We thank the reviewer for their detailed comments, constructive suggestions 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.

## Anonymous Referee #1

---

### 1. General: Model Evaluation:

***R:*** *The model evaluation is done looking at extensions of the datasets out to 2300, from a calibration of the scenarios up to 2100. To me, a more useful model evaluation would consist of applying the framework to other scenarios run by the same model. Most models presented here will have run one or more additional scenarios such as SSP-3.70 or SSP-1.19. In addition, many of them have run multiple ensemble members. Considering how the fits would do in that context to explore variability would make a lot of sense to me, at least producing some plots to see how the parametrization fits in the range of the climate models internal variability.*

**A:** We thank the reviewer for their interesting thoughts regarding possible emulator evaluation methods. We agree with the reviewer that an understanding of how our emulator captures internal variability is an important one. While we did consider this option along with a number of other evaluation methods, we decided to focus our evaluation on the long-term performance of our Arctic sea ice projections for the following reasons:

The long-term variability expressed across different ensemble members is comparatively small relative to the model to model spread. In addition, internal variability plays less of a role in long-term Arctic sea ice uncertainty than structural uncertainty (Bonan et al, 2020). We therefore chose not to evaluate the performance of our emulator by testing its ability to reproduce other ensemble members, as this might lead to the false impression of a close emulation given the strong similarities across different ensemble members. While internal variability is important, it is largely constrained within a single model's ensemble due to shared physics and parameterisations. Testing against another ensemble member may thus overstate emulator skill, as it would not demonstrate an ability to generalise across broader climate conditions. Instead, by validating against different external forcing scenarios over long time periods, we ensure that the emulator accurately represents the system's long-term trajectory, which is one of the main aims of this paper. Focusing the verification on time periods outside of the calibrated domain allows us to evaluate whether the emulator can faithfully represent both historical and future changes, making it a more robust test of emulator suitability.

In addition, one of the intentions of this study is to understand whether the physical relationship between SIA and temperature between 1850 and 2100 provides information on future sea ice loss. This provided another reason to test whether our emulator could capture the non-linearity of winter sea ice loss outside of the calibration period, as it provides a way to understand whether future sea ice loss can be understood from the physical mechanisms over the calibration period.

Although our emulator does not attempt to capture the internal variability, we have added a figure to the supplementary (Figure 1 below), applying our calibration parameters from ensemble member 1 to other ensemble members for CanESM5, to test their application. This simple test shows our emulator does capture other ensemble members for this model well. In addition, we have also added an evaluation of our CMIP6 parameterisations with SSP scenarios not used in the calibration process (specifically SSP3-7.0), as we agree with the reviewer that this is a valuable form of parameterisation evaluation. We have added this to the supplementary, as the focus of this paper is focused on how the observational constraints to the CMIP6 calibrated parameterisations and long-term projections further our understanding of Arctic sea ice possibilities, rather than an evaluation of our parameterisations.

---

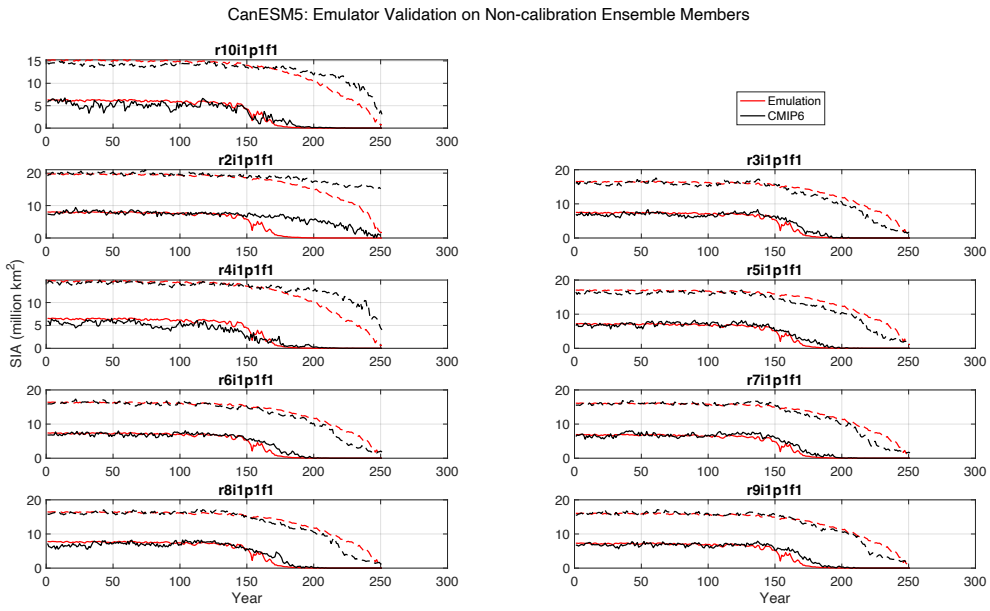


*Figure 1| Testing our calibrations on the internal variable of non-calibrated ensemble members from the CMIP6 model CanESM5.*

## 2. General: Overshoot:

***R:*** *Another point of particular interest to the applications for this tool which I am missing some discussion of, is overshoot. Can this emulation deal with that, or do you believe it is beyond the validity of the parametrizations? Is it only beyond that if accelerated tipping points occur before the maximal warming is achieved, or will it also be true otherwise. I think some discussion of the applicability of this tool to overshoot scenarios is necessary here.*

**A:** The reviewer suggests an interesting application for our emulator. One of the scenarios we train our emulator parameterisations on is SSP1-2.6, where temperature recovers slightly, leading to a corresponding recovery in our emulated sea ice (Fig. S13). Although this isn't an overshoot scenario it shows that if the temperature recovers, our projections follow the ESMs scenarios and also recover.

However, we don't train our model on actual overshoot scenarios. Some models that simulate overshoot suggest the Arctic and northern latitudes could remain cooler under overshoot scenarios due to the AMOC weakening and the longer time it takes to recover after peak temperatures (Liu et al, 2020; Jackson et al 2015, Schwinger et al, 2022). However, our sea ice projections are forced by temperature, so if the northern hemisphere temperature remained colder for longer, it's reasonable that our sea ice projections would show sea ice growth here. Furthermore, while it is likely scenarios with lower overshoot (small increases above 1.5℃ or 2℃) would have little impact on winter sea ice, it also is unclear how scenarios with larger overshoot (larger temperature increases before decline), could affect the non-linearity of winter sea ice.

As our emulator hasn't been trained on such conditions, we agree with the reviewer that it would be interesting to test our emulator on overshoot scenarios. While this is an interesting application, an analysis of overshoot applications using our emulator is outside the scope of this initial analysis. We believe an analysis of overshoot would be better placed in further emulator applications where an in-depth analysis and discussion can be undertaken. We do not believe a deep enough analysis can be discussed here, if included into the revised manuscript. We will add a few sentences to the revised manuscript to mention overshoot and its interest as a further application. Further work should therefore examine how overshoot is captured in our emulator in a first step. While a second step could train our emulator on overshoot

scenarios. This would allow our emulator to be used as a tool to understand Arctic sea ice under overshoot scenarios more robustly.

### 3. General: Framing

***R:*** *When I come into a model description paper, what I hope to see is a description a model that can be reused also by others outside the group of the authors, where the code is nicely reusable, with interesting scientific applications and somewhat of a finished product. In this case (and a lot of other cases), what I see is rather the description of a really interesting parametrization framework for a Arctic temperatures and sea ice, calibrated to model output, but with functional forms and based on empirical and mechanistic expert judgement. I also see a workflow in steps, but it is less clear how the code would allow me to apply that workflow to new datasets. Instead the code looks more like supporting material that would allow the reproduction of the current results, rather than something that is meant for further use (at least by outsiders). (An added problem here is the lack of a proper README-file for the code, and the choice of programming language, which requires licensed software (Matlab) possibly with an unknown number of add-on libraries with their own libraries to be run). All of this takes nothing away from the work presented, it is very nice and merits publication, but I think this is less the description of "an emulator" and more the description of an emulation framework (personally I'm not sure the word emulation is even the best here, as currently it seems to be used to mean practically everything, hence the meaning of it gets sucked out of the term). I will give more advice in detail below, on what would help in making this a better description paper regardless of whether we are describing an emulator or a framework, however, I'd like the authors to consider what they think this article is describing in the way that they write it, as that might also add to the clarity of the presentation.*

**A:** The reviewer makes a very important point regarding the framing of our manuscript. We have labelled the manuscript as a model description paper, however we agree it is more of a framework paper.

Our emulator code also does not currently support reusability. The code is not easily re-usable as it can only be used with specific datasets, as the reviewer mentions, to reproduce the current work. We therefore agree with the reviewer that we have outlined a parameterisation framework, rather than a body of code that can be used on a number of datasets and run easily.

As the purpose of this paper is to understand how the use of a simple model impacts long-term sea ice projections, we therefore intend for this manuscript to showcase an analysis using the described parameterisation collection on questions within the sea ice discourse, rather than a fully-fledged model that can be used on a range of datasets. We intend to update our emulator setup and code in further versions to be an easily reproducible model. Whereas here, we rather intend to highlight our emulator as an initial parameterisation analysis, that can be refined in further versions.

When revising the manuscript, we will ensure that our work is clearly framed as a parameterisation framework aimed at exploring whether the set of parameterisations presented can, in the first instance, capture the CMIP6 response and how this framework can be used to understand and address key biases in within the models. While our parameterisations can be adapted and applied by readers in their own work, our intention is not to present a reusable model in this manuscript. Instead, we provide an initial discussion on the potential of this framework to address current questions.

In reference to the .README and code, we discuss this in further detail under comment 9 below.

### 4. Title:

***R:*** *Starting with the title, I believe that one requirement of GMD model description papers is for model name and version to be included in the title. Is the model called "Arctic Sea Ice Emulator v.1"? If so, prefacing the name with the article "an" seems strange. I note that the github code repository for the code has the descriptive yet somewhat unusable name "Development-and-Application-of-an-Arctic-Sea-Ice-Emulator-*

*Journal-Article-2024", if the name is in fact "Arctic Sea Ice Emulator", that would be a better name even in the repository. For a model description paper, I expect a description of a model that can be used more generally, so the model shouldn't be named and coupled this closely to the paper, but this also feeds into my slightly more general point on what you mean for this article to be about.*

**A:** We agree with the reviewer. The title does not accurately reflect the papers intention or the model description requirements.

We did originally intend the model to be titled "Arctic Sea Ice Emulator v.1", however, as the paper is more of a parameterisation framework paper, a complete restructure of the title is probably necessary. The following updated titles may offer possible revisions:

"An Arctic Sea Ice Parameterisation Framework v.1".

Or, perhaps

"The Development and Application of an Arctic Sea Ice Emulator: A Parameterisation Framework v.1"

We will also update the github repository name, so that that it implies the parameterisations at least may be used more widely.

## 5. Abstract:

***R:*** *Just a couple of what I assume are typos here:*

***R:*** *Line 3: "we developement" -> "we develop"*

***R:*** *Line 5: "emitted to prevent" -> "emitted while preventing" (or something like that. I think nobody is emitting CO2 to prevent seasonally ice free conditions in the Arctic, and if they are that would be a very bad idea…)*

**A:** We thank the reviewer for picking up on these errors, we will correct them in the revised manuscript. In response to the last comment, we do indeed hope no one is emitting to prevent ice free conditions!

## 6. Introduction:

***R:*** *Line 39: Not sure "ascertain" is the right word here, I think I would prefer something like "find"*

***R*** *Line 49: "we aim assess" -> "we aim to assess"*

**A:** We agree with the author. We will again fix these grammatical issues in the revised manuscript.

## 7. Arctic Sea Ice Emulator Setup: Model description and data:

Firstly, the reviewer makes a number of great comments that will improve this section of the manuscript, we thank them. As there are a number of interesting comments regarding the model description and data, we have responded to each of the reviewer's comment separately below.

***R:*** *I would maybe not have a subsection headline for the overview but prefer to have that just immediately, but no big deal.*

**A:** We agree with the reviewer that the additional subsection is a little unnecessary, we will remove this in the revised manuscript.

***R:*** *When going through the overview I would also like references to the sections in which the various steps are described. I would also have preferred to have a separate section on the data used as this is quite instrumental to the workings of the methodology. Currently, the description of what is used for the CMIP6 models is described in the overview in brief, whereas the observational data is described in the data availability section. Ideally, I think both should be described with some discussion of choices made in a separate subsection here, while how to get those datasets should be described for both data types in the data availability section.*

**A:** The reviewer makes a very good point that is also highlighted by other reviewers. We did not include a good description of all data used and the reasoning for their use in the text. We agree it would be best to include an additional 'Data Collection and Processing' section which we will add below the 'Model Description Overview' section. We also agree that links to each of the sections within the manuscript we discuss in the model overview should be included, we will add this to the revised manuscript.

***R:*** *Tables of involved models and ensemble members used would be useful too. One option if you think this fits poorly in the main text is putting such a section including model and observation data tables in the supplement.*

**A:** We also agree that a list of all CMIP6 models used is useful to reference here. We will include a table in the supplement that shows each model used and the respective ensemble, however as we focus only on the first ensemble member, this may not be necessary. We will also add a link to this table in the Data Processing and Collection section.

***R:*** *Overall, to be a functioning model description paper I think each subsection should also mention the code/ scripts that implement the procedure described in the section.*

**A:** We also agree that from a usability perspective, a link to the code would be ideal, if our code repository represented an easily runnable model. However, as we have changed the approach to focus more on a parameterisation framework, we think this could cause confusion. We will instead update the README to point the reader to the appropriate sections in the manuscript.

## 8. Arctic Sea Ice Emulator Setup: workflow schematic

***R:*** *The workflow figure is very nice, but I'd like some more information explaining it in the caption, including information on what sections to refer to for more information on the various steps and workflows involved.*

**A:** We realise we did not add a significant explanation to the workflow schematic. We will update and add additional information explaining the schematic, with links to the appropriate sections in the revised manuscript.

## 9. Arctic Sea Ice Emulator Setup: code

We have combined all the reviewers' comments regarding the code runnability / REAMDME/ github repository, and provided our response here as there were a few comments regarding these area throughout the reviewers comments.

***R:*** *I am, however, missing a good connection between the text and the code base. How do I run it? How does it work? Given the right data, can I plug and play the whole workflow, or do I need to do each step separately. Can I run it for a different (set of) models or experiments? All of this information doesn't have to be given in detail in the overview, (especially details on how I run it or how it works), but since I also can't find it in the README in the repository, I don't really know where to look, and some of these questions should be answered in the overview, so I know whether I can expect this to be a runnable tool, or a description of a framework that I could possibly reimplement from the description (that is also fine, but again that speaks to what this is and isn't.)*

***R:*** *In the code availability section, I'd like there to be some more details. I think linking to the github repository (so one can easily see it exists) is useful, and some explanations on what requirements to run are (you could also solve that in the code README, at least detailing which Matlab extra libraries are needed), but I think here or somewhere you should let the reader know that your code is in Matlab so they know they might not be able to run it (this is less important if you mainly consider this to be a parametrization scheme see previous comments on this). Regardless, the README should include code flow and how to run to reproduce the various parts of the paper.*

**A:** The reviewers highlight a very important missing link regarding the usability of our code and the in-text discussion. Unfortunately, the whole workflow cannot currently be run from a 'single click', it must be run in separate steps. Given previous comments from the reviewer regarding the intention of the paper, we provide a similar response here, reiterating that as we have written the paper as a parameterisation framework, rather than a complete model that can be taken and used with a number of different datasets, this should be made clear in the paper text and the README. We will update the README in the repository to answer some of these questions, including making clear that we use MATLAB R2024b, and highlight that this is a framework that is intended to be used to reimplement the parameterisations presented. Since we are reframing the paper, we prefer to avoid extensive in-text references to the code, as this is not strictly a model description paper, we will therefore address most of these issues in the README, rather than in-text.

In response to the reviewer's comment regarding the possibility of using our code on other ensembles; while this code is only intended to be used as a method to 'reimplement' the process outlined in the manuscript, where we only use the first ensemble member, it is possible to use our code to calibrate or test other CMIP6 models and ensembles, as it is not solely set up for the models and ensembles we currently use. However, as we don't explicitly use this method in our manuscript, we don't highlight this in the paper text to avoid further confusion of the intention of the paper. However, we will mention in the README that this is possible.

We will address the bulk of these comments in the README rather than the data availability section, as we intend to keep the paper as a framework paper, in response to the reviewers' previous comments. However, we will also provide a brief overview of the usability in the model description to make this clear, as this is important information currently missing from the manuscript.

In regard to the presence of "thesis" within the code, this is an error on our part. This paper is an edited excerpt from my PhD thesis, however it appears we missed some references to the thesis when editing. We apologise for this and will edit some of the code to ensure the Section references are related to the manuscript and not a thesis.

## 10. Arctic Sea Ice Emulator Setup: parameterisations

***R:*** *Another general comment is that there are a lot of parameters and constants defined here, and it seems that the same letters are reused for different things here (though it is not always entirely clear).For instance, is the a in 3c and 3d, the same as the a in 3b? I would much prefer if no letters were reused. Referring early to the parameter tables in the supplement and stating the number of free parameters would be helpful, but also having more subscripts on the parameters to explain what they do, for instance a in 3b could be*

*a_linSIAloss or something like that, that would help in following the equations. I think only a relatively simple going over to clarify this more would help the readability of the equations a lot.*

**A:** We appreciate the reviewer for pointing these issues out. Updating the parameter names will definitely improve the readability of the manuscript. Firstly, the 'a' in 3c and 3d refers to the same parameter, as 'e' and 'a' are related, we therefore based 'e' on a simple linear regression of 'a'. However, a number of our parameters between other equations e.g the 'a' in the tas(AMAB) equation and the 'a' in the SIA equation are not related, which could definitely cause issues when reimplementing the parameterisations. We will therefore update a number of the variables in each equation and add descriptive subscripts for clarity and to ensure that it is clear which variables relate to each other.

***R:*** *Line 86: The equation for tas(AAAB) given here deserves an equation number and a separate line.*

**A:** We didn't initially give the Equation for tas(AAAB) on a separate line as we thought the equation was perhaps too simple to warrant a full equation. However, to make it easier to read, we agree with the reviewer that it should be presented in an equation line. We will add this to the revised manuscript.

***R:*** *Line 93: "for the first ensemble member of each", this turn of phrase confuses me here. My understanding was that you are only using the first ensemble member, hence the procedure her is that you find tas(AREF) for whatever model data you are trying to replicate, so in principle, I could choose some other member and do the same?*

**A:** The reviewer's comment regarding our use of the phrase "for the first ensemble member of each" is interesting. In this section we aimed to clarify that we calculate the absolute Arctic Annual mean temperature using the first ensemble member of each CMIP6 model used as this is the ensemble we calibrate to. However, as we mention this in the model calibration section it perhaps does not need repeating, and is also potentially misleading if the reader applies our framework to other ensemble members. We will therefore remove this section of the text in the revised manuscript.

***R:*** *Line 102: Equation 1, how is this tied to the equation for tas(AAB), is p matched up to fit the linear regression? There are so many parameters and functions floating around, please make it as clear as possible how they fit together.*

**A:** We provide a more detailed discussion of 'p' and 's' and how they relate to the first equation (tas(AAAB)) in the supplementary, however we agree that reference to this, improvement of our parameter names, and an added few sentences is necessary to make clear what they represent and how they link to other equations in the main text.

***R:*** *Line 126: I know this might be outside your control, but it would be very helpful if this equation came on the next page alongside the next equation and its description*

**A:** We agree that the positioning of Eq.2a would be ideal above 2b, unfortunately this is not in our control, given the nature of LatTeX. However, given the corrections in the manuscript and addition of a data section below the model overview sections, these are now on the same page closer together in the revised manuscript.

***R:*** *Line 129: "where tas(AAAB)", were you meaning to introduce tas(AMAB) here? That's the new term in the equations? The way it stands now, I got quite confused and had to reread several times…*

***R:*** *Figure 4, caption, second to last line: Duplicate "are" here, "2100 are are displayed"*

***R:*** *Line 161: "which is we define" -> "which we define"*

***R:*** *Line 198, equation 3b: As far as I can see, the term (1 + exp(tas(t=0 -b ))) is a constant (though model and parameter dependent) term. I would consider giving it it's own name and defining it in its own equation, at it makes it easier to read the functional form and evolution of the sea ice area from the equation.*

***R:*** *Line 205: "on the denominator" -> "in the denominator"*

**A:** Finally, in regard to all other minor corrections mentioned above, we thank the reviewer for highlighting these issues and will correct them in the revised manuscript.

## 11. Results

As several of the reviewer's comments from the general response are discussed further in this 'Results section comments', we refer to our discussion in these initial response sections throughout our responses here.

***R:*** *Line 246-254: A little more detail on this evaluation would be nice. Did you have data out to 2300 for all of the models and experiments considered? Was this part of your model selection criterion? This information could easily have fit into a dataset section in Methods, but if it isn't there, we need that info here (and repetition here might be useful anyway).*

**A:** We agree with the reviewer that perhaps our explanation is lacking here. We will elaborate on this in an added Data section (discussed in response 7).

***R:*** *All of 3.1: I'd really like a discussion/understanding of why you don't test using other experiments for the same models/ensemble members also. Is it because you would need separate calibration for each experiment? If so that again would make this much less useful as an emulator (but just as useful as a parametrization tool to understand Arctic Sea Ice loss, so no shame in that). Anyway, you should discuss this, discuss whether there could be some extension to other experiments, I'd be particularly interested in overshoot scenarios. Even if you need the scenario itself to get a new parameterization, checking whether the scheme works for overshoot would still be interesting, and I'd be very interested in such an application, or at least hear the arguments why you would think the parametrization might not work for that to understand the limitations of the scheme.*

**A:** As we discuss our reasoning for our validation approach through our response under the **General: Model Evaluation** section, here I will state briefly again that our parameterisations are flexible. While we have only shown a portion of its ability (as we intended for this paper to briefly showcase an emulator before using it to discuss observational constrain results and how this effects our application), our parameterisations would not require re-calibrating for each ensemble and scenario outside of the calibration scheme used. We calibrated over the 3 extreme scenarios, which allows our parameterisations to capture all experiments in between. We show this in the additional analysis we add to the supplementary where we test our parameterisations ability to capture SSP3-7.0 without calibration to these specific scenarios. We see our emulator captures these well, however as this isn't the focus of our analysis and questioning in this paper,

we put these into the supplementary. We did not make this clear in the original manuscript, but intend to make this clearer in the revised manuscript.

***R:*** *Line 311: From the previous text it looks to me like this should be -14Gt rather than 14Gt*

**A:** We did not initially add a negative sign as we indicate in the sentence that 'we have surpassed the remaining carbon budget by 14Gt', implying the sign is negative. However, we agree with the reviewer that it would improves readability by adding the negative sign in the revised manuscript.

***R:*** *Line 332-337: That is the t in these formulae? tonne of CO2?*

**A:** We thank the reviewer for their comment. The 't' in GtCO2 and $m^2$/t does indicate tonne. We will ensure we make this clearer in the revised manuscript.

## 12. Discussion

***R:*** *Line 367: "who use five" -> "who used five"*

**A:** We thank the reviewer and will correct this in the revised manuscript.

## 13. Data availability

***R:*** *I've commented on this further up, but just noting again that details on the CMIP6 data used and the availability of it should also be listed here.*

**A:** We thank the reviewer for highlighting this distinction between in-text data descriptions and data availability. We have, as discussed earlier, updated the data and data availability section to contain the appropriate information.

## 14. Supplement

***R:*** *Table S1 and table S2, the references in the last sentences should not be "bottom of the table". I assume you are referring to tables S3 and S4 respectively, if so, do that.*

**A:** We agree with the reviewer. We will remove the reference to 'Bottom of the table" from Tables S1 and S2 and place the sentence in the correct tables. We will also move the IPCC definitions to Section S2.3.

***R:*** *Some tables with error scores per model would be good to see.*

**A:** We have provided standard deviations for each of the models and equations used in the supplementary. In the revised thesis we will also provide RMSE values.

***R:*** *S2.2.: This section is strange. Are you just defining the IPCC terms here? If so that doesn't fit the headline very well. To me it looks like this should be a part of Section S2.3, i.e. definitions needed for that.*

**A:** The reviewer is correct, the supplementary section where we define the IPCC range definitions was labelled with an incorrect heading. This section was intended to provide more method and context to Figure 9. We will therefore re-label this header appropriately.