

Interactive comment on “Bayesian calibration of firn densification models” by Vincent Verjans et al.

Anonymous Referee #2

Received and published: 21 March 2020

The paper aims to evaluate and improve the parametrisations of 3 firn densification models, for the dry snow part. It is written in a concise and clear matter, but I find that the scope of the paper is too narrow. The authors use a sound method, and find weakly different results from the original publications for 2 of the 3 models. They make a good effort in discussing the implications of their findings in terms of densification physics, but still, to my mind, they consider only part of the problem, and it is difficult to make use of their findings without the rest of the picture.

For instance, in the conclusion, they say "As such, although model results can be improved by re-calibration methods, model tuning alone is insufficient to reach exact fidelity of firn densification models. The formulation of models' governing equations impacts the remaining errors with respect to observations, which highlights deficiencies in our understanding of dry firn densification." The problem is that they don't evaluate all parts of their models.

C1

I consider that the parts of the (inverse) problem are : 1. Input data. 1.1 Away from meteorological stations, surface temperature and accumulation are not well known. Even if RACMO is as good as it gets, there are biases that can be several °C in temperature. Simply assuming that RACMO is right is not acceptable. A range of scenarios, based on the known biases of RACMO, or uncertainty derived from also using MAR, and different reanalyses is necessary. In many cases, you can find a mean accumulation derived from the ice core the density was measured on, why not use that? 1.2 surface density. I support the idea to use measured density, but measuring surface density is not easy. It comes with an uncertainty of 20% at least. You need to do a sensitivity study of your whole process with randomly different surface density within a reasonable uncertainty range. Surface density is actually a pretty sensitive parameter.

2. Assumptions that go into the model, like the steady state assumption, and the 1D assumption. In some areas, we know these assumptions are not right. In your outliers to the models, when it is an outlier to all models, did you consider that maybe some of these assumptions were not right? Mizuo is for instance a very strange site. Places with snow dunes in Antarctica can do funny things too..

3. the physical formulation : you discussed this thing well, but did not combine all your model output to give the global uncertainty, or compare the within-model to the across-models uncertainty. For instance, it would be interesting to see a fourth row on figure 5 with the outputs of the 3 models on the same figure, and comment why the uncertainty ranges don't overlap. I am surprised by that, and it makes me think that your 95% confidence interval is underestimated.

4. The parametrisation of each model. You focused on this. My major problem with the method you used is that you assumed that your parameters were independent, when in reality they are not at all, as shown the covariance matrix on Fig S4 (e.g. k1 and E1 having a covariance of 0.94). You don't show your prior error covariance matrix, so it's difficult to assess exactly what you did. A better description in the prior assumptions, including a comparison of the prior and posterior probability distributions (for instance

C2

adding the prior to Fig 3) is needed. From what I read, you included no covariance in your parameters in your prior, and that is not right. For instance, in fig 3, showing k_1^* and E_1 independently makes no sense, because you could always compensate any error in E_1 by a change in k_1^* , it would make more sense to show these either in a 2D plot, or for a fixed E_1 in the case of k_1 and vice versa.

5. The metric used to compare models to data, and the quality of the data. 5.1 Data. Here, I'm with you, I don't want to go through checking each dataset again, but you should at least give some information about these data, citing the spencer paper, and maybe a few other, to illustrate what is the uncertainty in the data, given different methods (weighting a full core is not quite the same as doing gamma, or CT..). 5.2 Metric. Is DIP really the best metric? Is it faithful? You are the first to use this metric for the calibration of firn air models. Everyone else was using $\rho(z)$ directly. Does it give the same answer? You should demonstrate that. Is it a good metric also for other applications of firn modelign, such as the close-off depth estimation? A paragraph demonstrating the usefulness and validity of this metric would be nice.

The general objective of the paper is, if I quote the abstract to "demonstrate how model- and parameter-related uncertainties potentially affect ice sheet mass balance assessments". It's a great idea, and such a thorough assessment of firn models has not been done. Lundin et al. 2017 highlighted some model deficiencies, but did nothing to remedy those. I support a future version of this paper to be published with a quantitative answer to this question, but we are not there yet.

After you have properly addressed all the parts of the problem stated above, I'd like to see a quantitative comparison of the different sources of uncertainty. If I dont want to run the community firn model (CFM) 30,000 times but just a few, to get a gist of the uncertainty in my specific application, should I use a wide range of (T, accum) scenarios (step 1)? or rather use different model physics (step 3)? Or one of the models, but with a range of parameters (step 4)? This is a practical question that would be really useful to future users of the community firn model. And finally, when

C3

you have done that, it would be great to go back and recalculate the uncertainty in mass balance from altimetry, using the above mentionned decomposition. This is the great paper I'd like to read. It's marginally more work from what you have done, running a few more simulations on the same framework, but I think it would be really worth it.

I would like to finish with more detailed comments: 1. Method: Why did you go for Monte-Carlo, rather than a form of generalized least squares, which would have converged in <10 runs very likely, and could have easily dealt with covariance in model parameters? I agree that this problem is non linear, and underdetermined, but it is also monotonic, so least squares are applicable, and converge much faster. That being said, your method is valid.

2. choice of models. Why did you choose the Arthern model as one of your 2 models, rather than the IMAU versions (ligtenberg or kuipers)? We already know that the physical formulation of Arthern is not right (Lundin 2017). You later discuss the IMAU version. I think it would make more sense to publish an optimisation for these rather than the Arthern, which shows no sensitivity to the accumulation rate, something we know is wrong.

Interactive comment on The Cryosphere Discuss., <https://doi.org/10.5194/tc-2019-274>, 2020.

C4