This study conducted a comprehensive comparison with ICESat-2 data and CMIP6 sea ice model outputs. The writing and structure is general good. I have only a few minor comments here:

We sincerely thank the reviewer for taking the time to provide this constructive feedback and on our manuscript. Please see below for our responses (in blue).

Figure 1: the quality of this fugure should be improved by using vectorized format.

Yes, agreed. We will make this change.

Eqn 1: I suggest to use a table to include all model constants and variables so that we can be more clear which one is constant and which one is model outputs and which one is calculated in this study.

Yes agreed, we will add a table with model variables, constants and link this to the variables included in Table 1.

Eqn 3: use same math symbol format for h_s

Yes, agreed. We will make this change.

L188: does the sea ice mass include snow?

No, this is just the sea ice mass, snow mass is provided in a different variable.

L191: Table 1, not 2

Yes, agreed. We will make this change.

L214: what is sea area percentage?

This was a bit confusingly described, this is effectively the ocean fraction within a grid-cell. We will clarify this better in the revised manuscript.

A bit of background: The standard name of sftof is sea area percentage; when used as the land/sea mask, it is divided by 100 to give the ocean fraction within a grid cell. This is typically a value of 1 where there is ocean and 0 where there is land, which is what the regridding function from xESMF expects when providing a land/ocean mask (an ocean mask is also provided for the destination grids). xESMF treats NaN values as valid data and, without the mask, would expand those NaN values into the ocean domain as shown in this masking example: https://pangeo-xesmf.readthedocs.io/en/latest/notebooks/Masking.html. The land mask is crucial with conservative normed regridding, which accounts for the fractional overlap between source and destination grid cells. The regridding algorithm (1) finds all source cells that overlap a destination cell, (2) multiplies source values by their overlapping area fraction, and (3) sums

and normalizes to compute the destination value (e.g., if a NaN value is included in an overlap, then without a mask, the resulting destination would be NaN).

L240: Give the full name of NESOSIM if it appears for the first time

Yes, agreed. We will make this change.

L243: remove the comma after et al.

Yes, agreed. We will make this change.

L245: change vs to versus

Yes, agreed. We will make this change.

Fig 2: (a-c) is the total freeboard? Please clarify; (g-i) I think it should be sea ice concentration as the unit of colorbar is %. In the caption, please change (top) to (a, b, c), (middle) to (d, e, f), (bottom) to (g, h, i), same for other figures.

Yes, these refer to total freeboard and concentration. We will make these changes!

L266: so in the IS2SITMOGR4 dataset, there are two density approaches, a constant and a J22 parameterization?

Yes, J22 ice density and derived thickness is included in the IS2SITMOGR4 v3 dataset, but should be considered a highly experimental product variable. We will make describe this more clearly in the revised manuscript.

Fig 3: similar comments as for Fig 2

Yes, agreed. We will make this change.

Table 2: use consistent unit of freeboard through out the paper. Here is cm, but it is m in the figures

Yes, agreed. We will make this change.

Eqn 7: from this definition the plausibility index should be positive, then why there are negative values in Figure 9? and can we interprete this equation that the plausibility is big when \phi is small?

Yes there was a mistake here, we will remove the the mod term as we include the direction of the model bias associated with the plausibility index, as shown in the figures. The overall idea is that the lower the value the more plausible, with a zero value indicating perfect model agreement with observations. We will better clarify this in the revised manuscript!

Section 3.1: I suggest to put the analysis of total freeboard prior to bulk ice density, as the inversion of ice density rely on the data of freeboard.

Yes, agreed. We will make this change.

L407: Eq 4, not 5

We will make this change.

L410: Eq 5, not 6

We will make this change.

L416 and other places: use symbole ±

We will make this change.

Fig 5: use a and b instead of left and right

We will make this change.

Fig 7 and 8: have you tried the relative error instead of showing absolute errors? Also, add (a) and (b) for left and right subplots.

Thanks for the suggestion. We did have some earlier versions in this format but ultimately decided the figures were much more intuitive for readers with the absolute values included. We will add the letter labels to the revised figure.

L545: I would suggest to use “Arctic” and “Antarctica” instead of “Arctic Ocean” and “Southern Ocean” or something different, “Southen Ocean” feels like a much broader region than sea ice actually exists.

Southern Ocean is commonly used for referring to sea ice as its resides in the ocean rather than on the continent, and both “Southern Ocean” and “Antarctic” are used regularly throughout the sea ice literature. We appreciate the suggestion, but prefer to retain original terminology.

Fig 9: why there are negative values while \phi is defined as only positive

As our response above mentioned, this was an error. We removed the mod to include a sign here related to the bias direction.

Fig 11: put the explanation of those hatchings in the caption

Yes, agreed. We will make this change.

L644: fontsizes in “there” not consistent

We will make this change.

Section 4: Here I was expecting some more discussions of the difference between model outputs and observations, and also some suggestions to the modeling community. Afterall, there are quite a bit of implausible regions for both Arctic and Antarctica for some model results.

Discussion of implausible regions was omitted from the submitted draft as we were a bit concerned with the manuscript length. However, we agree more discussion is warranted. We will include a couple of suggestions for the modelling community in the revised manuscript, especially related to the stronger regional biases.