Response to referee comments

ICESat laser altimetry over small mountain glaciers

D. Treichler and A. Kääb

The Cryosphere Discuss., doi:10.5194/tc-2015-234, 2016

We would like to thank the two reviewers for their constructive feedback and valuable input that certainly helped to improve the article. Detail responses are provided below, together with a mark-up manuscript version where the changes made in response to the referees' comments are highlighted.

Anonymous Referee #1

The manuscript of the Treichler and Kääb discusses elevation changes of the southern Norwegian glaciers, based on differencing of DEM and ICESat data. Their main finding is that, after applying a number of corrections, this method results in credible elevation changes, but the accuracy is limited by uncertainties in the DEM reference data. Although the results and some of the input data (DEMS) are specific to this area, and the negative trend of the southern Norwegian glaciers has been reported elsewhere, I believe this manuscript deserves to be published, after a number of minor changes, as it provides a good road map for future studies using similar methods and that will run into the same limitations of the DEM reference data.

- *The authors choose to estimate elevation changes using the DEM differencing approach. However, another, popular method to estimate dh/dt is the plane-fitting approach (Howat 2008 and the Moholdt papers). Please provide a motivation why you prefer the DEM-ICESat differencing approach for this study.*

The plane-fitting approach requires constant surface slopes between neighbourhoods of footprints as found over large ice bodies and thus fails in rough mountain terrain. Using a reference DEM instead accounts for the more complex topography of mountain landscapes. The motivation for our choice including references to studies using the plane-fitting approach has been added in section 3.

- *The third research question, 'What is the minimum region size w.r.t. glacier density for ICESat GLAS data to ensure statistically significant results' isn't really answered. You do show that you can retrieve an elevation change signal for Myklebustbreen, but don't give a hard lower limit for the region size.*

The reviewer is correct that the manuscript only indirectly deals with that question. We found that there is no hard lower limit for a region size but the area/statistical sample size required depends on the combination of glacier/ICESat track density, homogeneity of the glacier signal, and sample representativeness, i.e. factors specific to each area – in that regard, the question may not be a good one to ask in the first place as there is only a qualitative answer to it. Since it is nevertheless

asked frequently by users of our method we decided not to remove it. We added a new paragraph in section 5.1 that emphasises the factors that need to be considered when grouping ICESat glacier footprints spatially. With this, we hope to provide a more satisfactory and direct – although still relative – answer to the minimum region size.

- *when assessing elevation bias and spatial shifts between ICESat and the reference DEMs (page 6, section 3.2), did you take into account vertical uplift due to GIA? Uplift rates are in the order of 5 mm/yr in Norway, so over a period of _40 years (1978-2016), this would result in about 20 cm difference.*

We did not consider vertical uplift due to GIA – the up to 20cm uplift the reviewer correctly calculated are much smaller than median tile/date offsets (up to +/-1m per tile, and +/- 5m per date) or elevation-dependent error cH (dm per 100m elevation). Also, we could not find a relationship between the DEM age (proxy) and elevation bias, leaving us with no evidence of a glacial rebound effect on the ICESat-Kartverket DEM elevation differences on stable terrain. On glaciers, the c_glac correction removes all offsets regardless of their origin. A note on the magnitude of GIA uplift compared to ICESat-reference DEM elevation differences has been added in section 3.2.

- *when determining the c_glac correction (page 6, line 31- : : :) did you check that the temporal coverage by ICESat is sufficient? If a glacier has only been sampled by one ICESat overpass, it's still possible to compute a c_glac correction, but the resulting dh will be _zero after applying the correction.*

The reviewer is correct that for glaciers that only experience one overpass (possible due to spatial variability of ICESat ground tracks) average differences will be zero. A priori, it can be assumed that mostly small glaciers (i.e. very few samples) would be affected by the problem of a single overpass only. The expected effect of the zero differences is to flatten out the trend. In our study we saw the contrary – surface elevation trends became steeper after application of c_glac. One can argue that the bias introduced by these samples is likely smaller than the considerably larger offsets we found from DEM age and vertical offsets, and that it should be captured by the trend confidence interval. However, the introduced bias from these single overpass samples is systematic, unlike other (random) error sources – especially if they occur in the beginning or end of the acquisition period. We concluded that this bias should be assessed and quantified, following the reviewers concerns.

A thorough analysis on the samples in question revealed the following:

- **128 samples on 36 glaciers are from single ICESat overpasses (only autumn campaigns), corresponding to ca. 10% of ice samples and glaciers hit.**
- **90% of the glaciers sampled only during one campaign (27 glaciers) were sampled in autumn 2003. This can be explained by the transition from an 8 day repeat orbit that was flown for the first 10 days of the autumn 2003 campaign to the 91 day repeat orbit used from October 4th, 2003 through the entire ICESat mission (Schutz et al., 2005). The two orbit repeat cycles have different ground tracks, resulting in 113 of 427 samples falling on glaciers sampled only by a single overpass for this particular campaign. A closer examination of the glaciers in question showed that a large fraction of the samples in question are from an overpass over northern**

Folgefonna where our DEM age proxy indicates reference elevations from the 1980ies. The (uncorrected) dh of these samples are considerably more negative. The option of not applying c_glac to one overpass samples (but all other samples) is therefore not a good solution in our case.

- **The other 15 single overpass samples are well distributed between the remaining campaigns as well as in space and with regard to reference DEM age.**
- **Exclusion of the 128 one-overpass glacier samples results in a slight shift towards larger dh of the median (but not the mean) of the 2003 autumn campaign and, consequently, slightly increases the trend slope (Figure 1). The effect this has on the trend slope lies within the trend standard error and is considerably smaller than the effect of applying c_glac in the first place.**
- **The trend slope difference corresponds to a steepening of 0.05 ma^{-1} for all subsets – except for where sample numbers are considerably smaller and especially where glacier size plays a role: glaciers <5km2, pre-2000 DEM source date and East of water divide. For the latter two the glacier size is an indirect cause (more small glaciers further east and outside areas where DEM updates were prioritised). There, the increase in trend slopes is > 0.1 ma^{-1}.This confirms how sensitive trends are to bias in dh – especially when sample numbers are small.**

In the revised manuscript we consider these new insights throughout the text, specifically in sections 4.1 and 4.4 in the results, in the discussion, as well as in Table 1.

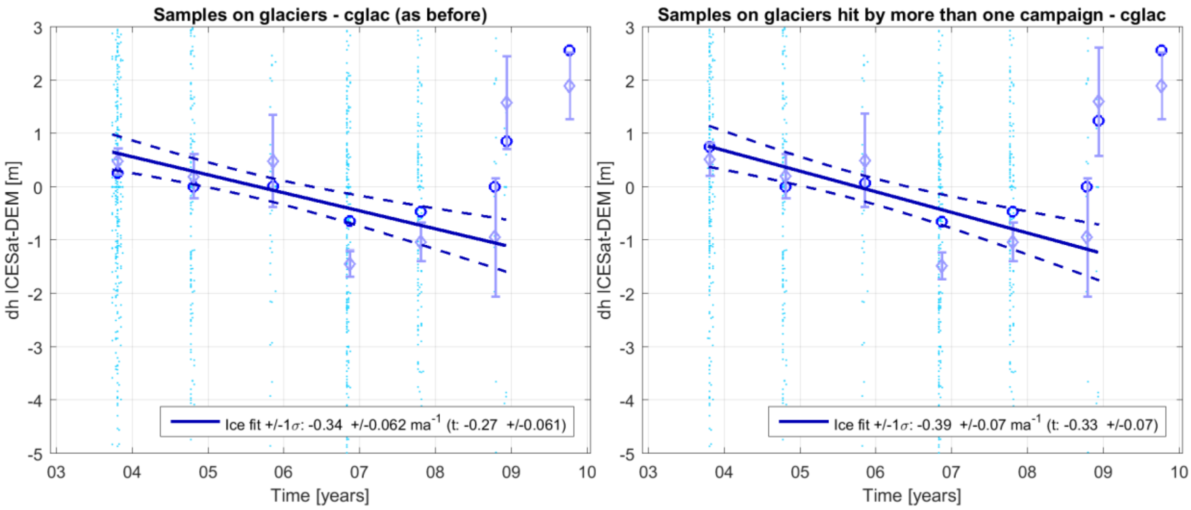

Figure 1: Glacier surface elevation trends with (left panel) and without (right panel) *ice* samples on glaciers sampled by a single overpass only.

- *page 13: the c_glac seems only to work if the DEM subset covering the glacier is based on data from one acquisition date (as you also point out on page 14 when discussing the Swiss Alps DEMs). It's worthwhile to point this out here.* **DONE on page 10 (assuming a typo in the page number above)**

- *page 10, lines 28-33: do to the increasing cumulative uncertainty in the in-situ mass balance measurements, it's hard to verify this claim. It would be helpful to include a 'mean' in-situ mass balance curve (after applying some weighting to ensure this 'mean' is representative).*

The range of cumulative glacier surface balances measured by NVE has been updated in the revised manuscript, using the harmonised/calibrated data that is now available from NVE (NVE, 2016;

Andreassen et al., 2016). A mean cumulative mass balance curve for these glaciers (weighed by respective glacier areas) has been added to the figures. The new harmonised data fit the ICESat trend even better than the non-harmonised data shown in the discussion manuscript (Figure 2). Note that in the corresponding figure in the revised manuscript, only 8 (instead of 10) in-situ glacier series are included as mass balance measurements on Folgefonna outlet glaciers started only in 2006, and homogenised data for these very short and recent time series are not yet distributed by NVE.

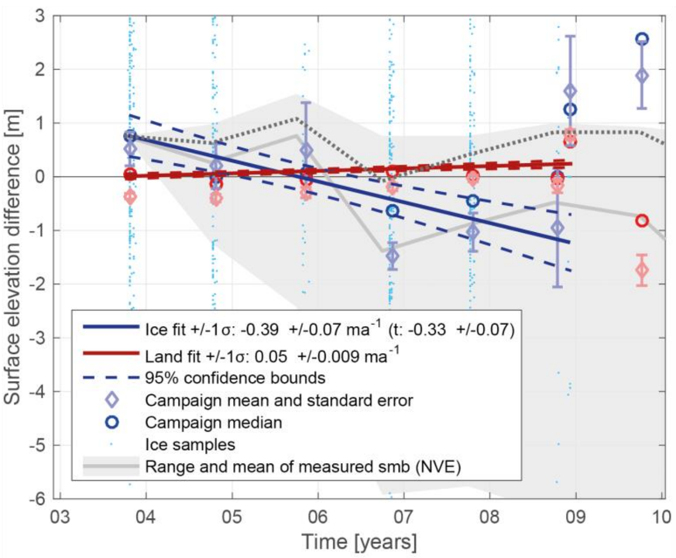

Figure 2: ICESat glacier elevation trend compared to surface mass balance (smb) of 8 glaciers in southern Norway measured by NVE (based on in-situ data and geodetic methods). Shown are the range of cumulative smb re-converted to ice volume using a density of 850 kg m^{-3} (Andreassen et al., 2016), and their weighed mean (solid line, weighed by glacier area) of the newly available harmonised data, as well as the area-weighed mean of the volumetric balance of 10 glaciers before harmonisation (dotted line, corresponding to the data shown in the grey spread in the corresponding figure in the discussion manuscript).

- page 11, lines 1-4 + figure 4: the upward jump in the 2009 campaign data is probably an artifact of poor sampling, but what does the in-situ data tell about this year?

From in-situ data we expect a slightly negative balance in 2009. A comment stressing this discrepancy has been added to section 4.4.

- page 13, lines 7: The trends for winter ice samples are indeed more negative, but the uncertainty is much larger, due to the interannual variability in accumulation, and differences with the autumn trends are non-significant. This should be pointed out. **DONE**

- page 13. limes 14-19: Whether or not the derived trends for such small glaciers are to be trusted depends to a large extent on the spatial sampling of the glacier. Samples across the entire elevation range are required, with a sampling density resembling the hypsometry distribution of the glacier. Without a further analysis it's impossible to tell what the 0.47 +/- 0.11 m/yr trend represents. Please discuss this in the manuscript.

The reviewer is absolutely correct that also such a local trend is only valid if the ICESat samples are representative for the glaciers in question. This is the case here (see Figure 3), and the

representativeness of this sub-sample is now mentioned in section 4.4 and discussed in a more general way in the new paragraph in section 5.1 (see also reply to the reviewer's second comment). However, we would like to emphasise that in our study, the main role of mentioning the Myklebustbreen/ Haugabreen surface elevation trend lies in explaining the 2009 jump and in stressing the need for per-campaign representativeness also in terms of good and consistent spatial distribution. In that sense it is a side product of our study. While we have no reason to assume that the trend is wrong we strongly advise that the surprisingly positive glacier surface elevation change trend on Myklebustbreen/Hansebreen is critically reviewed and, as far as possible, verified with other data in case it should be used in further studies.

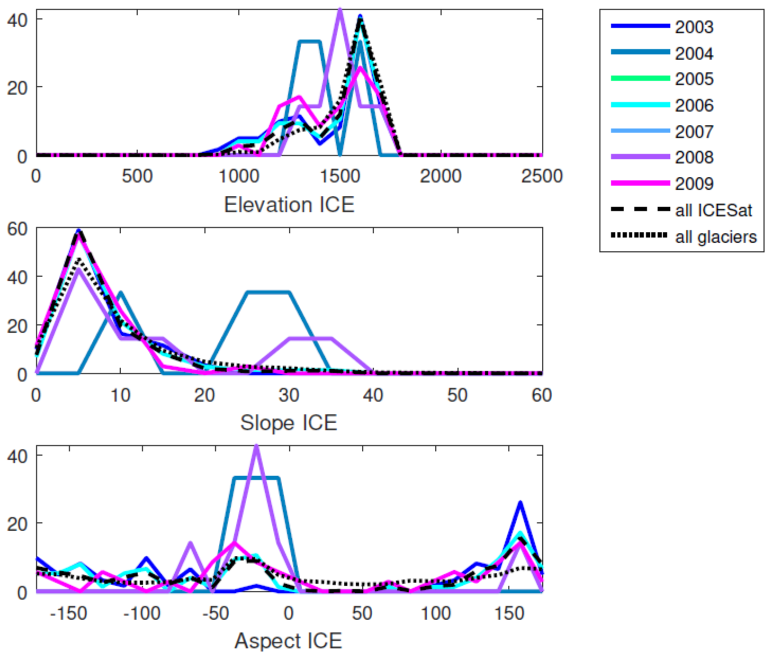

Figure 3: Representativeness of Myklebustbreen and Haugabreen (normalised frequency in %; elevation in m, slope in degrees, aspect in degrees from North). The dotted line corresponds to all DEM cells of these glaciers. Note that campaigns 2005 and 2007 contain none and campaigns 2004 and 2008 only 3 and 7 samples, respectively, reflected in worse fit of these curves compared to the ICESat full sample/entire glacier area from the reference DEM for these glaciers.

- figure 7: the uncertainties for '05 are huge. Did something go wrong during plotting, or are these real (if so, it deserves to be discussed in the manuscript).

The reviewer's comment concerns the trend for Myklebustbreen/Haugabreen. The autumn 2004 campaign (closer to the 2005 axis tick) consists of only 3 samples on these glaciers of which one obviously is an outlier with very large dh, resulting in a huge standard error of the campaign mean (error bar). A comment on these low campaign sample numbers has been added to section 4.4 and in the figure caption.

Technical/minor comments:

Page 1, lines 15-22: I would move this part of the abstract to line 13 (after, "rather than ICESat uncertainty"). Right now you're first discussing the DEM biases, then the ICESat elevation changes and then move back to the DEM biases. **DONE**

Page 2 line 10: Slobbe 2008 discusses the Greenland Ice Sheet, bot ice caps, so technically, it doesn't belong in this list.
The reference has been removed from the list.

Page 2 lines 20-34: I suggest to use bullet points here to present the list of research questions. **DONE**

Page 4, line 14: include references for previous ICESat studies **DONE**

page 5, line 30: start a new paragraph after "... removing footprints on clouds (false positive dh).".
DONE

Figure 2: the dotted line is really hard to distinguish (both on screen and print copy)
The figure has been updated to better distinguish the land spread. Note that the land spread may still be difficult to see as it is very narrow, and that both ice/land spread have slightly changed shapes in the updated manuscript since the autumn 2009 campaign (excluded from analyses) is now also excluded from the spread. This was, erroneously, not the case before.

page 8, line 2: what's the number of ice samples in the autumn 2003 campaign?
427 ice samples, this information has been added

Page 9 line: on average **DONE**

page 16, line 4: change 'volume loss' to 'elevation change' (or convert the -0.34/-0.27 m/yr height changes to volume changes) **DONE**

R.C. Lindenbergh

The authors use ICESat satellite laser altimetry elevations as available from 2003 to 2008/2009 to estimate glacial elevation change of small mountain glaciers in Norway. The authors consider several angles to this problem. First, part of the paper could be read as a report on how to extract such glacial elevation changes from the relatively sparse available ICESat elevations over the Norwegian glaciers with the help of locally and globally available auxiliary Digital Elevation Model (DEM) data. An important second angle the authors consider is the influence of the required reference DEM and its possible misalignment on the quality of the results. A third angle, as also the title suggests, is an assessment of using ICESat elevations in general to estimate elevation changes of small mountain glaciers, as can be found all over the world. For this angle it is crucial to assess to what extend local and sparse glacial elevation changes are representative for a glacial area as a whole.
We agree with the referee that the paper deals with several aspects of ICESat over mountain glaciers and readers might read it with a different focus in mind. Our motivation for this study was to thoroughly assess ICESat-derived glacier surface elevation changes – a method that has already

been applied globally in many places and thus begins to be relatively established, but has so far not been thoroughly validated. In that sense, we see our contribution as a road map to improve ICESat applications on small mountain glaciers in general – not only in Norway. This was reflected in the title, as the reviewer mentioned, too.

The paper has valuable contents that are interesting for a larger audience. Small mountain glaciers are present at many different locations on Earth. Monitoring their elevation change by satellite laser altimetry data from ICESat-1 and maybe later ICESat-2 using an additional reference DEM is useful, if detected changes are indeed representative.
My problem with the paper in its current state is its focus. If the paper is meant to guide how to extract glacial elevation changes for arbitrary mountain glaciers, at least an analysis on how ICESat is sampling glaciers as a function of latitude is missing: Norwegian latitudes are still relatively favorable, compared to e.g. many South American latitudes. The influence of DEM misalignment is clearly assessed in the manuscript, but how to identify and correct for such misalignment has already been discussed in existing articles. Therefore I suggest to focus the paper on the particular case the authors consider: detecting glacial elevation changes using ICESat and a reference DEM over small Norwegian mountain glaciers. Still, the discussion chapter could be used to generalize to other small mountain glaciers.
The selection of Norway as a test site was made due to the reference data available here, a key condition for a solid method assessment. We are certain – and this is also stressed in our manuscript – that the issues discussed (representativeness, DEM quality) are transferrable to other locations. The study builds on (and is also motivated by) extensive tests done on ICESat applications in High-Mountain Asia (Kääb et al, 2012; supplement). While based on a method assessment in Norway, our findings are not only specific to Norway. We very intentionally analyse our results from Norway with a broader, more global horizon in mind in the discussion section.

We agree with the reviewer that ICESat sampling in relation to latitude would certainly show that higher latitudes are more favourable. At the same time, latitude does only to some degree affect whether or not ICESat applications on glaciers are possible or not in a given region. Other factors – glacier density, size, position in relation to ICESat tracks, and homogeneity of the glacier signal – are equally important, if not more so. The key requirement is the representativeness of the samples, and this has to be assessed locally. We believe that a visualisation of ICESat glacier samples vs. latitude would give this particular aspect of ICESat applicability too much weight in comparison with other factors to consider. We prefer to keep the focus of the paper on the method, not on identifying the optimal places for ICESat glacier studies on the entire globe.
In the revised manuscript, we try to stress the dependence of track density on latitude more to make the reader aware of this fact. That this is not the only factor to consider when applying our method is now better stressed in a new paragraph on representativeness and minimum region size in section 5.1 in the revised manuscript.

We certainly don't claim that DEM misalignment is a new discovery of ours, but we emphasise that the commonly done global co-registration of entire DEM tiles with ICESat samples may not remove all bias. The fact that shifts of spatially unknown DEM sub-units can have a strongly biasing effect on ICESat analyses, as well as our proposed localised correction (c_glac), are new. We feel that the

issues we discuss have so far not received enough attention by users of the method and hope to make users of ICESat in other glacierised areas aware of them with our contribution.

We are afraid that the article would not reach all of the potentially interested scientific audience if the focus lay on Norwegian glaciers. These glaciers are well studied and our ICESat surface elevation trends only confirm the results from in-situ and geodetic mass balance studies. However, we agree with the structure the reviewer proposes: results from Norway, but discussed in a general, global context. We tried to make this now even clearer in the manuscript. We did not change the focus to Norwegian glaciers only, as the reviewer suggested, but kept it global/general. We hope the reviewer agrees.

In addition, the authors should address the following aspects: they don't distinguish between ICESat footprints sampling snow and ice. This should be discussed, and, the effect of this choice on the results should be assessed.

The reviewer is correct in that we don't distinguish between ICESat footprints on parts of the glacier that are snow-covered/bare. This is intentional: In order to capture a signal that may be related to geodetic mass balance we need to sample the entire glacier to consider both surface elevation changes from ice melt and glacier dynamics. Footprints on only snow-covered or bare ice parts of the glaciers would likely only lie on the accumulation or ablation parts of the glaciers, respectively, and do thus not fulfil the condition of mass continuity. It would therefore be physically incorrect to draw conclusions on the glaciers' mass balance from a trend based on such a subset of samples. This is an important and inherent condition of any volumetric-geodetic glacier method, including ICESat studies. It is also part of the reason why we stress the need for representativeness so much. An explanation on this has been added to section 3.3 where sample subsets are introduced.

The referee might also point to potentially different densities over ice and snow to convert elevation changes to mass changes. This issue is a tricky one as one cannot strictly know if an elevation change is due to a change in the ice column or in snow or firn thickness, new firn, or if the density profile changed over time (firn compaction, superimposed ice, etc.). These issues are discussed in Huss (2013) which we rely on for our density assumption. Further, ICESat dh density scenarios have also been evaluated in the cited Kääb et al. (2012). See also below response on density, and volume vs. mass.

The state of the glaciers during ICESat passes could be assessed using additional spectral data or by considering the raw ICESat full waveform signals.

The inclusion of additional remote sensing data to characterise ICESat footprints and the surface they fall on is a good idea that we very much support. We agree that for a follow-up study with a focus on seasonal changes or mass turnover such a distinction might be valuable. While we argue above why we think this is not appropriate in the way the reviewer proposes for the current study, it is to some degree also not possible, due to the following reasons:

- **Within optical remote sensing data, Landsat would be the most promising candidate to provide continuous data for land cover classification in Norway in 2003-2009. However, Landsat's repeat cycle of 16 days, combined with Norway's rather cloudy weather, makes it unfortunately impossible to ensure cloud-free coverage for operational classification of ICESat**

footprints on sufficient temporal time scale to reliably detect snow events and snow cover. MODIS fails as an alternative due to its coarse spatial resolution. A combination of both datasets complemented with modelled data, such as the snow cover maps distributed by the Norwegian Water and Energy Directorate NVE (available on senorge.no), could possibly work, although large uncertainty from interpolation and other modelling aspects would have to be expected.

- **Our impression from previous studies are that waveform classification methods may be used to classify land cover types in general but seem to be struggling to reliably distinguish different glacier surface types with sufficient accuracy (Molijn et al., 2011; Shi et al., 2013). We think these methods would need to be further refined to be used in a more operational way, i.e. without adding large uncertainties. Waveform analyses would be most interesting for applications where distinction of within-footprint snow cover or fine-scale surface topography are crucial. A possible approach to increase classification accuracy could be waveform fitting to within-footprint topography from a reference DEM. Snow cover (smoothing effect) might be detected where waveforms and reference DEM surface or other indicators (likelihood of smooth/rough surface from timing, elevation and slope of footprint) don't match.**

Footprint classification with the above methods would add uncertainty to our analyses that would be hard to quantify and, consequently, make validation of the method more difficult – which is the prime focus of the work. We prefer therefore not to do so in this study but see this as an interesting idea for further work.

As a side note, we did already a study on modelling waveforms from high-res elevation models and ground reflectivity data, and we analysed waveforms over rough mountain topography. Both unpublished studies didn't lead to very conclusive results that would clearly benefit the study under discussion here. Most likely, the rough and variable topography over mountains and mountain glaciers make it difficult to retrieve simple rules based on waveforms that could be applied to regional (i.e. not local) mountain studies.

Similarly, there might be an effect of terrain roughness and slope on the results, which is not discussed.

We agree with the reviewer that terrain roughness and slope will alter the ICESat return waveform and thus might affect extracted elevations. Already (Kääb et al., 2012; article supplement) did a thorough analysis on the potential effect of slope on ICESat elevations in the Himalayas. There, the most relevant finding was a positive relation between saturation of footprints and slope in particular for off-glacier terrain. It remains unclear which way round the causality works – sloping terrain causes saturated footprints, or the saturation classification algorithm misclassifies these waveforms as saturated due to their shape.

Also in Norway, we did extensive tests to discover potential systematic bias on dh from slope or within-footprint surface topography, using the standard deviations of slope, aspect and elevation of 10m DEM grid cells within an assumed 70m circular ICESat footprint as a proxy. We found no significant relationship between dh and slope or within-footprint topographic roughness, and no indication for a systematic bias from footprint slope or roughness, also not in combination with waveform saturation. Note that the Norwegian mountains present a different landscape than the Himalayas, and we experienced both much lower numbers of saturated samples and flatter slopes in this study than in the Himalayas, in particular for land samples.

Still, the reviewer is correct in that the slope of the footprints has an influence on derived trends. Bootstrapping methods on sample subsets using different slope thresholds show that ICESat glacier surface elevation trends in southern Norway are more negative for samples on larger slopes. However, it would be wrong to draw the conclusion that steeper glaciers experience stronger melt than flatter ones in our study area, as the phenomenon is explained by topography and glacier physics: Many of southern Norway's glaciers are small ice caps and have thus relatively flat accumulation areas on top of the rounded mountains, smoothened by the Scandinavian ice shield (an example is Jostedalsbreen ice cap the reviewer refers to below). Glacier tongues, on the other hand, extend into the steep fjords and valleys. A sample subset based on slope resembles therefore a sample subset of elevations. Trends derived from samples at high/low elevations (i.e. accumulation/ablation parts only) reveal the same effect, but even stronger: At high elevations, surface elevation change trends are considerably flatter than at elevations of the glacier tongues. This is a direct effect of glacier flow physics. Considering the fact that we see a trend at all we can be sure that the glaciers in southern Norway were not in balance between 2003 and 2008. Negative imbalance may result from increased melt – more pronounced at lower, warmer elevations – or decrease of precipitation. In both cases, glacier flow will eventually transport changes over the entire glacier due to ice flow, i.e. increased melt at the tongue will also result in lowering of the accumulation areas. The signal may though be delayed, and the interpretation of such non-trivial differences in surface elevation changes of different glacier parts is beyond the scope of this study.

Our goal with this study is to validate and improve the method of deriving volumetric glacier surface elevation changes from ICESat elevations. Consequently, we do not discuss aspects that were found to have a minor or no influence on derived trends in detail but rather focus on the main influencing/biasing factors.

In the revised manuscript, we added more explanations in section 3.3 and 4.4 to ensure readers are aware of glacier mass continuity where sample representativeness and sample subsets are discussed.

In addition, the authors confuse glacial elevation change with mass balance change, which are two different things. The authors should discuss why glacial elevation change can directly be linked to mass balance change, notably when one doesn't distinguish between ICESat footprints over snow and ice. Some more detailed remarks are given below.

See also above response on density.

Our study confirms that ICESat-derived glacier surface elevation changes indeed accurately reflect volumetric glacier balance – which may thereafter be converted into glacier mass balance with the use of ice/snow/firn densities. Since the focus of this particular study lies on the method (which is inherent to ICESat and contributing factors) and not on conversion between volumetric and mass changes (which applies also to e.g. DEM differencing, and which does not depend on ICESat parameters but local conditions mainly), we prefer not to discuss the problem of volume/mass conversion here. In contrast to e.g. Kääb et al. (2012) or Gardner et al. (2013) whose focus lay on the derived trends, not on the method itself, we do not attempt to convert ICESat surface

elevation trends to mass changes and draw conclusions thereof. We only relate ICESat trends to mass changes to compare to the NVE mass balances and use their choice of density for this purpose, i.e. in a way just back-convert NVE's results. We checked again and tried our best to be very clear and consistent in our related vocabulary throughout the manuscript.

The results of Kääb et al. (2012) differ by only 5% between two different density scenarios for conversion of ICESat elevation trends to mass change. One of the two applied scenarios distinguishes between firn and ice areas (600 and 900 kg m^{-3}, respectively) while the other assumes an average density of 900 kg m^{-3}. (Gardner et al., 2013) also assume a density of 900 kg m^{-3}. To compare the in-situ data with ICESat-derived elevation changes, we back-converted water equivalent to ice assuming a density of 850 kg m^{-3} based on the findings of Huss (2013) – which also NVE used for their geodetic data – and which can be seen as a new standard value for glacier volume/mass conversion for volumetric-geodetic mass balance studies. While the conditions for this number are not perfectly in place – given the imbalance and year-to-year variation of southern Norway's glaciers that indicates instable mass gradients – it should serve sufficiently well as a best guess for the validation purpose the in-situ data has in our study. A paragraph that highlights the difference between ice surface elevation and mass changes, and that justifies this choice more explicitly, has been added to section 3.

Detailed Remarks:

1. As above: I would focus the paper on Norwegian glaciers, which should be reflected by the title.

As explained above, we prefer to keep the title short and attractive for readers that should find relevant information for their work. And we believe the information given is by far not only relevant for Norwegian glaciers but rather for ICESat over mountain glaciers in general.

2. p2r26: A more general question that is still open: "Is ICESat track density (in combination with average cloud cover) high enough for sparse glaciers at arbitrary latitudes?"

We agree with the reviewer that the reader might like to get an answer also on more general questions, such as the main factors that govern ICESat applicability in glacierised regions. While we discuss many of these factors already now, the corresponding question is currently missing in the manuscript. Since our findings show that latitude plays an indirect role in the way that it affects sample numbers and, subsequently, representativeness, we formulated a new question that is even more general: *What prerequisites and conditions need to be fulfilled to make ICESat-derived elevation changes over a certain area a valid method to assess glacier volume changes?* This additional research question is discussed together with the minimum region size in a new paragraph in section 5.1

3. p3r20: "two to three month-long observation periods", you mean "two to three observation periods each year of about one month each" **CHANGED**

4. p3r21: "42 km" this may hold for Norway, but is in general latitude dependent.

We added a note on the relation between cross-track spacing and latitude.

5. p4r14: "ICESat tracks of more than one year": funny English, please reformulate. **DONE**

6. Section 2.3: what are the difference between: "vertical accuracy", "mean error" and "standard error", please define these notions.. **DONE**

7. p5r4: -> "The 2009 autumn campaign is excluded" (skip 'usually' to avoid confusion) **DONE**

8. p5r9: what is the influence of the 40m threshold for "ice border"? Apparently (Section 3.3) this threshold has a strong influence on the amount of ICESat elevations that are considered to fully cover glaciers (given the quotes of 2.5% on glacier points, and 0.9 % of border points)
We are not sure we understand the reviewer's question correctly. The goal of the 40m in- and outside buffer is to avoid footprints falling both on ice and land, i.e. giving a mixed elevation signal. There is little room for discussion for the threshold itself as it corresponds to roughly the diameter of one ICESat footprint (ca. 70m). The reasoning for this has been reformulated in section 3.1 to ensure better clarity. The buffer could potentially be larger to better account for spatial uncertainty of glacier outlines – both from potential changes over time, and from the limits in accuracy due to the fact that glacier outlines usually are derived from satellite data with a given pixel size. This would result in more *ice border* footprints and lower the *ice* sample numbers further. We believe that the 40m buffer is appropriate for the comparably high quality of the glacier outlines in southern Norway. The 0.9% share of *ice border* footprints of all samples reflects the fact that many glaciers in southern Norway are rather small, resulting in a comparably large buffer area around the many small ice bodies.
Ice border samples could both exhibit large dh where a glacier has melted/retreated/advanced since the reference DEM was acquired, or very small dh where there never has been ice within that footprint in the first place. The dh distribution of *ice border* samples was found to be wider than for *land* and *ice* dh and is not affected by c_glac correction (Fig. 2 in the manuscript). While one could argue that these large dh indicate changes in ice elevations, inclusion of *ice border* samples in trend calculations affects the trend slope if c_glac is not applied, and increases uncertainty if c_glac is applied (the trend uncertainty is the same as for *ice* samples only, but given the higher sample numbers it should be smaller if the additional samples contributed with dh representing accurate and valid measurements). We therefore advise users to exclude ice border samples also in other areas where the introduced uncertainties may be even larger. This recommendation has been added explicitly in section 5.1.
We hope this answers the reviewer's comment.

9. p5: "snow heights": (Kääb, 2012;2015) discusses Central-Asian glaciers. Why can conclusions on snow variations there be simply ported to a Norwegian setting?
And would there be no big differences for valley glaciers compared to icefields, as this figure of Jostedalsbreen suggests:
https://en.wikipedia.org/wiki/Jostedal_Glacier#/media/File:P1000290Jostedalsbreen.JPG
The reviewer's comment relates to our argumentation why spring and winter campaigns were excluded from trends. High-Mountain Asia (HMA) includes glaciers in a wide range of different environmental conditions. The argumentation in Kääb et al. (2012; 2015) is a general one that discusses the meaning of the signal and not, as the reviewer might have assumed, applicable to

HMA glaciers only. They postulate that yearly varying timing and magnitude of snow fall cannot ensure that the ICESat overpasses measures winter surface balances reliably. Snow densities are expected to vary highly depending on ICESat timing in relation to snow pack evolution. Yearly varying net glacier balance will affect glacier mass turnover which will also be reflected in winter surface elevations. As a consequence, winter elevations measured by ICESat reflect a mix of elevation changes in ice and snow surfaces that is hard to resolve. This is especially the case for time series as short as the five years where ICESat data is available. Cumulative net balances, such as our ICESat glacier elevation trends represent, should therefore be based on yearly net balances from the end of the hydrological year. For method comparability with studies of other authors (Gardner et al., 2013; Ke et al., 2015; Kropáček et al., 2014; Neckel et al., 2014) who include winter data in their ICESat-derived glacier trends, we computed winter trends for our study site, too. Our results are in line with the arguments brought forward already by Kääb et al. (2012; 2015). While winter campaigns (as well as samples from the accumulation/ablation areas, or samples on snow-covered/bare ice only) may contain a signal of changes in mass turnover potentially interesting for a future study, we advise against the use of winter campaigns to derive glacier surface elevation changes where these should be related to/used as glacier mass balance series.

Why is it not actively assessed if glaciers are covered by snow at the time of the ICESat passes? That could also assist in the issues on winter snow fall and December campaigns raised in Section 4.4.

We refer to our argumentation above for distinguishing snow-covered vs. bare ice samples. Concerning the December campaign: There are to date unfortunately no remote sensing methods that measure snow heights. Optical sensors are only capable to map snow cover. However, modelled snow depths from NVE (available from senorge.no) and observations from meteorological stations (data from the Norwegian Meteorological Institute, available from eklima.met.no) confirm the onset of snowfall in November/December 2008 – which is to be expected at that time of the year in Norway – and suggest the dh measured by ICESat are within plausible range. Since correction of the December 2008 dh with snow depths estimates based on December 2008 *land* samples results in an even flatter *land* trend and no noticeable change in the *ice* trend (Table 1, section 4.4), we believe this issue is sufficiently explained and expect no further insight from the use of additional data. We hope the reviewer agrees.

10. p5, IWD, what parameter? I.e. what power?

Power 1, i.e. linear inverse distance weighing. Information added in section 3.1.

11. p5: how did the outliers look like that were removed by the robust fitting? How did the spatial pattern of cloud affected ICESat elevations look like?

Robust fitting removed *ice* samples with |dh|>14m (Figure 4). After removal of one-overpass samples (see reply to reviewer 1), 76 of 1105 samples received a weight of 0.3 or lower (coloured) and 26 samples received weight 0. For footprints on slopes > ca. 25 degrees weights decrease with increasing slope (Figure 5) which is in accordance with the larger expected elevation uncertainty of footprints on sloping terrain. There is no clear dependency on footprint elevation, aspect, or area of the glacier sampled (Figure 5), and there is no visible pattern in the spatial distribution of the samples weighed less than 0.3 (Figure 4).

A note on the number of samples that received weight 0 has been added to section 3.1.

ICESat samples identified as cloud elevations (dh>100m, excluded from trend analyses) have a similar spatial distribution as the samples used in trend analyses (Figure 6).

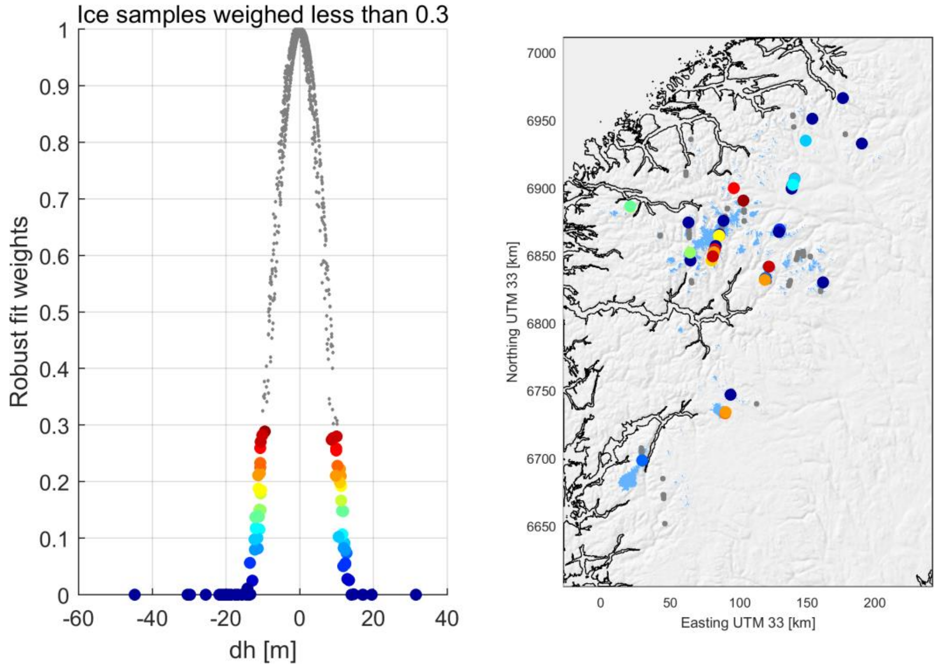

Figure 4: dh (left) and spatial distribution (right) of *ice* samples in relation to the weights assigned by robust fitting.

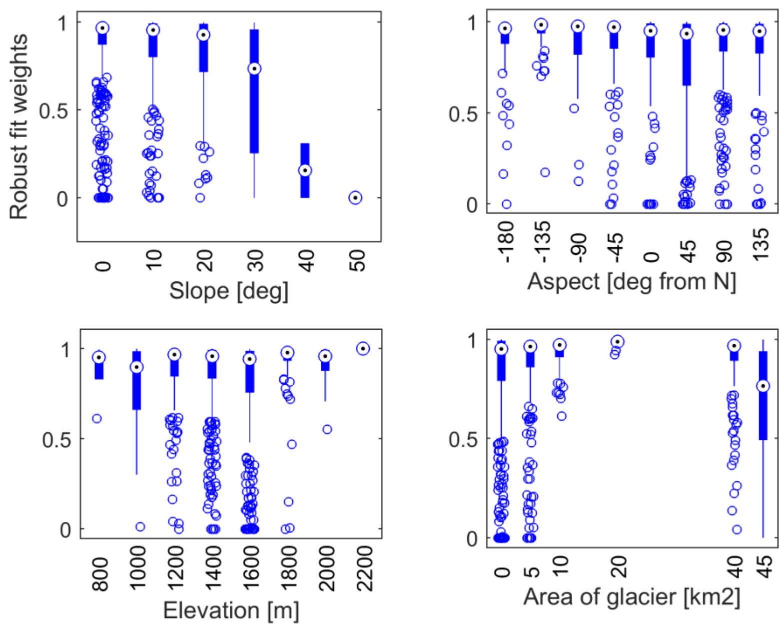

Figure 5: Weights assigned to *ice* samples by robust fitting as compared to glacier-governing parameters.

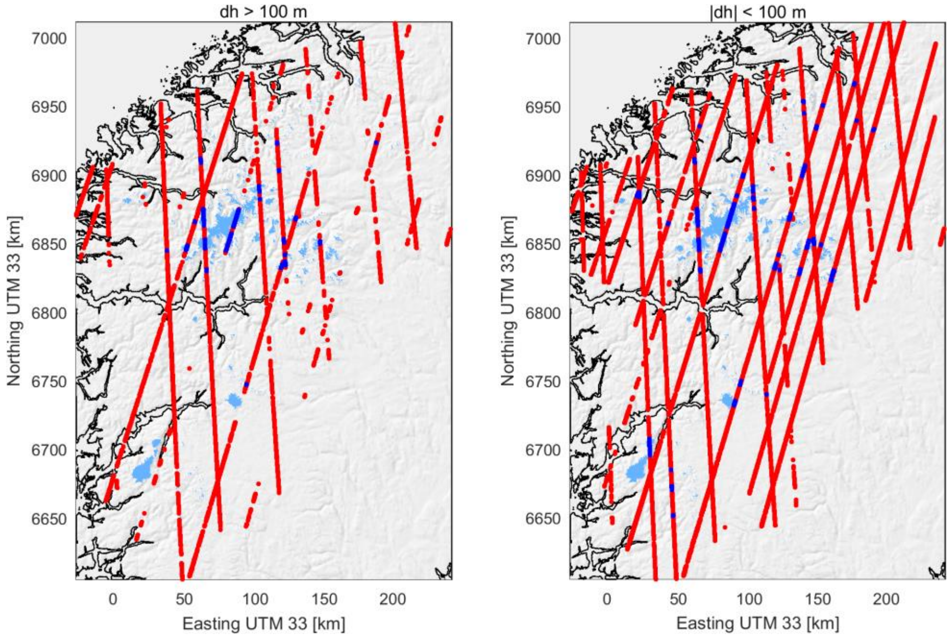

Figure 6: Spatial distribution of cloud samples (left) that were excluded from the study compared to samples included in the study (right, map corresponding to fig. 1 in the manuscript).

12. p6r1: can you quantify "larger number of outliers"?

A quantification of the number of "outliers" for any distribution requires a threshold dh or, alternatively, a p-value that corresponds to a fraction of samples considered valid. Figure 7 shows our *ice* dh distribution (c_glac applied) as well as fitted t- and normal distributions. In this direct comparison, it becomes obvious that the dh and t distributions have longer tails than the normal distribution with its bell shape. Consequently, the sample fractions of our ice dh distribution shown in Table 1 match the fractions of a fitted t distribution much better than the ones of a fitted normal distribution. Reading example for |dh|>20m: only 0.03% of normally distributed samples but 1.36 and 1.09% of the samples following a fitted t distribution and our dh distribution, respectively, exceed the threshold of |dh|>20m.

We included a reference to Figure 2 in section 3.1 and hope that statistically interested readers will be able to see the typical shape of a t distribution (of both *ice* and *land* samples) with their relatively larger number of outliers in this figure themselves.

Table 1: Example for the "heavier tails" of our sample distribution, as compared to a normal distribution. The numbers correspond to the fraction of samples (area under the tails of the curve) for |dh| exceeding the threshold values 5, 10 and 20m (symmetric against zero) for a fitted normal distribution, fitted t distribution, and the original ice sample distribution.

|dh|>	normal	t	samples
5	0.4026	0.2345	0.2434
10	0.0773	0.0654	0.0760
20	***0.0003***	***0.0136***	***0.0109***

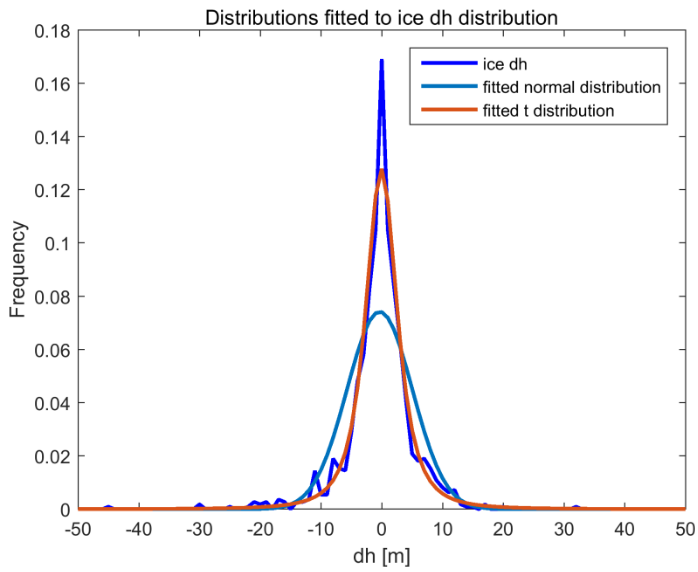

Figure 7: dh distribution of *ice* samples and corresponding t- and normal distributions fitted to the data.

13. p6: did you experience any issues in the LIDAR data due to not fully adjusted flight strips? (Remaining errors after strip-adjustment)

The LiDAR data was not used for glacier trends, in that sense the answer is no. Figure 4 in the manuscript shows that the LiDAR DEM elevations are closest to ICESat elevations, as compared to the Kartverket/SRTM DEM elevations, but an uncertainty remains nevertheless – which stems from uncertainties in ICESat elevations and, to some degree, likely also from remaining errors after strip adjustment. The sample number of only 184 LiDAR DEM autumn campaign samples over Hardangervidda is rather small and we could not detect a significant systematic spatial bias for any of the strips, or parts thereof, from splitting these samples into spatial subsets.

14. p6: what are possible reasons for the shifts in the Kartverkets DEMs?

We prefer to not discuss this further in the paper as the topic lies too much outside of the focus of this study. The main point is that the individual source units of ALL composite DEMs will in general not be perfectly aligned for various reasons depending on the DEM type and processing. In our case of the Kartverk DEM, we believe potential reasons for the shifts can be found in the general production processes of such national DEMs, at least older ones:

- **Norway is a big/long and sparsely populated country and maps (and thus DEM units) have to be compiled from different air photo series, collected and processed over a range of years, with a range of equipment both for data acquisition and photogrammetric compilation.**
- **All these air photo blocks are adjusted individually and connection to pre-existing blocks cannot be perfect unless the air photos over the entire country are adjusted as one block.**
- **The air photo blocks and maps generated from them were compiled many years ago, without or with only little computer methods available and thus reduced consistency over large areas.**
- **The accuracy demands at the time of map production (and thus production of the contours that lie behind much of the DEM) were much lower than today, and it is actually well possible that they fulfilled these requirements at the time of production. In our study we compare a modern high-precision data set (ICESat) to old or very old reference elevations.**

15. Section 4.1: how do you know the dh are t-distributed?

Visually, the dh distributions seem too narrow but have a larger number of outliers as compared to a normal distribution (Figure 7). Statistically argued, the assumption of a t distribution seems legitimate considering that a) the population standard deviation (of dh) is not known, and b) the measured dh correspond to the sum of the measured glacier surface elevation change (the signal) and multiple errors from both the DEM and ICESat as well as topography/clouds. Data with additive errors/resulting from additive processes are more likely to have outliers (compared to normally distributed data). In such cases regression based on a t-fit is suggested as a robust method (Lange et al., 1989).

We tested the data both for normal and t-distribution. The assumption of a normal distribution of our *ice/land* dh is rejected by an Anderson-Darling test (distribution fitted to the data). Both an Anderson-Darling test for a t-distribution fitted to the data as well as a Kolmogoroff-Smirnoff test in combination with a parametric bootstrap procedure to find a consistent estimate of the critical value (the standard critical values from tables are not valid if the test distribution is estimated from the data; Babu and Rao, 2004) are not able to reject the null hypothesis of a t-distribution. While that is no proof, it gives us as much confidence as possible for the assumption of a t-distribution of our data.

16. Saturation may occur along track when ICESat hits bare ice after rock (as it takes the gain _5 shots to reset after hitting the more reflective ice). Did you consider the spatial distribution of the saturated waveforms? (compare Molijn RA, Lindenbergh RC, Gunter BC. ICESat laser full waveform analysis for the classification of land cover types over the cryosphere. International journal of remote sensing. 2011 Dec 10;32(23):8799-822)

The reviewer's comment relates to the automated gain loop built in into the data acquisition which dynamically adjusts the gain based on received pulse intensities of the past laser shots (NSIDC, 2012). We think the reviewer has an important point here, given that a systematic bias at land/ice transitions would mainly affect parts of the glacier where we expect larger dh (assuming glacier melt/retreat). If this were the case it could affect and bias our trends.

The adjustment time of 5 shots the reviewer mentions corresponds to a ground track distance of ca. 860m. Many mountain glaciers (also in southern Norway) are smaller/narrower than this. While we believe this might be more pronounced with larger ice bodies, our studies indicate that the dynamic gain adjustment indeed seems very continuous in mountainous areas. There, the small mountain glaciers and rough topography never really allow the sensor to settle for a certain gain. In southern Norway we cannot see any pattern in the spatial distribution of the saturated samples, also not where ICESat is passing over glacier margins and experiences a land/ice surface type change. We don't see a consistent spatial pattern of the samples' gains either but find that as many as 40% of all samples have the maximum gain of 250, independent of their saturation flag. We also find that samples flagged as saturated have higher gains than non-saturated samples, both for *ice* and *land* samples – this is the contrary of what one instinctively would expect. Interesting is also that the gains of both *land* and *ice* samples increase with time, something that is beyond the scope of this validation study but could potentially be relevant for other studies where waveforms/surface types are the focus.

In this context, we would like to stress that the algorithm that flags samples as saturated is designed for the flat ice sheet surfaces. Rough mountain surfaces – including the steeper and more crevassed mountain glaciers – result in entirely different waveforms than laser returns from ice sheets. From our experience, we don't find the saturation flag to be a good indication of saturation – i.e. elevation bias – over mountainous topography. This finding, which is also in line with what Kääb et al. (2012) found, is supported by the NSIDC GLAS user guide that doesn't generally recommend saturation correction for the GLAH14 product (NSIDC, 2012), and which we refer to in section 2.2.
We added a reference to Molijn et al. (2011) in section 5.3 to make readers aware of the possibility of bias from saturated waveforms at land/ice transitions in other areas than southern Norway. Given that we can't draw any useful conclusions or improvements from saturation analyses in our study we hope the reviewer agrees with our decision not to discuss saturation (or sample gains) in more detail in the manuscript resubmission.

17. You state: "However, these differences cancel out (Fig4)". Could you help the reader seeing that in Figure 4.?
We rewrote the corresponding paragraph in section 4.3 to hopefully make the argumentation clearer.

18. From the material just in this paper it is difficult to understand what you mean by p10r25-27: "This stresses...weight". Could you explain this a bit more extensively?
The reviewer's comment concerns our statement that trends should be computed from individual dh and not campaign medians. With this we tried to account for a question that we often hear from the scientific community. Further explanations were added to create a better context in the corresponding paragraph in section 4.4.

19. p12r6: you say "terrain characteristics" are essential, but, as argued by me before you only consider these only in a very global way.
We are not sure what the reviewer would like us to change in that context. We found that "terrain characteristics that govern glacier behaviour" need to be well represented – which is also how the sentence in question is formulated in section 5.1. Thus, in the given context we refer to larger scale terrain, not within-footprint terrain. To be clearer about the scale of the terrain we refer to, we also added the term "topography" in the corresponding sentence in section 5.1.

20. Section 5.3: do you believe that indeed the age of the DEM is crucial, or rather the way it was constructed (photogrammetry, radar, LIDAR)?
This is very different for glaciers (surface elevation varies over time) and stable terrain. In general, vertical bias is a result from mosaicking of different datasets which have (different) elevation bias. The reasons for such bias are manifold and to a large degree depending on the quality of the data acquisition and post-processing (see above response).
On glaciers, however, it is primarily the age of the reference DEM which is crucial in this context: The glacier surface elevations of different years are expected to be different depending on annually changing mass balance. Even under the assumption that the majority of the glaciers in southern Norway follow the same cumulative mass balance curve (since they sit in the same

climate), reference elevations from different dates for different glaciers are not spatially uniform. Combining these with ICESat's spatially varying sampling can potentially lead to severe bias. This is why c_glac is more powerful – and more important – for glaciers than for the surrounding stable terrain.

We added a paragraph in section 5.3 to better stress the important difference of this correction for *ice* and *land* samples, respectively.

21. p15r13: saturation correction (and other flags). I would say this is an interesting topic for more study, to check how the rapid transitions between land-cover on small mountain glaciers influence the ICESat raw signal (and its corrections)

We agree with the reviewer that this could be an interesting topic for a study that focuses more on single footprints and also assesses their waveforms. From the findings of Molijn et al. (2011) we cannot exclude that there is a potential for a systematic bias from waveform saturation at ice/land transitions, even though we could not detect any such bias in our study area. A paragraph discussing this possibility has been added to section 5.3. For a more detailed argumentation on this matter we refer to our reply to point 16.

22. p15: quality of the reference DEM vs ICESat: should it not be only the quality,but also its spatial resolution compared to the ICESat footprints compared to the local relief variations?

We fully agree with the reviewer. We added a sentence to section 5.3 to make this clearer.

References

Andreassen, L. M., Elvehøy, H., Kjøllmoen, B., and Engeset, R. V.: Reanalysis of long-term series of glaciological and geodetic mass balance for 10 Norwegian glaciers, Cryosphere, 10, 535-552, 10.5194/tc-10-535-2016, 2016.

Babu, G. J., and Rao, C. R.: Goodness-of-fit tests when parameters are estimated, Sankhya, 60 (1), 63-74, 2004.

Gardner, A. S., Moholdt, G., Cogley, J. G., Wouters, B., Arendt, A. A., Wahr, J., Berthier, E., Hock, R., Pfeffer, W. T., Kaser, G., Ligtenberg, S. R. M., Bolch, T., Sharp, M. J., Hagen, J. O., van den Broeke, M. R., and Paul, F.: A Reconciled Estimate of Glacier Contributions to Sea Level Rise: 2003 to 2009, Science, 340, 852-857, doi: 10.1126/science.1234532, 2013.

Huss, M.: Density assumptions for converting geodetic glacier volume change to mass change, Cryosphere, 7, 877-887, doi: 10.5194/tc-7-877-2013, 2013.

Ke, L., Ding, X., and Song, C.: Heterogeneous changes of glaciers over the western Kunlun Mountains based on ICESat} and Landsat-8 derived glacier inventory, Remote Sens. Environ, 168, 13 - 23, doi: 10.1016/j.rse.2015.06.019, 2015.

Kropáček, J., Neckel, N., and Bauder, A.: Estimation of Mass Balance of the Grosser Aletschgletscher, Swiss Alps, from ICESat Laser Altimetry Data and Digital Elevation Models, Remote Sens., 6, 5614, doi: 10.3390/rs6065614, 2014.

Kääb, A., Berthier, E., Nuth, C., Gardelle, J., and Arnaud, Y.: Contrasting patterns of early twenty-first-century glacier mass change in the Himalayas, Nature, 488, 495-498, doi: 10.1038/nature11324, 2012.

Kääb, A., Treichler, D., Nuth, C., and Berthier, E.: Brief Communication: Contending estimates of 2003-2008 glacier mass balance over the Pamir-Karakoram-Himalaya, Cryosphere, 9, 557-564, doi: 10.5194/tc-9-557-2015, 2015.

Lange, K. L., Little, R. J. A., and Taylor, J. M. G.: Robust Statistical Modeling Using the T-Distribution, Journal of the American Statistical Association, 84, 881-896, doi: 10.2307/2290063, 1989.

Molijn, R. A., Lindenbergh, R. C., and Gunter, B. C.: ICESat laser full waveform analysis for the classification of land cover types over the cryosphere, Int. J. Remote Sens., 32, 8799-8822, doi: 10.1080/01431161.2010.547532, 2011.

Neckel, N., Kropáček, J., Bolch, T., and Hochschild, V.: Glacier mass changes on the Tibetan Plateau 2003–2009 derived from ICESat laser altimetry measurements, Environ. Res. Lett., 9, 014009, doi: 10.1088/1748-9326/9/1/014009, 2014.

NSIDC: GLAS Altimetry HDF5 Product Usage Guide, NASA DAAC at the National Snow and Ice Data Center, Boulder, Colorado USA, 2012.

NVE: Climate indicator products, Norwegian Water Resources and Energy Directorate, available at: http://glacier.nve.no/viewer/CI/, accessed 31 May 2016, online glacier database, 2016.

Schutz, B. E., Zwally, H. J., Shuman, C. A., Hancock, D., and DiMarzio, J. P.: Overview of the ICESat Mission, Geophys. Res. Lett., 32, doi: 10.1029/2005gl024009, 2005.

Shi, J. C., Menenti, M., and Lindenbergh, R.: Parameterization of Surface Roughness Based on ICESat/GLAS Full Waveforms: A Case Study on the Tibetan Plateau, Journal of Hydrometeorology, 14, 1278-1292, doi: 10.1175/Jhm-D-12-0130.1, 2013.

ICESat laser altimetry over small mountain glaciers

D. Treichler and A. Kääb

Institute of Geosciences, University of Oslo, P.O. box 1047, 0316 Oslo, Norway

Correspondence to: D. Treichler (desiree.treichler@geo.uio.no)

Abstract. Using sparsely glaciated southern Norway as a case study, we assess the potential and limitations of ICESat laser altimetry for analysing regional glacier elevation change in rough mountain terrain. Differences between ICESat GLAS elevations and reference elevation data are plotted over time to derive a glacier surface elevation trend for the ICESat acquisition period 2003-2008. We find spatially varying biases between ICESat and three tested digital elevation models (DEMs): the Norwegian national DEM, SRTM DEM and a high resolution LiDAR DEM. For regional glacier elevation change, the spatial inconsistency of reference DEMs – a result of spatio-temporal merging – has the potential to significantly affect or dilute trends. Elevation uncertainties of all three tested DEMs exceed ICESat elevation uncertainty by an order of magnitude, and are thus limiting the accuracy of the method, rather than ICESat uncertainty. ICESat matches glacier size distribution of the study area well and measures also small ice patches not commonly monitored in-situ. The sample is large enough for spatial and thematic subsetting. Vertical offsets to ICESat elevations vary for different glaciers in southern Norway due to spatially inconsistent reference DEM age. We introduce a per-glacier correction that removes these spatially varying offsets, and considerably increases trend significance. Only after application of this correction do also individual campaigns fit to observed in-situ glacier mass balance. Our correction has the potential to improve glacier trend significance also for other causes of spatially varying vertical offsets, for instance due to radar penetration into ice and snow for the SRTM DEM, or as a consequence from mosaicking and merging that is common for national or global DEMs. After correction of reference elevation bias, we find that ICESat provides a robust and realistic estimate of a moderately negative glacier mass balance of around -0.~~30m~~ 36m +/-0.~~06~~ 07 ice per year. This regional estimate agrees well with the heterogeneous but overall negative in-situ glacier mass balance observed in the area. ~~ICESat matches glacier size distribution of the study area well and measures also small ice patches not commonly monitored in-situ. The sample is large enough for spatial and thematic subsetting. Vertical offsets to ICESat elevations vary for different glaciers in southern Norway due to spatially inconsistent reference DEM age. We introduce a per-glacier correction that removes these spatially varying offsets, and considerably increases trend significance. Only after application of this correction do also individual campaigns fit to observed in-situ glacier mass balance. Our correction has the potential to improve glacier trend significance also for other causes of spatially varying vertical offsets, for instance due to radar penetration into ice and snow for the SRTM DEM, or as a consequence from mosaicking and merging that is common for national or global DEMs~~.

1 Introduction

The role of mountain glaciers and snow as source for drinking water, irrigation and hydropower is getting increasing attention, not least due to the significant population increase and economic development in a number of mountain regions and surrounding lowlands (Jansson et al., 2003; Viviroli et al., 2007). Retreat of mountain glaciers is also a major cause of eustatic sea level rise (Gardner et al., 2013). But the response of some large glacierised systems to climatic changes is still poorly quantified, especially in regions with large climatic variability. The glacier regions least represented in long-term in-situ glacier monitoring programmes are those with largest ice volumes (Zemp et al., 2015), which are less inhabited, difficult to access, and therefore not well studied. Regional estimates of ice loss recently gained importance, not least for assessing the current and future contribution of water stored in land ice masses to sea level rise (Gardner et al., 2013; Jacob et al., 2012; Marzeion et al., 2012; Radić et al., 2014; Radić and Hock, 2011) and for quantifying current runoff contribution from glacier imbalance (Kääb et al., 2015) or changes in the upstream cryosphere (e.g. Bliss et al., 2014; Immerzeel et al., 2010). Remotely sensed data is of special value in remote mountain regions where measurements such as in-situ mass balance measurements are sparse or lacking completely.

Elevation data from Geoscience Laser Altimeter System (GLAS) on board the NASA Ice, Cloud and land Elevation Satellite (ICESat) provides likely the most consistent global elevation measurement currently available (Nuth and Kääb, 2011). The use of this data to derive thickness changes of Arctic ice caps is well established (Nuth et al., 2010; Moholdt et al., 2010; Bolch et al., 2013; Nilsson et al., 2015; Slobbe et al., 2008). Kääb et al. (2012) have shown that, when combined with reference heights from a digital elevation model (DEM), ICESat data can successfully be used to derive regional-scale glacier mass balance even in rough topographies as the Himalayas. Subsequently, ICESat's elevation measures combined with the Shuttle Radar Topography Mission (SRTM) DEM were used to estimate sea level rise contributions from mountain glaciers globally (Gardner et al., 2013), regionally in High Mountain Asia (Neckel et al., 2014; Kääb et al., 2015), and even for local glacier mass balance studies in the Kunlun Shan (Ke et al., 2015) and the Alps (Kropáček et al., 2014).

The increased public interest in glacier retreat, not least due to its effects on water resources stored in mountain glaciers, requires that the performance of ICESat over such terrain is carefully evaluated and associated error sources well characterised. This is especially important given that using ICESat data over mountain topography is at (or even exceeds) the limits of what the mission was designed for. As a case study for this purpose we chose the mountains of southern Norway. With its comparably small and sparse glaciers, situated within a varied topographic setting of both steep and gentle mountains, we consider the region as a representative case for the limits of applicability of ICESat data for analysing changes of mountain glaciers. In contrast to large, remote areas like High Mountain Asia, the climatic framework and glacier responses are relatively well known and measured in southern Norway, and accurate, up-to-date glacier masks and a high-resolution reference DEM are available.

Specifically, we aim to address the following questions in our study:

- What prerequisites and conditions need to be fulfilled to make ICESat-derived elevation changes over a certain area a valid method to assess glacier volume changes?
- Is the ICESat track density high enough for the sparse glacier cover in the study region, and are the point samples along ICESat profiles representative of the whole glacier population in southern Norway?
- Can a realistic elevation trend be retrieved for the years 2003-2009 (glacier volume loss), and is it possible to detect climate-related patterns, namely the spatial transition from maritime towards more continental glaciers with increasing distance to the coast?
- What is the minimum region size with respect to glacier density for ICESat GLAS data to ensure statistically significant results? Are realistic annual glacier thickness changes visible over a sufficiently sampled single glacier?
- How do the findings compare to observed glaciological and geodetic glacier mass measurements?
- How does the reference DEM influence the quality of the results, and how to best model the footprint reference elevation?

2 Study site and data

2.1 Southern Norway

The study area referred to here as southern Norway extends over an area of 100’000km^2 at 59-63 degrees latitude. It comprises all areas of the Scandinavian Mountains south of Trondheim that are within a 20-km buffer around glaciers (Figure 1). While very steep especially at fjord flanks, the study area consists of both rounded and rough mountains but also includes high-elevation plateaus such as Hardangervidda. The climate of the study area is governed by a West-East gradient from a maritime climate at the coast with high precipitation amounts to dryer conditions further East in the rain shadow of the Scandinavian Mountains (Melvold and Skaugen, 2013). This is reflected also in measured glacier net balance magnitudes (Kjøllmoen et al., 2011). The Norwegian glacier area has recently been mapped by the Norwegian Water Resources and Energy Directorate (NVE) based on Landsat imagery from 1999-2006 (Winsvold et al. (2014); digital data available from the Global Land Ice Measurements from Space (GLIMS) database). Glaciers cover 1’522km^2 or roughly 1.5% of our study area. This includes 1’575 ice bodies ranging from small perennial ice patches of just over 0.01km^2 in size to the largest outlet glaciers (>40km^2) of the Jostedalsbreen ice cap. 50% of the glacierised surface in southern Norway consists of glaciers with <5km^2 spatial extension, and 20% of the glacier area of ice patches smaller than <1km^2. Some maritime glaciers advanced in the 90ties while glaciers located in more continental climate showed mainly frontal retreat (Nesje et al., 2008; Andreassen et al., 2005). After a culmination in 2000, most of the monitored glaciers in Norway experienced net mass deficit (Kjøllmoen et al., 2011; Andreassen et al., 2016).

2.2 ICESat

ICESat GLAS was a single-beam spaceborne laser altimeter operational between February 2003 and October 2009, sampling the surface elevation of the Earth within roughly 70m-footprints during two to three ~~month-long~~ observation periods each year of about one month each (Schutz et al., 2005). The laser footprints have 172m spacing along-track, and approximately 42km cross-track spacing between 91-day repeat reference orbits at 61 degrees latitude (Figure 1). Cross-track spacing increases at lower latitudes, making polar areas in principle more favourable for ICESat applications. Note that our study area already lies in the polar acquisition mask of the ICESat mission at >59°N, where the off-nadir pointing mode enabled near repeats of the tracks (ca. +/-150m), in contrast to a nominal orbit repeat precision of +/-1'000m for mid-latitudes (Schutz et al, 2005). In accordance with what Kääb et al. (2012) found to be the most suitable product for mountain glacier analyses, the ICESat data set used was GLAS/ICESat L2 Global Land Surface Altimetry HDF5 data (GLAH14), release 33 (Zwally et al., 2012). For GLAH14, elevation values were not changed between releases 33 and 34 (NSIDC, 2014). The data contains quality attributes and elevation corrections for each footprint. These attributes include a waveform saturation flag (attribute *sat_corr_flag*) to indicate saturation of the sensor when recording the returned pulse, and a correction for the potential bias in extracted elevations from these saturated waveforms (attribute *d_satElevCorr*). The flags and corrections are intended for improving elevation accuracy on ice sheets, the original main purpose of the mission, and are not necessarily valid in rough mountain topography (NSIDC, 2012).

2.3 Reference data

The reference elevation datasets used are the national DEMs provided by the Norwegian Mapping Authority (further referred to as Kartverket) in 10m and 20m spatial resolution (http://data.kartverket.no). In mountain areas, the Kartverket DEMs are based on source data at 1:50'000 map scale including elevation contours at 20m equidistance, resulting in a nominal absolute vertical accuracy of +/-4-6 m (defined as the standard deviation of elevation; Kartverket, 2016). Using the source date stamp of elevation contours as a proxy, the age of the DEMs was found to be highly variable geographically, ranging from 1978 to 2009 on southern Norway's glaciers, and from 1961 to 2011 on non-glacierised areas.

For the Hardangervidda area and up to approximately 60.3°N, the global DEM from the Shuttle Radar Topography Mission (SRTM, Farr and Kobrick, 2000) is available at 3 arc-seconds resolution (corresponding to 93m in y, and 45m in x-direction at 60°N) from the U.S. Geological Survey (https://dds.cr.usgs.gov/srtm/). The SRTM DEM used here is based on C-band radar data acquired in February 2000 and consists of a composite of four or more overpasses at latitudes that far north (Farr et al., 2007). The absolute vertical accuracy of the mission is stated as 16m (defined as 1.6 times the standard deviation of the error budget throughout the entire mission; Rabus et al., 2003) but found to be in the range of few metres as compared to ICESat elevations (Carabajal and Harding, 2006). The SRTM DEM featured as the reference DEM of choice for previous ICESat glacier trend analyses (e.g. Gardner et al., 2013; Kääb et al., 2012). Unfortunately, it does not cover glaciers ~~with~~

visited by more than one ICESat ~~tracks~~ overpass~~of more than one year~~ in southern Norway. In this study, SRTM version 3 serves as alternative reference DEM for, thus, only land samples.

For parts of the non-glacierised Hardangervidda plateau, high resolution LiDAR DEMs were provided by NVE (Melvold and Skaugen, 2013). The data consist of six east-west oriented 80km long stripes of 500m width and cell size of 2m, flown on 21 September 2008 (minimum snow cover, leaf-off conditions). Datasets were available as high-resolution gridded DEMs. From comparison to a kinematic ground GPS survey carried out in April 2008, Melvold and Skaugen (2013) found the absolute elevation errors of the LiDAR dataset to range from -0.95m to +0.51m, with a mean error of 0.012m and a standard ~~error~~ deviation of 0.12m.

Yearly net surface mass balance estimates from in-situ measurements of ~~10~~ 8 glaciers within the study area (see NVE's report series 'Glaciological investigations in Norway'; Kjøllmoen et al., 2011) were used as a reference for glacier behaviour during the ICESat acquisition period. The data series are the product of the recent homogenisation of in-situ measurements with geodetic measurements (Andreassen et al., 2016) and are available from http://glacier.nve.no/viewer/CI (NVE, 2016).

3 Methods

ICESat data points from the end of the hydrological year (autumn campaigns) are treated as a statistical sample of glacier surface elevations in southern Norway. We follow the double differencing method described by Kääb et al. (2012) where differences between ICESat elevations and a reference DEM (hereafter referred to as dh) are analysed. Direct comparison of ICESat elevations of different years, as done for larger Arctic glaciers and ice caps (plane-fitting methods, e.g. Howat et al., 2008; Moholdt et al., 2010), is not possible for small mountain glaciers. These methods assume a constant slope of the ice surface within the spatial variability of ICESat repeat ground tracks, which is not given for small mountain glaciers. The use of a reference DEM instead takes into account the more complex surface topography of small glaciers. When compared to elevations from a reference data set of a different source date, the dh will be negative if the surface has lowered over time between the DEM source date and ICESat acquisition time, and positive if the surface has risen. Differences should be zero if the surface elevation was constant, such as over stable ground. Uncertainties in elevation measures of both datasets, not least as a result of rough terrain within the ~70m circular ICESat footprint, raise the need for sufficiently large statistical samples to reduce the effect of random errors. The evolution of dh over time is used to investigate surface elevation change trends over the ICESat acquisition period 2003-2008. (The 2009 autumn campaign is ~~usually~~ excluded due to low spatial coverage before complete ICESat failure.) Note that ICESat captures a signal of volumetric balance that results from surface elevation changes rather than mass change directly. The same is also the case where geodetic mass balances are obtained from DEM differencing, which is a widely used method. Comparison of ice surface elevation change trends with in-situ measurements provided in metres water equivalent (m w. eq.) requires unit conversion that depends on ice density. To validate the ICESat-derived trends, we back-converted the in-situ data using the same density as NVE used for mass/volume conversion of geodetic data (Andreassen et al., 2016), which is based on the findings of Huss (2013) who suggested a value of 850 ± 60 kg m^{-3} as an average integrated over an entire glacier. (See also the discussion and density scenarios in the Supplement of Kääb et al., 2012.)

3.1 Pre-processing and filtering of ICESat data

ICESat surface elevations (height above reference ellipsoid) were converted to Norwegian height above mean sea level, in accordance with national DEM elevations. The ca. 170'000 data points within the study area were classified into *ice* and *land* footprints using the glacier outlines provided by NVE. Footprints lying partially on glaciers, i.e. with footprint centre locations within 40m of NVE glacier borders (both in- and outside original outlines), were classified as *ice border*, and excluded from further analysis. Apart from avoiding a mixed *land/ice* elevation signal from partly ice-covered 70m footprints ~~This was done to account~~this also accounts for the spatial uncertainty of glacier outlines and their potential change over time. For glacier analyses, spring and summer campaigns were excluded to avoid biased trends due to yearly varying snow heights (see argumentation in Kääb et al., 2012; 2015), and the 2009 autumn campaign was excluded due to insufficient spatial coverage caused by weakening of the laser over time. To account for differences in spatial distribution and potential elevation changes due to onset of snowfall, the split autumn campaign of October 2008 (laser 3K, ran out of power before the campaign was completed) and December 2008 (laser 2D, completion of the autumn 2008 campaign) were treated separately where appropriate. *Land* footprints on fjords and lakes were filtered out using shoreline data provided by the Norwegian Mapping Authority, as water levels may vary (tides, hydropower reservoirs).

Reference DEMs were corrected for elevation bias and spatially co-registered with ICESat (see Sect. 3.2). Reference elevations for each footprint were extracted from the DEMs by different statistical means: footprint centre elevation, mean, median, mode (rounded to the metre/decimetre for the Kartverket/LiDAR DEMs), inverse distance-weighted (IDW, linear weighing, i.e. power 1), and bilinear interpolation of elevation of DEM grid cells within an assumed circular footprint with 35m radius (i.e., 4 grid cells for SRTM, 12 for Kartverket 20m, 38 for Kartverket 10m and ~960 for the LiDAR DEM).

The elevation differences between ICESat and the Kartverket DEM were analysed to denote a cut-off threshold for maximum elevation differences. Mean dh were found to be ~-0.5m for *land*, and ~-2m for *ice* samples (i.e. ICESat elevations are lower than reference elevations over glaciers). Using bootstrapping methods and histogram analysis for thresholds between 50m and 250m for |dh|, we found that a cut-off threshold of +/-100m dh effectively removed cloud measurements. Footprints with |dh|>threshold were excluded from all further analyses. The conservative threshold allows for uncertainty in elevation measurements of both datasets (*land* and *ice*), while allowing for slightly skewed dh distributions. It ensures all negative dh from glacier surface lowering between DEM acquisition date and ICESat elevation measurements are included while removing footprints on clouds (false positive dh).

Robust linear regression (we used Matlab's robustfit function with default parameterisation) through all individual samples was performed to find a linear trend for surface elevation change over time. Robust methods iteratively re-weigh least squares to find and exclude outliers until regression coefficients converge. For our *ice* trends we found that ca. 2-3% of the samples received weight 0 and were thus essentially removed as outliers. As an alternative trend estimate, we used the gamlss package in R (www.gamlss.org) to perform regression using a fitted t-distribution. The t-fit accounts for the larger number of outliers in our distribution of dh (Figure 2) as compared to a normal distribution (Lange et al., 1989).

3.2 Sub-pixel shifts and corrections applied to the reference DEMs

Based on dh of autumn campaign *land* samples, elevation bias and spatial shifts between ICESat and the reference DEMs were quantified. The non-systematic spatial shifts of sub-pixel magnitude and biases were corrected, where possible. No corrections were applied to the LiDAR DEM. For the Kartverket and SRTM DEMs, directions and magnitudes of the shifts seemed to vary highly, also within single DEM tiles. Automated co-registration using the methods of Nuth and Kääb (2011) was performed to correct an overall 20 m south shift and -2.6 m vertical offset of the SRTM DEM, as compared to ICESat. However, additional shifts and biases that seem present in sub-units of the SRTM DEM could not be corrected. For the Kartverket DEMs, dh were found to be elevation-dependent (more negative with increasing elevation above sea level *H*). The relationship is in the order of decimetres per 100 m elevation and applies to both the 10m and 20m DEM as both are based on the same source data. To account for this vertical bias, a correction term c_H was applied to individual elevation values of both Kartverket DEMs:

$$c_H = 0.882 - 0.00158 * H \tag{1}$$

Automated co-registration of the individual nominal Kartverket DEM tiles (50x50km and 100x100km for the 10m/20m DEMs, respectively) was not applied systematically as it did not result in an overall positive effect. This is due to overlying shifts of (unknown) production sub-units within single tiles in different directions. To account for the apparently consistent vertical offsets in some areas, correction terms for each individual nominal tile (c_{tile}) and indicative source date (c_{date}) of the Kartverket DEM were computed (after *cH* correction). For each nominal DEM tile the median *land* difference between ICESat and the Kartverket DEM was removed, or alternatively the same was done for each temporal unit of the Kartverket DEM. Both corrections are meant to remove vertical spatio-temporal biases and bias patterns in the reference DEM. The values of the corrections correspond to the median dh of all filtered *land* footprints at minimum snow cover (autumn campaigns only) per tile and date and are in the order of +/-1m per tile, and +/- 5m per date, respectively. Potential physical causes such as vertical uplift due to post-glacial rebound in Scandinavia are in the order of decimetres for the last half century and cannot explain the large differences between ICESat and reference elevations on *land* surfaces. As a proxy for the reference DEM source date per ICESat footprint we used the time stamp of the closest elevation contour line to each footprint (elevation contours are the most important input dataset the Kartverket DEMs are based on; Kartverket, personal communication, 2013). However, these correction terms are approximate only as spatially confined units with unique source data/firm update dates do not strictly exist and the total DEM is thus a product of spatio-temporal merging (Kartverket, personal communication, 2013), not untypical for DEMs from national mapping agencies.

For glaciers, spatially varying DEM source dates add additional uncertainty. Surface elevation difference between Kartverket DEM acquisition and the first ICESat acquisitions varies for individual glaciers, resulting in different (additional) offsets for each glacier. A correction term c_{glac} for this effect was computed from the median dh of *ice* samples at the time of minimum snow cover (autumn campaigns only) for each individual glacier, as classified using NVE's glacier inventory. The values of c_{glac} range from -20m to +15m and reflect in this study mainly vertical glacier changes between the DEM and ICESat dates. For other areas potentially also other vertical biases from DEM production such as height datums or signal penetration could be addressed in a similar way. The latter are not relevant for the photogrammetric methods behind the Kartverket DEM, but for instance for radar wave penetration within the SRTM DEM.

3.3 Sample representativeness and trend sensitivity

In order to relate measured dh to actual net glacier mass balance, the ICESat sample has to mirror key characteristics of the area/terrain in respect to glacier driving processes. We assessed the representativeness of the ICESat glacier sample for the study area in terms of average elevation, slope, aspect, spatial distribution of the footprints, glacier size, and age of the reference DEM. Representativeness in respect to terrain parameters was tested by comparing the sample distribution to the respective distributions of all glaciers in southern Norway (we used all Kartverket DEM cells within the glacier mask). This was done both for the entire ICESat sample and for individual campaigns. Consistency in terms of reference DEM age distribution per campaign was assessed using the source date of the closest contour line for each sample as a proxy. Additionally, the size of the glaciers sampled by ICESat was compared to the entire glacier population of southern Norway.

To assure robustness of fitted glacier surface elevation difference trends, the effect of different data subsets and elevation corrections applied to either of the datasets were assessed. Subsets were created by including/excluding *a)* sets of footprints, as those classified as *ice border*, with specific DEM time stamps, or samples flagged as fully saturated (attribute *sat_corr_flag* >=3), *b)* spatial subsets, e.g. of glaciers east and west of the main water divide, and *c)* entire campaigns. The elevation corrections assessed include ICESat saturation elevation correction (attribute *d_satElevCorr*) in addition to the correction terms per Kartverket DEM tile/source date/glacier described above (c_{tile}, c_{date}, c_{glac}). Very intentionally, we did not divide our sample into footprints only in the accumulation or ablation parts of the glaciers, respectively. In order to capture a signal that translates into geodetic mass balance it is essential to sample the entire glacier to consider both surface elevation changes from ice melt/gain and dynamic glacier flow. If this is not ensured, the condition of mass continuity is violated, and it would thus be physically incorrect to draw conclusions on glacier mass balance based on surface elevation trends from a subset of samples in the ablation/accumulation areas only. The influence of separating footprints over ice and snow/firn for separate density scenarios is discussed in Kääb et al. (2012).

4 Results

4.1 ICESat sample overview

Roughly 75% of the nearly 170'000 ICESat footprints over southern Norway contain valid information of the Earth's surface elevation (125'312 samples after removal of footprints on clouds and water surfaces, see Table S1). Thereof, 2.6% lie fully on glaciers (versus an additional 0.9% that were classified as *ice border*). For glacier analyses, considering autumn campaigns only, a total of 1'268 *ice* and 48'854 *land* samples remain. These numbers are reduced by 2.8% (*ice*) and 1.6% (*land*) only by excluding the weak autumn 2009 campaign. Dh of the remaining samples rarely exceed +/-10m. The dh are t-distributed with a narrower peak but heavier tails as compared to a normal distribution. Before application of the correction terms to the Kartverket reference DEM, the dh distributions of *ice* and *ice border* samples are considerably wider and in average more negative than *land* dh (Figure 2 left). After application of c_H, c_{tile} and c_{glac} correction terms, 94% and 95% of the *ice*, and *land* autumn samples, respectively, but only 80% of *ice border* autumn samples, show less than 10m absolute elevation difference between ICESat and the (corrected) Kartverket 10m DEM elevations (Figure 2 right).

The spatial distribution and number of ICESat samples is not constant over time and decreases to as little as 10% of the number of samples of the autumn 2003 campaign, which includes most samples of all campaigns (427 *ice* samples). In autumn 2009, only 35 ice samples (vs. 792 land samples) remain over southern Norway. Other autumn campaigns with very small sample numbers are 2005 (65 *ice* samples) and 2008 (24 and 24 *ice* samples for the October and December campaigns, respectively). These periods with particularly few samples correspond to campaigns with few orbits flown (2008, 2009) or heavy cloud coverage (2005).

128 of the *ice* samples lie on glaciers that were sampled only during one single autumn campaign. After the application of c_{glac}, any glacier elevation change signal from these single overpass samples is cancelled out. The majority of these (113) occurred during the autumn 2003 campaign due to the a transition between two different orbit patterns in the middle of that campaign (Schutz et al., 2005). The single overpass samples with, in average, 0m dh may thus flatten out derived trends and were excluded where appropriate.

4.2 Representativeness of ICESat glacier sample

The entire ICESat glacier sample appears representative in terms of elevation, aspect, slope, spatial distribution, and glacier area of the glaciers sampled (Figure 3, Fig. S1). Compared to the frequency histogram of the entire glacierised surface in southern Norway, ICESat slightly oversamples east-facing glaciers and underrepresents the glacierised area in the southwestern parts of the area of interest due to the orbits not covering the Folgefonna ice cap (Figure 1). However, these deviations are of the same magnitude or less than deviations of the frequency histograms of the glacierised area monitored in-situ by NVE. Of the individual campaigns (autumn campaigns 2003-2008 shown within grey spread), those with fewest samples deviate most, but still follow the distribution of the full data set. Variability between campaigns is largest (wide grey spread) for easting, also for *land* samples, due to the sensitivity of the sample to exclusion of entire orbits (due to shorter campaigns / cloudy weather). The two autumn 2008 campaigns are only representative if combined as only a subset of orbits was flown each in October and December, respectively. The autumn 2009 campaign was found to include *ice* samples only for one overpass (orbit 30, Figure 1), resulting in sampling of only Myklebustbreen and Haugabreen, an outlet glacier of the Jotunheimen ice cap. All other campaigns have 5-13 different orbits with glacier samples. Severe spatial concentration and poor representation of southern Norway's glaciers confirmed that also for our study area, the entire autumn 2009 campaign should be excluded from further analyses.

Of the 1'575 ice bodies in southern Norway, 96 or 6.1% are hit by at least one footprint of our filtered ICESat *ice* sample. While not the same glaciers are sampled each year, for all autumn campaigns except for 2009, footprints are spread on 17 (2008) to 77 (2003) different glaciers across the study area. Our ICESat footprints seem to capture small ice bodies according to their relative share of the total glacierised area: 47% of the samples lie on glaciers smaller than 5km^2, 17% on ice bodies <1km^2 (Figure 3 right). Only the (combined) autumn 2008 campaign samples no glacier >12km^2, and the ice bodies sampled in December 2008 are distinctively smaller than those sampled in October in 2008. The smallest glacier within NVE's mass balance program in the area is 2.2km^2 large.

4.3 Error sources and corrections for ICESat and DEM elevations

Elevation errors in the DEMs were found to exceed ICESat footprint elevation uncertainty as well as the magnitude of corrections available in the ICESat products. ICESat elevation corrections from effects of waveform saturation (attribute *d_satElevCorr*) are in the range of decimetres; all other elevation corrections within the dataset are even smaller. Application of ICESat correction terms had no notable effect on dh distributions. The relative share of saturated samples (parameter *satCorrFlag* >=3 in the dataset) varies between 5-40% for the different campaigns, and is up to 15% higher for *ice* than for *land*. In contrast to the findings of Kääb et al. (2012) for High Mountain Asia, we found the number of saturated samples to decrease over time to as little 0-2% for the last three acquisition campaigns (laser 2D-2F). Filtering increased the relative share of saturated samples by ~~in~~ on average 5%, and mean absolute dh (after filtering) are smaller for saturated footprints than for non-saturated ones (95% confidence) for both *land* and *ice*, whether or not saturation correction was applied to the dh. Saturated samples were therefore not removed from the dataset for trend computation, and saturation correction was not applied.

In contrast to the ICESat elevation values that seem robust without any corrections, elevation correction terms applied to the Kartverket reference DEMs significantly narrowed dh distributions (Figure 2 right). The elevation-dependent correction term c_H successfully removed skewness towards more negative dh in dh-distributions, and per-glacier correction c_{glac} clearly caused a major reduction in *ice* dh. The correction terms c_{tile} and c_{date} were found to be interchangeable and resulted in minor improvements only on *land* and *ice* dh distributions. For single footprints, uncertainty in reference DEM elevation is on the order of metres.

Looking at single footprints, ~~Reference~~ reference DEM elevations differ by decimetres to metres between the different statistical measures (mean, bilinear interpolation etc.) applied to DEM grid cells within the ICESat footprint, for one and the same DEM. The method chosen matters most for the SRTM DEM with only four contributing cells, but differences resulting from the chosen elevation extraction method – from the perspective of a single footprint – are also higher for the high-resolution LiDAR DEM with ca. 960 contributing cells than for the 10/20m Kartverket DEMs. However, for larger sample numbers, these differences cancel out and dh distributions for reference elevations from the same DEM, but different elevation extraction methods, are approximately the same (Figure 4). Summarising statistical methods appear to produce slightly narrower dh distributions than centre DEM elevations only but the difference between the curves is not significant. Mode elevations differ most from reference elevations computed by the other methods, also for the 2m LiDAR DEM. We based our further analyses on median DEM elevations per footprint, or bilinear interpolation in the case of the low-resolution SRTM DEM.

Reference elevations between DEMs from different sources varied greatly. For the 184 autumn samples on Hardangervidda where all four reference DEMs were available, the LiDAR DEM matched ICESat elevations closest with a mean vertical offset of 0.03m and a narrow dh distribution (Figure 4). Elevation differences from the co-registered SRTM DEM are skewed with a heavier tail towards negative dh. Distributions of the (corrected) Kartverket DEMs, dating back to the 1970s in eastern parts of the Hardangervidda, are particularly wide for this subset of samples, including an average vertical offset of -1.3m. For other spatial subsets, widths and vertical offsets of dh distributions of the SRTM and Kartverket DEMs vary to the same degree in a seemingly random way. Distributions of dh based on the 10m vs. 20m Kartverket DEMs were the same, also for other spatial subsets, and no improvement in elevation precision per footprint could be found from the finer grid resolution.

Analysis of the DEM source dates for *ice* samples of the different campaigns (Figure 5) shows the representativeness of our sample in terms of Kartverket reference DEM age distribution. 70% of the samples have reference elevations from 2008-2009 (further termed 'post-2000'), and only approximately 20% and 10% date back to the 1990ies and 1980ies ('pre-2000'), respectively. Only two campaigns divert from this distribution: in autumn 2005, 60% of the ice samples have old reference DEMs, and in 2009, all ice samples have very recent reference elevations from 2008-2009. For the split autumn 2008 campaign, all but one of the October samples fall on reference DEMs from 2008 while 80% of the December samples have pre-2000 DEMs. If using uncorrected Kartverket DEM elevations, pre-2000 dh are significantly more negative (mean dh: -7.3m) than post-2000 dh (-3.1m). The per-glacier correction c_{glac} completely reconciles the two distributions as seen in Figure 2. Note that c_{glac} treats glaciers as spatial units with consistent source dataset. Where this is not given – and parts of a glacier surface are mapped on different dates or with different methods – the correction will be only partially effective.

4.4 Glacier thickness trends

We find a glacier surface elevation change of ~~-0.34~~ -0.39 ma^{-1} +/- ~~0.062~~0.07 standard error (1σ) for the years 2003-2008 (Figure 6 right) with all corrections to DEM elevations applied and samples on glaciers covered by only one single autumn overpass excluded. The trend slope decreases slightly to -0.34 +/- 0.062 ma^{-1} when such single-overpass samples are included. Using a t-fit instead, we found trends in general to be less sloping than robust trends for the same sample/set of applied corrections, and obtain ~~an~~ alternative *ice* trend estimates of -0.33 +/- 0.07 ma^{-1} and -0.27 +/- 0.061 ma^{-1} on the same datasets. Campaign means are more negative than campaign medians, which indicates slightly skewed dh distributions for both *ice* and *land* samples. *Land* campaign means/medians follow the near-zero trend as computed from all individual samples very closely (0.05 +/- 0.009 ma^{-1}, t-fit: 0.04 -/+ 0.009 ma^{-1}). An exception to that is the December 2008 campaign which indicates surface rise as compared to the October 2008 campaign due to onset of winter snow fall at higher elevations. Exclusion of the December 2008 campaign effectively sets the land trend to zero and renders the *ice* trend more negative ~~(-0.39 +/- 0.64ma^{-1})~~. On the other hand, however, the December *ice* samples are required for the autumn 2008 campaign to be representative (see section 4.1). Correction of December samples for increasing snow depth (estimated from October-December *land* dh differences per elevation) also removes the *land* trend, but does not affect the *ice* trend. If the per-glacier dh correction c_{glac} is not applied, the *ice* trend is reduced and uncertainty increases to -0.26 +/- 0.12 ma^{-1} (t-fit: -0.22 +/- 0.13 ma^{-1}). This decrease of thickness loss rate is due to the mixing of older and newer dates of the reference DEM that introduces biased dh and thus dilutes trends. Without the correction, *ice* campaign medians/means of uncorrected samples do not follow the assumed linear trend well and the standard errors of the campaign means just about overlap with 95% trend confidence bounds (Figure 6 left). Deviation and uncertainty are largest for campaigns with few samples and non-representative DEM age distribution: 2005, (split) October/December 2008, and 2009 (excluded from trends). If ICESat trends were fitted through campaign medians instead of individual samples, these biased/non-representative campaigns would get the same weight as all other campaigns and, consequently, have more power to alter the derived trend. This stresses that ICESat trends over glaciers should be computed based on the entire footprint sample, not based on campaign statistics (e.g. median dh) that give campaigns disproportionate weight compared to the actual number of samples included in that campaign.

After applying the per-glacier vertical correction c_{glac} to the *ice* dh, means/medians of single campaigns follow the pattern of NVE's in-situ mass-balance measurements remarkably well. The range of cumulative net surface mass balances, converted to surface elevation changes ~~ ~~(Huss, 2013), of ~~ten~~ eight glaciers in the study area is shown as grey spread in Figure 6 ~~(Huss, 2013).~~ (Note that ~~the~~ this data is a product of the recent~~ongoing~~ homogenisation of in-situ data of Norwegian glaciers with geodetic measurements (Andreassen et al., 2016) ~~is not yet reflected in the graph~~and thus differs from more positive glacier mass balance curves published earlier. ~~The~~ For some of the studied glaciers, the data homogenisation suggests ~~slightly~~ stronger mass loss and no or more moderate mass surplus for the glaciers with positive cumulative surface mass balance in the studied time period.) Campaign means are shifted up with the *ice* trend line crossing 0 m dh in autumn 2005 which corresponds to zero elevation difference between ICESat and reference DEM considering decreasing sample numbers (autumn 2005 corresponds to the mean date of all ICESat samples used). Noteworthy is the 2005 autumn campaign which – only after correction – fits well with the reported positive net balance for five of ten measured glaciers (Kjøllmoen et al., 2006). The 2009 campaign does not follow the trend or the in-situ measurements, regardless of the application of c_{glac}. In-situ measurements suggest moderately negative net surface mass balances for that year (Kjøllmoen et al., 2010).

The slopes of both *land* and *ice* trends are not significantly affected (< +/-0.01 ma^{-1} change in trend slope) by neither DEM correction terms (c_H, c_{tile} and c_{date}), the use of alternative statistical measures to extract DEM elevations per footprint, nor application of saturation correction to ICESat elevations. Exclusion of saturated samples and application of saturation correction to the remaining dh flattens out *ice* trend slopes by 0.03 ma^{-1} and increases uncertainty (see Table 1). Including *ice border* samples only affects the ice trend if c_{glac} is not applied, but does not increase trend significance despite the increased sample number. If winter campaigns are included, the *ice* trend becomes considerably more negative (-0.43 +/- 0.066ma^{-1}, t-fit: -0.41 +/- 0.070 ma^{-1}. The same accounts for fitting a trend through winter campaign samples only (-0.42 +/- 0.092, t-fit: -0.41 +/- 0.097 ma^{-1}). Note that for comparability between winter and autumn trends single overpass samples are not excluded in the numbers here. The 2003 winter campaign had a different orbit pattern than later campaigns (Schutz et al., 2005). We found yearly varying snow heights of between 3 to 7m on glaciers, and the maximum values in winter 2005 correspond well to the overall strongly positive winter mass balance of that particular year (Kjøllmoen et al., 2006). *Ice* trend slopes are considerably more sensitive to all changes in sample composition described above if c_{glac} is not applied.

Continental glaciers east of the water divide show a more negative trend than coastal glaciers. The same is true for small (area <5km^2) versus large glaciers, and *ice* samples with pre-2000 vs. post-2000 reference DEM. The latter corresponds to an arbitrary subset in size (with a tendency of older reference DEMs for smaller glaciers) and spatial distribution of glaciers rather than a selection based on any physically meaningful criteria. The increases in trend slope amount to ~~15-35~~50-150% between these respective ~~subsets~~ subset pairs (Table 1). However, we could not find a significant relationship between dh magnitude and distance to coast. Exclusion/inclusion of entire campaigns was found to affect trends only for campaigns at either end of the ICESat acquisition period.

Note that also subsets of samples of only accumulation/ablation zones, as well as certain elevation or slope classes, would result in different trends (not shown). Such sample subsets can obviously not fulfil the requirement of representativeness for the entire glacier area and are thus not comparable to in-situ glacier mass balance measurements. Glaciers that are not in balance adjust their geometry via glacier flow which causes additional surface elevation changes that may be different for the accumulation and ablation parts of a glacier. Only sampling of the entire glacier(s) ensure that both elevation changes due to surface mass balance as well as glacier dynamics are included in the volumetric mass balance signal measured by ICESat.

The problem of biased trends due to non-representative spatial sampling by ICESat is illustrated well by the spatially clumped autumn 2009 campaign. The only glaciers that are sampled in 2009 have a strongly positive trend (Figure 7, +0.47 +/-011.ma^{-1}, in total 181 samples from Myklebustbreen and Haugabreen for autumn campaigns 2003-2009). While this trend is based on fewer campaigns (missing data in 2005 and 2007, only 3 and 7 samples for the 2004/2008 campaigns, respectively), the trend slope ~~still lies within a realistic range~~is not unrealistic (2.05/0.14 m w.eq. cumulative balance before/after data homogenisation for nearby Nigardsbreen in 2003-2009; Kjøllmoen et al., 2009; Andreassen et al., 2016). The ICESat sample on these glaciers is representative (also for single campaigns) in terms of elevation, slope, aspect and spatial distribution (within a single track that roughly follows the glacier flowline) as compared to the entire glacier area of Myklebustbreen/Haugabreen from the reference DEM. The reference DEM for this area was updated in 2008, resulting in a positive offset of the *ice* campaign mean in autumn 2009 (Figure 6).The fact that these glaciers are not at all representative for the cumulative mass balance of the entire glacier population in southern Norway explains the large offset of the 2009 campaign mean to the 2003-2008 ICESat trend.

5 Discussion

5.1 Representativeness

When combined with reference elevations from a DEM, ICESat data provides realistic estimates for glacier surface elevation change in southern Norway. However, our results bring out the importance of ensuring representativeness of the sample as well as good control over biases in reference elevations.

The ICESat sample has to be representative not only in terms of terrain and topographic characteristics that govern glacier behaviour but also data quality aspects that vary spatially. Parameters with coarse spatial patterns have largest biasing potential. Consequently, reference DEM quality and age, glacier area, and severe variations in spatial distribution of the samples were found to have potentially largest impact on glacier trend estimates. This sensitivity is a direct result from interference of the non-uniform glacier behaviour within the study area with the (coarse) spatial pattern of these influencing parameters. In contrast, parameters that vary much more spatially such as elevation, slope or aspect were found to be of less concern. Also smaller sample subsets are representative in that respect. Campaigns with low sample numbers and spatial clumping are most prone to biases. Owing to the rapidly decreasing laser power, mostly campaigns towards the end of the acquisition period are affected. However, severe cloud cover and subsequent exclusion of too many orbits can result in poor spatial distribution also for other campaigns. An example for this is the autumn 2005 campaign in southern Norway for which the only few *ice* samples mostly lie on old reference DEMs.

When relating ICESat trends to traditional glaciological measurements it is important to keep in mind that the subset of in-situ monitored glaciers and the glaciers covered by our ICESat sample might not be fully comparable. Differences in estimated mass/volume changes are therefore likely not (only) caused by the methods used, but rather a result of different sample composition. This is in line with the findings of e.g. Zemp et al. (2015) or Cogley (2009) who assign differences in mass budgets as from glaciological ~~and~~ vs. geodetic measurements to sample composition rather than method-inherent causes. We find that with ICESat's random spatial sampling (with respect to glacier locations), we capture also many small ice bodies and snow patches. The share of samples, in terms of the area of the ice bodies where single footprints lie on, accurately reflects the size distribution of all glaciers and ice patches of the total glacierised surface in southern Norway. While such small ice patches are commonly not monitored in-situ, they are likely to be equally affected by climate change if not even more sensitive (Bahr and Radić, 2012; Fischer et al., 2014). Subsequent differences in glacier volume/mass changes as derived from ICESat, compared to traditional glaciological methods on selected valley glaciers, might therefore not agree if upscaled to the entire glacier population of a study area (Bahr and Radić, 2012).

The moderately negative glacier surface elevation change trends for the years 2003-2008 fit well with overall negative net cumulative mass balance series from glaciological measurements on glaciers in southern Norway. Trend slopes are robust against applied corrections or changes in sample composition as long as representativeness of the sample is guaranteed. Given the highly heterogeneous behaviour of Norway's glaciers and the varying age of some parts of the reference DEM, both the measured dh (up to 20m) and the resulting trend confidence intervals are within an expected range. We find that smaller glaciers, and glaciers to the (dryer) east of the water divide, experienced stronger changes than larger and coastal glaciers. This is in agreement with the individual reactions of the monitored glaciers in southern Norway to the increasing atmospheric temperatures during the last decade.

To fill gaps from missing campaigns, or to increase spatial resolution of estimated glacier trends, other authors have tried to obtain an alternative trend estimate fitted through winter *ice* samples (e.g. Gardner et al., 2013). However, our results for southern Norway show that ICESat is sensitive to – and even able to reproduce – yearly varying snow depths, and our glacier surface elevation change trends are more negative for winter *ice* samples. Even though the difference between the winter and autumn trends is not significant in our study, the standard error of the winter trend is 50% larger which reflects the uncertainty added from yearly/spatially varying snow depths. Moreover, the different orbit pattern of the winter 2003 campaign (and first phase of autumn 2003 campaign) as compared to all following campaigns may cause problems with representativeness and spatial distribution of the samples, especially if spatially varying elevation corrections such as our per-glacier correction c_{glac} are applied. Our results therefore advise against including winter samples in glacier trend analyses. We also recommend including only footprints lying entirely on glaciers, i.e. excluding footprints that we classified as *ice border* samples. The signal from mixed ice/land footprints adds unnecessary uncertainty to the derived trends that does not justify the increased sample numbers.

On the example of Myklebustbreen, we show that it may be possible to detect trends even for single glaciers. Unfortunately, no mass balance measurements exist to verify the positive surface elevation change found for this glacier. How confident we can be in such a local trend depends on appropriate temporal and spatial coverage. Our results show that the applicability of ICESat in arbitrary glacierised regions does not depend on a single factor only. Likewise, the minimum region size needed to derive valid estimates on glacier surface elevation change from ICESat cannot be expressed as a hard threshold but depends on a combination of factors specific to each area: Glacier density and ICESat track density (i.e. sample size), representativeness of the ICESat sample, and homogeneity of the glacier signal within the study (sub-) region. In general, ICESat track density increases with latitude, making areas closer to the poles more favourable for ICESat studies. However, size and spatial distribution of glaciers as well as less cloud cover in dryer areas may result in large enough sample numbers even in small mountain regions at lower latitudes – as long as the representativeness condition is fulfilled. Representativeness of the sample may be given also for lower sample numbers than we found in southern Norway where a glacier population is more homogeneous in respect to its topographic setting as well as mass balance changes/surface elevation trends. Spatially varying effects such as from DEM elevation bias or highly non-uniform glacier behaviour within the study area require larger sample numbers – and thus larger region sizes – to account for the introduced uncertainty. In that regard, southern Norway may not be an ideal location to test the limits of ICESat applicability, and in other mountain regions with more consistent reference DEMs even smaller study areas may potentially yield valid ICESat glacier surface elevation change estimates.

5.2 Glacier trend sensitivity

Given the temporal variability in annual surface mass balances from NVE's long-term measurements, the glacier surface elevation change derived from ICESat data is not likely to represent a long-term trend. Our results are only representative for the development within the five years covered. It is in general not recommended to extrapolate trends derived from such a short time interval, neither for ICESat-derived trends, nor mass-balance series in general.

~~On the example of Myklebustbreen, we show that it is possible to detect trends even for single glaciers. How confident we can be in such a local trend depends on appropriate temporal and spatial coverage.~~

Trend slopes are considerably less sensitive to missing/biasing campaigns in the middle of the ICESat acquisition period than to campaigns missing at either end. ~~Still, if the two missing campaigns (2005, 2007)~~Inclusion of the non-representative 2009 campaign which diverges strongly from the assumed linear trend (corresponding to an assumed constant mass balance)~~,~~ ~~they would have the potential to~~ significantly alters the trend slope. The considerable trend slope differences for our various sample subsets show that trends are even more sensitive to changes in sample composition or applied corrections when sample numbers are small. ~~Unfortunately, no mass balance measurements exist to verify the positive surface elevation change found for this glacier~~.

~~We~~ For our data, we found that robust fitting methods, as used by e.g. Kääb et al. (2012) for ICESat glacier trends, result in comparable but somewhat steeper trend estimates as when fitting a t-distribution to the data. The error estimates of both methods overlap for all subsets/sets of corrections applied to the dataset, thus the trends are not significantly different. A t-fit better captures the heavier tails of the sample distribution and includes the uncertainties in the data within the statistical model used to compute the fit. The iteratively lowered weighing of samples within the robust fitting technique (which assumes a normal distribution) results in a similar effect – although one can argue that the weights assigned to outliers are so small that data points that don't fit the trend essentially are removed, and thus sample numbers reduced. Consequently, according to Street et al. (1988), error estimates for the robust methods might not be correct. However, given that most outliers indeed correspond to erroneous measurement of either ICESat or reference elevations, exclusion of these samples from trend estimates might be desirable. We found that error estimates of both methods are very similar, and differences resulting from the different trend fitting approaches are of the same order as caused by changes to the sample composition or due to application of correction factors. We thus prefer to leave open if robust or t-fits are more appropriate to derive elevation trends from ICESat.

5.3 The role of DEM quality and elevation errors

Of all correction factors applied, the correction for constant offsets on glaciers introduced by DEM age (c_{glac}) deserves special attention as it considerably increased the statistical significance of glacier surface elevation trends. Not only is the trend standard error halved, but the correction also makes the trend slope much more robust to changes in sample composition/elevation corrections applied. The correction thus captures and eliminates errors in the dataset that have a far bigger effect on trends than for example different fitting techniques. By applying c_{glac} we see an increase in trend slope even though the correction decreases *ice* dh. The fact that also single campaigns fit measured mass balance after application of the correction strongly indicates that this correction is important to accurately capture glacier surface elevation development within the studied time period. ~~While t~~The estimated glacier surface elevation trend of the sample without accounting for DEM age offsets is not significantly different from the former trend estimate, but the wider confidence interval, trend sensitivity, and large offsets of single campaigns, are a clear sign that not all error sources were accounted for in the uncorrected dataset. It also illustrates very well the importance of representativeness in terms of maybe not immediately obvious factors such as spatially varying vertical offsets in the reference data. Note that a correction for "DEM age", as we do it here, has a different importance on glaciers as compared to stable ground. On glaciers that change their surface elevation over time the spatially varying bias we see in our dataset is likely indeed caused by different DEM ages. On top of that other spatially varying biases due to mosaicking of data from different sources may add additional bias on glaciers. On land surfaces, the contrary is the case and the latter type of bias would usually play the main role – while the age of the reference DEM is negligible except for areas and timescales where e.g. vertical uplift due to post-glacial rebound causes relevant age-dependent bias.

Where the correction is applied on spatial units with changing elevation – such as on glaciers – a certain consistency and repetition in spatial sampling is needed. The surface change signal contribution from a glacier sampled only by one overpass is removed by the c_{glac} correction. While we found that the error from keeping the single overpass samples in our trend estimates is smaller than the uncertainty from not applying c_{glac} we recommend removing these samples as the introduced bias corresponds to a systematic flattening of the trend. It should be kept in mind that for winter trends (summer trends on the southern hemisphere) this might affect most, if not all, of the March 2003 campaign samples due to the different ground track pattern of that campaign.

Correction of per-glacier offset is only possible in our study because the glaciers seem to mostly correspond to spatial units of consistent DEM age in Norway. The correction factor is independent of (not available) metadata for data quality and does not correspond to nor help to correct offsets of the surrounding terrain. In our case, zero *land* trend therefore does not guarantee the absence of a time-dependent bias for glacier samples (with different distribution in terms of source date stamp). The assumption of a constant vertical offset per glacier is not necessarily valid everywhere – e.g. Swiss glaciers were not considered as unities in the mosaicking of airborne DEM acquisition flight lines but sometimes cut right across (Martin Hölzle, personal communication, 2015). This resulted in differently timed outlines and elevation data for parts of the same glacier, further complicating DEM differencing with historic DEMs in the Alps, as done by Fischer et al. (2015). We faced similar challenges in our attempts to co-register ICESat and the reference DEMs. The spatial units (tiles or source time stamp of elevation contours) available to us did not correspond entirely with spatial units of data origin that would exhibit a constant spatial shift or elevation error. Other DEMs for larger areas, and especially national DEMs, are likely to contain similar inherent errors as we found for the Kartverket DEM, and Fischer et al. (2015) for historic Swiss DEMs, as they all consist of a patchwork of source datasets with various time stamps – especially in remote areas. Metadata on elevation data sources are rarely available, and DEMs might have been (post-) processed to optimize characteristics other than high elevation accuracy, for instance smoothness or realistic visual appearance.

Also global DEMs, for instance the ASTER GDEMs or the upcoming TanDEM-X DEM, might be a composite of numerous units of unknown or different age or elevation biases. While the radar-based elevations from the SRTM were acquired within a short time frame which eliminates DEM age error, the DEM still remains a patchwork from acquisitions from different overpasses, and elevation differences to ICESat elevations were found to vary spatially (e.g. Carabajal and Harding, 2006). Van Niel et al. (2008) found that shifts of sub-pixel magnitude result in artificially generated elevation differences of the same magnitude as the actual, measured elevation differences between the SRTM and national higher-resolution DEMs for two mountainous test sites in Australia and China. As an additional source of uncertainty for radar-based DEMs when serving as reference elevation, radar penetration into snow and ice is estimated to be in the range of several metres (Gardelle et al., 2012; Kääb et al., 2015) and can be considered another type of spatial pattern where our per-glacier correction could be of benefit. However, further analyses on this end would be necessary, given the strong gradients and differences in snow/ice consistency between accumulation and ablation zones of a glacier that make radar penetration vary strongly even within a single glacier (Dall et al., 2001; Müller, 2011; Rignot et al., 2001).

ICESat GLAS data comes with numerous correction terms which might signal uncertainty in the elevation values. On the example of saturation correction, which is in the order of decimetres, we showed that the effect of these corrections is negligible over rough mountain terrain and not affecting our results. Moreover, the saturation flag does not necessarily correspond with lower quality data over mountainous terrain, also not on ice surfaces in the mountains. The correction might not capture the effect of waveform saturation over such terrain appropriately. It is not generally recommended for land surfaces (NSIDC, 2012), and the error potentially resulting from waveform saturation is in the order of decimetres only. However, Molijn et al. (2011) found a larger occurrence of saturated samples at the transition from (rough) glacier-free terrain to (flat) glacier surfaces in the Dry Valleys in Antarctica. This can be explained with the adaptive gain setting of ICESat's GLAS instrument: The gain of the sensor is dynamically adjusted based on the recorded signal (NSIDC, 2012) and might not adapt fast enough for an abrupt change in the recorded waveform shapes between a footprint on dark, rough rocks and a flat, bright ice surface. A preferred occurrence of saturated samples and subsequent elevation error at glacier margins, where surface elevation changes are likely more pronounced, could potentially lead to a systematic bias in ICESat-derived glacier surface elevation change trends. In our study area we could not detect a systematic pattern in the spatial distribution of the saturated samples, also not where ICESat is passing over glacier margins and experiences a land/ice surface type change. We believe that this is due to the small size of mountain glaciers and the rough surface topography both on land and glaciers (as compared to large Antarctic outlet glaciers) that never really allow the sensor to settle for a certain gain. Nevertheless, from the findings of Molijn et al. (2011) we cannot exclude that there is a potential for a systematic bias from waveform saturation at ice/land transitions in other areas, and we recommend to consider this possibility when applying our method in an arbitrary glacier region.

Likewise, other available corrections and biases of even smaller magnitude, such as inter-campaign bias (<8cm, Hofton et al., 2013), the optional range increment for land samples (*d_ldRngOff*), and the GmC correction introduced in GLAS data of release 34, are of negligible importance compared to corrections applied to the reference DEM elevations. However, it cannot be excluded that these corrections might become relevant if a reference DEM without vertical bias were available – which would eliminate the current main error source.

On stable ground, the problem of time-dependent elevation differences due to surface elevation change is not present, but the artificial dh resulting from sub-pixel shifts or elevation-dependent errors were still found to compete with real, measured differences between the DEMs. This mainly has implications on the size of spatial and temporal units needed to aggregate footprints to get meaningful results. The example of Hardangervidda illustrates the potential of results on a local scale for areas with good quality reference elevations. Thereby, spatial resolution of the reference DEM is of less importance than the absence of (spatially varying) shifts or other biases in the data, resulting in narrower dh distributions of the low resolution SRTM DEM as compared to the Kartverket DEM which seems to be of poorer quality in this area. However, the DEM resolution has to be small enough to appropriately capture the local relief variations. ~~However, i~~In more rugged terrain with large elevation variation within a single footprint, the spatial resolution of the DEM would likely play a more important role than on rather flat areas like Hardangervidda. We found the reference DEM rather than ICESat to limit e.g. more localised results that would reflect spatial variation or patterns of glacier change within the study area.

For glacier trend applications, the time to collect better reference DEMs for improved retrospect ICESat analyses has likely passed where glaciers experienced large changes in volume over the past decade. Still, the biases in the old reference DEMs of our study, originating from 10 to 20 years prior to the ICESat acquisition period, obviously became detectable and quantifiable. This fact underlines that ICESat data fully bears the potential to serve as a sample of glacier surface elevations in the 2000s even for areas where we currently do not yet have very accurate reference DEMs.

6 Conclusion

On the example of southern Norway, we show that ICESat elevations normalised to a reference DEM are fully capable to provide robust and realistic glacier surface elevation trends for the years 2003-2008 in mountainous terrain with scattered small and medium size glaciers. We estimate an average ice ~~volume loss~~surface elevation change of -0.~~34m~~ 39m +/-0.~~062~~ 07 (robust fit) and -0.~~27m~~ 33m +/-0.~~061m~~ 07m (t-fit) ice per year in 2003-2008 for southern Norway's glaciers. Our estimate corresponds very well to the area-weighted average ~~lies well in the middle of the wide range~~ of observed cumulative mass balances from in-situ and geodetical mass balance measurements on ~~10~~ 8 glaciers in the study area.

Despite sparse glacier cover of the study area, the coarse spatial sample of ICESat represents southern Norway's glaciers accurately in terms of elevation, slope, aspect, spatial location, and area of the glaciers. Representativeness of the sample is given also for individual campaigns which is a prerequisite for robust trend results. Non-representative campaigns have the potential to alter trends. Especially in terms of glacier area, ICESat samples reflect the size distribution of all glaciers in southern Norway considerably better than the (predominantly large) glaciers included in the in-situ mass balance network in Norway.

The number of ICESat footprints on glaciers (1'233 after filtering) within the study area ~~is even~~ was found large enough to allow for spatial and thematic subsampling. The considerable differences between trends from different sample subsets reflect the wide range of observed cumulative mass balances in the study area. Reasonably, we see a ~~slightly~~ more negative elevation trend of continental and small glaciers as compared to coastal or large glaciers, respectively. Our glacier elevation change trends thus capture very varied glacier behaviour within the study area, and depict also glaciers with positive mass balance, as seen for Myklebustbreen and Hansebreen. On this example, we show that it ~~is~~ may be possible to detect trends even for single well-covered glaciers, however with increased uncertainty due to spatially clumped sampling and missing data for some campaigns.

The applicability of ICESat in arbitrary glacierised regions depends on a combination of factors rather than a minimum region or sample size. The number of samples is determined by glacier density in relation to ICESat track density and the topography/climate-determined fraction of valid elevation measurements in the study region. Their representativeness, however, depends on the homogeneity of both the glacier topographic setting and their mass balance signal within the study area, as well as other spatially varying effects such as from DEM elevation bias. These factors are inherent for each region (and reference DEM) and will affect the sample/area size needed for a valid surface elevation change estimate.

Uncertainties in reference DEM elevations exceed ICESat uncertainties by a magnitude. Elevation bias of unknown spatial units of the three assessed reference DEMs add noise that match, or exceed, measured elevation differences. These biases result from sub-pixel horizontal and vertical shifts, elevation-dependent bias, and varying source time stamps of the reference DEM of up to 20 years prior to ICESat acquisition. If not accounted for, spatially varying biases in combination with varying sample distribution over time may not cancel out, and can affect the results by causing false trends. Representativeness of the sample in terms of such spatially varying bias in the reference DEM was found to be more important (and less given) than for terrain parameters like elevation or aspect. Due to their coarse spatial pattern, the DEM errors add varying but systematic bias – in contrast to the random effects from geographic ICESat footprint distribution.

We developed a new per-glacier correction to harmonise the effect of age-dependent offsets between ICESat and the patchy reference DEM of unknown, but spatially varying source date. This correction greatly increased the statistical significance and robustness of our glacier change trend, and also single campaigns fit measured mass balance after application of the correction. For national or global DEMs in other regions, we see large potential from this correction, or modified versions of it, for reducing glacier trend uncertainty related to spatio-temporal biases such as from imperfect mosaicking, orbit inaccuracies, or radar penetration.

Our study shows that ICESat analyses in mountain terrain currently are limited by the reference DEMs rather than ICESat performance. ICESat provides an accurate sample of global glacier surface elevations in the 2000s. There is still large potential, even several years after the mission ended, that new upcoming DEMs could improve ICESat analysis in retrospect (e.g. TanDEM-X, new mapping agency DEMs). After its launch, ICESat2 with its denser cross- and along-track sampling and improved performance over rough surfaces (Kramer, 2015) will have the capability to provide even more detailed and accurate valuable sample of glacier surface elevations using the methods outlined here.

Author Contributions

D. Treichler designed the study, performed data analyses and prepared the manuscript. A. Kääb designed the study and edited the manuscript.

Acknowledgements

The study was funded by the European Research Council under the European Union’s Seventh Framework Programme (FP/2007–2013)/ERC grant agreement no. 320816, the ESA project Glaciers_cci (4000109873/14/I-NB) and the Department of Geosciences, University of Oslo. We are very grateful to NASA and USGS for free provision of the ICESat data and the SRTM DEM used, respectively, to the Norwegian mapping agency for their topographic DEMs, and to the Norwegian Water Resources and Energy Directorate for glacier outlines, ~~and~~ mass balance data, and the Hardangervidda LiDAR DEM.

References

Andreassen, L. M., Elvehøy, H., Kjøllmoen, B., Engeset, R. V., and Haakensen, N.: Glacier mass-balance and length variation in Norway, Ann. Glaciol., 42, 317-325, doi: 10.3189/172756405781812826, 2005.

Andreassen, L. M., Elvehøy, H., Kjøllmoen, B., and Engeset, R. V.: Reanalysis of long-term series of glaciological and geodetic mass balance for 10 Norwegian glaciers, Cryosphere, 10, 535-552, 10.5194/tc-10-535-2016, 2016.

Bahr, D. B., and Radić, V.: Significant contribution to total mass from very small glaciers, Cryosphere, 6, 763-770, doi: 10.5194/tc-6-763-2012, 2012.

Bliss, A., Hock, R., and Radić, V.: Global response of glacier runoff to twenty-first century climate change, J. Geophys. Res. Earth Surf., 119, 717-730, doi: 10.1002/2013jf002931, 2014.

Bolch, T., Sørensen, L. S., Simonsen, S. B., Mölg, N., Machguth, H., Rastner, P., and Paul, F.: Mass loss of Greenland's glaciers and ice caps 2003–2008 revealed from ICESat laser altimetry data, Geophys. Res. Lett., 40, 875-881, doi: 10.1002/grl.50270, 2013.

Carabajal, C. C., and Harding, D. J.: SRTM C-Band and ICESat Laser Altimetry Elevation Comparisons as a Function of Tree Cover and Relief, Photogr. Sci. Eng., 72, 287-298, 2006.

Cogley, J. G.: Geodetic and direct mass-balance measurements: comparison and joint analysis, Ann. Glaciol., 50, 96-100, 2009.

Dall, J., Madsen, S. N., Keller, K., and Forsberg, R.: Topography and penetration of the Greenland Ice Sheet measured with Airborne SAR Interferometry, Geophys. Res. Lett., 28, 1703-1706, doi: 10.1029/2000gl011787, 2001.

Farr, T. G., and Kobrick, M.: Shuttle radar topography mission produces a wealth of data, Eos Trans. AGU, 81, 583-585, di: 10.1029/EO081i048p00583, 2000.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., and Alsdorf, D.: The Shuttle Radar Topography Mission, Rev. Geophys., 45, doi: 10.1029/2005rg000183, 2007.

Fischer, M., Huss, M., Barboux, C., and Hoelzle, M.: The new Swiss Glacier Inventory SGI2010: relevance of using high-resolution source data in areas dominated by very small glaciers, Arct. Antarct. Alp. Res., 46, 933-945, doi: 10.1657/1938-4246-46.4.933, 2014.

Fischer, M., Huss, M., and Hoelzle, M.: Surface elevation and mass changes of all Swiss glaciers 1980–2010, Cryosphere, 9, 525-540, doi: 10.5194/tc-9-525-2015, 2015.

Gardelle, J., Berthier, E., and Arnaud, Y.: Impact of resolution and radar penetration on glacier elevation changes computed from DEM differencing, Journal of Glaciology, 58, 419-422, doi: 10.3189/2012JoG11J175, 2012.

Gardner, A. S., Moholdt, G., Cogley, J. G., Wouters, B., Arendt, A. A., Wahr, J., Berthier, E., Hock, R., Pfeffer, W. T., Kaser, G., Ligtenberg, S. R. M., Bolch, T., Sharp, M. J., Hagen, J. O., van den Broeke, M. R., and Paul, F.: A Reconciled Estimate of Glacier Contributions to Sea Level Rise: 2003 to 2009, Science, 340, 852-857, doi: 10.1126/science.1234532, 2013.

Hofton, M. A., Luthcke, S. B., and Blair, J. B.: Estimation of ICESat intercampaign elevation biases from comparison of lidar data in East Antarctica, Geophys. Res. Lett., 40, 5698-5703, doi: 10.1002/2013gl057652, 2013.

Howat, I. M., Smith, B. E., Joughin, I., and Scambos, T. A.: Rates of southeast Greenland ice volume loss from combined ICESat and ASTER observations, Geophys. Res. Lett., 35, doi: Artn L17505 10.1029/2008gl034496, 2008.

Huss, M.: Density assumptions for converting geodetic glacier volume change to mass change, Cryosphere, 7, 877-887, doi: 10.5194/tc-7-877-2013, 2013.

Immerzeel, W. W., van Beek, L. P. H., and Bierkens, M. F. P.: Climate Change Will Affect the Asian Water Towers, Science, 328, 1382-1385, doi: 10.1126/science.1183188, 2010.

Jacob, T., Wahr, J., Pfeffer, W. T., and Swenson, S.: Recent contributions of glaciers and ice caps to sea level rise, Nature, 482, 514-518, doi: 10.1038/nature10847, 2012.

Jansson, P., Hock, R., and Schneider, T.: The concept of glacier storage: a review, J. Hydrol., 282, 116-129, doi: 10.1016/S0022-1694(03)00258-0, 2003.

Kartverket: Terrengmodeller - land, Kartverket; available at: http://www.kartverket.no/Kart/Kartdata/Terrengmodeller/Terrengmodell-10-meters-grid/, accessed: 6 June, 2016.

Ke, L., Ding, X., and Song, C.: Heterogeneous changes of glaciers over the western Kunlun Mountains based on ICESat} and Landsat-8 derived glacier inventory, Remote Sens. Environ, 168, 13 - 23, doi: 10.1016/j.rse.2015.06.019, 2015.

Kjøllmoen, B., Andreassen, L. M., Engeset, R. V., Elvehøy, H., Jackson, M., and Giesen, R. H.: Glaciological investigations in Norway in 2005. NVE report 2 2006, edited by: Kjøllmoen, B., Norwegian Water Resources and Energy Directorate (NVE), 2006.

Kjøllmoen, B., Andreassen, L. M., Elvehøy, H., Jackson, M., Giesen, R. H., and Tvede, A. M.: Glaciological investigations in Norway in 2008. NVE report 2 2009, edited by: Kjøllmoen, B., Norwegian Water Resources and Energy Directorate (NVE), 2009.

Kjøllmoen, B., Andreassen, L. M., Elvehøy, H., Jackson, M., and Giesen, R. H.: Glaciological investigations in Norway in 2009. NVE report 2 2010, edited by: Kjøllmoen, B., Norwegian Water Resources and Energy Directorate (NVE), 2010.

Kjøllmoen, B., Andreassen, L. M., Elvehøy, H., Jackson, M., and Giesen, R. H.: Glaciological investigations in Norway in 2010. NVE report 3 2011, edited by: Kjøllmoen, B., Norwegian Water Resources and Energy Directorate (NVE), 2011.

Kramer, H. J.: ICESat-2 (Ice, Cloud and land Elevation Satellite-2), available at: https://directory.eoportal.org/web/eoportal/satellite-missions/i/icesat-2, accessed: 20 November, 2015.

Kropáček, J., Neckel, N., and Bauder, A.: Estimation of Mass Balance of the Grosser Aletschgletscher, Swiss Alps, from ICESat Laser Altimetry Data and Digital Elevation Models, Remote Sens., 6, 5614, doi: 10.3390/rs6065614, 2014.

Kääb, A., Berthier, E., Nuth, C., Gardelle, J., and Arnaud, Y.: Contrasting patterns of early twenty-first-century glacier mass change in the Himalayas, Nature, 488, 495-498, doi: 10.1038/nature11324, 2012.

Kääb, A., Treichler, D., Nuth, C., and Berthier, E.: Brief Communication: Contending estimates of 2003-2008 glacier mass balance over the Pamir-Karakoram-Himalaya, Cryosphere, 9, 557-564, doi: 10.5194/tc-9-557-2015, 2015.

Lange, K. L., Little, R. J. A., and Taylor, J. M. G.: Robust Statistical Modeling Using the T-Distribution, Journal of the American Statistical Association, 84, 881-896, doi: 10.2307/2290063, 1989.

Marzeion, B., Jarosch, A. H., and Hofer, M.: Past and future sea-level change from the surface mass balance of glaciers, Cryosphere, 6, 1295-1322, doi: 10.5194/tc-6-1295-2012, 2012.

Melvold, K., and Skaugen, T.: Multiscale spatial variability of lidar-derived and modeled snow depth on Hardangervidda, Norway, Ann. Glaciol., 54, 273-281, doi: 10.3189/2013AoG62A161, 2013.

Moholdt, G., Nuth, C., Hagen, J. O., and Kohler, J.: Recent elevation changes of Svalbard glaciers derived from ICESat laser altimetry, Remote Sens. Environ, 114, 2756 - 2767, doi: 10.1016/j.rse.2010.06.008, 2010.

Molijn, R. A., Lindenbergh, R. C., and Gunter, B. C.: ICESat laser full waveform analysis for the classification of land cover types over the cryosphere, Int. J. Remote Sens., 32, 8799-8822, doi: 10.1080/01431161.2010.547532, 2011.

Müller, K.: Microwave penetration in polar snow and ice: Implications for GPR and SAR, Ph.D. thesis, Department of Geosciences, University of Oslo, Norway, 2011.

Neckel, N., Kropáček, J., Bolch, T., and Hochschild, V.: Glacier mass changes on the Tibetan Plateau 2003–2009 derived from ICESat laser altimetry measurements, Environ. Res. Lett., 9, 014009, doi: 10.1088/1748-9326/9/1/014009, 2014.

Nesje, A., Bakke, J., Dahl, S. O., Lie, Ø., and Matthews, J. A.: Norwegian mountain glaciers in the past, present and future, Global Planet. Change, 60, 10-27, doi: 10.1016/j.gloplacha.2006.08.004, 2008.

Nilsson, J., Sandberg Sørensen, L., Barletta, V. R., and Forsberg, R.: Mass changes in Arctic ice caps and glaciers: implications of regionalizing elevation changes, Cryosphere, 9, 139-150, doi: 10.5194/tc-9-139-2015, 2015.

NSIDC: GLAS Altimetry HDF5 Product Usage Guide, NASA DAAC at the National Snow and Ice Data Center, Boulder, Colorado USA, 2012.

NSIDC: GLAS/ICESat L1 and L2 Global Altimetry Data, Version 34, NASA DAAC at the National Snow and Ice Data Center, Boulder, Colorado USA, available at: http://nsidc.org/data/docs/daac/glas_icesat_l1_l2_global_altimetry.gd.html, 2014.

Nuth, C., Moholdt, G., Kohler, J., Hagen, J. O., and Kääb, A.: Svalbard glacier elevation changes and contribution to sea level rise, J. Geophys. Res. Earth Surf., 115, doi: 10.1029/2008jf001223, 2010.

Nuth, C., and Kääb, A.: Co-registration and bias corrections of satellite elevation data sets for quantifying glacier thickness change, Cryosphere, 5, 271-290, doi: 10.5194/tc-5-271-2011, 2011.

NVE: Climate indicator products, Norwegian Water Resources and Energy Directorate, available at: http://glacier.nve.no/viewer/CI/, accessed 31 May 2016, online glacier database, 2016.

Rabus, B., Eineder, M., Roth, A., and Bamler, R.: The shuttle radar topography mission – a new class of digital elevation models acquired by spaceborne radar, ISPRS J. Photogramm., 57, 241-262, doi: 10.1016/S0924-2716(02)00124-7, 2003.

Radić, V., and Hock, R.: Regionally differentiated contribution of mountain glaciers and ice caps to future sea-level rise, Nat. Geosci., 4, 91-94, doi: 10.1038/Ngeo1052, 2011.

Radić, V., Bliss, A., Beedlow, A. C., Hock, R., Miles, E., and Cogley, J. G.: Regional and global projections of twenty-first century glacier mass changes in response to climate scenarios from global climate models, Climate Dynamics, 42, 37-58, doi: 10.1007/s00382-013-1719-7, 2014.

Rignot, E., Echelmeyer, K., and Krabill, W.: Penetration depth of interferometric synthetic-aperture radar signals in snow and ice, Geophys. Res. Lett., 28, 3501-3504, doi: 10.1029/2000gl012484, 2001.

Schutz, B. E., Zwally, H. J., Shuman, C. A., Hancock, D., and DiMarzio, J. P.: Overview of the ICESat Mission, Geophys. Res. Lett., 32, doi: 10.1029/2005gl024009, 2005.

Slobbe, D. C., Lindenbergh, R. C., and Ditmar, P.: Estimation of volume change rates of Greenland's ice sheet from ICESat data using overlapping footprints, Remote Sens. Environ, 112, 4204 -4213, doi: 10.1016/j.rse.2008.07.004, 2008.

Street, J. O., Carroll, R. J., and Ruppert, D.: A Note on Computing Robust Regression Estimates via Iteratively Reweighted Least Squares, Am. Stat., 42, 152-154, doi: 10.2307/2684491, 1988.

Van Niel, T. G., McVicar, T. R., Li, L., Gallant, J. C., and Yang, Q.: The impact of misregistration on SRTM and DEM image differences, Remote Sens. Environ, 112, 2430 - 2442, doi: 10.1016/j.rse.2007.11.003, 2008.

Viviroli, D., Dürr, H. H., Messerli, B., Meybeck, M., and Weingartner, R.: Mountains of the world, water towers for humanity: Typology, mapping, and global significance, Water Resour. Res., 43, W07447, doi: 10.1029/2006wr005653, 2007.

Winsvold, S. H., Andreassen, L. M., and Kienholz, C.: Glacier area and length changes in Norway from repeat inventories, Cryosphere, 8, 1885-1903, doi: 10.5194/tc-8-1885-2014, 2014.

Zemp, M., Frey, H., Gärtner-Roer, I., Nussbaumer, S. U., Hoelzle, M., Paul, F., Haeberli, W., Denzinger, F., Ahlstrøm, A. P., Anderson, B., and others: Historically unprecedented global glacier decline in the early 21st century, Journal of Glaciology, 61, 745-762, doi: 10.3189/2015JoG15J017, 2015.

Zwally, H. J., Schutz, R., Bentley, C., Bufton, J., Herring, T., Minster, J., Spinhirne, J., and Thomas, R.: GLAS/ICESat L2 Antarctic and Greenland Ice Sheet Altimetry Data, Version 33, GLAH14, NASA DAAC at the National Snow and Ice Data Center, Boulder, Colorado USA, 2012.

Table 1. Trends and trend standard error (se), as computed from different subsets and corrections applied to the dataset ($c_{H,}$ c_{tile} and c_{glac} are applied unless specified otherwise). Footprints on glaciers sampled only during one autumn campaign are excluded except for the subsets marked with an asterisk, i.e. * corresponds to all 2003-2008 (autumn) *ice* samples.

Dataset	**Correction / subset**	**robust trend**	***se (1σ)***	**samples**	**t-trend**	***se (1σ)***
ice	(c_H, c_{glac},~~)~~ only > 1 overpass)	**-0.39**	**0.07**	**1'105**	**-0.33**	**0.07**
land	(c_H, c_{tile}/c_{date})	***+0.05***	***0.009***	*48'089*	***+0.04***	***0.009***
ice	(c_H, c_{glac}) all *ice* samples *	**-0.34**	**0.062**	1'233	**-0.27**	**0.061**
ice	c_{glac} not applied *	-0.26	0.12	1'233	-0.22	0.13
ice	Dec 2008 excluded	-0.44 ~~-0.39~~	0.072 ~~0.064~~	1085 ~~1'209~~	-0.37 ~~-0.31~~	0.071 ~~0.063~~
land	*Dec 2008 excluded*	*-0.003*	*0.010*	*44'568*	*-0.004*	*0.010*
ice	Corr Dec 2008	-0.4 ~~-0.35~~	0.07 ~~0.062~~	1105 ~~1'233~~	-0.34 ~~-0.28~~	0.069 ~~0.061~~
land	*Corr Dec 2008*	*+0.001*	*0.009*	*48'089*	*-0.003*	*0.009*
ice	Incl 2009	-0.25 ~~-0.22~~	0.065 ~~0.058~~	1'140 ~~1'268~~	-0.22 ~~-0.19~~	0.066 ~~0.058~~
land	*Incl 2009*	*+0.03*	*0.008*	*48'854*	*+0.03*	*0.008*
ice	*Sat_corr* applied, saturated samples excluded	-0.35 ~~-0.31~~	0.072 ~~0.066~~	1'001 ~~1'106~~	-0.3 ~~-0.25~~	0.075 ~~0.067~~
ice	East of water divide	-0.55 ~~-0.43~~	0.14 ~~0.12~~	242 ~~264~~	-0.54 ~~-0.39~~	0.14 ~~0.12~~
ice	West of water divide	-0.36 ~~-0.33~~	0.08 ~~0.072~~	863 ~~969~~	-0.29 ~~-0.25~~	0.08 ~~0.072~~
ice	Pre-2000 DEM source date	-0.72 ~~-0.47~~	0.16 ~~0.12~~	298 ~~373~~	-0.64 ~~-0.33~~	0.17 ~~0.12~~
ice	Post-2000 DEM source date	-0.29 ~~-0.28~~	0.076 ~~0.072~~	807 ~~860~~	-0.26 ~~-0.24~~	0.076 ~~0.073~~
ice	Including ice border samples	-0.36 ~~-0.33~~	0.07 ~~0.063~~	1'541 ~~1'669~~	-0.33 ~~-0.29~~	0.07 ~~0.063~~
ice	Including winters 03-08 *	-0.43	0.066	2'536	-0.41	0.070
ice	Only winters 03-08 *	-0.42	0.092	1'303	-0.41	0.097

ice	Samples on glaciers > 5km2	-0.28~~-0.27~~	0.089~~0.083~~	621~~654~~	-0.26~~-0.24~~	0.091~~0.085~~
ice	Samples on glaciers < 5km2	-0.53~~-0.41~~	0.11~~0.09~~	484~~579~~	-0.43~~-0.30~~	0.11~~0.088~~
ice	Myklebustbreen/Haugabreen (03-09)	+0.47	0.11	181	+0.47	0.12

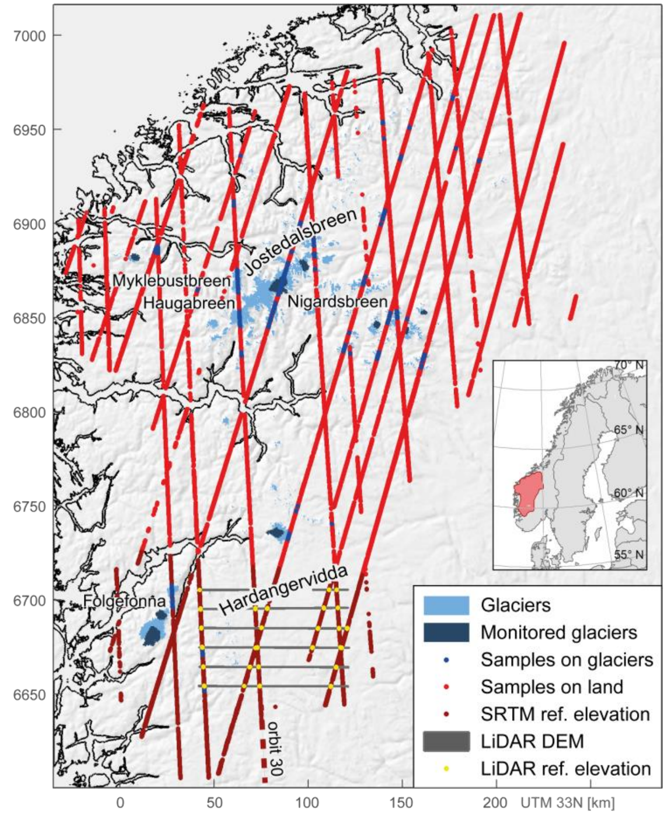

Figure 1. ICESat samples over glaciers and stable ground (land) in southern Norway. Only used footprints are displayed (no footprints on clouds or water). Glaciers with on-going monitoring by NVE are emphasised.

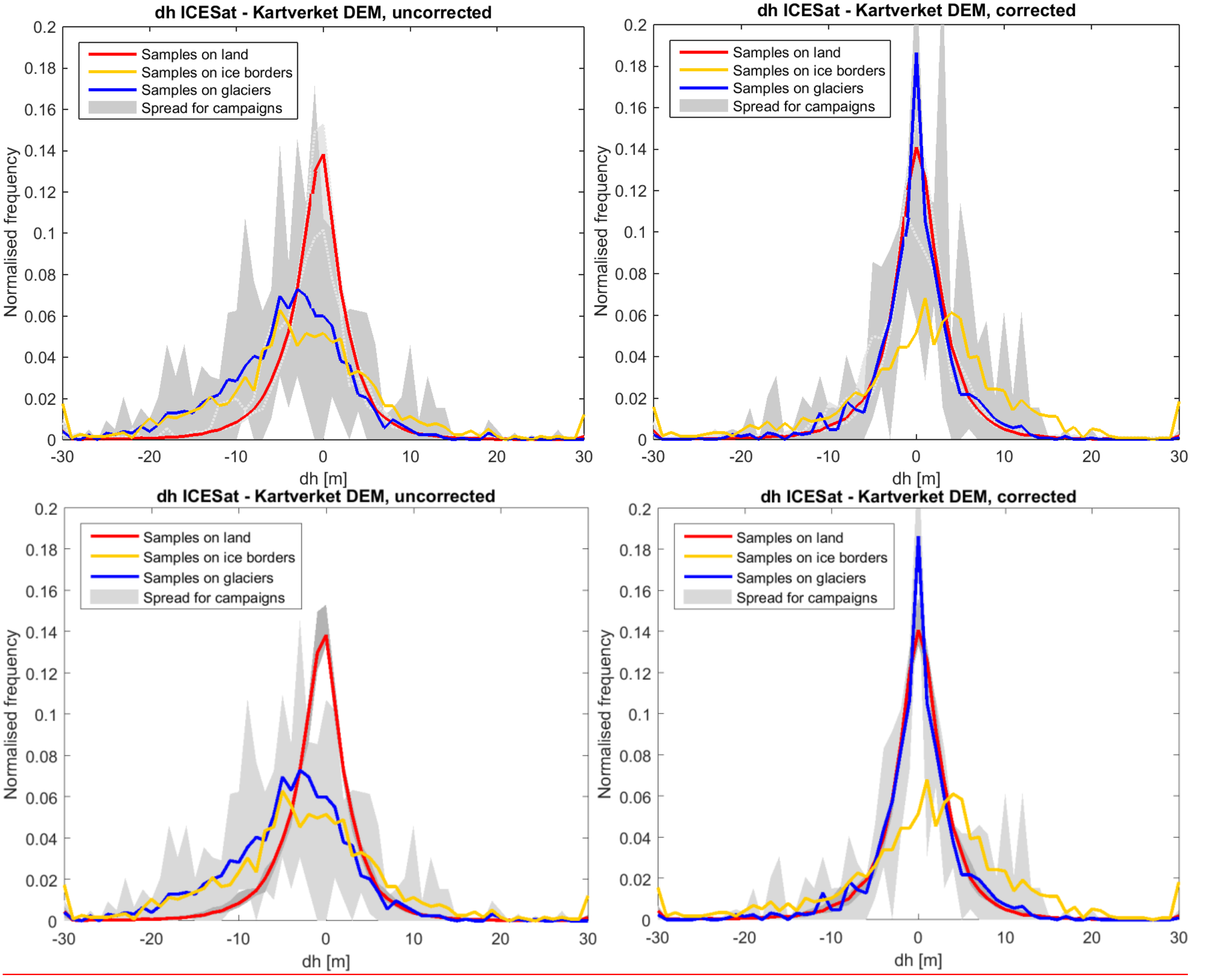

Figure 2. dh of *land* and *ice* (autumn campaigns 2003-2008)~~(all campaigns)~~ for the uncorrected Kartverket 10m DEM elevations (left) and with DEM elevation corrections (c_{tile}, c_H) and per-glacier correction (c_{glac}) applied (right). The grey spreads shows the range of distributions for *ice* (wide spread, light grey) and *land* dh (narrow spread, darker grey~~ with dotted outlines~~) of individual campaigns.

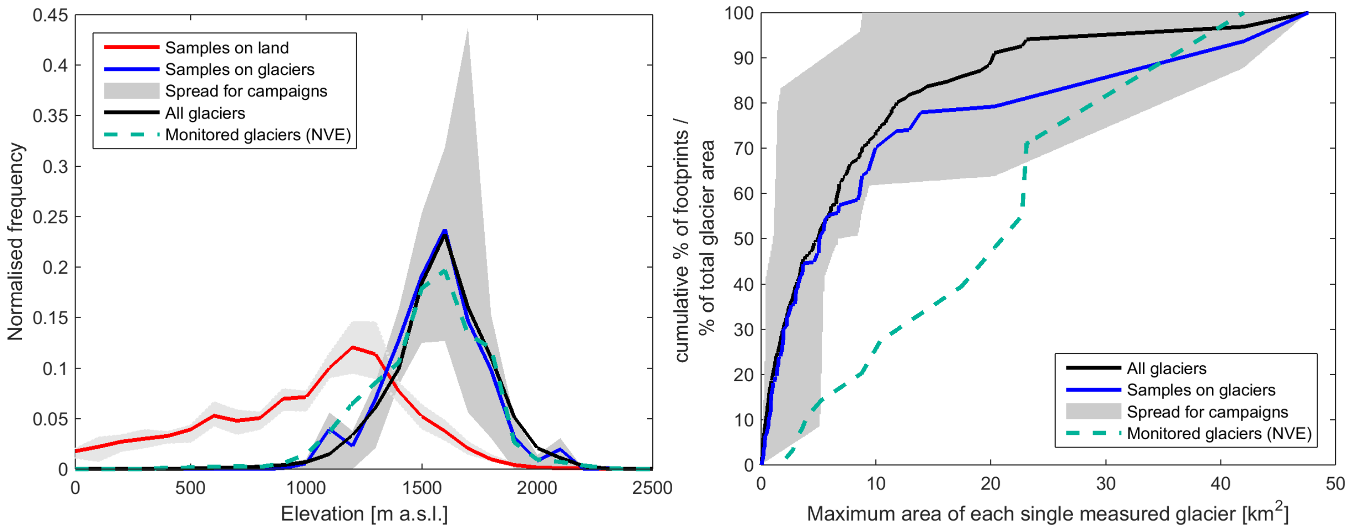

Figure 3. Representativeness of 2003-2008 ICESat autumn campaign samples in terms of footprint elevation (left) and area of glaciers sampled (right), compared to the entire glacierised surface in southern Norway, and to monitored glacierised surface (mass balance program by NVE). The grey spread encompasses the distributions of single ICESat autumn campaigns; where it is wide, the difference between individual campaigns is largest. (Reading example for glacier area comparison: 50% of the entire glacierised surface in southern Norway is made out of glaciers <5km^2, 50% of the glacierised surface where NVE runs a mass balance program is made of glaciers <23km^2, and 50% of all ICESat autumn ice samples lie on glaciers <5.1km^2.)

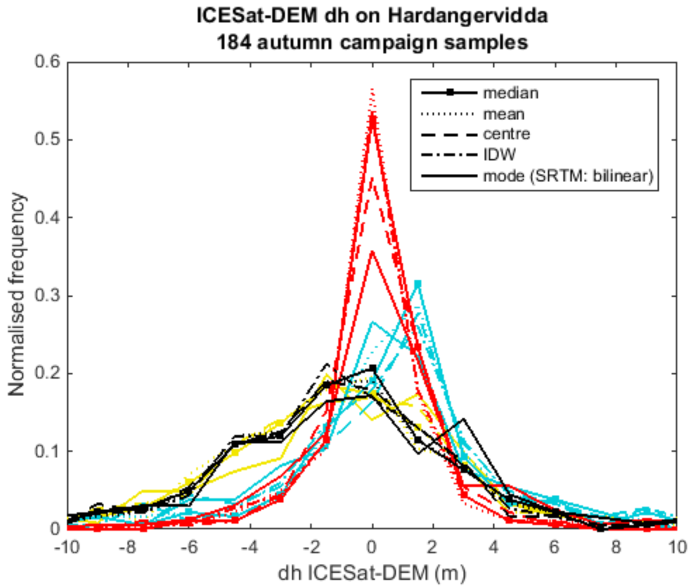

Figure 4. dh from different reference DEMs and statistical measures to summarise elevations within footprints (184 *land* samples): LiDAR 2m (red), Kartverket 10m (black) and 20m (yellow), SRTM ~90m (blue, bilinear interpolation shown instead of mode).

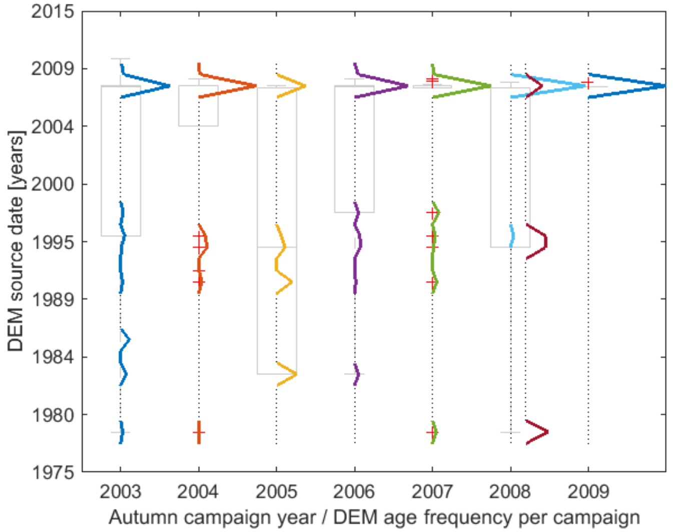

Figure 5. DEM source date distributions for ICESat autumn campaign samples on glaciers. The boxplots emphasise the average DEM age per campaign while the frequency histograms (coloured curves) reflect the relative DEM age distributions. In 2008, the October (blue) and December (brown) campaigns are shown separately (frequency histogram) and grouped (boxplot).

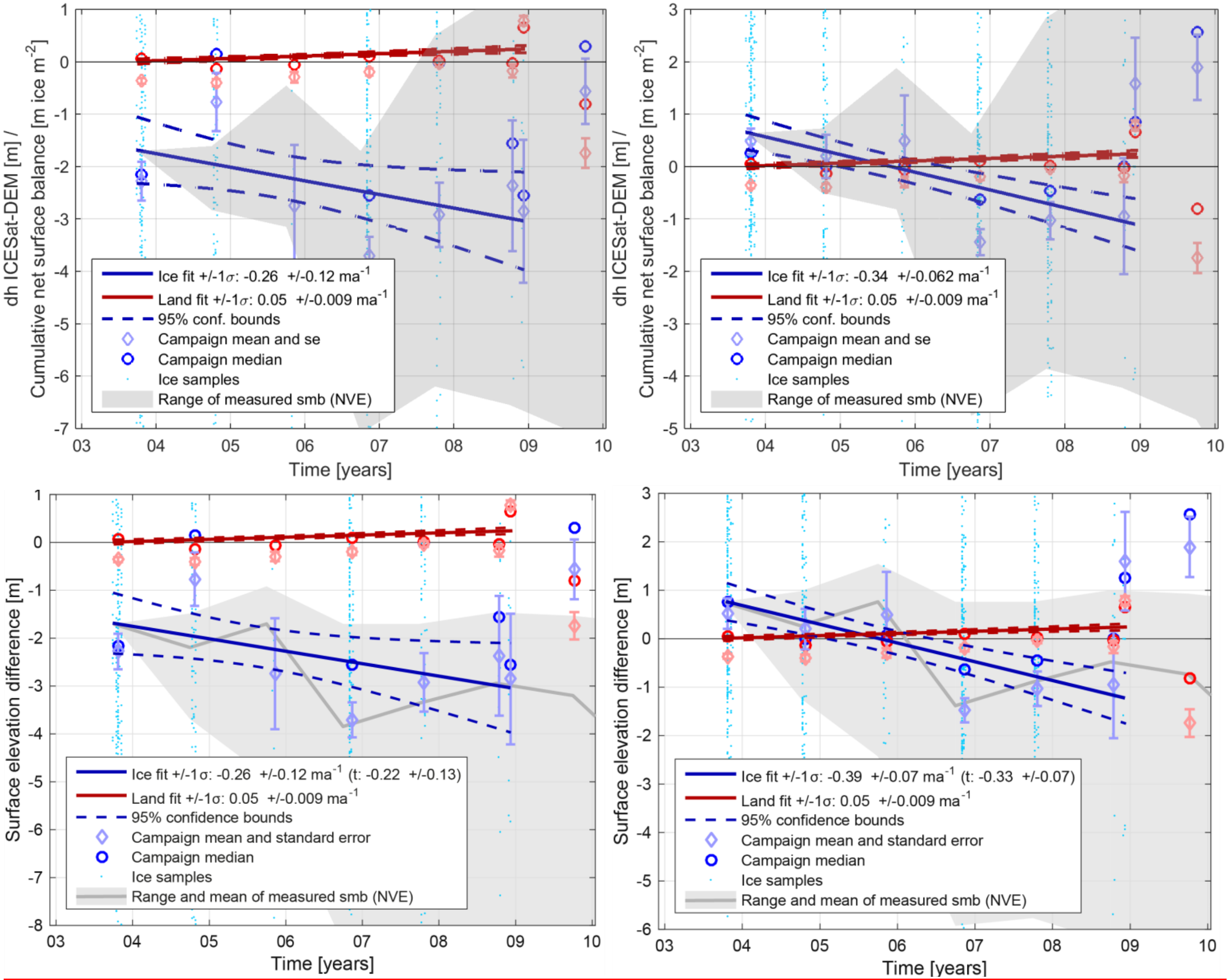

Figure 6. Surface elevation difference trends for *land* (red) and *ice* (blue) samples, respectively, for autumn campaigns 2003-2008. Left: per-tile and -elevation corrections (c_{tile}, c_H) applied, 1'233 samples; right: also per-glacier correction (c_{glac}) applied, 1'105 samples. Trends are computed from individual dh samples using robust linear regression. Campaign median and mean +/- standard error per campaign and class are shown to indicate the variability in dh per campaign. The grey spread corresponds to the measured range of cumulative surface mass balances of 8 glaciers in the area, re-converted to ice volume changes using a density of 850 kg m^{-3} (Andreassen et al., 2016), and their area-weighed mean. The data provided by NVE are based on in-situ and geodetic measurements.

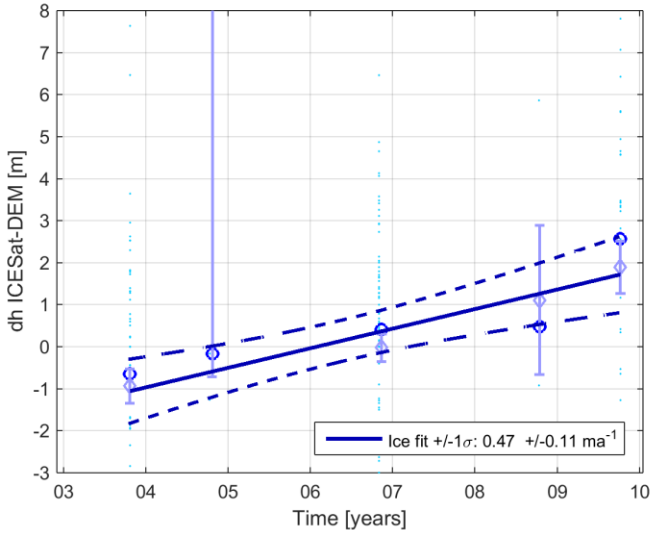

Figure 7. The autumn 2003-2009 trend for samples only on those glaciers that are covered by the autumn 2009 campaign (Myklebustbreen and Haugabreen) is strongly positive. The large error bars in 2004 and 2008 result from the very low campaign sample numbers of only 3 and 7 samples, respectively.