Response to reviewers

Thank you for reading our manuscript and taking the time and effort to leave a community comment to strengthen the manuscript. We have responded to each comment in full and outlined the changes we will make to the manuscript to address your comments in this document. Our responses are in black font in response to review comments in blue, and where we quote new text, this is in italics.

Reviewer comments

First: Kudos to the authors for pulling together a great review of existing reduced complexity model structures, components, and other factors. This will be very much appreciated by those of us who work with reduced complexity models.

Thank you for your kind words. The intention of this manuscript, which evolved from a literature review made by the lead author as part of his PhD, was precisely to serve as a guide to newcomers to the field as well as SCM users/developers. Hearing that this objective may have been fulfilled is very encouraging. We respond to each comment in turn here and in the revised manuscript.

Second: I have a couple minor comments:

Minor comment one regards this statement:

"Moreover, ESMs operate at finer scales, benefitting local and regional analysis, although downscalling approaches to generate regional climate emulators from SCMs have also been explored (Beusch et al., 2020; Mitchell, 2003; Mathison et al., 2025)"

I think this sentence, as written, is slightly misleading because most (all?) regional climate emulators depend on ESM data. All three cited approaches - Beusch et al (MESMER), Mitchell (pattern scaling), and Mathison et al. (PRIME) rely on ESM output in order to build their databases. I'd also recommend adding Tebaldi et al. (2022) (STITCHES) to your list of regional climate emulators (and which also relies on ESM data). Perhaps if the sentence was reframed as, "Moreover, ESMs operate at finer scales, benefitting local and regional analysis. Pattern scaling approaches that leverage ESM data to generate regional climate emulators, which can then be coupled with reduced complexity models to provide higher resolution functionality, have also been explored (citations)."

We agree that the statements could be misread as it was initially written. We have rewritten it to make sure we acknowledge these downscaling approaches are employing ESM data, and added a reference to STITCHES as suggested. The text now reads:

"Moreover, ESMs operate at finer scales, benefitting local and regional analysis, although downscaling approaches that leverage ESM data to generate regional climate emulators have also been explored, partially mitigating this limitation (Beusch et al., 2020; Mitchell, 2003; Mathison et al., 2025; Tebaldi et al., 2022; Sandstad et al., 2025). These regional emulators can then be coupled with SCMs to provide higher resolution functionality."

As a more general statement (and less relevant to this manuscript), I would like the ESM community to lean into the role of understanding the patterns of responses to forcing by running stylized scenarios (e.g., isolating GHG, aerosol, and land use change forcing components) rather than running scenarios produced by IAMs, which add complexity to comparing different

scenarios as they differ on all 3 forcing types. Then, these various pattern scaling approaches could be coupled to reduced complexity models to project more realistic scenarios as well as important probabilistic assessments.

This is a valid point, although some efforts in this area are underway. We have just submitted a manuscript, RCMIP Phase 3 (https://egusphere.copernicus.org/preprints/2025/egusphere-2025-5775/) that will run idealised experiments including some of the single-forcing experiments that you propose. While the scope of this intercomparison is global (and SCM-focused), we are aware of some efforts to isolate the regional patterns of different forcings, particularly the METEOR model (https://gmd.copernicus.org/articles/18/8269/2025/), which we have also added a citation for. New ESM experiments such as the Regional Aerosol Model Intercomparison Project (RAMIP; https://gmd.copernicus.org/articles/16/4451/2023/) will provide useful regional response data that can be used to calibrate regional single-forcing responses.

Minor comment two is that I would be interested in understanding in what fashion (if at all) box models and IRM approaches differ in terms of behavior. After looking at RCMIP and various other tests, I haven't identified any fundamental differences between IRMs (as a class) and box models (as a class). Perhaps the discussion at lines 390-409 regarding mathematical equivalence is the answer to this question? If that's true, maybe this equivalence could be emphasized even more.

They are indeed equivalent models displaying the same dynamics. We have updated the text (~L421) to make this more explicit:

"While the n-time-constant temperature IRMs and the n-layer temperature models are mathematically equivalent, and could therefore be fundamentally considered the same model describing the same dynamics, the IRM formulation has the advantage of a simple relationship (Geoffroy et al., 2013a) between the parameters in Eq. 15 and two of the most critical and widely discussed quantities in climate science, the Equilibrium Climate Sensitivity (ECS) and the TCR:"

Minor additional notes:

Typo: downscaling, not scalling

Thank you, now corrected in the text.

Figure 10 typo: atmopshere should be atmosphere

Thank you, now corrected in the text. While revising this figure I also detected an additional typo, which said "*greenshouse*" instead of "*greenhouse*" in "*concentration of greenhouse gases in atmosphere*". This is also now corrected.