ANALYSIS OF FILTERING TECHNIQUES FOR INVESTIGATING LANDSLIDE-INDUCED TOPOGRAPHIC CHANGES IN THE OETZ VALLEY (TYROL, AUSTRIA)

I. Gutierrez[1,*], E. Før Gjermundsen [2], W. D. Harcourt[3], M. Kuschnerus[4], F. Tonion[5], T. Zieher[6]

[1] Department of Energy Resources, University of Stavanger, Kristine Bonnevies Vei 22, 4021 Stavanger, Norway
ivan.gutierrez@uis.no
[2] GIS-group, School of Business, University of South-Eastern Norway, Lærerskoleveien 40, 3679 Notodden, Norway
endre.f.gjermundsen@usn.no
[3] School of Physics and Astronomy, University of St Andrews, St Andrews, KY16 9SS, United Kingdom
wdh1@st-andrews.ac.uk
[4] Department of Geoscience and Remote Sensing, Delft University of Technology, Stevinweg 1, 2628CN Delft, The Netherlands
m.kuschnerus@tudelft.nl
[5] Department TESAF, University of Padua, Viale dell'Universita', Legnaro (PD), Italy
filippo.tonion@phd.unipd.it
[6] Institute for Interdisciplinary Mountain Research, Austrian Academy of Sciences, Technikerstr. 21a, 6020 Innsbruck, Austria
thomas.zieher@oeaw.ac.at

Commission II, WG 10

KEY WORDS: LiDAR, Change detection, Landslide, Terrain extraction, ISPRS Innsbruck Summer School of Alpine Research

ABSTRACT:

Landslides endanger settlements and infrastructure in mountain areas across the world. Monitoring of landslides is therefore essential in order to understand and possibly predict their behavior and potential danger. Terrestrial laser scanning has proven to be a successful tool in the assessment of changes on landslide surfaces due to its high resolution and accuracy. However, it is necessary to classify the 3D point clouds into vegetation and bare-earth points using filtering algorithms so that changes caused by landslide activity can be quantified. For this study, three classification algorithms are compared on an exemplary landslide study site in the Oetz valley in Tyrol, Austria. An optimal set of parameters is derived for each algorithm and their performances are evaluated using different metrics. The volume changes on the study site between the years 2017 and 2019 are compared after the application of each algorithm. The results show that (i) the tested filter techniques perform differently, (ii) their performance depends on their parameterization and (iii) the best-performing parameterization found over the vegetated test area will yield misclassifications on non-vegetated rough terrain. In particular, if only small changes have occurred the choice of the filtering technique and its parameterization play an important role in estimating volume changes.

1. INTRODUCTION

The monitoring of landslides is an important tool for assessing their activity in time and space. Only remote sensing techniques are suitable for an area-wide quantification of landslide-induced changes in topography. Current techniques used include (Scaioni et al., 2014): (i) Light Detection and Ranging (LiDAR), (ii) optical cameras and photogrammetry, (iii) passive thermal infrared sensing and (iv) Interferometric Synthetic Aperture Radar (InSAR) from ground and satellite-based sensors. LiDAR sensors have been most widely used for landslide investigations due to their high spatial resolution and accuracy (Glenn et al., 2006; Jaboyedoff et al., 2012). In particular, Terrestrial Laser Scanning (TLS; Pfeiffer et al., 2018) and Unmanned aerial vehicle-based Laser Scanning (ULS; Zieher et al., 2019) are flexible and cost-efficient platforms for monitoring landslides.

A striking advantage of LiDAR sensors is that they can penetrate vegetation, which enables topographic changes in vegetated areas to be detected (e.g. Eeckhaut et al., 2007). Ground-based InSAR sensors tend to operate at high radar frequencies (12-30 GHz) where vegetation penetration is minimal (Monserrat et al., 2014). So while InSAR techniques can penetrate

vegetation, the backscatter received from vegetation tends to be in the form of volume reflection. Extracting the points that represent the terrain in a 3D point cloud has a long tradition (e.g. Kraus and Pfeifer, 1998) and a variety of techniques and tools have been proposed and evaluated to achieve this (Sithole and Vosselman, 2004; Chen et al., 2017; Roberts et al., 2019). The general principle of terrain filtering is to select those points which represent the bare earth and remove objects in the intervening space. Following Briese et al. (2010) and Pfeifer et al. (2018), the techniques can be classified into: (i) morphological, (ii) progressive densification, (iii) surface-based, and (iv) segment-based filters.

In general, surface-based filters tend to out-perform other techniques but the complexity of the landscape ultimately determines the accuracy of the bare-earth extraction (Sithole and Vosselman, 2004). However, recent reviews on the subject have challenged this view (Chen et al., 2017; Roberts et al., 2019), calling instead for a multi-method approach to filtering. For example, the popular Triangular Irregular Network (TIN)-based densification method (Axelsson, 2000) performs well in steep terrain but is less accurate when the surface has many sudden discontinuities (Chen et al., 2017). In comparison, methods that filter at multiple scales (Evans and Hudak, 2007) are able to classify these discontinuities more readily. The simplest

---

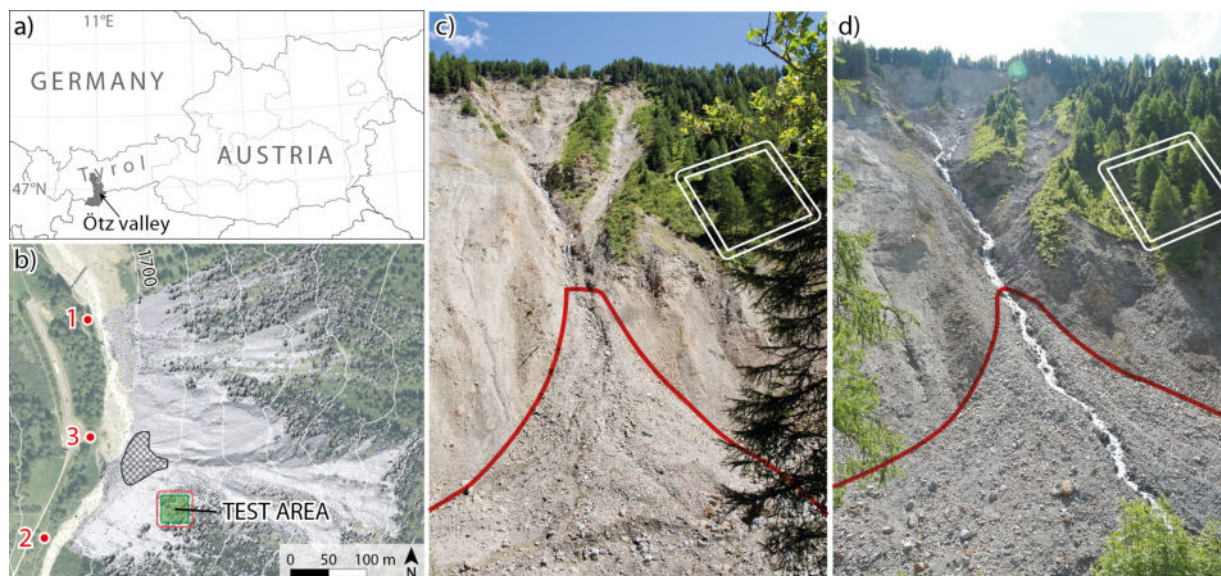* Corresponding author: ivan.gutierrez@uis.no

Figure 1. Overview map highlighting the Oetz valley (a); a zoomed in view of the landslide area showing the measurement setup at the landslide site including the scanning positions (red dots), the area covered by the acquired 3D point clouds (shaded relief), the selected 35x35 m test area for calibrating and comparing the filtering techniques, and the debris cone (hatched area) (b); photo of the landslide slope taken on 17/07/2017 (c); and photo of the landslide slope taken on 19/06/2019 (d). Orthophoto (2015) and contour lines (2010): Federal state of Tyrol, division of geoinformation. Photos: T. Zieher (c), M. Kuschnerus (d)

morphological techniques characterise ground and non-ground points using slope thresholds but struggle to classify terrain that contains multiple different objects. However, our understanding of the performance of each technique in different environmental settings remains limited. A common theme amongst all filters is that they become less accurate where the terrain is steep (Kraus and Pfeifer, 1998; Chen et al., 2017; Zhao et al., 2018; Pfeifer et al., 2018), making the discrimination between vegetation and bare-earth on landslides particularly difficult.

The aim of this study is to better understand the effects of different ground filtering techniques in quantifying volume changes caused by landslides. Here, ground points are defined as the lowest points acquired by TLS which likely represent the terrain surface below vegetation. The study was carried out on a partly vegetated slope prone to landslide processes in the inner Oetz valley (Tyrol, Austria) with TLS campaigns in July 2017 and June 2019. The parameterizations of three surface-based ground filtering techniques were systematically tested within a selected, vegetated test area and compared against a manual classification. Based on these tests, the sets of parameters which most closely matched the manual classification were identified for each technique and campaign. Subsequently, the performance of the filtering techniques was evaluated over rough terrain where no vegetation is present. Conclusions are then drawn on the effects of ground filtering for quantifying landslide-induced volume change. The objectives of the study are to:

1. Investigate the performance of different filtering techniques in determining ground and non-ground points in TLS 3D point clouds.

2. Identify optimum parameter values for each filtering technique by comparing the classification result to a manually classified test area.

3. Assess the effects of ground filtering on quantifying volume change in a debris cone between July 2017 and June 2019.

2. MATERIALS AND METHODS

2.1 Study area

The study area is located in the Oetz valley (Tyrol, Austria) between the villages of Zwieselstein and Obergurgl (Fig. 1). The valley bottom in this area is v-shaped and deeply incised into the Quaternary sediments and the underlying rocks of the Oetztal-Stubai-Basement complex. The west-facing slope of the study area is characterized by active landslide processes, including repeated rockfalls and debris slides and flows. On the valley bottom the slope is bound by the river Gurgler Ache at ca. 1700 m and the crest elevation of the landslide slope is ca. 2000 m. The foot slope forms part of the river embankment, which is subject to permanent fluvial erosion, further destabilizing the slope. The whole landslide can be viewed from several locations on the east-facing slope along a trail parallel to the road.

2.2 Data acquisition, pre-processing and registration

Figure 2 shows the main processing steps used in this study. The 3D point clouds were acquired with a Riegl VZ-6000 long-range TLS in July 2017 and June 2019. The scanner's wavelength of 1064 nm has the potential to cause eye damage and so safety precautions had to be taken to ensure no-one had a direct line of sight to the scanner. Scanning repetition rates were set to 150 kHz in July 2017 and 300 kHz in June 2019. Both, frame and line resolutions were set between 0.005° and 0.014°. The specified beam divergence of the Riegl VZ-6000 is 0.12 mrad resulting in a nominal footprint diameter of 1.2 cm at a distance of 100 m, considering a Gaussian beam with its diameter defined at $1/e^2$ irradiance (Riegl LMS GmbH, 2020).

Firstly, outliers were identified and removed as those points with less than five neighbours within a radius of 0.5 m. The number of points in the data set was then thinned using a 3D block thinning approach where only those points closest to
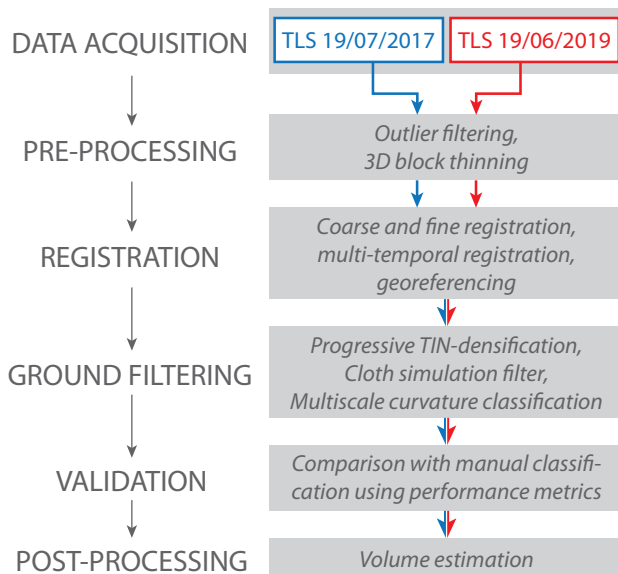
Figure 2. Conceptual workflow of the main processing steps.

the centre of a voxel with edge length 0.1 m were kept. Subsequently, the acquired 3D point clouds of the two epochs were registered respectively using the Iterative Closest Point algorithm (ICP; Besl and McKay, 1992) using flat areas extracted based on the locally computed planarity feature (1 m neighbourhood; planarity threshold 0.90). Then, the multi-temporal 3D point clouds were registered based on identified bedrock outcrops which are considered stable over time.

For comparing the performance of the considered ground filtering techniques, a test area with high vegetation density and sufficient coverage in the 3D point clouds acquired in 2017 and 2019 was selected (Fig. 1b,c,d). The test area has an edge length of 35 m (green square in Fig. 1b) and an additional buffer of 5 m (red polygon in Fig. 1b) around it to exclude edge effects in the filtering results. The 3D point clouds within the test area acquired in 2017 and 2019 were manually classified into ground and non-ground points on a point-by-point basis using the CloudCompare software (version 2.9.1). Subsequently, systematic tests of the filtering parameters were conducted using the 3D point clouds within the test area.

2.3 Applied terrain filtering algorithms

2.3.1 Progressive TIN Densification

The Progressive TIN Densification (PTD) algorithm starts from an initially coarse TIN of local minima and iteratively refines the classification by adding points which meet defined criteria (Fig. 3). The edge length of the triangulated network is iteratively reduced until reaching a defined Minimum Edge length (ME). Points within each triangle are considered if they are within a defined maximum distance constrained by the Maximum vertical Angle (MA) spanned by the network. The resulting classification is thus mainly controlled by the parameters ME and MA. A range of values were tested for each. For ME, values between 0.1 and 1.0 m were tested, and for MA, values between 5° and 80° were tested. The PTD algorithm was proposed by Axelsson (2000) and has been revised several times since then (e.g. Zhao et al., 2016; Nie et al., 2017). In this study the original version of the PTD implemented in SAGA LIS following Axelsson (2000) was used.
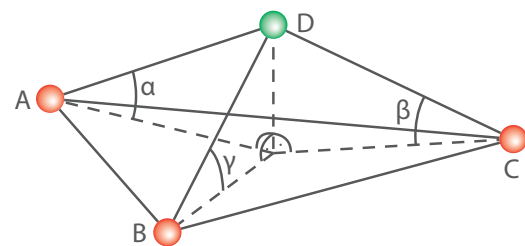

Figure 3. Sketch of the applied progressive TIN densification filter, modified after Axelsson (2000).

2.3.2 Cloth Simulation Filtering

The Cloth Simulation Filter (CSF) is a surface-based classifier which iteratively adjusts a flexible cloth to an inverted 3D point cloud by mimicking the laws of gravity. The simulated cloth is supported by cloth particles with a defined spacing, the Cloth Resolution (CR), and begins as a plane surface (Fig. 4). Its shape is then iteratively refined considering gravity and inner forces of the cloth until it is sufficiently supported by the 3D point cloud. Points located within a specified distance, the Classification Threshold (CT), to the simulated cloth are classified as ground points. The CSF was introduced by Zhang et al. (2016) and is available as a plugin in the CloudCompare Software (version 2.9.1). The test area was used to refine values for CR and CT. In this study, values between 0.1 and 0.5 were tested for CR, while values between 0.1 and 1.0 were tested for CT.
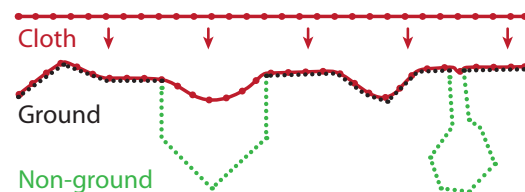

Figure 4. Sketch of the cloth simulation filter, modified after Zhang et al. (2016).

2.3.3 Multiscale Curvature Classification

The Multiscale Curvature Classification (MCC) algorithm is a surface-based classifier that determines ground and non-ground points in a 3D point cloud based on a user-defined curvature threshold (CUT). The surface curvature is calculated across multiple scales, defined by the Scale Parameter (SP), using an interpolated thin-plate spline surface. Ground and non-ground points are classified at three scale domains until the number of non-ground points being classified reaches <1%, when the iteration stops and the classification is complete (Fig. 5). The MCC algorithm was proposed by Evans and Hudak (2007) and is available as a command line tool (MCC-LiDAR v.2.1, https://sourceforge.net/projects/mcclidar). In this study both parameters have been systematically tested across ranges between 0.1 and 1.0 for both CUT and SP.
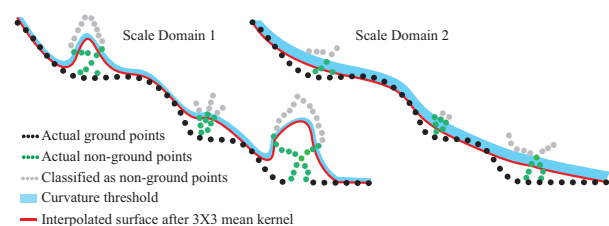

Figure 5. Sketch of multiscale curvature classification principle.

Table 1. Confusion matrix used to compare the manual classification and the classification from the filtering technique.

| | | FILTER RESULTS | |
		(1) Non-ground points	(2) Ground points
MANUAL CLASSIFICATION	(1) Non-ground points	Correctly classified non-ground points TRUE NEGATIVES (TN)	Non-ground points classified as ground FALSE POSITIVES (FP)
	(2) Ground points	Ground points classified as non-ground FALSE NEGATIVES (FN)	Correctly classified ground points TRUE POSITIVES (TP)

2.4 Validation procedure

Before analysing the results, the filtered points within the buffer area were removed. The remaining classified points were validated against manually classified 3D point clouds from 2017 and 2019. The following formulas correspond to the performance metrics (e.g. Xia and Wang, 2017; Roberts et al., 2019) that were calculated from a confusion matrix of classified ground and non-ground points (abbreviations refer to Table 1):

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \qquad (1)$$

The overall accuracy (ACC) relates the number of correctly classified ground (TP) and non-ground points (TN) to the total number of points.

$$PRE = \frac{TP}{TP + FP} \qquad (2)$$

The precision metric (PRE) is the ratio between the correctly predicted ground points (TP) and all points which were predicted as ground points by the filter (TP and FP). It reflects the ability of the filter technique to avoid commission errors i.e. by identifying ground points and differentiating them from non-ground points (Evans and Hudak, 2007).

$$REC = \frac{TP}{TP + FN} \qquad (3)$$

The recall metric (REC) is the ratio between correctly classified ground points (TP) and the total number of manually classified ground points (TP and FN). High values of the recall metric suggest that the filter minimizes omission errors by preserving valid ground points (Evans and Hudak, 2007).

$$F1 = 2 * \frac{PRE * REC}{PRE + REC} \qquad (4)$$

The F1 score is the harmonic mean of the precision and recall metrics. It ranges from 0 to 1 (1 = perfect precision and recall) indicating the efficiency of the filter to classify ground and non-ground points. The F1 score can be used to find the optimal trade-off between precision and recall and hence the best-performing parameter set for each filtering technique.

Table 2. Number of manually classified ground and non-ground points in the vegetated test area without the buffer.

Manual Classification	2017	2019
Ground Points	146785	203135
Non-ground Points	1330537	1076819
Total	1477322	1279954

2.5 Data post-processing

After identifying the best-performing parameter value sets for each technique, the values were then applied to the whole 3D point clouds. The filtering results were then evaluated on the debris cone (Fig. 1b), where large volumes of sediment of varying size have accumulated in the 2 years between measurements. The 2.5D volume change on the debris cone between 2017 and 2019 was assessed (i) based on the unfiltered 3D point cloud, assuming all translocated objects such as boulders should be considered, and (ii) based on the filtered 3D point clouds with the optimal parameter sets. By comparing (i) and (ii) the volume bias introduced by the filtering techniques can be quantified. For computing the volume change, the 3D point clouds were gridded at a spatial resolution of 0.25 m and taking the mean elevation in each raster cell.

For data processing, the following software packages were used: RiScan Pro (v.2.8), CloudCompare (v.2.9.1), SAGA LIS, MCC-LiDAR (v.2.1). The processing workflows have been automated using Python 2.7 (Python Software Foundation, 2019) and the R open source statistical software (R Core Team, 2019).

3. RESULTS

3.1 TLS accuracy assessment

The accuracy of the multi-temporal ICP registration between the 2017 and 2019 3D point clouds was assessed at different distances to the scanner by extracting bedrock outcrops assumed to remain stable. Point-to-plane distances were computed along locally fitted normal vectors resulting in an accuracy (2-fold standard deviation) of ± 2.5 cm at 200 m, ± 2.7 cm at 300 m and ± 5.2 cm at 550 m. Topographic changes of at least 10 cm could therefore be quantified. This value is above the diameter of the LiDAR footprint which is approximately 6.0 cm at 500 m (assuming a 90° incidence angle on a plane).

3.2 Filtering results for the vegetated test area

Performance curves in Fig. 6 show the derived F1 score metrics of the three filters applied to both 3D point clouds acquired in 2017 and 2019. The performance of each filter varies markedly over the ranges of values considered for each parameter. In general, the filters perform better in 2019 compared to 2017. This could be explained by differences in the ratio of points classified as ground or non-ground between the two measurement epochs. Although the total number of points is higher in the 2017 data set, its total number of ground points is lower compared to the 2019 data set (Table 2).

Generally, the three techniques depict a similar trend. Initially, there is an upward trend in the curves. Then, they level off and tend to stay constant, except for the steady downward trend observed for the MCC results (Fig. 6c,f). However, the filters
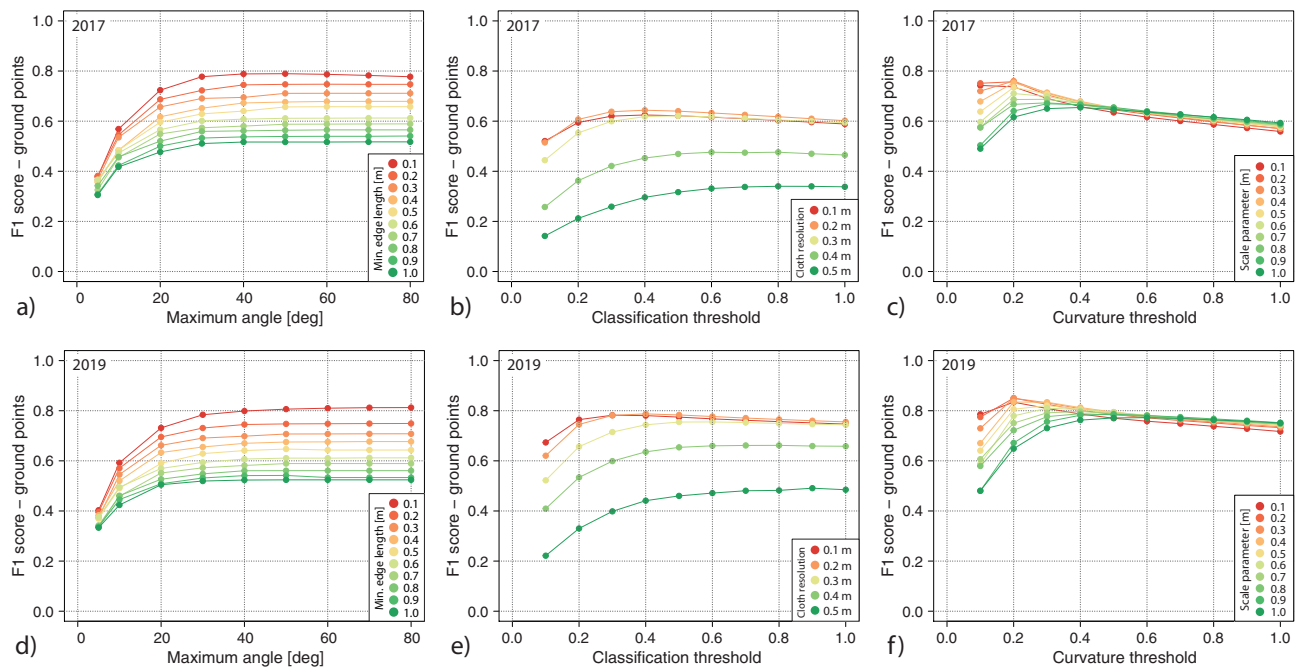
Figure 6. F1 scores for classifying ground points with the PTD filter (a,d), the CSF (b,e) and the MCC filter (c,f) for the selected vegetated test area. Results are shown for the 3D point clouds acquired in 2017 (a,b,c) and 2019 (d,e,f).

Table 3. Performance metrics for the identified best parameter set for each technique applied to the 3D point clouds acquired in 2017 and 2019. Acronyms are defined in the text.

Filter	Year	Parameter A	Parameter B	ACC	PRE	REC	F1
PTD	2017	MA 50°	ME 0.10 m	0.96	0.83	0.75	0.79
	2019	MA 80°	ME 0.10 m	0.95	0.91	0.73	0.81
CSF	2017	CT 0.40	CR 0.20 m	0.91	0.51	0.86	0.64
	2019	CT 0.40	CR 0.20 m	0.92	0.71	0.89	0.79
MCC	2017	CUT 0.20	SP 0.30 m	0.95	0.69	0.85	0.76
	2019	CUT 0.20	SP 0.20 m	0.95	0.84	0.86	0.85

perform differently when considering the parameter representing the neighbourhood sizes (depicted by the colour scale in Fig. 6). The F1 score of the PTD filter increases when ME is reduced. In contrast, applying the smallest chosen value of 0.1 m to CR in the CSF and SP in the MCC does not yield the maximum F1 score. For the MCC, using a small value for SP yields a high F1 score if combined with a small value for CUT.

The parameter sets providing the best classification, defined here as the optimal trade-off between precision and recall, can be derived from the highest F1 scores for each technique (Fig. 6, see Table 3). In case of the PDT filter the smallest tested ME of 0.10 m combined with an MA of 50° (2017) and 80° (2019) produce the highest F1 score. For the CSF, the combination of 0.20 m for CR and 0.40 for CT yields the highest F1 score for both data sets. The MCC performs best when using SP values of 0.30 m (2017) and 0.20 m (2019) combined with a value of 0.20 for CUT.

Table 3 shows the performance of each filter when applied to the 2017 and 2019 data sets, using the optimized parameter sets identified previously. The achieved ACC of each filter is comparable for each data set. However, comparing PRE, REC, and the F1 score, differences in the classification performance become clear.

The differences regarding PRE and REC are shown as maps in Fig. 7. The points are colour-coded following the confusion matrix. Maps displaying green coloured points (True Positive) along with magenta coloured points (False Positive) represent the PRE metric (Fig. 7a-f). The maps with a concentration of magenta points suggest low PRE (e.g. in the north-northwestern part of Fig. 7b, Fig. 7c, and Fig. 7e), particularly for the CSF and MCC classifications (both 2017 and 2019). Conversely, maps with fewer magenta points suggest a high PRE, which is particularly true for the PTD filter (Fig. 7a,d). Likewise, maps displaying green coloured points (True Positive) along with blue coloured points (False Negative) represent the REC metric (Fig. 7g-l). The generally less False Negatives resulting from the CSF and the MCC are mainly located in the southwestern part of the test area in both data sets. For the PTD filter results, the larger number of False Negatives are well distributed over the test area (Fig. 7g,j).

The PTD filter yields the lowest number of False Positive points, thus avoiding commission errors. Therefore, it is the technique that achieves the best performance in terms of PRE. However, the PTD filter underestimates the number of manually classified ground points. Hence, it does not avoid omission errors efficiently and it yields the highest number of False Negative points as consequence. The PTD filter shows the poorest performance in terms of REC. In contrast, the CSF results show the lowest PRE because the number of manually classified ground points is systematically overestimated (Fig. 7b,e). Nonetheless, it is the technique with the best REC (Fig. 7h,k). The values for PRE and REC for the MCC filter are more balanced compared to those resulting from the CSF and PTD filter. The F1 score indicates that the MCC is the technique that achieves the most optimal trade-off between PRE and REC, particularly in 2019 (Fig. 6f, Table 3).

3.3 Effects of filters in quantifying volume change over a debris cone

Figure 8 highlights the points classified as non-ground on the debris cone using each technique, applying the best parameter sets identified in the vegetated test area. Despite the absence
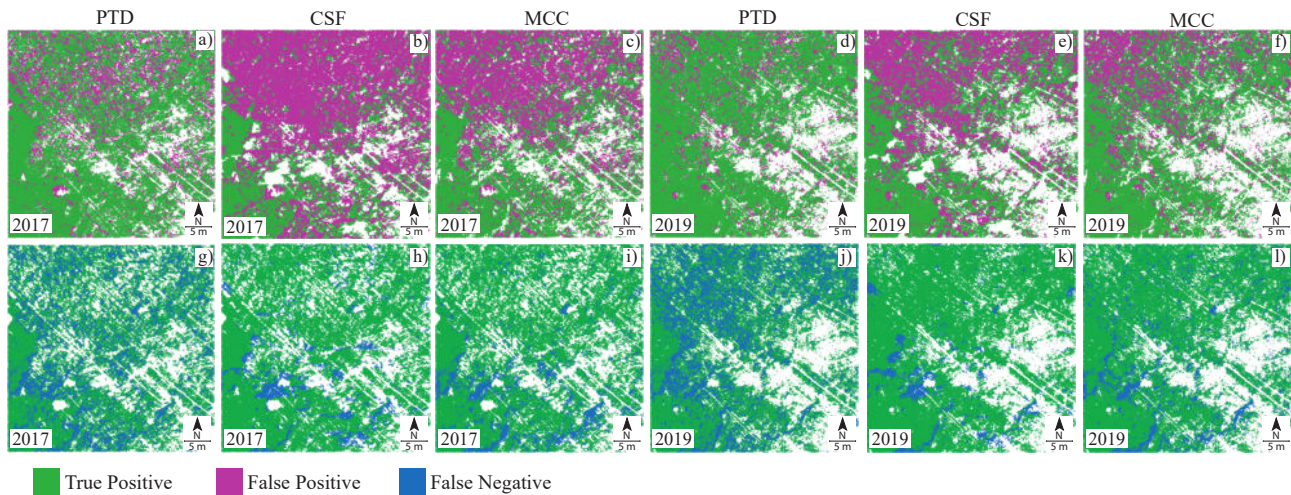
Figure 7. Point cloud maps of the vegetated test area derived from the classified data sets acquired in 2017 (a,b,c,g,h,i) and 2019 (d,e,f,j,k,l). The test area was scanned from the west and north-west direction. The points are coloured following the confusion matrix classification code for each ground filtering technique. The results are shown for the following filtering techniques: PTD (a,g,d,j), CSF (b,h,e,k) and MCC (c,i,f,l). PRE metric (a-f) and REC metric (g-l). See text for full discussion.

Table 4. Number of classified Ground Points (GP) and Non-Ground Points (NGP) on the debris cone. The filter acronyms are defined in the text.

Filter	Number of GP		Number of NGP		Percentage NGP	
	2017/07	2019/06	2017/07	2019/06	2017/07	2019/06
PTD	296149	357373	59743	127199	16.8%	26.2%
CSF	351402	475875	4490	8697	1.3%	1.8%
MCC	343359	454330	12533	30242	3.5%	6.2%

Table 5. Volume change and derived volume bias on the debris cone for the different filtering techniques. The filter acronyms are defined in the text.

Filter	Volume change [m³]			Volume bias [m³]		
	positive	negative	total	positive	negative	total
NONE	3373.6	-25.5	3348.1			
PTD	3348.6	-26.6	3322.0	-25.0	-1.1	-26.1
CSF	3371.4	-25.4	3346.0	-2.2	0.1	-2.1
MCC	3363.0	-25.9	3337.1	-10.6	-0.4	-11.0

of vegetation, the filtering techniques yield non-ground points over the debris cone, although their number and location varies between techniques. This can be attributed to the rough terrain across the debris cone which introduces discontinuities interpreted as vegetation by the filtering techniques. The respective number of classified ground and non-ground points are shown in Table 4. Generally, more non-ground points are classified in the 3D point cloud acquired in 2019. Comparing the three filtering techniques, the PTD filter classified the most points as non-ground, and these were distributed all over the debris cone including the top surface of boulders but also small blocks (Fig. 8a,d). The MCC yields less non-ground points than the PTD filter. The points mainly cover the tops of large boulders (Fig. 8c,f). The smallest number of non-ground points resulted from the CSF (Fig. 8b,e), and these were mainly located on top of large boulders.

To quantify the volume change between the two epochs, the 2017 and 2019 unfiltered 3D point clouds were differenced from each other. Based on this calculation, more than 3000 m³ of loose material accumulated across the landslide over the 2 year study period. This included the numerous large boulders that were present on the surface of the debris cone. Only selected areas show a negative elevation change, which together equate to -25.5 m³ (Table 5). Assuming the 3D point cloud coordinates have a vertical uncertainty of ±2.5 cm at the range of the debris cone, the uncertainty of the derived total volume change is ±60 m³ within the covered area of 2400 m².

Using different filtering techniques alters the volume change estimate (Table 5). Because of the larger number of non-ground points classified by the PTD filter, which are omitted when computing the respective DTMs, this technique also yields the

highest volume bias of -26.1 m³. On the contrary, the bias introduced by the CSF of -2.1 m³ is considerably lower. However, the derived volume bias is within the accuracy of the acquired data.

4. DISCUSSION

This study has illustrated the effects of applying different ground filtering techniques when quantifying volume change on a landslide slope in the Oetz Valley, Austria. A subset of the data, the test area, was manually classified and compared to the classified 3D point cloud produced by each filtering technique. This analysis identified the optimum parameter sets for each filtering technique. When considering the F1 score which was used as a global performance metric, both the PTD and the MCC filtering techniques perform better than the CSF (PTD performs better for the 2017 data set and MCC better for the 2019 data set, see Table 3), which makes them comparable in terms of performance.

It should be noted here that since the number of non-ground points in the test area is higher than the number of ground points, the accuracy does not necessarily reflect the effectiveness of the filtering techniques. However, these results show that even on a small, well known test area, choosing the best performing algorithm with the best set of parameter values is not trivial. The choice of algorithm may ultimately depend on whether False Positives or False Negatives are acceptable or the chosen metric (ACC, PRE, REC or F1 score), and both can be subject to 'over-fitting'. Whilst it may be possible to classify vegetation as non-ground points over this test area, it does not necessarily hold for other regions of the data set.
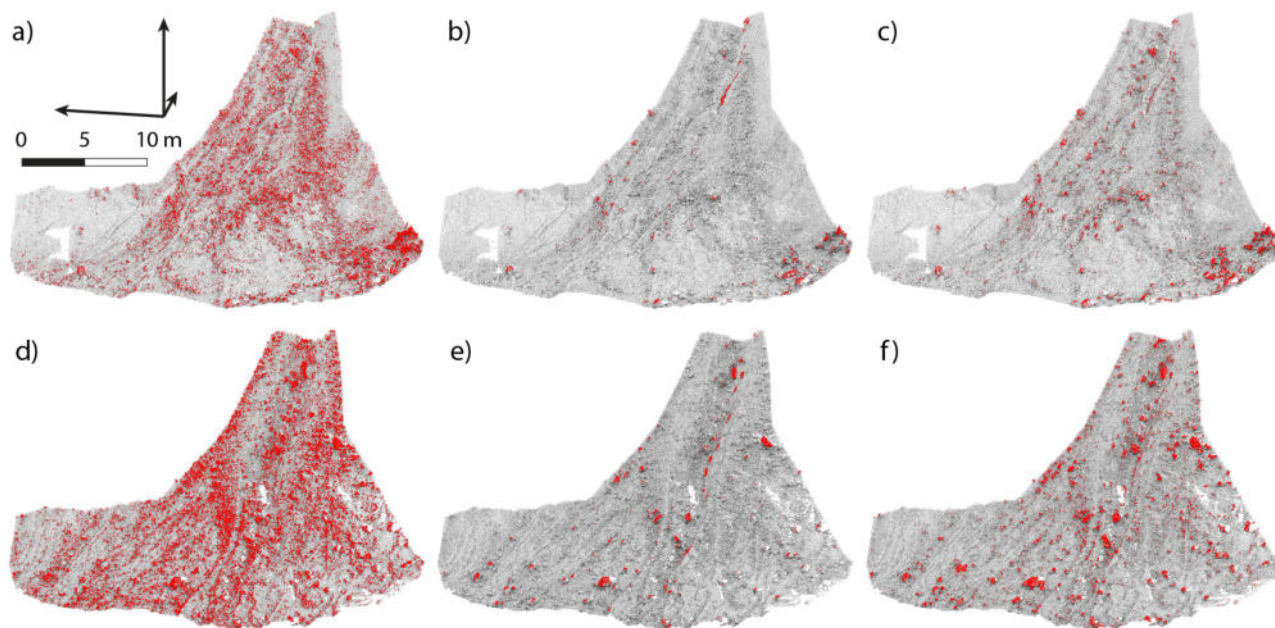
Figure 8. Comparison of the classified non-ground points (red) on the debris cone. Results are shown for the points clouds acquired in 2017 (a,b,c) and 2019 (d,e,f) filtered with the best parameter set using the PTD filter (a,d), the CSF (b,e) and the MCC (c,f).

When applying each optimized filtering technique, there are examples of False Negative results (i.e. valid ground points classified as non-ground) across the study area, particularly over the debris cone. Here, no vegetation is present but selected areas of loose debris, the tops of boulders and other loose material are repeatedly classified as non-ground. This is because these areas act as discontinuities which the filtering techniques determine to be related to objects above the ground, thus demonstrating a clear weakness in the tested classifiers. Further, the results in Table 4 suggest that the CSF performed best over the debris cone, assuming only ground points are present here, even though it performed the weakest over the test area. This suggests that the CSF performs better when less vegetation is present, while the PTD and MCC are preferable when the vegetation is more dense.

The results indicate that the choice of ground filtering technique impacts the estimation of volume change on a landslide, particularly when aiming at quantifying minor volume changes (e.g. occurring during a shorter period of time). Therefore, choosing the optimal parameter values is not a simple task. This study demonstrates a data driven approach to choosing the optimum parameter sets for a landslide. However, optimizing these parameter values over vegetation leads to false classification of non-ground points over the debris cone. Whilst this impacts the volume change estimate, the difference between filtering techniques is small (Table 5) i.e. they are all within the uncertainty bounds of the total volume estimation. Thus, all three filtering techniques provide a reasonable solution to the classification of bare-earth where manual classification is not an option.

In the vegetated test area, removing valid ground points may be preferred over introducing valid non-ground points to the classified ground points. As demonstrated in section 3, this can be best achieved using the PTD filtering technique. In comparison, if the study area contains almost no vegetation, as was the case for the debris cone in this study, the CSF is the best alternative. The MCC appears to perform better when the vegetation is more sparse since it delivers a good balance between REC and

PRE (high F1 score). Given that the performance of these techniques changes based on the surface conditions, a multi-method approach may be more suitable.

The implementation of the optimized parameter sets derived in this study might not readily apply to other studies where the landslide has comparable surface conditions. Acquisition aspects such as the TLS acquisition settings and the location and number of scan positions may influence the performance of the filtering techniques.

5. CONCLUSIONS

This study has compared the performance of three widely used filtering techniques for classifying ground and non-ground LiDAR point clouds. The key results are as follows:

1. Each technique performs differently depending on the density of vegetation present on the surface.

2. The performance of each technique depends on the choice of paramaterization used.

3. Despite optimizing the parameters for each filtering technique, boulders and loose debris on the debris cone were often misclassified as non-ground owing to the use of surface-based classifiers.

When the filter overestimates the number of ground points the precision (PRE) performance reduces as the number of False Positive classification points increase. This means that the filtering technique cannot efficiently avoid commission errors. Conversely, the filtering technique achieves a high recall (REC) performance by avoiding omission errors. In this situation, the filtering technique has a high probability of detection of ground points and delivers fewer False Negative points as consequence.

For quantifying topographic change, the parameterization of the filtering technique must be adapted to the terrain conditions. In

particular, areas partly covered by vegetation requires a trade-off between filter performance and the correct classification of non-ground points (True Negatives). Alternatively, areas where non-ground points are not expected to be present (e.g. on the debris cone) could be identified a priori and omitted from the filtering.

The workflow and performance metrics employed in this study to obtain the optimized parameter sets, could be easily adapted to similar studies dealing with terrains partially covered by vegetation. Utilising the optimized parameter sets and the knowledge of how each filtering technique performs, a multi-method approach to ground filtering could be achieved.

ACKNOWLEDGEMENTS

References

Axelsson, P., 2000. Dem generation from laser scanner data using adaptive tin models. *International Archives of Photogrammetry and Remote Sensing* 33(B4/1; PART 4), pp. 111–118.

Besl, P. and McKay, N., 1992. A method for registration of 3-D shapes. *IEEE Transactions on pattern analysis and machine intelligence* 14(2), pp. 239–256.

Briese, C., Vosselman, G. e. and Maas, H.-G. e., 2010. *Airborne and Terrestrial Laser Scanning*. Taylor & Francis, chapter Extraction of Digital Terrain Models, pp. 135–167.

Chen, Z., Gao, B. and Devereux, B., 2017. State-of-the-art: DTM generation using airborne LIDAR data. *Sensors*.

Eeckhaut, M. V. D., Poesen, J., Verstraeten, G., Vanacker, V., Nyssen, J., Moeyersons, J., Beek, L. P. H. v. and Vandekerckhove, L., 2007. Use of lidar-derived images for mapping old landslides under forest. *Earth Surface Processes and Landforms* 32(5), pp. 754–769.

Evans, J. and Hudak, T., 2007. A multiscale curvature algorithm for classifying discrete return lidar in forested environments. *IEEE Trans. Geosci. Remote Sens.* 45, pp. 1029–2018.

Glenn, N. F., Streutker, D. R., Chadwick, D. J., Thackray, G. D. and Dorsch, S. J., 2006. Analysis of lidar-derived topographic information for characterizing and differentiating landslide morphology and activity. *Geomorphology* 73(1), pp. 131 – 148.

Jaboyedoff, M., Oppikofer, T., Abellán, A., Derron, M. H., Loye, A., Metzger, R. and Pedrazzini, A., 2012. Use of lidar in landslide investigations: A review. *Natural Hazards* 61(1), pp. 5–28.

Kraus, K. and Pfeifer, N., 1998. Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing* 53(4), pp. 193 – 203.

Monserrat, O., Crosetto, M. and Luzi, G., 2014. A review of ground-based sar interferometry for deformation measurement. *ISPRS Journal of Photogrammetry and Remote Sensing* 93, pp. 40–48.

Nie, S., Wang, C., Dong, P., Xi, X., Luo, S. and Qin, H., 2017. A revised progressive TIN densification for filtering airborne LiDAR data. *Measurement* 104, pp. 70 – 77.

Pfeifer, N., Mandlburger, G., Shan, J. e. and Toth, C. e., 2018. *LiDAR Data Filtering and Digital Terrain Model Generation*. CRC Press, Boca Raton, chapter Topographic Laser Ranging and Scanning.

Pfeiffer, J., Zieher, T., Bremer, M., Wichmann, V. and Rutzinger, M., 2018. Derivation of Three-Dimensional Displacement Vectors from Multi-Temporal Long-Range Terrestrial Laser Scanning at the Reissenschuh Landslide (Tyrol, Austria). *Remote Sensing* 10(11), pp. 1688.

Python Software Foundation, 2019. Python Programming Language. Python Software Foundation, Hampton, New Hampshire. http://www.python.org/ (last access: 16 December 2019).

R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ (last access: 7 January 2020).

Riegl LMS GmbH, 2020. Riegl VZ-6000 data sheet. Horn, Austria. www.riegl.com/uploads/tx_pxpriegldownloads/RIEGL_VZ-6000_Datasheet_2019-09-02.pdf (last access: 27 January 2020).

Roberts, K. C., Lindsay, J. B. and Berg, A. A., 2019. An analysis of ground-point classifiers for terrestrial lidar. *Remote Sensing*.

Scaioni, M., Longoni, L., Melillo, V. and Papini, M., 2014. Remote sensing for landslide investigations: An overview of recent achievements and perspectives. *Remote Sensing*.

Sithole, G. and Vosselman, G., 2004. Experimental comparison of filter algorithms for bare-earth extraction from airborne laser scanning point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing* 59(1), pp. 85 – 101.

Xia, S. and Wang, R., 2017. A Fast Edge Extraction Method for Mobile Lidar Point Clouds. *IEEE Geoscience and Remote Sensing Letters* 14(8), pp. 1288–1292.

Zhang, W., Qi, J., Wan, P., Wang, H., Xie, D., Wang, X. and Yan, G., 2016. An Easy-to-Use Airborne LiDAR Data Filtering Method Based on Cloth Simulation. *Remote Sensing*.

Zhao, X., Guo, Q., Su, Y. and Xue, B., 2016. Improved progressive tin densification filtering algorithm for airborne lidar data in forested areas. *ISPRS Journal of Photogrammetry and Remote Sensing* 117, pp. 79 – 91.

Zhao, X., Su, Y., Li, W. K., Hu, T., Liu, J. and Guo, Q., 2018. A Comparison of LiDAR Filtering Algorithms in Vegetated Mountain Areas. *Canadian Journal of Remote Sensing* 44(4), pp. 287–298.

Zieher, T., Bremer, M., Rutzinger, M., Pfeiffer, J., Fritzmann, P. and Wichmann, V., 2019. Assessment of Landslide-induced Displacement and Deformation of Above-ground Objects Using UAV-borne and Airborne Laser Scanning Data. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* IV-2/W5, pp. 461–467.