We thank the reviewer for their detailed comments, constructive feedback and appreciate the time taken to review this work. We have carefully considered all comments and have provided our response below.

In the following, reviewer comments start with a **R:** and are set in grey italics, while our responses start with a **A:** and are in red.

Junichi Tsutsui:

This paper describes a newly developed Arctic sea ice emulator for use with a more comprehensive climate and carbon cycle emulator such as MAGICC. This sea ice emulator can integrate insights from complex multi-model climate projections combined with observational constraints regarding the Arctic response to changes in global surface temperature. Elaborated parameterizations dealing with key features of Arctic warming patterns and sea ice behaviors result in the successful emulation of complex models' response of the sea ice area to a wide range of warming level pathways over the 21st century and beyond. Emulators can provide a probabilistic approach to the scientific basis for climate change mitigation, and this paper includes such an application study as well. Thus, this paper is suitable for publication in the GMD. However, the current manuscript may require revisions for clarification and consistency within the scope. The following are my concerns and suggestions for consideration.

---

1. **L53:**

R: *The acronym SSP has not been described yet. However, it is not necessary to limit the range of scenarios to a specific set of scenarios in this context.*

A: We thank the reviewer for catching this, we will define the SSP acronym in the revised manuscript.

The reviewer makes a good point regarding the range of scenarios used. We chose the scenarios presented in this study as they represent a range of high emission and low emission scenarios, while leaving a number of scenarios un-calibrated for testing. Successfully calibrating to a range of scenarios that capture the range of sea ice area response to high emission and low emission warming will ensure our scenarios also capture all warming scenarios in-between. We hope this provides some clarification regarding our choice of scenarios used in this particular study. However, we agree that in future uses of our emulator, it is not necessary to limit the range to SSP scenarios. In the revised manuscript we will remove the 'SSP' acronym from this sentence and define it at the beginning of the Model Description section instead.

---

2. **Subsection 2.2:**

R: *Are there any drifting errors in the preindustrial control runs for the CMIP6 model variables used in this study? Are these erroneous trends removed?*

A: This is a very interesting comment from the author and has provided an interesting discussion for us. As it is largely unknown if PiControl drift errors are a function of global warming level, we have opted to not remove drift here. In addition, as the studies we compare our results with also do not correct for drift, for continuity reasons we also chose not to here. We also test the SIA PiControl which produces negligible drift. However, this has definitely given us food for thought.

---

3. **Subsections 2.4 and 2.5:**

***R:** Given that air temperatures over the sea surface greatly depend on whether the surface is covered by sea ice, decreases in the amplitude of the seasonal cycle, at least partly, reflect decreases in the sea ice area. Is it appropriate to process Steps ii and iii independently?*

Likewise, is it not necessary to address the temperature biases described in Section 2.4.1, regarding possible sea-ice biases?

A: The reviewer raises an interesting point. We separate Step ii and Step iii primarily to validate each parameterisation independently, making it easier to catch any errors that may occur. While Arctic temperature and sea ice are closely linked, by keeping them distinct, we can systematically address the different biases between models and observations at each stage.

Furthermore, Step ii establishes the seasonality for our SIA parameterisation (Step iii). Forcing SIA with seasonal Arctic temperature allows us to generate seasonal SIA projections. If Steps ii and iii were combined, an additional parameterisation step would be required after the Arctic Amplification parameterisation to reintroduce seasonality. Given this, it is more logical to incorporate seasonality directly through a seasonal temperature amplification.

Our intent is for these parameterisations to be used to understand Arctic sea ice projections and relationships. By separating these steps, we can better investigate the sources of observed trends. Additionally, this separation provides a more flexible emulator, allowing us to analyse Arctic temperature independently and explore a wider range of applications.

Keeping these steps distinct helps refine our understanding of the relationship between seasonal Arctic temperature and sea ice area. Capturing this relationship is crucial, as it enables our emulator to account for the non-linearity of winter SIA beyond the calibration period—a key objective of our study. While the relationships between global temperature and SIA, or Arctic annual temperature and SIA, are well-studied, the connection between Arctic seasonal temperature and seasonal SIA remains less understood. Separating these steps allows our parameterizations to be applied more effectively in addressing this knowledge gap.

In regard to combining the seasonal Arctic temperature and SIA biases, while this could apply to one of the biases, the second temperature bias is essential for the functioning of our observational constraint on SIA. This again helps us understand where the CMIP6 models do not capture observations. Combining the two could incorrectly attribute bias to the SIA, or SIA parameterisation.

While we appreciate the reviewer's valuable feedback, we believe it is more fitting for our framework to handle these steps separately.

4. Equations (2a), (2b), and (2d):

***R:** It is necessary to clarify the dimensions of the parameters. Although Equations (2a) and (2b) imply that P, f, and h have the dimension of temperature, if so, Equation (2d) makes h dimensionless, which is a contradiction.*

A: We thank the reviewer for providing great feedback, clarification of the dimensions will help improve our framework.

As 'h' is a scaling factor that is dependent on temperature, it therefore does not have any units and is dimensionless. 'f' represents the amplitude of our tas(AMAB) function and therefore has temperature units (degC). We realise our equation 2 is a bit misleading in this regard. We will edit equation 3 to show 'h' is dimensionless and is a scaling factor using; tas(AMAB) = P(1+h). We will also add a sentence to the equation description stating that 'h' is a dimensionless function that modifies 'P' and that (2b) is written in a simplified form where 'h' implicitly depends on a dimensioned quantity. We will also make this clear in the supplementary tables where we provide with the parameter values for each model. This may shine some light on the dimensions of each of the parameters.

5. L250-251 and Figure 6:

R: *The acronym RMSE is referred to inappropriately regarding its standard definition, that is, the root mean squared error. The goodness-of-fit statistics are confusing because they are indicated by different indexes: RMSE correlation, correlation coefficient, and RMSE fit.*

A: The reviewer highlights an important distinction between the different indexes we use to describe the RMSE. We will correct the goodness of fit statistics in the revised manuscript, to ensure we state we are looking at the root mean square error without the addition of non-standard mathematical indexes. We will also ensure our description of the RMSE is correct.

6. Figure 6:

R: *Figure 6 shows relatively large emulation errors for IPSL-CM6A-LR in May. Is there any need to mention anything about this point?*

A: We agree with the reviewer, that the quality of fit is slightly worse in May compared to other months. We will add a few sentences of discussion around this, we provide an explanation for this below.

We attribute the poorer fit in May is due to our weighting scheme in the SIA parameterisation. Our approach assumes each month requires the same fraction of the previous month's temperature to capture sea ice loss. However, melting in spring (April–May) requires more energy than ice growth in autumn (September–October), meaning May should have a greater weight. Since our calibration applies a single weight across the seasonal cycle (1850–2100) for all SSPs, the emulator does not account for this seasonal variation. While future versions could refine this with an additional weight, we chose a simpler approach, prioritising the fit in the growth months, which are more relevant for understanding the impacts of seasonal sea ice loss. Despite a poorer fit in the melt season, we find it remains sufficient for our study.

7. Subsections 3.3 to 3.5 and the second paragraph of Section 4:

R: *Given that this paper is submitted as a model description paper, the results and discussion in these parts, while providing an interesting application study, give the impression of being out of the scope. The introduction should provide the context of the application use presented in this paper. The related results and discussion should focus on methodological matters rather than mitigation issues, such as the remaining carbon budget, to be consistent with a model description paper.*

A: The reviewer makes an important point. We appreciate their feedback regarding the scope of the results and discussion in relation to the paper's classification as a model description paper.

However, our intention is not to present a standalone, runnable model, but rather a parameterisation framework designed to understand Arctic sea ice projections. Specifically, we aim to assess whether our approach can reproduce CMIP6 projections and, using these as a basis, correct sensitivity biases when comparing models to observations.

Given this focus, we believe it is appropriate to illustrate the impact of observational constraints on projections, as this directly informs the utility of our framework. While methodological considerations remain central, the broader discussion of model biases and their implications for sea ice projections is integral to demonstrating the relevance of our parameterisation approach.

We therefore respectfully maintain that including the mitigation issues discussed align with the purpose of the paper and provides necessary context for its application within sea ice discourse. However, we acknowledge the importance of clarity in framing the study and will refine the introduction to better emphasise the intended scope and objectives.

8. L289-291:

R: *'slightly larger' is valid for sensitivity to temperature, but not for sensitivity to CO2 emissions. The right panel in Figure 7 shows 'slightly smaller' for the latter.*

A: We agree with the reviewer, they are correct in saying that our observationally constrained emulator projects a sensitivity to global warming that is slightly larger, but a sensitivity to CO2 emissions that is slightly smaller. While we do specify we are discussing the global temperature when we say "slightly larger", we will make it clear in the revised manuscript, that the against CO2 emissions it is "slightly smaller".

9. Code and Data availability:

R: *The README document in the code archive should include a description of the programming environment and where to obtain the input data necessary for execution.*

A: The reviewer provides valuable feedback, aligning with other reviewer comments.

We will update and improve the README in the repository to clarify the programming language and where to obtain the data. We have cited the observational data we have used, rather than providing a link as these can change regularly, however we realise we did not cite the Earth System Grid Federation (ESGF) as the source of our CMIP6 data. We will also make clear that we use MATLAB R2024b, and highlight that this is a framework that is intended to be used to reimplement the parameterisations presented, rather than a 'single-click' runnable model.

R: *The current manuscript describes what data were used but does not describe how they can be obtained. The observational and reference data used in this study are described in the main text.*

A: We have not included a sufficient description of all data used and how they are obtained. To address this, we have included an additional section 'Data Collection and Processing', which we will add below the 'Model Description Overview' section. Here we will discuss in further detail which data is used, from all CMIP6, observational, RCMIP6 and MAGICC sources, while in the Data Availability section we will provide citations for where we obtained the data.

10. Figure 1:

R: *The figure caption would be more suitable with 'calibration and constraining processes' than with 'conceptual model.' Font size should be increased for readability.*

A: We agree with the reviewer that a better description of the model may be necessary. We will take the reviewers advice and update the figure title/ caption.

In combination to another reviewer's comment, we will add more information to the Figure 1 caption and update the caption to 'Conceptual schematic of the parameterisation framework', or 'A work-flow of the calibration and constraining processes in our parameterisation framework'.

11. Figure 2:

R: *Is the description 'Ranges indicate the 17th-83rd percentile of the scenario range' correct?*

I assume that it is a model-structure uncertainty range rather than a scenario range.

A: The reviewer is correct, we are referring to the 17th-83rd percentile of the structural uncertainty for each of the SSP scenarios presented, rather than the 17th-83rd percentile of the scenario range. We will update our figure caption appropriately.

12. Figure 3:

R: *The figure caption should include a description of each panel. What is the histogram at the right end of panel (a)? Panel (a) shows that the values for CMIP6 models fall to zero from place to place. What does this mean?*

A: We will take the reviewer's advice on board and update the figure caption for figure 3 to more thoroughly describe the plot. We discuss this further in the supplementary however, the histogram represents the distribution of Arctic Amplification values, highlighting their spread and frequency, as derived from the randomly sampled 's' and 'p' variables.

Additionally, 'Panel (a) shows that the values for CMIP6 models fall to zero from place to place' as the Arctic Amplification represents the warming ratio between the Arctic and global temperature anomalies. The low signal to noise ratio for small warming grades in some models cause the Arctic Amplification to near 0 when global warming anomaly or the Arctic temperature anomaly is still relatively small in a given year. This therefore makes the AA fluctuations more pronounced.

13. Figure 7:

R: *If the quantity of CO2 emissions is cumulative, it should be clearly indicated from which point in time it has accumulated.*

The unit of weight should be clearly stated as tCO2 or tC.

A: The reviewer makes a very good point, we will include the point in time (1750) from which the CO2 emissions have accumulated.

We will also update figure 7 to m^2/tCO2 if it makes the figure clearer, however we initially used m^2/t as this has generally been the standard unit used in other studies discussing the sea ice sensitivity to CO2.

14. L126:

R: *In Equation (2a), the outermost and innermost parentheses are redundant.*

A: We have removed the outermost parentheses from Equation (2a), however the innermost parentheses are required to group the variables we apply the cosine to. Unless we have misunderstood the reviewer's comment, in which case we would be happy to discuss further.

15. L496:

R: *In this journal, references may be provided with full author lists. I notice that the author list of Nicholls et al. (2020) is incomplete.*

A: While this is true, for aesthetic reasons we choose to keep the authors in the reference list as short as possible, as we believe the list reference list could become messy and unreadable as I have cited a number of papers with a large number of authors.

16. L129:

R: *Considering the difference in the number of days in each month, the equally spaced m values do not exactly represent each month.*

A: The reviewer has highlighted an interesting point. While we agree, we technically emulate the average temperature in each month given by 'm', and our plotting interpolates between these points. However, we agree that our wording in this sentence is misleading. We will correct it to a more appropriate description of 'm'. Perhaps the following could be a possible revision: 'm' is an equally spaced value between 0 and 2pi representing the average point of each month of the year (0 and 2pi represent January of year 't' and January of year 't+1' respectively).

We would be happy to discuss this further with the reviewer if our revision suggestion does not adequately describe 'm'.

17. Minor comments and technical corrections:

A: In regard to the following minor comments and technical corrections, we will address all the minor revisions in the revised manuscript, and thank the reviewer for picking up on them.

Figure labels and markers:

R: *The labels in the figures and the median markers in the box plots are too small.*

L18:

R: *The acronym SIA is not explicitly defined.*

L129

R: *Considering the difference in the number of days in each month, the equally spaced m values do not exactly represent each month.*

L136

R: *To avoid 'moves ... vertically' that implicitly assumes the orientation of a graph plot, how about rewriting 'a non-optimised parameter that moves the temperature curve vertically to ensure ...' to 'a non-optimised offset value to ensure ...'?*

Likewise, 'vertically shifted' on L207 should be rewritten appropriately.

L180:

R: *The description of the seasonal asymmetry of SIA changes may need some literature.*

L285

R: *Probably, 'than' is an error for 'that.'*

L289

R: *Delete 'm' just after '-2.5'*

Figure 4

R: *The caption includes duplicated 'are'*

Figures 4 and 5

R: *Although the selected CMIP6 models are arbitrary examples, it appears preferable to show the same set of models in these figures.*

Supplementary tables

R: *The units of the values are not stated.*