

We thank the referee for the invested time to assess our manuscript and for pointing out specific issues to improve upon. We reply individually to all specific points below in blue but start with the general intended changes and comments.

We agree that the initial manuscript was unsatisfying in terms of intercomparison. Previously, only the final figure item offered an intercomparison and only for one region without any comparison to an ESM. Every other assessment was based on the output of a single emulator only. Further, without an explicitly stated common protocol, there was no basis for judging how much of the results reflected the emulator choice or the sample of ESMs behind it. Thus, the initial manuscript failed to answer its main question: How much does the choice of regional emulator, and the choice of which ESMs are emulated, affect the computed consequences for emission scenarios of IAMs?

We intend to make four major changes to address this. Later referred to as C1 to C4.

C1: The intercomparison becomes the central element.

Former Fig. 6 (one region, one scenario, three emulators, no ESMs) is removed and replaced by two new figure items built on the common core set of emulations (C3). New Fig 1 shows the distribution (boxplots) for each AR6 region of annual mean temperatures from all three emulators and for the seven CMIP6 models behind the common core set (C3). New Fig 2 focusses on the upper percentiles (exceedance frequencies and absolute values) of annual temperature, for the same regions and again all three emulators & CMIP6 ESMs and additionally contrasts each emulator on the common core ESM set with the emulator output based on its full ESM set. The new Fig 2 responds directly to RC1's point that agreement at the center of the distribution is of limited interest, and its second element responds to RC1's point on lines 318–320 by measuring the effect of ESM selection.

C2: The use cases are recomputed to integrate all three emulators.

The regional emulator use cases become part of the intercomparison. Former Figs. 3, 4 and 5 (the scenario-to-climate outcome space, exposed population, and exposed GDP against mitigation cost) were previously computed from MESMER output alone. They will be recomputed and shown for each of the three emulators now based on upper percentile “tas” instead of “TXx”. This aims to bridge the two major inquiries of the manuscript. (i) The differences in climate alone of the emulated tas established in the two new figures 1 & 2 (C1), and (ii) what those differences mean for derived climatic consequences of emission scenarios like GDP or population exposure.

C3: The shared core protocol will be described explicitly.

Our experiment did run a shared core: Identical SCM-derived global-mean temperature trajectories, the same five probability levels of the same SCM ensembles, a common scenario set, a core set of seven emulated ESMs, with annual mean near-surface air temperature delivered by all three emulators and all output harmonised to a common 2.5° grid. The initial submitted manuscript never said so explicitly and only stated the overall emulated variables, ESMs, etc. The revision describes the core protocol explicitly and separately from the extended set of ESMs and variables each emulator contributed beyond it.

C4: The manuscript is restructured so that the intercomparison comes first.

The revised section order that follows from this is: 1 Introduction; 2 Design and protocol (C3); 3 Intercomparison of emulated regional climate (C1, new Figs. 1–2, replacing former Fig. 6); 4 What the choice of emulator and ESM set implies for scenario assessment (C3, former Figs. 3–

5, all three emulators); 5 Towards endogenous links (former Sect. 3.1, shortened); 6 Discussion and conclusions; 7 Methods.

RC1

This paper is both interesting and unsatisfying. Interesting in the angles that it brings, and unsatisfying because it is trying to do multiple things and does not do those well enough. Here are my specific points

- If this is a paper that discusses the potential of coupling IAMs with regional information, then it should at the very least discuss existing efforts (e.g., <https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2024MS004806> and other references from that paper) and discuss the actual potential for IAMs to integrate the regional information. Just saying that regional information will be available does not mean that the IAMs will be able to use it, regardless of how good the generated information is. There has to be much more discussion from the IAM side on how coupling would work and whether it might make a difference. Lines 115-116 seem to indicate that there is not much that can be done at this point.
- If this is a paper that discusses a MIP, then it needs to be much more extensive about the reason for the MIP and the actual protocol to be followed. The three models that are being discussed here do **not** follow a protocol, they all use a different set of models and/or variables. How are we supposed to understand the differences in results, which is a key element to a MIP? Furthermore, how do we know that any of those are generating good results? What is the method that will be used to assess that? What is the expectation in terms of generating statistics consistent (or not) with the underlying GCM? For example, I would have expected that a protocol would follow a benchmarking, similar to what is described in this paper (<https://arxiv.org/abs/2606.29172>)

On the former point:

We agree that more recent literature on coupling efforts needs to be taken into account and intend to tackle this in both introduction and discussion. Recent studies have begun to use emulated climate information to model impacts within an IAM context (Tebaldi et al., 2026b; Rouhette et al., 2026), and others approach the same problem through damage functions (van Vuuren et al., 2026), but none of them closes the loop literally, by hard or soft coupling while the IAM produces the scenario. We will also not pursue to couple the emulated climate consequences produced here to one or more IAMs in this manuscript. Our intended contribution here is narrower: Regional emulators make it compelling to compute climatic consequences of emission scenarios without any ESM simulations for as many scenarios as an IAM produces. This manuscript aims to provide a first assessment of how strongly those climatic consequences and derived socioeconomic consequences depend on which emulator is used and on which ESMs it emulates.

We intend to make this intended contribution much clearer in writing in the introduction and discussion and further with the complete overhaul of the structure (intercomparison first) and the main assessments (C1 to C4).

On the latter point:

The emulation effort behind the initial manuscript did follow a protocol detailing minimum provided standards to allow comparison. We failed to explicitly describe them as such. All three

emulators were driven with identical SCM-derived global-mean temperature trajectories, taken at the same five probability levels of the same SCM ensembles, across the same scenario set, for a core set of seven emulated ESMs. Annual mean near-surface air temperature is provided by all three, and all output was harmonised to a common 2.5° grid. We will explicitly describe the core set that all emulators provided and compare it to the full provided emulation (C3).

The new Figs 1 and 2 are built on that common core set, so that a difference between emulators is attributable to the emulator and not to the sample of ESMs behind it. Further, both figures allow a comparison to the emulated ESMs themselves. In addition, the new Fig 2 relaxes the constraint on the core ESM set deliberately and shows each emulator on its full ESM set beside its core-set result, so that the effect of ESM selection can be assessed.

On whether the emulators generate good results, both new figures place the emulated distributions alongside the CMIP6 simulations of the emulated ESMs themselves, so that the emulated spread can be read against the simulated spread it is meant to represent. We will not attempt a full validation of each specific emulator against CMIP6 as each emulator comes with its own published evaluation (Beusch et al., 2020; Tebaldi et al., 2022; Quilcaille et al., 2023; Mathison et al., 2025).

The intercomparison will become the main theme described throughout the manuscript. Directly in the intercomparison of climatic outputs between emulators and CMIP6 in new figures 1 & 2 and also in an integrated way joined with socioeconomic consequences in updated figures 3 to 5 which will include data from all emulators based on the core set (C1 & C2)

Minor points;

Line 29: the fact that all three distributions agree on the center of distribution is probably of limited interest since the main goal of the discussed downscaling is to provide better extreme values. Additional points should be made.

We concur that the agreement of the emulators around the center was not the best choice given the manuscript is aimed towards deriving socioeconomic consequences meaningful to emission scenarios. We intend to produce two new intercomparison assessments (C1) that replace the former subsection that produced this statement here. Especially, the planned figure 2 will directly provide results for the more extreme percentiles produced by the emulators and how they compare to ESM results.

Line 78: back to my main point #1, why isn't there a tighter connection with the ScenarioMIP effort?

We intend to provide direct comparison with the ESM results of the previous ScenarioMIP cycle leading to CMIP6 (using the SSP-RCP simulations) in new assessments and figure items described in C1. A coordinated emulator effort for the current ScenarioMIP cycle leading to CMIP7 is underway but will not be made part of this manuscript.

Section 3.1: why focus on fire-weather? This is already a very difficult process to represent on the ESM side, why is it thought that this approach would be representing those processes better?

We agree that fire weather is a complex metric and further is only produced by one emulator in this manuscript. We argue fire weather is a key candidate of a meaningful climatic consequence

of the scenarios computed by IAMs. Possibly the main goal of an IAM is the computation of a consistent emission scenario. Forests are a major emission source and sink. Hence, results towards better constraining this major source/sink is of interest to the main goal of the IAM modelling endeavor.

We intend to add to this section that the emulation of any fire weather metric cannot improve our process understanding of fire weather and cannot represent these processes any better than an ESM. If at all it carries forward the biases the emulated ESM has towards the complex metric. We also intend to work out more clearly the potential role of fire weather to IAM modelling and their key results, consistent emission scenarios.

Lines 166-168: how would that happen?

We intend to add that the described potential “..exploration of alternative scenarios through spatial constraints or penalty implementations” hinges on the IAMs integrating information within their own (cost-) optimization processes. We intend to further describe what we specifically mean by both mentioned options. “Spatial constraints” could be integrated in the form of a regional mask applied within the IAM excluding heavily fire weather prone regions. “Penalty implementations” could lower the cost incentive of establishing new forests with the purpose of mitigation in the emission scenario. (e.g. by lowering the carbon price that incentivizes the establishment of new forest in an IAM. E.g. see Kreidenweis et al. 2016, DOI 10.1088/1748-9326/11/8/085001).

Lines 318-320: this is most likely just a reflection of the model selection, is it not?

We agree that model selection could’ve played a significant role here. As described in C1 above, we intend to remove this assessment and replace it with two new intercomparison assessments. Especially, new figure 2 and its results will specifically look at the role that ESM selection has on the outcome of the emulation.

Lines 347-349: I am not sure it demonstrates it. I think it shows there is potential, but I would want to see how GCAM/REMIND (since those are explicitly discussed in the paper, Figure 5) are planning to include this in their projections

We agree that we can only show the potential here as we do not intend to “close the loop” within this manuscript by producing updated emission scenarios with one or more IAMs within this manuscript. As stated in our response to comments on lines 166-168, we intend to make it more explicit how a potential implementation could look like with the help of two examples. One being the spatial constraint using a regional mask to exclude fire weather prone regions in the IAM itself. The other targeting the cost optimization of an IAM directly through a cost penalty assigned to the establishment of mitigation oriented forest establishment.