Response to Referee Comments on "Constraining net long term climate feedback from satellite observed internal variability possible by mid 2030s"

We appreciate the referees' constructive feedbacks. Below, we provide our detailed responses and the modifications made in response to their comments and suggestions.

- **Referee 2.**

The authors investigate the relationship between internal variability feedbacks and forced climate feedbacks across a range of CMIP6 models. They explore the feasibility of using this relationship, along with observed internal variability feedback estimates derived using CERES, to establish an emergent constraint on Equilibrium Climate Sensitivity (ECS). The authors find a robust relationship between internal variability and forced feedbacks, particularly for shortwave and longwave components, whereas the relationship seen for the net feedback is weaker. To address this, the authors explore how the relationship strengthens over longer time periods (50 years). To provide an estimated constraint on ECS, the authors combine satellite observations with a reanalysis dataset to provide an observed estimate of internal variability feedbacks. However, in order to provide a constraint based on observations only, continuous satellite observations until the mid-2030s would be necessary.

I found this paper enjoyable to read and I believe it would be a useful addition to the literature in this field. I have one major comment and a number of minor comments.

Major Comment:

The authors suggest that the relationship between internal variability feedback and forced feedback could be used as an emergent constraint on ECS. However, the utility of this relationship could be challenged were there to be a bias in modelled estimates of internal variability feedbacks compared to observations.

Armour et al. (2024) investigated the relationship between historical temperature trends and ECS, showing that this relationship was not suitable for use as an emergent constraint due to a known systematic bias in modelled estimates of historical temperature trends.

They show that since coupled climate models do not simulate observed temperature patterns, modelled historical warming was systematically warmer compared to observations.

Would the results of Armour et al. (2024) impact the conclusions reached in this analysis? Would this suggest that there may be a systematic bias between observed and modelled internal variability feedbacks due to different SST patterns?

For example, if AOGCMs are biased in their simulation of SSTs patterns and feedbacks due to internal variability, then it is plausible that this biases the emergent constraint with long term feedbacks proposed here.

Either way, I would expect some discussion on the limitations and potential for biases in the results.

The referee is asking whether the conclusions of Armour et al. (2024) —that coupled climate models, due to their inability to accurately simulate observed temperature patterns, render the relationship between modeled historical warming and ECS unsuitable as an emergent constraint— suggest that there may be a systematic bias between observed and modelled internal variability feedbacks due to different SST patterns.

As noted by Referee 1, these biases in warming patterns are more likely to lead to discrepancies between modeled and real-world forced climate feedbacks, rather than being directly tied to internal variability feedbacks. Such discrepancies may reflect differences in the relationship between internal variability and forced climate feedbacks in models versus the real world, which the emergent constraint methodology assumes do not exist. This is a caveat in our study and has now been explicitly highlighted in the abstract, discussion, and notes of caution in the revised manuscript.

Minor Comments:

- Line 2-3 – Aren't the changes in top-of-atmosphere flux in response to surface temperature changes how we often define feedbacks in general (not just internal variability feedbacks). Could the definition of internal variability feedbacks and forced climate feedbacks be more explicitly defined?

  You are correct in noting that changes in top-of-atmosphere flux in response to surface temperature changes are often how we define feedbacks in general. In our approach, the distinction between internal variability and forced climate feedbacks lies in the drivers of those surface temperature changes. For internal variability, changes in TOA fluxes result from natural variations in surface temperature due to internal variability. In contrast, for forced climate feedbacks, the changes in surface temperature

are driven by radiative external forcings. We have revised the sentece in line 2 to clarify this distinction. The phrase "top-of-the-atmosphere flux variations in response to surface temperature fluctuations" has been corrected to "top-of-the-atmosphere flux variations in response to natural surface temperature fluctuations"

- Line 50 (and in the introduction in general) – I think it might be beneficial to formally define somewhere in the introduction what is meant by forced feedback and internal variability feedback.

  The definition of forced climate feedbacks is now provided in line 21, while the definition of internal variability feedbacks is given in line 46. Details on their calculation begin in line 84.

- Line 69 – Could the historical and amip experiments used be more clearly defined.

  We have now defined the historical CMIP and AMIP experiments in the manuscript line 70.

- Line 82 – "TOA flux anomalies" – Could the authors write "R" given they have shortened surface temperature anomalies to "T".

  We have now replaced "TOA flux anomalies" with R.

- Line 84 – Could the authors expand on how the historical members are detrended?

  We removed the linear trend from each historical member individually to detrend the data. A more detailed description of this methodology has been added to the "Materials and Methods" section.

- Line 85 – "various" – Could the authors be a bit more specific?

  We have added a more detailed description of this methodology to the "Materials and Methods" section and the word "various has been removed.

- Line 91 – "the transformed datasets" – It isn't completely clear what datasets are being referred to here.

  We refer to the transformed temperature time series for each model realization. A more detailed description of this methodology has been added to the "Materials and Methods" section.

- Line 81 Paragraph – I think in general, if this paragraph could be rewritten to be much more thorough with the details it would help. Further questions I am left with are... How is F calculated in order to calculate the lambda in the historical experiments? Is the idea that the internal variability feedbacks have no forcing effecting them? And if so, how is this achieved given there is forcing over the historical period? Is this why the timeseries have been detrended? If the detrending has had a linear trend removed, given the historical temperature and flux timeseries' are often highly non-linear, is this definitely an appropriate method? Would that leave an imprint of the forcing still in the timeseries?

I think this paragraph was the main unclear bit in the paper for me.

I was also curious about whether this is a different method compared to that used in Uribe et al. 2022? (i.e. the use of GLS). Has that contributed to the slight change in the results?

Thank you for your insightful questions. We address each point below:

- **How is F calculated to determine lambda in the historical experiments?: In our methodology, lambda is calculated as the regression slope between R and T. Thus, F is not explicitly needed for its calculation.**
- **Is the idea that the internal variability feedbacks have no forcing affecting them?: The concept of internal variability feedbacks pertains to the response of TOA fluxes to natural temperature variations driven by internal climate variability, rather than external forcing. Therefore, these feedbacks are considered independent of external radiative forcing.**
- **How is this achieved given there is forcing over the historical period?: We achieve this by removing the linear trend from the historical temperature and TOA fluxes time series to isolate internal variability. This detrending helps to eliminate the forcing effects in the relatively short time periods used in the study, allowing us to focus on short-term variability.**
- **If detrending has removed a linear trend, and given that historical temperature and flux time series are often non-linear, is this an appropriate method?: We understand the concern regarding non-linearity. Although radiative forcing is indeed non-linear over the extended period, the linear detrending approach is commonly used for short time scales. This method is supported by previous studies for similar periods (e.g., Uribe, Bender, and Mauritsen 2022; Chao and Andrew E. Dessler 2021; A. E. Dessler 2013; Lutsko et al. 2021; Mauritsen and Stevens 2015; Zhou et al. 2015).**
- **Would this leave an imprint of the forcing still in the time series?: In cases with large temporal datasets, there might be residual effects of forcing. However, in our study, we found that the leading EOF mode of CERES TOA variability (Figure 1a in this document) is highly correlated with**

the El Niño 3.4 temperatue anomalies (Figure 1b in this document), indicating that the variability in the data series is predominantly driven by internal variability rather than residual forcing effects.**

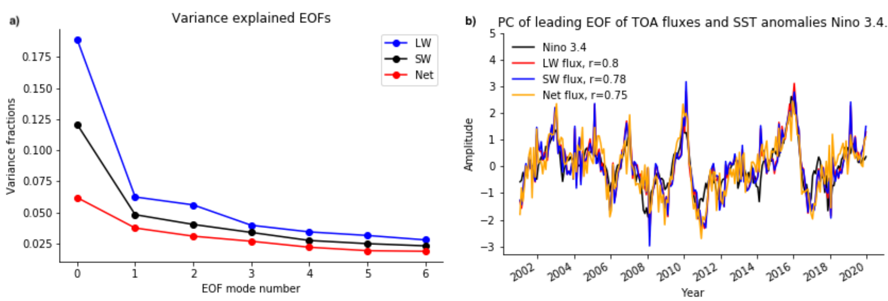

Figure 1: a) Displays the dominant modes of variability in the CERES TOA flux dataset. b) Shows the principal component of the leading EOF mode of CERES TOA and the Niño region 3.4 temperature anomalies.

- **I was also curious about whether this is a different method compared to that used in Uribe et al. 2022? (i.e. the use of GLS). Has that contributed to the slight change in the results?: In our study, we analyze up to 5 realizations of historical ensemble members to capture a broader range of climate outcomes and obtain more robust estimates of internal variability feedbacks. This approach requires addressing autocorrelation in the temperature time series within each ensemble member, which we handle using GLS. Conversely, Uribe et al. (2022) used only a single ensemble member per model, so GLS was not needed in their analysis. Therefore, the differences in results between our study and Uribe et al. (2022) are primarily due to the use of multiple ensemble members rather than differences in the estimation methodology.**

- Line 125 – This isn't completely clear what has been done here. How were the datasets randomly permuted?

  We acknowledge that our initial description of the method lacked detailed explanation, a point also noted by Referee 1. Here is our response: In summary, our method estimates the likelihood of obtaining a correlation as high as, or higher than, the observed correlation between internal variability and forced climate feedbacks in climate models. Specifically, we randomly permuted the feedback datasets, disrupting the correspondence between models for internal variability and forced climate feedbacks (for

example, by pairing the internal variability feedback from one model with the forced climate feedback from another). We then recalculated the correlation coefficient using this shuffled data. This procedure was repeated 10^5 times, generating a null distribution of correlation coefficients that reflects the range of values expected if no real relationship exists. Finally, we compared the observed correlation to this null distribution to estimate how often a correlation of equal or greater magnitude could arise by chance, providing a p-value as a measure of statistical significance. We have added more details to this method in line 135 in the manuscript to clarify the procedure. This methodology has been more clearly incorporated in now line 135.

- Figure 3 – All text in Figure 3 is very small and smaller details are rather hard to read. Could the authors increase the font size and perhaps consider re-arranging the subplot to help make it more readable.

  We have rearranged now Figure 2 vertically to improve its readability.

- Figures in general – Although the other figures are not as hard to read as Figure 3, some may benefit from larger text. Figure 5 is an example of this. Figure 5c also has the legend partially obscured by some of the lines in the plot.

  We have resized now figure 4 for better readability.

Very Minor Comments:

- Line 1 – ", crucial climate regulators," - I would remove this or restructure the sentence as it is a little unclear whether the authors are describing the act of observing climate feedbacks or the climate feedbacks themselves. Obviously it is the latter, but I kept reading it as crucial climate regulators is not feasible. – This is a very very minor comment, but I think it would help make the first line of the abstract more punchy.

  We have reworded the sentence for better clarity.

- Line 85 – "However" – This doesn't seem like quite the right word, "however" introduces a statement that contrasts or seems to contradict something. Would saying instead "Here, it is crucial…". Again, a very minor comment, but it stood out to me.

  The sentence has been reworded.

- Line 91 – "a OLS" –¿ "an OLS".

  The spelling mistake has been corrected.

References

Chao, Li-Wei and Andrew E. Dessler (2021). "An Assessment of Climate Feedbacks in Observations and Climate Models Using Different Energy Balance Frameworks". In: *Journal of Climate*, pp. 1–30. DOI: 10.1175/JCLI-D-21-0226.1.

Dessler, A. E. (2013). "Observations of Climate Feedbacks over 2000–10 and Comparisons to Climate Models". In: *Journal of Climate* 26.1, pp. 333–342. DOI: 10.1175/JCLI-D-11-00640.1.

Lutsko, Nicholas J. et al. (2021). "Emergent Constraints on Regional Cloud Feedbacks". In: *Geophysical Research Letters* 48.10, e2021GL092934. DOI: https://doi.org/10.1029/2021GL092934.

Mauritsen, Thorsten and Bjorn Stevens (2015). "Missing iris effect as a possible cause of muted hydrological change and high climate sensitivity in models". In: *Nature Geoscience* 8.5, pp. 346–351. DOI: 10.1038/ngeo2414.

Uribe, Alejandro, Frida A.-M. Bender, and Thorsten Mauritsen (2022). "Observed and CMIP6 Modeled Internal Variability Feedbacks and Their Relation to Forced Climate Feedbacks". In: *Geophysical Research Letters* 49.24. DOI: https://doi.org/10.1029/2022GL100075.

Zhou, Chen et al. (2015). "The relationship between interannual and long-term cloud feedbacks". In: *Geophysical Research Letters* 42.23, pp. 10, 463–10, 469. DOI: https://doi.org/10.1002/2015GL066698.