Overview: This paper integrates altimetry, conventional field and GRACE observations of glacier mass changes to develop a high resolution assessment of Greenland and Canadian high Arctic glacier changes. It builds on a previous method developed by the lead author that performs Monte-Carlo simulations aimed at extracting higher resolution from GRACE by incorporating glacier inventory information as a first-order constraint on the total flux of mass possible from any given region. The current manuscript incorporates additional independent information by including elevation change data from airborne and satellite altimetry. Given the extensive usage of GRACE within the community to assess glacier and ice sheet changes, and a variety of recent attempts to exploit GRACE data at higher spatial resolutions, this work is important and timely.

General comments:

1) Incorporation of altimetry data: the authors choose to incorporate satellite/altimetry-derived elevation changes as an additional independent observation to guide their iteration scheme. The use of a volume change measurement to constrain iterations of a mass balance solution raises several concerns, such as biases due to firn compaction, elevation-dependent biases in regions with highly variable hypsometries, and difficulty in sampling the few glaciers undergoing rapid tidewater retreat. The authors recognize many of these issues, but their overall argument is that the elevation changes are used to discern relative trends between nodes rather than to constrain the absolute magnitude of the mass variation. I question whether this is actually the case because the iterative update term (Eq 2) now contains both a factor for the surface area of glaciers in the node, and the average elevation change of those glaciers, the product of which is a very coarse estimate of mass balance, if one includes a density correction. Given all of the additional information the authors have complied for this effort, it is not clear why they did not take the additional step to calculate the mass balance within each node by combining these two pieces of information. Otherwise the justification they provide for only using average elevation changes (essentially just a single statement on page 543, lines 14-18) is not convincing and would require further justification. I would argue that taking these additional steps probably matters more for the Canadian Arctic than for Greenland. Over the Canadian Arctic there is generally more variability in regional hyposmetries and the issue of elevation-dependent sampling biases (and also possibly biases due to over- or undersampling of large dynamic losses) is of greater concern (see comment (2) below regarding ICESat processing for Canadian Arctic). Over Greenland I would assume elevation changes provide a better independent measure of the mass variations seen by GRACE, and that averages over 26 km nodes are more appropriate (at least for the interior parts of the ice sheet). In general I would like to see clearer statements to justify the use of elevation changes in this approach, and ways in which the data may need to be treated differently given the differences in the regions being studied.
2) Altimetry data processing: Additional details are needed (page 542) to better assess the processing and usage of the altimetry data products. It appears that the GC-offset corrections have not been applied to the ice sheet calculations of Schenk and Csatho (2012)? If so this should be justified, at least with a calculation of the expected magnitude of the offset relative to other terms in the error budget. Next, were all available airborne laser campaigns used in the

elevation change assessment, or only those within the GRACE measurement window? If the latter, provide information on the minimum start date of the altimetry observations (i.e. there will be many airborne campaigns that only partially overlap with the GRACE observations). Describe whether altimetry from various platforms (the high and low elevation ATM and the high elevation LVIS) were included or not.

Given that the focus of this paper is on high resolution mass balances assessment, the authors should justify not including the extensive airborne altimetry data from the Canadian Arctic. They should base their decision and subsequent evidence on results from both Gardner et al. (2011, Nature) and Gardner et al. (2012, TC 6). The 2012 reference includes an assessment of regional mass balance assessments derived from sparse ATM data. The 2011 reference illustrates the spatial sampling of ICESat and assesses its errors relative to other methods. Make it clear whether the additional spatial sampling provided by airborne laser in this area would fundamentally change your results.

The authors state that elevation changes for the Canadian Arctic are calculated following the method of Gardner et al. (2011,2012). In that approach, polynomial equations are fit to elevation changes, as a function of elevation over a region, and then applied to the regional hypsometry. Here the reader is left to assume that these last steps are not carried out because the focus is instead on calculating regionally-averaged elevation changes. The authors should be more explicit as to which elements of these previous approaches have been implemented in the present effort. Given that the full mass balance calculation has already been carried out for this region in the previous studies, it is not clear why that is not being utilized here.

On page 543, the authors take the 1-sigma standard deviation of the mean elevation changes within a grid cell as the uncertainty in a grid cell. This implies a correlation distance of one grid cell dimension (26 km), which is much larger than used in previous ICESat assessment, including the Gardner references listed above. The authors should reconcile their error calculation with the standards set out in these previous publications.

3) Terminology: the authors follow a standard of mass balance terminology they established in their earlier paper (Colgan et al, 2013). They distinguish between cryospheric mass loss per unit grid cell area from mass loss per unit ice area, calling the former a "mass balance per unit area" and the latter a "specific mass balance (per unit ice area)". This is confusing in light of the specifications laid out in Cogley et al (2011, mass balance glossary), where the term "specific" simply describes if a measurement is expressed per unit area or not. Given the history of usage in your previous paper, I do not suggest a major change, but would like to see a clear definition of terms, with reference made to equivalent standard terminology, laid out early in the manuscript.

Further terminology problems occur with the definition of the continuity equation (Eqn 7). The continuity equation should have dot_h on the left-hand side of the equation. It is a formulation of the mass balance equation applied to a column of ice in which density is assumed constant, and all changes in surface elevation reflect a change in mass (due to surface balance and/or flux convergence). Do the authors intend to equate dot_h to their dot_m? I assume this is not the case considering Equation (6) shows the formulation in which density variations are correctly

incorporated. Thus it seems your Equation (7) is Equation (6) for the case of dot_rho equals zero?
Further clarity would be attained if the authors stated, in the terminology of the equations they present, what it is that GRACE is actually measuring at each of their nodes? I assume it is all changes in mass associated with surface mass balance and ice flow in or out of each node.

4) Comparison to *in situ* observations: Comparisons between point observations of dot_m (which I understand to be ice equivalent changes in glacier elevation, according to your definition in Eqn 7) with GRACE-derived dot_m is problematic. Dot_m at a point on the glacier will depend strongly on elevation and the local velocity field, whereas GRACE sees an integrated balance that averages out much of this local variability. This is particularly problematic for smaller glaciers in the Canadian Arctic, raising again the issue that different analyses might be necessary for these vastly different systems. Consider the hypothetical case in the Canadian Arctic where all *in situ* data were clustered at a specific elevation range, thereby oversampling high rates of mass loss due to higher temperatures. If point data on glaciers and ice caps is going to be directly compared to GRACE, I recommend first integrating the mass changes to the full glacier extent. You can then assume that the integrated glacier balance is representative of that 26km region, and then proceed with the comparison to GRACE. The implications of clustered samples out in the middle of the ice sheet might have much less of an effect on this kind of comparison.

Specific Comments:

Page 538, lines 15-18: list a specific finding of the ice dynamic experiments, rather than this general statement.

Page 539, line 6: "…at THE ice sheet scale…"

Page 539, lines 16-19: what is the difference between "mass changes" and "absolute mass changes"? I recommend keeping things as simple as possible: altimetry measures volume change and GRACE measures mass change.

Page 539, line 24: "…as well as the fundamentalLY COARSE spatial…"

Page 540, line 1: An annual balance field for some time span is understood to be the average over that period, thus "mean" is somewhat redundant.

Page 540, line 2: What is being "combined"? This is unclear.

Page 540, line 8: "…derived from satellite AND AIRBORNE altimetry."

Page 540, lines 20-21: delete "object-oriented"

Page 541, line 27: "…when nearly all signal IS removed…". Make it more obvious to the reader that the Luthcke et al. (2013) iteration to minimize KBRR residuals accounting for mismodeled signal is completely separate from your iteration to correctly locate the mass variations in space.

Page 542, line 1: Remove “residual”. This creates confusion relative to the previous sentence referring to the KBRR residuals.

Page 548, line 1: remove comma

Page 549, line 16: “…with distance inland or WITH elevation”

Page 550, lines 6-7: The HIGA values agree to within error bounds, and are therefore not “significantly less than” the IMBIE values.

Page 550, line 12: You refer to mass losses in the Canadian Arctic as “peripheral”, a term which is usually reserved for the glaciers surrounding the Greenland ice sheet. Suggest deleting “peripheral”.

Page 550, line 25: Same comment here: don’t call Canadian Arctic “peripheral”.

Page 551, Section 4: I find the inclusion of significant analysis and results in the discussion to be structurally confusing. I would prefer to see this integrated into the methods/results sections of the manuscript.

Page 553: Change “geodetic-derived” to “geodetically-derived” throughout.

Page 553, description of geodetic method: Overall this section is unclear. What is an OSU cluster? The two references provided were difficult to locate for further details. If this is the first peer-reviewed presentation of these data, then that should be indicated more clearly. According to the Jezek report the measurements do include surface mass balance observations (stake data and snowpits), so this is strictly speaking not only a geodetic method. Your general expression of the mass balance of a column (Eq 5 and 6) would be better placed earlier in the manuscript to ensure consistent terminology and accounting of the various mass balance terms.

Page 554, lines 8-9: Is it correct that there are only 23 published point mass balances for the Arctic? There are long-term stake observation on Devon and Meighen ice caps, and on White glacier, that continue today. I assume these do not fit your definition of mass balance, but this should be clarified.

Page 554, lines 23-24: “which is consistent with the notion that the mass balance has generally decreased…”. This is not necessarily true. What if all the historical point mass balances were taken at low elevations?

Page 555, lines 27-28: The authors state that the MAR-derived surface mass balance field is not available over the Canadian Arctic. Note that this approach of subtracting the HIGA balance from the surface mass balance would not be appropriate anyhow for mountain glaciers, at the grid resolution you are using here. Related to this is that the Gardner et al. (2011) paper includes a modeled mass balance field. Have you considered summing this over the same 26 km fishnet for direct comparison to your HIGA values?