Reply to T. James:

Original review in gray italic.
Response as black standard text.

The manuscript by Nuth et al presents a complete and comprehensive inventory of the glaciers of the Svalbard Archipelago using the satellite record of the 2000s and basing its structure on the previous benchmark inventory from Hagen et al. (1993). The new inventory is nicely supplemented with the compilation and digitisation of older records to form two older inventories with which important change detection can be undertaken. This work represents an important step in the long term monitoring of the archipelago's ice and is the first such record of its kind in the region. It is of course important that these inventories be published and made available to the community as they will form a key future reference. The text and figures of the manuscript are of good quality and generally I believe the manuscript to be of publishable quality for The Cryosphere. However, there are a number of issues that must be addressed which I outline below, first as general comments and then more specific comments.

Many thanks to T. James for this highly specific review. Many of the points here helped clarify the poorly described sections of this manuscript. Furthermore, the detailed investigation of figures and tables truly found a number of mistakes that would have otherwise been overlooked. Thank you very much. Finally, a website address for the data has now been incorporated into the text, such that any reader may download the data presented in this paper, either in the RGI, GLIMS or from the NP data archive.

General comments:

- A main concern is the handing of the data accuracy issue. The accuracy section that is raised at the end of the results needs to be raised in the methods so this is understood by the reader before being presented with the results. These errors need to be better incorporated into the results section especially with respect to the insignificance of any length changes below 14 maˆ-1. It is difficult to interpret the results as presented without this information at hand and incorporated nto figures where appropriate.
We find it difficult to place the Accuracy section before the results since much of the accuracy is based upon comparisons discussed in the results. We use the results to derive sensitivity tests to our change parameters. If this were to come before the results, then the approach we have taken may not be understood well by the reader and will most likely add more questions rather than answer some which it does now. We understand the point that it is hard to judge the significance of the results with the accuracy section after the results, but we feel that the information provided in the results provides a significant amount of information required to understand the accuracy section, and that is more important than logical flow.

Also, the exclusion of ice cored morains is significant. Is there a way to quantify potential error introduced in excluding (the very likely) ice cored moraines?
The question of whether to include ice-cored moraines in an inventory of glaciers is questionable. The decision is dependent upon the specific use of the inventory. To calculate the total volumes of a glacier using scaling methods may require their inclusion while using the inventory to model future glacier evolution may require their specific separation. We included an additional sentence, "Without widespread ground truth information about ice below debris, it is not possible to quantify potential error introduced."

- The inclusion of a surge vs non-surge type designation in the inventory. This seems like a relatively simple thing to include based on your observations and previous publications (i.e. Jiskoot

et al., 2000). It seems that during the manual editing of the glacier outlines for GI00, each glacier was 'visited' and signs of surges may have been identifiable. It's difficult to interpret change patterns in the absence of this information. If this is not within the scope of this project than the fact that this was not completed should at least be acknowledged.

We acknowledge that incorporation of a surge database into this inventory would make it more than complete. However, it is difficult and demanding to derive a surge inventory based upon this inventory and inventory change data. While some surges may be easily visible on the imagery, surge detection is heavily dependent upon the date of the surge, size of the surge, in respect to the date of the imagery. For example, a number of glaciers have surged in Epoch 1 which is not visible distinctly in our change parameters. Furthermore, the definition of what qualifies as a surge is rather uncertain, and many smaller surge-like events have been observed on glaciers that never resulted in changing of the glacier front position and originally was not classified as surge type (see e.g. Sund et. al., 2009). Therefore, given these uncertainties, we refrain from providing a likely-flawed surging glacier inventory, and keep this work to other projects looking specifically at surging. A comment on this is now placed in the discussions.

Furthermore, we are fully aware of the difficulties to interpret the changes without the relation to a potential surge history. Especially for those glaciers that may be surge type, but have not yet been observed to surge. These shortcomings are specifically addressed in the Discussion.

- It would make this inventory and paper significantly more impactful (to the high TC standard) if the recorded parameters for each GI were listed in a table along with those of the global inventories. That would make it easier for the reader to know at a glance what was available in each, how they compare and how they can best be utilised collectively and independently.

Table 3 is now created with this information.

- Clarification is needed of how the annual change rates were calculated in terms of the multiple dates used in the different inventories. A paragraph needs to be dedicated to this issue somewhere. Dividing the total area/length changes by the averaged time span is questionable. The range of time spans is presumably large given H93 ranges from 1960 to 1980, GIold ranges from 1936 to 1977 and GI00 ranges from 2000 to 2010 but there is no mention of this. Were the average times spans calculated on a per glacier basis or overall? Does it make more sense to measure the total change at each location, divide it by the actual time span and average the change rates? This at least needs to be clarified.

The time spans were calculated for each glacier outline individually since each glacier outline contains its own dates. For regional and total inventory comparisons, these individual change rates were summarized. This is now clarified in section 3.4.

- The authors discuss the present glaciation of the archipelago through topographic and glaciological inventory parameters but much of this information is not particularly interesting on its own. There are several examples. We know the islands to the east have more of an ice cap geometry. We know what parts of Svalbard are dominated by tidewater glaciers and small glaciers etc. This general description information has not really changed since Hagen et al. (1993) and I don't think it's necessary to rehash it. The interesting things are of course the changes in the inventories and I think the manuscript results and discussion need to focus more on that and especially on how to manage the highly varying time spans.

Thank you for this comment. There are two sections that specifically address this information, Section 4.1 and the first paragraph in the discussion. While it may not be so very interesting, since this is an inventory paper, we find it necessary to aid the description of the inventory in Section 4.1.

Furthermore, this is in slight contradiction to the short comment by M. Pelto. Therefore, this information will remain a part of the manuscript.

Specific Comments

Page 2490
L5 – it's not really terminus width you have measured here though is it? Not in its conventional sense. Maybe summed average terminus width?
We agree that the definition of terminus width may be obscure. Nonetheless, we prefer not to define the word in the abstract, but hope to have made clearer in Section 3.3. Moreover, we have also used the suggestion by Anonymous Reviewer #2 to make it "total terminus width".

P2493
L13 – Not sure what you mean by coherent here? Can you rephrase to be more specific?
Have rephrased to: "We prioritize data from sensors that obtain stereo optical imagery for creation of orthophotos that are temporally and spatially consistent with the digital elevation models (DEMs) used to generate them."

L17-19 – Can it be made clearer to the reader that these DEMs are produced by SPIRIT rather than in house?
Replaced "provided within the framework of" with "generated by"

L19-20 – These are the same product but for different epochs. . . not five separate products.
Replaced "products" with "acquisitions"

P2494
L9-10 – Likewise my experience with GDEM is that it is of fairly low quality in the poor contrast areas of glaciers and ice caps. My impression is that the 'bumpy' texture describe is the result of failures of the image correlation algorithm in areas with little image information. These areas would be bumpy without merging with other DEMs and I would expect that merging DEMs would smooth these out somewhat even if the DEMs were from different times. You discuss later in the manuscript the importance of DEM errors on delineating glacier outlines/drainage basins but what is the magnitude of the errors you are filtering out? This is discussed in some detail in the appendix but I think in the methods itself a statement about the magnitude and sign of the errors that are being removed is necessary.L17-20 – Do you compare these two drainage basin data sets to assess quality? If so how do they compare?
The effects of the filtering is described in detail in the Appendix and summarized in the main text, "a low-pass Fourier filter is applied over glacier surfaces to remove the high frequency noise and minimize the size of the blunders that occur at the highest elevations." Have added a few more sentences to address these questions further: "The results of filtering reduced standard deviations in differences with ICESat and SPOT5-HRS DEMs and improved visual appearance of the GDEM without changing the overall structure of the surface. Moreover, visual comparisons between the GDEM-derived hydrological basins with those derived from the NPI topographic maps/DEMs, the SPOT-SPIRIT DEMs and individual AST14DMO DEMs reveal small variations which verify the use of the GDEM for this purpose and infer that rough DEM quality does not have a large impact on drainage basin generation. The largest discrepancies (blunders) occur on the flattest upper regions of ice cap like geometries where small differences in the elevational surface lead to large differences in the determination of a hydrological divide. These blunders are manually adjusted to the drainage basins derived from the other DEMs."

P2495

L1 – 'satellite instrument pointing'? Do you mean accuracy of the orbital parameters? The term 'spatially coherent' is vague. Can you be more specific? Does this mean there are positional errors relative to some ground truth?

It is the accuracy of determination of the satellite/instrument line of site vector, which includes both the satellite position (orbital parameters), and the instrument 3D angular pointing information (auxiliary attitude information). 'Spatial Coherency' was meant to refer to the fact that the DEMs are relative to the true ground surface (due to e.g. inaccuracies in the 3D angular pointing information of the sensor). The text has been re-written to better describe these points: "While the SPOT5 and ASTER orthoimages are internally consistent with the associated DEMs, the geolocation accuracy is dependent upon the accuracy of the satellite position determination (orbital parameters) and instrument pointing (auxillery attitude information) and thus the relative DEM/orthophoto may not necessarily be located precisely on the ground.".

L2-3 – it's not clear here or from the citations you provide how you co-register a DEM in XYZ using such a low spatial resolution data set like ICESAT. Can you either explain this here or explicitly refer to the publication that explains this. I've not read Nuth and Kääb (2011) in great detail but it's not obviously explained there.

Given the subject matter and length of this manuscript as it is now, we prefer not to describe the details of co-registration but rather explicitly refer the reader to Nuth and Kääb (2011) which is dedicated to explaining co-registration. For particular questions regarding the use of "low spatial resolution data" like ICESat, see Figure 8 and 9 of Nuth and Kääb (2011), which exemplifies the applicability of using data such as ICESat for co-registration (in fact using an ASTER DEM that is used in this study as well), and further recommends the use of ICESat for co-registration (georeferencing) because of its global consistency.

L16 – Can you specify where the automatically generated hydrological basins and topographic contours come from? The former from the ASTER GDEM and the latter from the S100?

This is now specified in the text. We use hydrological basins and topographic contours from all the data available described in the data section. That means the GDEM, SPIRIT DEMs, ASTER DEMS and the S100 maps.

P2496

L17 – why were these adjusted?

They were adjusted because the previous divides (H93) were not completely coherent with the updated velocity fields that were used in this study.

L23-24 – what descriptive, glacier and topographic attributes are available for GIold and GI90?

Added Table 3 describing this information.

P2497

L1-2 – glacier hypsometries come from what topographic data base? S100? The ASTER GDEM? If the former, wouldn't you expect significant changes since the 1990s as suggested in James et al. (2013)? If the ASTER GDEM, how does this compare to S100 and what are the effects of the 'bumpy' errors?

Added "from the ASTER GDEM". The historic elevation data is not used for hypsometry, and moreover, we prefer not to compare the hypsometries from the historic data as that would be an elevation change study and a bit out of scope of this already lengthy inventory paper.

L9-10 – why did you choose the lowermost 10% of 10 the centerline for GI00s? Is this arbitrary or based on something more substantial? Please state which/why.

Revised sentences: "The threshold is chosen visually to best represent the varying tongue shapes of both small and large glaciers. Varying the threshold by 5% has little effect except for those glaciers

that have very point glacier tongue shapes. For GI90 and GIold, if the centerline length change is greater than 10% of the earlier centerline, the average width along the area of change is used to ensure estimates are representative for the area of change within an epoch.

L27-2 – what are the known benefits of this approached as found in the citations you provide? Is the tongue width used here that which was previously calculated using the lower 10% of the centreline? Clarify.
Added, "which provides an average change across the glacier tongue rather than a single estimate dependent upon the location of the centerline". The tongue width is estimated as described in Section 3.3.

P2498
L7 – Not clear what you mean by distributions of glacier numbers.
Revised sentence, "The distributions of glacier lengths and sizes are…"

L10 – how are you defining glaciers, glacierets and snowpatches?
This is described in the second half of this sentence, "smaller than 1 km^2" and we do not distinguish between them. Glacieret is defined in the new "Glossary of Glacier Mass Balance" (Cogley et. al.2011) as, "a very small glaciers, typically less than 0.25 km^2". This has now been clarified distinctly in Section 3.2 Glacier delineation and identification.

L21-22 – I don't understand the last sentence of this paragraph (and do you mean 11 and 5?). Please clarify.
Yes, this is confusing. The 5 glacier difference actually included 11 glaciers that were not labeled tidewater in GI00s and 6 glaciers that were not labeled tidewater in Blaszczyk (2009). The sentence was rewritten to clarify this.

P2499
L3 – Worth a mention here that your fronts are shorter due to the averaging of width of the bottom 10%?. Worth stating here? It makes slightly less sense to sum the fronts in the way Blaszczyk et al., (2009) and others have but still an interesting comparison I guess.
Interesting point. I guess the way the fronts are summed does matter depending upon whether one wants the tongue width for estimating calving fluxes, or rather to study ocean/calving front interactions… Nonetheless, our smaller estimate is not due to averaging the bottom 10% since both ours and Blaszczyk's estimate are similar for Spitsbergen but rather show the largest differences on Austfonna and the ice caps of Edgeøya and Kvitøya. This lets to infer (and confirmed by visual inspection) that the difference is as stated, a result of the lobate tongues from calving fronts off the ice caps which is not captured by this method.

L8 – Does 'truncated' suggest data has been cut off. Steep tails?
No, this is a natural characteristic of ice cap geometries. I.E. Ice caps flatten off at the top, while valley glaciers become steeper towards the heads of the glaciers. This results in glacier hypsometries of valley glaciers that tail off at higher elevations while those for ice caps become "truncated".

P2500
L5 – ". . .to control coherence between the glacier upper boundaries" Not sure what this means or if it is necessary. Delete or be more explicit.
The raw glacier outlines from H93 are not available in a GIS (as now clarified Section 2.1), and therefore, one cannot compare the physical outline vertexes to determine whether the upper glacier margins are consistent between H93 and GI00s. Have modified the sentence to read, "…to control that the upper glacier boundaries are consistent between the inventories".

L27 – What is the range of time spans that are used? I'm not sure the average time span works here for calculating annual change rates.

The range of time spans can be seen in Figure 1. Each individual glacier outline has a timestamp associated with it. A paragraph is now included in Section 3.4. to describe how we handle the time stamps. We describe this change as an average time span of ~32 years, however, the calculation of this number uses each individual glacier's time span between outlines. We sum these changes to provide the average change per year, and simply state the ~32 year time span. We prefer to generalize the text given the complexity of dates and time inherent in the available data we had to generate these inventories.

L27 – so this 7% is 100% change and no error?
Yes, this is 7% change plus a random error. It does not include omission, commission errors between the dataset since we only use "comparable" glacier IDs for this comparison.

P2501
L15-16 – again, the calculation of these annual changes needs to be made clearer in the methods, in particular how the varying time spans were handled.
A paragraph is now included in Section 3.4. to describe how we handle the time stamps.

L17-18 – how did you choose this threshold and how do the interpretation errors you describe affect the quality of your results?
We wanted to be certain that the smallest glaciers and snow patches could not bias the results, partly due to the larger sensitivity of these polygon areas to error. To be conservative, we chose 2 km^2 as a cutoff. Since these glaciers have been removed, the interpretation errors do not affect the quality of these results and our interpretations.

P2502
Sec 4.4 – This is very late in the paper to be addressing the accuracy issue. This should be raised in the methods and the results of the assessment given here.
L10-20 – all this needs to be raised in the methods.
This comment was addressed above.

P2508
L10-12 – again. . . it would be useful if the relationship between these data bases and those presented here were made clear in a table.
Revised this sentence according to Anonymous Reviewer #2

P2509
L13-14 – If you don't know then I think the following statement is not necessary ". . .*which may result from a smaller matching template in the original parallax determination of the DEM generation."*
Sentence revised.

L21-22 – as previously, I'm not convinced this is due to the merging of DEM tiles but more due to failures of the image correlation in areas of low image texture.
Sentence revised.

L24-25 – how do you do this given the low spatial resolution of ICESat? I presume you are using all the ICESAT tracks? Explain or provide citation.
Reference added. [Repeated from above] The method details are explicitly described in Nuth and Kääb (2011), and will not be repeated here. Please read this article to understand the details. For particular questions regarding the use of "low spatial resolution data" like ICESat, see Figure 8 and 9 of Nuth and Kääb (2011), which exemplifies the applicability of using data such as ICESat for co-registration (in fact using an ASTER DEM that is used in this study as well), and further recommends the use of ICESat for co-registration (georeferencing) because of its global consistency.

P2519

Table 2 – First time we've seen the term 'glacieret'. Can this be defined in the text?
It is now defined specifically in Section 3.2. Glacieret is used together with snow patches to describe all the glaciers smaller than 1 km^2. Glacieret is defined in the new "Glossary of Glacier Mass Balance" (Cogley et. al.2011) as, "a very small glaciers, typically less than 0.25 km^2".

It is not clear the difference between 'glaciated area' and 'comparable area'. There are a few issues raised here in the differences between H93 and GI00 that should be discussed. If my understanding is correct, the Total Area should be the same in most cases (unless area of ice has been lost and replaced by the sea), so some of these differences are errors. What causes the discrepancies in Glaciated Area other than glacier shrinkage which is presumably represented in 'Comparable glacier area'? It is this area that is shown as Percent Area Change? I guess what I'm saying is that it's important to state what's change, what's error and what's omission in these differences.
Thanks very much for this comment, have re-written the caption to address the lack of clarity.

*"**Table 2.** Glacier statistics for the major drainage basins of Svalbard for H93 (in bold) and GI00s. Total Area is the size of the drainage basin, both glacier and land. "Glacierized area" is the total glacierized area in each atlas. Differences in this parameter include glacier change and ommission/commission errors between the datasets. "Comparable glacier area" is the area corresponding to similar IDs in both Atlases excluding snow patches and glacierets (see Section 3.2). Differences in this parameter include glacier changes and an error associated with delineating glaciers. The number of glacier units is provided only for GI00s as the number of unique 5-digit IDs (no decimals) provides the total number for H93. For GI00s, this is the number of merged integer IDs. Also shown is the number of individual snow patches (GI00s IDs=XXX99.XX) and glaciers less than 1 km2 (H93) along with the area sums. All area estimates have unit km2."*

To specifically determine the three components of change, omission and error is difficult because the physical outlines of H93 are not available and IDs were only given to glaciers larger than 1km^2 in H93 and thus it is impossible to determine which snowpatches are included in both inventories. Therefore, the use of "Comparable glacier area". Again, thanks for pointing out this lack of clarity, and we hope that the caption and text are more clear now.

P2520
Fig 1. – I like this figure; very informative. A few suggestions: (i) width of the 1990's bars are wider than the other epochs suggesting a different x axis scale which should not be the case. This shows that most images are from 1990 and some from 1991?
Corrected. The south coast of Austfonna was mapped by helicopter in 1992 (Section 2.1).

(ii) on page 2491 you say that the GIold is composed of images from 1936, 1960, 1961, 1966, 1969, 1970, 1971 but this figure suggests that only three years are used. I assume the contribution of the years not shown are small? Is there a way to include these in the figure? A compressed y axis scale? Maybe an accompanying Table?
Yea, this is difficult, as the other years besides 1936, 1966 and 1971 contain such small percentages of glacier area. I have changed the overlay of the bars, which makes it visible, but still very small. Also, the figure to appear in TC should be larger than the one in TCD.

(iii) the previous might be helped if the figure could be made larger. The smaller areas on the maps are basically invisible (i.e. areas of GI00s from the early part of 2000s). Kvitøya could go in an inset to save space; Maybe a simpler and narrower outline of Svalbard would make these areas less obscured.
Those smaller areas are the regions that are the least glacierized as well. I believe the figure in TC will be larger than that set in the TCD.

(iv) specify in the caption that grey is no data.
Thanks. Added.

P2521

Fig 2. – Kvitøya is not included in the hypsometry figures but the % sums to 100%. Source and date of the hypsometry are necessary.

Kvitøya is now included. The percentages included Kvitøya, but I forgot to include the hypsometry in the figure. Thanks for picking up on this.

P2522

Fig. 3 – This figure is very useful for describing the classification system. The inset is in the way a bit. Is it necessary? Can the boxes be shown in another figure or just add lat-long to the figure. Maybe it can be moved to the left corner. Can the elevation scale be stretched?

The inset has been made smaller. It is not covering any important information. The elevation scale has been enlarged.

P2523

Fig.4 – The exact meaning of the panels in this figure are not clear. For example, in a) initially I thought this plot was telling me that 40% of the glaciers have an area of 1 km2 and at the same time that 40% of the glaciers have an area of <100 kmˆ2. After some thought that in the latter case you mean 40% of the total area is made up of glaciers that have an area of <100 kmˆ2. Maybe two axes; one with glacier numbers the other with area would be more obvious? Also the y axis scale labels are misaligned. In b) and d) do you mean the percent of the total area that is (tidewater) glaciated? Make all three of these more clear in the caption and in the text on Page 2498.

Again, Thanks for the suggestions. The figure and caption has been modified.

P2524

Fig. 5 – Is the aspect degrees from north clockwise? I think this would tell a different story if it was by area rather than by number of glaciers. The area of north facing glaciers seems quite small in comparison to the area of south facing glaciers. Units should be added to the x axis of the histograms.

We agree that the story is different if using all pixels of the DEM for this histogram (this represents area rather than number of glacier) rather than the average per glacier. The last sentence in the caption addresses the DEM pixel histogram of aspect which is uniform. The histogram thus reflects the dominance of small glaciers facing northward. . Figure and caption has been revised and the units are clarified in the caption.

P2526

Fig. 7b) – hard to get anything out of this figure. Maybe making the grey dots red would help? They are hard to see. Here you show change in relative units but in the text in Section 4.3 you discuss length changes in mˆa-1 which is more intuitive.

The dots are made red, as suggested. Fig. 7a. shows the units as m per year, while Fig 7b was just an example of the relative changes as discussed in Section 4.3.

P2527

Fig. 8 – This is another example where it's not clear how the varying time intervals were handled. In a), the dashed line divides between a longer and shorter time interval. Were two separate time intervals used to calculate the annual changes? Nice highlighting of the likely surge-type glaciers here.

A paragraph is now included in Section 3.4. to describe how we handle the time stamps. The line here is to show roughly where the time intervals vary greatly.

Comments accepted and modification made:

Page 2490
Line 2 – Is this 100% of Svalbard's ice cover which covers 57% the total land area? This needs to be made clearer.
L12 – Is 'temporal' necessary here? Maybe 'annual retreat rates' if just 'retreat rates' is not enough.
L25 – are 'outlines' or 'polygons' better terms than 'borders'?

Page 2491
L4-5 – are these citations examples? If so they should be preceded by e.g.
L22 – Kohler et al (2007) only demonstrated volume losses in a relatively small part of the west. James et al. (2013) measured increasing volume loss at various locations in the archipelago including the higher rates in the west. Isn't this a more appropriate citation?

P2492
L5 – estimated
L6 – consider moving 'derived for two epochs' before the list so it's clear you are not just doing this for the 3rd parameter.
L10 – consider using analogue or analytical photogrammetry. The term manual photogrammetry is a bit meaningless since even with the most modern digital photogrammetry, some components are manual.

P2493
L25 – . . .required to complete coverage of the archipelago.

P2497
L8 – intersected with the glacier outlines for each GI?
L9 – I think 'measured' is more appropriate than 'estimated'
L21 – maybe specify percentage change rates?

P2498
L17 – with a calving front
L19 – tidewater terminating
L23 – you're not really summing actual front widths here. You have averaged the widths over an area. Here and elsewhere in the manuscript it reads as if the number you provide are actually length of the calving front which is not true.
L25 – 'perimeter' implies a closed polygon. Maybe length, width or extent?

P2500
L7 - 60% of the archipelago's land area
L17-24 – this is really for the methods. You elude to the difference in the treatment of snow patches this in the methods but it's not clear to the reader until this section is read. L28-2 – should this also be specified in the methods?

P2501
L4-6 – This statement is pretty obvious. For meaningful changes it really has to be absolute area loss. I've never been sure of the significance of relative changes in this context.
L12 – Is this 30% for all of Svalbard or just southern and western Spitsbergen?
L21-24 – It should be mentioned in the methods when you introduce the area/width parameter what the benefits of this approach are over centreline changes alone. It's hard to know how to interpret this and what it is telling us that the centreline width does

not.

P2503
L26-29 – something is wrong with this sentence.

P2506
L13 – I think it's redundant to use the term 'temporal' with 'rates'. You use this combination elsewhere in the manuscript. I think just area/width retreat rates is sufficient.

P2508
L19 – for a sample of _400 glaciers in the south/west Spitsbergen

P2531-2
Fig. 11 – Move the letters for the sub figures before the text to be consistent with the other figures. x and y axis of e) and f) need labels.