EGUsphere-2025-4899 Response Document

Below is a list of the responses to all comments made to the paper.

Reviewer 1

Main Comments:

- In numerous places results are compared to IPCC AR6 as if that was some gold standard or independent estimate. But that is nothing but an assessment of literature available at the time of the cut-off date, and this literature includes studies of the kind used here (Jimenez, Nijsse, Tokarska), along with estimates based on longer term instrumental record warming. I would strongly suggest to go through the paper and carefully revise all these instances.
  - Response: Rightly or wrongly, the IPCC assessments are seen as the gold statement by many climate scientists, so we think it's important that we compare our results to the latest IPCC assessment (IPCC AR6) which contains an agglomeration of all available literature at the time. However, we do recognise that the assessment is informed by many of the studies already referenced (e.g Mauritsen 2019, Nijsse 2020, Tokarska 2020, Sherwood 2020) and so we have reduced the number of references (e.g line 60). We also note that this is the first IPCC report that gives a TCR best estimate that is constrained by additional constraints (paleoclimate, recent observations), rather than using the raw model range.

- The method of smoothing down weighs the warm 2023 and 2024 years. I would like to see results without any smoothing, comparing up to 2019 to an analysis including 2024. This is important since the point of the paper is that these years do not change the results substantially, but the chosen analysis method implicitly down weighs these very years. It would also be good to try using the much simpler delta method, and years of Jimenez-de-la-Cuesta who stopped in 2005 to avoid stitching errors with scenario runs. Ideally, since the inclusion of the recent warm years is the central theme of the study, it would be good to completely take apart how much of the change originates from the inclusion of 2023 and 2024, how much comes from the updated data sources (e.g. HadCRUT5), and if possible how much is due to methods.
  - Response: The supplemental material includes the result of how the TCR prediction (both central estimate and likely range) changes when a 5 year smoothing period, terminating on the central year of 2022 is used. Unsurprisingly there is more variability in the 5-year plot, however the TCR prediction remains reasonably steady after the year 2000. The figure shows a very similar result if the (central) end year is 2005, for both 5 and 11 year mean plots. This would not have had any stitching issues as the 2005 point in the 5-year plot would have used the 2003-2007 average and the 2005 point in the 11-year plot would have used 2000-2010 average. Since CMIP6 historical runs stop in 2014, none of the SSP runs would incur into this result. We have taken this piece of feedback into account, by referencing more frequently throughout the paper how the result is not dependent on the smoothing window. Additionally, all figures in the main paper have now been updated to show a more direct comparison with and without the inclusion of the years 2020-2024.

Minor Comments:

- 3, also pattern effects matter for TCR
  - Revision: “A key metric is the Transient Climate Response (TCR), which incorporates aspects of Equilibrium Climate Sensitivity (ECS), ocean heat uptake and pattern effects, and is closely correlated with historical global warming in Earth System Models (ESMs).”
- 16, as per IPCC AR6 the surface air temperature is used
  - Revision: I changed all instances of GMST to GMSAT (this is true - in the IPCC it is defined with the word air in it) and updated the definition of ECS and TCR.
- 23, "Both metrics can be computed"
  - Revision: Updated this.
- 32, not often recognised, but 'very likely' means greater than 90 percent confidence.
  - Revision: Updated this to say "at most" 90% confidence”. In future instances where I say ‘very likely’, I do specify exactly 90% - though this convention is used in the IPCC.
- 35-36, this statement misrepresents the cited paper.
  - Revision: I updated this to make the wording more similar to the abstract of the Tokarska paper. "As expected, models that simulate stronger warming trends in recent decades have higher TCR values, and project larger future temperature increases (Tokarska, 2020).”
- 68, I would have called emissions and estimated aerosol forcing stable for decades from the 1970s onwards. Perhaps be more specific about the timing of the decline, and also provide references.
  - Revision: Was more specific about the history of aerosol emissions in this paragraph - saying they rose from the mid-century, peaking in the 1970s before declining from 1980 onwards. Also referenced Samset et. al 2018 which also went through this historical breakdown.
- 74, Paleoclimate estimates are also observational.
  - Revision: Changed this sentence to "observations constrained by recent decades", which differentiates it from Paleoclimate studies.
- 78, consider deleting "Oceanic Nino Index of +2.0 K occurring"
  - Deleted it.
- 78-79, not sure how this statement is relevant, could be deleted.
  - Deleted it.
- 80-82, This is a problematic statement. It is unclear who is meant to perceive, who is applying pressure, and who in 'climate science' is under pressure. Either things should be clarified with documented data and/or scientific references, or I would suggest deleting.
  - Revision: Changed this sentence to "Nevertheless, the occurrence of these warm years, in addition to the larger TCR estimates from CMIP6 models in comparison to CMIP5, has raised uncertainty regarding whether the Earth's climate sensitivity could in fact be greater than current estimates."
- Table 1 caption contains a double negation.
  - Updated this.

- 107-108, The sentence seems misplaced as it opens up a new theme, or perhaps starting a new paragraph would be good for readability?
  - Added a paragraph break.
- 190, please replace 'significant' with something else, e.g. 'excessive' or 'stronger than observed'. The word 'significant' is usually used in statistical tests.
  - Revision: Used “stronger than observed”
- Table 2, perhaps include results from Jiménez-de-la-Cuesta (2019).
  - Revision: Added this in.
- Table 2, The row "IPCC AR6 (Unconstrained)" is misleading. These values taken from Chapter 7 are constrained by observations, including information from the above mentioned studies.
  - Revision Changed this to "Chapter 7 Assessment"
- 201-203, this is unsurprising since these two studies used more or less the same models and methods.
  - Revision: Mentioned that all studies used very similar methodology.
- 225, the stated difference is incorrect, according to Table 2 the difference is 0.21 K.
  - Revision: It turns out that Tokarska's paper had a TCR value for 1981-2014 and a TCR value for 1981-2017. I put the latter in as it encompassed more years, so the difference is still around ~0.1K (this is likely a result of the El-Nino event in 2015/16 which produced 2 warm years). I also compared with Jiménez and Mauritsen 2019 in addition to Nijsse 2020 and Tokarska 2020.
- 233, "it was not used"
  - Revision: Updated it to this.
- 235-236, I would strongly suggest to delete this statement as it is unsubstantiated speculation.
  - Revision: Deleted this.
- 239-240, This is not a standardised initialisation, the initial conditions are simply some state taken from the piControl and it happens to be the first.
  - Revision: Updated this to "a prescribed set".
- 246-251, this text could be deleted as it contains nothing concrete.
  - Revision: Removed the speculative sentence.
- 252-258, the purpose/conclusion of this paragraph is unclear, it would benefit from rewriting and focusing on physics. What you expect here is rather irrelevant.
  - Revision: We removed the final sentence of this paragraph which was arguably speculative “In the case of TCR, we expect this effect to be relatively small on the decadal timescales we are interested in here.”
- 259-260, rather unsurprising.
  - Added ‘unsurprisingly’ so that it’s emphasised.

Reviewer 2

- From Table 3 to Fig. 5, you present the results of emergent constraints using data from 1975 to 2024. However, readers would like to know how these results change (or remain stable) when including data from recent record-breaking warm years. Even if the differences are small, this information is valuable to the community. Therefore, please compare the results of “EC 1975-2019” and “EC 1975-2024” in these tables and figures.
  - Revision: All figures and tables have now been updated in the paper to show a more direct comparison when the emergent constraint is performed using years up until 2019, and that when using years up until 2024. For maximal consistency, and as indicated in all figure captions, the dashed lines represent the result up until 2019, and the solid lines represent the result up until 2024. We hope that this change more directly illustrates that adding these 5 years on has only made a marginal difference to both the TCR central estimate as well as the temperature projections into the future. Please find attached an example of how the figures (in this case, figure 5) has been updated.
- L174 “The uncertainty associated with internal variability is estimated as the standard deviation of all points in time series (with respect to the smoothed series), divided by $\sqrt{n}$, where n is the number of points in the smoothing window.”: This method excludes the decadal-scale internal variability. You can include it by using pi-Control simulation data.
  - Response: To assess internal variability associated with longer, decadal timescale, we have performed additional analysis of how the variability of the control runs of the ESMs varies in comparison to the detrended historical runs. This has now become an additional section (2.2.2) in the main paper, and we include figures in the supplementary material. We found that the non-de trended piControl runs - on average - had slightly larger variability than the detrended historical runs, although there was variation amongst the models. The model with the largest discrepancy was the CNRM-CM6 model, which showed 2.7 times the amount of internal variability in the un-detrended control run to the detrended historical run. To test how sensitive our emergent constraint on TCR is to this additional factor, we calculated the TCR with this additional factor of 2.7 applied to the observational uncertainty associated with internal variability. We found that the TCR very likely range did not significantly broaden (only by ~0.09 K either side), and this was due to the uncertainty being largely dominated by the emergent relationship, rather than the uncertainty in the observational data.
- Fig.2: The horizontal axis of Fig. 2a should compare the global mean temperature trends of 1975-2019 and 1975-2024 in the K/10-year unit. Please show correlation values. In Fig. 2b, “EC 1980-2014” and “EC 1980-2019” should be “EC 1975-2019” and “EC 1975-2024”.
  - Revision: I have rescaled the axis so that it now shows the decadal warming rate on the x-axis. I have also changed the caption so that it states the whole period (rather than the centred years), e.g 1975-2024 rather than 1980 - 2019.
- Fig. 3: Please denote correlation values in each panel.
  - Revision: These have now been added to the figure.