In this manuscript, Ashokkumar and Harig present a new estimate of recent mass loss (2002-2017) based on GRACE gravimetric observations. These observations are then used to calibrate the parameters of a glacier evolution model that relies on volume-area scaling and for which the surface mass balance (SMB) is calculated from a degree-day approach. The authors compare the so-modelled mass losses with estimates from other research groups (Hock et al., 2019). For the Gulf of Alaska and for Arctic Canada North, they find mass losses that are in line with other estimates from the literature. For Arctic Canada South, their newly estimated mass loss is substantially higher than previous estimates from the literature.

The new gravimetric data presented in this paper seems sound and will likely be of interest to the glaciological community. However, the manuscript mostly focuses on using these data to calibrate a glacier evolution model and to simulate the future evolution of glaciers. There are some rather fundamental problems with this main part of the manuscript:

1. Limited novelty:
   a. The model relies on volume area scaling to simulate the future evolution of glaciers. V-A scaling was a widely used technique in regional glacier evolution studies at a time when glacier thickness distribution was largely unknown and when computational constraints did not allow for more elaborate approaches. Over recent years, several methods have been developed to estimate the ice thickness distribution over a large sample of glaciers (see e.g. Huss & Farinotti, 2012; Farinotti et al., 2019). In combination with a better characterization of glaciers and increasing computational resources, more sophisticated methods (vs. V-A scaling) have been developed and successfully used to simulate a large ensemble of glaciers. This includes methods in which the glacier geometry is explicitly accounted for and on which changes are imposed based on parameterizations relying on observations (e.g. Huss & Hock, 2015; Rounce et al., 2020a) and methods in which the ice dynamics are explicitly included to simulate the future evolution of glaciers (e.g. Clarke et al., 2015; Maussion et al., 2019; Zekollari et al., 2019). Some of these methods have already been applied at a global scale (Huss & Hock, 2015; Hock et al., 2019; Maussion et al., 2019; Marzeion et al., 2020). In this respect, using a V-A scaling approach at a regional scale is far from being novel and may even be considered to be a bit outdated...
   b. Also the climatic conditions used in this study are slightly outdated. The authors rely on ERA-Interim data, while now a more sophisticated and higher-resolution product is available: ERA5 (see https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5).
   Also for the future simulations, it is a pity that the authors relied on CMIP5 GCM output and not on CMIP6 GCM output. In fact, given the regional focus
of their study, the most logical option would have been to rely on regional climate model (RCM) output from the ‘North American CORDEX Program’ (https://na-cordex.org/)

c. The authors compare their results to the output of the GlacierMIP (Hock et al., 2019). The problem with GlacierMIP is the fact that the simulations are difficult to compare given the large difference occurring not only in terms of glacier model, but also in terms of boundary conditions (different glacier volume, different forcing,...etc.). This has now in part been solved in GlacierMIP2 (Marzeion et al., 2020), which is the first coordinated experiment for glacier evolution modelling. I am well aware that this is a brand-new study and that the authors could not have been aware of this. It is therefore not really a point of critique, but I think that it would nevertheless be good if the comparison could be made with these new results.

d. Calibrating model parameters with regional values is tricky in general (reasoning behind this is elaborated in bullet point 2). The best method to calibrate a large-scale glacier evolution model is to perform a calibration of model parameters at the glacier level by reproducing glacier-specific observations. Glacier specific mass change estimates are now becoming widely available (e.g. Brun et al., 2017; Braun et al., 2019; Dussaillant et al., 2019; Shean et al., 2020), and have been used to calibrate regional glacier evolution models (e.g. Zekollari et al., 2019; Rounce et al., 2020a, 2020b). In this sense, relying on regional estimates of glacier mass change, even if these are probably more accurate than previous regional estimates (which seems to be the case with the new GRACE data you present), is not ideal (see bullet point 3).

2. Calibrating parameters at regional scale is always quite tricky. It was done until recently, but now the priority is to transition to glacier-specific observations when calibrating models. The main reasons for this are that (i) individual glaciers within a region are subject to different mass changes and that (ii) no single combination of model parameters (to be applied for all glaciers within a region) can accurately describe this. In general, there are two options when working with regional mass changes for model calibration:

a. One can assume that all glaciers in a given region are subject to the same mass balance and then perform a calibration of the model parameters at the glacier scale (i.e. different parameters for every glacier). As you explain in your manuscript, this is not ideal, as in reality large glaciers tend to have a more negative mass balance (more out of balance compared to climatic conditions), compared to smaller glaciers. This is the method used by Huss and Hock (2015) and it has the advantage that you can match the mass changes for every glacier. But as you match a mass balance that is ‘off’ at the individual glacier level, you tend to underestimate future volume losses (as the mass balance of the large glaciers, which make up for most of the volume, is positively biased when assuming that all glaciers have the same mass balance) (e.g. Zekollari et al., 2019).

b. The second option is to match the regional mass changes by using a single set of parameters for all glaciers within a region. This is the option you opt for by matching the new GRACE derived mass changes with your model. For this, you need to rely on the same model parameters for all glaciers, and in an ideal case
you would be able to match regional loss (calibration) and the mass balance at the individual glacier level (evaluation). However, Huss and Hock (2015) have shown that this does not work well, as by working with the same parameters for every glacier you can get very strange mass balances at the individual glacier level: i.e. the sum of the mass balances match the regional mass balance, but the reason for this is wrong. Through this, it is likely that you have very strange mass balances at the individual glacier level: e.g. some glaciers may even have a very positive mass balance for the observational period. When you then project these in the future, these glaciers may even be growing... It would be good to have an overview of how the mass balance looks like at the individual glacier level:

- Comparison to direct mass balance observations. You perform this, but while doing you still modify some parameters (which you can’t do!). Even then, the match is not very good so far.
- It would be good to describe and show the mass balances that you obtain for all glaciers (e.g. through box plot). I hope this is not too bad, but I am afraid that some glaciers may have a very strange mass balance.

3. So what is the role of GRACE? You present it like a major advantage of that you match the GRACE derived regional mass change. OK, this estimate is probably better than relying on a very rough approach in which mass balance observations from a few glaciers are used to derive a regional mass balance, but in the end this is just a number for an entire region, which is a strong limitation for regional glacier modelling. The difference between relying on these GRACE observations and rougher regional estimates is relatively small in the end. If a regional mass change estimate is used, I would advise you to use another approach for the calibration (bullet point 2a). And when doing so, make sure that you match the GRACE observations (see next bullet point)

4. You do not match the GRACE derived mass balance!
   a. While allowing to have a wide range of values for your parameters, you are not able to reproduce the mass balance derived from glaciers. This is quite concerning. This may be solved by changing the calibration setup (bullet point 2a).
   b. By allowing the model parameters to take any value, one can match the GRACE derived mass balance (at the regional scale). If this would have been done, the projected future glacier changes would likely have been very close to the values from the literature. In the result you present:
      - Gulf of Alaska: is the region with the best match between GRACE and modelled MB → modelled future glacier evolution relatively close to estimates from literature
      - For Arctic Canada North: the modelled MB is more negative than GRACE observations → modelled future glacier loss is higher than the estimates from the literature
      - For Arctic Canada South: modelled MB is far more negative than in GRACE observations → modelled future glacier loss is much higher than the estimates from the literature
By matching the observed MB (from GRACE), your projected mass changes will be close to the other projections from the literature. This is not surprising, as the GRACE derived mass balance is not that different from previous estimates based on the extrapolation of field measurements and/or remote observations.

I think it is a brave effort of the authors to ‘jump’ into a research field (glacier evolution modelling) that is new from them. It is such efforts and new impulses that will help a research field – in this case the field of glacier evolution modelling - forward. When doing so, one must first gain a good overview of this field in order to avoid making some basic conceptual mistakes and to ensure that the work is an added value. I think the authors should have focused on this, rather than pointing at wrong reasons for explaining discrepancies between their results and results from the literature. A correction based on comments above will help tackling some problems (rethink the calibration, make sure to match the GRACE observations, use state-of-the-art climate output), but even then, the novelty of results put forward will be very questionable given the model architecture. In GlacierMIP and GlacierMIP2 models of the same complexity and more sophisticated models have been used at larger regional scales.

References


Maussion, F., Butenko, A., Champollion, N., Dusch, M., Eis, J., Fourteau, K., et al. (2019). The


