We thank Reviewer 2 for reviewing our work and for providing constructive comments and recommendations to improve our manuscript. We have worked on updating the manuscript to include the suggestions of Reviewer 2 as best as possible. We believe that the revisions have improved the quality of the manuscript, and are addressing the concerns raised by Reviewer 2. This document consists of a point-by-point reply to all the comments. In this document, the comments from Reviewer 2 are in blue, while our responses are in black. New text that has been included in the manuscript is in *black italic*.
A revised version of the manuscript will accompany this response. We will also upload a "tracked changes" version of the manuscript where all the modifications are highlighted. Line numbers in this document refer to line numbers of the revised manuscript without tracked changes.

General comment 1:
Accounting for uncertainty in ice-sheet modeling is of paramount importance and tools like the one presented here are important and worthy of publication. However, I find the exposition hard to follow. Details about the stochastic model are at times buried in the presentation of the numerical experiments and some important details are missing. In my understanding the StISSM provides two stochastic processes in time (potentially different in each sub-domain): an autoregressive process and an one based on Gaussian noise.

We agree entirely with this comment. We have added extensive emphasis on the difference between additive Gaussian white noise processes and stochastic autoregressive processes. In the manuscript, we consistently use the generic variable "y" to represent the former and the generic variable "eta" to represent the latter. We now specify the difference between both explicitly before introducing the equation for autoregressive processes (l 213-215):
We use the notation η to represent a variable governed by a stochastic AR process, in contrast to y, which is governed by an additive Gaussian white noise process (see Eq. (1)).

In Sect. 2.3, we have also clarified that both y and eta are ISSM variables that are represented as random processes in StISSM v1.0 (l 239-240):
It should be noted that η (Eq. (7)) and y (Eq. (1)) are both the realization of a random process used as an ISSM variable, with the former being an autoregressive stochastic process and the latter an additive Gaussian white noise process.

General comment 2:
It also allows to run ensemble members in parallel, although it is not clear how the computational resources (cores, memory) are used.

We have expanded Sect. 2.4 to provide more details about these aspects. Please note that most of the management of computational resources, of the output generation, of the debugging files, and of the run monitoring are identical to what is done in the standard ISSM model. The only difference is that different ensemble members are run separately (i.e., in parallel on different nodes). Because ensemble members do not need to share information with each other and the overhead associated with stochastic generation is negligible, the computational performance of StISSM v1.0 is effectively identical to ISSM as documented in prior studies (e.g., Larour et al. 2012). See l 254-261:
The runs of the different members are executed on different nodes, and each separate member run can further be parallelized on different processors using the usual ISSM parallelization capabilities (Larour et al., 2012). StISSM v1.0 allows any of the possible output variables of ISSM to be saved, and at a user-specified frequency in order to manage output size. Output variables can be scalar (e.g., total ice mass) or multi-dimensional (e.g., ice thickness at each mesh element) fields. From the outputs, ensemble statistics can be computed (example codes are provided, see Sect. Code and data availability). Log files are automatically generated for debugging purposes, as is usual for ISSM runs. The implementation of ensemble simulations in StISSM v1.0 is straightforward, as illustrated in Algorithm 1.

Furthermore, we now also explicitly state in Algorithm 1 that the user can specify the requested outputs and the temporal frequency at which outputs are saved.

General comment 3:
Statistical quantities (e.g. moments, p-values) are computed and reported in the Results section, but it is not clear whether these are computed directly by StISSM or how the data from the ensembles are collected.

Output from the ensemble runs are retrieved as is usual in the ISSM model (Larour et al., 2012). Statistical quantities can be computed from the outputs. We provide the scripts that we used to compute these quantities from the model outputs in the Zenodo dataset associated to this publication (see Code and Data availability section). This has been specified in l 254-261 (see text in the response to General comment 2 above).

General comment 4:
Finally, how is the stochastic layer implemented? How is it coupled to ISSM? Is it a driver to ISSM? Is it implemented in C++, Python or other languages?

All the StISSM v1.0 capabilities are implemented within the source code of ISSM itself. Using stochasticity or not is specified by the user when configuring their simulations. This is specified at the very start of our Methods section (l 104-106):
The new stochastic capabilities are implemented within the core of the source code of ISSM. We refer readers to Larour et al. (2012) for a general description of ISSM. Usage of stochasticity is optional: if turned off, ISSM simulations are fully deterministic.

To further address this concern from Reviewer 2, we specify in subsection 2.2 (l 199-201):
All the stochastic schemes are implemented in the C++ source code of ISSM and are integral parts of the core of the model, but the schemes are not called if stochasticity is not required by the user.

General comment 5:
A good part of the paper is devoted to using the StISSM for applying different stochastic parametrizations to two synthetic ice problems and a real ice sheet (Greenland). The numerical experiments are well thought out but I don't think it is particularly useful to target three different applications. I think it would have been better to target only one application (maybe a glacier) and show the effect of different choices of the parametrizations on the glacier evolution and mass balance. Targeting different (and complex) problems makes it harder to understand the impact of different parametrizations, without adding much in terms of explaining or demonstrating the stochastic model.

Our choice of model experiments is motivated by the fact that this publication is a model development study. Therefore, the main goal of the experiments is to demonstrate the range of capabilities of StISSM v1.0. We agree with the reviewer that there is a large scope for using StISSM v1.0 to addressing more specific science questions, in which case it would be relevant to target one system and do a comprehensive sensitivity study to various parameterizations.
Here, our three model experiments have specific purposes. We believe that they are all relevant to demonstrating the features of StISSM v1.0 and/or to demonstrating the importance of stochastic processes in idealized ice sheet model experiments. Thus, we prefer to preserve all three experiments in the main text as they provide the broadest possible demonstration of the capabilities of StISSM v1.0 which will be useful to the audience of Geoscientific Model Development.
MISMIP+: These experiments show the effect of different choices of autocorrelation timescales in the melt parameterization on the evolution and mass balance of an idealized glacier. We believe that this addresses the request mentioned by Reviewer 2.
IQIS: The stochastic SMB experiment (IQIS_1) demonstrates that StISSM v1.0 can use different subdomains with a prescribed correlation between them. This is an important tool of the model and it needs to be

demonstrated in a simple experiment. The stochastic calving experiment (IQIS_2) shows that even an idealized ice sheet configuration is subject to strong noise-induced drift. The stochastic calving and SMB experiment (IQIS_3) demonstrates that correlation can be prescribed between different variables, which is another important feature of StISSM v1.0. This experiment also demonstrates that there is no one-to-one correspondence between the ensemble spread of IQIS_3 and those of IQIS_1 and IQIS_2, despite IQIS_3 only combining the stochastic forcings of IQIS_1 and IQIS_2. Finally, IQIS_4 shows how the ensemble PDF is affected by temporal persistence in frontal ablation compared to white noise in frontal ablation (which is shown in IQIS_2).
GrIS: The purpose of this experiment is to show that StISSM v1.0 can be applied at the scale of a realistic ice sheet. We have used reasonably realistic climatic forcing fields for the purpose of demonstration.

General comment 6:
Moreover, given that the main novelty introduced by StISSM is that it provides parametrizations, I would have expected more emphasis on 1) why to choose a specific stochastic (e.g autoregressive) processes for modeling, e.g., the surface mass balance (SMB), 2) how the stochastic process compares, in a statistical sense, with available time-series of SMB,

We agree with Reviewer 2 that this is an important scientific question. However, the purpose of this study is to develop the numerical capabilities to exploit such knowledge in ice sheet model simulations. Constraining precisely the statistics of variability in climate and in glaciological processes is, in our opinion, beyond the scope of this work. However, we are working actively on these research questions. Two future publications are currently in preparation, and we are involved in cross-institution research projects to better quantify variability in climatic forcing and poorly constrained ice sheet processes. We agree with the reviewer that this important aspect should be more emphasized in the manuscript of StISSM v1.0. For this reason, we have added some information in the Discussion section (l 569-572):
In this first version of a stochastic ISM, we have implemented simple forms of stochastic processes and statistical generators of climate forcing: additive Gaussian white noise and autoregressive time series models, respectively. This lays the groundwork for future, more sophisticated schemes specifically calibrated to represent the details of variability in glaciological and climatic processes. In particular, priorities are to implement seasonality in the statistical models, more complete time series models (e.g., autoregressive moving average, ARMA), and to allow for other forms of noise forcing in order to represent non-Gaussianity in components of the climate and ice sheet systems (e.g., Perron and Sura, 2013).

Please also note that we reiterate the importance of constraining the parameterizations in the last paragraph of the Conclusion (l 640-643):
In the future, calibration work will be needed to constrain the statistical models for climate forcing, as well as the variability in unresolved glaciological processes such as calving and hydrology. Such an effort will require combining observations, theory, and results from high-fidelity model experiments to understand the internal spatiotemporal variability of processes of interest.

General comment 7:
(…) 3) what is the impact of using a first-order versus an higher-order autoregressive process, and so on.
We provide now a better physical intuition of how first-order autoregressive processes compare to higher-order ones (l 220-222):
The order p of an AR process allows to capture multiple degrees of freedom influencing η, and that may act on different timescales (von Storch and Zwiers, 1999). Using a higher-order AR model thus allows to capture more complicated temporal variability, but implies the risk of overfitting if the calibration time series are too short.

Additional comments:
Eq. (1). Is the Gaussian noise uniform in space even if the mean value is not? Please specify this in the text and discuss this choice. Please specify that the "mean" is intended in time, not in space (if I understand correctly).
We have clarified that the mean is intended in time (l 123):

If y has a prescribed temporal mean value y^bar,
We have also clarified upfront that the noise can vary spatially across the model domain, and refer the reader to more details in Section 2.2 (l 127-128):
As explained in Sect. 2.2, σ_y is fixed in time but can be variable in space, hence allowing ε_y to vary across the model domain.

line 153: would it make sense to have different stochastic time steps for different parameters?
This is a future development that can be considered. This was also pointed out by Reviewer 1. It is not straightforward to implement different stochastic time steps for correlated variables, because noise terms cannot be generated independently in this case. However, we have included this prospect in the manuscript (l 169-171):
At this stage, StISSM v1.0 uses an identical stochastic time step for all variables modeled with additive Gaussian white noise (Eq. (1)), but implementing different stochastic time steps is a possible avenue for future development.
line 164: OK, so the spatial stochasticity is introduced at the subdomain level. This should have been explained before, in the introduction and, in more detail in section 2.1 where it should be explained that eq. (1) is at the subdomain level.
See our response to the comment on Eq. (1) (first Additional comment).

eq. (7): I do not fully understand the purpose of the intercept and trend terms. Also, what is the choice for beta_0 and beta_1 in the numerical experiments in sections 3?
We agree with Reviewer 2 that the explanations on the intercept and trend terms in the previous version were insufficient. We have reformulated Eq. (7) to make it rigorous, and to make it clear how an autoregressive variable is computed. The purpose of the trend (beta_1) term is to allow for a deterministic linear trend in time (e.g., warming of ocean waters or decrease in SMB). The purpose of the intercept (beta_0) term is to allow for a non-zero baseline of the variable. In particular, if beta_1 is equal to zero, beta_0 represents the temporal mean of the variable. We have now specified the values of beta_0 and beta_1 in all the experiments introduced in Section 3.

line 235: can you detail how you manage resources (nodes, cores, memory) when you run in parallel multiple members of the ensemble (each of the members might need to be distributed on several ranks). Do you use any strategy to reduce I/O and storage when running large ensembles? Any strategy to monitor the runs (e.g. what happens if a few of the 500 simulations in the ensemble fail?)
Please see our response to the General comment 2.

eq (8) and (11): this is very minor, but I think that the use of squared terms "C_W^2" and "C_B^2" is poor notation. I know it is somewhat common, but it is misleading because it sort of implies that C_W and C_B have some physical meaning. Using the square to denote positive quantities (if that's the reason for the square) is hardly defensible because there are a lot of other physical variables (e.g. thickness) or coefficients (flow factor) that are positive (or nonnegative) and they are not denoted with a square of some other quantity. I would suggest dropping the square and using directly the coefficient C_W and C_B.
We understand the point of view of Reviewer 2. We have changed all the terms C_W^2 and C_B^2 to C_W and C_B, respectively.

Sections 3 and 4: The rigid separation of the "Model experiments" and "Results" sections makes it harder to follow the exposition. I think that the Results part should follow the Model Experiments part for each of the three examples.
We believe that this is a valid concern. We did consider this possibility, even before the first submission of the manuscript. However, we find it preferable to keep the "Model experiments" and "Results" sections separate. Some readers may be particularly interested in how StISSM v1.0 works, and how to set up experiments with the model. These readers can focus on the Model experiments section. In contrast, some other readers may be interested in understanding the consequences of stochastic forcings on idealized ice sheet model experiments,

without feeling the need to go through the details of the model experiment setup. These readers can focus on the Results section. Finally, we believe that our use of subsections makes it easy to go back-and-forth between the explanations on the model experiment setup and the corresponding results. For all these reasons, we have preferred to keep these two sections separate.

We thank you for your constructive comments.
Vincent Verjans, on behalf of all authors