Author Response to Reviews of

A stochastic parameterization of ice sheet surface mass balance for the Stochastic Ice-Sheet and Sea-Level System Model (StISSM v1.0)

L. Ultee, A. Robel, S. Castruccio
Geoscientific Model Development

RC: ***Reviewer Comment***, AR: *Author Response*, ☐ Manuscript text

We thank the reviewer and editor for their comments. We respond to the remaining points below.

1. Reviewer 1

1.1. Summary

RC: ***The revised version of the manuscript addresses all of my earlier concerns (except two) and has greatly improved. I congratulate the authors to their thorough work. I only have a few minor remaining points that I want to (re-)raise.***

AR: *We thank the reviewer for their renewed attention to our manuscript. We know that the manuscript and review response are long, and the reviewer undoubtedly has many demands on their time.*

1.2. Specific comments

RC: ***l93. "sum to annual time scales" reads like a repetition to l90. It could be made clearer that l90 is a summary statement of what is described in more detail below, like so: "We aggregate each SMB model output field for each outlet glacier catchment at an annual time scale. In order to achieve that, we overlay each field ..."***

AR: *Good point. Revised to "We aggregate each SMB model output field for each outlet glacier catchment at an annual time scale.* ***To achieve that,*** *we overlay each field..."*

RC: ***l175 "Figure 3b shows example best fits for four model types and their BIC (see legend)." – I still find it very difficult to understand that the black line (AR0) in Figure 3b comes out as a good fit to the time series. If that is true, than Fig. 3b is clearly not a good illustration of that fitting result. Could you explain in the context of l175 why a model producing the time series without any variability is a good fit to the noisy time series in 3b? If reproducing the noise is not relevant for a good fit, what is?***

AR: *The reviewer is asking about fitting, which is in essence a question about what we are selecting for by optimizing the Bayesian Information Criterion. We address that in the Discussion, lines 323-326:*

> *Moreover, among low-order autoregressive models, white noise AR(0) models with a trend are preferred over higher-order models in most basins, for all seven process models tested (Figure A1). Low-order AR models could have a low BIC despite relatively greater error than higher-order models, as seen in Figure 3, because the BIC penalizes excess parameters (Equation 2).*

We interpret that there is also a question about what the resulting generator will (re)produce, relative to the training series. We have added the following clarification to lines 179-181:

> *We note that the models capturing only a trend in Figure 3b will still generate stochastic series with temporal variability; the distinction is that almost all of the temporal variability in the final generator will come from the spatial noise generation process described in the next section.*

RC: ***l295 "The same principle could be adapted for training data provided at even finer temporal resolution (i.e., weekly or daily)." I have remarked before that going to different time resolution may not be obvious because of other types of variability and difficulties to produce enough training data. ... I couldn't find a discussion item that specifically addresses problems related to going to a daily time scale. As it stands, I still find the statement above to easily made and not justified without proof that this is actually possible without additional problems.***

AR: *We have revised line 295 to be more general:*

> *The same principle could be adapted for training data provided at even finer temporal resolution,* ***though a large training data set may be needed to capture the relevant variability in sub-monthly SMB.***

The point about choice of training dataset is further addressed in lines 391-397:

> *The example presented here illustrates the possibility of inferring a downscaling function from process-model output. It would be possible to infer similar downscaling functions at different temporal or spatial resolutions, using reanalysis or reconstructed data, or computed over a different reference period. Ultimately, the choice of a reference period and the best spatial dataset to infer such a function depends on the user's intended application, and this selection may be non-trivial. Further, our simple downscaling does not capture changes in elevation dependence of SMB over time, for example due to changes in precipitation phase or local atmospheric lapse rate. Users seeking improved fine-scale performance may wish to implement more granular statistical downscaling methods (e.g. Noël et al., 2016).*