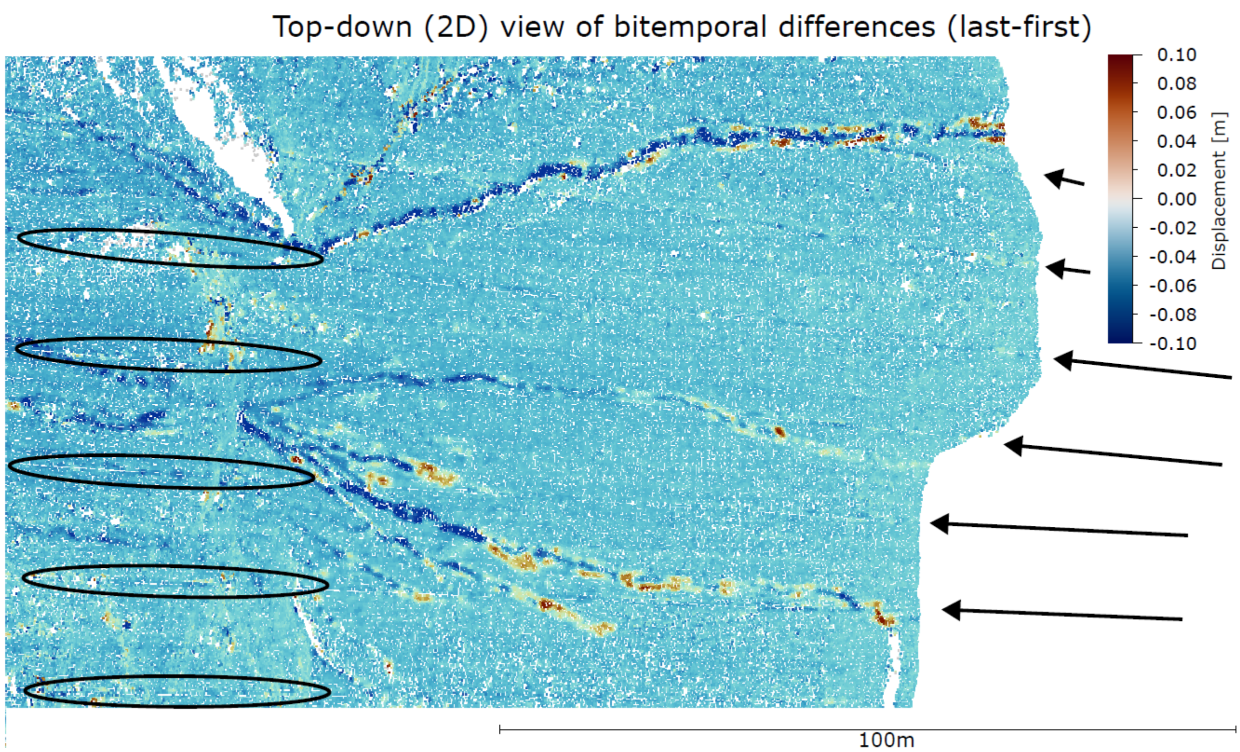
Authors’ comment: ESURF-2021-103

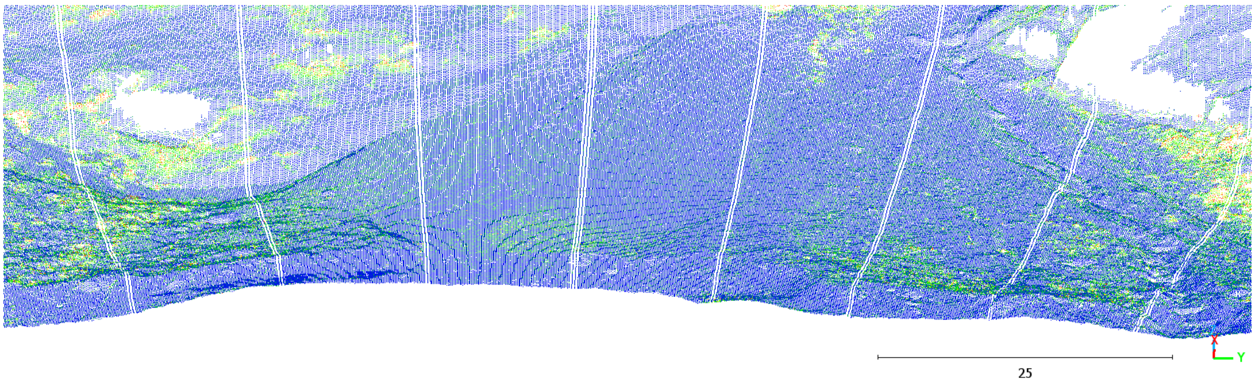
Full 4D Change Analysis of Topographic Point Cloud Time Series using Kalman Filtering

This authors’ comment accompanies a revised version of the manuscript originally submitted on December 18, 2021. It extends the comments of the AC1 (first response to the reviewer’s comments) submitted on April 25, 2022 (https://doi.org/10.5194/esurf-2021-103-AC1).

In response to the comments by Reviewer 2 (Dr. D. Lague), we have improved to coregistration procedure of the multitemporal dataset. We now use a Helmert transform derived from retroreflective prisms installed at stable positions around the area of interest. As a consequence, the derived level of detection has dropped to a point where the original dataset (acquired in 2020) started to show changes resulting from sensing errors in the data. The source of these errors, showing as missing scan lines, is unclear. However, they dominate the resulting segmentation and the change image plots, shown below:



Perspective view at the bottom rim of a single epoch dataset (in the direction of the arrows of the previous figure). The missing data lines are clearly visible:



To avoid putting a focus on these missing scan lines in the paper, we decided to use an updated dataset, acquired at the same location and in the same setting in 2021 (one year later). The dataset covers a longer period and shows multiple different surface change processes (incl. snowfall and avalanches). We believe that this adaptation is in the best interest of presenting the method of Kalman filtering for 4D change detection.

In the following, we write the reviewer's comments in *italics and grey font*, and our responses in black. We have numbered the comments for better reference. Direct quotes from the updated manuscript are highlighted in **dark green and bold font**.

We thank the reviewers for their valuable input and are looking forward that the revised, strongly improved manuscript will be considered for publication.

Aug 28, 2022; Dr. Lukas Winiwarter, on behalf of all authors

1. Review by Dr. R. Lindenbergh and M. Kuschnerus

1. *The presented work contains a lot of interesting ideas and visualizations and is definitely pushing information extraction from time series of 3D data a good step forward. Especially the processing steps of spatial smoothing (M3C2-EP) in combination with temporal smoothing (Kalman filter) to generate regularly sampled, smooth time series are very innovative. These smoothed time series could be used for a variety of applications and in combination with many other methods. Here the authors choose to use feature extraction and clustering to find regions of similar deformation behavior. The explanation of these last two steps lacks focus and should be concentrated on one (maximal two) sets of features and one clustering method. A separate section of the results should then deal with the comparison to other methods.*

As suggested, we have removed the calculation of engineered features and now use only kMeans clustering (following Kuschnerus et al., 2021) on the smoothed and on the raw time series data. In addition, we have included Kromer et al. (2015) as a comparison method of temporal averaging, resulting in a total of four methods to be compared: Kalman Filtering + kMeans clustering, spatiotemporal smoothing (Kromer et al., 2015) + kMeans clustering, kMeans clustering without smoothing (simple linear interpolation) and bitemporal surface change analysis. In the discussion section, we discuss the differences and similarities between the methods based on selected surface processes contained in the data.

2. *The Kalman filter is one way of interpolating a time series. Directly related is Kriging, which aims at assessing and exploiting spatial and/or temporal correlation. Kriging also enables error propagation. Time series could also be approximated using Fourier or BSpline polynomials. This could be better discussed in the Intro (as it is related work).*

We have added the following alternative time series interpolating methods to the introduction:

To smooth observed time series, (B-)splines are commonly employed (Lepot et al., 2017). Splines are piece-wise approximations of the signal by polynomial functions. Depending on the degree n of the polynomials, the continuity of derivatives is guaranteed up to order n−1, resulting in smooth estimates. For example, with commonly used cubic splines, the second derivative is continuous. In general, splines are interpolators, meaning they will pass through every data point. In the presence of noise, this might not be justified, and approximative splines utilizing least-squares methods have been presented (Wegman and Wright, 1983). For time series of 3D point clouds, a moving average filter has been successfully used to reduce daily patterns and random effects in time series (Kromer et al., 2015; Eltner et al., 2017; Anders et al., 2019).

The geostatistical prediction method of Kriging (Matheron, 1963; Goovaerts, 1997) has been applied in the analysis of time series of geospatial data (e.g., Lindenbergh et al., 2008). Kriging allows to estimate the uncertainty of the predicted (interpolated) value, an important measure when attempting to separate change signals from noise (e.g., Lloyd and Atkinson, 2001). For example, when the distance between sampling locations increases, the uncertainty for predictions between these locations will also increase, following the variogram derived in the Kriging process.

3. *In your Kalman filter implementation you use three parameters, location, velocity and acceleration. First it should be clearly stated somewhere that you use these parameters to model change in the vertical direction (right?). That said, using velocity and acceleration to model change at a location that changes due to digging is not directly intuitive to me, as such change would better modeled as a step function, please comment. Or more general, how should instantaneous change be incorporated in your setup? And do you really need acceleration as a third parameter, would location and velocity not provide similar results in an easier way?*

The Kalman filter models change in the direction of the M3C2 distances, which is not necessarily the vertical. We have clarified this in the manuscript:

We present the use of a Kalman filter, which can be used to incorporate multiple observations (in our case the change values for each epoch, quantified along the local 3D surface normals using M3C2-EP, cf. Section 2.1) and obtain predictions about the displacement at arbitrary points in the time series, analogous to the median smoothing.

It is correct that the Kalman filter is ill-suited for sudden changes (i.e. change occurring between successive epochs). Here, a re-initialisation of the Kalman filter could be carried out once the locations are detected. We now point this out in the discussion:

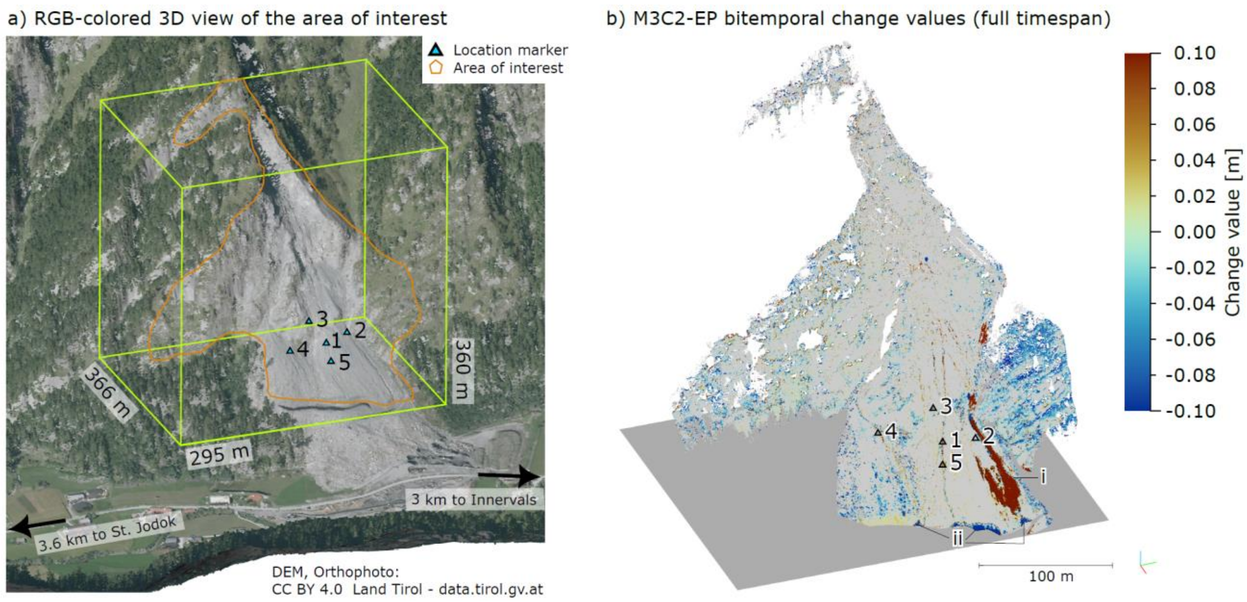
Future research could investigate how discrete change events can be identified and modeled appropriately by re-initializing the Kalman filter just after such an event. Such a re-initialization resets the estimated displacement, velocity, and acceleration (depending on the chosen order of the model), which increases the uncertainty until more observations become available and the filter converges

again. In line with this consideration is the choice of uncertainty at the beginning of the process. At the start of the time series, the displacement must be - by definition - zero, and we, therefore, assign an uncertainty of zero to this initialization. This also ensures that all trajectories pass through the point at 0 at the beginning of the timespan. For subsequent initializations, this argument does not hold, and a larger uncertainty (e.g., derived from the bitemporal comparison) should be assumed.

Furthermore, we have added a comparison with a Kalman filter model that only uses position, and one that uses velocity and position, and present the results, as well as discuss them (see new Section 4.1)

4. *I find Figure 1 difficult to understand at first sight, would be good to also include a photo or point cloud colored by height as a fist impression of the site.*

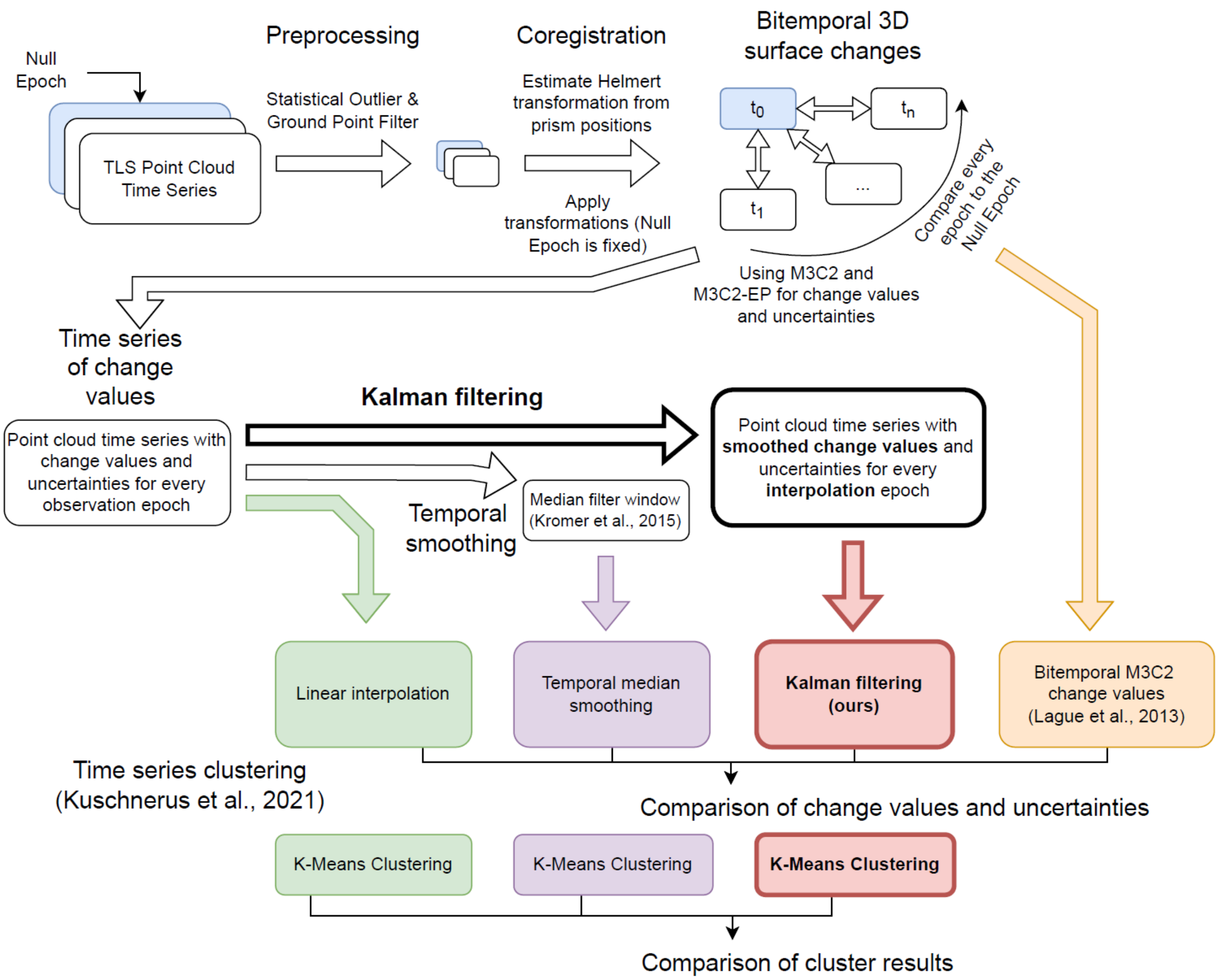
We have improved the visualisation of the study site in Fig. 1 following the suggestion:



5. *Sect. 2.5, from line 263 onwards and Sect. 2.6: Looks like you are trying many different methods at once. Why not choose one clear method of (1) pre-processing (2) uncertainties with M3C2-EP (3) smoothed time series with Kalman filter (4) feature extraction (5) clustering. Where you choose one set of (most relevant) features for the clustering. In my understanding the main goal of this paper is not to compare different features and/or clustering algorithms but to introduce the two previous steps and highlight the improvements that they yield during clustering. Possibly also add the workflow chart as shown in the readme file that comes with your code on Github. The comparison with clustering on unfiltered time series and without feature detection can then be part of the discussion*

In line with our response to comment #1, we have decided to completely remove the extraction of engineered features to focus on the Kalman filter as the method we are presenting in this paper. This results in a complete rework of the discussion section.

As suggested, we have further added a workflow graphic to the method section (Figure 3):



6. *It would be good if some or all of the features in Tables 1 to 4 could be illustrated on 2 two 3 example time series (e.g. RTS-SE-0.5) of representative locations, one in the excavation area, one in a rockfall gully and some third one.*

As we have removed the feature calculation, this comment no longer applies. However, we included more plots of smoothed time series to show how the Kalman filter deals with surface processes at different velocities (Figs. 4, 5. 6).

7. *Results section and Figures 4-8: some subsections are needed here to make it more accessible. The first part deals with the results of processing steps 1-4. From line 300 it goes into feature extraction and comparison/visualization of different features. As mentioned above: better focus on one set of features. Then a subsection called 'comparison' is needed. Here it should be clear, what is 'your own presented method' and what other methods do you compare with. I would suggest to only compare end results and not different steps in between. Make a selection out of Figures 4 to 8 and show the most relevant results.*
8. *P14, Fig 7 is hardly discussed, discuss if relevant, or omit the figure.*

We have restructured the results and discussion sections and introduced the following subsections: “Results – Impact of model and parameter choice” and “Results – Comparison with other methods”. The figures have been reworked.

9. *The testing framework you mention in the 3rd paragraph of Ch.1 we applied to two epoch TLS data iDeformation Analysis of a bored tunnel by means of Terrestrial Laser Scanning, Rinske van Gosliga, Roderik Lindenbergh and Norbert Pfeifer, IASPRS Volume XXXVI, Part 5, Dresden 25-27 September 2006*

We have added the reference to the introduction section.

10. *For significant change extraction, also the terrain roughness could be incorporated as a variance value, compare Kraus, K., Karel, W., Briese, C., & Mandlburger, G. (2006). Local accuracy measures for digital terrain models. The Photogrammetric Record, 21(116), 342-354*

We have added these considerations to the introduction:

The variance of point distances to the fitted surfaces is typically used as a measure for the uncertainty in the estimated position in elevation models (Kraus et al., 2006) and M3C2 change values (Lague et al., 2013).

11. *In Figure 2, the velocity and accelaration could also be omitted, (or shown once, in a separate image) as the graphs have a lot of details now.*

We changed the graphs to show only change value but included different rates of smoothing into a single plot instead.

12. *Line 90: data set from 2020*

This line was removed as we changed to the 2021 dataset.

13. *Line 97: not clear what kind of comparison is meant here. Comparison of uncertainty estimation, clustering approach or change detection in general?*

We have clarified that we refer to the extracted change clusters:

Second, we show how different smoothing methods for topographic point cloud time series influence the results of clustering to derive change patterns in the observed scene.

14. *Figure 1: caption discusses II and III, but these are not in the figure. No reference to Fig. 1b in the text, suggested to add to section 2.3*

With the different area of interest, we changed the figures and made sure the areas are correctly labelled. Fig. 1b is now referenced:

While most of this snow melted again by 2021-11-15, an avalanche led to accumulation of snow, which persisted throughout the observation period. This deposition can be seen in Figure 1b on the bottom right in red.

15. *Line 124: ‘methods […] are based on a part of recorded data […]’ -> Methods are applied to the data, tested on the data, or similar.*

This sentence was removed during revision.

16. *Line 133 – 139: part of 2.3? not clear why it is mentioned here before M3C2-EP has been introduced.*

We believe that bitemporal M3C2-EP can be considered a preprocessing step in this analysis, and therefore opt to keep it with the data. However, we recognize that a detailed explanation of M3C2-EP is required for readers not familiar with the method (also following the request by Reviewer #2,). We therefore now explain the method in Section 2.3.

17. *Sect. 2.4: not explicitly mentioned in the text what are t and x_t*

The explanation on t and x_t have been added:

It allows the integration of measurements over time into a state vector x_t describing the system at a specific point in time t.

18. *p8, r205: what exactly is the "uncertainty in point cloud distance obtained by M3C2-EP?*

We have added a more in-depth explanation on M3C2-EP, which includes the obtained uncertainties:

This error propagation is carried out by taking the mathematical model of how point cloud coordinates are obtained from transforming measured quantities (range, azimuth angle, and elevation angle) and computed quantities (transformation parameters). This model is then linearized by a Taylor approximation. Following Niemeier (2001), the uncertainty in the target variables (C_{xyz}) can then be estimated by multiplying the linear approximation model in the form of the Jacobian matrix A onto the covariance matrix of the input quantities $C_{r\phi\theta}$ from the right, and the transpose of the Jacobian from the right, respectively (cf. Eq. 1).

19. *P8r227: no variance of position in null epoch: would it not be more realistic to involve a measurement error?*

We think that an initial value of zero is appropriate for the variance in the null epoch's position, as the actual change is by definition zero, and not distributed around zero (in a Bayesian sense). We do note, however, that the Kalman filter typically takes a few epochs to converge, and that the results during this phase are less reliable. One could argue that this ought to be represented by a higher uncertainty. In reality, it does not make a difference: the filter result is bound by the first distance observation, and the uncertainty of this observation includes the uncertainty of the null epoch (as it is the uncertainty of the difference between these two epochs). We added this point to the discussion:

[…] In line with this consideration is the choice of uncertainty at the beginning of the process. At the start of the time series, the displacement must be - by definition - zero, and we, therefore, assign an uncertainty of zero to this initialization. This also ensures that all trajectories pass through the point at 0 at the beginning of the timespan. For subsequent initializations, this argument does not hold, and a larger uncertainty (e.g., derived from the bitemporal comparison) should be assumed.

Figure 2: where is the example core point located in the area? You could mark the location in Figure 1, so it is visible what kind of change to expect.

We have added markers for the locations of the displayed time series in Figure 1.

20. *From Section 2.4 it was no clear to me why RTS was discussed, later I found out that this was actually used to obtain (additional) results, this could be better announced.*

We now better explain what the final resulting time series is in Section 3.3:

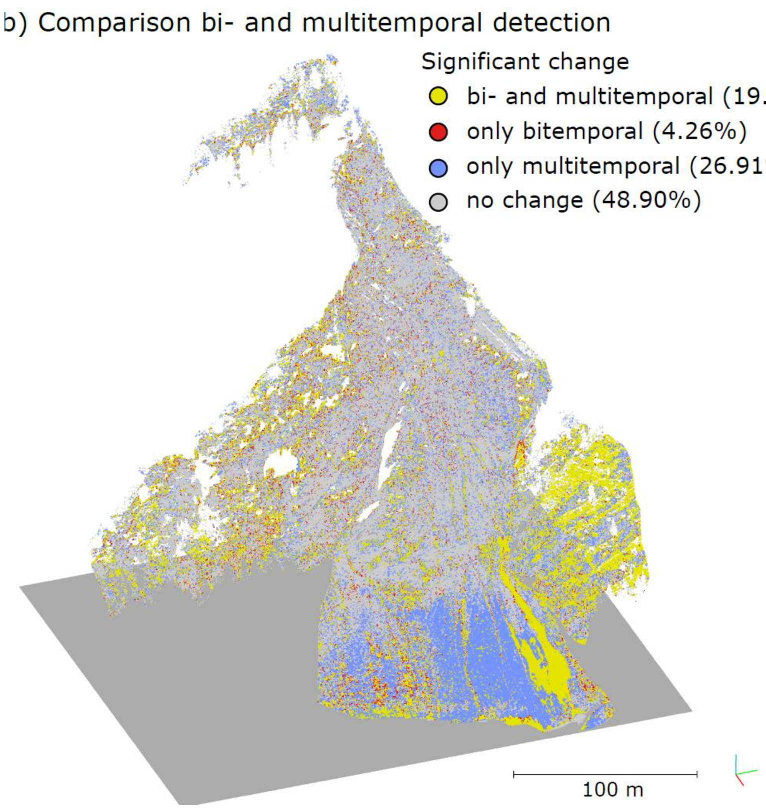
While the Kalman filter is an "online" method, which allows updates by adding new data points, we consider a post hoc analysis and assume that all measurements are available at the time of analysis. This allows us to not only consider previously observed change values at a given location, but also future ones. That way, we can make use of the full 4D domain of the dataset.

21. *Is the last series of features (last paragraph of Section 2.5) necessary for this paper? In my opinion these could be omitted and focus could be on the features in Tables 1 to 4.*

We agree with the reviewer and have omitted the extraction of features, as explained with comment #1.

22. *Figure 3b: red points (only bitemporal change) are difficult to see and it is a bit confusing that the borders of the area have the same color.*

We have improved the coloring of Figure 8b:



23. *Caption Fig.4: -> "At grey points no significant change could be detected'*

Fig. 4 was removed from the updated manuscript.

24. *Caption Fig. 5: 'residuals: between what and what?*

Fig. 5 was removed from the manuscript.

25. *P16: "Fig 8 depicts a bird's eye view": this is the same view as all the other figures, and is not focusing on the lower part of the slope: wrong figure?*
26. *P16: there is no 'II' in Fig.1.*
27. *P16: I could not find Fig. 8c unfortunately*

We corrected the error in the manuscript layout (Figs. 7 and 8 were partially mixed up in the text due to a double identifier in LaTeX.).

28. *P19: "Recovered velocities and accelerations": I would use the word "estimated"*

We have followed the reviewers' suggestion throughout.

29. *P19: what do you mean by "manually extracted features"? I though all work was automated?*

Yes, the work was automated, yet the design of the features was done by hand (i.e., the selection of which features to extract). "Engineered" would have been a better term, however, we have removed the feature extraction altogether.

---

2. Review by Dr. D. Lague

1. *while the general idea of smoothing temporally the signal to improve the signal to noise ratio and the detectability of potentially smaller events is interesting (but not new in itself, e.g., Komer et al., 2015), I find that the paper do not demonstrate clearly the benefits of the complex Kalman filtering and its associated error model compared to previous approaches (Kromer et al., 2015) or more simpler approach such as bi-temporal analysis, or simple linear interpolation when regular temporal sampling of data is needed. The paper also lacks information and discussion on key aspects of the clustering approach., and use a very complex set of features derived from time series without clear justification and in-depth analysis of the results.*

In our paper, we present the Kalman filter as an alternative approach for point cloud time series data, in line with the concept of error propagation for temporal data. In the revision, we take care to better compare the obtained results with the state-of-the-art. We therefore now compare four different methods: bitemporal M3C2, kMeans Clustering on the raw change time series, smoothed time series (Kromer et al., 2015) and ours. We are convinced that the restructured results and discussion sections show a clearer picture now.

2. *The introduction is very good, but the result section is not well organized and many figures are not informative, or of limited quality, or not fully described in the text. A simple figure illustrating the principle of the method is also lacking.*

We have added a figure showing the principle of the method, as well as the comparisons we perform (see response to Reviewer #1, Comment #5)

Furthermore, we have strongly restructured the results and discussion.

3. *I think it is possible for this MS to be published at some point, but it needs very significant work to better present the results (both in terms of figure quality and analysis), better demonstrate the*

advantages of the method compared to simpler approaches, which could be done for instance on synthetic data. Also focusing the clustering approach on one method with a meaningfull set of features that would be easy to interpret would make the paper simpler.

In line with the comments by reviewers #1, we opted to remove the feature extraction and clustering based on these features. Instead, we now focus on kMeans clustering on the time series directly, as presented by Kuschnerus et al. (2021).

We have also added a synthetic dataset to this comparison, showing the performance in the same way as for the real data.

4. *Kalman filtering is used for predicting system states that vary smoothly. I do not see why it would apply to an excavator removing rocks, rockfall or climate-driven surface erosion given that these events tends to be highly discrete in time, and thus inconsistent with a smooth evolution. Moreover, the use of a backward pass limits the ability of the method to accurately detect the timing of an event. Why can't you simply define local velocities or accelerations from a finite-difference calculation (e.g., v=(P(t2)-P(t1))/(t2-t1) where P is your point location. You'd get a better temporal localization of events, at the expense of a lower detectability of small events. As for the clustering, as you mention in the discussion, a simple linear interpolation would suffice.*

We have introduced the study site as a rockfall-affected slope, but actually focus on the surface processes acting on the debris at a much slower and less discrete pattern. To remove the dominance of the excavation works, we have removed this area from the analysis and focus on the upper slope. At the end of the 2021 time series, which we are now using, snowfall and an avalanche occur. We agree that these events are not ideally modelled by the Kalman filter, yet want to show that they *can* be represented. The discussion now includes a section on how such sudden changes, once detected, could be modelled by re-initialization of the Kalman filter state.

As already discussed in our first Authors' comment, the backwards pass does not limit temporal detection accuracy. It ensures in a global optimum (in a Bayesian filtering sense) for minimization of acceleration changes. For any discrete event, these changes will therefore be equally distributed to before and after the event, resulting in optimal detection in time. The issue with numerical differences, however, is the drastic decrease of the SNR, rendering a maxima/minima detection difficult.

We now also include clustering on the time series smoothed following Kromer et al. (2015), which allows to interpolate over data gaps. The results are compared in the discussion section.

5. *The part of the paper using features extracted from time series is quite superficial. There is no discussion on which features are actually important in the clustering.*

We have removed the feature extraction from the manuscript, following all reviewers' suggestion.

6. *The choice of the number of cluster is not discussed at all. This is a critical point as the issues of over or under-segmentation are critical in clustering, and not addressed at all here in the paper.*

In our analyses, we found that from a certain point onwards (e.g., 10 clusters), the clusters on the hill slope do not change much anymore. We assume this is due to core points in the adjacent forested areas having a higher variability, and therefore over-segmentation occurring in these areas. We have added a section on the cluster choice explaining these considerations and showing exemplary results. Note that the cluster numbers are significantly lower in this revision than previously, as the area of interest was reduced and now excludes most forested areas.

7. *Some figures have poor quality, with details that are difficult to see (fig. 7,8)*

We have taken care to improve the quality of all figures.

8. *Figure 8 seems to miss 1 sub-pictures that is mentioned in the text, but not shown*

Unfortunately, references to Figure 8 in part actually referred to Figure 7. We have taken care to avoid this in the revision.

9. *The results section does not have a clear organization, and many figure are not described and exploited to their full extent.*

We have restructured the results section and improved the figures as well as their discussion.

10. *The discussion do not address the choice of the number of clusters. Also it does not discuss the limits of the approach, when it comes to the reduced temporal resolution, the need to choose a state variance, and the general complexity of the approach. In particular, it is ill adapted to detect precursors which can be critical in real-time monitoring because of the smoothing effect of the signal. In general, I find that there is a tendency in the discussion to assume that because the method is more complex, it is better. However, it is not clearly supported by the evidence shown in the paper. Apart from one figure (fig. 3), the superiority of the new approach compared to a simple bitemporal approach is not clearly demonstrated (and I actually have doubts on the results of fig. 3). The benefits in terms of lower level of change detection are not obvious and would benefit from synthetic data simulation to evaluate quantitatively how each method is able to recover a known change.*

As per comment #6, we have added a section on the number of clusters. In the discussion, we now include a more in-depth review of the limits of the approach and that the choice of state variance should depend on the type of observed and investigated process.

We state the limitation of pre-cursor detection in an online system, which is in line with a probability-based approach for filtering outliers.

We provide a detailed response to the results of Fig. 3 in comment #30.

Also, we have added a synthetic dataset to the comparison to show a clearer picture of the differences between the individual methods (Section 4.3).

11. *The discussion does not describe the benefit of Kalman filtering compared to the Kromer et al., (2015) approach.*

We have added a comparison to Kromer et al. (2015) in Section 4.2.

12. *Also the fact that the clustering is done in 2D, while the core points are inherently 3D is not discussed.*

As we have removed the feature-based clustering in the revision, we are now only working with kMeans clustering following Kuschnerus et al. (2021). This clustering is not spatially based, but only operates on the time series. However, as spatially contiguous areas are subject to the same geomorphological processes, spatial clusters form.

The results of the clustering were shown in 2D views solely for visualisation purposes. We have ensured to add this critical information throughout the methods and results sections.

13. *The introduction is very good, and states clearly the objectives of the paper with thenecessary references to previous work.*

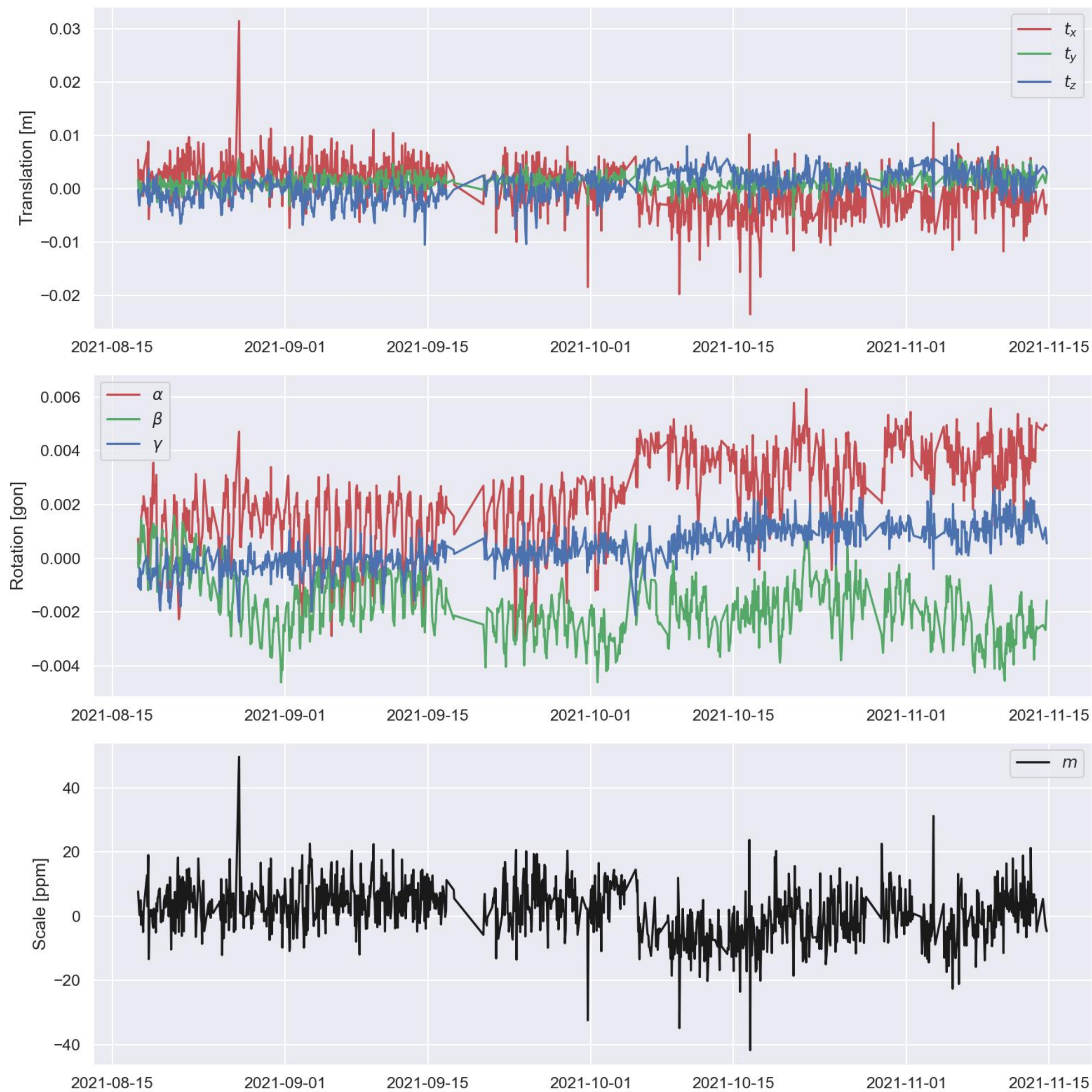
14. *L113 : please specify the typical point spacing. This is a critical information that is missing to understand why you are not able with bi-temporal analysis to detect a 5-10 cm change with a sensor with 0.005 m precision !*

We have added the point spacing information in the data section. In the original manuscript, the large uncertainty in the alignment was a result of the ICP alignment, which mainly used stable patches in the valley (i.e. closer to the scanner than the actual target area on the slope). As a result, angular uncertainties at longer ranges increased, which yielded level of detections of ~8 cm derived using M3C2-EP.

In the revised version, we are using retroreflective prisms mounted around the area of interest, avoiding extrapolation of the transformation parameters. Still, the bitemporal level of detection is at ~5 cm for many locations. This is a combination of (a) remaining alignment uncertainties, (b) ranging uncertainty (0.005 m is One-Sigma, and t-tests at $p<0.05$ are approximately Two-Sigma) and (c) angular uncertainty resulting from the beam divergence. The footprint (One-Sigma) at 800 m range is approximately 5.4 cm in radius. At an incidence angle of 30 deg (typical for the study area), 2.7 cm remain just from the angular uncertainty.

15. *L127 : could you explain why you needed to realign the data if the sensor was on a fixed pillar, and arguably, all scans were acquired in the same reference frame ? or is it specifically related to using M3C2-EP and estimating the alignment uncertainty ? In that case, mention it in the text.*

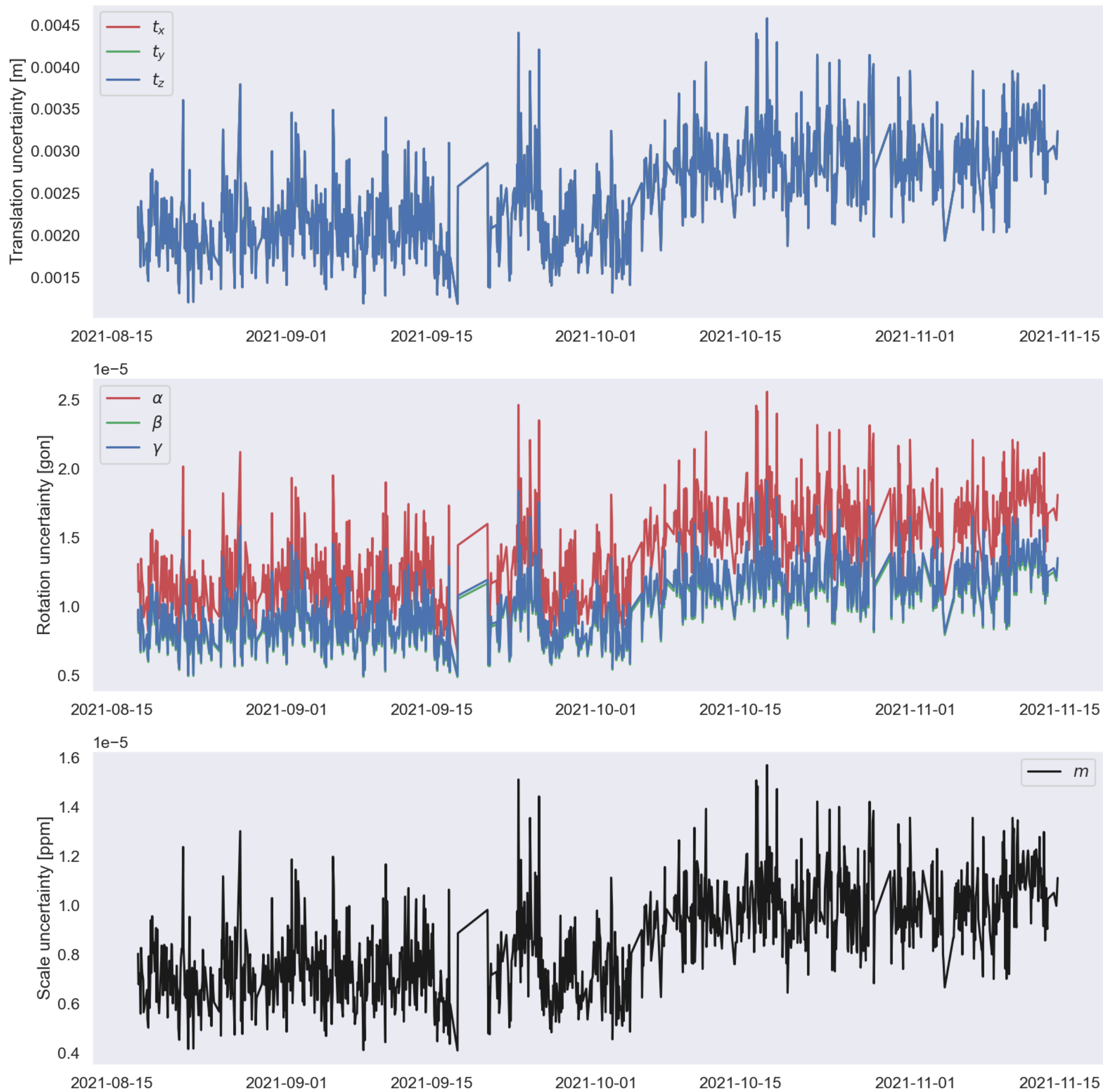
Similar to TLS time series from fixed positions in, e.g., Kromer et al. (2017), Williams et al. (2018), and Anders et al. (2019), we also observe that our data is not perfectly aligned, even though the sensor was installed on a fixed pillar. Especially with the lowered level of detection through temporal smoothing, the effects from imperfect alignment become relevant. With the now-included prism extraction, we can show the course of the transformation parameters over time, which clearly shows (a) daily (temperature-dependent) patterns and (b) a general increase of the yaw rotation (i.e., the axis of the main screw fixing the TLS to the pillar):



(for the order of rotations, refer to Joeckel, R., Gruber, F. J. (2020). Formelsammlung für das Vermessungswesen. Germany: Springer Fachmedien Wiesbaden.)

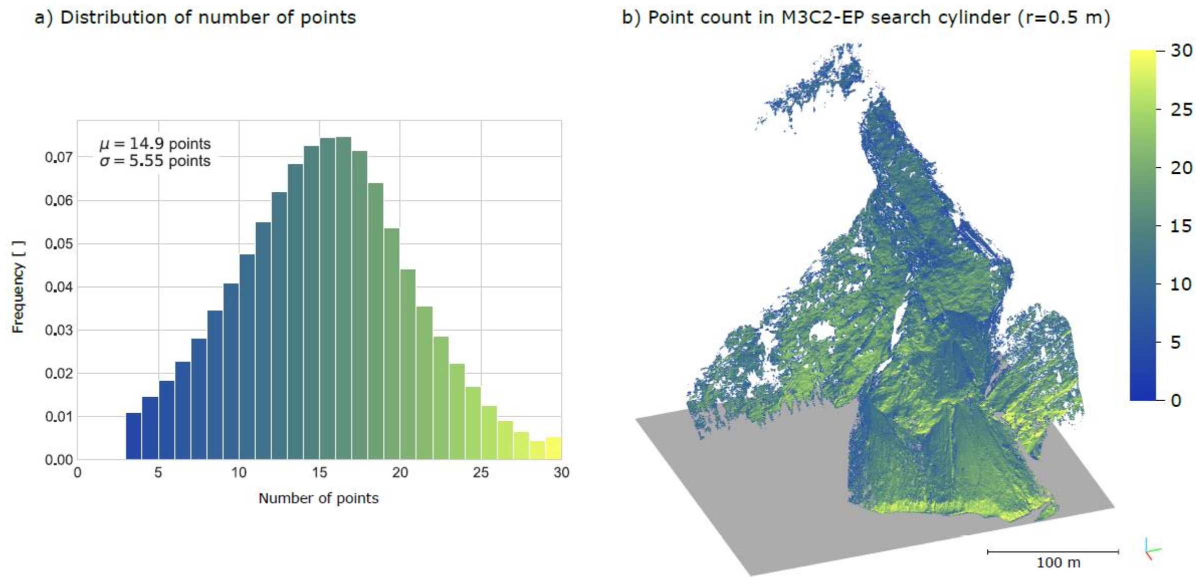
The maximum effect of singular rotations (α at the end of the time series approx. 0.005 gon) at 800m range corresponds to 6.3 cm tangential shift. For more information on the prism extraction, refer to Gaisecker, Schröder (2022): RIEGL V-Line Scanners for Permanent Monitoring Applications and integration capabilities into customers risk management (http://www.riegl.com/uploads/tx_pxpriegldownloads/Whitepaper_RIEGL_DMT.pdf, last accessed 2022-07-22).

Above are the actual transformation parameters – the corresponding uncertainties are shown here, where the maximum rotational uncertainty of 2.5e-5 gon corresponds to 3/10 mm at 800 m range.



16. L136 : why 0.5 m ?

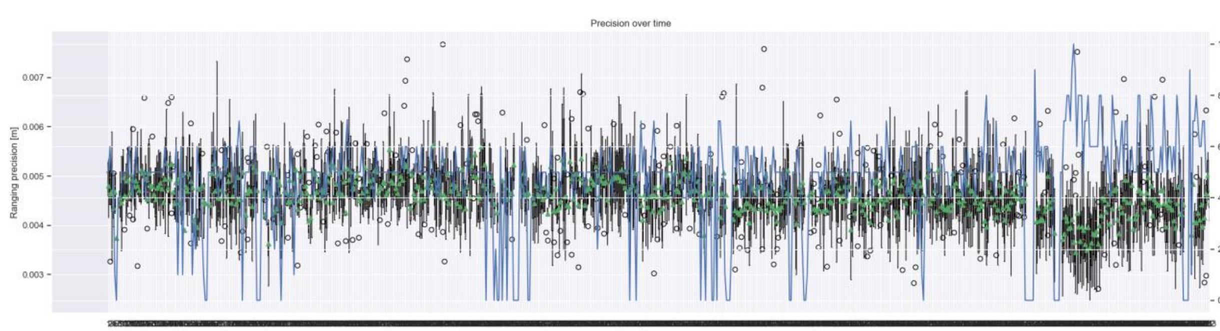
We choose a search radius that allows for meaningful statistics (n ~ 15) in most parts of the dataset (see also the updated Figure 2 showing the point count in the M3C2 cylinder).



17. *L138 : are there any correction for temperature effects on ranging measurement (that start to be significant over 800 m) ? Also, I suspect the 0.005 m ranging accuracy is certainly not at 800 m distance ! have you better constrained on the actual ranging accuracy at 800 m ?*

Linear dependent temperature effects are considered by including a scale parameter in the Helmert transformation (cf. Williams et al., 2018), which we use.
The TLS manufacturer only provides 0.005 m as a ranging accuracy value ("tested under RIEGL conditions"). Additionally, there have been some studies investigating the accuracy of TLS over longer ranges (e.g., Fey and Wichmann, 2016), who have noted that the increase in ranging uncertainty mainly stems from the finite footprint and non-zero incidence angles.

To get an estimate of the actual ranging uncertainty, we employed a data-driven approach: For each retroreflective prism, we selected points within certain amplitude thresholds. We then calculated the quality of a plane fit to these points: As the prism represents the highest-energy reflector in the footprint for many points, the ranging component should be the same – and any deviations from it, i.e., from the best fitting plane, are due to ranging uncertainty. We applied this method for every epoch separately, resulting in a histogram of ranging uncertainties (over the different prisms) for each epoch:



We now use the mean values of ranging uncertainty derived for each epoch as the ranging error component in the error propagation. The obtained values are all very close to, or even below, the nominal accuracy of 0.005 m. Any linear effects in scale, which are not well obtainable through this approach, will be covered by the Helmert transformation.

Overall, it is important to consider here that at 800 m range, the influence of the angular uncertainty is about an order of magnitude larger than the one of the ranging uncertainty.

18. *Section 2.1 : could you give an estimate of the mean point density of the scans ?*

Point density strongly varies with range and incidence angle – we therefore have added a figure showing the point count in the M3C2 search cylinder (Fig. 2b), which also shows this variance.

19. *Section 2.2 seems like an introduction to the algorithm you present to analyses PC series, with a bit of state of the art in spatio-temporal clustering. Then subsequent section (M3C2-EP etc…à) should be sub-section on this one (2.2.1, 2.2.2….) otherwise section 2.2 by itself is not part of the method.*

We have restructured the data and methods sections accordingly.

20. *L174 : while I know M3C2-EP, I suspect it would help less specialist readers to have a bit more explanation on the extra steps needed for the uncertainty calculation in M3C2-EP, and the benefits compared to the standard uncertainty model of M3C2. No need to go into too much detail, but the M3C2-EP paper being a tough one to read, it would help to have a self consistent paper*

We have gladly added more introduction to M3C2-EP in the method section:

This error propagation is carried out by taking the mathematical model of how point cloud coordinates are obtained from transforming measured quantities (range, azimuth angle, and elevation angle) and computed quantities (transformation parameters). This model is then linearized by a Taylor approximation. Following Niemeier (2001), the uncertainty in the target variables (C_{xyz}) can then be estimated by multiplying the linear approximation model in the form of the Jacobian matrix A onto the covariance matrix of the input quantities $C_{r\phi\theta}$ from the right, and the transpose of the Jacobian from the right, respectively (cf. Eq. 1).

[…]

While M3C2 itself also quantifies the uncertainty of the estimated bitemporal differences, this estimate is derived from the data and influenced by non-orthogonal look angles, and object roughness within the M3C2 search cylinder (Winiwarter et al., 2021).

21. *L179 : you should specify how k is going to be defined, as it needs to be manually chosen for k-means clustering.*

We have added a section (Section 4.4) on the choice of cluster classes (cf. response to comment #6)

22. *L180 : I'm roughly familiar with Kalman filtering owing to airborne LiDAR data processing, however, I suspect many readers won't, and they may have trouble following this part. Maybe a sketch of the basis of kalman filtering applied in your specific case would help.*

We have added a paragraph on the basic idea of Kalman filtering to the methods section:

We present the use of a Kalman filter, which can be used to incorporate multiple observations (in our case the change values for each epoch, quantified along the local 3D surface normals using M3C2-EP, cf. Section 2.1) and obtain predictions about the displacement at arbitrary points in the time series, analogous to the median smoothing. A main advantage of the Kalman filter is its potential to consider uncertainties both in the inputs, allowing for observations of different qualities to be combined, as well as in the output. Here, an uncertainty value for each point in time can be estimated, which allows for statistical testing of the obtained smoothed change values (as in the M3C2 for bitemporal change values).

Typical applications of Kalman filtering include sensor integration settings, e.g. in the integration of GNSS and IMU (inertial) measurements, when the target trajectory is smooth. A famous application was the guidance computer in the Apollo missions (Grewal and Andrews, 2010). Kalman filters are commonly used today in trajectory estimation, e.g. for direct georeferencing of airborne laser scanning data (El-Sheimy, 2017). In our case of 4D point cloud change analysis, not all changes are smooth. The limitations arising thereof are discussed later (Section 5).

L220 : see major comment 1. I really have trouble reconciling the smooth nature of Kalman filtering with the highly discrete nature of erosion events

We agree that the Kalman filter is not ideal for discrete events, cf. our response to comment #1. We discuss this limitation in the revised manuscript (Section 5) but also point out that there is no apparent effect in further processing (i.e., clustering).

23. *L241 : this sentence is not clear to me. How do you turn the 4D data into 2D ?-> ok I get it, it's an introduction to the subsequent section. Maybe rephrase to make things clearer.*

We have rewritten the section on feature extraction and do not focus on the creation of 2D maps further.

24. *L254-258 : making sense of the attributes in relation to the expected geomorphic processes would be great. For instance, it is not obvious at this stage why the total curvature is importante (compared to a more straightforward measure such as cumulative change) ?*
25. *L263 : FFT on a signal which is have periodic pattern does not really make sense especially if you're not detrending the signal and using filters to account for the finite dimension of the time series. Maybe theres's a reason I don't see, but in that case it seems important to give a little intuition as to why you suggest such features. Wavelet analysis might make more sense as it combines temporal location (when an event happen) and frequency analysis (~ duration of an event), but it's hard to come up with simple integrative features to be used for subsequent clustering.*
26. *L275 : I do not see at all, how the clustering based on the features, which are potentially very numerous and do not contain any relation to "physics" or "drivers" of cliff erosion (precipitation, local cliff geometry, ….) can actually lead to a more "physical interpretation" than analyzing the estimated change directly. The authors need to back this statement.*

We have removed extraction of engineered features as we think it would go beyond the scope of the now extended paper – thereby comments #25-27 are solved.

27. *L277 : you should mention that the number of clusters need to be specified, in case nonspecialist readers think that unsupervised clustering is just pushing a button and getting a result. An you should explain here, how you choose the number of clusters (as you did for GMM. It's critical.*

Solved with our response to comment #6.

28. *L293-299 / figure 2 : the description of the figure needs to be improed. You're first sentence stating "appropriately filters daily effects" gives a sense that 0.005 m/day² is initially the best value, while indeed you choose 0.05 m/day². Also for such an important parameter, your search of the optimum is rather qualitative. I don't think plotting acceleration helps at all. You do not discuss the occurrence of clear oscillations in the signal prior to the change. Are they real signal, or variations of the scanner position (+-2 cm, that's huge) ? It seems that another criteria for choosing σ is that it must be large enough to not trigger a detection for these oscillations.*

With the new dataset and the improved alignment, we have chosen new examples to present in our manuscript. They demonstrate that a value of 0.005 m/day^2 does appropriately filter these daily effects. Unfortunately, it also removes parts of the signal, and has large residuals around the 2020-08-22 12:00 mark. The second choice, 0.05 m/day^2 still removes the daily signal, whereas the third choice (0.5 m/day^2) overfits the daily noise.

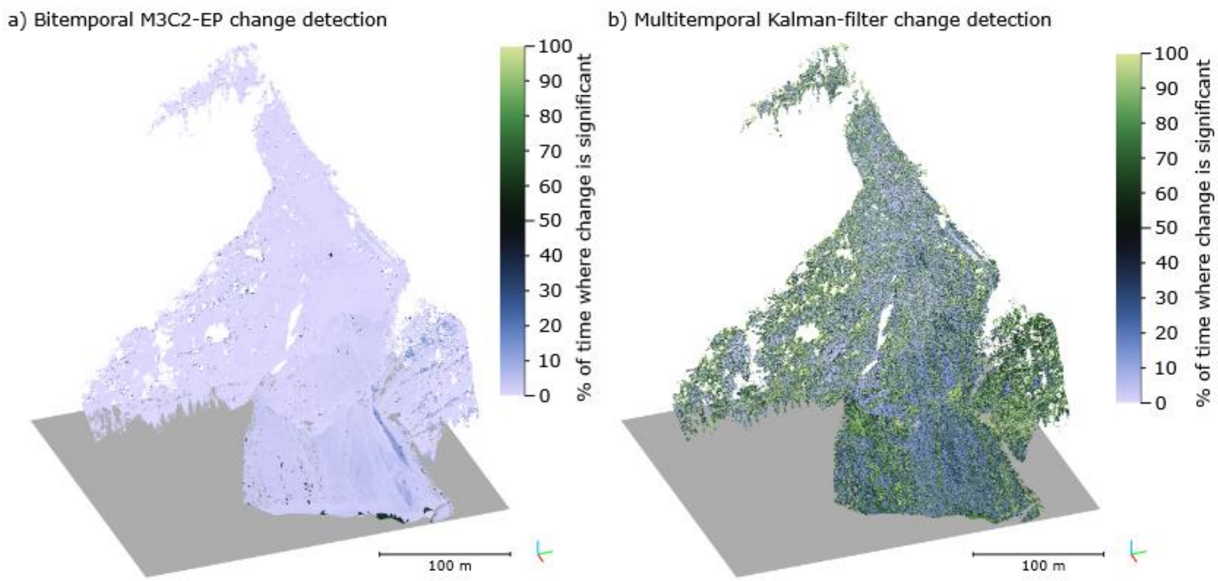
In the updated manuscript, using different locations allows us to better discuss these effects in much more detail.

29. *Fig3 : It is hard to tell without having the information on the point cloud spacing, but I'm extremely surprised that a bi-temporal analysis is not able to detect change in the channels where distances more than 5 cm are measured by the multitemporal approach. It might be that the ICP registration has an issue on the two epochs used for testing for significant change and translated into a large registration error increasing the LoD. But 5 cm over a few cm² should be detected easily with a sensor with 1 cm ranging error (an estimate at 800 m) and a 1 cm registration error. It's very odd.*

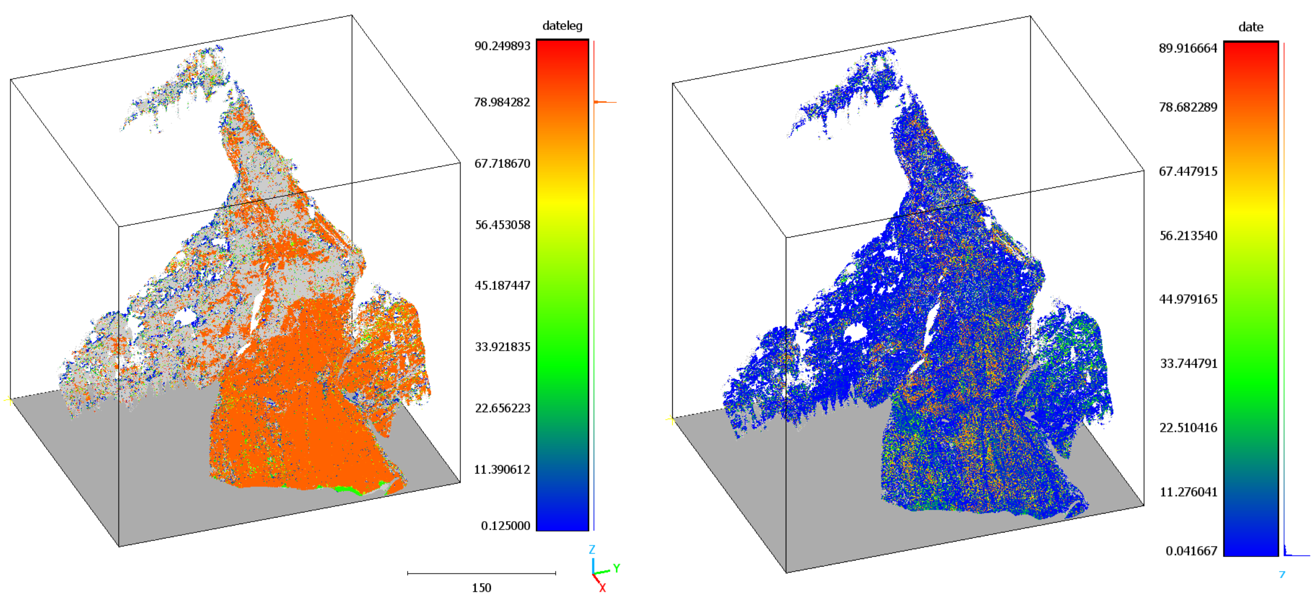
See our response to comments #14 and #17. The positional error of sensed points are between 2 and 3 cm (std. dev., due to the footprint) and the alignment error contributes another 0.5 cm (see #15). The alignment uncertainty needs to be considered as systematic, whereas the positional error can be considered as random. With 30 points in the search radius, the resulting level of detection (according to the Equation given in Lague et al., 2013) is 2-3 cm, which is in line with the current results. Applying bitemporal M3C2 gives (on average) levels of detection between 1 cm and 10 cm (mean = 6 cm) when assuming 0.5 cm alignment error.

30. *Fig 4b: use also greyed color for the area with non significant change to facilitate comparison with 4a. It would be interesting, following fig. 3 to show if the onset of change detection differs significantly from the bi-temporal approach compared to the multitemporal approach. This would better emphasize the interest of your method.*

We have added a plot that shows the relative amount of time of the whole time series where significant change is detected for both the Kalman approach and bitemporal M3C2 (Figure 7):



The Kalman approach detects different surface-change inducing processes as significant, therefore a plot showing the difference in time between the first detection of significant change does not show the desired result. In the plot below, the bitemporal M3C2-EP mainly detects the snowfall event 79 days after the start of the time series (orange), whereas the Kalman approach detects small-magnitude change processes much earlier (i.e., within the first few days).



Left: point in time (in days) of first significant change using bitemporal M3C2-EP, right: point in time (in days) of first significant change using Kalman filtering (order 1 model, σ=0.02m/day)

We opted to include only the first of the discussed plots in the manuscript, as we believe the second one being misleading due to different change processes being detected.

31. *L318 : which "value" ? it's not clear*

The "value" referred to the RTS smoother residuals. We have removed the figure in response to comment #33.

32. *L321 : I fail to grab the interest in showing fig. 5. What do you learn and how important it is ?*

Without more in-depth analysis of the different surface processes acting on the terrain, and with the removal of the area of anthropogenic change from the dataset, Fig. 5 becomes less informative and obsolete. We have removed this part from the analysis.

33. *L325 : Ok, figure 6 tells us you use 50 clusters in one case and 100 in the second case, but it is not even mentioned in the text and you do not justify your choices. It's a critical point to discuss. Also why can't you simply create a linear interpolation between two epochs to fil in the gap for your clustering ? it would solve your problem of temporal spacing without having to rely on a complex Kalman filtering.*

We have added a section on the choice of cluster numbers (see our responses to comments #6 and #21)

34. *Question: is there correspondence between the large pink area and the area where no change is detected ?*

Referring to Fig. 6, the large pink area is closely resembling the area where no significant change has occurred. These results have changed in the revision.

35. *L330 & fig 7 : this description of figure 7 is insufficient. It is not up to the reader to analyze the results. You must highlight much more key results, otherwise it means the figure is useless (actually I'm not convinced it's actually useful, because the quality of the visualization is poor, and we don't know why you 150 clusters and not a lower or larger number).*

The clustering approach based on engineered features was removed from the manuscript.

36. *L334 : the whole subsequent section is really hard to follow.*
37. *L340 : How dependent are the so-called "distinct" features on the number of cluster. As you are using a large number of cluster, you are artificially producing many features. But this may simply result from over-segmentation. Here the choice of your number of cluster should be discussed in depth.*

Obsolete, as we have removed this section and the clustering analysis using engineered features.

38. *Figure 8 : the visualization is extremely poor, and this figure is not really usable.*

This figure was removed and we have reworked the figures in the updated manuscript to improve readability.

39. *L343 : Fig 8c is missing*

Unfortunately, some references to Fig. 7 were incorrectly labelled Fig. 8. We have thoroughly checked the updated version for such errors.

40. *L345 : this last statement seems to contradict previous sentences in the very same paragraph. So in the end, your method detect the same things than the others. What is really its interest beyond interpolating slightly (which could simply be done with linear interpolation between 2 epochs…) ?*

We have more clearly pointed out the benefits of filtering for time series analysis, especially with respect to error propagation. Kalman filtering is – to date – the only known time series analysis method incorporating geodetic error propagation through time. These are results from the comparisons to the other methods on both the synthetic and the real dataset.

41. *L355 : smoothing of the time series is debatable advantage, as it decreases the temporal resolution of event detection. Also "predicting" future states when it comes to natural environments seems hardly feasible, especially when considering rockfalls or rain-related erosion which are by nature not really predictable.*

We have carefully revised the discussion to shift the focus to (a) post-event analysis and (b) prediction of the development of uncertainty for non-discrete events. We agree that the Kalman-filter is ill-formed to detect rockfall precursors:

The Kalman filter is ill-suited to represent sudden changes, as caused by discrete events of mass movement. However, gradual motions such as rockfall precursors as studied by Abellán et al. (2009), could be detected well even without the backward pass of the RTS smoother, given that repeated observations show such a trend. In such use cases, the reduction in the Level of Detection is especially crucial.

42. *L362 : velocity and acceleration are meaningfull for estimating a plane trajectory as it by nature smooth, however it is not useful, and probably not desirable, for interpolating the occurrence of discrete erosion events.*

We agree with the reviewer that this is the case for discrete events, but it is in fact useful when investigating gradual movements, i.e. changes represented in the time series over multiple epochs. We discuss how such types of surface activities could be represented in the Kalman filter in Section 5.