Reviewer 1:

Thank you for thoughtfully addressing my initial round of feedback, either within the manuscript text or through the author's notes where appropriate. It is evident that you have taken the feedback from both reviewers seriously and have invested significant effort into revising the manuscript, resulting in substantial improvements to its quality and clarity. Your dedication to enhancing the manuscript is greatly appreciated.

That said, there are still a few minor outstanding issues that should be addressed prior to publication. I am confident that resolving these will further strengthen the final version of your work. My comments below are grouped into the following categories, more general vs. specific line by line comments.

Thank you again for your hard work on this manuscript. I look forward to seeing the polished final version.

> Thanks for these very kind words. We sincerely appreciate the constructive comments of this review.

General Comments

The repeated use of phrases such as "our EM–GC model" and "our model" struck me as somewhat unusual in the context of model description and application papers. Typically, authors in such papers refer to "the model," use the model's name directly, or specify the version to ensure clarity and consistency. While there is nothing technically incorrect about the phrasing used, it did make me wonder if the authors were intentionally aiming to differentiate their model from another existing EM–GC model. Clarifying whether this is the case or adopting more standard language might help avoid any potential confusion for readers.

> We have reduced the use of the word "our" from 53 occurrences in the most recently reviewed version of the paper, to 25 occurrences in the revised manuscript.

The methods section has noticeably improved, and I appreciate the effort put into enhancing its detail and structure. However, as a reader, I still have some questions regarding certain aspects of the methodology. Addressing these points could significantly enhance the reproducibility and transparency of the overall manuscript.

Radiative Forcing Time Series: Since subsection 2.1.2 Radiative Forcing of Greenhouse Gases is part of 2.1 Model Inputs, does this mean that the radiative forcing time series are considered inputs to the EM–GC model? This seems to be implied by the statement: "The EM–GC relies on time series of the RF due to GHGs." If so, what role do the greenhouse gas (GHG) concentrations play within the model?

> GHG concentrations are used as input for our calculation of the RF due to GHGs, which is then input to the model.

The sentence that had read:

> The EM-GC relies on time series of the RF due to GHGs, computed from time series of the atmospheric abundance of each gas.

has been changed to read:

> The EM-GC uses as input time series of the RF due to GHGs, computed from time series of the atmospheric concentration of each GHG.

to address this comment.

Role of HadCRUT5 GMST Anomaly: Could you clarify whether the HadCRUT5 global mean surface temperature (GMST) anomaly serves as an input to force the model simulation during the historical period? Alternatively, does the EM–GC model use historical GHG concentrations to drive the simulation from 1850 to 2014 and then use the HadCRUT5 GMST anomaly to constrain the model runs? Expanding on this distinction would help clarify the methodological framework.

We are sympathetic to this comment, as the model description is terse. The terse nature of the model description was a change made, in between submission of a much longer manuscript, and the revised version that was most recently reviewed. The two sentences noted below have been edited to address this comment.

We have changed the sentence that had begun on line 196 that had read:

> The EM-GC output shown below relies entirely on simulations constrained to match the HadCRUT5 GMST anomaly (ΔT) record (Morice et al., 2021) over the years 1850–2019.

to now read:

> The EM-GC output shown below relies entirely on simulations that we constrain to match the HadCRUT5 GMST anomaly (ΔT) record (Morice et al., 2021) over the years 1850–2019.

Also, the sentence that began on line 172 that had read:

> A hallmark of this approach is to span a wide range of possible time series of ERF_{AER}, as well as a model parameter λ_{Σ} that represents the sum of all climate feedbacks, retaining for further analysis the members of this ensemble that satisfy three goodness-of-fit constraints, to the: 1) 170-year GMST record; 2) GMST record over the past 8 decades (formally, 1940 to 2019); 3) the ocean heat content record that begins in 1955.

has been changed to now read:

> A hallmark of this approach is to span a wide range of possible time series of

> ERF_{AER}, as well as a model parameter λ_{Σ} that represents the sum of all climate feedbacks, retaining for further analysis the members of this ensemble that satisfy three goodness-of-fit constraints, to the observed: 1) 170-year GMST record; 2) GMST record over the past 8 decades (formally, 1940 to 2019); 3) ocean heat content record that begins in 1955.

Use of AR5 Simulations: In section 3.2 Probabilistic Forecast of Future Warming, there is some discussion comparing results with those from AR5. Were any AR5 simulations specifically run with the EM−GC model? If so, it would be helpful for the authors to provide details about which AR5 simulations were conducted. For example, was the CMIP5 historical dataset used instead of the CMIP6 historical GHG concentrations from ScenarioMIP?

> No, we did not run any AR5 simulations in this manuscript. AR5 relied on the Representative Concentration Pathway (RCP) scenarios, which are the basis of our earlier Mascioli et al. (2012) and Hope et al. (2017) studies. In the current manuscript, we make important reference to CMIP5 ESM output by the inclusion of the trapezoid placed on Figure 3, which originates from Figure 11.25b of AR5. This trapezoid represents the likely range of warming over the time period 2016 to 2035, provided by the authors of Chapter 11 of AR5, in recognition of the CMIP5 "hot model" problem. We have not made any changes to the paper, in response to this comment, because: a) we had addressed CMIP5 output extensively in our prior studies; b) the most important aspect of this prior analysis is used in our paper, via the trapezoid.

This is more of a suggestion than a required revision, but I believe the manuscript could benefit from a clearer articulation of the motivation and novel aspect of this research. While the study certainly contributes to the existing literature, particularly on the AR5 and AR6 results (which is sufficient justification in my opinion), it currently lacks a concise statement explaining how this specific analysis advances the field. Adding to the substantial research on AR5 and AR6 outcomes is a valuable addition to the publication record but highlighting why it was necessary to conduct this analysis using EM−GC could elevate the paper and its overall impact.

> We have added the following new sentence, to the end of the second paragraph of the Introduction:
>
> > The projections of GMST shown throughout this paper are motivated by quantifying the likelihood of achieving either the target (1.5 ^{0}C) or upper limit (2.0 ^{0}C) of global warming under the PA, for various assumptions regarding GHGs and aerosols.

Specific Comments

L25: Is large ensemble generation a unique capability of EM−GC? The phrasing of this sentence sort of implies that this capability is unique to only EM−GC.

> We have not changed this sentence in the abstract, which reads:

The hallmark of our RCM is the ability to conduct large (here, 160,000 member) ensemble forecasts of global warming.

because this sentence highlights an important aspect of the research shown in the paper. The ability to run ensemble members of this size, while not unique to the EM-GC, is certainly not possible to accomplish with any ESM. The text does not state “unique”. We have left the text as is because the ability to run truly large ensembles is a “unique” feature of the class of models known as “Reduced Complexity Models”, as detailed in Table 1 of Nicholls et al. (2021).

L41: The authors might want to consider a different reference here, while Meinshausen et al., 2020 demonstrates the capability it is not a model documentation manuscript. Perhaps a lit review of RCMs like the RCMIP manuscripts would be more appropriate here. Or consider referencing additional RCMs and using the respective documentation manuscripts instead of only citing a specific model application manuscript.

We have added citations to the RCM papers Nicholls et al., 2020 and Nichols et al., 2021 to line 41.

L54: The text beginning here is a tad unclear and I would like to make sure I am understanding this section correctly - so in the previous study (McBride 2021) EM–GC results were from the SSP scenarios published between AR5 and AR6. But the EM–GC RF parameter values were from AR5. This manuscript uses the AR6 parameterizations for RF and the AR6 scenarios? So essentially this paper is EM–GC uses different equations/parameters for RF and different scenarios compared to McBride?

Table 1 has been added to summarize the model inputs used for the Baseline and AR6 frameworks.

L90: The text “The second number in the name of the SSP scenario is the target RF at the end of the century in units of W m−2, commonly referred to as the “nameplate RF” (O'Neill et al., 2014)” seems a bit out of place and could probably be removed.

We have decided to retain this sentence, as written in the second paragraph of the “Data and methods” section, because the word “nameplate” appears four more times in the paper. For this sentence in question, we feel a definition is important, even though most of the readers might understand the latter context had this sentence been removed.

L109: The phrase "permafrost melt" is informal, I encourage the authors to use "permafrost thaw" instead.

Change made

L265 - in the text “Two of these metrics quantify how well the modelled GMST anomaly represents the observed temperature anomaly of the atmosphere for the entire training

period (1850– 2019, χ2 ATM) and over the last 80 years (1940 – 2019, χ2RECENT)" is the 1940-2019 temperature observations being counted twice? Is it using the same temperature observations or different ones? Some more details would be helpful here, or a clear justification as to why there are two goodness of fit metrics for temperature anomaly.

L266 - "The χ2RECENT metric is used because without this constraint" what constraint? The χ2 ATM ≤ 2 or the χ2 OCEAN constraint?

Here we are replying to the comments on both L265 and L266.

The sentence that had read:

> The χ^2_{RECENT} metric is used because without this constraint, some solutions with values of $\chi^2_{ATM} \leq 2$ have a visually poor simulation of the rise in GMST over the past 4 to 5 decades, due to the large uncertainty associated with early measurements of ΔT (McBride et al., 2021).

has been changed to read:

> The χ^2_{RECENT} metric is used because without this particular constraint, some solutions with values of $\chi^2_{ATM} \leq 2$ have a visually poor simulation of the observed rise in GMST over the past 4 to 5 decades, due to the large uncertainty associated with early measurements of ΔT, as described in Sect. 2.1 of McBride et al. (2021).

Reviewer 2:

Farago et al. present an assessment of how the updates to concentrations of GHGs, parametrisations of their effective radiative forcings (ERFs), and tropospheric aerosol ERFs, made between the 5th and 6th IPCC assessment reports, influence projected 21st century warming in a semi-empirical reduced complexity model. They show that, collectively, these changes increase the projected warming substantially. This is an impressive and clearly written study. It makes a useful contribution which I suspect will be of wide interest. I support it's publication. I have suggested some minor changes below, all of which I leave to the discretion of the authors whether to implement.

> Thanks for these very kind words. We also sincerely appreciate the constructive comments of this review.

General comments

It would be valuable to quantify the relative contributions to warming of the different changes made between the frameworks. It may be that this is not possible to do without significant extra analysis, but I suspect many readers will be left wanting to know the answer, and if only a qualitative sense of the relative contributions could be given, this would still be valuable. More broadly, I think the Conclusions section could benefit from some more qualitative explanation to guide the reader to the right interpretation of what is driving the main findings. There's a lot in this paper, and the reader (or at least this reader) could have benefitted from some help on this point. I assume that one major contributor to the extra warming is the combination of a larger positive GHG forcing and a more negative aerosol forcing during the recent observed period under the AR6 framework (i.e. more aerosol masking of warming under the AR6 framework, with the same total recent warming), which then leads to more rapid warming as aerosols decline and GHGs rise over the coming decades. Do the authors agree?

> Thanks for this great suggestion. A new paragraph towards the end of Section 3.2, as well as a new Table 4, has been added to address this important point. The new paragraph reads as follows:
>
> > Larger values of ΔT_{2100} are found in the AR6 framework compared to the Baseline (Table 2). As detailed in Table 4, the more aggressive warming within the AR6 framework is due mainly to three factors: 1) stronger cooling over the historical time period by tropospheric aerosols in the AR6 framework relative to Baseline (Fig. S1e); 2) larger future concentrations of CO_2 projected by AR6 compared to the SSP database (Fig. S1a); 3) greater ERF due to CO_2 using AR6 formulae compared to the AR5 formulae. Table 4 shows the change in the median value of $\Delta T_{2100,}$ found for a series of full ensemble model simulations conducted using the AR6 framework, except for replacement of individual model inputs from the Baseline run. The entry labeled "CO_2 PPM" for SSP4−3.4 represents the difference

between a computation of ΔT_{2100} found using AR6-based model inputs (2.10 ^{o}C for SSP4–3.4, as shown in Table 2) and a new median value of ΔT_{2100} found from a simulation that uses the Baseline projection of the atmospheric concentration of CO_2 (dotted line, Fig. S1a) and AR6 values for all other model inputs (in this case, yielding a median value for ΔT_{2100} of 1.986 ^{o}C for SSP4–3.4). Similarly, the entry labeled "CO_2 RF Formula" shows the difference in ΔT_{2100} from the full AR6 simulation compared to a run that uses the ERF formulae for CO_2 from Table 8.SM.1 of AR5. The fact that the single, largest impact on ΔT_{2100} is driven by changes to the ERF of tropospheric aerosols, for both the SSP4–3.4 and SSP2–4.5 scenarios, underscores the importance of reducing the current uncertainty in this quantity, to better constrain future projections of global warming. Furthermore, the most important GHG-related factors affecting our forecasts of ΔT_{2100} are the projections of the future atmospheric concentrations of CO_2 and CH_4, which are notably different within Annex III of AR6 compared to the SSP database. Finally, changes to other inputs between the AR6 and Baseline computations, that are not considered in Table 4 such as the ERF due to tropospheric ozone, halocarbons, and LUC, make small contributions to the differences in the model projections of ΔT_{2100}.

- Given that the changes between AR5 and AR6 frameworks are so fundamental, I wonder if a summary table or figure in the main article in Section 2.1 Model Inputs, to give the reader a quick sense of the most important of these changes, would be useful?

Thanks for this suggestion. A new table, Table 1, has been added to summarize the model inputs used for the Baseline and AR6 frameworks.

Minor comments:

Abstract

- ".. to the state of knowledge that existed" I am not sure that the pre-AR6 framework is synonymous with the existing knowledge. Perhaps it would be better here to make a limited comparison to e.g. 'the AR5 framework' or similar?

We have left the phrase "we compare and contrast these AR6 updates to the state of knowledge that existed prior to the publication of AR6" in the abstract, because we prefer to not define and then use the term "Baseline" within the abstract. This phrase, perhaps imperfect, reflects our best attempt to succinctly and accurately define the term "Baseline" as used in the paper, without using the word "Baseline" in the abstract.

- "the rate of warming is … within the AR6 framework" having read the paper, it is clear what this means, but as a standalone sentence it reads strangely. There is only one observed rate of warming over a historical period. Consider rewording – e.g. "the modelled rate of warming". There is a broader point here that the abstract might benefit from a

statement of the relationship between observations, RCMs, and ESMs, to give the non-expert reader a sense of the motivation for the study (i.e. to provide "data-driven probabilistic" projections, as is noted in the Introduction).

The word "modeled" has been added, as suggested.

We have also added "data-driven" in front of "probabilistic" projections on line 17 of the abstract

1 Introduction

- Line 55 "our": here and later in the paragraph "we" and "our" etc. are used to refer to the McBride et al., 2021 paper. I found this a little confusing as I expect these terms to imply a statement made from the perspective of this study (even if authors are the same on both studies)

The other reviewer also commented on the over-use of "our", which we have addressed.

The sentence on line 55 that had read:

> In the McBride et al. (2021) paper, our projections of GMST

has been changed to read:

> In the McBride et al. (2021) paper, their projections of GMST

2.Model inputs

- Is there any reason not to use "AR5" throughout instead of "Baseline"?

We have decided to retain baseline, rather than "AR5", because the SSP scenarios that define future GHG concentrations were developed after the publication of AR5.

3.2 Probabilistic forecast on future warming

- Line 500: this point, that given our updated knowledge, the highest emissions scenario with a good chance of achieving 2°C has essentially moved down one SSP, from 3.4 to 2.6, is I think the single most significant finding of the study, and will be of broad interest. I would suggest adding a sentence stating this to the **abstract**, even if that meant removing some of the other quantitative details given there.

Great point; much thanks. We have added the following last sentence to the abstract:

> For society to have high confidence in achieving at least the upper limit of 2 ^{0}C warming of the PA, the radiative forcing of climate due to GHGs must be placed close to the SSP1−2.6 pathway over the coming decades.