

Review of “Lessons for multi-model design drawn from emulator experiments: application to a large ensemble for future sea level contributions of the Greenland ice sheet” by Rohmer et al.

This article is a revised version of a previously submitted paper. While some changes have been made to address my comments –for which I would like to thank the authors– I do not believe that the current version is ready for publication, as it does not yet meet the standards of clarity and readability required for publication. I continue to believe that the content of this article is noteworthy and valuable for the scientific community, but further work is needed to make it publishable.

My review is structured in two parts. First, in the general comments section, I address the major changes made by the authors in response to my original remarks. Then, I provide a list of specific comments. These are mostly minor suggestions and corrections, but they are relatively numerous. In particular, I have not gone through the entire text in detail. To facilitate the next round of revisions, I would therefore ask the authors to carefully proofread their revised manuscript, which would help streamline the revision and correction process. As a side note, the article would be more pleasant to read if the figures were included in vector format or, at least, in higher resolution, so that they do not appear blurry.

General comments

General comment 1. I would like to thank the authors for their efforts to reorganize the paper to make it more accessible to a broader audience.

General comment 2. I am pleased to see that a brief discussion on the weight attribution of each ensemble member is now included in the paper. However, unless I am mistaken, this appears primarily in the interpretation of results, and thus appears rather late in the main text (it is first mentioned on page 8). Evaluating the predictive quality of the members is crucial in the context of future ensemble studies; indeed, it is a key step in linking numerical simulations to observations and constraining the former using the latter. It is also a quite natural step from a Bayesian approach, as it allows for relaxing the assumption of a uniform prior. Therefore, it would make sense to introduce this question earlier in the main text, so that the reader clearly understands how each member is to be compared. I suggest adding such a discussion when the inputs are presented in subsection 2.1, at the end of page 5. The authors could also briefly mention it in the introduction when outlining the scope of the paper.

General comment 3. I would like to thank the authors for the additional details concerning the interpretation of the parameter κ . While I am not entirely convinced by the rationale of comparing a single parameter to a full forcing scenario choice, I am fine with the authors retaining this aspect of their analysis. I still wonder whether the meaning of this parameter could be made clearer by explicitly renaming it as ‘ocean forcing’ parameter instead of ‘retreat’ parameter, particularly if it is associated with uncertainties in ice–ocean coupling, rather than being a parameter intrinsic to the ice-sheet model itself. This would make it clearer that it represents a forcing. Furthermore, it seems to me that the structural uncertainty of the ice-sheet models has not been addressed in the paper. This omission might bias the results by underestimating the impact of uncertainties related to ice-sheet physics and models.

Specific Comments

Note: I am using the author’s tracked changes document for the line numbers.

- (1) [Lines 32–33] ‘(IPCC: e.g. Lee et al., 2021)’ → ‘(IPCC; e.g., Lee et al., 2021).
- (2) [Line 36] ‘(e.g. Knutti et al., 2010)’ → ‘(e.g., Knutti et al., 2010)’.
- (3) [Line 37] The fact that each member evenly spans a representative set of plausible realizations

is somewhat misleading, as this is only true if no additional information is not available (e.g., observations). This would benefit from further clarification; see general comment 2.

- (4) [Line 40] ‘(e.g. Merrifield et al., 2020)’ → ‘(e.g., Merrifield et al., 2020)’.
- (5) [Lines 52–55] This sentence is too long. Consider splitting it in two, maybe after the ‘thoroughly’.
- (6) [Line 82] ‘the main model parameter’ is ambiguous: do you mean κ ?
- (7) [Line 94] Consider adding what you mean by ‘as best as possible’, as it is vague on its own. Maybe mention that the misfit between computed and observed surface velocities and/or thicknesses is minimized?
- (8) [Line 98] ‘(Slater et al., 2020, 2019)’ → ‘(Slater et al., 2019, 2020)’.
- (9) [Line 108] ‘parameter values’ is ambiguous: do you mean the ice-sheet parameter values?
- (10) [Lines 109–112] I would remove entirely the discussion about the merge of the two inputs and present your final setup more directly, namely, the use of a GSAT that corresponds to a combination of SSP-RCP and GCM. This would be easier to follow.
- (11) [Line 114] ‘The inputs from the double line’ is not clear. Do you mean the inputs below the double line?
- (12) [Table 1] Please be consistent in your system of notations. Some of the names start with capital letters (e.g., ‘Sliding’), others do not (e.g., ‘thermodin.’).
- (13) [Table 1] ‘thermodin’ → ‘thermodyn’ or even ‘thermo’.
- (14) [Table 1] One way to simplify the reading of the table would be to use math symbols to clarify whether the variables are categorical or continuous. For example, the ISM models would become {CISM, Elmer/Ice, GISM, IMAUICE}, while the resolution would become [1, 40] km.
- (15) [Line 118] ‘as a particular’ → ‘a particular’.
- (16) [Line 121] ‘expressed in meters sea level equivalent SLE’ → ‘expressed in meters sea level equivalent, SLE’.
- (17) [Figures 2–4] I think it would make more sense to first present the inputs (Figures 3 and 4) before the output (Figure 2).
- (18) [Figure 2] Vertical label: ‘density’ → ‘PDF’.
- (19) [Figure 2] Caption: ‘Probability density function of the sea level contribution in 2100 (with respect to 2014) from the Greenland ice-sheet (in cm sea level equivalent, SLE) based on the raw MME ensemble data considered in this study’ → ‘Probability density function of the sea level contribution of the Greenland ice sheet in 2100, with respect to 2014, based on the raw MME ensemble data considered in this study’.
- (20) [Line 132] ‘highest importance’ is ambiguous. Do you mean ‘that contributes the most to the uncertainty’?

- (21) [Line 137] Consider adding one sentence that quickly explains why the design of experiments was indeed unbalanced.
- (22) [Figures 3, 4, C1 and C2] Consider drawing the plots with the count number in the y axis, as that is more common.
- (23) [Line 154] ‘(named emulator)’ is not necessary here as you have already introduced several times the notion of emulator previously.
- (24) [Line 156] Consider putting the reference to the overview as new separate sentence.
- (25) [Line 169] I found the reference to a conditional mean to be not very clear. On what is the mean conditioned here?
- (26) [Line 184] What is meant here by ‘tolerance’?
- (27) [Line 184] ‘in Sect. 2.4’ → ‘in the next section’.
- (28) [Line 204] You mention AR6 here, but this has not been introduced/defined before.
- (29) [Line 209] Consider adding an adverb at the beginning of the sentence here (e.g., ‘consequently’), so that it is clear that the 200 number is directly linked to the 50 number before, and not a new parameter for the validation exercise.
- (30) [Lines 212–218] Given that you introduce three performance criteria, you should use three distinct items, not two.
- (31) [Line 219] Consider removing ‘for a GSAT scenario’ from the title of the subsection. Both subsections 2.4.1 and 2.4.2 depend on the GSAT scenario.
- (32) [Line 266] [4.6; 7.4cm] → [4.6 cm; 7.4 cm] or [4.6; 7.4] cm.
- (33) [Line 266] [10.4; 17.0cm] → [10.4 cm; 17.0 cm] or [10.4; 17.0] cm.
- (34) [Figure 6] Vertical label: ‘density’ → ‘PDF’.
- (35) [Line 269] ‘constructed the Monte-Carlo-based procedure’: there seems to be a missing word here.
- (36) [Line 281] ‘whatever the performance criteria’ → ‘for every performance criterion’.
- (37) [Line 286] ‘Table 1’ → Table 2’.
- (38) [Line 367] ‘(based on Goelzer et al. (2020))’ → ‘(based on Goelzer et al., 2020)’.
- (39) [Lines 410–411] ‘(Merrifield et al., 2023; Evin et al., 2019)’ → ‘(Evin et al., 2019; Merrifield et al., 2023)’.
- (40) [Line 411] ‘take’ → ‘took’.

- (41) [Line 609] You use slc with a superscript for the index, and then with a subscript later in the text. Please be consistent in your system of notations.
- (42) [Line 612] ‘By nature’ $\rightarrow$ ‘By construction’.
- (43) [Line 615] ‘squared errors’: errors of what?