We would like to thank Dr. Helm and the anonymous reviewer for the time and care they have taken with their reviews. Their comments are very helpful and will certainly improve the paper. Their contribution, and the help of the editor, Dr. Berthier, will be acknowledged.
We have copied all the comments from both reviewers below, and added our response in italics.

Anonymous Referee #1
Received and published: 9 July 2015
This paper develops estimates of surface elevation change for five Arctic ice caps, using radar altimetry data acquired by the CryoSat-2 satellite since 2010. The authors provide a detailed analysis of CryoSat's capability to measure elevation changes at both a monthly and annual sampling frequency, and evaluate their findings using a range of field and airborne datasets. Particular attention is given to the impact of changing snowpack conditions on the retrieved elevations.
I found the paper interesting, informative and comprehensive. The manuscript is very well written and has clearly been thoroughly proofed by the authors prior to submission. I would expect the work to be of relevance to a wide audience of The Cryosphere, in particular to readers with an interest in the current evolution of Arctic ice caps and the performance of satellite radar altimeters. More widely, the results are relevant to anyone with an interest in geodetic estimates of ice sheet, glacier and ice cap mass balance. I have no major comments and would therefore recommend the manuscript for publication. I have listed some minor comments below, which I hope will help to clarify certain aspects of the work.

Minor comments
P2822 line 6: Suggest "contribution of subsurface to surface" => "ratio of subsurface to surface".
Wording changed as suggested.

P2823 line 4-6: ICESat and NASA acronyms are not defined.
Acronyms defined.

P2827 line 21: 0.13 m w.e. / yr?
Fixed
P2828 line 6: "indicate" => "indicates".
Fixed

P2829 line 11: "variation in the time history of the illuminated area". Not sure I understand exactly what you mean here. Do you mean variation with time of the area illuminated by the wavefront?
Yes, the wording has been changed as follows to make this clear: … 'This complexity is not entirely unexpected and arises due to the nature of the surface being measured, in particular the possible variation of the illuminated area at the sampling times in the receive window, and the possibility of reflections from sub-surface layers.'

P2829 line 19: Worth noting that Davis's measure of repeatability was based on a single-cycle cross-over analysis and so, as I understand it, his low threshold retracker showed greater repeatability with respect to variations in antenna orientation between the ascending and descending passes (perhaps because of reduced sensitivity to subsurface backscattering anisotropy), but not necessarily improved repeatability with respect to changes in time (as is desirable for elevation change detection).
Thanks for the clarification. The Davis et al. paper has been re-checked and the wording has been changed appropriately…
'The choice of the threshold level retracker used by Helm et al. (2014) for their work in Greenland and Antarctica followed that of Davis (1997), who advocated a threshold retracker to help minimize the influence of subsurface returns on the detected elevation.'

P2832 line 20: By 'better' do you mean that it gives a lower standard error of the mean, because of the larger sample size? This would presumably reflect a better precision on the measured mean elevation change, but not necessarily the most accurate measurement of the actual surface elevation change, because the measured elevation change may include the effects of changes in the dominant scattering horizon? In some cases, when melt occurs every year, then perhaps differencing summer elevations might give a more accurate measure of the real surface elevation change because of the negligible bias between the real and detected surface. I'm not expecting the authors to redo any analysis but I am more interested in their opinion on how to best minimize the impact of time variations in the bias.
Year-to-year surface height change can be achieved, as stated, by differencing summer minimum heights (small sample but, with surface melt, less uncertainty as to the level being measured) or by differencing average winter heights (large sample but now with increased uncertainty as to the dominant reflecting layer). It is dangerous to generalize too much but our results suggest that the latter is normally better as the 30 day averages encompass many fewer points, are consequently nosier, and may not capture the true minimum summer elevation. Perhaps the best approach is to use all available information, including field work and meteorological records. Some wording has been added to help clarify this…
'This is a consequence of the advantage obtained by averaging the many samples obtained over the larger time period in comparison to the fewer acquisitions possible in the 30 day period, which are then noisier and may not capture the true minimum surface elevation.'

P2833 line 13: 'This' => 'These'. *Fixed.*
P2833 line 14 'was' => 'were'. *Fixed.*

P2835 line 12: 'Changing historical meteorological conditions' I don't find particularly clear. Do you mean year-to-year changes in meteorological conditions?
The wording has been changed to clarify this sentence. We were not referring to just year-to-year changes but also the possibility that changes during the winter could influence the bias between the surface and the CS2 detected 'surface'…

'Changing meteorological conditions; accumulation, storms, heavy snow falls, etc., could change the bias between the CS2 detected surface and the true surface, even during the winter. We expect that the magnitude of this variable bias may be dependent on the winter accumulation and the variability in conditions.'

P2835 line 21: 'CryoSat-2' => 'CS2'. *Fixed.*
P2836 line 17: 'Devon Ice Cap and Austfonna'. *Fixed.*

P2837 line 13: If I understand correctly, you are comparing an AWS point measurement with a CS2 spatial average calculated over several thousand square km. Do you think the different sampling scales could explain some of the observed discrepancies, for example through spatial variability in meteorological conditions within the NW sector?
There is certainly a huge scale difference between the essentially point AWS measurement and the CS2 average measurement. Nevertheless, other AWS data and the mass balance pole measurements imply that the spatial variability in conditions across the NW sector is not large, although the local topography around the AWS could (slightly) affect the local accumulation.

P2838 line 17: If the backscatter is dominated by the previous end of summer layer, then could another possible contribution to the decreasing height over winter be from the downward motion of the previous summer layer as a consequence of firn compaction?
Yes, this is a possibility, and we have added text to provide a more complete explanation of the possible causes of the CS2 height decrease…
'It is possible that the changing nature of the winter accumulation reduces the surface reflectivity in relation to the volume component, such that the bias between the surface and the CS2 detected height increases during the winter. If the previous summer melt layer remains as the dominant backscatter layer then the apparent height could decrease because of firn compaction and the additional two-way path length due to the permittivity of the winter snow layer. This could then contribute to the apparent decrease in surface height seen in the 2012/13 winter.'

P2838 line 28: Here you could refer to http://onlinelibrary.wiley.com/doi/10.1002/2015GL063296/full.
Yes, this recently published reference is certainly relevant, and has been added…
'The influence of changing conditions on the apparent CS2 detected elevation was also observed with the low resolution mode (LRM) data in Greenland after the extensive 2012 melt (Nilsson et al., 2015).'

P2839 line 17: These are referred to as height changes and not height decreases, and so the values should presumably be negative?
Yes, the sign has been changed in the text. This problem also existed elsewhere in the text, and has been fixed.

P2840 line 4: Is the trend with respect to elevation significant given the dispersion of the height change data?
Yes, we believe so. Both polynomial fits and segmenting into different elevation bands leads to the same conclusion. This conclusion is also consistent with field measurements.

P2841 line 19: Why July 2010 to Dec. 2011 and not an integer number of years?
This choice is related to illustrating the height loss in basin 3 associated with the surge. The larger the data set (time wise) prior to the beginning of the rapid motion the more likely there will be adjacent pairs of points with which to estimate the average height loss.

P2842 line 8: I can see that a specular surface would increase backscattered power, but wouldn't moisture in the snow result in more microwave absorption and reduced backscattered energy?
Yes, the wording here was poor. As you suggest, we think that moisture in snow would initially lead to a reduction in the backscatter. However, when there is sufficient melt to create a truly wet layer then strong coherent reflection is possible. The wording has been improved to reflect this…
'On Barnes Ice Cap the relative maximum power of each return waveform shows increased power and dynamic range in the summers (Fig. 13a), which we interpret to be a consequence of significant melt and the possibility of a specular return from a wet surface. Initially moisture in the snow can reduce the backscatter but with continued melt and the creation of a wet surface there is the possibility of relatively strong coherent reflection.'

P2843 line 24: remove 'at'. *Done*

P2844 line 12: I appreciate this probably has no straightforward answer, but I wonder whether the authors have any thoughts on the extent to which their conclusions are specific to their chosen retracker, or are generalizable to other retrackers?
A complete answer to this question would require systematic testing of different retrackers, something beyond the scope of our work. Although some of the details of the seasonal change in elevation, e.g. summer-winter, may change slightly with the form of the retracker (c.f. Ricker et al., 2014, 'Sensitivity of CryoSat-2 Arctic sea-ice freeboard and thickness on radar-waveform interpretation'), we suspect that any CS2 detected elevation will be more dependent on the changing conditions than on the detailed form of the retracker. Consequently, we think our conclusion that the bias between the surface and CS-2 detected elevations depends on the conditions of the surface and near-surface is generalizable to other retrackers.

P2844 line 14: 'between the mean CS2 height'?
Changed to '… variable bias between the mean detected CS2 elevation and the surface elevation…'.

P2844 line 25: Might be worth commenting on the value of the length of the time series.

I'd expect that the magnitude of the bias is constrained by the physical properties of the snowpack and microwave penetration, and that the bias varies most over seasonal to annual timescales depending upon changing meteorological conditions. Therefore over longer time periods, would you expect the relative influence of the variable bias on the measured height change to diminish?
Yes, we would expect that over a sequence of winter-to-winter CS2 height comparisons that the effect of a possible variable bias would diminish, but we should still be aware that slowly changing conditions may have an influence on the year-to-year CS2 detected height change. In the light of the comment we have expanded the text as follows…
'Notwithstanding the uncertainty in the bias between the surface and CS2 elevation, the winter-to-winter CS2 height change results can give a credible estimate of the year-to-year ice cap surface height change, particularly as more years are added to the time series. The largest uncertainty in these estimates, and the most difficult to quantify, comes from the fact that the conditions winter-to-winter may change in a manner that affects the bias between the surface and the CS2 elevation. Surface field measurements under cold spring conditions may help identify differing winter conditions that could lead to a changing bias between the surface and average CS2 elevations.
The results for the Canadian ice caps…'

P2844 line 27: 'A change: : : volume change'. This sentence seems a little out of place as the manuscript has focused on elevation change and not mass change.
This is true, so we have removed the sentence.

P2845 line 16. I don't find this final sentence particularly clear. Do you mean the interferometric capability to locate POCA within the beam footprint, or something different?
We believe that the improved resolution and interferometric capability of the SARIn mode of Cryosat allows the user to identify the POCA position more accurately than with previous altimeters, and that the temporal height changes we have shown in this work are only possible thanks to the ability to better geocode the POCA footprint. The text has been expanded to help make this clear…

'In summary, we believe that the improved resolution and interferometric capability of the SARIn mode of Cryosat allows the user to identify the POCA position more accurately than with previous altimeters, and that the temporal height changes we have shown in this work depend to a large extent on the ability to better geocode the POCA footprint.'

Table 1: Consider splitting glacier facies types with commas or new lines.
Commas added.

Table 1: Would be helpful to give a brief description of the data type that each DEM is derived from.
Added as footnotes below table

Table 1. Is ‘high temporal resolution’ 30 days? If so might be worth mentioning this explicitly.
Done.

Figure 2: Replace CS with CS2 to maintain consistency with acronym used in text.
Done.

Figure 4 Replace CS with CS2 and STD with SD to maintain consistency with acronyms used in text. To avoid any misinterpretation, suggest you define more explicitly, perhaps in the caption, what <delH> is.
Done.

Figure 6 caption: ‘stations is indicated’ => ‘stations are indicated’.
Corrected.

Figure 6: I know ‘A’ is described in figure 7 but would be helpful to also mention what it refers to here.
Done.

Figure 7: Panels a and b are lacking vertical axis labels to specify units.
Fixed.

Figure 7: My understanding was that the horizontal black dashes in panel a were meant to mark the 30 day sampling periods. However, the Nov-Dec 2011 dash seems to cover a longer time period.
Yes, this is correct. Sometimes the data in any one 30 day period is so small that it was better to reduce the time resolution and amalgamate with adjacent data time-wise. The text both in the paper and the figure caption has been changed to explain this. The caption now includes… ‘The short dashed black lines at the top indicate the time periods encompassing the CS2 passes which have been combined for the high temporal resolution plots. In one case (Nov.-Dec. 2011) passes were missing so the two groups were combined.’

Figure 7: I can’t work out where the blue AWS B data are from. It doesn’t appear to be specified in the caption or visible in Figure 6.
Label added to Fig. 6, and appropriate text to the captions.

Figure 7: Why have you chosen to compute year-to-year height change for AWS A rather than AWS B?
The elevation of AWS ‘A’ is closer to the average of the elevations used for the average CS2 height change data.

Figure 7: AWS B height change looks extremely stable between Dec 2012 and May 2013. Is this real or have the data been interpolated?
While the plot does looks quite ‘suspicious’ we have checked the AWS B data and believe it to be real..

Figure 8 caption: ‘along the north-south transect shown’?
Fixed.

Figure 8a: Labels on colour bar are difficult to read.
Fixed.

Figure 10: ‘has been removed’ => ‘have been removed’.
Fixed.

Figure 10: Is background image a DEM?
Yes, and this has now been acknowledged in the caption.

Figure 13a: dashed black line is hard to see, consider changing to a more visible colour.
Fixed.

Figure 14: Suggest adding units to elevation ranges in panels a-d.
Units added.

Figure 14 caption: ‘The winter-to-winter average height change’.
Fixed.

V. Helm (Referee)
veit.helm@awi.de
Received and published: 12 July 2015
In this paper monthly and inter-annual surface elevation changes of 5 Arctic Ice Caps are presented. The results are derived by a detailed and comprehensive analysis of CryoSat2 data acquired in the period 2010 to 2014 and evaluated to field, meteorological” and airborne data sets.
Special focus is given to seasonal changing surface conditions causing a shift of the main radar scatter horizon, where the radar elevation is tracked. This is a very important finding, which needs to be considered when estimating elevation change time series from radar altimetry, since it demonstrates that observed elevation change might not necessarily be a true surface elevation change.
C1188
The paper is well written, clear, concise, well-structured and understandable and the data processing and analysis of high quality. I think the paper is of interest to a broader community with interest in evolution of Arctic ice caps and altimetry.
I have no major concerns and would recommend the paper for publication in TC. In the following I have added some comments and questions that came into my mind when reading the paper.

P2827L4: accumulation rates of 0.5m ! m/a
Fixed.

P2827L21: accumulation rate of 0.13m ! m/a
Fixed.

P2829L10-15: I would expect that internal layers are not resolved by CS2 and therefore not show up as single peak. Maybe they broaden the waveform due to increased volume scattering. I would expect that peaks in the later part of the waveform are more due to undulating surface within the CS2 beam width.
Thanks for the comment. We suspect that more studies of different areas will help in understanding the shape of the waveform leading edge, in particular the balance between the influence of the topography and volume backscatter. We agree that the variations in return power beyond the leading edge are related primarily to the surface topography.

P2829L21-24: It would be of great value if you show the results you obtained by the comparison between the two methods. Are they within the error bar of the derived elevation changes or do they differ significantly? From the sentence it seems to me that your new approach using the max gradient in the leading edge to re-track the elevation eliminate the problem of a variable bias between the detected elevation and the physical surface. But in the rest of the paper your findings are clearly showing that this is not the case.
A systematic comparison of retrackers is beyond the scope of this work. Indeed we acknowledge that the sentence referred to implies more testing than was actually done. Consequently we have changed this to… 'Limited tests on some of the ice caps in our study have shown that a threshold retracker also produces satisfactory results but still does not totally eliminate the problem of a variable bias between the detected elevation and the physical surface.'

P2830L12-16: This method is an excellent way to improve the quality of the SIN CS2 elevation data used in the analysis, since you get rid of large outliers without neglecting them. To demonstrate the performance of your method you could add a comparison when using the method with and without a DEM.
Unfortunately, in our code it is difficult to turn off the reference DEM feature, so it isn't straightforward to do as you suggest. However, we know that the official L2 product for these ice caps does include some poor results that are clearly related to 2π phase ambiguities, resulting in the footprint being mapped to the wrong side of the sub-satellite track. With a reference DEM one can usually avoid this situation. Also, the advantage afforded by a reference DEM will be ice cap specific as the potential benefit depends on cross-track slopes. In defense of the ESA L2 product it would be very difficult to include reference DEMs for all the areas covered by the SARIn mode.

P2830L19: Reference DEMs: Could you please add some more details about origin, time of creation, the quality, and resolution (slightly update table 1 should be fine).
Some of these details, and relevant references, have been added to Table 1.

P2830L28: How close are the CS 30 day sub-cycle tracks (are they within 1km or 5km)?

At the latitude of Agassiz or Austfonna (~ 80° N) the 30 day ascending or descending tracks are separated by ~ 5 km but this separation increases as the latitude decreases. At 70° N, the latitude of the Barnes Ice Cap, the separation has increased to ~ 10 km.

P2831L9: How many data points usually fall within a 400m footprint within a 30 day subcycle?
Normally a few hundred, but this varies a lot depending on the size of the area being studied, the latitude, the number of passes in the time window, and the temporal separation of the 2 time windows (the middle of which are usually separated by a multiple of ~ 30 days). Having both ascending and descending passes helps provide sufficient height comparisons.

Are you averaging all elevations within one footprint of one time period and subtract the averaged elevation of the second period or do you average all possible combinations of height differences for the two time periods within one footprint? Maybe add an equation to make it clearer.
The average CS2 height change for a particular area between the two time periods is estimated as follows: Initially each point (the centre of a CS2 footprint) in one time period is compared to all the points in the other time period on a point-by-point basis. If the distance between the points is within the preset limit (usually 400 m), the height difference is stored and corrected for the slope between the two footprint centres. This is done using the reference DEM. When all the height differences are collected the mean and standard deviation (SD) are calculated and any pair with a height difference greater than ~ 10 m from the mean (which is larger than 3 standard deviations) is discarded. The mean and standard deviation are recalculated and stored. This is done for all the possible time period combinations. This approach has the important advantage that bad height estimates can be identified and discarded easily. In this regard there is no guarantee that 'bad' CS2 points (arising from 2π ambiguities, high slopes, low coherence, etc.) will have the same mean as 'good' values, and it doesn't take too many 'bad' points with the same sign to affect the mean. The disadvantage of the above approach is that it can be computationally tedious. The appropriate text (second paragraph in section 3.2) has been expanded, like the above, to make this clear.

P2831L14 and P2832L9 When you derive an elevation rate (m/a or m/month) for temporal height change analysis (month to month) it is important to know the averaged time tag next to the averaged elevation difference. I could imagine that this difference of the averaged times are not exactly 30 days and could vary from a couple of days.
Are you considering this as well?
Yes, initially the temporal distribution of data is examined and if there are missing passes then the temporal window can be expanded to create a larger data sample. The black lines at the top of lower panel in Fig. 13 illustrates the time periods over which data were acquired for the Barnes Ice Cap. Note that the October - November time period is longer than the others, this is a results of missing passes. Consequently, the date associated with each time period is always the average of the dates of all the data in that time window. Again we have modified the text in the paper to make sure this is clear.

P2832L9: “relatively small number of data samples possible in 30 day periods:“ ! What means small number? Do you use any criteria to neglect points with too low coverage?
Typically there are a few hundred points in each ‘30 day’ period depending on the size of the area being studied. If the data set for the height differences between any pair of time periods contains less than 10 values it is discarded. Again some text has been added to clarify this step in the process.

P2833L15: Technical University of Denmark (TUD) acquired the ALS data during the ESA CryoVex campaign. TUD also processed the GPS and ALS data, which is a lot of work. In the acknowledgments TUD is listed but I think it would be good to add a sentence directly in the text for this effort.
Yes, the contribution made by TUD with the ALS and GPS data (and AWI with ASIRAS) is significant and does warrant further acknowledgement. This has been done.

P2833L25: As mentioned before. Could you show the statistics without using your CS2 SIN processing DEM approach to detect and correct for phase uncertainties? Do you see similar mean and SD when applying the re-tracking at a certain threshold like Helm et. al. 2014 as you discussed before?
As explained above we are reluctant to include results or statistics in a situation where there hasn’t been systematic study. We believe that the improvement obtained by including a reference DEM will be ice cap specific, indeed it is likely that the DEM is most important in situations in which the cross-track slope can be close to, or above, the slope that leads to phase ‘wrapping’ (~ 0.55°).

P2834 Error estimation: This is a good error discussion but I miss the explanation how you derived the uncertainty for the elevation changes. You mentioned 5 points, which need to be considered. It’s not clear to me if you have considered those points and if yes, how exactly you did this.
Good comment! Yes, we did consider all the points listed in section 4.2 but admit that the errors included in Table 1 have not been carefully justified. The various uncertainties for each ice cap were all estimated and then root square summed. The trouble is some of our potential errors are very hard to quantify and perhaps are best described as ‘best estimates’. While we admit this is less than ideal we have added text to justify the approach. Many papers in this field assume random errors, and then quote an error based on the standard deviation divided by the square root of the number of samples (the ‘standard error’). In our case, for some of these ice caps, we believe that this would lead to an optimistic estimate of the errors due to our concern that some of the errors are not random.

E.g. in Fig. 7 and 11 there are no error bars but in Fig. 13 you show the error.
The error bars are shown in Fig. 13 for the temporal height change for the Barnes Ice Cap, because in this case the absence of firn makes the glaciological situation more straightforward than for the other ice caps, and we can use the standard deviation and number of samples for each height change estimate to define a ‘standard error’. The error bars then reflect ± two standard errors, and as there is some variation in this

number the error bars are included in the graph. For the other ice caps we feel that this approach could lead to an underestimate of the potential error. Consequently, we estimate one number for the CS2 height change error for each ice cap and add this to Table 1. Adding constant bars to the upper panels in Figs. 7 and 11 would un-necessarily complicate the figures, and we felt the easiest way to document our errors was to include an 'all-inclusive' estimate for the two temporal resolutions for each ice cap in Table 1.

Does your error estimate reflect your assumption of larger uncertainty in winter than in summer?
Except for the Barnes ice cap results the answer is no. However, the error estimate for the high temporal resolution height change does include the potential contribution from the change in the bias between the surface and the CS2 detected elevation in winter.

The scatter in Fig. 5 is pretty large. Are there any assumption why this is the case? Could medium scale roughness within the CryoSat2 POCA footprint, like sastrugies cause such differences? Could you please add a figure showing the elevation difference versus surface gradient and roughness derived from the reference DEMs. Maybe this could give some more information. Another idea might be to plot the coherence extracted at the retracked bin versus elevation difference.
What you are suggesting is actually part of some on-going work. For those areas where there is good spatial and temporal overlap with surface GPS and/or airborne laser altimeter data we have been comparing the CS2 'height' minus the reference surface height with cross-track and along-track slopes, coherence, and some other parameters related to the waveforms. Your suggestion re roughness and sastrugi is very interesting and would be best approached by looking at the statistics of the ALS and ATM data, and then comparing these with the CS2 data. Again, we feel that this is beyond the scope of this paper but should be addressed in future work.

P2837L5 You mentioned ASIRAS data before and CRESIS Ku band airborne data in this section. Who provided ASIRAS data, who processed it - not mentioned in the acknowledgments? Did you use ASIRAS or only CRESIS in your analysis – not clear to me?
While the paper uses data only from the CReSIS airborne altimeter, our overall work has certainly involved study of ASIRAS data. In Canada we processed it from the raw files (from the aircraft) but the comment is certainly pertinent, and we should have made better specific acknowledgement to TUD (for the ALS and the ESA sponsored CryoVex flights), and to AWI for their efforts to process all the CryoVex ASIRAS data. This has been done.

P2838L14 It would be helpful to add the elevation profile as a subpanel as well as the different glacier facies (like percolation zone) along the profiles.
The primary purpose in including Fig. 8 was to show that the waveforms from the high resolution airborne Ku band altimeter flown over cold spring conditions on Devon varied year-to-year in a way that wasn't directly related to the elevation or surface facies. We doubt that adding an elevation profile for the NS and EW lines illustrated in Fig. 6 would add useful information.

P2839L17 Please add uncertainty of AWS.
The operational accuracy of the AWS acoustic distance ranger is ~ 0.03 m. This has been added to the text.

P2840L14 How does your result for Austfonna compare to McMillan 2014?
Where a comparison is possible the results are quantitatively consistent but the spatial and temporal resolution in our results appear to be better.

P2840L26 elevation change of 1 to 1.5 m
Fixed in 2 places.

P2840L28 you mention AWS temperature data. Could you also add this in Fig 11, as it is used to explain the April/May elevation change?
We think Fig. 11 becomes too busy and complicated with the additional temperature data, see additional comments below.

P2841L4 Why does Duvebreen AWS show an increase in elevation and CS2 a decrease between Sep2012 to March 2013? It seems that CS2 completely penetrates the fresh snow and still tracks the surface of the last summer which densifies and moves down. With the onset of the warm air the new surface is tracked by CS2. Could you comment a bit more on this? Are there density profiles of snow pit data available?
The Duvebreen AWS is on the north coast of Austfonna and does show an increase in surface elevation with the winter snow accumulation. The purple trace in the upper panel of Fig. 11 is the appropriate CS2 line for comparison. This also shows a more modest increase and yes, it is probable that the difference is due to an increasing bias between the surface and the CS2 detected elevation. The decrease in CS2 height for the low elevation southern data set (fawn trace in Fig. 11a) may be due to the strong subsurface component, compaction and the added two-way path component due to the permittivity of the snow. Note the south coast warm snow event referred to in the text is visible as the small blue spike towards the end of April in the CPDD (cumulative positive degree day) plot in Fig. 11b. Also, at the same time the Basin 3 height sensor (red trace in Fig. 11b) does pick up a height increase due to the snow fall.

P2841L28 Please add uncertainty (0.18 +/_ ?).
This sentence was worded poorly, and has been replaced by…
'The three-year surface height gain at elevations above 600 m was measured by finding pairs of surface GPS points within 1 m of one another where one point was from the spring of 2011 and the other from the spring of 2014 (blue points in Fig. 12). The mean height increase was 1.19 m with a standard deviation of 0.32 m. The CS2 height change was also estimated by using pairs of height estimates adjacent to the GPS transect with time windows Nov. to April in the winters of 2010/11 and 2013/14. In this case the CS2 height increase was 0.96 m with a standard deviation of 1.15 m. Although the approaches are very different this does provide added credibility that the 3-year CS2 height change illustrated in Fig 12 does bare a strong resemblance to the surface height change.'

P2842L25 Which error estimate is used for the other ice caps?
Text has been added to sections 5.1, 5.2 and 5.4 to justify the errors quoted in Table 1 for the other ice caps.

P2844L12 I think one should add a citation to Ricker (2015) and Nilsson (2015) who also explained the effect of "variable bias between the physical surface and the heights derived from CryoSat-2"
Nilsson, J., P. Vallelonga, S. B. Simonsen, L. S. Sørensen, R. Forsberg, D. Dahl-Jensen, M. Hirabayashi, K. Goto-Azuma, C. S. Hvidberg, H. A. Kjær, and K. Satow (2015), Greenland 2012 melt event effects on CryoSat-2 radar altimetry. Geophys. Res. Lett., 42, 3919–3926. doi: 10.1002/2015GL063296.
Ricker, R., S. Hendricks, D. K. Perovich, V. Helm, and R. Gerdes (2015), Impact of snow accumulation on CryoSat-2 range retrievals over Arctic sea ice: An observational approach with buoy data. Geophys. Res. Lett., 42, 4447–4455. doi: 10.1002/2015GL064081.
This would add some more weight to the argument that the so called "penetration bias" need to be considered when deriving elevation change in areas observing summer melt or freeboard change over sea ice. Would you expect that this kind of seasonal variability is important for longer time series as well?
Agreed, and both these references will be added to the paper.

Fig1 - Use larger characters
Done

Fig6 - AWS positions hardly visible. Use different colour.
The AWS positions have been made easier to see.

Where is AWS B located?
Fig7 – Labeling of y-axis is missing. Why are you not showing results of AWS 3 and 4
AWS 'B' is identified now in Fig. 6.
The labelling for the y axis in Fig. 7 has been added.
The results from the 2 higher elevation AWSs on the Devon ice cap don't really add relevant information to compare to the CS2 data. AWS 'A' is close to the average elevation of the blue CS2 and provides the best comparison.

Fig11 - Why does CS2 not show the summer melt in 2012 as the AWS are indicating? It is not explained in text. Please add temperature of the AWS - similar to Fig. 7.
There are dips in the CS2 'elevation' for all 3 areas centred at the end of July in 2012. However, as noted by the reviewer, the magnitude of the CS2 detected melt in 2012 is much less than that in 2011, but all 3 AWS show comparable summer height losses in 2011 and 2012. We are not certain why this is but have added the following tentative explanation in the text. … 'In comparing the CS2 and AWS summer height loss data it appears that CS2 indicates less melt in the summer of 2012 than in 2011, but the 3 AWS surface height sensors show comparable melt. It should be noted that the positions and elevations of the AWS sensors (Fig. 10; Etonbreen, E; elevation 369 m, Duvebreen, D; 304 m and Basin 3, B; 175 m) may not be truly representative of the CS2 data. In

particular, the average CS2 elevations for the low elevation north and south data sets (459 m and 380 m) are significantly higher than the relevant AWS elevations. Consequently, it is possible that the melt at higher elevations in 2012 was really less than that in 2011.'
We feel that the important surface temperature variation is better captured by the cumulative positive degree day data already present in the graph, and that the temperature variation for the three AWS positions would add little useful information.

Comment:
Maybe it is worth to scale all height changes to a height rate (m/month), because e.g. in Fig. 7 the black bars, which are hard to identify have different lengths indicating the time used for a monthly average is not constant.
We have added text to better explain the variation in the short time periods used for the high temporal resolution height variation plots (see also the response to reviewer 1). The monthly height change rate would be very noisy, and we feel that the height change plots, using the average time of each data set, is the clearest way to show the temporal evolution of CS2 'height'.