Review of « *Spatial probabilistic calibration of a high-resolution Amundsen Sea Embayment ice-sheet model with satellite altimeter data* » by Wernecke *et al.*

I was not a reviewer of the previous versions of the manuscript.

This paper presents a dimension-reduced approach to calibrate model projections against observations of surface elevation rates of change. Following the comments of Reviewer 1, the new version includes new benchmark simulations to show that the method allows to recover correct known parameters values. The method is clearly described and the applications convincing, providing a valuable contribution to the field.

However, I still have few major comments that the authors should consider before publication.

Major comments :

- Following comments from Reviewer 2, the distinction between the simulations presented in the paper (using constant forcing) and « projections » is still unclear. The word « projections » is still used in several places to describe the simulations and this really needs to be clarified. I suggest to avoid the term « projection » in the abstract, discussion and conclusion. I suggest to split the section "**2.1 Ice sheet model ensemble"** in two subsections, the first to describe the model initialisation and the set of perturbed parameters, the second to describe the transient simulations, with the spin-up, calibration and forecasts periods. This would be the good place to discuss the assumptions in the forecast period and why the results differ from « projections ».
- It is not exactly clear which observations are used and what is their equivalent in the model. The dataset is a compilation of surface elevation changes from 1992 to 2015, but we understand that only observations from a 7 years period are used. Which period? How is it chosen? Does the initialisation of the model correspond to a given date? What exactly are the model outputs that are used for the comparison with the observation, i.e. the mean surface elevation change during the 7-year calibration period, the surface elevation change ate the end of the calibration period or the average of the annual (or bi-annual) surface elevation changes during the calibration period? By the way, the term "surface elevation changes" is used for the observation, but "thickness change" is use for the model. It seems that only observations on the grounded part are used so that "surface elevation changes" should correspond to "thickness changes", but better discuss this point and check that it is consistent throughout the manuscript.
- Finally, I encourage the authors to discuss with more details the benefits of using the surface elevation changes with their experimental design for the calibration. The model is first calibrated using spatial observations of surface velocities to tune the basal friction and viscosity. This point should be made more clear for readers that are not familiar with the initialisation of ice sheet models, i.e. in a sub-section to describe the model initialisation as suggested above. As the ensemble design implies multiplicative perturbations of these inverted fields, the best fit is obtained with the default values (0.5) and all other combinations should degrade the fit to the observed velocities. The fact that the calibration recovers values that are close to the default

means that it is the configuration of the model that best fit the velocities that give the best fit to the surface elevation changes. Any other result would mean that the model is not able to reproduce both the velocities and elevation changes, and indicate a problem in the model or in the ensemble design. So a question is how much additional informations do we get from using the surface elevation change field as an additional observation for the model calibration/initialisation? I think it would be difficult to answer this question in a quantitative way, but it would be interesting, in the discussion section, to group and improve the parts discussing the limitation of the experimental design (l 5-10, p18), with the discussion on what we can expect from including the temporal component (l 21-23, p18), even if the calibration period (7 years) seems too short to discriminate the friction law exponent and basal melting scaling.

Minor comments :

- Everywhere ; better to use « *friction law* » instead of « *sliding law* ».
- Abstract, line 11 : *« while a net sea level contribution calibration imposes only weaker constraints* ». Maybe not very clear, suggestion *« while calibration against an aggregated observation, as the net sea level contribution, imposes only weaker constraints* ».
- Page 2 , line 8 : *« basal melting is expected to continue for the next few years to decades* », not sur what do you mean, maybe « *high rates of basal melting* » ?
- Page 5, line 4 : « *and use the following 7 years as calibration period* » ; see major comment above, explain how the model results are used.
- Page 5, line 4-5 : « *Other calibration periods have been tested and show small impact on the results for calibrations in basis representation* », give more details for the meaning of « *others* » : longer, shorter, different spin-up duration, etc… ?
- Page 5, line 5 : « *We regrid the simulated surface elevation fields* » ; Please clarify ; is it surface elevation or surface elevation rates of change ?
- Sec. 2.2 Observations : please provide more informations on the data that are used. Dates ?
- Page 6, line 8-9 : « *The first k columns of U) are illustrated in Figure 1 which are related to the PCs (Bi) by multiplication with the singular values*. » ; Please check this sentence and how it relates with Eq. (2). It is said in lines 4-5 that the principal components (PCs) are the first columns of B, and caption of Fig.1 says that it shows the 5 PCs .
- Page 6, line 12-13 : *« This decomposition reduces the dimensions from m grid cells to just k principal components.* ». B' and V' still have *m* lines corresponding to the grid cells but *k* columns, so the dimension reduction is from the *n* ensemble members to the first *k<<n* PCs ?
- Figure 2, caption : « *Mean observed ice thickness change* ». Date ?
- Page 7, line 5 : « *observations over a seven year period* » ; see above ; provide details in Sec. 2.2.
- Page 7, line 10-11 : « *The spatial variance of the difference between the reprojected and original fields is substantially smaller than from $z_{(xy)}$ alone*: ». What are the implications ?
- Page 8 : Define that N represent the normal distribution.

- Page 12, line 12 : « *fast simulations* » ; Needs reformulation : « *simulations with high velocities* » ?
- Page 12, lines 21-26 : I think this is hardly understandable for a non specialist of ice flow modelling, especially the part « as C compensate for v ». It is maybe better to move Eq. 13 in Sec. 2.1 and give more details there on how the frictions coefficients *C* are tuned with respect to the observed velocities. Expressed in a simple way, the explanation is that the model has been tuned to give the same initial state, however as the friction laws have a different non-linearity, differences will only become apparent in areas where changes in velocity or stresses are significant. The authors might also be interested by the study from Brondex *et al., Sensitivity of centennial mass loss projections of the Amundsen basin to the friction law*, Cryosphere, 2019.
- Page 12, line 29 : *« From this test we conclude that basal sliding law and ocean melt scaling cannot be inferred from this calibration approach* ». As explained, the problem seems not to be the calibration approach itself but the fact that the changes have not been sufficiently large during the calibration period to distinguish between different sliding laws and different melt scaling.
- Page 16, line 15 : « *However, no satellite observations have been used for the bedrock modification, nor has there been a quantitative probabilistic assessment*. ». Do you mean "radar observations" instead of "*satellite observations*" ? It would be interesting to compare with the BedMachine bed topography.