We greatly appreciate the two helpful reviews, and will use the comments to improve the manuscript. Below, we provide a detailed response to the reviewer comments, outlining how we will revise the manuscript. We have responded to each comment (except for some very minor ones regarding typos), with the referee comments in blue italics and our responses in normal text.

Referee 1

The short communication introduces a workflow to process multi-temporal data for accurate change detection although no GCPs are available. Thereby, images from multiple survey campaigns are processed at once. Afterwards, the orientated images of the individual surveys are split to retrieve the corresponding point clouds of each campaign for change detection. The idea is simple but very effective. The manuscript is well-structured and easy to follow. The results are presented comprehensively and support the introduced method. However, some issues remain regarding the explanation of the approach (especially terminology) that should be addressed in a revised manuscript. Furthermore, the authors should consider the Time-SIFT publication by Feurer & Vinatier (2018) because it describes a similar approach more detailed for applications to archival imagery. Please, see below for some detailed comments.

Thanks a lot for pointing us towards the Feurer and Vinatier paper. We had not seen this, and it is indeed quite similar to our approach. It is a little bit of a relief to see that we are not the only ones to have thought of this nontraditional approach. We will certainly cite it and add it to our discussion of previous work.

P1L23: It is not clear what the authors refer to with camera optical parameters. Are these the interior orientation parameters. If yes, it should be mentioned that the GCPs are also used to refine the parameters of the exterior orientation besides the interior parameters.

Yes, this was unclear. We will restate that GCPs are used to georeference the model and to improve the calculation of both camera interior parameters and camera positions and orientations.

P1L29: It might be better to refer to dGNSS instead of dGPS as also other satellites can be used for geo-referencing.

Good point, will be changed.

P2L33-35: Are these control points tie points or ground control points? If they are GCPs, where does the reliable/accurate 3D information come from? And if they are tie points, I would avoid the term control points.

Peppa et al., 2019 refer to them as pseudo-GCPs. We will use that instead of the term control points.

P3L65-71: This paragraph seems to be a little bit off-topic if it is left as it is. A better explanation why these challenges are displayed should be provided. For instance, why is the changing appearance of the cliff relevant? Does that potentially impact feature detection and matching? Furthermore, a final statement might improve that paragraph, as well, highlighting that this study at the cliff is a very suitable study to demonstrate the usability/necessity/benefits of the authors' approach. Although, this intention is probably meant in the paragraph it might be suitable to mention this explicitly.

Good point. We will remove the part about the changing appearance, as it's not really relevant. Then we will explicitly state that these challenges are the reason why this is a good test case.

P3 chapter Methods: I would suggest to include sub-headings for data acquisition and co-alignment processing to improve the readability.

These will be added

P3L75-76: Did the authors also consider check points as an independent reference of the reconstruction accuracy? With 14 and 12 GCPs this should be possible.

We did not do this, as we rely on the accuracy study in Cook, 2017, which was conducted at the same site using the same control points, to estimate uncertainties (as stated in the text).

P3L94: I thought, only the Mavic Pro was used for data acquisition (but also a Phantom is mentioned here)?

The Phantom 3 was used for the Daan River surveys (this was mentioned in line 73).

P4L97: What is the unit for the reconstruction uncertainty? According to Agisoft, the reconstruction uncertainty somehow relates to the base-height ratio. But how is the reconstruction uncertainty calculated?

This value has no unit, as it is the ratio between the variation in the direction of maximum variation and the variation in the direction of minimum variation. We are reporting the parameters used in Photoscan for completeness, not because they are particularly important for the method. For the case of surveys with a lot of oblique photos, we have found that filtering by reconstruction uncertainty is the best way to remove tie points that are clearly erroneous. Other users may clean and optimize the survey differently; it doesn't really matter in terms of the co-alignment method. Because this is not a particularly important step, we don't think it's necessary to give an introduction to how Photoscan calculates it.

P4L98: How is the adaptive camera model fitting working? What is the difference to the approach without adaptive fitting?

As with the comment above, we are not sure that this manuscript is the place to explain the details of Photoscan's methods, but can provide a reference to the user manual.

P4L99: The fine registration in CloudCompare is done via ICP (iterative closest point) fitting. Maybe, it might be preferable to state the actual performed algorithm rather than the tool name.

This will be changed.

P4L105: The alignment optimization is actually also a bundle adjustment, however considering some refined parameter settings and/or referencing information. Thus, this might be rephrased to avoid confusion of the reader.

This will be rephrased.

P4L116-118: Is it possible to express these differences between both change maps in numbers, e.g. considering the average of deviations between both maps? This question would also be relevant for the

Rügen analysis. Furthermore, did the authors also check accuracies at check points? They might be helpful to assess how well changes are detectable with the reference in general.

We can provide the average change for each map, but we feel that the histograms shown in the figures are more informative. One issue is that some of the differences calculated are real, so a smaller average difference does not necessarily mean a better result. Unfortunately, we cannot directly compare the two Daan River change maps because the models without GCPs are warped relative to the models with GCPs, so the change maps don't align with each other.

For Rügen, as mentioned in the text, we have no independently measured check points. If we had the ability to have such points, then we would also be able to use GCPs and wouldn't have the need for this workflow. At the Daan River site, we rely on the accuracy study in Cook, 2017, which was conducted at the same site using the same control points, to estimate uncertainties.

P4L124-125: However, this depends on how the models are aligned. If GCPs or stable areas are used, I am not certain if this statement still holds. Of course, if ICP is used than these distortions can lead to difficulties in the alignment (depending on how strong these distortions are).

If alignment involves just rotation and transformation, then distortions will prevent good alignment of the whole model no matter what method is used. Perhaps alignment is not the best term to use here, as we are talking about only transformation and rotation of dense point clouds or meshes; we can see how this can be confused with camera alignment. We can substitute co-registration for alignment.

P5L128-129: I am not sure if I understand that sentence correctly. Changes between 1 and 2 m are common at the observed cliff on Rügen? Thus, the noise in the data is higher than the common changes at the cliff?

We will rephrase this. In this comparison, up 1-2 meters of change are erroneously detected in many stable areas, indicating that real changes of this magnitude would be below the level of detection.

L123-129: Maybe the entire paragraph can be rewritten to improve clarity regarding model related distortions and issues due to alignment approaches.

Hopefully this will be more clear by using "co-registration" rather than "alignment." But basically, if models are distorted relative to each other, then they will not perfectly fit together no matter what method you use.

P5L130-131: Maybe it is worth to extent the explanation that the simultaneous alignment of all campaigns leads to the circumstance that the highly spatially correlated errors (James et al., 2017), which also depend on the image observations (i.e. tie points), are potentially situated at the same locations in the individual models (because image orientation across surveys are constrained to the same tie-points) and therefore mitigated during point cloud differencing.

Yes, this is exactly what we are trying to convey – that the models contain errors, but they are consistent across the different surveys, so they don't influence the change detection. We will add a sentence to say this explicitly.

P5L153: I would not state that edges are the issue but rather areas outside the tie point region.

In these surveys, it does seem to be more an issue of edges, as the extents of the dense and sparse clouds are the same. The points near the edge are generally only seen on two or three photos, and they are only seen on the same edge of these photos, so they are less well-constrained (compared to points which are visible on many photos and which are located at a range of positions on different photos).

P6L160-163: Maybe this statement should be separated more clearly from the previous because another aspect is discussed. The first aspect is referring to too strong changes of the surface and therefore failing to find matches and the second refers to changes of the entire surface but remaining a general similar appearance and thus falsely retrieving matches.

Yes, we can clarify this. The method requires matches (the first aspect), and it requires accurate matches (the second aspect). We can't comment too much on false matches because we didn't find any, but we raise this as something to watch out for.

P6L164: What do the authors refer to when they are talking about scaling between numbers of photos?

This refers to the nonlinear increase in processing time when more photos are added – doubling the photos will more than double the time required for point matching and camera alignment.

P6L164-168: I have a little bit difficulty to understand that sentence. Do the authors mean that with each new campaign all the campaigns have to be re-processed?

Yes, this is what we mean. Any previous campaigns that you would like to compare to the new campaign must be reprocessed.

P6L168-169: Might it not be possible to only compare from one survey to the next to avoid increasing the processing time with each new survey, although this might be less favorable for error propagation? Maybe it might worth testing in a future study how well campaign to campaign processing performs compared to reprocessing everything.

Of course, and this is what we typically do with the Rügen surveys. But this means that for four surveys A, B, C, and D, you will need to do all of the processing three times – A+B, B+C, and C+D. So if it takes 10 hours to generate the dense cloud for survey A, you will have to do that twice. And you can't compare A vs. C or B vs. D. Basically, the issue is that when you conduct a new survey, you can't re-use any of the models you have previously generated to compare with it. This is quite different from the typical workflow, where you create a "final" model that can be compared to any future models.

P6L177-178: Maybe the combination of both is most suitable (e.g. as discussed by Feurer & Vinatier, 2018). Align all campaigns in one workflow (this might also improve general model accuracy as more image observations will be available) and scale/georeference the whole project with GCPs (from just one campaign).

This is a great suggestion, thanks. We will add a sentence about this, with a citation to Feurer and Vinatier.

Referee 2

This is an interesting study on the possibility of improving the comparative accuracy of multiple surveys by co-processing the image sets when stable areas can be found and matched in a particular area. Rather than the workflow itself (which is hardly a proper workflow but just a modification of the standard SfM pipeline), I found the greatest merit of this work drawing the attention to this co-alignment possibility, that in many cases my be discarded or overlooked and can help to improve the quality of the results. My main comments to this work are the following (please check the annotated pdf for specific comments throughout the manuscript): 1. I would strongly suggest to include in the manuscript title the main limitation of the workflow. i.e. the presence of stable areas. The authors have acknowledged this in the limitations and conclusions sections and should be specified in the title since is a major requirement. 2. The authors are too focused on the geomorphological settings (cliffs, rivers and such), which is not bad, but a relevant part of the SfM community works on more artificial environments such as agricultural settings where hardly stable areas can be found. How applicable would be the co-aligment in these cases? A quick literature review could give the authors a general view of the types of scenarios in which the SfM approaches are being applied and maybe they could comment in more detail to what extent their method is feasible to be applied.

The limitations of the method are made quite clear in the manuscript; we disagree that it's necessary to include mention of stable areas in the title and feel that it would make the title long, awkward, and too detailed. We will add a sentence to the abstract specifying that the method relies of the presence of stable areas.

The goal of this manuscript to introduce a solution for people working in environments where ground control is impossible or very difficult to obtain. Thus, we are not focused on places like agricultural settings or other artificial environments, as traditional methods work just fine in these areas. Plus, since we have no experience with or data from such areas, we have no idea how well the method will perform. It may be that peripheral stable infrastructure such as buildings, roads, fences, etc. may provide sufficient common tie points, and maybe not. It may also vary depending on the location. We hope that people working in different settings can give it a try and evaluate whether it works for their particular area.

Número: 5 Autor: Asunto: Resaltado Fecha: 17/07/2019 7:31:01 1. Please justify the selection of these two study areas 2. Rephrase the sentence to be shorter and better structured

We will rephrase this. The appropriateness of the study areas are justified in more detail in the area descriptions below.

Número: 6 Autor: Asunto: Resaltado Fecha: 17/07/2019 7:30:56 Include company and country or reference.

Agisoft is the company, so that is already there. We have never seen the country listed in other publications that use Photoscan. We did neglect to provide the version number, so we will add this.

Número: 8 Autor: Asunto: Resaltado Fecha: 17/07/2019 7:38:04 No information of this study area has been included as a figure. Please provide a map and picture if appropriate, similarly to the Rugen site.

We felt that this information wasn't necessary, as it is available in a previous publication and isn't critical for interpreting the results.

Número: 9 Autor: Asunto: Resaltado Fecha: 17/07/2019 7:33:21 Autor: Asunto: Nota adhesiva Fecha: 17/07/2019 7:33:40 Why not describing this first, being the primary area?

We use the Daan case as a kind of proof of concept, where we apply the method to a more traditional type of survey in an area where traditional methods are possible, so we present it first. We describe it first because we present it first in the results. We consider Rügen to be the primary area, as it is the setting where traditional methods can't be used and our co-registration workflow is really necessary to get useful change detection results.

This was made after or before the fine registration and M3C2 algorithm?

This was done after the M3C2 calculations – the steps were done in the order that they are presented in the text.

Número: 2 Autor: Asunto: Resaltado Fecha: 17/07/2019 7:47:23 Can you provide the results when both surveys are processed independently (no co-alignment)?

We will add this analysis to figure 3.

Página: 5 Número: 1 Autor: Asunto: Resaltado Fecha: 17/07/2019 7:56:33 I recommend discussing in more detail each of the results in Fig. 5 and 6.

We are not sure what additional detail is expected. To discuss the patterns of cliff collapse and retreat on Rügen is far beyond the scope of this manuscript, particularly since it is a short communication and not a full research paper.

Número: 2 Autor: Asunto: Resaltado Fecha: 17/07/2019 8:00:42 This table must be improved: using capital letters at the beginning of the column titles, better structure, etc... The references to the study sites are confusing, please include always the main name of the site and then the particular name of the area. Why not being consistent with Daan river results first and then Rugen? It is confusing.

We will reorganize the table and clarify the study site names.

The examples provided by the authors are typically focused on geomorphological settings which include stable areas. What about other scenarios such as agricultural settings where SfM is being frequently used? I would recommend revising other settings in literature where UAV SfM is being extensively used and comment the feasibility of the workflow accordingly.

As we say above, we don't have the data or experience to evaluate the feasibility of the approach in such settings, and it may vary depending on the details of the agricultural setting and the surveys. Such a discussion would only be speculation, which we don't feel is very useful. We are open about what is required to make the approach work, and we hope that readers can evaluate their own areas on the basis of that.

Número: 2 Autor: Asunto: Resaltado Fecha: 17/07/2019 8:03:23 What does this mean? Different surveys may have different target accuracies depending on the aims of the study

Survey-grade accuracy is a term that is commonly used in the surveying community and typically refers to accuracy on the order of cm or better.