Response to Reviewer 2:
Interactive comment on "Greenland Ice Sheet seasonal and spatial mass variability from model simulations and GRACE (2003-2012)" by P. M. Alexander et al.
Anonymous Referee #2

Review of paper by Alexander et al.

General comment

The paper by Alexander et al. entitled "Greenland Ice Sheet seasonal and spatial mass variability from model simulations and GRACE (2003-2012)" employs surface mass balance from the regional climate model MAR v3.5.2 and the ice sheet model ISSM, to obtain a good representation of the overall Greenland Ice Sheet Mass Balance changes between 2003 and 2012. The authors process model outputs performing a spatio-temporal filtering in order to make a fair comparison with mass changes obtained from the GRACE data using the mascon strategy developed by Luthcke et al. (2013). They find a quite good agreement over the entire ice sheet and several sub-regions. Some discrepancies remain to date unexplained.

The paper is well written and illustrated. Succeeding in comparing GRACE and model outputs, the paper is a significant step towards the understanding of mass changes over the Greenland Ice Sheet but also at a more regional scale which is fundamental to understand the overall ice sheet system and its response to climate change.

We thank the reviewer for taking the time to review the paper and for comments that improved the quality of the manuscript.

Specific comments

p. 6359 – l. 26-27: "the Gaussian filtering procedure does not incorporate changes in mass". It looks to me that it is not entirely correct, or I may have misunderstood the filtering strategy. Indeed, the parameters sigma I, lambda I and sigma time seem to be determined using "the aggregated unfiltered MAR v2.0 data". However, leakage is usually a function of the mass change in mascon j and mascon j. So I wonder why the parameters mentioned above would not change using the aggregated unfiltered MAR v3.5.2 data given the differences mentioned in the manuscript between the two versions of the model. The entire GrIS MB may not significantly be affected because the differences between v2.0 and v3.5.2 of MAR may only result in the spatial distribution of the MB but not the total MB of the GrIS. However, for each mascon, MB might change from one model to another. Can you clarify this up?

The choice of parameters for Gaussian filtering is determined through comparison between aggregated unfiltered MAR v2.0 data, to which Gaussian filtering is applied, and the GRACE-filtered MAR v2.0 data. The Gaussian filtering is a means of

redistributing or smoothing the mass change from the unfiltered MAR v2.0 data in space and time so that the patterns of mass change from the Gaussian-filtered data match the GRACE-filtered patterns of mass change. It is therefore a way of approximating the effects of the GRACE resolution operator without a high degree of processing. The resolution operator is independent of the dataset to which it is applied, and therefore our derived parameters should also be independent of the dataset being filtered, since they depend not on the original mass change values from MAR v2.0 but on the relationship between the unfiltered and GRACE-filtered MAR v2.0 data. We should therefore be able to derive the same parameters using unfiltered and GRACE-filtered MAR v3.5.2 data. We have clarified this as follows:

"These differences do not affect our ability to filter MAR v3.5.2 outputs, as the Gaussian filter does not depend on mass changes, but approximates the GRACE-LM resolution operator, which serves to redistribute mass changes subject to specified constraint regions."

p. 6360, l. 6: " to a daily temporal resolution" is not quite correct, you are not increasing the temporal resolution, you are just increasing the number of time sample which leads me to the question, can you really give in the following sections of your manuscript seasonal timing at a daily/weekly (p. 6365, l. 21: "roughly 1 week earlier") accuracy given that GRACE-LM and GRACE-like filtered models have a time resolution of 10 days?

The resolution of the GRACE solution is roughly but not exactly 10 days. Therefore GRACE results are not provided on exactly the same day every year. For this reason it was necessary to interpolate the data to daily timesteps. The GRACE solution is therefore sampling the seasonal cycle on different days every year, reducing some of the error associated with the GRACE temporal resolution. The maximum error in timing associated with a peak is roughly 10 days, but this is nearly always smaller than the range of timing that we have estimated from the uncertainty in GRACE data. Given the relatively large uncertainty on GRACE outputs, we generally provide the median date or likely range of dates for the timing of GRACE peaks, and most of the differences discussed are larger than 1 month. We agree that we cannot say with certainty that there is a difference of 1 week between any two peaks. We have added an additional 10 days of uncertainty on each end of our error bars, have revised the text where necessary to avoid discussion of differences smaller than 10 days. The following text has been added to section 2.4.4:

"The temporal resolution of the GRACE-LM dataset can also lead to errors of roughly ±10 days for the timing of any estimate. Because the GRACE-LM timesteps are not regular, the uncertainty on the timing of peaks for the average seasonal cycle due to temporal resolution is generally smaller than 10 days. Given that the error could be as large as 10 days, however, we calculated our error on the timing of seasonal cycle peaks as the 95% confidence interval from the Monte Carlo simulations, ±10 days. If model peaks fell outside of this error range, the timing of the GRACE-LM and model peaks was deemed to differ. "

p. 6360, l. 8: "two-year composite seasonal cycle" why two years ? need justifications here.

See the response to the first general comment from Reviewer #1. A two-year wrapped seasonal cycle was constructed to calculate maximum and minimum values in the seasonal cycle. However, we now calculate the maximum and minimum values from the annual average seasonal cycle.

p. 6361, l. 2-8: As you say, "it is not possible to compare GRACE-derived mass changes directly to MAR" however you can compare MAR v3.5.2 with ISSM, so why not doing it earlier so that your assumption regarding ice discharge changes can be justified?

Because ISSM outputs are a model result, we do not feel that we can use the ISSM results as definite proof that variability in ice discharge is small relative to variability in SMB at this point in the manuscript, especially since the ISSM simulation does not include processes that may contribute to higher seasonal variability, such as the influence of meltwater on ice velocities and ice-ocean interactions. We agree that the model results do provide some support to this assumption, and have added this sentence to the discussion:

"(As will be seen in Section 3.2, modeled ice-sheet wide seasonal variability from ISSM is also less than 10% of variability from MAR.)"

p. 6368, l. 28: What type of GRACE errors might influence the variability in that context? Are, for instance, loading effects of surface mass changes of the GrIS and atmospheric changes taken into account when inverting for the KBRR data? Could it have an important impact on the signal, e.g. the amplitudes? Is the atmospheric model used in the GRACE-LM processing strategy the same as in the SMB model?

We have left open the possibility that some process unaccounted for might lead to the observed variability, but we do not believe that the processes taken into account in processing of GRACE data could have an impact on this variability, and their estimated uncertainty is taken into account in the calculation of uncertainty for each mascon. Loading effects of surface mass changes are not expected to have an impact on variability at the relatively short timescales examined here. Uncertainty associated with the atmospheric and ocean models used in the GRACE processing are estimated from differences between two different sets of models (as discussed by Luthcke et al., 2013). Given relatively sparse data for validation in this region, it is possible that model errors are larger than the inter-model differences, but we do not know what the magnitude of these errors might be. The atmospheric model used in GRACE-LM processing is the ECMWF reanalysis, which is the reanalysis used to drive MAR at the lateral boundaries. As ECMWF is used for estimating variations in atmospheric pressure, while MAR is used to estimate SMB, it seems unlikely that the use of ECMWF vs. another model could substantially impact the observed

differences. We have clarified in the text in section 3.4 that we do not know of any likely source for the differences, but leave open the possibility that some unknown factor could be responsible:

"Additionally, it is possible that although the GRACE-LM solution includes error estimates associated with the forward models used in GRACE processing, unaccounted for errors or processes, such as errors in model simulations used to correct for variability in atmospheric or ocean circulation (for which observations for validation are limited) may contribute to the differences. However, we cannot envision any obvious reason for the discrepancy other than the potential errors in ISSM or MAR v3.5.2 that have been noted."

Technical corrections
p. 6347
l. 4f: "While several studies…are still lacking." A short sentence to explain why it is important to examine sub-annual and sub-basin-wide changes would be nice here.

We have added a short sentence as suggested:
"At these scales, processes responsible for mass change are less well understood and modeled, and could potentially play an important role in future GrIS mass change."

p. 6349
l. 6-9: MB in l. 6 only defined in l. 9.

MB on l. 6 has been changed to "rate of GrIS mass change"

p.6350
l.4: you should precise the starting and ending months of the period of interest at least once but best everywhere the period appears.

We have added the starting and ending months when describing model simulations and GRACE outputs for the sections prior to the results, and for table and figure captions. In the results and conclusions sections we have left out the months for readability. Generally in all cases the period begins in January and ends in December, and we feel this is implied by the range of years. The only exception is the GRACE dataset, which begins in January 2003 and ends in June 2013. However, we have chosen to only consider data through 2012 in our analysis, as we are interested in capturing complete years for the purpose of evaluating the seasonal cycle. This has been clarified in Section 2.1. We have also revised Fig. 6 to include the only timeseries' through 2012, to avoid creating confusion, as all analysis is performed for the January 2003 – December 2012 period.

l. 15: "between GRACE-derived and simulated" -> would estimated be better than simulated here?

This sentence was confusing. We have replaced "simulated mass changes not accounted for by the simulations" with "modeled mass changes".

p. 6352:
l. 4-5 you should precise the starting and ending months of the two mentioned periods.
These have been added as suggested.

p. 6353
l. 19-23: Following these lines I expected some results of the evaluation and the comparison just mentioned. Instead it comes much later (p. 6364). May worth to modify a little bit the text here for instance by moving these lines to p. 6364.

As suggested, we have moved this material to Section 3.1.

l. 24: Is the period 2003-2013 correct? Isn't it 2012? Can you also precise the starting and ending months please?

Indeed, the period is January 2003 – December 2012. We have corrected the text accordingly.

Could be nice for clarity to summarize that to represent GrIS MB and given the ISSM forcing you have to combine ISSM with MAR v3.5.2 and not MAR v2.0

In response to reviewer #1, we have clarified in Section 2.2 that MAR v2.0 is only used for derivation of parameters used in filtering, while MAR v3.5.2 is used in the comparison with GRACE. We have also added this sentence near the end of Section 2.3:
"The cumulative mass change from MAR v3.5.2 and ISSM are then combined for comparison with GRACE."

p. 6354
l. 1: "at a high resolution" -> temporal resolution? Spatial resolution? Both?

"High resolution" has been changed to "the spatial and temporal resolution of the GRACE-LM solution"

Last sentence: I don't understand this sentence here. Isn't the job done by MAR and ISSM models? Or may be the sentence should be moved to the beginning of the paragraph as an introduction?

Since the GRACE solution is estimating ice sheet mass change, and hydrological changes over land areas are accounted for during GRACE processing, one might expect that GRACE processing would include removal of changes in snow cover over land. This is not the case, meaning that the GRACE solution also includes mass changes from tundra areas along the margins of the ice sheet. The sentence is

probably more appropriate in the section describing MAR data, so we have moved it to Section 2.2 and have modified it as follows:

"For comparison with GRACE, we include MAR SMB for the entire island of Greenland, including the GrIS, peripherial ice-covered areas, and tundra areas, as Greenland mass changes related to snow and ice cover outside of the ice sheet boundaries are not removed in the GRACE solution."

p. 6356
l. 2-4: Any reference for the 2000 m elevation limit between positive and negative MB?

The constraint regions are defined by Luthcke et al. (2013). The 2000 m boundary is a rough estimate of the boundary between positive and negative SMB (the equilibrium line altitude) and is not exact. A reference is added for Luthcke et al. (2013).

p. 6357
l. 1: Explain x and mu.

We have clarified the meaning of x and μ:
"…where x is the x-coordinate and μ is the mean of the distribution…"

l. 21-22: It looks to me that this sentence is a repetition of lines 16-18. Is there a physical meaning/interpretation for lambda i?

We have removed the statement on lines 16-18 at the suggestion of Reviewer #1. We have explained that the leakage parameter "represents the fraction of mass in mascon *j* outside the constraint region that influences the mass change in mascon *i*."

p. 6358
l. 12: explain t0.

Time t_0 is not a single point in time, but represents a point in in the timeseries for which temporal filtering is conducted. We have clarified that "t_0 is a point in time along the timeseries" and have noted that "The filtering was applied at each time t_0 to produce a timeseries of filtered cumulative mass change."

l. 21: "iteratively adjusted" using the least squares method? Also sigma l or l is not explained or present in formulas (same in p. 6359 l. 1). How was the iteration initiated and modifications performed?

"I" was a typographical error and σ_l has been changed to σ_i. The procedure was iterative in the sense that we repetitively applied filtering with different combinations of each parameter, but not in the sense that the parameters at a given

point were chosen based on the level of agreement for the previous set of parameters. We simply tried all combinations of the specified range of parameters, and subsequently found the combination of parameters that yielded the minimum RMSE. We have slightly changed and reorganized the text for clarity:

"We iteratively adjusted the values of σ_i, σ_{time} (in the case of temporal filtering), and λ_{ij}, for each mascon *i*. Values of σ_i were varied at 10 km increments over a range of 1 to 600 km, while values of σ_{time} ranged between 1 and 91 days at increments of 5 days, and λ_{ij} ranged between 0 and 1 at increments of 0.01. We tried all combinations of the three parameters over these ranges. The combination of parameters that yielded the minimum root mean squared error (RMSE) between the Gaussian-filtered and GRACE-LM-filtered cumulative mass timeseries were taken as the optimal set of parameters."

p. 6359
l. 1 what was the initial values of sigma i, sigma time and lambda i?

σ_I has been changed to σ_i. The range of parameters has been noted in this paragraph. The "initial" value does not matter in the identification of the best combination of parameters, as we simply tried all combinations over the specified set of ranges. The range of parameters was the same for the case where values of σ_i, σ_{time} and λ_i were the same for all mascons. We have modified the sentence for clarity:

"We also tried applying the same values of σ_i, σ_{time}, and λ_{ij} across all mascons *i* over the specified range of each parameter, but it was found that by spatially varying the values of these parameters the errors were reduced."

p. 6360
l. 5: You should precise the type of interpolation you use here.

We now note that we used linear interpolation.

l. 6: please explain how you remove the linear trends (least squares adjustment, moving window average…)

For clarity we have modified the text as follows:
"We then subtracted the long-term linear trend for the entire timeseries (2003-2012) obtained from least-squares regression, and averaged the data for a given day of the year across all available years."

p. 6352
l. 1: Would "Additional temporal Gaussian-filtering improves the agreement between the spatial Gaussian-filtered…" be better ?

Agreed. The text has been modified to read: "Adding temporal Gaussian-filtering…"

p. 6363
Don't you have uncertainties on your trend estimates?

We have calculated uncertainties on the trend estimates and have added these to the text, as well as a description in Section 2.4.4:

"We also compared modeled and GRACE-LM trends for the 2003-2012 period. To calculate the uncertainty in trends from GRACE-LM, we employed a similar procedure to estimate uncertainty in trends, conducting 10,000 Monte Carlo simulations and obtaining a distribution of trends and uncertainty values (from the 95% confidence interval for calculated each trend). The error on the trend was calculated as the average of the 2.5% and 97.5% deviations from the trend added to the 97.5% (upper) bound on the distribution of uncertainty values. For all model estimates, uncertainty on trends is reported as the 95% confidence interval obtained during linear regression."

l. 17: "2000-2012" isn't it "2003-2012"?

Yes, 2003-2012 is correct. The text has been changed.

p.6364
l. 1-3: Consistency -> -150-30 = -180 Gt/yr different to -179 Gt/yr in l. 14, p. 6363. Same for -242 + 3 = -239 Gt/yr w.r.t. -240 Gt/yr.

These errors have now been corrected.

l. 4-end: It looks like only models can be wrong, what about GRACE-LM ? Are you accurately taking all geophysical effects into account?

Errors associated with processing GRACE-LM data are accounted for in the error estimates for the GRACE-LM solution. The GRACE-LM solution produces trends comparable to those of other GRACE solutions, within the range of specified errors for the different solutions (Luthcke et al., 2013). We have added a mention of this in Section 2.1. We have added estimates of uncertainty on the trends and the differences between the filtered model results and GRACE exceeds the uncertainty on the difference. We have noted this in Section 3.1.

p. 6365
l. 7-8. Can you suggest examples for independent evaluations of the models?

We have added some suggestions as follows:

"In these areas, high spatial variability of topography can strongly influence SMB. To properly identify the source of the differences, further independent evaluations of

MAR SMB and ISSM DMB are needed, for example, comparison between ISSM and remote-sensing derived discharge estimates (e.g. Rignot and Kanagaratnam, 2006), or comparison between MAR and additional in situ measurements in ablation areas."

p. 6368
l. 1: "the maximum modeled mass occurs" -> "the maximum modelled mass CHANGE occurs"?

For clarity we have changed "the maximum modeled mass" to "the maximum in the cycle of cumulative mass change".

p. 6369
l. 15: Can you give examples of non-ice-sheet-related processes that may contribute to the discrepancy?

See response to the last point in the general comments. We have added the following here:

"Additionally, it is possible that although the GRACE-LM solution includes error estimates associated with the forward models used in GRACE processing, unaccounted for errors, such as errors in model simulations of atmospheric or ocean circulation (for which observations for validation are limited) may contribute to the differences. However, we cannot envision any obvious reason for the discrepancy other than the potential errors in ISSM or MAR v3.5.2 that have been noted."

l. 17: I would change "Discussion and conclusions" to "Concluding remarks".

Changed as suggested.

l. 23-25: "We also find that … this effect." What about something like "We also applied a temporal Gaussian filter to the models to reproduce the attenuation inherent to the GRACE-LM processing strategy."?

The sentence has been changed to:

"We have therefore also applied a temporal Gaussian filter to the model outputs to reproduce the attenuation inherent to GRACE-LM processing."

p. 6370
l. 13-14: How many in situ stations? How many SMB measurements are available?

There are eight stations along the Kangerlussuaq Transect in west Greenland (compared with MAR in Colgan et. al. 2015), there are also a handful of stations from the PROMICE network with stakes in areas of net ablation on the ice sheet (www.promice.org). The sentence has been modified as follows:

"A comparison at eight in situ stations at the Kangerlussuaq ransect on the southwest GrIS suggests that MAR v3.5.2 SMB is closer to in situ measurements (Colgan et al., 2015), but such measurements are limited to this transect with the exception of a comparable number of ablation stake locations from the Programme for Monitoring of the Greenland Ice Sheet (PROMICE; www.promice.org)."

l. 19: Any suggestion for conducting independent evaluation of each model?

We have added some suggestions as follows:
"The only means of determining the relative contribution of ISSM and MAR v3.5.2 to underestimated mass loss would be to conduct an independent evaluation of each model against DMB (e.g. using the methods of Rignot and Kanagaratnam, 2006) and SMB estimates over large portions of the GrIS. These analyses are beyond the scope of this study, and the evaluation of SMB is limited by sparsely available data, although radar measurements of snow accumulation may help to fill this gap. A preliminary comparison by Koenig et al. (2015) suggests that there is a good agreement between MAR v3.5.2 and radar-derived accumulation estimates over the interior of the GrIS during the years 2009-2012, but that MAR tends to overestimate accumulation along the southeastern coast."

l. 27: "represented" -> observed? "results -> predictions?

"as represented by" has been changed to "from". "results" has been changed to "outputs". We feel that "predictions" implies a representation of future behavior, which is not the case for these model outputs.

p. 6372
l. 8: "and may transition between"?

The sentence has been modified for clarity:
"Different glaciers exhibit different patterns of flow variability, and a single glacier may exhibit different patterns of flow in different years."

fig. 3: RMSD -> RMSE?

RMSD has been changed to RMSE as suggested.

fig. 5 and where it appears: errors bars are not that easy to identify.

We are not sure exactly which error bars the reviewer was referring to. We did not have vertical error bars on the 95% range of the distribution; these have now been added.

fig. 6: can't see the pink shading surrounding the GRACE time series. 2003-2013? Be consistent with the manuscript where it is mentioned 2003-2012.

Fig. 6 has be changed so that the timeseries extends from 2003-2012 for consistency with most of the text. "2012" was changed to "2013" in the caption. The error is small relative to the trend, and pink shading was obscured by the timeseries line. The width of the line has been reduced to allow the pink shading to be more visible.

fig. 9: 2003-2013 or 2003-2012? In b), missing an x in maximum. It would be easier for comparison purpose to have the same colour scale in all figures.

"2013" has been changed to "2012". The typo in "maximum" has been corrected. Most of the time, the dates for the maximum and minimum peaks do not overlap. We have adjusted the color scales so that the same scale is used for images of maximum and minimum peaks.

fig. 10: "and blue colours indicate AND earlier date" -> typo?

Yes, this was a typo. "and" has been changed to "an"

fig. 12b: Could satellite altimetry data help assessing the difference between models and GRACE data?

Previous studies have focused on annual trends rather than seasonal trends, as the temporal resolution of altimetry data are limited. The historical altimetry measurements available for the period used in our study will likely not help to assess the differences shown in Figure 12. However, newly launched sensors, such as Cryosat-2 and Sentinel3 might provide higher temporal resolution and we appreciate the suggestion by the reviewer in this regard. Still, translating elevation changes into mass changes is still strongly dependent on the firn compaction models adopted and a more detailed, long-term study is required in this regard.