The Cryosphere Discuss.,
https://doi.org/10.5194/tc-2018-18-AC2, 2018

Interactive comment on "Darkening Swiss glacier ice?" *by* Kathrin Naegeli et al.

Kathrin Naegeli et al.

kan1@aber.ac.uk

Received and published: 17 April 2018

We would like to thank the referee 2 for this careful and detailed review. We appreciate the in-depth comments and are convinced that thanks to the respective changes, the manuscript will improve substantially.

In response to this review, we elaborated the role of supraglacial debris in more detail by: (1) manually delineating medial moraines and areas where tributaries separated from the main glacier trunk and debris has become exposed to obtain a complete supraglacial debris mask based on the Sentinel-2 image acquired in August 2016, and (2) applying this debris mask to all data and, thus, to exclude areas with debris coverage from all consecutive analyses. Additionally, we further developed the uncertainty assessment of the retrieved albedo values by providing more information about the

datasets used as well as their specific constraints and uncertainties that may result thereof in a separate sub-section in the methods section. Based on these revisions, the conclusions will most likely be slightly adjusted, however, stay in line with the original aim of our study to investigate possible changes in bare-ice albedo in the Swiss Alps based on readily available Landsat surface reflectance data.

Below we respond to all comments by anonymous referee 2. The responses (normal font style) are following the *referees' comments* (displayed in italic font style) directly. The corresponding revised sentences in the manuscript are given in quotation marks.

My main concern is that the method claims to assess the changes of albedo on bare ice. However, most of the ablation areas of the glaciers under consideration exhibit large medial moraines and changing debris covers on sides and terminus, capable to affect strongly the albedo signal. It comes that much of the changes found to be significant enough appear obviously related to changes in the spatial distribution of debris rather than a "darkening of bare-ice" surfaces as suggested by the authors. I am however concerned that the study gives only little acknowledgement to the fact that the target has changed in most instances, but rather insist on the fact that significant areas of bare-ice are perceived to darken. I find this insufficiently supported, if not misleading in view of the data and results provided. Removing from the analysis all areas where the significant change in albedo may be associated with the surface not being bare-ice at some point of the chronology has the potential to change substantially the message of this study, and I believe can compromise the significance and robustness of its current conclusion.

We agree that many areas with strong albedo changes are somehow linked to supraglacial debris cover (e.g. medial moraines). However, a clear definition of bare-ice or debris-covered ice is not existent in glaciology as the transition is smooth and strongly site-specific with many intermediate stages of dirty ice. Thus, with the remote sensing data used in this study a clear separation of bare- and debris-covered ice is difficult. However, to acknowledge the referee's concern, we manually delineate medial moraines as well as areas where tributaries separated from the main glacier trunk and debris has become exposed. These debris-covered areas will be consecutively excluded from the study area. Other clearly debris-covered areas of the glacier (e.g. the tongue of Unteraar or Zmutt Glacier) had already previously been excluded from the analysis.

Specific comments

P1L17: Is it "to" or "at higher altitude"? I suppose the authors mean "to" but since the meaning would be different with either preposition, it is important to correct this.

Changed.

"Increasing air temperatures and changing precipitation patterns provoke snowlines to rise to higher altitudes and thus a spatially greater exposure of bare-ice surfaces."

P2L13: remove "necessarily"

Deleted.

P2L15-18: Although it is true that the use of MODIS data to retrieve surface albedo on mountain glaciers is complicated by a relatively coarse resolution, it is not "unsuitable" as the authors claim. Since Dumont et al. (2011) the use of MODIS data to measure temporal variations in glacier surface albedo has proven to be successful to inform about changes occurring on alpine glaciers.

We added the use of downscaled MODIS data in long-term albedo studies and respective references.

"To date, most long-term studies either used point data from automatic weather stations located in the ablation area of a glacier [Oerlemans et al., 2009], coarsely spaced satellite data from the Moderate Resolution Imaging Spectroradiometer (MODIS) [e.g. Stroeve et al., 2013; Mernild et al., 2015], downscaled MODIS data [Dumont et al., 2012; Sirguey et al., 2016; Davaze et al., 2018] or other remote sensing datasets [e.g. Wang et al., 2014] to infer trends in ice albedo."

P2L28: The study focuses on glaciers which exceeds 5km2, yet Table 1 reports three glaciers smaller than that.

We changed the respective wording.

"All of them are characterised by a surface area of roughly 5 km^2 and larger, (. . .)."

P2L33: I don't think "of high accuracy" is meaningful or well used in this sentence. I suggest removing it.

Deleted.

P3L1: I believe there is a contradiction between the emphasize put on the fact that the study seeks to characterize changes in surface albedo of "bare ice only" (mentioned 8 times in the abstract and 6 times in the introduction) to discover now that the medial moraines have been kept in the analysis. A number of glaciers such as Aletsch exhibit relatively large medial moraines that are expected to affect strongly albedo estimates. Furthermore, over the time period of the study, it is reasonable to think that the location of the moraines may vary within or beyond the 30-m pixel resolution of the Landsat images (not mentioning the modulation associated with co-registration variability between images), thus convoluting their spectral response with that of bare ice, and potentially creating perceived changes in albedo value.

As mentioned in our response to the referee's general concern at the beginning, we delineated all medial moraines and areas where tributaries separated from the main glacier trunk and debris has become exposed based on the Sentinel-2 image from August 2016. These debris-covered areas will consecutively be excluded from the investigated study area. Furthermore, we agree that the location and extent of medial moraines over the study period might change. Thus, we will strengthen our statement about this process affecting albedo changes in both directions (decreasing and increasing albedo values along medial moraines) in of the discussion section in the revised version of the paper.

P3L5: Landsat 7 sensor should be named ETM+ throughout the text, tables and figures.

Changed.

P3L5: Although it leaves little doubts that the authors are referring to the Level 2

surface reflectance product LEDAPS for Landsat 5-7 for and LaSRC for Landsat 8 OLI, I would suggest these products are named accordingly to their rightful designation for clarity. Since those surface reflectance are not correcting for topography nor shadow effects and that this can compromise the accuracy of albedo estimates, I believe the product being used deserve a more comprehensive description in relation to the context of the study. In particular, it would be important to review what is corrected for in those products, the expected accuracy and limitations of those corrections, and how suitable it is for the study at hand in the present context.

We added more information about the used Landsat products, in particular on the level of correction and the expected accuracy of retrieved surface reflectance values.

"We used the Landsat Surface Reflectance Level-2 science products of the USGS for Landsat 5 and 7 (TM/ETM+) and 8 (OLI) as a basis to obtain broadband shortwave albedo (see Section 3.2). For Landsat TM and ETM+, the product is generated from the specialized software Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS), whereas the Landsat OLI product is based on the Landsat 8 Surface Reflectance Code (LaSRC). These data products consist of six (TM/ETM+) or seven (OLI) individual spectral bands in the wavelength range of around 440 nm to 2300 nm, with slight deviations of the individual band widths for the specific sensors. Detailed information about these products can be found in [Masek et al., 2006] for Landsat TM/ETM+, and in [Vermote et al., 2016] for Landsat 8, as well as in the product guides provided by the USGS. In the context of this study, it is important to mention that both products are neither corrected for topography nor shadow effects. Claverie et al. [2015] investigated the accuracy of retrieved surface reflectance values based on the LEDAPS algorithm by inter-comparing the product with data from the Aerosol Robotic Network (AERONET) and MODIS data obtained on the same day. This comparison showed good results overall with the poorest performance in the blue band, which is known to have the greatest atmospheric sensitivity [Vermote and Kotchenova, 2008]. Most importantly, they found no trend or significant year-to-year variability, suggesting this data product to be highly valuable for temporal analysis. Similarly, Vermote et al. [2016] analysed the performance of the Landsat 8 surface reflectance product, concluding with high correlations between the MODIS and OLI surface reflectance values, with worst results found again for the blue band, and a general improvement of Landsat OLI surface reflectance product over the ad-hoc Landsat TM/ETM+ LEDAPS product."

P3L5: As the authors must know and have experienced, all ETM+ data post May 2003 are affected by the SLC failure leaving significant areas of the glaciers missing observations. I must admit I am somewhat surprised that this is not mentioned once, despite a number of images being affected as illustrated in Figure 3, and despite this possibly having noticeable consequences on the computation of mean bare-ice albedo in Figure 4. To me, the inclusion of those data in the analysis would need a specific test to quantify how much the mean albedo may be affected by SLC-off data. This could be done relatively easily with a SLC-on dataset from which a SLC-off would be simulated using the mask of a SLC-off epoch.

Thanks for this friendly reminder to add some details about the SLC failure in the ETM+ data post May 2003. We added respective information in the data section of the manuscript.

Furthermore, we evaluated the referee's concern about the impact of the SCL failure on mean bare-ice albedo by using the ETM+ SLC-off data for Findelengletscher from 12.09.2011 as a mask and the three scenes from 12.08.2000 (ETM+, SCL-on), 13.08.2003 (TM, SCL-on) and 30.08.2015 (OLI, SCL-on) as test data to obtain sensitivity values. This analysis revealed that the impact of the SLC failure on mean bare-ice albedo is negligible with a difference of 1.2 to 2.2% (e.g. 12.08.2000 SLC-on mean bare-ice albedo 0.204 versus SLC-off mean bare-ice albedo 0.209 indicating a difference of 2.2%).

"Missing data in some of the Landsat ETM+ data, generated due to the scan line corrector (SLC)

failure post May 2003, also occurs in our albedo retrievals (e.g. 09.09.2004 in Figure 3). We tested the impact of the SLC failure by simulating missing data for three scenes with an intact SLC for Findelengletscher. SLC failure resulted in slightly higher mean bare-ice albedo values (1.2 to 2.2%, e.g. 12.08.2000 SLC-on mean bare-ice albedo 0.204 versus SLC-off mean bare-ice albedo 0.209 indicating a difference of 2.2%), which is a negligible impact."

P4L8: "All reflectance data were downloaded"

Changed.

P4L10: The authors relied on their own cloud classification approach. Since cloud masks are provided as part of the LEDAPS and LaSRC products albeit having known issues, the authors could explain and discuss the reason why they favored a custom algorithm, and how this was assessed to deliver more useful images. I did download a number of the LaSRC images used in this study and left wondering why a custom cloud detection algorithm was needed here, especially given the relatively limited number of images and the visual assessment being made to select those less of not affected by clouds at all.

As mentioned above, the provided cloud masks are known to have several limitations, especially concerning bright surfaces such as snow and ice. Furthermore, based on our assessment the provided masks usually strongly misclassify medial moraines and lateral debris along glaciers as clouds too. We thus used Spectral Angle Mapper as an independent classification algorithm for cloud classification. To justify our reasoning, we added some clarifying statement in the manuscript.

"As cloud masks provided with the science products are known to have certain limitations, in particular for bright targets such as snow and ice, but also misclassified medial and lateral moraines, we used a semi-automatic classification approach based on the Spectral Angle Mapper (SAM, [Kruse et al., 1993]) implemented in ENVI to detect and delineate clouds obscuring the glacier surfaces."

P4L16: the citation should be Liang (2001), not 2000, same in P19

Changed in all places.

P4L18: It would be desirable to use the same notation as Liang (2001) with αi used to represent narrowband ground reflectance of TM/ETM+ in band i.

Changed.

P4L19: the symbol bn is not the spectral band number but should be the narrowband ground reflectance of TM/ETM+ in band i. As commented above, using α would make it consistent with Liang (2001) and general understanding of the quantity used.

Changed.

P4L22: The narrow to broadband conversion assumes ground reflectance on horizontal surfaces. The LEDAPS and LaSRC products account for topography only in terms of the control of elevation on atmospheric effects, yet the control of topography on the modulation of irradiance on varying slope and aspects, hence on the measured satellite radiance in mountainous regions, is ignored. In other words, the surface reflectance products assume a flat surface and the relative geometry between irradiance, the target, and observation direction is not accounted for, let alone the higher orders

of effects related to the topography such as terrain reflected irradiance and modulation of the observed reflectance by the BRDF of the target under the said geometry. Although this effect may be limited on relatively flat glacier tongues and/or when considering only variations rather than absolute values of albedo through time, this and the disconnect between the variable topographic setting under study and the relevance of the empirical formulation of the albedo cannot just be ignored by the authors.

We agree that the fact of the missing topographic correction in the Landsat science products LEDAPS and LaSRC must be mentioned. This is now done (see answer to comment above). Moreover, we will include a statement about the impact of neglecting BRDF based on the investigations from Naegeli et al. [2017b] in the uncertainty assessment.

P4L22: I find the claim that the albedo products are of "very high accuracy" and "deviate by less than <0.001 on average from more sophisticated approach" not well informed, if not misleading. It suggests that albedo values retrieved by the empirical method, on the basis of atmospherically but not topographically corrected data may be 100 times more accurate than reference albedo estimated from an albedometer. I can find the sentence in Naegeli et al. (2017) claiming such "accuracy". I however understand this is the mean difference between Liang albedo applied to synthetic APEX data compared to a rigorous albedo derived from the full spectrum. At least in this case the source of data is all APEX and thus topographically corrected. All in Naegeli et al. (2017) suggest that albedo derived from L8 can hardly meet this target even on average. If to comment on the accuracy of this formula (and provided the limitations associated with the use of non-topographically corrected data and other sources of uncertainty is justified), a more useful number would rather be the standard error of this comparison between APEX and APEXLiang, namely 0.11. This is far more realistic with what can reasonably be expected from albedo retrieval from satellite.

We agree that the statements made so far are not sufficient enough to clarify the performance of the narrow-to-broadband (NTB) formula by Liang [2001] to the reader. We will thus extend the explanations about the NTB formula performance, based on the detailed assessments made by Naegeli et al. [2017], and incorporate them into the new written sub-section “3.5 Uncertainty assessment” at the end of the methods section.

However, concerning the referee’s suggestion to use “the standard error of this comparison between APEX and APEXLiang, namely 0.11”, we note that this seems to have been a misunderstanding. The number 0.11 refers to the standard deviation within this albedo product and, thus, not a comparison between APEX and APEXLiang. Experiment 3 in Naegeli et al. [2017] investigates the comparison between APEX and APEXLiang, and reports a mean glacier-wide albedo of 0.41 $\pm$ 0.17 for L8 and 0.41 $\pm$ 0.18 for L8Liang for Findelengletscher and 0.15 $\pm$ 0.03 for L8 and 0.17 $\pm$ 0.09 for L8Liang for Glacier de la Plaine Morte. Moreover, we refer to the discussion and conclusion made in Naegeli et al., [2017] about the performance of the NTB formula by Liang applied to Landsat 8 data: “From Experiment 3 (Table 3) it becomes evident that the Liang [2001] formula provides good estimates for glacier-wide mean albedo for both glaciers and all datasets, whereas the Knap et al. [1999] formula is subject to stronger deviations.”

P4L25: the definition of such hard threshold on albedo derived from nontopographically corrected data is disputable. Again and related to my previous comment, this should be reviewed and/or critically discussed in view of variability of albedo associated with the topographic control on irradiance.

We agree that a hard threshold is questionable. However, as our surface type evaluation is a multi-step classification, the initially set threshold has a limited effect on

the final result as it only provides a zero-order classification that is afterwards adjusted for grid cells for which the surface type is unclear.

P4L9: “(0.25<α<0.55)” not “(0.25>α<0.55)”

Changed.

P7L9: snapshot is singular

Changed.

P7L10: The use of “ideal case” is value-laden. This sentence should be rephrased more objectively.

Replaced “in an ideal case” with “at most”.

P8L2: The authors seem to consider equally epochs one (sometimes two) year(s) apart and sometimes just a few days apart to compute trends. I don’t think this is an acceptable methodology. When several images are available for the same summer on a grid cell, I think only one estimate should prevail (or an average maybe) for this year and trends only derived from a “yearly” record. This also applies to the linear regression through all available data points. Note also that it is recommended that the Mann-Kendall test be conducted with only one data point per time period (see Chapter 12 in Helsel and Hirsch, Statistical Methods in Water Resources, U.S. Geological Survey, Techniques of Water-Resources Investigations Book 4, Chapter A3).

Thank you for this valuable comment. We admit having used multiple scenes within the same year for three instances. To account for this statistical mistake, we've now selected one scene per year only based on optimal quality (largest bare-ice area exposed) and will derive trends from "yearly" records as suggested.

P8L14: the areas of nodata in ETM+ are not to be referred as striping (which is usually associated with radiometric calibration) but the Side-line-correction failure that occurred in May 2003.

Rephrased.

"Missing data in some of the Landsat ETM+ data, generated due to the scan line corrector (SLC) failure post May 2003, also occurs in our albedo retrievals (e.g. 08.09.2004 in Figure 3)."

P8L15: the 2004 image seems to be 8/9/2004, not 9/9/2004. All dates in Figure 3 must also be double-checked for erroneous date reported in Table 2 as stressed in a later comment.

Thanks for the careful review of this information. We apologize for the incorrect reporting of the respective scene dates. We double-checked, and changed where necessary, all dates in Table 2 and 4, Figure 3 and the entire manuscript.

P8L16: The authors acknowledge cloud shadow exemplified by one image in the upper accumulation area. They then claim that "bare-ice area is almost always well represented and inferred albedo is realistic thus allowing monitoring through time". I find this claim particularly vague and not supported. What do the authors mean by

"realistic"? In what sense would realistic "allow" monitoring? Just looking at both 1999 ETM+ images on Aletsch glacier, the 10/08 is unusable due to cloud cover on most glaciers of interest and in particular Aletsch, while the 11/09 image shows Aletsch tongue severely impacted by cloud shadow, not even to mention Fiescher glacier. Right from this first date, the quality of the Image begs the question about how much consideration was given to the "realistic" retrieval of albedo. Beyond the inherent accuracy of the albedo retrieval method which I think is misrepresented here, there is also no consideration of factors such as the different radiometric quality of TM/ETM vs OLI (whiskbroom vs pushbroom, 8b vs 12b radiometric resolution) or imperfect coregistration between image, and the potential of these factors on the quality of the albedo signal. To me the accuracy of the albedo retrieval method once all source of uncertainties are considered is the main factor that must be given consideration before seeking to make inferences on changes. So far and despite what the authors may suggest earlier, the albedo retrieval for a single pixel can hardly be proven to perform at better than 0.1 accuracy if not worst. A discussion of all environmental factors that may even degrade this further would be welcome. For example how cloud shadow as well as topographic shading whether in the upper reach or on the tongue of glaciers is handled by the methodology remains far too obscure. Yet the relative share of any of such phenomena on some dates could potentially result in variation of albedo. Furthermore, obviously the average bare-ice-albedo are computed on varying areas depending on the classification of the snow/ice limit. Naegeli et al. (2017) themselves reported on substantial variation in albedo up the tongue of Findelengletscher. This should be put into perspective with a mean albedo obtained from a variable number of "bare-ice" pixels at different years.

We agree that the treatment of cloud shadows in our analyses has not been explained clear enough. To account for very low signal values and thus unrealistic albedo values, we generally excluded albedo values < 0.05. However, we will apply a Spectral Angle Mapper (SAM) classification for cloud shadows and calculate, similarly as the cloud coverage column in Table 2, a cloud shadow percentage per scene. This will help to eliminate albedo values that are affected by cloud shadows and, thus, should not be included in the analyses.

Regarding the specific sensor differences, we note that, in particular the differences between Landsat TM/ETM+ and OLI, are incorporated in the respective surface reflectance retrieval algorithms LEDAPS and LaSRC. Furthermore, the used science products are Tier 1 products offered by the USGS. They are, according to the USGS, suitable for time-series analysis with well-characterized radiometry. Moreover, Tier 1 data are geo-referenced with ≤12 m root mean square error and inter-calibrated across the different Landsat sensors. We will add this information in the revised manuscript.

The analysis of mean bare-ice albedo over time is based on varying spatial extents per glacier indeed. We will perform a similar analysis based on a constant spatial extent per glacier (i.e. a minimal bare-ice extent) over the study period and compare the resulting time-series and trends to the reported data so far.

P10L5: In the context of the relatively large uncertainties being involved in the albedo estimates and interpretation of Table 3 and Figure 5, I find the phrasing that "trends were significant yet at low level" rather misleading. When focusing on those pixels with trends significant at 80% confidence (most of the bare ice), Table 3 reveals a symmetrical distribution (arguably Gaussian) of occurrences exhibiting either positive or equally negative trends. The trends themselves are estimated within the .05/decades or .1 overall magnitude over the approximate 20 years of the study; in other word, barely what we could hope as the uncertainty of the albedo retrieval method. To me this is rather showing that the level of detection (or signal to noise ratio) is simply not enough to be conclusive. Presenting this results and Figure 5 suggesting there are trends significant at least at 80% level everywhere is to me contrary to an alternate interpretation being that this level of detection is not suitable to evidence any obvious trend

at all. I think it would be fairer that the interpretation of this result and its significance stress the limitations of the methods rather than suggesting that there are indeed trends, potentially driven by some physical cause. At this stage, my interpretation of the results given the proposed methodology is simply that it can only be inconclusive.

We agree on the fact that some of the uncertainties are not yet fully considered and might force the general conclusions to be re-evaluated. However, the calculated significant trends, even though over large areas of the glacier only at a low significance level, are based on a profound methodology and should not be judged, neither as conclusive nor inconclusive. Moreover, our uncertainty analysis revealed that trends also persist with albedo values randomly perturbed with estimated uncertainties. The consecutive holistic discussion may reveal a clearer interpretation of the observed trends, but this is part of the discussion. We thus intend to keep the presentation of trends and their significance level in the results section as a finding of our analysis. However, we will highlight the impact of all uncertainties on the obtained trends in a more detailed way in the discussion of the revised manuscript.

Turning now to the small portions of pixels exhibiting high confidence (95%) of a trend. In this case, the negative trend apparently prevails. Simply looking at the spatial patterns of occurrences of those pixels in Figure 5 and 6 exemplifies what could be expected of the redistribution and spatial variation of debris and medial moraine on most glaciers. A clear example of this is Gorner glacier. Comparing the 11/9/1999 image to recent 25/9/2013 or 30/8/2015 immediately reveals that all those areas of "highly" significant darkening actually don't qualify as "bare-ice" but rather occur mostly due to spatial variation in the distribution of debris. Visually, there is an obvious widening of the medial moraine on the main trunk, retreat of the glacier front and that of tributaries that are exposing rocks, and obvious down wasting with lateral moraines material falling on the glacier. The fact this drives some pixels to appear as exhibiting a *strong decrease in albedo is in fact not so much reliant on the analysis presented in the paper to be revealed. More importantly, I am concerned that the paper may suggest a darkening of "bare ice" as if this was a subtle trend associated with increasing concentration of LAI on glaciers when the areas where the changes occur appear mostly to be those exhibiting a step-change in surface type altogether. In view of this, the way the authors elaborate in P10L10 on probable causes associated with such a progressive darkening such as the growth of algae and bacteria is to me far too speculative at this stage, and finally not supported by any new data in the present work.*

As outlined above, we delineated and excluded medial moraines and areas where tributaries separated from the main glacier trunk and debris has become exposed. Moreover, as mentioned in several other answers, the importance of debris redistribution and thus possible causes of bare-ice darkening will be discussed more critically in a new sub-section of the discussion section.

P10L12: Further to my comments above, I find that the authors far underestimate or seem to lower the role of the debris and changes in the distribution of moraine material in their observations of albedo changes. Based on my visual interpretation of the images used by the authors, it is obvious that there is more than a mere "possibility" than the areas of significant changes are associated to increase in debris cover and changes in medial moraines. I find the authors suggesting that this may only affect "certain grid cells" not supported by observations.

Please see our answers and suggested solutions/changes about this topic further above. Furthermore, we will extend the discussion about possible darkening effects in the discussion section.

P12L5: The authors claim that assessing the uncertainties associated with the albedo estimate is beyond the scope of this work. I don't think this is acceptable and as stressed several times in my comments, I believe this and would require a far more thorough consideration of the uncertainties than presently offered, for any inferences being made about processes at work to be deemed robust.

To account for uncertainties in the retrieved albedo values in a clearer way, we moved the section "uncertainty analysis" from the discussion section to the methods section. Moreover, we added detailed information about the datasets used in the data section (see answers above).

P12L8: What do the authors mean by "better result". This is unspecific, value-laden, and should be unsupported by a stronger argument.

Rephrased.

"Moreover, it became clear, that in contrast to Landsat TM and ETM+ data, for which a saturation problem over snow-covered areas exists, the most recent Landsat OLI data has a higher quality and albedo values can also be retrieved for snow."

P12L9: As demonstrated above, this study does NOT "focus on bare-ice areas" only. Furthermore, the data quality issue (e.g., SLC off, cloud shadows, saturation) can severely affect that albedo retrieval.

We will clarify our definition of bare-ice in the introduction. Furthermore, as elaborated in several previous answers, we adjusted our debris exclusion, added information about the impact of SLC-off, cloud shadows and saturation issues at several places in the manuscript.

P12L13: The authors claim that "manual checks" revealed low frequency of misclassified pixels compromising the albedo retrieval. My own check alone on the first two dates revealed immediately that the biggest glacier in the study (Aletsch) is severely affected by clouds and cloud shadow to the extent that I don't believe a realistic estimate of albedo could have been obtained across many parts of the glacier tongue. In view of this, one cannot be satisfied by the unsupported claim of the authors.

We agree that our statement was too vague. Actually, very low albedo values (α<0.05) were generally excluded from the analysis. This already minimized the incorporation of grid cells affected by cloud shadows. Furthermore, we will apply a SAM classification for cloud shadows to eliminate grid cells affected by cloud shadows before the albedo retrieval. We will update Table 2 to show details about this analysis.

P13L1: While it is true that using 2016 outlines of glaciers in the context of the current glacier demise would have reduced the dominance of ground becoming exposed dominating the change in albedo at the terminus and lateral moraines, Figure 5 and 6 still reveal that the significant changes in albedo remain associated to areas of probably thick debris deposition. It begs the question about what would have been the signal and conclusions of this work if the analysis had focus on a (conservative) mask of bare-ice, meaning a mask where pixels throughout the study period can be observed as free of thick debris. I think this is the main methodological issue that the authors should address to revise this work.

As outlined in previous comments, we manually generated a mask of supraglacial debris with a relevant thickness that was used to exclude these grid cells from all con-

secutive analyses. However, we note that a clear distinction between debris-covered and dirty ice does not exist and that the transition between the two is smooth and often unclear.

P13L5: what does "most scenes" means in this context? This sentence is ambiguous and should be clarified.

Findelen is located within two different Landsat scenes per overpass, thus more data is available for Findelen than e.g. Aletsch. Clarified.

"The analysis was performed for one glacier, Findelen, as more scenes were available for this glacier due to the overlapping coverage of this glacier by two different Landsat scenes (path/row 194/28 and 195/28)."

P13L8: This and later paragraphs relate to a specific methodology and analysis that should be introduced earlier. The steps taken to assess uncertainty should be integral of the research design and reporting of results.

To account for this comment, we moved the entire section concerning the uncertainty analysis to the end of the methods section.

P13L11: The authors stated in the previous paragraph that the uncertainty analysis was performed on Findelen glacier, but now indicate that the 39 glaciers were considered. Please clarify.

The uncertainty analysis was in fact only performed for one glacier, Findelen, due to the greater data availability (Findelen is located in two Landsat footprints, see answer above). However, the uncertainty estimates were taken to be representative for all glaciers. We added a respective statement in the manuscript.

"Assuming that bare-ice albedo remains constant over this short time period in reality, this value provides a direct uncertainty estimate for local satellite-retrieved albedo representative for all investigated glaciers in this study."

P14L1: The repeatability of mean albedo determination on (supposedly) bare-ice areas of Findelen glacier is reported in Table 4, leading to an assessment of uncertainty claimed by the authors on the albedo (pixel-wise) being 0.026. It should be clearly indicated that this reports on the precision (repeatability) of the albedo retrieval only, not its accuracy. I also note that 0.026 is obtained by simply averaging the four estimates corresponding to each year. I am not convinced this simple averaging is providing a fair assessment of the repeatability that can be obtained from this approach. As highlighted in my comment on Table 4, I have concerns about the very small value reported for 2016 and it would contribute to lower the perceived precision. I could not find images in table 4 for 2014 so it leave only two other instances in 2013 and 2015 together suggesting quite a substantially larger precision than 0.026.

Our uncertainty analysis indeed estimates the snap-shot uncertainty and, hence, the seasonal/temporal variability and robustness of using only one scene per year. It clearly shows that the closer to each other scenes are acquired, such as in 2016, the lower the snap-shot uncertainty. This emphasizes that there can be substantial temporal variations of bare-ice albedo within one ablation season. However, as we only make use of scenes acquired in August or September (end-off summer scenes) this uncertainty remains relatively small.

Regarding the referee's concern of the very small value from 2016 that might lower the

perceived precision, we argue that if only taking years 2014 and 2015, when about a month lies between the acquisition dates and a medium number of pixels were snow-free, a similar overall snap-shot uncertainty of local albedo of 0.027 is found. The computed average uncertainty of local albedo of 0.026 is thus representative for end-off summer albedo snap-shot uncertainty within the frame of this study.

P14L1: The fact that the assessment of trends does not appear to change in view of the perturbations is not truly surprising. As discussed above, I don't think the methodology and level of detection is conclusive enough to elaborate on areas exhibiting trends at the 80% CI in the context of the current methodology. It would be more informative to test how this level of significance changes when increasing the perturbation on albedo to more realistic uncertainty levels given all other environmental factors. The area exhibiting most change, generally a strong albedo decline are apparently greatly controlled by a redistribution of debris and moraines, hence does not qualify as bare-ice throughout the study period. It is expected that those areas remain confidently detected as areas of strong change in albedo. In conclusion, I cannot agree with the author's statement that the "inferred trends in local bare-ice albedo are considered to be robust despite the uncertainty in the albedo retrieval".

Following our answer further above, we would like to focus on the separation of results and discussion. The calculated trends, also with albedo values randomly perturbed with estimated uncertainties, are based on a profound methodology. Again, we agree with the referee that not all uncertainties are properly considered and discussed so far. In our revision we will update all uncertainties with more realistic uncertainty levels as suggested by the referee and repeat the checks for the robustness of the trend. This might change the general conclusion about the robustness of the obtained albedo trends. Moreover, some areas with strongly negative albedo changes will be removed based on the delineation of medial moraines as well as areas where tributaries separated from the main glacier trunk and debris has become exposed.

P14L11: the expression "snap-shot uncertainty" is vague and unspecific. I recommend that this unfamiliar and uncommon wording is revisited to express more plainly what uncertainty the authors are referring to.

We agree that there was no clear definition of snap-shot uncertainty given in the manuscript. We thus added a clarifying statement in the uncertainty analysis section.

"To account for the uncertainty introduced by the use of one end-off summer scene only and thus the exclusion of sub-seasonal variability in albedo, the snap-shot uncertainty, we performed a comprehensive uncertainty analysis based on ten end-of summer Landsat 8 scenes acquired between 2013 and 2016 (Table 4)."

P15L17-24: This far-reaching interpretation is not supported by any tangible results presented in present study. Presented as it is now and provided some of the weaknesses of the methodology, this equates more to a relatively general hypothesis rather than one directly informed and supported by the data and results presented here.

We agree that the added value of this analysis was not clearly stated, and the interpretation as presented maybe reached too far and was based on rather weak/loose statements. However, we note that the analysis is based on valuable data and methodology and adds an interesting and reasonable component to the discussion of possible causes and dependencies of the observed bare-ice albedo changes as a whole. We will thus strengthen our motivation to include this analysis in the introduction and rephrase the discussion and interpretation of its results under causes and dependencies in the discussion section in the revised manuscript.

P17L9: I believe the authors mean "snap-shot" and not "snap-short" yet I maintain that the use of this "new" terminology is not specific enough to make it informative of a clear source of uncertainty, and hence would advise against the use of it.

As explained in the answer above, we clarified the definition of snap-shot uncertainty. We corrected the typing error.

P17L11: The authors insist in their conclusion that "meteorological conditions preceding the acquisition (. . .) need to be considered". I however saw no such consideration in the manuscript despite most images exhibiting various stages of snowline retreat, some of them with obvious short-lived snowfall.

We admit that the statement wasn't properly formulated, and thus rephrased the respective section. However, in our study we examined the data, in particular concerning fresh summer snow on the glacier surfaces.

"Specifically, the meteorological conditions preceding the acquisition of the satellite data can influence albedo, e.g. summer snow fall events, and so should be taken into account."

P17L12: I am actually quite concerned that one of the main conclusion point is that "highly significant darkening" affect about 10% of the ablation areas. Based on my own interpretation of many of the images used, much of this darkening can be attributed to change in surface type from bare-ice to debris, rather than a "darkening" of bare-ice via the accumulation of LAI for example. To me the point raised by this study rather stresses the potential accumulation of debris on the lower reaches of mountain glaciers in the context of retreat.

Based on the exclusion of supraglacial debris, by manually delineating medial moraines and areas where tributaries separated from the main glacier trunk and debris has become exposed, from our analyses, our conclusions focus more clearly on bare-ice and are thus not misleading. However, we would like to mention again that is generally difficult to clearly disentangle thick debris from Light Absorbing Impurities (LAI) coverage, especially in relation to possible darkening phenomena.

Table 2: Landsat 7 sensor should be named ETM+

Changed.

Table 2: Although it is understandable that rounding may bring variations in numbers at the decimal level, it would be preferable that in 2008 when 0 km2 cloud is reported, this correspond to 0% as well.

We agree that the reported, rounded number is somewhat misleading. Changed.

Table 2: Some of the information reported in the table 2 appears to be incorrect. I could not find images for quite a few dates: I suggest 8/9/2004 (not 9/9/2004); 22/9/2006 (not 20/09/2006); 27/08/2014 (not 26/08/2014); 12/09/2014 or 28/09/2014 (not 1/9/2014 and 27/09 /2014). The sensor in 9/9/2013 appears to be L8 and not L7. This also affects Table 4.

Thanks for this careful review. We adjusted the information in Table 2 accordingly.

Table 2: I am perplex about the reporting of clouds in 25/9/2013 and the reason for reliance on a secondary date 9/9/2013. I downloaded both dates and it appears that clouds in 25/9/2013 indeed marginally affect the east of area (d), however barely above the terminus of Bachi and Minsti glaciers, both outside the scope of this study. I could not see anywhere else where clouds may have caused an issue and the need to rely on an alternate date. This begs the question about the performance of the cloud classification algorithm used by the authors. It is even more confusing since the 9/9/2013 image (L8 and not L7 as reported by the authors) is obviously far worse with many clouds and the fact that the authors specifically mention this image in P4L13. Looking at both dates, it is obvious that the snowline retreated over the period, thus exposing more of the bare ice, at least for those glaciers not obscured by clouds and as reported in Figure 3. It suggests the use of two dates in this instance may be done to retrieve albedo over larger areas of bare ice, at least on some glaciers, but the reader can only wonder. Even in this case, it would beg the question about the consistency of the data being used and the potential effect on mean albedo. Table 4 shows that several scenes are used for assessing uncertainty, but also shows an inconsistency with table 2 as multiple acquisitions in 2013 and 2014 are consistent, while those in 2015 and2016 are not reported in Table 2.

We agree that the there is a need to better clarify the choice of the scene used in our analysis for the year 2013. The later date (25. September) is considerably less affected by clouds, however, fresh snow seems to be present on many glaciers in the southern part of the study area, but also on Glacier de la Plaine Morte in the northern part for example. In contrast, the early date (09. September) is, as you mentioned correctly, more affected by clouds, especially in the Aletschgletscher region. We thus decided for each glacier individually which end-of summer scene in 2013 is more appropriate. This will be explained and clarified in the manuscript.

Concerning the application of our own cloud classification algorithm, please see the answer to a comment further up. Similarly, we clarified the inconsistency that you mentioned between Table 2 and 4 (see also above).

Figure 2: "snowline altitude" not "snowline altitdue"

Changed.

Figure 2: what are SLAconst and rcrit?

We apologize to have missed the explanation of both parameters and thus will add some explanation in the figure caption.

Table 4: I could not find any image on 1/09/2014 nor 27/09/2014.

The given dates were incorrect – sorry! Changed.

Table 4: In 2016, the number of pixels used for the assessment is 5495, thus representing 5km2 of supposedly bare-ice surface. I obtained both images and could only map 3km2 of ice at the most, the rest being mostly the accumulation area still obviously covered by snow. Note also that the lower part of the glacier tongue is severely affected by cloud shadow on 1/9/2016. Beside an issue of size being considered as bare-ice which I can't reconciliate with what the images depicts and shedding doubt on the performance of the snow/ice classification, the albedo variability appears surprisingly small (0.008) when any other years yield about five times larger albedo precision. I believe some clarification is required here.

We double-checked the bare-ice classification based on our surface type evaluation for both scenes used in the uncertainty analysis for the year 2016. It revealed that for the earlier date (25.08.2016) about 5.4 km^2 were mapped as bare-ice with a maximum albedo of 0.315. Similarly, for the later date (01.09.2016) also about 5.4 km^2 of bare-ice were mapped with a maximum albedo of 0.295. We thus are not sure how the referee's estimate of only about 3 km^2 has been obtained.

The low albedo variability between the two scenes used in the snap-shot uncertainty analysis for the year 2016 results from the fact that the two scenes are only acquired seven days apart. In contrast, the scenes used in 2013, 2014 and 2015 were 16, 32 and 32 days apart, respectively. If the obtained standard deviations per year are normalized by the time period in-between the acquisition of the individual scenes used per year (in days), highly similar (0.001 for the years 2014, 2015 and 2016) and slightly higher (0.003 for the year 2013) in-between scenes albedo variability are obtained. The slightly higher normalized standard deviation for the snap-shot uncertainty in year 2013 most likely results from the small bare-ice area available (1190 pixels) to perform the analysis.

Interactive comment on The Cryosphere Discuss., https://doi.org/10.5194/tc-2018-18, 2018.