Comment on egusphere-2022-435
Anonymous Referee #2

Following is a review of, “Improving interpretation of sea-level projections through a machine-learning-based local explanation approach” by J. Rohmer et al.

In this manuscript, the authors describe a strategy for the interpretation of an ensemble of Greenland Ice Sheet model projections, in particular, those created for the ISMIP6 experiment. The goal is to use a novel machine-learning approach (SHAP) to analyze the existing ensemble, and bring insight to the ice sheet modeling community, in terms of what modeling choices affect model results and when. This manuscript focuses on the ISMIP6 high end (RCP8.5) projections forced with MIROC5 output, which are provided through the year 2100. The authors find that different modeling assumptions influence results during varying epochs of the projection. In particular, model results are sensitive to the retreat parameter, especially after the first 30 years of simulation, as well as the choice of ice flow equation. A significant dependence of results on minimum grid cell spatial resolution is also found. The authors conclude that the SHAP approach is a promising method for analysis of earth system multi-model ensembles (MME), especially in terms of extracting information about how modeling assumptions may drive simulation results. They note that, with caution, the analysis can offer valuable insight, and they offer suggestions on how to improve upon the approach for future studies. The manuscript is well written and organized, and the figures are of good quality. The authors especially take care in describing the methods, including a schematic to describe the procedure adopted for this study.

Overall, the manuscript is successful in illustrating that the SHAP approach can be used to help researchers interpret results of MME experiments, like ISMIP6. The methods are novel, especially in the adaption of relatively new machine learning techniques to ice sheet model projections of sea-level change. The introduction offers a thorough explanation of the background of the adopted approach and the data section adequately describes the ISMIP6 experiments and model assumptions chosen for this study. However, I find that some additional explanation could be added to the application section, to help lead the reader through the analysis results. I also find that the discussion could be expanded to
add context to the analysis results, particularly with respect to how the results might be compelling for the ice sheet modeling community or how they might impact future ice sheet model intercomparison projects. Overall, I recommend publication of this manuscript with minor revisions.

Below I have some specific comments and suggestions for the authors:

Line 35: Could you please explicitly define what is meant by global vs. local for this context? This is not necessarily terminology that some readers would be familiar with.

Section 2 title, figure 7 caption, and line 468 (and maybe others): Since these are produced from simulations, they are not really "Data", but model output.

Line 98: more accurately, this can be referred to as "global mean sea-level equivalent"

Line 126: maybe "predicted" or "modeled" or "simulated" sea-level change?

Line 357: "sea-level" contribution

Line 360: Could you add a statement about what this might mean from the modeling perspective (similar to what is done to conclude the next paragraph about Fig 7b)? For example, does having a negative sea-level contribution result from a low κ value suggest anything to modelers, or is it possibly too dependent on the specific warming scenario being tested (i.e. RCP8.5 from MIROC5)?

Line 372: First, please take care to lead the reader through your logic, in particular, it would be helpful to explicitly remind the reader that conclusions are based on the red envelope derived from your analysis. In terms of the stated conclusions in this paragraph, it looks to me that 3cm is a value suggested by the fitting curve envelope. It also appears that 3cm is significantly above the plotted interval (maybe ~2.5 cm SLE is a more accurate number here?). Also, if we follow the logic of using the fitted curve to drive conclusions, then it looks like to me that there are values for >5 km resolution that should still be considered negligible. The curve actually suggests that perhaps >7.5km might be a more appropriate cutoff for the >5km statement? In addition, it seems that since the few results from the 3-4km range are driving the 2 km conclusion (as noted in the text). Because of this, I suggest softening the statement to say that results support a minimum grid size of 5km for sure, but they also reveal that a minimum resolution of as fine as 2km may be required, with more investigation needed. Overall, please consider revising this paragraph’s wording, in general, with more accurate statements to reflect the plotted output. (This is important as the results are highly pertinent to ice sheet modelers and
may be referenced to support modeling decisions in the future.)

Line 376: This final statement is a bit awkward. Please consider revising. Maybe something like: “if spatial grid resolution is too coarse, this choice may highly influence the results of sea-level projections.”

Lines 383-384: This sentence is awkward, and it is not explicit what is meant by “mask” in this context. Please reword. For example, “Fig. 8 further suggests that the contribution of minimum grid size might dominate over (?) all the other modeling choices, since they do not exceed those contributions associated with minimum grid sizes of 8km (?) or greater.” Or something similar. With respect to this statement, don’t the ice flow extreme responses technically exceed that of the extremes of the available minimum grid size? That is, aren’t the results in Fig 7b within 10km and 15km artificially high, as an artifact of the fitted curve? This might just be an issue of reworking this paragraph to lead the reader through the stated conclusions in a clear way, but as written, the logic is not obvious.

Line 396: be noticeable in the -> “to impact the”?

Figure 9f title: For consistency, please reference κ in addition to (or instead of) Retreat parameter

Line 424: At this point in the discussion, it would be helpful to add some sentences to put these results into context for ice sheet modelers. That is, what are the implications for some of these findings and suggestions? Fleshing out some of these ideas and expanding upon them during the discussion would broaden the audience who can benefit from this type of study.

Lines 425-435: It would also be beneficial to the manuscript to lead the reader and explicitly explain about how these points pertain to the particular study case here (as opposed to only referring back to earlier sections). For example, is the choice of method mostly appropriate because of the specific model assumptions that were chosen? Though these assumptions do have inter-dependencies, there are many other model choices which could be studied but are much more inter-dependent (for example different physically-based parameters involved with various processes related to ice dynamics). Can you say anything about what type of ice sheet model parameters this method would and would not be appropriate for diagnosing, based on the experience gained in this study? While the language currently included is cautionary, I would like to see more discussion geared towards ice sheet models in particularly, like whether the results may be highly specific to the chosen ensemble (e.g., RCP8.5 MIROC5 climate forcing) or what type of “right” or “wrong” conclusions an ice sheet modeler designing a new intercomparison project might take from the method presented here.
Minor notes:

Line 23: GCM is more typically used to stand for General Circulation Model. If a more general acronym is desired here, I recommend using something like ESM (Earth System Model) for this context.

Line 91: relevance “of”

Line 145: units

Line 226: gets

Line 323: allows “us” (?)

Line 353: values

Line 359: values

Line 364: “indication of” where

Line 374: “a” few

Line 419: setting “of” the minimum grid size

Line 425: helps “alleviate”