Review Malles and Marzeion 20th Century global glacier mass change…

This paper is about a method to find out which data set is most appropriate to calculate the volume change of glaciers and to provide a realistic uncertainty estimate on this value. As such the title is misleading. The title suggests that this is going to be the best estimate of the glacier volume loss. At the same time this is probably true but the paper needs to be reframed in that case which I would favor. Part of the methods can go to an appendix to free up some space to discuss results more extensively in context. There is very much emphasis on the results and limited context on other estimates in the literature aiming the same. The abstract for instance has only one sentence (the last) with results and is nearly completely about the methodology.

Major points

IF the paper is going to focus more on the result rather than on the method, one needs to address the fact that this paper is based on reconstructions of the climate based on data sets which are partly observational driven whereas projections are based on model results and the question is whether they can be connected. For this reason it would be interesting to use the CMIP ensemble over the historical period as an additional data set to see what that brings applied to the proposed method here.

I would be interested to see which fraction of the volume loss is related to temperature and which to precipitation. You could easily calculate that with your calibrated model by switching off the variation of one of the two.

I cannot understand the assumption on line 90 where you argue that mu_star and beta_star follow from the assumption that the present-day geometry is a steady state. We all know that the majority of the glaciers is not in a steady state? So even if you find that t_start is not present-day this seems a flawed approach. In addition I can buy the argument that t_star is not only a function of regional climatological history, but then you conclude based on that argument that it is allowed to take a global value for t_star??. I would conclude based on the same observation that t_star is not a global value, but requires to be estimated for each individual glacier and in fact you seem to do that around line 145. So I have serious problems with the concept outlined in lines 90-95. Following up you argue that u_star for each glacier follows from a global t_star (line 98) why?? and that b_star maybe spatially interpolated why?? Why not the other way around? or both spatially interpolated or based on t_star. I feel lost on the model description.

You use different data sets for temperature, but for precipitation you only use the anomalies of the different data sets based on the holy CRU CL 2.0 data set. Thereby you ignore all the uncertainty in the CRU data and consider it as the holy truth that doesn't seem to be correct to me. Furthermore you don't stress this at all.

Line 380. Table 4 is biased towards a comparison over the recent period, there is a variety of estimates for the entire century it would be interesting to compare you results also with those.

Minor points

Line 12: A little unclear what is meant with the different ensemble members. I presume you imply to say that most variation is caused by the input climatological rather than by the uncertainty in the parameters to constrain each model. Please rephrase to clarify.

Line 16: I don't understand the difference between total uncertainty and reconstruction uncertainty. Is ensemble uncertainty + reconstruction uncertainty the total uncertainty? Please rephrase to clarify. I am also unsure what the difference is between ensemble spread and ensemble uncertainty, both are used in the abstract or are they the same?

Line 19: The total uncertainty yields an uncertainty of only 8%, that is extremely small in my view and I doubt whether that is in line with budget closure studies of sea level over the last century.

Line 22: The paper of Slangen et al 2017 is on attribution that does not seem to be a very logical paper to refer to. Moreover the idea that glacier are important is much older. So an older reference seems more appropriate.

Line 28: densities can be left out.

Line 31: It is not trivial that an ensemble based data reconstruction adds to the uncertainty in modelling the future. I think that is only true for an ensemble based on models for the historical period but that is not what you do. You have to make this separation clearer. Though I would prefer you take such an ensemble CMIP5 or 6 onboard in your reconstruction.

Line 36: The word additionally suggest you take changes in the geometry and hypsometry on board, but do you really do that? And how? Having a response time scale is something different. Explicitly state that the time scale mimics only part of the effect caused by the dynamics, the height and length changes are not captured by a time scale, but normally embedded in the dynamical adjustment.

Line 40: strange sentence whether you resolve the energy balance or the dynamics are different entities. So the fact that an energy balance does not include the dynamics is not an argument to dismiss an energy balance approach.

Line 44: In the context of the paper it is very odd to argue that the additional degrees of freedom of an energy balance models are a limitation. You are aim is to get to an adequate estimate of the uncertainty, so the uncertainty space can much better be explored in an energy balance model than with a temperature-index approach.

Line 46: Computational limitations are not a constraint. The constraint is the lack of data on the geometry and they probably add little due to the large uncertainty in the forcing.

Line 51-55. I would expect here a summary of the outcome of Marzeion 2020 specifying how large model uncertainty is versus forcing uncertainty as discussed in Marzeion 2020.

Line 75. You can not assume the reader to know the Marzeion et al. 2012 paper, please explain in a few sentences the concept or explicitly refer to the next section. Then wrap up in line 84 a bit more the concept of the Marzeion 2012 model.

Line 96. Unclear what you mean with leave-one-glacier-out cross validation. Explain in more detail.

Line 110. Do you include all the uncertainties in parameters around precipitation. What does " with T_i__zmax(t) " mean. I think you can leave that out the fact that you use (t) already indicates that this parameter is time-dependent.

Line 110: Unclear what to do with line 4 of the equation adding it or multiplying it with line 3? Syntax ambiguous.

Line 125: Why can you not find an initial volume which leads to the RGI volume at the end of your run. It is a pretty linear system, so I would expect that there is always an initial volume leading to the RGI volume. Please explain why this is apparently not the case.

Line 135. Can you explain why there should be a condition of B(t_hat)=0. Many glaciers are not in steady state for the given geometry so why would this condition hold for at least one value of t_hat.

Line 155. Figure 1 is not clear I don't understand the difference between the direct and spatially interpolated method.

Line 190. Mention for which time interval the RGI data are valid.

Line 220. MSE is not minimized if sigma_M=sigma_O but in case R=0 then they are independent.

Line 222. I don't buy the argument that this is a complex model. Please provides stronger arguments.

Line 251 margin is not the right English word.

Line 266 twice than

Line 314 part D of *Figure* 3

Line 333 and for and?

Line 340. For completeness you have to mention that you assume e_ensemble to be independent of sigma_t such that you can pool them. This is not necessarily true and maybe one of the reasons why you end up with a total uncertainty of only 8%. You need to discuss the later in this context.

Line 358 specify the most recent period 2015-2018?

Line 370 and 371 consider leaving out “is”

Line 394 What is the explanation for the large difference with Zemp et al. 2019.

Line 399. You can easily be more specific as you have the Parkes and Marzeion numbers.

Line 410. I guess you need to take the marine terminating on – board in a final estimate in order to prevent the suggestion that glacier mass loss is extremely accurately known as you suggest with the 0.05 mm/yr value.

Line 426. It is worth bringing this to the abstract.