

The study introduces a number of important concepts and a potentially very powerful framework for policy-relevant analysis. I also enjoyed the hierarchical approach with the use of a Budyko-Seller model to illustrate the methodology.

On the other hand, there are some aspects of the analysis and discussion that I believe need to be altered or addressed in greater detail. I would recommend the paper for publication in ESD provided the authors implement the appropriate changes.

Major Comments:

1. My main concern is the applicability of the methodology to CMIP5 (and soon CMIP6) simulations. It is unrealistic to expect 200-member ensembles of CMIP5 models to become the norm anytime soon. What conclusions could the authors actually draw from the simulations currently stored in the CMIP5 archive or those already being run/planned for CMIP6? I am looking for a detailed, critical evaluation of the real applicability of the methodology. Just adding a one-line statement will not be sufficient.
2. While reading the paper, I often found the wording a bit awkward (a non-exhaustive list: ll. 53 and following, ll. 67 (Giving → Given?), l. 143 (non-viable **for**?), ll. 176 and following etc.). I would encourage the authors to carefully revise the text as part of their revisions.
3. In order to ensure an unbiased approach to the study, I did not read the comments of the other Reviewers until after having noted down my own. However, I generally agree with the feedback provided in the first two reviews, and note that the authors could have addressed more extensively some of the points that were raised. In fact, in many places the authors seem to have performed a minimal-effort revision. For example, Reviewer #1 correctly questions the assumption of Gaussianity of the GMST. In my opinion, a brief mention of this limitation in the discussion section is insufficient. What do instrumental/reanalysis datasets tell us about the GMST distribution in the recent/present climate? Since, if the authors themselves or others will (hopefully) apply this method in the future, GMST might not be the only variable of interest. How easily can the method be adapted to other types of distributions? Similarly, Reviewer #2 highlights a potentially important improvement that could be applied to the methodology (Page 15, l. 384), but the authors seem to have taken little notice of this. This could also help to address my comment #1 above, which I think is one of the crucial limitations of the method.

Minor points:

- 1) There are lots of acronyms in the abstract, which make it unnecessarily unwieldy. While it is perfectly fine to use acronyms in the paper, I would encourage the authors to limit their use in the abstract.
- 2) l. 41 Explain what the Burning Embers diagram is or remove the reference altogether.
- 3) l. 46 Define pCO₂.
- 4) The methodology is spread across the different sections of the paper. While in some cases this works well in terms of logical flow, it makes it very difficult for the reader to obtain an overview of what type of analysis is being carried out. For example, the authors could consider moving some of the more technical parts of Section 4.1 to the Methodology Section.
- 5) The description of the PLASIM model is very dry. A few more details would help the reader to better understand the type of data being used here without having to refer to other papers.