Supplement of Earth Surf. Dynam., 9, 1013–1044, 2021
https://doi.org/10.5194/esurf-9-1013-2021-supplement

Earth Surface Dynamics
Open Access
EGU

Supplement of

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

Thomas G. Bernard et al.

Correspondence to: Thomas G. Bernard (thomas.bernard@univ-rennes1.fr)

Section S1. Treatment of misclassified points

Remaining vegetation points in the ground-classified pre-EQ point cloud can be identified as (1) they are locally high above the surrounding ground and (2) it concerns generally one or a patch of points. Thus they are mainly characterized by a high local roughness and a low point density. To remove suspicious remaining vegetation points, we first reconstructed a mesh from the ground-classified pre-EQ point cloud using the Poisson surface reconstruction algorithm developed by Kazhdan et al. (2006). We then removed outlier points of the pre-EQ point cloud by first computing the cloud-to-mesh distance (EDF R&D, 2011) for each point and then filter points for which the positive cloud-to-mesh distance was higher than 4 standard deviations. This threshold has been chosen, by trials and errors, to preserve points located on steep ridges and steep valley bottoms that will be removed if chosen too low due to the smoothing of the surface by the interpolation. This procedure was repeated 3 times while refining the surface reconstruction between each iteration as further iterations removed significantly points located on ridges and valley bottoms.

Figure S1. Map of flight line extents from the pre-EQ LiDAR data. Flight lines are overlaid on the post-earthquake orthoimagery (12-15-2016, Aerial survey, 2017)

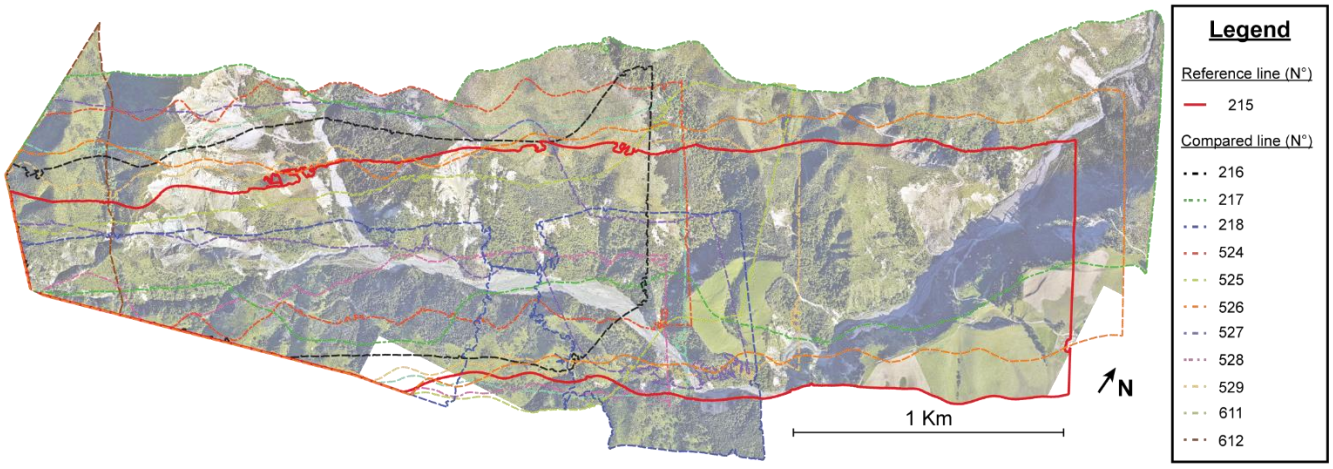


Table S1. Mean 3D-M3C2 distance and standard deviation between each flight line and a reference line estimated from the pre-EQ (n° 215, Fig. S2) and post-EQ LiDAR point cloud (n° 301).

Data	N° flight line	Mean 3D-M3C2 distance	Standard deviation
Pre-EQ point cloud	**216**	0.00	0.18
	217	0.00	0.15
	218	0.01	0.15
	524	-0.03	0.20
	525	-0.03	0.17
	526	-0.03	0.15
	527	-0.03	0.19
	528	-0.03	0.13
	529	0.00	0.14
	611	0.03	0.15
	612	0.03	0.14
Post-EQ point cloud	**108**	**-0.01**	0.06
	110	**-0.01**	0.05
	304	**0.01**	0.12
	303	**0.00**	0.07
	111	**0.00**	0.06

Table S2. Value of the transformation apply on the post-earthquake LiDAR data along the X, Y and Z axis.

	X	Y	Z
Coarse registration (m)	-	-	- 1.36
ICP fine registration (m)	-1.40	0.14	0.85

Section S2. Segmentation results with density based clustering algorithms

Density based spatial clustering algorithms aim at identifying clusters of high point density representing a signal, from areas of low point density representing noise. As for the connected component algorithm, the number of clusters to detect does not have to be specified. DBSCAN (Ester et al., 1996) is an algorithm now classically used in the segmentation of 3D rockfall inventories (e.g. Benjamin et al., 2020; Carrea et al., 2021; Riquelme et al., 2014). DBSCAN has two main parameters: (1) ε,

is a spatial search radius, (2) *minpoints* is the minimum of points within a distance ε of a point to define it as a cluster and to consider a point as member of an existing cluster. Both *minpoints* and ε are related, and are not easy to a priori define (Carrea et al., 2021). In our context, the main difference between a connected component and a density based clustering is that in the first case, cluster-membership is based on the distance to the nearest neighbor in 3D with respect to D_m, while in the second it is based on the density of neighbors within a radius ε. No cluster is defined if it does not have *minpoints.* As our minimum detectable landslide size is 20 m², and to be consistent with the parameters used in the Connected Component we set *minpoints*=N_p=20. Following Carrea and co-authors (2021), we use OPTICS (Ankerst et al., 1999) an improved version of DBSCAN available in the python scikit-learn library. OPTICS does not consider a fixed ε value to identify a cluster, although practically a maximum ε is generally set to reduce computation time. We test the impact of ε on the segmentation.

We also compare our results to HDBSCAN (McInnes et al., 2017), with the dedicated python library. HDBSCAN can better capture clusters of varying density compared to DBSCAN and ε does not have to be specified. HDBSCAN has two main parameters: (1) *min_cluster_size* which is equivalent to *minpoints* in OPTICS, and that we set to 20, our minimum detectable size; (2) *min_samples* which evaluate how conservative the clustering will be. As *min_samples* is increased, more points will be declared as noise. We test the impact of *min_samples* on the segmentation. To our knowledge, HDBSCAN has never been used in the context of rockfall segmentation.

OPTICS and HDBSCAN were applied on the significant core point sources and compared to the segmentation obtained by Connected Component. Computation time are given for a laptop using a core i7-7820HQ, clustering 370,708 points in 3D.

Results: Table S6 shows the number of clusters detected, the percentage of points not clustered and the size of the largest landslide. Figure S7 shows segmentation results. Figure S8 the pdf(A) resulting from various segmentations. The results slightly differ from the main paper as we used the standard LoD definition of M3C2 with t=1.96 in eq.(2).

In our reference segmentation with connected components (D_m=2 m), the number of clusters is 1270, the largest source area is 40,272 m², and 4.2 % of core points are not clustered. Decreasing D_m to 1.5 m does not reduce the largest landslide, but increasing D_m reduces the number of clusters while increasing slightly the size of the largest source. A connected component is calculated in 3 seconds with Cloudcompare.

OPTICS: As ε increases, OPTICS creates more clusters while fewer points are considered as noise, but the largest source area remains around 34,350 m². The segmentation map (Fig S7a and b) obtained for the case that discards the less points (ε=10 m), show that OPTICS manages to segment one branch of the largest landslide (A in Fig S7a), but does not significantly improve the segmentation of this landslide compared to a connected component. However, OPTICS tends to oversegment intermediate landslides into smaller ones, generating nearly 50 % more clusters than our reference case. This translates into a *pdf(A)* with a slightly lower exponent γ than the reference data (-1.88 vs -1.82) for ε=10 m (Fig. S8a). OPTICS computation time with scikit-learn is 45 minutes (for ε=10 m).

HDBSCAN: For the most conservative clustering parameter (*min_samples*=1), HDBSCAN creates 35% more clusters than the Connected Component, rejects only 2.5 % of points, and segments the largest landslide exactly as the reference dataset. As *min_samples* increases up to 20, the largest source area and the number of clusters decreases. For min_samples > 20, the

number of cluster decreases, but amalgamation starts to be very pronounced and the largest source area increases. For *min_samples*=20, HDBSCAN creates a similar number of clusters than our reference dataset, but the largest landslide is only 31,776 m², 21 % smaller than the largest landslide of the connected component. The segmentation shows that the largest landslide is not sub-segmented (Fig. S7c), and the size reduction is in fact due to removal of border core points. This size reduction happens for all landslides and is actually detrimental as these points correspond to statistically significant change. HDBSCAN segmentation does not result in a scaling exponent of *pdf(A)* significantly different from the connected component. HDBSCAN computation time with the python library is 45 seconds.

Discussion and conclusion

OPTICS and DBSCAN generates different segmentation than a connected component, which in some cases marginally improves the segmentation of large landslides compared to a connected component (e.g., landslide B in Fig. S7). However, they do not outperform a Connected Component in segmenting large strongly amalgamated landslides into several smaller ones. This is expected as the density of core points is actually nearly uniform in very large landslides. Density based clustering mostly affects clustering and noise detection on landslides close to the minimum cluster size (e.g., OPTICS) or border points where the core density drops (e.g., HDBSCAN). Yet, their impact on the pdf of source area remains marginal. Given that a trial and error approach is systematically needed to evaluate the best clustering parameters; the much longer computation time of OPTICS is hardly suitable for large datasets. HDBSCAN is faster, however choosing the right set of parameters is difficult, and border points of large landslide tend to be removed. Consequently, we choose the fastest and simplest solution that has the benefit of being highly scalable to much larger dataset and is consistent with segmenting a dataset where within landslide point density is uniform.

Table S3: summary of parameters tested for the 3 clustering algorithms used for source segmentation

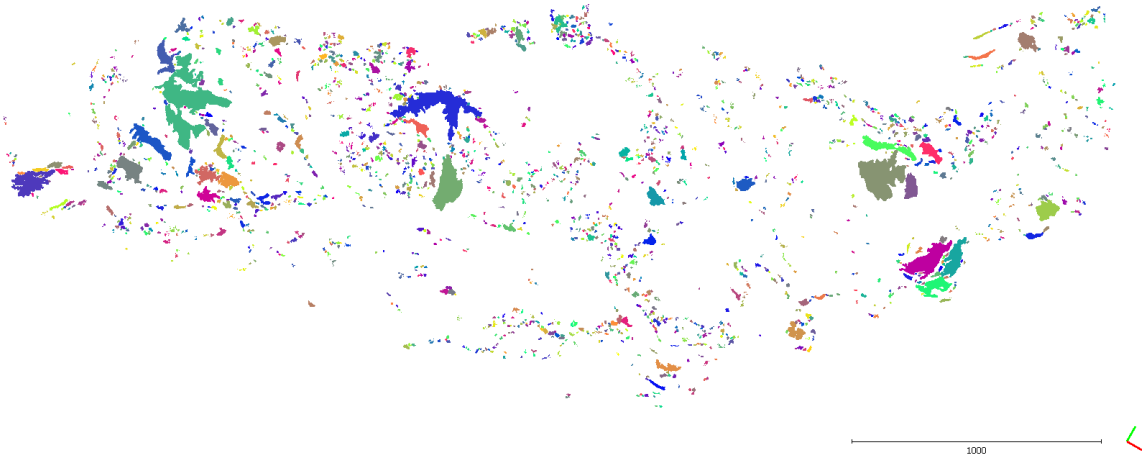
	Algorithm parameter	Nb_cluster	% of pts not clustered	Largest source area (m²)
Connected Component with N_p=20 and variable *min_samples*	1.5	1370	5.9	40,269
	2*	**1270**	**4.2**	**40,272**
	3	1061	3.7	40,819
	4	871	2.8	42,547
	6	634	1.7	42,809
OPTICS with min_points =20 and variable *epsilon*	3+	1335	17.6	34,350
	4	1632	13.7	34,350
	6	1793	11.7	34,346
	10	1873	10.6	34,346
HDBSCAN with min_cluster_size=20 and variable *min_samples*	1	1714	2.5	40,274
	10	1469	3.5	39,945
	15	1410	6.25	36,862
	20	1279	8.6	31,776
	25	1045	5.9	42,577

*: reference case used in the study before SNR and CDD filtering. +: minimum value for which the algorithm creates several clusters.

Figure S2: Best segmentation results of statistically significant core points with erosion obtained with Connected Component (a, reference dataset prior to SNR and CDD filtering), OPTICS (b), and HBDSCAN (c). The largest source is systematically A. Note the varying degrees of segmentation of sources B. Colors are random for each segmentation and not comparable.

a: Connected components, with D_m=2 m (reference dataset prior to SNR and CDD filtering)

b: DBSCAN-OPTICS with minimum cluster size = 20 and ε = 10



c: HDBSCAN with minimum cluster size = 20 and minimum number of samples = 20

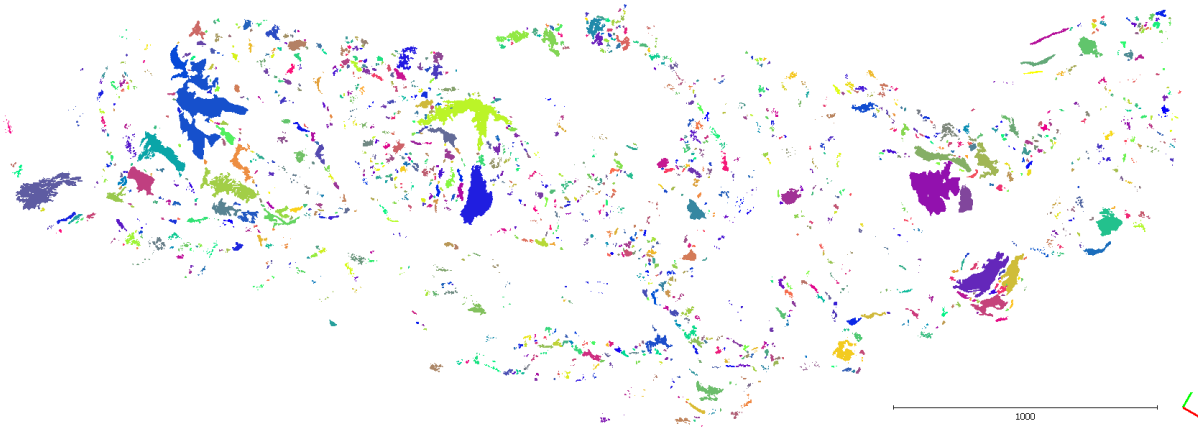


Figure S3: Probability density function of source area calculated for OPTICS, HDBSCAN and compared to the reference segmentation (Connected component, N_p=20, D_m=2 m) before SNR and CDD filtering. Fits for the segmentations shown in Fig. S1.

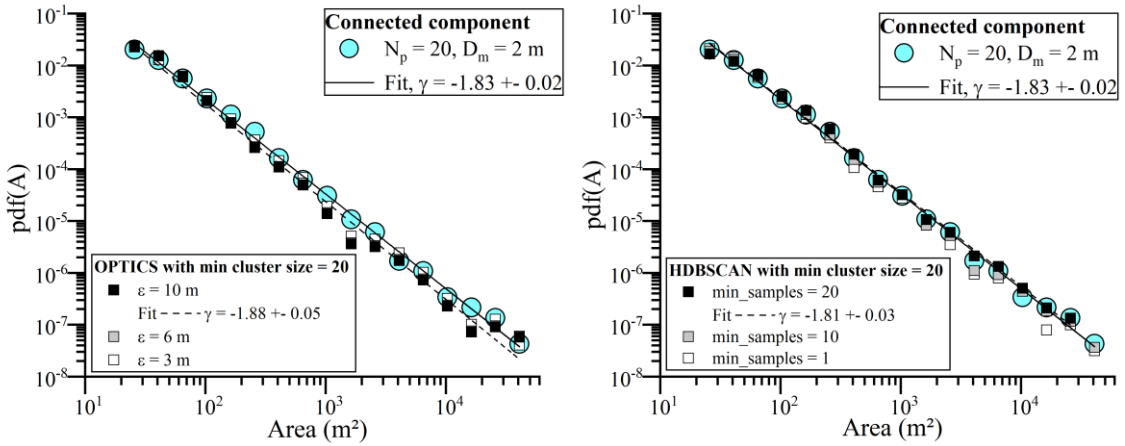
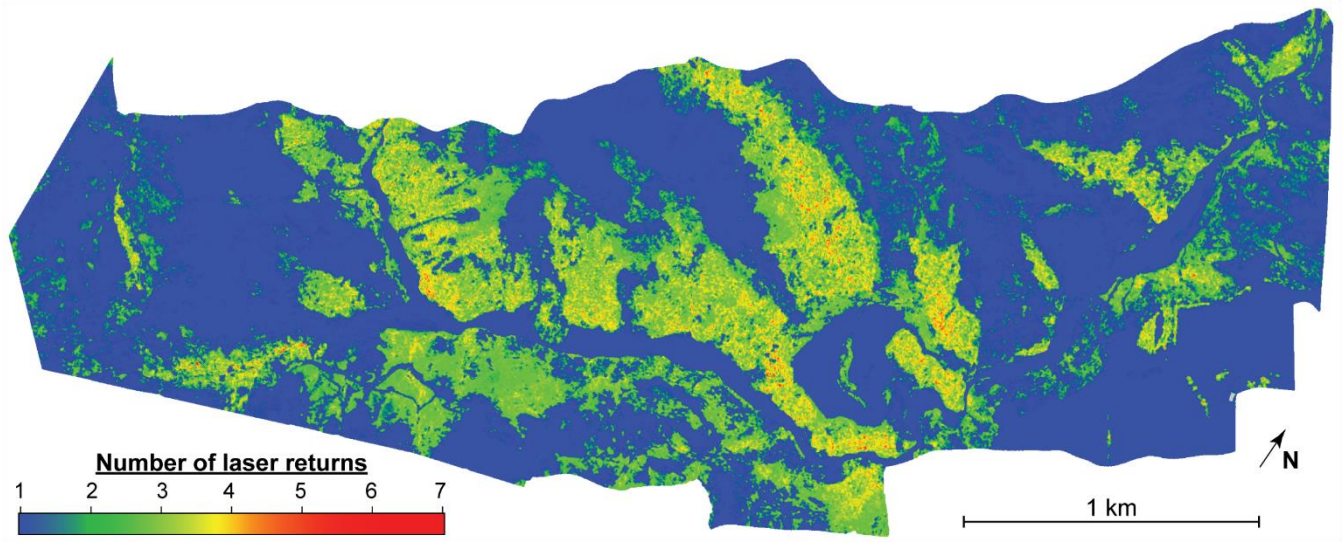


Figure S4. Map of the average number of laser returns from the post-EQ LiDAR data. The laser returns correspond to the number of targets a laser pulse has intercepted.



100

Figure S5. Map of the landslide source area only detected in the 3D predicted inventory with the classes corresponding to the table 4. Landslide sources are overlaid on the post-earthquake orthoimagery (12-15-2016, Aerial survey, 2017).

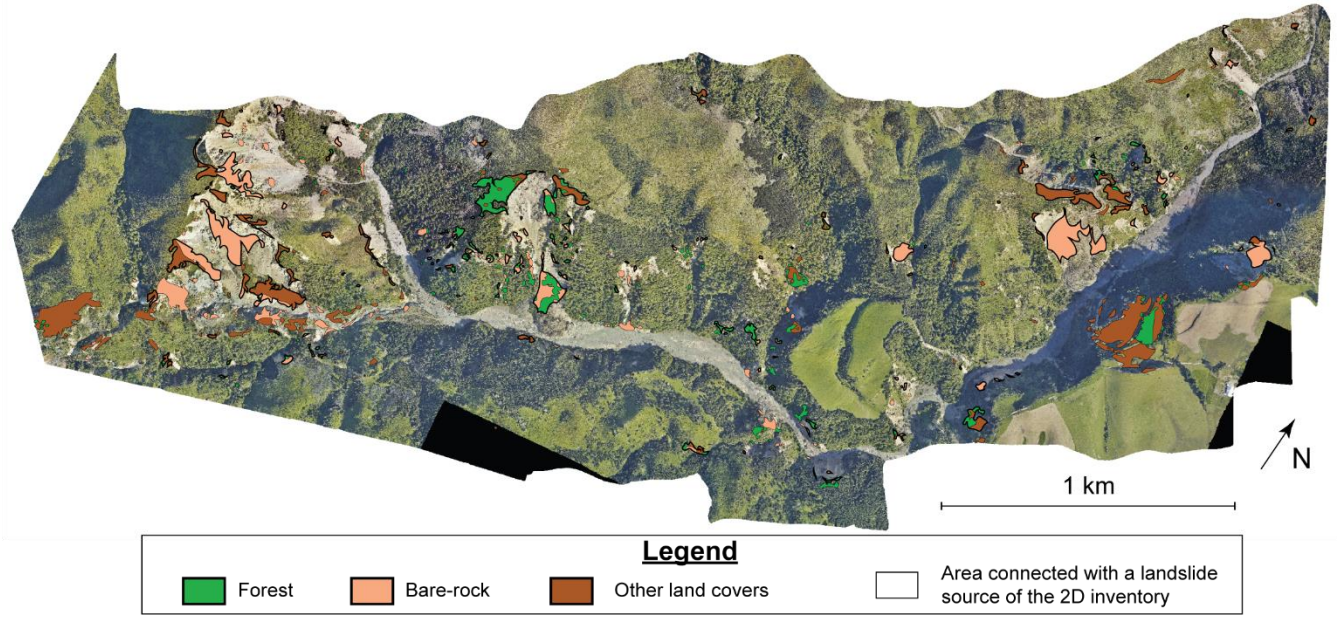


Figure S6. Slope distribution of landslide sources detected from 3D point cloud differencing

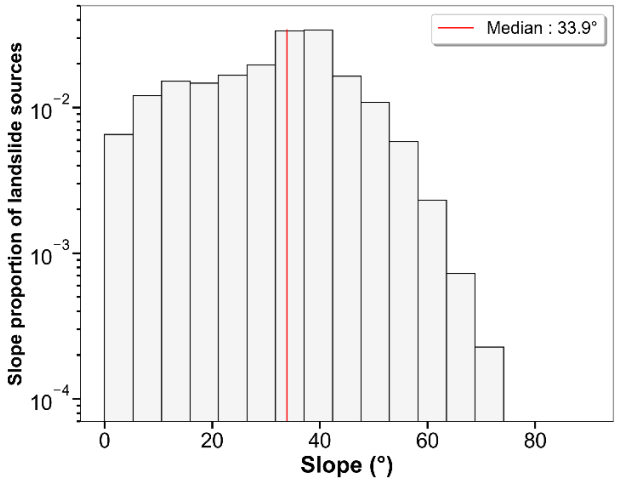


Fig S7: Sensitivity analysis to a) the registration error, b) the minimum segmentation distance and c) the SNR threshold to the landslide volume distribution.

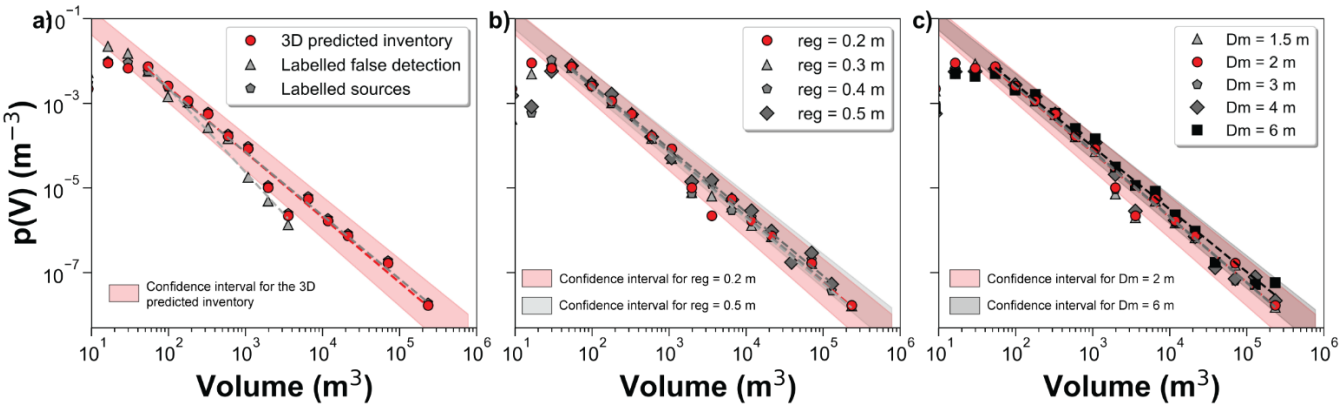
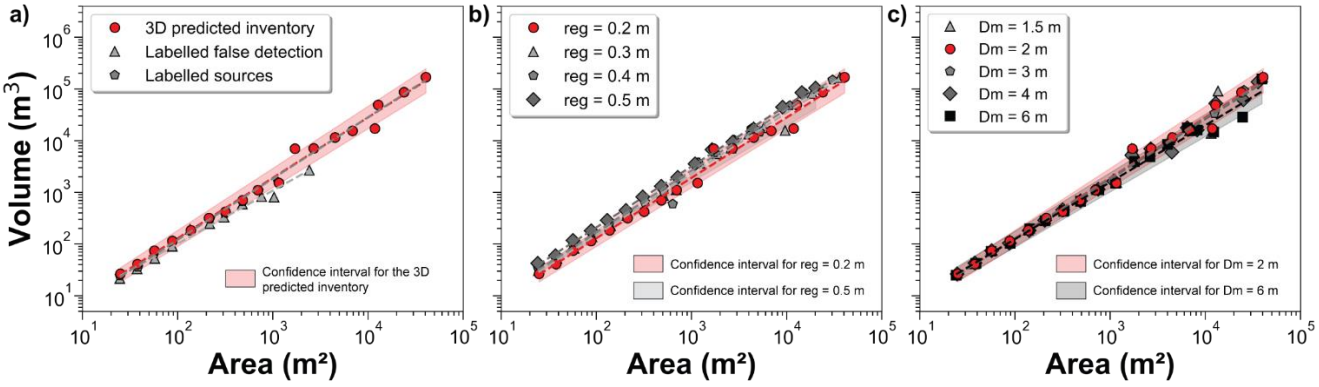


Fig S8: Sensitivity analysis to a) the registration error, b) the minimum segmentation distance and c) the SNR threshold to the landslide V-A relationship.



115

Table S4: Table of the result of the sensitivity analyses to the landslide inventory and the workflow parameters: *reg*, D_m. Units for the registration error *reg* and the minimum segmentation distance D_m are in meter.

Test	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
Inventory	**Predicted**	**433**	**0.65 ± 0.11**	**-1.64 ± 0.03**	**0.99**	**0.45 ± 0.23**	**-1.54 ± 0.07**	**0.98**	**-0.22 ± 0.10**	**1.17 ± 0.03**	**0.99**
	Label FD	355	1.13 ± 0.20	-1.89 ± 0.08	0.98	1.32 ± 0.33	-1.98 ± 0.12	0.98	-0.05 ± 0.09	1.02 ± 0.04	0.99
	Label TP	384	0.60 ± 0.11	-1.61 ± 0.04	0.99	0.35 ± 0.23	-1.50 ± 0.06	0.98	-0.29 ± 0.10	1.18 ± 0.03	0.99
reg	**0.2***	**433**	**0.65 ± 0.11**	**-1.64 ± 0.03**	**0.99**	**0.45 ± 0.23**	**-1.54 ± 0.07**	**0.98**	**-0.22 ± 0.10**	**1.17 ± 0.03**	**0.99**
	0.3	448	0.68 ± 0.13	-1.67 ± 0.04	0.99	0.45 ± 0.17	-1.54 ± 0.05	0.99	-0.19 ± 0.09	1.18 ± 0.03	0.99
	0.4	361	0.60 ± 0.09	-1.63 ± 0.03	0.99	0.42 ± 0.19	-1.52 ± 0.06	0.99	-0.15 ± 0.10	1.19 ± 0.03	0.99
	0.5	251	0.40 ± 0.16	-1.54 ± 0.05	0.98	0.44 ± 0.17	-1.50 ± 0.05	0.99	-0.04 ± 0.03	1.18 ± 0.01	0.99
Dm	1.5	486	0.69 ± 0.10	-1.67 ± 0.03	0.99	0.45 ± 0.25	-1.55 ± 0.07	0.98	-0.24 ± 0.12	1.18 ± 0.04	0.98
	2*	**433**	**0.65 ± 0.11**	**-1.64 ± 0.03**	**0.99**	**0.45 ± 0.23**	**-1.54 ± 0.07**	**0.98**	**-0.22 ± 0.10**	**1.17 ± 0.03**	**0.99**
	3	387	0.49 ± 0.14	-1.57 ± 0.04	0.99	0.50 ± 0.22	-1.55 ± 0.06	0.98	-0.22 ± 0.08	1.16 ± 0.03	0.99
	4	337	0.43 ± 0.15	-1.54 ± 0.05	0.99	0.55 ± 0.19	-1.56 ± 0.05	0.99	-0.21 ± 0.08	1.16 ± 0.03	0.99
	6	252	0.25 ± 0.17	-1.45 ± 0.06	0.98	0.41 ± 0.16	-1.48 ± 0.05	0.99	-0.10 ± 0.09	1.10 ± 0.03	0.99

References

120 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., Breuing, M. M., Kriegel, H. P. and Sander, J.: OPTICS: Ordering Points To Identify the Clustering Structure, ACM SIGMOD, 28(2), 49–60, 1999.

Benjamin, J., Rosser, N. J. and Brain, M. J.: Emergent characteristics of rockfall inventories captured at a regional scale, Earth

125 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.

EDF R&D, T. P.: Cloudcompare (version 2.12), [online] Available from: http://www.danielgm.net/cc/, 2011.

Ester, M., Kriegel, H. P., Sander, J. and Xu, X.: A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, in Kdd, vol. 96, pp. 226–231., 1996.

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.

Riquelme, A. J., Abellan, A., Tomas, R. and Jaboyedoff, M.: A new approach for semi-automatic rock mass joints recognition from 3D point clouds, Comput. Geosci., 68, 38–52, doi:10.1016/j.cageo.2014.03.014, 2014.