrectly map their contour.

It isn't clear why reproducibility is favoured over skill. If you think that manual segmentation would outperform connected component segmentation it seems strange to continue with automated approach because it is reproducible. You say that 2D manual mapping "misses a very large number of landslides" but this is on the assumption that the 3D inventory is correct.

In this new version of the paper, the landslide amalgamation can be visualized with a map of the landslide source colored by individual landslide as defined by the method (section 4.2 Fig.7). Moreover, the comparison between both inventories shows that while 171 of 2D-sources are shared with 3D-sources, it represents 144 sources 3D-sources. This highlight that 25 landslide sources are amalgamated in the 3D inventory (L-. As in the previous version of the paper, we perform a sensitivity analysis of the impact of D_m showing that landslide statistics are not severely affected by this parameter for $1.5 < D_m < 3$. We also explored density based spatial clustering algorithm used in 3D rockfall segmentation, derived from DBSCAN (Ankerst et al., 1999; Martin Ester, Hans-Peter Kriegel, Jiirg Sander, 1996; Tonini and Abellan, 2014) and HDBSCAN (Carrea et al., 2021; McInnes et al., 2017). None of them managed to provide a significantly better segmentation of the largest landslide and are significantly longer to run than the connected component algorithm we use. We now thoroughly discuss about this in section 5.1.2.

The analysis that you have added demonstrates that the problem is not that your particular segmentation approach is worse than the alternatives but that automated segmentation itself is problematic. You make the argument above the subjective segmentation is equally problematic. I think making this point in a more detailed discussion of the problem of segmentation (both automated and manual) would help to address my concern here.

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.*

We added further analyses to the method to deal with topographic errors and erroneous landslide. We also added an entirely new 2D landslide inventory as suggested by the reviewer. We are now confident that the actual landslide inventory corresponds to real changes. Please see our reply to the MC2) and MC8) comments for a detail answer.

Most of my concerns in this comment have been left unaddressed. MC2 focusses on topographic errors but your response to MC2 doesn't deal with the problem of spatially correlated errors. Because you are interested in the size of patches generated by thresholding the difference surface it is essential that you examine the spatial structure of the errors. I will deal with each of my comments in turn reflecting on whether they have been addressed in the new manuscript even if they have not directly been addressed by your response to this comment.

First, you did 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)". **You remove reference to an SSDS test to set the minimum area but retain that test to optimise the SNR. However, when you introduce it you do not recognise that it synthesises uncorrelated noise while the two surfaces that you are differencing both include spatially correlated elevation errors. You do not include any description or explanation in the text for the "stable areas error" shown in Figure 4, these errors appear to be spatially structured on multiple length scales from tens to hundreds of metres.**

Second, I argued that *"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"* **and that** *"You need to quantify these detection biases before you can make these claims."* **I'm still not convinced by this claim because I still don't think you have excluded the possibility that the lack of rollover is solely due to detection errors. You need to quantify the size dependent detection bias in the 3D inventory and/or to considerably tone down your claims about rollover in this and other 2D inventories being due to under-detection.**

Third, I suggested that you: *"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.* **I don't see a response to this suggestion here but when I raised the same point in a minor comment (related to L196). you responded that** *"Applied on stable area, the workflow does not detect any landslide."* **This result would definitely be worth reporting! However, I think we must have misunderstood one another, I can see many patches of significant change (>150) within the stable areas, most of these patches of significant change are removed in Figure 5c. Is that because they are smaller than 20 m^2? However, even after this filtering I can still see several landslides within the stable zones in Figure 5c. I phrased this as a suggestion in my previous review but I really think this is one of the few opportunities that you have to build confidence in your method. It remains a weak test because you chose the stable areas based on areas of limited change in the difference maps and because they oversample non-forest vegetation but in the absence of field checks to the inventory this remains one of the best tests I can come up with. My second suggestion to** *"collect a landslide check dataset using independent observations"* **is dealt with in a separate major comment MC8 and doesn't need further discussion here.**

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.

First: contrary to the reviewer experience, this result does not surprise us, in particular when the runout of landslides is not long. The filtered data clearly support this finding.

This has not been addressed. I commented that you needed to compare your results with those from previous studies with respect to scar and deposit depths and scar and deposit areas. I didn't find this new discussion nor a response to explain why it was not necessary.

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.

Third: We now filter landslides by a signal-to-noise ratio (section 3.3.4) and are confident that the actual landslide inventory corresponds to real changes. Some very small deposit areas may not have upslope erosion as we expect deposit are to be easily detectable by our method than source areas (section 5.1.2).

Why do you expect that deposits are more easily detectable than source zones? I would have expected the opposite. In my experience deposits can be very thin, (<50 mm) patchy and extensive whereas source zones are far more coherent.

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.

Fourth: we partially disagree with this statement. As now explained in the discussion large landslides that strip out vegetation are obviously mapped in 2D inventory, but small ones that occur on less dense area are extremely difficult to map in 2D imagery as our inventory shows. Moreover, vertical subsidence due to upslope propagation of landslides is entirely missed in forest, while it is detected in our approach. We think the comparison with the new 2D inventory will resolve the reviewer's reserve.

How do you know that small ones that occur on less dense area are difficult to map? Is this on the assumption that your LiDAR inventory is correct? If you go to the landslide locations predicted by the 3D method do you find evidence in the orthophotos that there is indeed a landslide at that location (even if it wasn't independently mappable)? This would help to build confidence in your method. The point about vertical subsidence is important and you do a nice job of demonstrating the plausibility of the claim that this is real change. It prompts a series of questions about representation of these landslides within an inventory derived from surface change but you deal with this nicely in the discussion. My only suggestion is that you prepare readers for this finding in the introduction by adding a more complete explanation of the types of landslides that 2D and 3D inventories might include (see MC1).

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 timeconsuming (L255), confirming their existence and characteristics (e.g. area, shape, depth) would not.

This lack of comparison between the 3D differencing method and a more classical approach has been addressed by adding a manual mapping of landslides based on 2D images. We added a section that specifically explore the differences between the methods (section 4.2). Moreover, the resulted landslide area distribution mapped manually has been added in the figure in section 4.3 and compared with those obtained with the 3D differencing method.

The 2D inventory considerably strengthens the paper and addresses the comments in the first paragraph above. However, much of the second paragraph remains unaddressed. You have added analysis of the two inventories and discuss false positives for the 2D inventory where deposit is incorrectly mapped as source zone, this is a secure result and is exactly the type of analysis I was looking for. You use your observations and theory/logic to argue that one method is correct and the other is in error. The remaining areas of disagreement you assign as false negatives for the method that has not identified a landslide at that location. This is not a secure result. You have no objective way of establishing which method is in error (i.e. whether this is a false positive for one or a false negative for the other) and you don't provide any justification for why disagreements should always be treated as false negatives. In fact there is good evidence to suggest that the 3D inventory should contain false positives and that these false positives likely have a strong size bias. First, the SSDS test (which itself is a very weak test because it assumes that errors are uncorrelated in space) results in artefact landslides. Second, the inventory contains landslides in the 'stable' areas of the study area. This problem propagates into the discussion where you describe disagreement between inventories as error in the 2D inventory under the implicit (but untested) assumption that the 3D inventory is correct. I don't think you could use this language in the paper even if the tests above generate only small numbers of false positives from the 3D inventory.

Outstanding Minor Comments from previous review

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

The survey report of the LiDAR data only mention that the ground points have been automatically classified using the Terrascan software. The reference to the survey report (Dolan and Rhodes, 2016) have been added to the text.

I don't see Dolan and Rhodes (2016) cited in the text. The sentence above should be added to the text. You have added a manual quality check and reprocessed the data to remove non-ground-points as a result. That is a good additional step that you have introduced since the last version. However, I don't think you can simply point to the SI for the details of this analysis. This is a key step in your method and should be included in the article itself.

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.

To calculate the distance between the two point clouds, the average positions i_1 and i_2 of the point clouds are first defined and then the distance is computed between the two positions along the normal vector. The average positions are defined as the arithmetic mean of the distance distribution of each point of the subset of points (created by the intersection of the cylinder to the point cloud) to the normal vector (or the cylinder axis; see Lague et al., 2013). As this is part of the M3C2 algorithm that we did not modify, we don't discuss the choice of a uniform kernel rather than a non-uniform one.

You have not changed the manuscript in response to my comment. Your response is useful, particularly: *"the distance is computed between the two positions along the normal vector"* and *"average positions are defined as the arithmetic mean"*. You could easily amend the sentence to clarify this: *"…as the distance of the arithmetic mean positions of the two point clouds along the normal vector"*.

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?

In LiDAR datasets, the density of points is non-uniform over the entire point cloud. Consequently, missing data or very low point density (<5 pt/m²) can occur inside the point cloud due to the absence of laser impact on the ground during the data acquisition. This mostly occur in dense vegetated areas or water surface areas (for topographic lidar). When performing M3C2, it is thus possible that the cylinder cannot intercept points or just a few (< 5). In both cases, the M3C2 distance will not be considered significant. In areas with low point density (<5 pts/m²) a solution is to perform M3C2 with a larger projection scale *d* to include more points in the distance and statistic calculations.

You have not updated the text to reflect this discussion. You do mean: 'if the cylinder intersects <5 points in the second point cloud'. If so, this should be added to the paper. Your explanation above is useful but I understand it to mean something different to what you say in the paper.

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)?

We refer the reviewer to the original M3C2 paper which has an extensive discussion on the confidence interval, and how to consider surface roughness, point cloud errors, point density and registration in the context of change detection on 3D point clouds. The threshold has been set to 95% to build the segmentation on as many good points as possible. We do not believe that changing this threshold to 90 or 99% significantly change the landslide statistics given the results of the sensitivity analyses of parameters (reg, D_m and SNR threshold) that mainly control the landslide inventory.

First, I think you can evidence your statement above from your sensitivity analysis and doing so will strengthen the paper. You demonstrate that changes to the LoD do not alter your main findings and the changes you explore cover the range that you would expect from changing the threshold CI from 90-99% (i.e. 0.39-0.52 m). This strengthens your argument that the statistics are robust to model choices.

Second, having read the original M3C2 paper (Lague et al., 2013), it discusses the confidence interval and registration error but took a different approach to estimating registration error so it is difficult to translate directly between the two papers. They provide only a brief description of the theoretical basis for the LoD equation citing a statistics textbook.

I found the description of James et al (2017) who you cite and who cite Lague et al. (2013) very useful: *"where reg is the relative overall registration error between the surveys, assumed isotropic and spatially uniform (Lague et al., 2013). Note that Lague et al. (2013) took a conservative approach by adding reg directly (as a potential systematic bias)"*. A similar statement would be useful in your paper to explain that you estimate local uncorrelated random errors using the first two terms and systematic errors under the assumption that they introduce a spatially uniform bias with the final term.

Anderson (2019) describes an approach similar to yours, but highlights spatially correlated random errors as a key component within error analysis for surface differencing and includes this as a term in his total error calculations (eqn 21). This term is missing from your error propagation but seems likely to be very important, particularly because you are interested in the size of thresholded difference patches. Anderson (2019) also argues for direct characterisation of errors within 'stable areas' similar to my suggestion in MC4.

Anderson, S.W., 2019. Uncertainty in quantitative analyses of topographic change: error propagation and the role of thresholding. *Earth Surface Processes and Landforms*, *44*(5), pp.1015-1033.

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.

We choose to keep the "reg" notation to specifically refer to the registration error when needed in the manuscript. We now provide an explanation in the text as to why reg is not converted as a standard error (L 166). Reg is not measured over a length scale, it is based on the standard deviation of the 3D-M3C2 distance between 2 clouds. This part is now explained in greated detail in section 3.3.1 where we discuss the notion of intra-survey and inter-survey registration quality.

You now say: *"The M3C2 definition of the LoD95% makes the conservative choice of adding the registration error to the combined standard error related to point cloud roughness, rather than taking the square root of the sum of squared standard error and squared registration."*

First, I agree that your equation is conservative in the sense that it results in a larger LoD but I don't see a justification for this functional form either here or in Lague et al. (2013). The best description of a framework for propagation of both random and systematic errors that I can find is in Anderson (2018). It is similar to yours in its approximation of random and systematic errors (see Anderson's eqns 12 and 20) but differs from yours in how these are combined (see eqns 21-22). Can you explain the difference?

Second, you have not explained why standard deviation rather than standard error is used for the systematic error (reg). I think this is because standard errors are used to approximate random errors (under the assumption that they are uncorrelated) but standard deviations are used for systematic errors under the assumption that these are perfectly correlated (see Anderson's eqns 1-4, 12 and 20). If that is the case it would be helpful to explain it in the text.

Third, I'm happy for you to retain a notation that is specific to registration error but suggest sigma with reg as subscript would make it clearer that this is a standard deviation.

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?

The sentence has been modified. Applying d=10 m increases the landslide source and deposit by 1% and 0.7% respectively. In terms of volume this represents an increase of 2% and 0.89% respectively.

I do not see where you have modified the text to reflect this response.

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

This can be identified in the bottom of narrow valleys and top of very steep divides where no evidence of mass movement processing can be identified by visual inspection of orthophotos. These cases are now filtered by using the SNR metric as they have very large $LoD_{95\%}$ (see section 3.3.4).

This is still not clear both in the response and in the text (which has changed very little). L211 you say *"This is generally obtained by trial and error"*. This was what I was referring to when I asked *"what is the objective was that you are optimising"* by trial and error. Re-reading the text I am not sure whether it was obtained by trial and error in this paper. If not then how did you find out that the maximum observed change in the study area was 30 m? If p_{max} is designed to prevent anomalously large changes how do you verify that a change is an anomaly? Something seems circular in the argument as it is currently presented. I don't understand the connection to SNR filtering and that doesn't currently feature in the text.

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.

This analyse has not been performed due to the changes we made on how we manage artefacts with SNR filtering (see section 3.3.4). Applied on stable area, the workflow does not detect any landslide.

This result is definitely worth reporting. If there are any remaining patches then you should report the size distribution of these patches as they give an indication of the expected bias that elevation errors will introduce into your size distribution.

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

We disagree with this comment as the registration error "reg" can be defined differently depending on the application of M3C2. In section 3.1, we aim to give a general description of the M3C2 algorithm.

I agree that it can be defined differently but you are reporting your method rather than the method in general so you should define it as you have used it.

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

The percentage of these points on steep hillslope has been revised and actually represents up to 12% of steep slopes.

It would still be useful to show the location of the points on one of your maps.

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

Some deposit areas can be detected without an associated upslope erosion. A possible reason is that the surface change associated to these deposit areas is sufficient to be above the $LoD_{95\%}$ but not for the upslope erosion area.

This is useful discussion, I couldn't see it in the revised manuscript. It would be useful in your description of Figure 5.

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.

Indeed, and we now discuss this. In particular, the relationship is barely different if we consider a SNR=1, meaning that compared to the 2D inventory, we miss 32 shallow landslide sources over an inventory of 1270 sources (SNR >=1). Hence, even if the 3d inventory had captured the very shallow landslides that the 2D mapping did capture, it would hardly change the scaling relationship. One could also say that previous 2D inventories have significantly underevaluated the number of small landslides, which in turns affect the representativity of published V-A relationships from 2D inventories.

This is useful discussion, I couldn't see it in the revised manuscript but I wasn't sure whether I was looking for parts of the text in your response above or something different.

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

We think it is not surprising given the small runout of the landslides in this area, and is actually backed by the data in terms of mean 3D thickness of the deposits and sources.

This comment has not been addressed. I cannot find new text comparing your results with expectations from other studies.

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

The reference has been added in the text.

This comment has not been addressed. You have not altered the text to recognise the debate around the form of the right tail. I don't think you can simply add Medwedeff to the current citation. It reads as though Medwedeff et al. are among those arguing for power law scaling when I understand their paper to argue the opposite.

Figures

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. Actually, what it is illustrated here is the 3D minimum volume that can be measured here given the minimum area (20 m^2) that we consider and the minimum significant depth (~0.4 m).

The minimum area and depth act together to set an absolute minimum detectable volume (when both area and depth are minimised). However, there is also a depth dependent minimum detectable volume that is set by the depth constraint alone. At present the horizontal line that you use to highlight the absolute minimum volume might be misinterpreted by some to be the area dependent minimum volume. You can easily fix this by adding the area dependent minimum volume. It will be a straight line in log-log space and will pass through the points (20,8 and 20000,8000). It would be useful to include this on the Figure and would also bring make the interpretation of the dashed lines in the figure and inset internally consistent.

Supplement

S1: this text should definitely be included within the paper itself this is a key part of your method. You should also report the parameters that you used for this analysis and the parameter values that you chose, preferably with a justification.

L11: "4 standard deviations" I have three questions here: 1) what sample is the standard deviation being calculated from? 2) Do you remove points that differ by >4sigma in positive and negative? I'd expect vegetation to result in only positive residuals. 3) How did you choose the 4 standard deviations threshold?

L12: did you repeat three times because there were no outliers after that? If so you should report this, if not you should explain why you chose to stop after three iterations.

Figure S2: This is a useful Figure but I can't distinguish the flight lines based on the legend information, the line styles are not sufficiently distinct. Dashed lines are clearly visible and can be distinguished from the solid line. I think you will need to use different line styles to enable the reader to distinguish this many lines.

I don't understand how you can have only one reference line in Figure S2 and Table S3. Is this a flight line from the post-EQ set? Or is this some combination of points from multiple lines? Either way I think you need to explain this, it will affect how the reader interprets Table S3.

S9: "Determination of forested area" Is this a standard technique? It would help if you could give a reference for the technique. I would like to know how returns are classified (i.e. how different do returns need to be to be classified as two distinct returns). However, if this is a very standard exercise it is fine for you simply to point to a reference.

Figure S10: It would be useful to give more detail in the caption. Something like "corresponding to the number of targets a laser pulse has intercepted" would probably be sufficient.

Figure S12: This is an interesting plot. How do you calculate slope here? Is this based on the gradient of the core points? How do you explain the large number of sources with very low slope? Are these all associated with deep seated failures? This slope data provides another really useful way to check your dataset and therefore to build confidence in your results. You should plot slope for each core point against the size of the patch to which each core point belongs. Core points with low slopes associated with large patches may indicate deep seated landslides; those associated with small patches are likely to be errors because landslides require a steeply dipping failure surface (>20 degrees?) to move. Another useful approach would be to examine average slope for each patch plotted against patch size. As above, small patches with gentle gradients are likely to be errors.

References

Aerial Surveys, Aerial photographs derived from two surveys of the study area carried out in 2014 to 2015 and in 2016 to 2017, by Aerial Surveys Ltd, 2017.

Dolan, J.F.: Data collection and processing report LiDAR survey of five fault segments (Estern Clarence, Western Clarence, Central Eastern Awatere, West Wairau and East Hope-Conway) of the Malborough fault system on the Northwestern portion of New Zealand's south island. Ph.D., University of southern California, 11 pp., 2014

Dolan J.F, Rhodes E.J.. Marlborough Fault System, South Island, New Zealand, airborne lidar. National Center for Airborne Laser Mapping (NCALM), distributed by OpenTopography. http://dx.doi.org/10.5069/G9G44N75, 2016.

Ankerst, M., Breunig, M. M., Kriegel, H.-P. and Sander, J.: OPTICS, ACM SIGMOD Rec., 28(2), 49–60, 775 doi:10.1145/304181.304187, 1999.

Benjamin, J., Rosser, N. J. and Brain, M. J.: Emergent characteristics of rockfall inventories captured at a regional scale, Earth Surf. Process. Landforms, 45(12), 2773–2787, doi:10.1002/esp.4929, 2020.

Carrea, D., Abellan, A., Derron, M., Gauvin, N. and Jaboyedoff, M.: MATLAB Virtual Toolbox for Retrospective Rockfall Source Detection and Volume Estimation Using 3D Point Clouds : A Case Study of a Subalpine Molasse Cliff, 2021.

Lague, D., Brodu, N. and Leroux, J.: Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (N-Z), ISPRS J. Photogramm. Remote Sens., 82, 10–26, doi:10.1016/j.isprsjprs.2013.04.009, 2013.

Larsen, I. J., Montgomery, D. R. and Korup, O.: Landslide erosion controlled by hillslope material, Nat. Geosci., 3(4), 247–251, doi:10.1038/ngeo776, 2010.

Martin Ester, Hans-Peter Kriegel, Jiirg Sander, X. X.: A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, Compr. Chemom., 96(34), 226–231 [online] Available from: https://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf?source=post_page, 1996.

Massey, C. I., Townsend, D., Jones, K., Lukovic, B., Rhoades, D., Morgenstern, R., Rosser, B., Ries, W., Howarth, J., Hamling, I., Petley, D., Clark, M., Wartman, J., Litchfield, N. and Olsen, M.: Volume characteristics of landslides triggered by the M W 7.8 2016 Kaikōura Earthquake, New Zealand, derived from digital surface difference modelling , J. Geophys. Res. Earth Surf., 0–3, doi:10.1029/2019jf005163, 2020.

McInnes, L., Healy, J. and Astels, S.: hdbscan: Hierarchical density based clustering, J. Open Source Softw., 2(11), 205, doi:10.21105/joss.00205, 2017.

Tonini, M. and Abellan, A.: Rockfall detection from terrestrial lidar point clouds: A clustering approach using R, J. Spat. Inf. Sci., 8(1), 95–110, doi:10.5311/JOSIS.2014.8.123, 2014.

Beyond 2D landslide inventories and their rollover: synoptic 3D inventories and volume from repeat LiDAR data

Thomas G. Bernard, Dimitri Lague, Philippe Steer
Univ Rennes, CNRS, Géosciences Rennes - UMR 6118, 35000, Rennes, France

Abstract.

Efficient and robust landslide mapping and volume estimation is essential to rapidly infer landslide spatial distribution, to quantify the role of triggering events on landscape changes and to assess direct and secondary landslide-related geomorphic hazards. Many efforts have been made to develop landslide mapping methods, based on 2D satellite or aerial images, and to constrain the empirical volume-area (V-A) relationship allowing in turn to offer indirect estimates of landslide volume. Despite these efforts, major issues remain including the uncertainty of the V-A scaling, landslide amalgamation and the under-detection of landslides. To address these issues, we propose a new semi-automatic 3D point cloud differencing method to detect geomorphic changes, obtain robust landslide inventories with an uncertainty metric and directly measure the volume and geometric properties of landslides. This method is based on the M3C2 algorithm and was applied to a multi-temporal airborne LiDAR dataset of the Kaikoura region, New Zealand, following the M_w 7.8 earthquake of 14 November 2016.

In a 5 km² area, our 3D workflow detects 524 landslide sources and 304 deposits with a minimum size of 20 m² and a total volume of 752,616 ± 154,165 m^3 and 949,742 ± 150,014 m^3, respectively. Geometric properties of the 3D inventory, including the V-A relationship, are consistent with previous results, except for the lack of the classically observed rollover of the distribution of source area. A 2D inventory hand-crafted from aerial image comparison only identified 258 landslide scars, exhibits a rollover in the distribution of source area and severely underestimates the total area and volume of 3D detected sources by 75 % and 60 %, respectively. Detection and delimitation errors in 2D occurs in areas with limited texture change (bare rock surfaces, forests) and at the transition between sources and deposits that the 3D method accurately captures. Large rotational/translational landslides are missed in the 2D inventory owing to the dominant vertical topographic change. Our data show a systematic size-dependent under-detection in the 2D inventory below 200 m² that may explain all or part of the classical rollover observed in 2D landslide source area distribution. While the 3D segmentation of complex clustered landslides sources remains challenging, we demonstrate that 3D point cloud differencing offers a greater sensitivity to detect small changes than a classical difference of DEMs (digital elevation models). Our results underline the vast potential of 3D-derived inventories in quantifying exhaustively and objectively the impact of extreme events on topographic change in regions prone to landsliding and to explore in new ways the scaling properties of landslides.

1. Introduction

In mountain areas, extreme events such as large earthquakes and typhoons can trigger important topographic changes through landsliding. Landslides are a key agent of hillslope and landscape erosion (Keefer, 1994; Malamud et al., 2004) and represent a significant hazard for local populations. Efficient and rapid mapping of landslides is required to robustly infer their spatial distribution, their total volume and the induced landscape changes. Such information are crucial to understand the role of triggering events on landscape evolution and to manage direct and secondary landslide-related hazards (Guzzetti et al., 2012; Hovius et al., 1997; Marc et al., 2016; Parker et al., 2011). Following a triggering event, total landslide volume over a regional scale is classically determined in two steps: (i) individual landslide mapping using 2D satellite or aerial images (e.g. Behling et al., 2014; Fan et al., 2019; Guzzetti et al., 2012; Li et al., 2014; Malamud et al., 2004; Martha et al., 2010; Massey et al., 2018; Parker et al., 2011) and (ii) indirect volume estimation using a volume-area relationship (e.g. Simonett, 1967; Larsen et al., 2010):

$$V = \alpha A^{\gamma} \quad (1)$$

with V and A the volume and area of individual landslides, α a prefactor and γ a scaling exponent ranging between 1.1 and 1.6 (e.g. Larsen et al., 2010; Massey et al., 2020).

A first source of error comes from the uncertainty on the values of α and γ which tend to be site specific and potentially process specific (e.g. shallow versus bedrock landsliding). This uncertainty could lead to an order magnitude of difference in total estimated volume given the non-linearity of eq. (1) (Larsen et al., 2010). Two other sources of error arise from the detectability of individual landslides themselves and the ability to accurately measure the distribution of landslide areas due to landslide amalgamation and under-detection of landslides. Landslide amalgamation can produce up to 200 % error in the total volume estimation (Li et al., 2014; Marc and Hovius, 2015) and occurs because of landslide spatial clustering or incorrect mapping due, for instance, to automatic processing. Indeed, automatic landslide mapping (Behling et al., 2014; Marc et al., 2019; Martha et al., 2010; Pradhan et al., 2016) relies on the difference in texture, color and spectral properties such as NDVI (Normalized difference vegetation index) of multispectral 2D images between pre- and post-landslide images, assuming that landslides lead to vegetation removal or significant texture change. During this process, difficulties in automatic segmentation of landslide sources can result in incorrect estimate of individual landslide area, which propagates into a much larger estimate of volume owing to the non-linearity of eq. (1). Manual mapping and automatic algorithms based on geometrical and topographical inconsistencies can reduce the amalgamation effect on landslide volume estimation (Marc and Hovius, 2015), but it remains a source of error due to the inherent spatial clustering of landslides and the overlapping of landslide deposits and sources.

Under-detection of landslides can occur because the spectral signature of images is not sufficiently altered by a new failure. Notably, under-detection of small landslides is one hypothesis put forward to explain the rollover classically observed in the distribution of landslide area (e.g., Stark and Hovius, 2001), even if the emergence of a rollover is predicted by mechanical models of landslide failures at the regional scale (e.g. Jeandet et al., 2019; Tanyaş et al., 2019 and references therein). Under-detection can be particularly important in areas with thin soils and sparse or missing vegetation (Behling et al., 2014). It can

be further complicated when using different image sources with different resolution, spectral resolution, projected shadows and consequent ability to detect surface change. Yet, the level of under-detection of landslide in a given inventory remains generally largely unknown. The delimitation of landslides in areas with poor or total lack of vegetation is therefore critical to robustly infer landslide area distribution and total volume. To deal with poor vegetated areas, Behling et al. (2014, 2016) developed a method using temporal NDVI-trajectories which describes the temporal footprints of vegetation changes but cannot fully address complex cases when texture is not significantly changing such as bedrock landsliding on bare rock hillslopes.

Addressing these three sources of uncertainty - volume-area scaling uncertainty, landslide amalgamation and the under-detection of landslides - requires new approaches to obtain and analyse landslide inventories. In the last decade, the increasing availability of multi-temporal high resolution 3D point cloud data and digital elevation models (DEM), based on aerial or satellite photogrammetry and Light Detection and Ranging (LiDAR), has opened the possibility to better quantify landside volume and displacement (Bull et al., 2010; Mouyen et al., 2019; Okyay et al., 2019; Passalacqua et al., 2015; Ventura et al., 2011).

The most commonly used technique is the difference of DEM (DoD) which computes the vertical elevation differences between two DEMs of different time (Corsini et al., 2009; Giordan et al., 2013; Mora et al., 2018; Wheaton et al., 2010). Even though this method is fast and works properly on planar surface, a vertical difference can be prone to strong errors when used to quantify changes on vertical or very steep surfaces where landsliding typically occurs (e.g., Lague et al., 2013). The "Multiscale model-to-model cloud comparison" (M3C2) algorithm implemented by Lague et al. (2013) rather considers a direct 3D point cloud comparison. This algorithm has three main advantages over a DoD: (i) it operates directly on 3D point clouds, avoiding a phase of DEM creation that is conducive to a loss of resolution imposed by the cell size and potential data interpolation, (ii) it computes 3D distances along the normal direction of the topographic surface, allowing better capture of subtle changes on steep surfaces, and (iii) it computes a spatially variable confidence interval that accounts for surface roughness, point density and uncertainties in data registration. Applicable to any type of 3D data to measure the orthogonal distance between two point clouds, this approach has generally been used for terrestrial lidar and UAV photogrammetry over sub-kilometer scales. In the context of landsliding, it has been used to infer the displacement and volume of individual landslides, using point clouds obtained by UAV photogrammetry (e.g., Esposito et al., 2017; Stumpf et al., 2015), as well as for rockfall studies (Benjamin et al., 2020; Williams et al., 2018) and sediment tracking in post-wildfire conditions (DiBiase and Lamb, 2020). To our knowledge, systematic detection and segmentation of hundreds of landslides from 3D point clouds have not yet been attempted.

Here, we produce an inventory map of landslide topographic changes using a semi-automatic 3D point cloud differencing (3D-PcD) method based on M3C2 and applied to multi-temporal airborne LiDAR data. We use the generic term of "landslide" to define the spatially coherent changes detected by our method on hillslopes that result in at least several decimeter of erosion (i.e.,scars or sources) or deposition. This definition therefore includes all the types of mass wasting processes which involves the rapid displacement of soil, rocks and debris. The 3D landslide inventory is then compared to a traditional manually mapped

inventory of landslide scars based on aerial image comparison, hereafter called the 2D inventory. We apply our method to a complex topography located near Kaikoura, New Zealand, where the 2016 Mw 7.8 earthquake triggered nearly 30,000 landslides over a 10,000 km² area (Massey et al., 2020). A large part of these landslides have been mapped by Massey et al. (2020) providing a region based landslide distribution area and V-A scaling relationships. We choose a 5 km² area characterized by a high landslide spatial density along the Conway segment of the Hope fault, inactive during the earthquake, where pre- and post-earthquake LiDAR and aerial images were available (Fig. 1). This area has a variety of vegetation cover (e.g. forest, sparse or low vegetation, bare bedrock) and typically represents a challenge for conventional 2D landslide mapping. We illustrate the benefits of working directly on 3D data to generate landslide source and deposit inventories, and discuss the methodological advantages to operate directly on point clouds with M3C2 compared to DoD in terms of detection accuracy and error for total landslide volume.

The paper is organized as followed: first, the LiDAR dataset is presented followed by a detailed description of the 3D-PcD method. Second, results of the geomorphic change detection and identification of individual landslides in the studied area are presented. The remaining part of the paper focuses only on landslide sources. First, the comparison with conventional 2D landslide mapping is presented. Then, the statistical properties of the 3D and 2D landslide source inventories are investigated in terms of area and volume. Finally, current limitations of the method are discussed as well as knowledge gained on the importance of landslide under-detection on the co-seismic landslide inventory budget and landslide source geometry statistics.

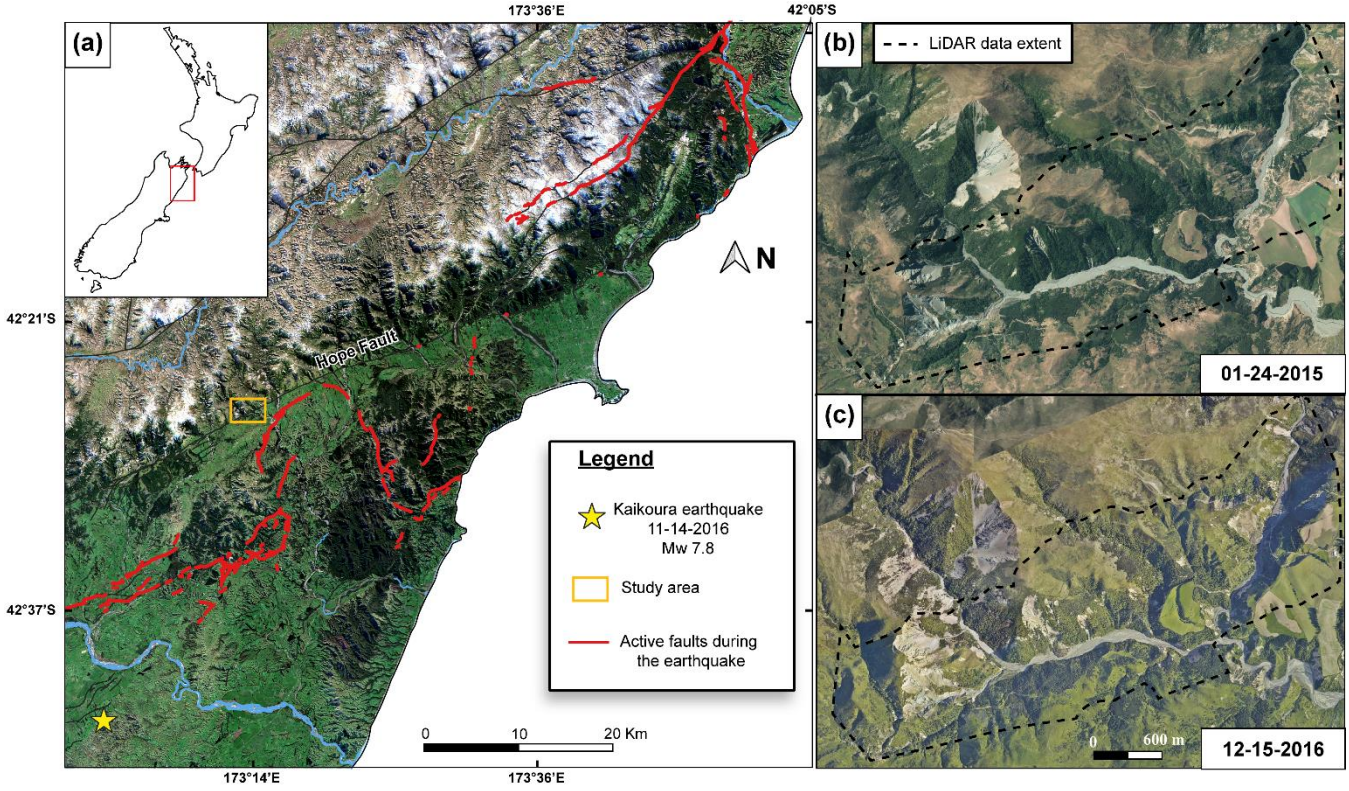

Figure 1: Maps of the regional context and location of the study area. (a) Regional map of Kaikoura with the location of the 2016 M_w 7.8 earthquake, associated active faults and the study area. (b-c) Orthoimages focused on the study area dated before and after the earthquake with the 5 km² LiDAR dataset extent used in this paper (all images are available at https://data.linz.govt.nz/set/4702-nz-aerial-imagery/, Aerial survey 2017).

2. Data description

In this study, we compare two 3D point clouds obtained from airborne LiDAR data collected before and after the November 14 2016 Kaikoura earthquake (Table. **Erreur ! Source du renvoi introuvable.**). Both airborne LiDAR surveys were acquired during summer. Pre-earthquake (pre-EQ) LiDAR data were collated over six flights performed from March 13, 2014 to March 20, 2014 for a resulting ground point density of 3.8 ± 2.1 pts/m². The vertical accuracy of this dataset has been estimated at 0.068 m to 0.165 m as the standard deviation of the difference between the elevation of GPS points located on highways and the nearest neighbour LiDAR shot elevation (Dolan and Rhodes, 2016). The post-earthquake (post-EQ) LiDAR survey took place rapidly after the earthquake from December 3, 2016 to January 6, 2017 for an average ground point density of 11.5 ± 6.8 pts/m². The vertical accuracy of this dataset has been estimated following the same protocol than the pre-earthquake LiDAR data with a mean of 0.00 m and a standard deviation of 0.04 m (Aerial survey, 2017). The difference in acquisition dates represents a period of 2 years and 8 months. For both LiDAR point clouds, only ground points defined by the data provider are selected (see details of the classification in Dolan, 2014). Manual quality control showed that some points located on vegetation remained in the pre-EQ data. As these points are located a few meters above the ground, they can lead to false landslide detection. We thus reprocessed this dataset to remove incorrectly classified points (details in section S1 in the supplements). Removed points represent 0.3% of the pre-EQ original point cloud.

In addition, orthoimages were used to perform a manual mapping of landslides to compare the detection of landslides from the 3D approach and a more classical approach. The pre-EQ orthoimage was obtained on January 24 2015 (available at https://data.linz.govt.nz/layer/52602-canterbury-03m-rural-aerial-photos-2014-2015) and the post-EQ one on December 15 2016. The resolutions are 0.3 and 0.2 m, respectively.

Table 1 : Information about LiDAR data used in this study

	Pre-earthquake LiDAR	Post-earthquake LiDAR
Date of acquisition	13/03/2014 – 20/03/2014	03/12/2016 – 06/01/2017
Commissioned by/provided by	USC-UCLA-GNS science/NCALM	Land Information New Zealand/AAM NZ
Availability	https://doi.org/10.5069/G9G44N75	On request at gisbasemap.ecan.govt.nz
Original point density (pts/m²)	9.02	19.2 ± 11.7

Number of ground points	10,660,089	63,729,096
Ground point density (pts/m²)	3.8 ± 2.1	11.5 ± 6.8
Vertical accuracy (m)	0.068 – 0.165	0.04 m
Study area (m²)	5,253,133	5,253,133

3. Methods and parameter choice

3.1. 3D point cloud differencing with M3C2

The method developed here to detect landslides consists of 3D point cloud differencing between two epochs using the M3C2 algorithm (Lague et al., 2013) available in the Cloudcompare software (Cloudcompare v2.11, 2020). This algorithm estimates orthogonal distances along the surface normal directly on 3D point clouds without the need for surface interpolation or gridding. While M3C2 can be applied on all points, the algorithm can use an accessory point cloud, called core points. In our case, core points constitute a regular grid with constant horizontal spacing generated by the rasterization of one of the two clouds. In the following, all the M3C2 calculations are done in 3D using the raw point clouds, but the results are "stored" on the core points. The use of a regular grid of core points has four advantages: (i) a regular sampling of the results allows computation of robust statistics of changes, unbiased by spatial variations in point density; (ii) it facilitates the volume calculation and the uncertainty assessment; (iii) it can be directly reused with 2D GIS as a raster (rather than a non-regular point cloud); and (iv) it speeds up calculations, although in the proposed workflow, computation time is not an issue and can be done on a regular laptop.

The first step of M3C2 consists in computing a 3D surface normal for each core point at a scale *D* (called the normal scale) by fitting a plane to the core points located within a radius of size D/2. Once the normal vectors are defined, the local distance between the two clouds is computed for each core point as the distance of the average positions of the two point clouds at a scale *d* (projection scale). This is done by defining a cylinder of radius d/2, oriented along the normal with a maximum length p_{max}. Distances are not computed if no intercept is found in the second point cloud. M3C2 also provides uncertainty on the computed distance at 95% of confidence based on local roughness, point density and registration error as follows:

$$LoD_{95\%}(d) = \pm 1.96\left(\sqrt{\frac{\sigma_1(d)^2}{n_1} + \frac{\sigma_2(d)^2}{n_2}} + reg\right) \quad (2)$$

where $LoD_{95\%}$ is the Level of Detection, $\sigma_1(d)$ and $\sigma_2(d)$ are the standard deviation of distances of each cloud, at scale *d*, measured along the normal direction, n_1 and n_2 are the number of points and *reg* is the co-registration error between the two epochs. *Reg* is assumed to be spatially uniform and isotropic in our case, but could be spatially variable and anisotropic (James et al., 2017). Because point density and surface roughness are spatially variable, $LoD_{95\%}$ is also spatially variable. For instance,

in forested steep hillslopes, points located under the canopy, with a lower point density, or vegetation points incorrectly filtered out creating locally high roughness, result into a higher $LoD_{95\%}$ and therefore require a larger topographic change to be detected as significant change. The M3C2 definition of the $LoD_{95\%}$ makes the conservative choice of adding the registration error to the combined standard error related to point cloud roughness, rather than taking the square root of the sum of squared standard error and squared registration. This choice is dictated by the simplistic hypothesis of using a spatially uniform and isotropic *reg*, while in reality *reg* varies spatially due to intra-survey registration errors of flight lines, and inter-survey rigid registration (Passalacqua et al., 2015). M3C2 has the option to compute the distance vertically which bypasses the normal calculation, and we use this option several times in the workflow. We use the abbreviation vertical-M3C2 in that case and 3D-M3C2 otherwise.

3.2. The Same Surface Different Sampling test

Following the approach proposed in Lague et al. (2013), we use a test based on using different sampling of the same natural surface to tune parameters of the workflow and apply the entire workflow. To this end, we create two randomly sub-sampled versions of the post-earthquake LiDAR data (which has the largest point density) with an average point density equal to the pre-EQ data. The resulting point clouds corresponds exactly to the same surface (i.e., *reg*=0), with roughness characteristics typical of the studied area, but with different point sampling. Applying the entire workflow to these two point clouds, allows to evaluate the prevalence of false landslide sources and deposits only related to the difference in sampling of a rough surface. We subsequently refer to this type of approach as a Same Surface Different Sampling test (SSDS test).

3.3. Parameter selection and 3D point cloud differencing performance

In this section, we explain how to select the appropriate normal scale *D* and projection scale *d* to detect landslides using M3C2. The normal scale *D* should be large enough to encompass enough points for a robust calculation, and smooth out small-scale point cloud roughness that results in normal orientation flickering and overestimation of the distance between surfaces (Lague et al., 2013). However, *D* should also be small enough to track the large-scale variations in hillslope geometry. By studying roughness properties of various natural surface, Lague et al. (2013) proposed that the ratio of the normal scale and the surface roughness, measured at the same scale, should be larger than about 25. We thus set *D* as the minimum scale for which a majority of core points verify this condition. As roughness is a scale and point density dependent measure, we explore a range for *D* from 2 m to 15 m for the pre-EQ dataset, which has the lowest point density (Fig.2.a). We found that $D \sim 10$ m represents a threshold scale below which the number of core points verifying this condition significantly drops.

The projection scale *d* should be chosen such that it is large enough to compute robust statistics using enough points, but small enough to avoid spatial smoothing of the distance measurement. Following Lague et al. (2013), M3C2 computes eq. (2) only if 5 points are included in the cylinder of radius *d/2* for each cloud. In our case, the pre-EQ data with the lowest point density will thus set the value of *d*. We use a SDDS test applying M3C2 with D=10 m and *d* varying from 1 to 40 m. Results show

that (Fig.2.b): (i) when it can be computed the $LoD_{95\%}$ actually predicts no significant change for at least 95 % of the time, indicating that the statistical model behind eq. (2) (Lague et al., 2013) is correct for this dataset; (ii) the fraction of core points for which the $LoD_{95\%}$ can be calculated rapidly increases between 1 and 8 m at which point it reaches 100 %. We choose d=5 m as it represents a good balance between the ability to compute a $LoD_{95\%}$ on most core points (here, ~ 97 %) and the smallest projection scale possible. To be able to generate M3C2 confidence intervals for as many points as possible, in particular on steep slopes below vegetation, we use a second pass of M3C2 with d=10 m using the core points for which no confidence interval was calculated at d=5 m. We note that d could theoretically be set as a function of the lowest mean point density of the two LiDAR datasets, *res*, by $d \sim 25/\pi res$. In our case the pre-EQ dataset has *res* = 3.8 pts/m² and would predict d = 1.3 m. However, the presence of vegetation significantly reduces the ground point density in some parts and the overlapping of flight lines creates localized high point density. Examining the mean ground point density of the entire dataset thus gives an incomplete picture of the strong spatial variations in point density. These changes in point density, critical to the correct evaluation of the $LoD_{95\%}$ (eq. (2)), are generally lost when working on raster of elevation (e.g., DEM).

The spacing of the core point grid should be smaller than half the projection scale d to ensure that all potential points are covered by at least one M3C2 measurement, while being larger than the typical point cloud spacing of the lowest resolution dataset. Because the ground point density on steep forested hillslope of the 2014 survey is of the order of 1 pt/m², we set a core point spacing of 1 m.

Finally, the maximum cylinder length p_{max} was set to 30 m as it encompassed the maximum change observed in the study area. This is generally obtained by trial and error. Setting p_{max} too large increases computation time significantly and may result in two different surfaces of the same point cloud being averaged (e.g., near very steep divides or in narrow gorges).

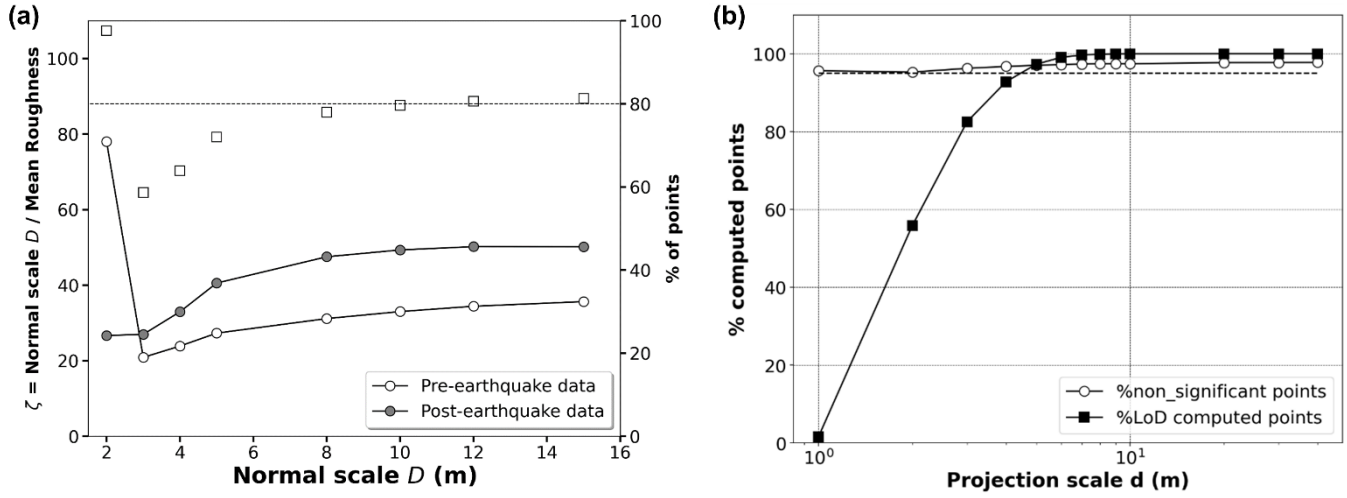

Figure 2 : Analysis of two main parameters of the M3C2 algorithm: the normal scale *D* and the projection scale *d*. (a) Ratio between normal scale and mean roughness for different normal scale values, and fraction of the pre-earthquake core points for which the normal scale is 25 times larger than the local roughness. (b) Percentage of computed points with a confidence interval of 95% versus projection scale d. The percentage of non-significant points is represented as well as the percentage of points where the Level of Detection ($LoD_{95\%}$) was computed (i.e., with at least 5 points on each point cloud).

3.3. 3D landslide mapping workflow and parameter selection

Our 3D landslide mapping workflow is divided in five main steps (Fig. 3).

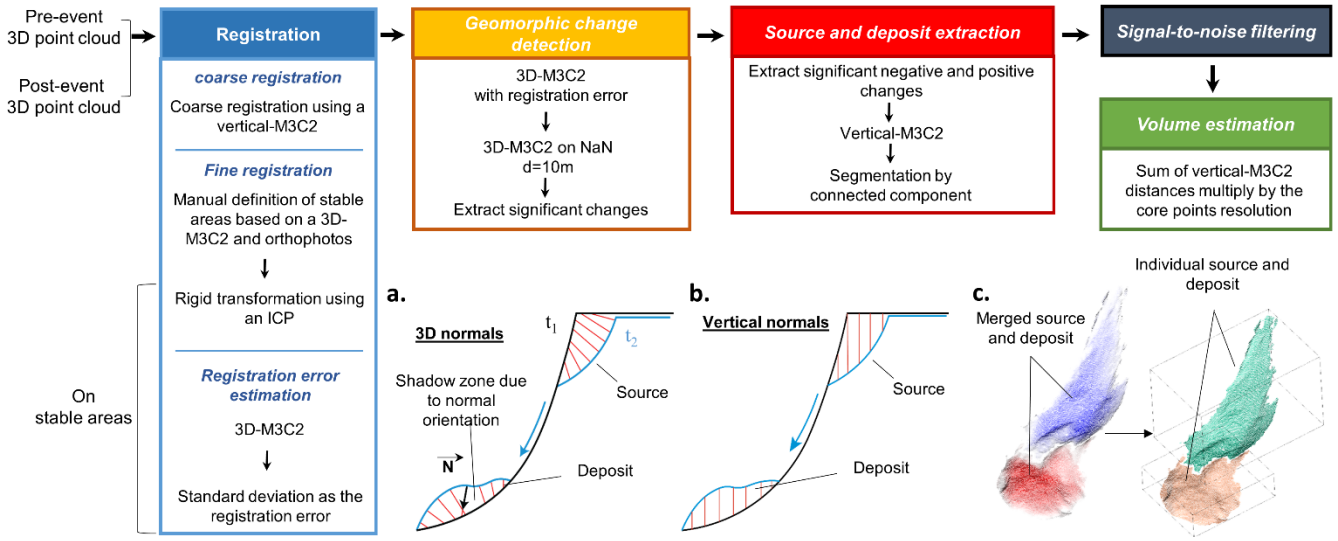

Figure 3: Workflow of the 3D point cloud differencing method for landslide detection and volume estimation with schematic representations of the different steps (a,b,c). (a) 3D measurement step with the shadow zone effect, where the red lines show the normal orientation. (b) Vertical-M3C2 step. (c) Segmentation by connected component. The resulting sources and deposits are individual point clouds illustrated in the figure by different colours.

3.3.1. Registration of the datasets and registration error estimate

To detect geomorphic changes and landslides, the two datasets need to be co-registered as closely as possible and any large-scale tectonic deformation need to be corrected. The registration error to be used in eq. (2) must also be estimated.

First, a preliminary quality control is performed to evaluate the internal registration quality of each dataset. This is feasible if the individual flight lines can be isolated, by using for instance, the pointID information specific to each line and provided in the las file format. The intra-survey registration quality can be investigated with 3D-M3C2 measurements of overlapping flight lines using a 1 m regular grid of core points, from which we define the registration bias and error as the mean and standard deviation of the 3D-M3C2 distances, respectively. The point cloud of the pre-EQ dataset results from 12 flight lines (Fig S2 in supplements), while the post-EQ corresponds to 5 flight lines. Result shows that for each dataset, no significant bias is measured between lines (maximum of 3 cm for the pre-EQ survey and 1 cm for the post-EQ survey; Tab. S3 in the supplements), but the registration error ranges from 13 to 20 cm for the pre-EQ survey, and is typically around 6 cm, with one instance at 12 cm for the post-EQ survey. Hence, the internal registration quality of the pre-EQ is significantly worse than the post-EQ dataset, a likely consequence of differences in instrument precision and post-processing methods.

Second, the registration between the two surveys must be evaluated, and in general improved. As delivered, the LiDAR datasets have a vertical shift between 1 and 2 m. To correct for this shift, a grid of core points is first created by rasterizing the dataset with the largest point density with a 1 m grid spacing. Then, a vertical-M3C2 calculation is performed and the mode of the

resulting distribution is used to adjust the two datasets by a vertical shift of 1.36 m. This approach is valid only when the fraction of the surface affected by landsliding is small. A subsequent 3D-M3C2 calculation is performed to obtain a preliminary map of geomorphic change. At this stage, a visual inspection of the pre-EQ and post-EQ orthoimages and of the preliminary 3D-M3C2 distances allow us to determine that there is no significant internal tectonic displacement. Then, we manually define areas deemed stable, 25 % of the studied area (Fig.4a), to perform a cloud matching registration. The stable areas were defined as surfaces (1) with a 3D-M3C2 distance smaller than 1 m, (2) with no visual texture changes in orthoimages, and (3) away from mass-wasting processes and forested area. Attention has been paid to select areas uniformly distributed in terms of location and slopes in the studied region to maximize the registration quality.

An Iterative Closest Point (ICP) algorithm (Besl and McKay, 1992) is then performed on the stable areas, and the obtained rigid transformation is applied to the entire post-earthquake point cloud to align it with the pre-earthquake one (Tab. S4). The mean 3D-M3C2 distance on stable areas is -0.01 m, showing that there is almost no bias left in the registration, and the standard deviation of 3D-M3C2 distances is 0.17 m (Fig. 4b). At this stage, the two datasets are considered optimally registered for the stable areas but with an unknown registration error *reg*. We propose to define *reg* as the maximum of the standard deviation of the intra-survey and inter-survey 3D-M3C2 distances. In the ideal case of two very high quality lidar datasets, *reg* would be equal to the inter-survey registration error. In the studied case, the pre-EQ intra-survey registration error is locally worse (0.2 m) than the inter-survey registration error (0.17 m). We thus set *reg*=0.2 m. Consequently, and according to eq. (2), with *reg*=0.2 m, our workflow cannot detect a 3D change that is smaller than 0.39 m in the ideal case of negligible roughness surface. At this stage, a 3D map of topographic change is available, but the significant geomorphic changes and individual landslides have not been isolated.

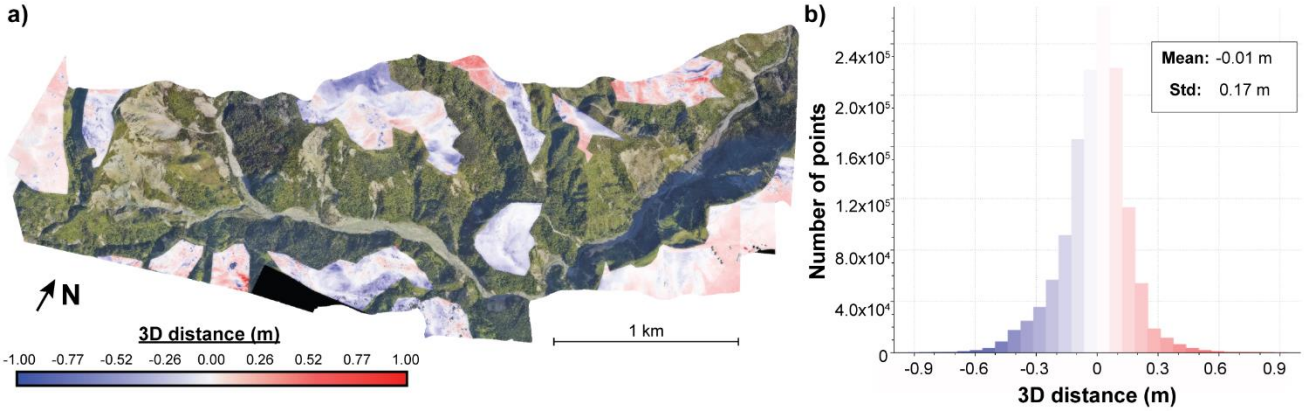

Figure 4: Map of 3D-M3C2 distances on stable areas error and the associated histogram. Map is a point cloud colored with the post-earthquake orthoimage (Aerial survey, 2017).

3.3.2. Geomorphic change detection

The registration error *reg* is then used in a first application of 3D-M3C2, using the pre-determined projection scale *d*=5 m, to estimate the spatially variable $LoD_{95\%}$ according to eq.(2). For problematic core points, where a confidence interval could not

be estimated due to insufficient points, a second application of 3D-M3C2 is performed at a larger projection scale d=10. These problematic core points generally correspond to ground points under canopy on steep slopes and represent typically 9.5 % of the entire area and up to 12% of steep slopes prone to landsliding. Significant geomorphic changes at the 95% confidence interval are then obtained by considering core points associated to a 3D-M3C2 distance larger than the $LoD_{95\%}$. Significant geomorphic changes can be associated to any geomorphic processes, including landsliding, but also fluvial erosion and deposition. Changes located in the river bed, and specifically related to river dynamics and not to landslide deposits, are manually removed using the post-EQ orthoimage. With the selection of the stable area, this is the only manual phase of the workflow.

3.3.3. Landslide source and deposit segmentation

Core points with negative and positive significant changes are first separated into tow point clouds of sources and deposits, respectively. A vertical-M3C2 is performed on each of these point clouds to estimate the volume of landslide sources and deposits (see section 3.3.5). As for any 2D landslide inventory, a critical component of the workflow is to segment each point cloud into individual landslide sources and areas. Segmenting complex patterns of erosion and deposition in 3D, with a very wide range of sizes is still a challenge. Here, for the sake of simplicity we use a classical clustering approach by a 3D label connected component algorithm (Lumia et al., 1983), available in CloudCompare (Fig.3.c). The point cloud is segmented into individual clusters based on two criteria: a minimum number of points *Np* defining a cluster and a minimum distance D_m below which neighbouring points, measured in a 3D euclidian sense, belong to the same cluster (Lumia et al., 1983). *Np*, or the minimum detectable landslide source or deposit area, was set to 20 m² to be consistent with the area of the projection cylinder used to average the point cloud position in the M3C2 distance calculation, $\pi(d/2)^2$ = 19.6 m² with d=5 *m*. D_m is an important parameter which, if chosen too large, will favour landslide amalgamation in identical clusters, and if is too small, in relation to the core point spacing, may over-segment landslides. In any case, D_m must be larger than the core point spacing. As there is no objective way to a priori choose D_m, we explore various values and choose D_m=2 m as an optimal value between landslide amalgamation and over-segmentation. The impact of D_m on the statistical distribution of landslide sources is addressed in the discussion.

We note that density based clustering algorithms based on DBSCAN (Martin Ester, Hans-Peter Kriegel, Jiirg Sander, 1996) have been used for 3D rockfall inventories segmentation (e.g., Benjamin et al., 2020; Tonini and Abellan, 2014). These algorithms separate dense clusters of points, considered as coherent topographic change, from areas of low point density, considered as noise. As shown in supplementary material (S5), density based clustering approaches do not yield a significantly better segmentation than a connected component. However, they have several drawbacks ranging from slow computation time, to less intuitive selection of parameters. We have therefore not used density based clustering in the following.

3.3.4. Definition of a confidence metric for each source and deposit: the SNR

Applying the previous workflow to an SSDS test results in 89 sources and deposits detected with an area between 20 and 37 m². This represents less than 1 % of the total surface, consistent with our definition of the $LoD_{95\%}$. However, in a context of very low landslide activity, artefacts (either negative or positive false detection) may represent a large fraction of the final inventory. Moreover, despite attempts to minimize the source of errors, false detection of sources and deposits could occur due for instance to classification errors of the LiDAR data or to the large M3C2 cylinder length p_{max} required to detect deep landslides intercepting different part of the same cloud in narrow gorge. To minimize these potential errors, that a label connected component cannot detect, we compute for each segmented deposit or source an equivalent signal-to-noise (SNR) ratio measured as the mean of the ratio between the 3D-M3C2 distance and the associated $LoD_{95\%}$ for each core point. The SNR can then be used to filter the landslide inventory. We propose to estimate an optimal SNR once the landslide mapping workflow is performed on the real dataset, by comparing it to the number of sources and deposits generated via a SSDS test. The optimal SNR is then chosen to minimize the number of artefacts compared to the number of real landslide.

3.3.5. Landslide area and volume estimation

While 3D normal computation is optimal to detect geomorphic changes, it is not suitable for volume estimation which requires to consider normals with parallel directions for a given landslide. Considering 3D normals can lead to "shadow zones", due to surface roughness, which would result in a biased volume estimate (Fig. 2a). Therefore, distances and in turn volumes are computed by using a vertical-M3C2 on a grid of core points corresponding to the significant changes (Fig.3.b). As the core points are regularly spaced by 1 m, the landslide volume is simply the sum of the vertical-M3C2 distances estimated from the individualized landslides. While the distance uncertainty predicted by the vertical-M3C2 could be used as the volume uncertainty, it significantly overpredicts the true distance uncertainty due to non-optimal normal orientation for the estimation of point cloud roughness on steep slopes (i.e., the roughness is not the detrended roughness). For each landslide source and deposit, we thus compute the volume uncertainty from the sum of the 3D-M3C2 uncertainty measured at each core point, not the vertical-M3C2 uncertainty. The volume uncertainty is specific to each landslide sources and deposits and depends on the local surface properties such as roughness, the number of points considered and the global registration error, but not on the volume itself. For each individual landslide source, the area A was obtained by computing the number of core points inside the source region. This represents the vertically projected area, to be consistent with the existing literature based on 2D studies of landslide statistics. The difference between planimetric area and true surface area (i.e., measured parallel to the surface) is addressed in the discussion.

3.4. Comparison with a manually mapped inventory based on orthoimagery

To estimate the potential in terms of landslide topographic change detection between the 3D-PcD method (3D inventory) and a traditional approach, we created a second inventory (2D inventory) by manually delineating landslide sources based on a

visual interpretation of the pre- and post-EQ orthoimages. The 3D data was not used in the process, and the mapmaker did not have a detailed knowledge of the 3D inventory. Deposits were not mapped. The 2D and 3D landslide source inventories are then compared in terms of number of landslides and intersection of mapped surfaces in planimetric view using GIS software. For areas only detected by manual mapping, we define 4 classes: (1) areas located on deposit zone detected by the 3D-PcD method, (2) areas under the $LoD_{95\%}$, (3) areas filtered by the minimum area of 20 m² and (4) areas filtered by the application of the SNR threshold. For areas only detected by the 3D-PcD method, we first distinguish landslide areas located on forested regions from those located in forest-free areas based on the number of laser returns of the post-EQ dataset (Fig. S9 and S10). We then separate the forest-free areas into 2 classes corresponding to typical textures found on the orthoimages, and delineate manually: (1) bare-rock areas and (2) sparse or low vegetation areas. We finally analyse the proportion of under-detected areas that are connected to a landslide source in the 2D inventory.

4. Results

4.1. Geomorphic change and 3D landslide sources and deposit inventory

The map of 3D-M3C2 distances (Fig. 5a) prior to statistically significant change analysis and segmentation provide a rare insight into topographic changes following large earthquakes. At first order, it highlights areas of coherent patterns of large (3D-M3C2 > 4 m) erosion (i.e. negative 3D distances) and deposition areas (i.e. positive 3D distances) located on hillslopes and corresponding to major landslides. Simple configurations with one major source area and a single deposit area can be easily recognized. A more complex pattern of intertwined landslides and rockfalls occur on a bare rock surface in the western part of the study area with a large variety of source sizes and aggregation of deposits. Most of the deposits are located on hillslopes while the deposits of three large landslides have reached the river and altered its geometry. At second order, a variety of patches of smaller amplitude (< 2 m) is visible on hillslopes. Erosion/deposition patterns in relation to fluvial activity can be documented on the river bed. The flight line mismatch, identified during the preliminary quality control, leads to low amplitude and long wavelength patterns of negative and positive 3D distances, notably visible on the central northern part of the studied area.

The area extent of significant changes, where the absolute amplitude of change is greater than $LoD_{95\%}$, represents 15 % of the study area (Fig 5.b). After the manual removal of changes in the fluvial domain related to fluvial processes, the minimum and maximum 3D-M3C2 distances on significant change areas are 0.40 ± 0.40 m and 29.97 ± 0.67 m, respectively (Fig.5.b). Due to surface roughness, the minimum $LoD_{95\%}$ observed is thus 0.40 m.

The point cloud of significant changes was segmented to identify the landslide sources (i.e. net erosion) and deposits (i.e. net deposition). During this step, clusters smaller than the detection limit of 20 m² are removed. They account for an area of 29,823 m², that is 4.2 % of the total area of significant change. Before application of a SNR threshold, the 3D inventory contained 1270 sources and 748 deposits. Applying the workflow to evaluate the optimal threshold SNR, we found that for a SNR = 1.5, the percentage of false detection in the inventory is expected to be minimum, equal to 2.1 % (Fig.6). Applying a SNR threshold

of 1.5 removes 746 sources and 442 deposits, characterized by a mean area and standard deviation of 88.6 ± 98 m² and 234 ± 198 m², respectively. Hence, the SNR filtering removed about 59 % in number of the sources and deposits, but only 17.6 % and 27.6 % of their total surface, respectively. This latter highlights the small individual area of the sources and deposits filtered. A large fraction of small patches in steep forested hillslopes are removed by the SNR filtering (Fig.5.b). The pre- and post-EQ raw point clouds of 20 patches with SNR < 1.5 were visually inspected in detail. It was difficult for a human expert to confirm a topographic change with high confidence owing to the very low point density and the rough topography of steep forested hillslopes On the contrary, the final filtered dataset shows coherent patterns of upslope erosion and downslope deposition in forested areas, indicative of landslides or large rockfalls. The orthoimagery cannot however help in confirming this due to the dense canopy cover and/or the pronounced projected shadows. The SNR filtering also removed large patches of very low amplitude topographic changes on open ground (< 0.5 m, either positive or negative), located in the pre-EQ data flight line mismatch, and did not correspond to a visible change in the orthoimagery (e.g., in the central-northern part of the studied area, Fig.5.b). While the SNR filtering removes more than half of the initial number of sources and deposits, these patches are of small size, close to the $LoD_{95\%}$ and only correspond to 5.96 % and 3.4 % of the total source and deposit volumes, respectively.

The final landslide inventory (3D inventory) contains a total of 524 sources and 304 deposits (Fig. 5c and 7a), with many sources sharing the same deposits at the toe of hillslopes. For sources, the mean absolute vertical-M3C2 distance is 2.69 m, the standard deviation 2.91 m and the maximum absolute value 23.06 ± 0.53 m. For deposits, the mean absolute vertical-M3C2 distance is 3.40 m, the standard deviation 3.68 m and the maximum absolute value is 27.9 ± 0.49 m. The area of detected landslides ranges from 20 to 40,679 m² for sources and from 20 to 28,037 m² for deposits, and the total source and deposit areas are 286,445 and 283,661 m², respectively. The resulting individual landslide volume ranges from 0.56 ± 41.5 m^3 to 169,890 ± 20,188 m^3 for source areas, with a total of 752,616 ± 154,165 m^3, and from 3.5 ± 19.5 m^3 to 151,706 ± 15,082 m^3 for deposits, with a total of 949,742 ± 150,014 m^3. The uncertainty on total volume estimation represents 20% for sources and 16% for deposits.

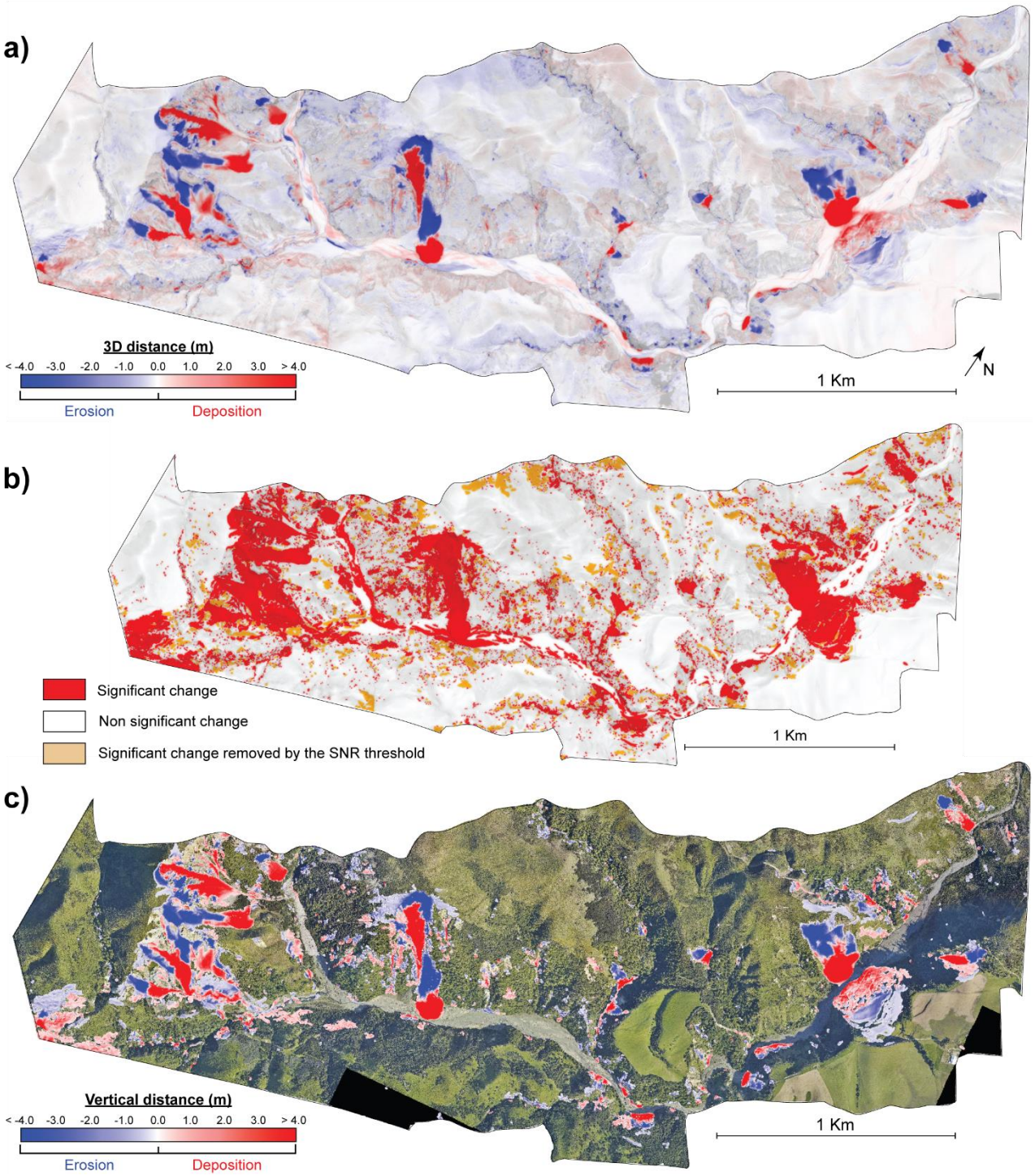

Figure 5: Map of the different steps of the workflow to generate the final landslide inventory. A) 3D-M3C2 distances from the geomorphic change detection step. B) Significant changes (>LoD95%) with indication of areas filtered by the SNR threshold. C)

Final map of vertical-M3C2 distances after application of a minimum area of 20 m² and a minimum SNR =1.5. Data is overlaid on the post-earthquake orthoimagery (12-15-2016, Aerial survey, 2017). Landslide sources are in blue and landslide deposits are in red.

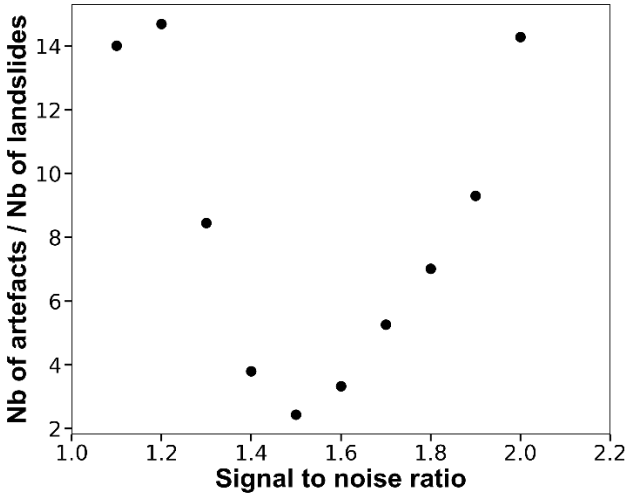

Figure 6: Percentage of the number of artefacts compared to the total number of landslides as a function of the signal-to-noise ratio.

4.2. 3D vs 2D landslide source inventory

In the following analysis we separate two types of errors: detection errors, corresponding to landslides present in only one of the 2D and 3D inventories, and delimitation errors, corresponding to differences in the planimetric contours of landslides. 258 landslide sources, called hereafter *2D-sources* as opposed to *3D-sources* derived from 3D-PcD, were mapped from visual inspection of pre-EQ and post-EQ orthoimages (Fig. 7b). The *2D-sources* represent a total area of 147,039 m² (Tab.2) with a minimum area of 7.2 m² and a maximum of 40,679 m². The minimum area detected shows that the resolution capability of the 2D inventory is finer than the 3D-PcD workflow. From the 258 *2D-sources*, 171 intersect *3D-sources* and 87 are not detected by the 3D-PcD method. These 87 *2D-sources* range from 14 m² to 599 m² with 63% smaller than 100 m². However, 22 are actual deposits in the 3D inventory, highlighting detection errors in the 2D inventory. These detection errors represent 16 % of the surface of the 2D inventory and are removed in the following, leading to 65 *2D-sources* not detected by the 3D-PcD method. The 3D-PcD method thus detects 72.4 % of the *2D-sources*, and 57.4 % of the total surface of the 2D inventory (Tab.2). The 65 *2D-sources* not detected by the 3D-PcD method correspond to 32 *2D-sources* located in areas with no statistically significant change (i.e., 3D-M3C2 distance < $LoD_{95\%}$), 29 *2D-sources* filtered by the SNR threshold and 4 *2D-sources* below the minimum detectable size of 20 m². In terms of planimetric surface however, the area not captured by 3D-PcD is overwhelming dominated by non-statistically significant change (40.9 % of the total *2D-sources* surface are < $LoD_{95\%}$), as opposed to SNR filtering (1.5%) or the minimum detectable size (0.2%). The surface of non-statistically significant change corresponds to delimitation errors located on the edges of sources and deposits, owing to the averaging effect of the M3C2 approach, and the transition between landslide sources and deposits (Fig.7c). The volume missed in *3D*-sources was computed

by using the intersection between the contours of the *2D-sources* not shared in 3D and the vertical M3C2 field of the core points. We find that the volume that would be missed in the 3D inventory is 0.6% of the total volume of *3D-sources*.

While 171 *2D-sources* are common to *3D-sources*, this corresponds to 144 *3D-sources* owing to the difference in landslide segmentation in the two inventories (Fig.7a). The 2D inventory misses 72 % (380) of the landslide sources detected in 3D, including landslides as large as 12,928 m² (blue polygon in the frame of figure 7a). The detection errors are predominantly in forest (222, 58% of detection errors) and bare rock (88, 23% of detection errors). They also occur on sparse or low vegetation (70, 18 % of detection errors) as is the case of two large landslides where pronounced shadows in the post-EQ, and where the topographic change is mostly vertical (Fig.7c). 75% of the total surface of *3D-sources*, is not detected in the 2D inventory, with 26% under forest, 34% on bare rock and 40% on sparse or low vegetation areas (Tab. 2 and Fig. S11). The landslide scars in this latter domain should be generally visible owing to strong spectral contrast between pre-EQ vegetation and post-EQ rock surface. However, the large source of error is explained by the incorrect delimitation of upslope topographic subsidence related to large scars (Fig.7c), as well as the under-detection of vertical subsidence related to translational and rotational landslides (Fig. 10b). These vertical movements of metric amplitude or less do not correspond to a clear change of orthoimagery texture, or create scarps that are too small or not easily detectable if they do not generate a shadow. Detection and delimitation errors contribute roughly equally to the 2D area under-detection (47 % detection error, 53 % delimitation error). The *2D-sources* misses 60 % of the total volume of the *3D-sources*. In contrast to the missed planimetric surface, the volume missed is predominantly on bare rock (32.8 %), nearly three times larger than in forest (13.9 %) or sparse/low vegetation (13.4 %).

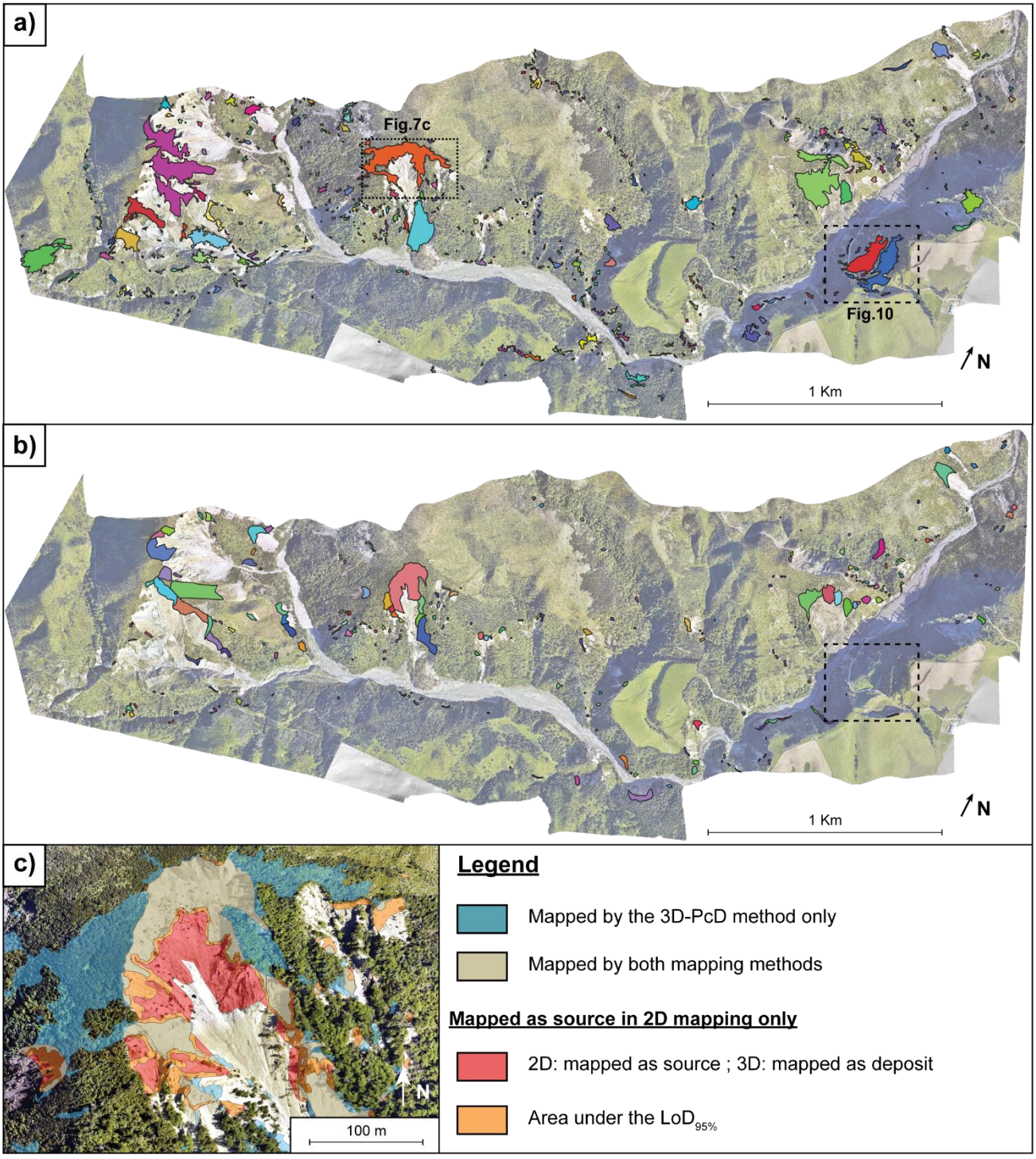

Figure 7: Comparison between landslide source inventories from a) the 3D point cloud detection workflow and b) a manual mapping based on 2D orthoimage comparison. Each landslide source is shown as a single colored polygon. c) Detailed comparison of typical

mapping differences between the 2D and 3D approach. Data are overlaid on the post-earthquake orthoimagery (12-15-2016, Aerial survey, 2017).

Table 2 : Summary of the comparison between 2D and 3D landslide source inventories. For the 2D inventory, the percentages are calculated with respect to the corrected total in which the 2D sources corresponding to 3D deposits are removed.

		Category	landslide sources	Area		Volume*	
				m²	%	m³	% of 3D total
2D inventory		Total	258	147,039		NA	
		On 3D deposit	*22*	*23,497*	16	NA	
		Corrected total	236	123,542	**100**	305,095	40.5
		Shared	171	70,936	57.4	300,249	39.9
	Not in 3D	< min area (20 m²)	4	295	0.2	179	0.6
		< SNR threshold	29	1848	1.5	1,790	
		< $LoD_{95\%}$	32	50,463	40.9	2,877	
3D inventory		Total	524	286,445	**100**	752,616	**100**
		Shared	144	70,936	24.7	300,249	39.9
	Not in 2D	Forest	222	56,288	19.7	104783	13.9
		Bare rock	88	73,191	25.5	246851	32.8
		Sparse or low vegetation	70	86,124	30.1	100731	13.4

***: volumes for the 2D inventory are computed from the vertical-M3C2 of core points located within the sources delimitations**

4.3. Landslide sources area, depth and volume analysis

The area distribution of landslide sources is computed as follow (Hovius et al., 1997; Malamud et al., 2004):

$$p(A) = \frac{1}{N_{LT}} \times \frac{\delta N_L}{\delta A} \quad (3)$$

where $p(A)$ is the probability density of a given area range within a landslide inventory, N_{LT} is the total number of landslides and A is the landslide source area. δN_{LT} corresponds to the number of landslides with areas between A and $A + \delta A$. The landslide area bin widths δA are equal in logarithmic space.

First, the area distribution of landslide sources inventories obeys a power-law scaling relationship consistent with previous studies (e.g., Hovius et al., 1997; Malamud et al., 2004). The exponents are respectively c = - 1.78 ± 0.07 and c = - 1.72 ± 0.04 for the 2D and 3D inventory, respectively (Fig. 8a). The landslide area distribution of the 2D inventory shows a cut-off around 100 m² and a roll-over, characteristic of landslide distributions, around 20 m² (Guzzetti et al., 2002; Malamud et al., 2004; Malamud and Turcotte, 1999; Tanyaş et al., 2019). The landslide area distribution of the 3D inventory does not exhibit a

rollover but a cut-off of the power-law behaviour around 40 m². This behaviour differs from the one observed from the landslide area distribution from Massey et al. (2020) in the broader Kaikoura region for which a cut-off appears around 1000 m² with a rollover at 100 m².

The volume distribution of the landslide sources of the 3D inventory was defined using equation (3), replacing A by the volume V, and also exhibits a typical negative power-law scaling (Fig.8b) of the form: $p(V) = dV^e$. The exponent of the power-law relationship is e = -1.61 ± 0.08. A roll-over is visible on the landslide volume distribution around 20 m³. Considering that the minimum $LoD_{95\%}$ observed in 3D is 0.40 m, and that the minimum landslide area is 20 m², the minimum volume that we can confidently measure should be 8 m³, a value close to the observed rollover. 10 landslides are smaller than 8 m³ in our inventory. They correspond to peculiar cases of very small landslides where negative 3D distances close to the $LoD_{95\%}$ are positive when measured vertically and thus reduce the apparent volume of eroded material.

With a direct measurement of landslide volume, it is possible to compute the volume-area relationship (eq. (1); Simonett, 1967; Larsen et al., 2010) and to compare it with previous results in New Zealand (Larsen et al., 2010, Massey et al., 2020). Here we determine V-A scaling coefficients using two methods: by fitting a linear model (1) on log-transformed data and (2) on averaged log-binned data. While the first method leads to a V-A relationship best describing the volume of each landslide, the second one is not affected by the varying number of landslides in each range of landslide area and leads to a V-A relationship that best matches the total landslide volume. Using the first approach, we find a volume-area scaling exponent of $\gamma = 1.16 \pm 0.02$ and an intercept $\log \alpha = -0.28 \pm 0.04$ m$^{0.7}$ with a determination coefficient $R^2 = 0.88$ (Fig.8.c). Using the second method, we find $\gamma = 1.17 \pm 0.02$, an intercept $\log \alpha = -0.26 \pm 0.08$ m$^{0.7}$ and a determination coefficient $R^2 = 0.99$. We also obtain a good correlation R² of 0.83 and 0.79 with the Larsen et al. (2010) relationships derived from soil landslides and from mixed soil landslides and bedrock landslides, respectively (Tab.333). R² of 0.85 is obtained when considering the parameters of the V-A relationships of the Kaikoura region, derived by Massey et al. (2020), including all their mapped landslides. At first order, the V-A relationships we obtained are thus consistent with previous studies. Yet, if the relationships from Larsen et al. (2010) and Massey et al. (2020) were applied to our landslide area inventory, the total volume would vary from 0.376x10^6 m^3 to 1.005x10^6 m^3 (Tab.3), compared to 0.753x10^6 ± 0.154x10^6 m^3 that we estimate directly. The closest evaluation of the total volume is based on the Massey et al. (2020) V-A relationship, that predicts a total volume of 0.653x10^6 m^3. The farthest evaluation of the total volume is the V-A relationship from Larsen et al. (2010) for all landslides (1.005x10^6 m^3), while their soil-dominated landslide relationship only predicts half of the total volume.

We presented the V-A relationship as it classically used in co-seismic volumes estimate from 2D inventories, however the volume being the product of mean depth and area, the V-A relationship hides an indirect correlation with area which may hinder subtle variations of depth with landslide size. Fig 8c shows that mean depth increases on average with area as a power law with an exponent 0.16±0.03 consistent with the V-A relationships we derived. However, the binned data suggests potentially two different trends: one for landslides smaller than 1000 m² in which, mean depth is nearly constant and exhibits a factor 4 of variation for given area (~ 0.7 m to ~ 3 m), while depth increases with area for sources larger than 1000 m².

However, the limited number of landslides in our inventory is insufficient to be conclusive on the existence of two different regimes.

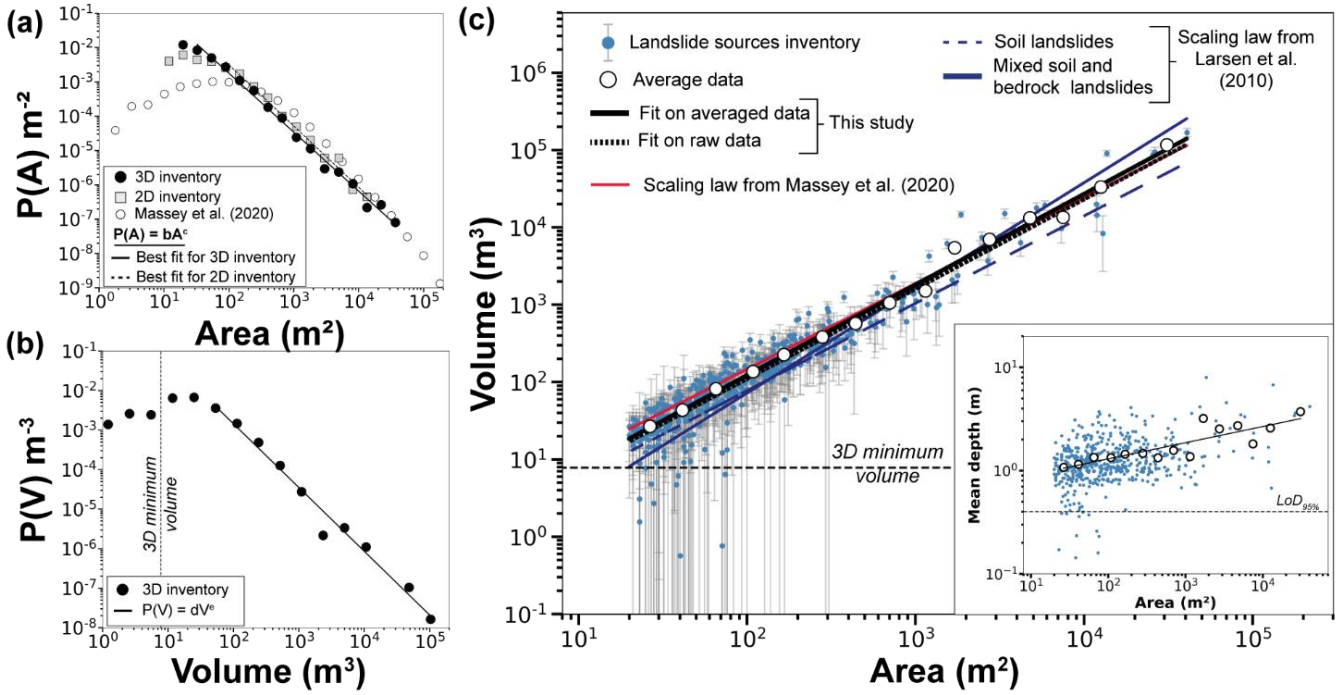

Figure 8: Landslide sources inventory analysis of the study area. (a) Landslide area distribution of both 3D and 2D inventory and Massey et al. (2020). (b) Landslide volume distribution of the 3D inventory. (c) Volume-area scaling relationships with uncertainty on volume and comparison with Larsen et al. (2010) and Massey et al. (2020) relationships obtained in New Zealand. The landslide mean depth vs area is also presented in inset. All scaling parameter values are summarized in Table 3.

Table 3 : Power-law scaling parameter values of the relations show in figure 8. Log α and γ are scaling parameters from the landslide area-volume relationship. Unit of α is [L(3-2γ)] with L in meters. Landslide source area and volume distribution coefficients are b and d while exponents are c and e respectively. The coefficient of determination R^2 is also given for each power-law fit function. The total volume refers to the application of the V-A relationship to the landslide areas of the 3D inventory.

	log b, log d or log α	c, e or γ	R^2	Total Volume (m^3)
Landslide area distribution (3D inventory)	0.70 ± 0.12	-1.72 ± 0.04	0.99	-
Landslide area distribution (2D inventory)	1.07 ± 0.20	-1.78 ± 0.07	0.99	-
Landslide volume distribution	0.39 ± 0.26	-1.61 ± 0.08	0.98	0.753×10^6 (direct measurement)
V-A relationship from averaged log-binned data (this study)	-0.26 ± 0.08	1.17 ± 0.02	0.99	0.643×10^6
V-A relationship from log-transformed data (this study)	-0.28 ± 0.04	1.16 ± 0.02	0.86	0.524×10^6
V-A relationship for soil landslides	-0.37 ± 0.06	1.13 ± 0.03	0.83	0.376×10^6

(Larsen et al., 2010)				
V-A relationship for mixed soil and bedrock landslide (Larsen et al., 2010)	-0.86 ± 0.05	1.36 ± 0.01	0.79	1.005×10^6
V-A relationship for the landslide inventory of Massey et al., 2020	-0.05 ± 0.02	1.109 ± 0.01	0.85	0.653×10^6

5. Discussion

The aim of this paper is to present a semi-automatic workflow, called 3D-PcD, for the detection and geometric characterization of landslide sources and deposits from repeated airborne LiDAR data. We specifically aim to overcome issues such as under-detection of landslides in inventories based on imagery analysis, landslide amalgamation and V-A relationship biases on total volume calculation. In the following, we discuss 1) the benefits and limits of the 3D-PcD method, 2) the benefits of 3D change detection to create landslide inventories, and 3) how 3D landslide inventories shed new light on the scaling properties of landslide sources.

5.1. 3D point cloud differencing and landslide detection

5.1.1 Vertical versus 3D change detection capability, and the M3C2 algorithm

The importance of detecting changes in 3D (3D-M3C2), as opposed to vertically (vertical-M3C2), in steep slopes can be illustrated by a SSDS test applied to the post-EQ point cloud (Fig. 9). Typical of change measurement methods on rough surfaces with random point sampling (e.g., Lague et al., 2013), a non-null mean distance is often measured, even though the two point clouds are samples of exactly the same surface. The distribution of measured distances is centred near zero, with a means of -2.10^{-4} and 1.10^{-4} m, for the vertical and 3D approach respectively. However, the 3D approach results in a standard deviation, σ=0.05 m, four times smaller than using a vertical differencing, σ= 0.20 m. The map of distance shows that vertical differencing systematically results in much larger distances on steep slopes than the 3D approach, while they both yield similar low distances on horizontal surfaces.

We thus find that the 3D-PcD method offers a greater sensitivity to detect changes compared to classical vertical DoD. This difference is particularly important as it propagates into a lower level of detection and uncertainty on volume calculations. Using the M3C2 algorithm in 3D (Lague et al., 2013) also offers the benefit of accounting for spatially variable point density and roughness in estimating a distance uncertainty for each core point, that can be subsequently used in volume uncertainty calculation. For instance, 3D-M3C2 reduces the sensibility of change detection in vegetated areas to a lower ground point density and potentially to a higher roughness due to vegetation misclassification. In turn, this advantage prevents in part the detection of false sources or deposits by using 3D-M3C2. By using a regular grid of core points as in Wagner et al. (2017), our workflow combines the benefits of working directly with the raw unorganized 3D data, as opposed to DoD where the relationship with the underlying higher point density data is lost. This approach also produces results with a regular sampling

that can easily be used for unbiased spatial statistics, volume calculation and easy integration into 2D GIS software. Compared to DoD, if an interpolation is needed, it is performed on the results rather than on the original DEM which can lead to uncontrolled error budget management.

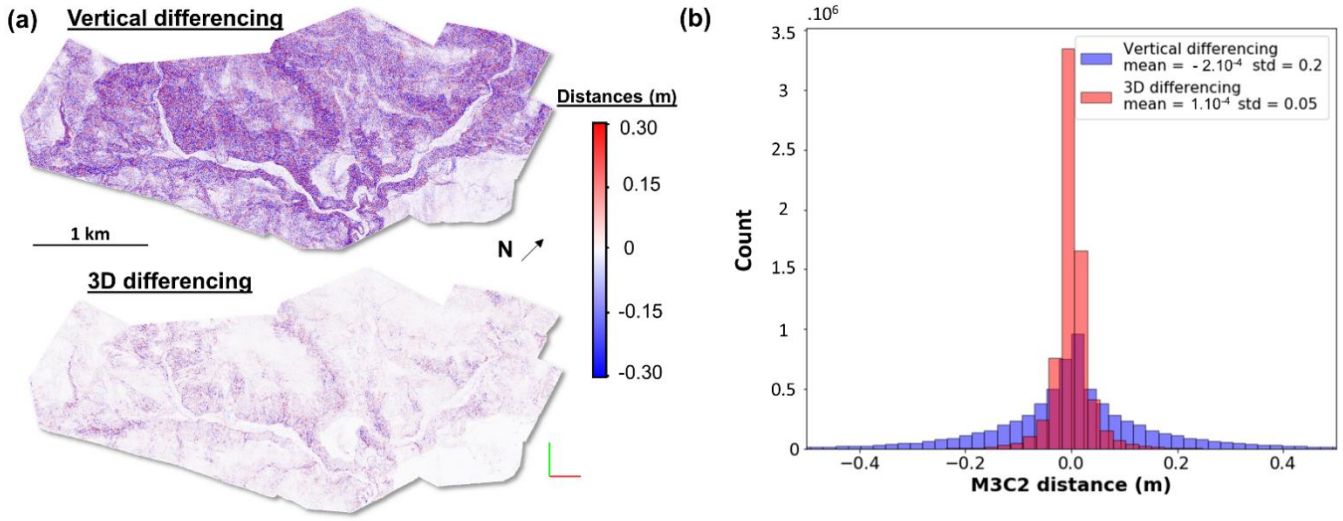

Figure 9: Comparison between vertical differencing (vertical-M3C2) and 3D differencing (3D-M3C2) on the post-EQ point cloud, sub-sampled randomly two times to generate two point clouds of the same surface with a different sampling (Same Surface Different Sampling test). A) Resulting change detection maps of the two different techniques. B) Histogram of the computed distances with the two techniques.

5.1.2. Current limits of the method

Registration: A critical aspect of the comparison of 3D point clouds is their co-registration, in particular in the context of co-seismic landsliding. In this study, a rigid transformation is applied to the entire datasets using an ICP algorithm (Besl and McKay, 1992), assuming that internal deformation induced by the earthquake is negligible. The 3D-M3C2 map does not exhibit any systematic horizontal shift either north or south of the Hope fault. We thus conclude that internal deformation, if any, was below the typical registration error in our study area. For larger studied regions with internal deformation and in the absence of a 3D co-seismic deformation model that could be applied to the post-EQ point cloud (e.g., Massey et al., 2020), our workflow should be applied in a piecewise manner with boundaries corresponding to the main identified faults or deformation zones. For landslide inventories following climatic events, the application to large datasets should be straightforward as no significant internal deformation is expected. Similarly, we also note internal flight line height mismatches of 0.13-0.20 m in the pre-EQ survey that are difficult to correct after data delivery and generate some apparent large scale low amplitude topographic changes (Fig. 4, Tab. S3 and section S2 in supplements). Interestingly, in the M3C2 calculation, flight line mismatches are averaged out in the distance measurement but leads to a higher local point cloud standard deviation, and thus to an increase of the $LoD_{95\%}$ and to a lower probability of incorrect topographic change detection. Despite some significant flight line mismatches in the pre-EQ dataset, using a SNR filtering approach efficiently removes the few false positive sources related to this issue. This highlights 1) the need for a detailed quality control (e.g., by applying M3C2 on overlapping lines) to

ensure the highest accuracy of the LiDAR data, 2) the importance of the statistical significance tests performed at the core point scale, and 3) the need for confidence metrics at the landslide scale, such as the SNR, to filter out a variety of potential false landslides. Ideally, a spatially variable model for point cloud errors and registration should be developed for each survey and combined into a more accurate and complete form of *LoD* than what the M3C2 approach currently offers (e.g., Glennie, 2008; Passalacqua et al., 2015). However, the position and attitude information of the sensor (e.g., Smoothed Best Estimate of Trajectory file) and raw LiDAR data area rarely available on LiDAR data repositories. Additionally, a dense network of ground control points is hard to get in mountainous environment. It is thus frequently impossible either to reprocess the LiDAR data to improve its quality (e.g.,Glennie et al., 2014), or to create a spatially variable registration and point cloud error model.

Landslide segmentation: The connected component segmentation is a simple, objective and rapid way to separate landslides in 3D that can be scaled up to much larger datasets. However, given the complexity of the 3D data, and in particular the very large range of landslide sizes (i.e., 4 orders of magnitude in the studied case), it inevitably exhibits some drawbacks and is subject to improvement. In particular, landslide amalgamation occurs, between two sources or deposits, if two of their core points are closer than D_m. Hence, landslides occurring on the two sides of a collapsed divide can be connected. This is the case for the largest landslide of our database located on the rock cliffs on the western part of the study area (Fig. 7a and Fig. 10a). In this example, the landslide source could reasonably be segmented in at least 5 smaller landslides. However, there does not seem to be a unique way to segment such a complex set of amalgamated events, even manually, underlining that landslide segmentation, can never be fully objective. We have explored the use of fast implementations of density based spatial clustering algorithm derived from DBSCAN (Martin Ester, Hans-Peter Kriegel, Jiirg Sander, 1996), an algorithm used for segmentation of 3D point clouds of rockfalls and removal of noisy points (e.g., Benjamin et al., 2020; Tonini and Abellan, 2014; details in section S5 in supplements). We applied OPTICS (Ankerst et al., 1999), recently used for rockfall segmentation of 3D lidar data (Carrea et al., 2021) and HDBSCAN (McInnes et al., 2017) which has a better ability to detect clusters of various sizes compared to DBSCAN, and is much faster. However, none managed to provide a significantly better segmentation of the largest landslides of our database, and the density probability of source area they produce is very similar to the one generated by a connected component. These approaches are however significantly longer to run than a connected component in Cloudcompare (S5 in supplementary material), and have parameters which are less intuitive to set than D_m, which is a distance directly comparable to core point spacing. New segmentation approaches accounting for normal direction, divide organization and 3D depth maps of amalgamated sources are needed to improve the segmentation of complex cases. We note however that segmentation issues do not affect the total landslide volume calculation in our study and that a sensitivity analysis of the impact of D_m shows that landslide source statistics are not severely affected by this parameter as long as it is close to the value we have used (see section 5.2).

Landslide volume calculation: Landslide volume is computed using a vertical-M3C2 on regular core points. This facilitates volume calculation on potentially complex 2D landslide geometry, but may lead to incorrect volume estimates on very steep slopes. Yet, the median slope distribution of the source core points (measured on the pre-EQ surface, Fig S12) is 34.6°, and only 0.74 % of the core points have slopes higher than 60°. We thus expect this effect to affect a very small fraction of our

inventory. Measuring landslide volume in 3D would be preferable, for instance along a constant surface normal direction defined for each source or deposits, but such simple approach is not better than a vertical measurement for the complex surface geometry of large landslides observed in the dataset and which are properly segmented (e.g., Fig. 7c). New approaches based on 3D mesh reconstruction have been used recently for rockfall volume estimation (Benjamin et al., 2020) and represent a future improvement of our workflow.

Landslide surface area: Another simplification of our approach is the calculation of planimetric surface areas, rather than true surface area. This choice was made to be consistent with previous results based on 2D inventories and to facilitate the comparison with our image based inventory. Measuring surface parallel area with 3D data would potentially help unravel new relationships between normal depth and area that are independent of topographic slope. Yet, this calculation is not trivial for complex landslide geometries in which assuming a unique normal orientation to get a surface parallel area measurement could result in a strong bias (e.g., Fig. 7c). Approaches based on 3D mesh surface calculation could help resolve this. Given that the median of the slope distribution of our landslide source inventory is 34.1°, a back of the envelope estimate of the true area gives a total landslide source area of 356,876 m² rather than 286,445 m².

Translational landslides: The 3D-PcD workflow we have designed is not designed for the measurement of translational landslides for which the dominant movement is parallel to the topographic surface. As in figure 10b, these landslides will appear as negative surface elevation in the source area and positive in the downslope accumulation area with little or non-significant 3D-M3C2 distance over much of the landslide body. These landslides can be detected with the 3D-PcD workflow, but the corresponding volume and area of sources may not be fully relevant to the bulk of the landslide inventory for which the source material has travelled further downslope exposing a large part of the slip surface. For mostly translational landslides, the surface parallel component of the deformation may be evaluated with feature tracking approaches as long as there are features to track (e.g., Aryal et al., 2012; Teza et al., 2007). The only elements that could be easily tracked in the 3D-PcD workflow are the barycenter of the source and associated deposit of each landslide, to explore runout dynamics, but we have not investigated this option yet.

Significant changes and geomorphic processes: While not a limitation per se, the 3D-PcD workflow detect changes, but cannot classify the nature of this change into various types of geomorphic processes. Given the current $LoD_{95\%}$ (i.e., > 0.40 m) only large topographic changes, correspond to landslide type processes on hillslopes and fluvial processes, are detected. Debris-flow processes could be detected, and may actually be part of the processes that remobilize landslide debris, however, they potentially create erosion in narrow steep channels that are likely below our spatial resolution capability, or will generate very small sources. They could however contribute to generate very large deposits. Fluvial processes are removed by the only manual operation performed in 3D-PcD, deemed necessary to preserve landslide deposits that have reached the river. While we do not have independent field constraints for all our detected sources, we are confident that the strict handling of error and SNR filtering approach filters out artefacts, and that our inventory of sources only contains landslide processes. Our approach does not directly resolve the typology of the landslides, including their failure mechanism (sliding, flow, fall), the failed materials (rock, soil, debris) and the velocity of the displacement (Hungr et al., 2014). Yet, combining the 3D-M3C2 distance

field with orthoimages (Fig. 10), we have identified the presence of rock avalanches, slumps (rotational failures), debris slides (translational failure) and we suspect the occurrence of some large rockfalls, although pre-EQ slopes steeper than 60° are extremely rare in the detected sources. We did not try to separate these as: (1) we were primarily interested in co-seismic volumes rather than detailed landslide mechanics which would have required field data; (2) there is no way to univocally identify, for the vast majority of our sources, the dominant landslide mechanism with either the 3D-M3C2 distance field and/or the orthoimages; (3) large landslides for which a dominant mechanism can be identified are too few in our inventory to draw a robust inference on scaling properties and geometry .

5.1.3. Landslide topographic change detection compared to manual passive imagery mapping

We presented for the first time a comparison between a classical handcrafted inventory of landslide sources from 2D orthoimagery comparison, and a 3D inventory based on LiDAR change detection where landslides are detected according to the topographic change they produce, not a change in optical passive imagery. Results show how different two landslide inventories of the same region, constructed from fundamentally different data sources (passive vs active remote sensing), can be. While the 3D inventory cannot be considered exhaustive, as it has a non-null $LoD_{95\%}$, it nonetheless detects roughly 3 times more landslides than the 2D imagery and a planimetric area affected by landsliding nearly two times larger. Most importantly, the 3D-PcD workflow knows its detection limit as one of its outcome is a spatially variable confidence interval ($LoD_{95\%}$) and confidence metrics (SNR) for each segmented source and deposit. While the resolution capability of 2D image analysis can be evaluated based on pixel size and is better than the LiDAR based approach in our study case, the detection capability is much more difficult to quantify, especially if the inventory is manually handcrafted.

Both detection and delimitation errors equally contribute to under-detection of the total area in the 2D inventory. They are, as expected (Zhong et al., 2020), frequent in areas with poor spectral contrasts between successive orthoimages such as bare rock surfaces or forests. But under-detection also occurs in sparse or low vegetation zones where some very large areas corresponding to vertical subsidence at the top of rotational or translational landslides were not detected (e.g., Fig. 10b) or incorrectly mapped (e.g., Fig. 7c). We note that under-detection in forest, while very significant (58 % of all detection errors), only corresponds to 19.7% of total area and 13.9% of the total volume. This is explained by the small size of missed landslides, as large ones strip out vegetation and are easily detectable. Hence, we do not expect that under-detection on forested area represent a large contribution in previously published co-seismic landslide volume estimates. However, it may limit the detection of subsidence area associated to new retrogressive slip planes (Fig 7c) which may prove important for subsequent landslide hazard management. Delimitation and detection errors are dominant on sparse and bare rock surfaces corresponding to 55.6% of the total landslide area. In particular, it is extremely difficult to map the transition between sources and deposits, especially on large and amalgamated landslides (e.g., Fig 7c). Here the ability of the 3D-PcD approach to not only detect sources but also deposits is essential. Our results thus indicate that existing landslide inventories, manually mapped from 2D images, may significantly suffer from under-detection of landslide area at least in regions dominated by sparse or absent vegetation cover.

We show that the main reason the 3D-PcD method did not detect surfaces mapped on the 2D inventory is that these surfaces are located in areas where the 3D-M3C2 distance is below the $LoD_{95\%}$. The detection limits of the 3D-PcD will improve in future years, by using the latest generation of LiDAR instrument generating dense (> 10 pts/m²) and more accurate 3D point clouds (< 5 cm Z error). With such data, the registration error could become of the order of 5 cm or less, further improving the detection capability of 3D-PCD both in terms of spatial resolution and $LoD_{95\%}$.

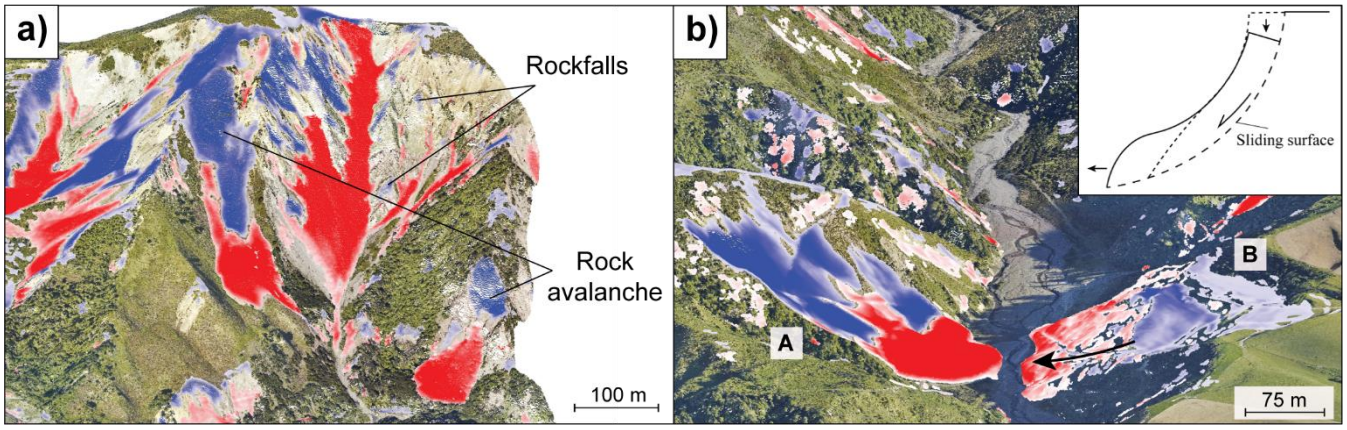

Figure 10: Two different points of interest of the 3D inventory illustrating various type of landslide mechanisms. a) Area mostly dominated by rock avalanche, where large rockfalls are also expected. b) Debris slide with mostly translational movement (A) and slump with likely rotational to translational displacement (B). The post-earthquake orthoimage is overlaid on the point cloud (December 15, 2016; Aerial survey, 2017)

5.1.4. Toward a limitation of amalgamation and under-detection biases on total landslide volume estimation

By a direct measurement from topographic data, the amalgamation effect is no longer an issue for total landslide volume estimation of an inventory even though our segmentation approach cannot resolve the amalgamation of individual landslides perfectly. Bypassing the use of a non-linear V-A relationship also avoid uncertainty inherent to the choice of the best suited scaling parameters. As we show, the total landslide volume vary significantly (from 0.376×10^6 m^3 to 1.005×10^6 m^3 ; Tab.3) depending on the V-A scaling relationship applied to our landslide inventory. For instance, we observe a difference of 18% in total volume estimation only due to the method used to fit data (i.e. on log-transformed or on averaged log-binned data). We also note that total landslide volume estimation from such relationship can get close to the volume estimated from the 3D-PcD for wrong reasons. As instance, applying our V-A relationship to both versions of the 2D inventory with and without deposit areas (Tab.2) lead to a difference of 17% in total volume. These results highlight the overarching sensitivity of the total volume of eroded material to the V-A relationship biases (Li et al., 2014; Marc and Hovius, 2015).

Our 3D-PcD approach also allows to estimate total landslide volume without the issue of under-detection of landslides. Due to the difference in the type of under-detection and delimitation errors between both 2D and 3D inventories, these issues do not propagate into total landslide volume estimate in similar ways. The area not detected by the 3D-PCD method in the 2D inventory represents only 0.6 % of the total volume. This is a negligible component owing to the fact that only very shallow

landslides, or shallow parts of very large landslides are missed. On the contrary, the area not detected by the 2D inventory represents 60 % of the total volume, highlighting the pronounced underestimation of total volume estimate if one uses image based detection followed by volume calculation. Most of this missed volume is due to the landslide delimitation errors on bare rock and sparse vegetation cover surfaces which represent 42% of the total volume while the under-detection of entire landslides only represents 18%. We also note that a third of the total volume is missed on bare rock surfaces. Our study area was chosen based on LiDAR data availability and contains a particularly high proportion of under-detected landslides in the 2D inventory due to the presence of actively eroding bare bedrock hillslopes. We expect this proportion to significantly vary when considering other landscapes with potentially varying proportions of vegetation cover, vegetation density and type (e.g. grass, shrubs, trees), lithology and ground shaking intensity. Nonetheless, our finding represents a first approach to the issue of considering the under-detection of landslides in total landslide volume estimates. We show that extreme caution should be put on co-seismic volumes estimated on landscapes where a large fraction of bare rock surfaces and sparse vegetation cover are present before earthquake, such as the Kaikoura ranges.

5.2. Landslide source scaling properties

The use of 3D data opens up a very large range of new geometric analysis of landslide sources and deposits. Here, we revisit traditional size-distributions and scaling relationships of landslide sources generated from 2D inventories. We systematically perform sensitivity analysis of these relations to the main parameters of the 3D-PcD workflow: the registration error *reg*, the minimum distance for segmentation D_m and the SNR threshold for removing landslides with limited confidence (See Fig.11, Appendix A, S13 and S14 in supplementary materials).

5.2.1. Total volume of landslide sources and deposits

Over the studied area of ~5 km^2, 524 landslide sources and 304 landslide deposits were detected with the 3D point cloud processing workflow. The scaling of pdf(V), with an exponent of -1.61 ± 0.08, indicates a slight tendency for the overall eroded volume to be dominated by the largest landslide (151,706 m^3, that is 20 % of the total volume). The uncertainty on total landslide volume, 16% to 20 % for deposits and sources, respectively, might appear large, as it is based on a conservative 95% confidence interval that we use throughout our analysis. These uncertainties are dominated by the registration error (*reg* = 0.2 m) and by the lower point cloud density of the pre-earthquake LiDAR data (Table.**Erreur ! Source du renvoi introuvable.**). Within these uncertainties, the total volume of sources (752,616 ± 154,165 m^3) and deposits (949,742 ± 150,014 m^3) are not statistically different. The larger volume of deposit is however consistent with rock decompaction during landsliding, which could be constrained by using a joint gravity survey in future studies (Mouyen et al., 2020). On top of this effect, we also expect an increased likelihood for sources to be more systematically filtered out than deposits as they are thinner (mean 3D-M3C2 distance are 2.69 m for sources and 3.4 m for deposits). This is consistent with the tendency for debris deposits, located on convergent parts of the landscape (e.g., hillslope hollows, debris flow channels), to collect different upslope sources in thicker patches than the initial individual sources (e;g., Fig. 10a)

5.2.2. Distribution of landslide source area and lack of rollover

We obtain a range of landslide area over 3 to 4 orders of magnitude (20 to 42,679 m²) which obey a clear power-law relationship for A > 40 m² with an exponent c = -1.72 ± 0.04 (Fig.8a). The negative power law behaviour for landslide area is generally observed for 2D landslide inventories, although only for source areas typically larger than 500- 5000 m² (Guzzetti et al., 2002; Malamud et al., 2004; Malamud and Turcotte, 1999; Medwedeff et al., 2020). Our exponent is roughly consistent with the exponents obtained over the entire Kaikoura coseismic landslide inventory of -1.88 (N_{LT} = 10,195; Massey et al., 2018) but differs significantly from the most recent estimate of -2.10 (N_{LT} = 29,557; Massey et al., 2020, Fig. 7a) for which the power-law scaling is expressed for A > 500 m². A sensitivity analysis of the impact of the workflow parameters (Fig. 11), in particular D_m which affects the level of amalgamation in the dataset, does not yield values of *c* smaller than -1.85 and cannot reconcile our results with Massey and co-authors (2020). Either, our limited study area overemphasizes, by chance, the occurrence of large landslides generating a smaller value of c, the manual inventory of Massey et al., 2020 may miss a large fraction of intermediate and small landslides, especially in bare rock hillslopes which are frequent in the high mountains of the Kaikoura range.

Most importantly, the landslide area distribution that we derive does not exhibit a rollover classically observed in 2D landslide inventories. Only a small deviation of the power-law behaviour appears for A < 40 m². Varying *reg* or D_m does not change this behaviour (Fig. 11a and 11b), nor using a density based clustering approach (Fig. S8). Increasing the SNR threshold to 2 (Fig. 11c), and thus censoring a larger number of small/intermediate landslides with small depth, increases the deviation from the power-law behaviour, but never creates a rollover. Hence, we are confident that our probability density of source area, generated by a purely objective and automatic approach, does not exhibit a rollover. If there is any, it would occur for sizes much smaller than 20 m².

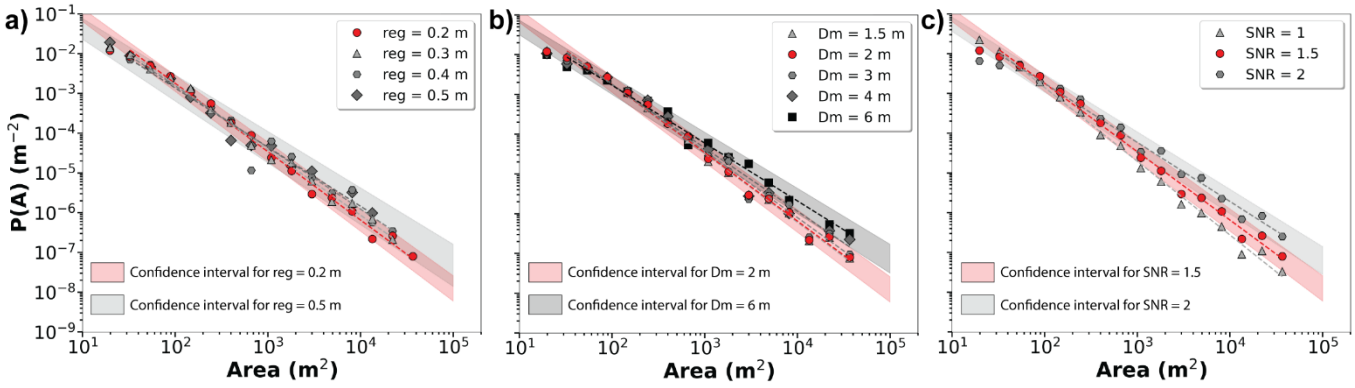

Figure 11: Landslide source area distributions for different (a) registration error *reg*, (b) minimum segmentation distance D_m and (c) signal-to-noise (*SNR*) values. All plots share the same y-axis. Value of the parameters used for this study are colored in red.

Several hypotheses, related to landslide mechanics or to landslide detection capabilities, have been put forward to explain the rollover behaviour for small landslide area. These include the transition to a cohesion dominated regime reducing the likelihood of rupture (Frattini and Crosta, 2013; Jeandet et al., 2019; Stark and Guzzetti, 2009), a cohesion gradient with depth (Frattini

and Crosta, 2013), landslide amalgamation (Tanyas et al., 2019) or the under-detection of small landslides (Hovius et al., 1997; Stark and Hovius, 2001). Our segmentation approach tends to amalgamate landslides rather than over-segment large ones and cannot explain the lack of rollover. On the contrary, this would create or accentuate a rollover by suppressing small landslides through amalgamation. The lack of rollover may also hint at a transition towards a different landsliding process, where rockfall dominates for instance. However, core points in sources with slopes > 60° represent only 0.74 % of the source area, pointing at an extremely limited contribution of rockfall processes originating from near-vertical cliffs.

To evaluate the degree of under-detection as a function of landslide size, we can leverage the two inventories we have created. For this, we compute a completeness ratio as the number of detected sources in 2D over the number detected in 3D, per range of source area. Fig.12 shows that the completeness ratio is around 0.25 for areas ~20-40 m² and systematically increases with landslide size up to 0.8-0.9 for sizes larger than 200-500 m². The behaviour above 500 m² suggests a slight increase of the completeness ratio which would asymptotically tend towards 1 for very large landslide source that cannot be missed both in 2D and 3D. However, we caution that our inventories contain too few landslides above 1000 m² (~ 20 each) to draw robust conclusions on this behaviour or derive a functional relationship. As some very shallow landslides detected in 2D are not detected in 3D, we cannot consider the 3D inventory as complete for small sizes and the true completeness ratio may actually be slightly overestimated at very small sizes. Yet, the 3D inventory is however far more complete than the 2D inventory. As such, our results demonstrate that in this study area, the rollover behaviour of the 2D inventory is caused by a size-dependent under-detection of small landslides below 500 m², existing even when using high resolution imagery with a better resolving capability than our 3D-PcD workflow (7 m² vs 20 m²) (Hovius et al., 1997; Stark and Hovius, 2001). This size-dependent under-detection of small landslides is expected to be systematically present in other image-based landslide inventories, even if carefully hand-crafted. Whether this effect systematically explains all the rollovers observed in past landslide inventories, or if other hypothesis such as a transition to cohesion dominated regime also contribute or are only expressed at even smaller scales remain to be explored. In any case, the number of landslides potentially missed in previous studies can be important given the level of under-detection that we report for small sizes. For instance, we note that in the first manual inventory of the Kaikoura EQ landslides (Massey et al., 2018), only 27 landslides were detected in our study area. The volume corresponding to under-detected small landslides may actually not matter in terms of total volume produced by earthquake derived landsliding. However, the presence or not of a rollover significantly matters in terms of hazards management (i.e. impact on the exposed population, infrastructure damage etc.) owing to the very large differences in the probability of small landslides.

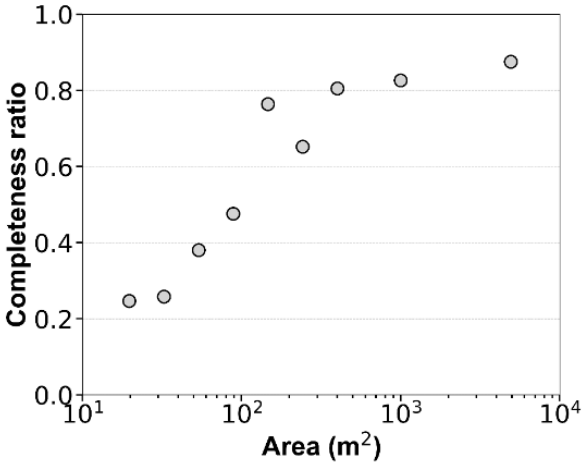

Figure 12: Number of 2D sources over number of 3D sources as a function of the source area. Assuming that the 3D inventory is nearly complete, this measure represents the ratio of completeness of the 2D inventory.

5.2.3. Distribution of landslide volume

We present here one of the first co-seismic landslide volume distribution derived directly from 3D topographic data (Fig.8b), rather than inferred from the combination of the landslide area distribution, based on 2D data, and an estimated V-A relationship. Our direct measurements show that the landslide volume distribution indeed obeys a power-law relationship for $V > 30$ m^3 with an exponent $e = -1.61 \pm 0.08$, consistent with the very broad range of exponents estimated in previous studies of $-1.0 \leq e \leq -1.9$ and $-1.5 \leq e \leq -1.9$ for rock and soil landslides, respectively (e.g., Brunetti et al., 2009; Malamud et al., 2004)). The sensitivity analysis to the workflow parameters (Fig. S13 and Appendix A), shows that the exponent e will decrease with *reg* and the *SNR* threshold as these parameters will censor progressively thinner landslides which are statistically the smallest ones. A *SNR* threshold of 2 strongly reduces the size of the inventory, in favour of larger landslides, and $e = -1.38$. While a SNR =1, results in $e = -1.69$ and a complete lack of a rollover above the minimum detectable volume. Contrary to the distribution of source area, the segmentation distance D_m has little impact on e.

The lack of a pronounced rollover above the minimum volume that we can theoretically detect (~ 8 m^3), makes the comparison with rockfall volume statistics relevant. The probability distribution of rockfall volume generally obeys a power-law relationship with an exponent e_R ranging from -1 to -2.2 (e.g., Malamud et al., 2004; Benjamin et al., 2020). If we restrict existing inventories to those having at least 500 rockfalls and the largest rockfall at least of 20 m^3, the range of exponent e_R narrows to -1.5/-2 with a majority of inventories around -1.6 ± 0.1 (Benjamin et al., 2020). Although, we do not expect rockfalls to be a dominant mechanism in our database given the lack of very steep slopes and given that rupture mechanisms (e.g., fragmentation, sliding, slumping…), rock heterogeneity and topographic constraints (e.g., hillslope size) are not expected to be similar (Dussauge et al., 2003), the consistency of the exponent we find is striking. This may suggest a much large range of scales over which the volume of landslides, encompassing rockfalls in this definition, obeys a unique scaling behavior. Dataset

specifically acquired to bridge the gap between large scale airborne lidar and terrestrial lidar are needed to get a better handle on the volume distribution of landslides, a critical information with respect to risk analysis and landslide erosion calculation.

5.2.4. Landslide depth and volume-area relationship

Our 3D-PcD approach opens the possibility to directly quantify the variations of landslide depth with size. We show that landslide mean depth does not vary for landslide area smaller than 1000 m². The same behaviour has been observed by Larsen et al. (2010) for soil failures suggesting that our landslide inventory may be relevant to shallow landslide. This is consistent with the fact that 50% of the landslide thicknesses are lower than 1.2 m and that the landslide volume-area (V-A) scaling relationships obtained in this study are close to that of Massey et al. (2020) and Larsen et al. (2010) for soil landslides. Moreover, for bedrock failures, Larsen et al. (2010) did observe an increase of landslide depth with size. Our landslide inventory may exhibit a slight increase of landslide depth for landslide area larger than 1000 m² that may hint at the transition from shallow to deeper bedrock landslide. However, the limited number of large landslides in our inventory does not allow to draw robust conclusions on this point.

The sensitivity analyses to the workflow parameters show that the V-A exponent γ is not significantly affected by the variations of the *reg* and SNR values we explored. Respectively, γ vary from 1.16 ± 0.03 to 1.19 ± 0.01 with the explored range of *reg* and varies from 1.17 ± 0.02 to 1.21 ± 0.02 with the SNR threshold (Fig S14, Appendix A). It is also not affected by the segmentation distance for $D_m < 4$ m, beyond which landslide amalgamation becomes significant and γ decreases to 1.1 for D_m=6 m.

6. Conclusion

In this paper, we introduce a new workflow for semi-automated landslide sources and deposit detection using 3D differencing based on high resolution topographic point cloud data. This method uses the M3C2 algorithm developed by Lague et al. (2013) for accurate change detection based on the 3D distance normal to the local surface. Landslide sources and deposits are segmented, using a 3D connected component approach, and their volumes are computed by a vertical-M3C2. Spatially variable uncertainties on distance and volume are provided by the calculation and used in the workflow to evaluate if a change is statistically significant or not, for volume uncertainty estimation and to define a confidence metric per source or deposit (Signal to Noise Ratio). The SNR is used to filter out potentially remaining artefacts. We provide various tests and recipes to estimate the registration error and to choose the parameters of the M3C2 algorithm as function of the point cloud density to ensure the lowest level of change detection, and the best resolution of the 3D map of change. Applied to a 5 km² area located in the Kaikoura region in New Zealand with pre- and post-earthquake LiDAR, we generate the first automatic inventory of landslide sources and deposits based on repeat 3D airborne LiDAR data. We show that:

- A minimum level of 3D change detection at 95% confidence of 0.40 m can be reached with airborne LiDAR data, which is largely set by the registration error. In our case, the limited quality of flight line alignment of the pre-EQ

data was the dominant source of registration uncertainty. Because it operates on raw data, M3C2 accounts for sub-pixel characteristics such as point density and roughness that are not accounted for when working on DEMs, and results in more robust statistics when it comes to evaluate if a change is significant or not. 3D point cloud differencing is critical on steep slopes and allows a lower level of change detection compared to the traditional DoD.

- Considering 3D topographic change for landslide detection removes the amalgamation effect on the total landslide volume by directly measuring it in 3D rather than considering an *ad hoc* V-A relationship. Amalgamation in 3D is still an issue when exploring individual landslide area and volume statistics given the simplistic segmentation approach that we have used. However, our approach has the benefits of more systematically capturing small landslides than traditional approaches based on 2D imagery with manual landslide mapping.
- Landslides on surfaces with low or no vegetation cover are classically missed with 2D imagery processing due to the lack of texture or spectral change. In our study case, 75 % of surface area was missed when considering a 2D inventory, corresponding to 60 % of the total volume determined with the 3D inventory. Missing area both correspond to detection error (landslide fully missed) and delimitation error (uncertainty in contours). Our method also shows the ability to detect subsidence related to slip failure propagation and the initiation or displacement of translational and rotational large landslides, which cannot be detected with 2D imagery.
- As this method provides direct 3D measurement, landslide geometry properties such as volume, area, depth and their distribution can be explored. Our results are broadly consistent with the V-A relationship scaling parameters determined by Larsen et al. (2010) for soil landslides and Massey et al. (2020), with a scaling exponent of 1.17.
- No rollover is observed in the landslide area distribution down to 20 m², our conservative resolution limit, using the 3D landslide inventory. However, we demonstrate, for the first time, a size-based under-detection in landslide mapped from repeat 2D images, which in turn results in a rollover of the landslide source area distribution for the 2D inventory. This result lends credit to the hypothesis that the rollover systematically observed in landslide area distributions generated from 2D images is entirely or partially related to an under-detection of small landslides (Stark and Hovius, 2001).

Our 3D processing workflow is a first step towards harnessing the full potential of repeated 3D high resolution topographic surveys to automatically create complete and accurate landslide inventories. However, high density LiDAR flights are not always available in landslide-prone regions for which a 2D image-based approach remains the most suited approach. Nevertheless, we recommend to systematically perform a 3D-PcD approach where repeat LiDAR data exist. This is critically needed to improve landslide science and managing the cascade of hazards following large earthquakes or storm events, by automatically identifying landslide deposits, and subtle features such as subsidence developing around landslides missed in 2D inventories. Current bottlenecks to apply this workflow over larger scales, beyond the availability of high-quality 3D data itself, are the registration of pre- and post-EQ data when complex co-seismic deformation patterns occur, and limitations of the segmentation method in high landslide density areas. While airborne LiDAR is best suited to vegetated environments and

currently results in the best precision compared to aerial or spatial photogrammetry, the workflow operates for any kind of 3D data.

840

Appendix A: Table of the result of the sensitivity analyses to the workflow parameters: *reg*, D_m and *SNR*. Unit for the registration error *reg* and the minimum segmentation distance D_m is in meter and *SNR* has no unit.

Workflow parameter	Value	N_{LT}	Landslide area distribution			Landside volume distribution			V-A relationship		
			Log b	c	R^2	Log d	e	R^2	Log α	γ	R^2
reg	**0.2***	**524**	**0.70 ± 0.12**	**-1.72 ± 0.04**	**0.99**	**0.39 ± 0.26**	**-1.61 ± 0.08**	**0.98**	**-0.26 ± 0.08**	**1.17 ± 0.02**	**0.99**
	0.3	329	0.50 ± 0.12	-1.64 ± 0.04	0.99	0.25 ± 0.16	-1.53 ± 0.05	0.99	-0.13 ± 0.10	1.16 ± 0.03	0.99
	0.4	200	0.28 ± 0.27	-1.55 ± 0.09	0.96	0.20 ± 0.22	-1.49 ± 0.06	0.98	-0.08 ± 0.06	1.17 ± 0.02	0.99
	0.5	234	0.11 ± 0.19	-1.49 ± 0.07	0.98	0.17 ± 0.21	-1.50 ± 0.17	0.99	-0.08 ± 0.04	1.19 ± 0.01	0.99
D_m	1.5	550	0.70 ± 0.12	-1.73 ± 0.04	0.99	0.40 ± 0.26	-1.62 ± 0.08	0.98	-0.13 ± 0.10	1.17 ± 0.02	0.99
	2*	**524**	**0.70 ± 0.12**	**-1.72 ± 0.04**	**0.99**	**0.39 ± 0.26**	**-1.61 ± 0.08**	**0.98**	**-0.26 ± 0.08**	**1.17 ± 0.02**	**0.99**
	3	443	0.61 ± 0.16	-1.68 ± 0.05	0.99	0.35 ± 0.17	-1.58 ± 0.05	0.99	-0.26 ± 0.04	1.17 ± 0.01	0.99
	4	373	0.42 ± 0.18	-1.59 ± 0.06	0.98	0.44 ± 0.23	-1.61 ± 0.06	0.99	-0.08 ± 0.06	1.14 ± 0.03	0.99
	6	259	0.13 ± 0.14	-1.45 ± 0.04	0.99	0.30 ± 0.23	-1.54 ± 0.07	0.98	-0.08 ± 0.04	1.10 ± 0.03	0.99
SNR	1	1270	0.86 ± 0.11	-1.85 ± 0.04	0.99	0.32 ± 0.24	-1.69 ± 0.07	0.99	-0.42 ± 0.06	1.21 ± 0.02	0.99
	1.5*	**524**	**0.69 ± 0.12**	**-1.72 ± 0.04**	**0.99**	**0.39 ± 0.26**	**-1.61 ± 0.08**	**0.98**	**-0.26 ± 0.06**	**1.17 ± 0.02**	**0.99**
	2	166	0.22 ± 0.14	-1.49 ± 0.04	0.99	-0.1 ± 0.26	-1.38 ± 0.08	0.98	-0.25 ± 0.09	1.19 ± 0.03	0.99

*: reference case used in the study.

Code availability

845 The code producing the landslide inventory in this study is available as a jupyter notebook at https://github.com/Thomas-Brd/3D_landslide_detection, and is also archived in Zenodo: http://doi.org/10.5281/zenodo.4010806.

Data availability

The landslide source and deposit information supporting the findings of this paper can be found in Zenodo: 10.5281/zenodo.4558905

850 **Supplement link**

The supplement related to this article is available online at: .

Author contribution

D.L. and T.B. designed the landslide detection workflow. T.B. and all co-authors participated to the discussion, writing and reviewing of this paper. T.B produced the figures and the code.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 803721), from the Agence Nationale de la Recherche (grant no. ANR-14-CE33-0005), from the LiDAR topobathymetric platform of the Rennes University and from the Brittany region (project LIDAREAU).

References

Aerial Surveys, Aerial photographs derived from two surveys of the study area carried out in 2014 to 2015 and in 2016 to 2017, by Aerial Surveys Ltd, 2017.

Ankerst, M., Breunig, M. M., Kriegel, H.-P. and Sander, J.: OPTICS, ACM SIGMOD Rec., 28(2), 49–60, doi:10.1145/304181.304187, 1999.

Aryal, A., Brooks, B. A., Reid, M. E., Bawden, G. W. and Pawlak, G. R.: Displacement fields from point cloud data: Application of particle imaging velocimetry to landslide geodesy, J. Geophys. Res. Earth Surf., 117(1), 1–15, doi:10.1029/2011JF002161, 2012.

Behling, R., Roessner, S., Kaufmann, H. and Kleinschmit, B.: Automated spatiotemporal landslide mapping over large areas using rapideye time series data, Remote Sens., 6(9), 8026–8055, doi:10.3390/rs6098026, 2014.

Benjamin, J., Rosser, N. J. and Brain, M. J.: Emergent characteristics of rockfall inventories captured at a regional scale, Earth Surf. Process. Landforms, 45(12), 2773–2787, doi:10.1002/esp.4929, 2020.

Besl, P. J. and McKay, N. D.: A Method for Registration of 3-D Shapes, IEEE Trans. Pattern Anal. Mach. Intell., 14(2), 239–256, doi:10.1109/34.121791, 1992.

Brunetti, M. T., Guzzetti, F. and Rossi, M.: Probability distributions of landslide volumes, Nonlinear Process. Geophys., 16(2), 179–188, doi:10.5194/npg-16-179-2009, 2009.

Bull, J. M., Miller, H., Gravley, D. M., Costello, D., Hikuroa, D. C. H. and Dix, J. K.: Assessing debris flows using LIDAR differencing: 18 May 2005 Matata event, New Zealand, Geomorphology, 124(1–2), 75–84,

doi:10.1016/j.geomorph.2010.08.011, 2010.

Carrea, D., Abellan, A., Derron, M., Gauvin, N. and Jaboyedoff, M.: MATLAB Virtual Toolbox for Retrospective Rockfall Source Detection and Volume Estimation Using 3D Point Clouds : A Case Study of a Subalpine Molasse Cliff, 2021. CloudCompare (version 2.11) [GPL software]. (2020). Retrieved from http://www.cloudcompare.org/

Corsini, A., Borgatti, L., Cervi, F., Dahne, A., Ronchetti, F. and Sterzai, P.: Estimating mass-wasting processes in active earth slides - Earth flows with time-series of High-Resolution DEMs from photogrammetry and airborne LiDAR, Nat. Hazards Earth Syst. Sci., 9(2), 433–439, doi:10.5194/nhess-9-433-2009, 2009.

DiBiase, R. A. and Lamb, M. P.: Dry sediment loading of headwater channels fuels post-wildfire debris flows in bedrock landscapes, Geology, 48(2), 189–193, doi:10.1130/G46847.1, 2020.

Dolan, J.F.: Data collection and processing report LiDAR survey of five fault segments (Estern Clarence, Western Clarence, Central Eastern Awatere, West Wairau and East Hope-Conway) of the Malborough fault system on the Northwestern portion of New Zealand's south island. Ph.D., University of southern California, 11 pp., 2014

Dolan J.F, Rhodes E.J.. Marlborough Fault System, South Island, New Zealand, airborne lidar. National Center for Airborne Laser Mapping (NCALM), distributed by OpenTopography. http://dx.doi.org/10.5069/G9G44N75, 2016.

Dussauge, C., Grasso, J.-R. and Helmstetter, A.: Statistical analysis of rockfall volume distributions: Implications for rockfall dynamics, J. Geophys. Res. Solid Earth, 108(B6), doi:10.1029/2001jb000650, 2003.

Esposito, G., Salvini, R., Matano, F., Sacchi, M., Danzi, M., Somma, R. and Troise, C.: Multitemporal monitoring of a coastal landslide through SfM-derived point cloud comparison, Photogramm. Rec., 32(160), 459–479, doi:10.1111/phor.12218, 2017.

Fan, Y., Clark, M., Lawrence, D. M., Swenson, S., Band, L. E., Brantley, S. L., Brooks, P. D., Dietrich, W. E., Flores, A., Grant, G., Kirchner, J. W., Mackay, D. S., McDonnell, J. J., Milly, P. C. D., Sullivan, P. L., Tague, C., Ajami, H., Chaney, N., Hartmann, A., Hazenberg, P., McNamara, J., Pelletier, J., Perket, J., Rouholahnejad-Freund, E., Wagener, T., Zeng, X., Beighley, E., Buzan, J., Huang, M., Livneh, B., Mohanty, B. P., Nijssen, B., Safeeq, M., Shen, C., van Verseveld, W., Volk, J. and Yamazaki, D.: Hillslope Hydrology in Global Change Research and Earth System Modeling, Water Resour. Res., 1737–1772, doi:10.1029/2018WR023903, 2019.

Frattini, P. and Crosta, G. B.: The role of material properties and landscape morphology on landslide size distributions, Earth Planet. Sci. Lett., 361, 310–319, doi:10.1016/j.epsl.2012.10.029, 2013.

Giordan, D., Allasia, P., Manconi, A., Baldo, M., Santangelo, M., Cardinali, M., Corazza, A., Albanese, V., Lollino, G. and Guzzetti, F.: Geomorphology Morphological and kinematic evolution of a large earth fl ow : The Montaguto landslide , southern Italy, Geomorphology, 187, 61–79, doi:10.1016/j.geomorph.2012.12.035, 2013.

Glennie, C.: Rigorous 3D error analysis of kinematic scanning LIDAR systems, J. Appl. Geod., 1(3), 147–157, doi:10.1515/jag.2007.017, 2008.

Glennie, C. L., Hinojosa-Corona, A., Nissen, E., Kusari, A., Oskin, M. E., Arrowsmith, J. R. and Borsa, A.: Optimization of legacy lidar data sets for measuring near-field earthquake displacements, Geophys. Res. Lett., 41(10), 3494–3501, doi:10.1002/2014GL059919, 2014.

Guzzetti, F., Malamud, B. D., Turcotte, D. L. and Reichenbach, P.: Power-law correlations of landslide areas in central Italy, Earth Planet. Sci. Lett., 195(3–4), 169–183, doi:10.1016/S0012-821X(01)00589-1, 2002.

Guzzetti, F., Mondini, A. C., Cardinali, M., Fiorucci, F., Santangelo, M. and Chang, K. T.: Landslide inventory maps: New tools for an old problem, Earth-Science Rev., 112(1–2), 42–66, doi:10.1016/j.earscirev.2012.02.001, 2012.

Hovius, N., Stark, C. P. and Allen, P. A.: Sediment flux from a mountain belt derived by landslide mapping, Geology, 25(3), 231, doi:10.1130/0091-7613(1997)025<0231:SFFAMB>2.3.CO;2, 1997.

Hungr, O., Leroueil, S. and Picarelli, L.: The Varnes classification of landslide types, an update, Landslides, 11(2), 167–194, doi:10.1007/s10346-013-0436-y, 2014.

James, M. R., Robson, S. and Smith, M. W.: 3-D uncertainty-based topographic change detection with structure-from-motion photogrammetry: precision maps for ground control and directly georeferenced surveys, Earth Surf. Process. Landforms, 42(12), 1769–1788, doi:10.1002/esp.4125, 2017.

Jeandet, L., Steer, P., Lague, D. and Davy, P.: Coulomb Mechanics and Relief Constraints Explain Landslide Size Distribution, Geophys. Res. Lett., 46(8), 4258–4266, doi:10.1029/2019GL082351, 2019.

Keefer, D. K.: The importance of earthquake-induced landslides to long-term slope erosion and slope-failure hazards in seismically active regions, Geomorphology, 10(1–4), 265–284, doi:10.1016/0169-555X(94)90021-3, 1994.

Lague, D., Brodu, N. and Leroux, J.: Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (N-Z), ISPRS J. Photogramm. Remote Sens., 82, 10–26, doi:10.1016/j.isprsjprs.2013.04.009, 2013.

Larsen, I. J., Montgomery, D. R. and Korup, O.: Landslide erosion controlled by hillslope material, Nat. Geosci., 3(4), 247–251, doi:10.1038/ngeo776, 2010a.

Larsen, I. J., Montgomery, D. R. and Korup, O.: Landslide erosion controlled by hillslope material, Nat. Geosci., 3(4), 247–251, doi:10.1038/ngeo776, 2010b.

Li, G., West, A. J., Densmore, A. L., Jin, Z., Parker, R. N. and Hilton, R. G.: Seismic mountain building: Landslides associated with the 2008 Wenchuan earthquake in the context of a generalized model for earthquake volume balance, Geochemistry, Geophys. Geosystems, 15(4), 833–844, doi:10.1002/2013GC005067, 2014.

Lumia, R., Shapiro, L. and Zuniga, O.: A new connected components algorithm for virtual memory computers, Comput. Vision, Graph. Image Process., 22(2), 287–300, doi:10.1016/0734-189X(83)90071-3, 1983.

Malamud, B. D. and Turcotte, D. L.: Self-organized criticality applied to natural hazards, Nat. Hazards, 20(2–3), 93–116, doi:10.1023/A:1008014000515, 1999.

Malamud, B. D., Turcotte, D. L., Guzzetti, F. and Reichenbach, P.: Landslide inventories and their statistical properties, Earth Surf. Process. Landforms, 29(6), 687–711, doi:10.1002/esp.1064, 2004.

Marc, O. and Hovius, N.: Amalgamation in landslide maps: Effects and automatic detection, Nat. Hazards Earth Syst. Sci., 15(4), 723–733, doi:10.5194/nhess-15-723-2015, 2015.

Marc, O., Hovius, N. and Meunier, P.: The mass balance of earthquakes and earthquake sequences, Geophys. Res. Lett., 43(8),

3708–3716, doi:10.1002/2016GL068333, 2016.

Marc, O., Behling, R., Andermann, C., Turowski, J. M., Illien, L., Roessner, S. and Hovius, N.: Long-term erosion of the Nepal Himalayas by bedrock landsliding: The role of monsoons, earthquakes and giant landslides, Earth Surf. Dyn., 7(1), 107–128, doi:10.5194/esurf-7-107-2019, 2019.

Martha, T. R., Kerle, N., Jetten, V., van Westen, C. J. and Kumar, K. V.: Characterising spectral, spatial and morphometric properties of landslides for semi-automatic detection using object-oriented methods, Geomorphology, 116(1–2), 24–36, doi:10.1016/j.geomorph.2009.10.004, 2010.

Martin Ester, Hans-Peter Kriegel, Jiirg Sander, X. X.: A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, Compr. Chemom., 96(34), 226–231 [online] Available from: https://www.aaai.org/Papers/KDD/1996/KDD96-037.pdf?source=post_page, 1996.

Massey, C., Townsend, D., Rathje, E., Allstadt, K. E., Lukovic, B., Kaneko, Y., Bradley, B., Wartman, J., Jibson, R. W., Petley, D. N., Horspool, N., Hamling, I., Carey, J., Cox, S., Davidson, J., Dellow, S., Godt, J. W., Holden, C., Jones, K., Kaiser, A., Little, M., Lyndsell, B., McColl, S., Morgenstern, R., Rengers, F. K., Rhoades, D., Rosser, B., Strong, D., Singeisen, C. and Villeneuve, M.: Landslides triggered by the 14 November 2016 Mw 7.8 Kaikōura earthquake, New Zealand, Bull. Seismol. Soc. Am., 108(3B), 1630–1648, doi:10.1785/0120170305, 2018.

Massey, C. I., Townsend, D., Jones, K., Lukovic, B., Rhoades, D., Morgenstern, R., Rosser, B., Ries, W., Howarth, J., Hamling, I., Petley, D., Clark, M., Wartman, J., Litchfield, N. and Olsen, M.: Volume characteristics of landslides triggered by the M W 7.8 2016 Kaikōura Earthquake, New Zealand, derived from digital surface difference modelling , J. Geophys. Res. Earth Surf., 0–3, doi:10.1029/2019jf005163, 2020.

McInnes, L., Healy, J. and Astels, S.: hdbscan: Hierarchical density based clustering, J. Open Source Softw., 2(11), 205, doi:10.21105/joss.00205, 2017.

Medwedeff, W. G., Clark, M. K., Zekkos, D. and West, A. J.: Characteristic landslide distributions: An investigation of landscape controls on landslide size, Earth Planet. Sci. Lett., 539, 116203, doi:10.1016/j.epsl.2020.116203, 2020.

Mora, O. E., Gabriela Lenzano, M., Toth, C. K., Grejner-Brzezinska, D. A. and Fayne, J. V.: Landslide change detection based on Multi-Temporal airborne LIDAR-derived DEMs, Geosci., 8(1), 6–8, doi:10.3390/geosciences8010023, 2018.

Mouyen, M., Steer, P., Chang, K.-J., Le Moigne, N., Hwang, C., Hsieh, W.-C., Jeandet, L., Longuevergne, L., Cheng, C.-C., Boy, J.-P. and Masson, F.: Quantifying sediment mass redistribution from joint time-lapse gravimetry and photogrammetry surveys, Earth Surf. Dyn., 8(2), 555–577, doi:10.5194/esurf-8-555-2020, 2020.

Parker, R. N., Densmore, A. L., Rosser, N. J., De Michele, M., Li, Y., Huang, R., Whadcoat, S. and Petley, D. N.: Mass wasting triggered by the 2008 Wenchuan earthquake is greater than orogenic growth, Nat. Geosci., 4(7), 449–452, doi:10.1038/ngeo1154, 2011.

Passalacqua, P., Belmont, P., Staley, D. M., Simley, J. D., Arrowsmith, J. R., Bode, C. A., Crosby, C., DeLong, S. B., Glenn, N. F., Kelly, S. A., Lague, D., Sangireddy, H., Schaffrath, K., Tarboton, D. G., Wasklewicz, T. and Wheaton, J. M.: Analyzing high resolution topography for advancing the understanding of mass and energy transfer through landscapes: A review, Earth-

Science Rev., 148, 174–193, doi:10.1016/j.earscirev.2015.05.012, 2015.

Pradhan, B., Jebur, M. N., Shafri, H. Z. M. and Tehrany, M. S.: Data fusion technique using wavelet transform and taguchi methods for automatic landslide detection from airborne laser scanning data and quickbird satellite imagery, IEEE Trans. Geosci. Remote Sens., 54(3), 1610–1622, doi:10.1109/TGRS.2015.2484325, 2016.

Stark, C. P. and Guzzetti, F.: Landslide rupture and the probability distribution of mobilized debris volumes, J. Geophys. Res. Earth Surf., 114(2), 1–16, doi:10.1029/2008JF001008, 2009.

Stark, C. P. and Hovius, N.: The characterization of landslide size distributions, Geophys. Res. Lett., 28(6), 1091–1094, doi:10.1029/2000GL008527, 2001.

Stumpf, A., Malet, J. P., Allemand, P., Pierrot-Deseilligny, M. and Skupinski, G.: Ground-based multi-view photogrammetry for the monitoring of landslide deformation and erosion, Geomorphology, 231(October), 130–145, doi:10.1016/j.geomorph.2014.10.039, 2015.

Tanyaş, H., van Westen, C. J., Allstadt, K. E. and Jibson, R. W.: Factors controlling landslide frequency–area distributions, Earth Surf. Process. Landforms, 44(4), 900–917, doi:10.1002/esp.4543, 2019.

Teza, G., Galgaro, A., Zaltron, N. and Genevois, R.: Terrestrial laser scanner to detect landslide displacement fields: A new approach, Int. J. Remote Sens., 28(16), 3425–3446, doi:10.1080/01431160601024234, 2007.

Tonini, M. and Abellan, A.: Rockfall detection from terrestrial lidar point clouds: A clustering approach using R, J. Spat. Inf. Sci., 8(1), 95–110, doi:10.5311/JOSIS.2014.8.123, 2014.

Ventura, G., Vilardo, G., Terranova, C. and Sessa, E. B.: Tracking and evolution of complex active landslides by multi-temporal airborne LiDAR data: The Montaguto landslide (Southern Italy), Remote Sens. Environ., 115(12), 3237–3248, doi:10.1016/j.rse.2011.07.007, 2011.

Wagner, W., Lague, D., Mohrig, D., Passalacqua, P., Shaw, J. and Moffett, K.: Elevation change and stability on a prograding delta, Geophys. Res. Lett., 44(4), 1786–1794, doi:10.1002/2016GL072070, 2017.

Wheaton, J. M., Brasington, J., Darby, S. E. and Sear, D. A.: Accounting for uncertainty in DEMs from repeat topographic surveys: Improved sediment budgets, Earth Surf. Process. Landforms, 35(2), 136–156, doi:10.1002/esp.1886, 2010.

Williams, J. G., Rosser, N. J., Hardy, R. J., Brain, M. J. and Afana, A. A.: Optimising 4-D surface change detection: An approach for capturing rockfall magnitude-frequency, Earth Surf. Dyn., 6(1), 101–119, doi:10.5194/esurf-6-101-2018, 2018.

Zhong, C., Liu, Y., Gao, P., Chen, W., Li, H., Hou, Y., Nuremanguli, T. and Ma, H.: Landslide mapping with remote sensing: challenges and opportunities, Int. J. Remote Sens., 41(4), 1555–1581, doi:10.1080/01431161.2019.1672904, 2020.