

# Response to reviewers

Thank you to the reviewer for your detailed review of our 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 italic.

## Reviewer general comments

Simple Climate Models (SCMs) make up a critical component of the climate model hierarchy and have been used for decades for climate assessment. Although many SCMs with varying levels of complexity exist, there have been few efforts to clarify differences in model structure. In this work, the authors review the 14 SCMs participating in the reduced-complexity model intercomparison project (RCMIP), organizing them by increasing complexity and creating a guide for developers and users.

Overall, this is a strong manuscript with a clear contribution in providing a comprehensive overview of some of the most widely used SCMs. The methodological choices of both which SCMs to include and what components to focus on (e.g. choosing structure over performance) are clear and sufficient. I have one major comment about the stated goals/conclusions of the manuscript. The rest of my questions/suggestions to improve the quality of the manuscript are relatively minor.

Thanks for your comments, we respond to each comment in turn here and in the revised manuscript.

### **Specific Comments:**

The paper aims to support informed use of SCMs, e.g. 1571-1572: "... while also informing about the implications of selecting one model over another." While the detailed descriptions achieve this, the discussion could be strengthened by adding a paragraph that informs practical guidance for model selection. For example, you could briefly summarize which models are best suited for specific applications based on their features. This type of synthesis would make the paper's thorough exploration of SCMs more accessible to the user community.

We agree that a paragraph briefly summarising the nature and focus of the different models reviewed in the paper would be valuable for the casual reader that does not necessarily want to read the detailed descriptions of all the models. However, the problem with assigning a "good-for" label for each model is that it may be clear for certain SCMs (e.g., OSCAR for the carbon cycle, GREB for grid-based process resolving), but there are others without clear areas of focus. This is why originally the paper did not include such a discussion. However, prompted by this comment (and a similar one from RC2), we wrote a synthesis in the discussion section that divided SCMs into "generalist" models, without any strong focus, and "specialist" SCMs that do possess such as focus. This allowed us to point out such specialist focus when it clearly exists, while avoiding assigning it when it is not present or it is not as clear. Given the nature of the review, which only examines the structure of the models without delving into the performance/accuracy, we think this is a suitable method to present this information to the user community. The discussion now includes the following text (which precedes the more detailed model summary already present in the previous iteration of the text):

*“The aim of this review was to provide clarity on the current SCM landscape by identifying the processes represented by each model, their respective implementations, and the commonalities shared across different models. This was achieved by reviewing the suite of SCMs participating in RCMIP, detailing their components and development history. Ultimately, we hope this text serves as a valuable guide for the difficult task of SCM selection. While other considerations such as accuracy, calibration or usability are important when selecting a model, clarifying which processes a model resolves is a critical first step to assess model suitability for a particular use. Consequently, we offer here a brief summary of the model descriptions presented in section 4, first classifying SCMs in two broad families based on design philosophy, and then summarising the commonalities and differences of the models included in this review. Finally, we conclude with a brief discussion of the limitations of our review.*

*When selecting an SCM, it is important to recognise that different models were developed with different intended applications and design philosophies. Broadly, SCMs can be grouped into what might be termed “specialist” and “generalist” models. Specialist models are developed around clearly defined objectives or processes. Examples include AR5-IR, which provides a minimal framework to estimate warming from CO<sub>2</sub> concentrations; EMGC, which explicitly represents natural variability; ESMICON, designed within a system dynamics framework; GREB, whose spatio-temporal resolution and process representation are designed to support a physical understanding of climate and its teaching; OSCAR, which focuses on the carbon cycle and LULCC disturbances; and WASP, with an emphasis on ocean dynamics. In contrast, generalist SCMs can be viewed as models developed without prioritising any particular component or objective beyond providing climate simulations. Note, however, that the design philosophy should be evaluated across the whole model lifetime, as this distinction is likely violated in the short-term (e.g., FaIR was initially developed as an extension of AR5-IR consisting of a small set of equations to produce warming estimates, arguably qualifying as an “specialist” SCM, but has evolved considerably since then). The most representative example of a generalist SCM is MAGICC, as the oldest SCM in this review with an extended history of usage to generate climate projections. Other models in this category arguably include ACC2, CICERO-SCM, Hector, FaIR, MCE, and SCM4OPT. Reflecting their broad scope, several generalist models (ACC2, Hector, MAGICC, SCM4OPT) have been coupled as climate modules within IAMs, while others (FaIR, MAGICC) have been widely used across multiple IPCC assessment reports. This specialist-generalist split can be a useful first step to evaluate which SCM to use, particularly if a given specialist design philosophy aligns with the requirements of the user. However, a more detailed understanding of the processes and methods used by different SCMs is likely still required after this first step, which we offer below.”*

*In a similar vein, there are points in the manuscript that explicitly address model advantages, but as a developer/user, I'd also love to see disadvantages/failure modes of these different models. Is there a consistent/rigorous way to define a failure mode of an SCM? If so, why have you chosen not to discuss them?*

We could not find any clear, objective way to discuss model disadvantages or model failures by solely looking at model structure and implementations. Model advantages are usually considerably easier to discuss because different schemes are commonly implemented to address certain limitations in previous models. As a result, advantages are usually explicitly stated, while limitations are more difficult to identify. We think model disadvantages or failures are more easily identified when comparing model performance and dynamics, hence better accomplished under the RCMIP efforts. There is currently an ongoing third phase of RCMIP

(<https://doi.org/10.5194/egusphere-2025-5775>) that will inform about potential model disadvantages in the medium future. We have included a paragraph in the (new) limitation section to explicitly explain this:

*“The discussion of model differences in this review has focused on the processes represented in each SCM and the advantages those representations provide, rather than on potential disadvantages. This was a deliberate choice: we found no consistent or objective method for assessing model drawbacks solely from their technical specifications. Different methodological approaches, most notably RCMIP efforts (Nicholls et al., 2020, 2021; Romero-Prieto et al., 2025) with their systemic evaluation and benchmarking of model outputs, are better suited to address questions of model disadvantages.”*

Minor:

1. The historical overview of the field is comprehensive but ends somewhat abruptly around 2020, and doesn't acknowledge the rapid rise in statistical/machine learning-based approaches. Since that isn't the focus of the paper, it's not a major issue, but a comprehensive review such as this should acknowledge the current state of the field.

Thank you for pointing this omission. We agree that some more discussion around the recent innovations in machine learning is warranted. Towards the end of the historical overview section, we have added the following paragraph:

*In the early 2020s, following the strong demand for climate emulation established in the previous decade and fuelled by the rapid popularisation of artificial intelligence (AI), data-driven emulators rose sharply in prominence (Watson-Parris et al., 2021, 2022; Watt-Meyer et al., 2023; Bassetti et al., 2024). These models typically provide regional rather than global outputs and follow methodological approaches that differ substantially from those of SCMs, using machine learning and AI algorithms. For these reasons, they fall outside the scope of the present review. Readers interested in data-driven climate emulation are referred to the comprehensive review by Tebaldi et al. (2025).*

2. In a similar vein, climate emulators are referenced on line 23 in the introduction, but aren't addressed again. For both this and the previous point consider the question: how would someone coming from the ML/emulator literature connect to your manuscript?

Thank you for raising this point. We agree that the distinction between SCMs and the broader emulator family is important, as it is often neglected in the literature and may lead some readers to confusion. We have updated the text to frame SCMs within the broader emulator family, and mentioned statistical models as the other big family of emulators. The text now reads:

*“Their emulation capability positions SCMs within the broader family of climate emulators, at least when they are operated in an emulator configuration. Although these two terms are sometimes used interchangeably, the “emulator” label strictly encompasses a greater number of models than SCMs, as it applies to any model designed to approximate the output of another model. In particular, the emulator label also includes any machine learning and AI approach able to mimic the output of more complex models through statistical learning. The boundary between SCMs and data-driven emulators is not always sharp—some SCMs employ techniques such as*

*impulse-response functions that could be viewed as statistical. Nevertheless, SCMs generally adopt a more mechanistic emulation strategy, grounded in physical reasoning and parametrisation.”*

3. 352: What is meant by ‘a higher number of layers provides the model with... an enhanced capability to reproduce non-linearities’? Unless I’m misunderstanding Equation 11 (please correct me if this is the case!), this is still a linear model, so what type of non-linearities do you mean?

Thank you for pointing this out. We agree that Eq. 11 still describes a linear model, regardless of the number of layers. The point we wanted to raise here was the enhanced ability to describe complex dynamics in the transient response, which we have now amended the text with:

*“A higher number of layers increases the number of distinct timescales in the system (see  $\tau_i$  in Eqs. 12 and 13), and therefore increases the complexity the model is able to display in its transient response. The equilibrium state, however, remains independent from the number of layers, as Eq. 11 reduces to  $T1 = F/k1$  when  $T1 = T2 = \dots = Tn$ .”*

4. 366-367: “... they (IRFs) are not derived from fundamental physical principles but through mathematical approximations of the relationships we aim to model.” I’m not sure what is meant by this statement, as response functions are, by definition, the exact solution to the type of systems defined by Equations 9-11. You explicitly state this in lines 390-392 as well. The second statement implies they’re only approximations to the same extent that 9-11 are approximations as well. Can you be more precise in what you mean by the former statement?

We agree that EBM (Eq 9-11) and IRFs are both equivalent approximations to the heat dynamics in. The point we wanted to raise here was that the EBM framework is usually derived applying physical arguments to heat transport, while IRFs are a broader mathematical framework that allows an approximation of any non-linear system, independent of physical arguments or the heat transport phenomenon. The paragraph now reads as follows, which hopefully clarifies this point:

*“In the preceding sections, methods were explored that aimed to achieve this emulation through the development of computationally efficient approximations of the Earth system maintaining physical intuition. IRMs (Joos and Bruno, 1996) take a different route; they are not derived from fundamental physical principles but rather, offer a mathematical framework able to approximate the dynamics of any non-linear system through empirical parameter tuning.*

5. 418 (and elsewhere): the term ‘complexity’ is used with respect to models frequently, but without a rigorous definition. You seem to imply complexity is proportional to the number of physical processes represented, but can you clarify this against e.g. computational complexity?

We agree this clarification is warranted. We have added the following text in the introduction (~L78):

*Throughout this review, “complexity” refers to the conceptual or process-level complexity of a model—i.e., the number and intricacy of physical processes it represents—rather than to other notions such as computational complexity.*

6. Figure 1: I don’t follow what ‘None’ means just from looking at the figure, consider clarifying this in the caption.

We have added the following text to the caption to clarify the meaning of “None”: *“No public implementation was found for AR5-IR, so no specific programming language was associated with this model (None).”*

7. Figure 5: Unclear what is meant by ‘Earth’ from just the figure, consider clarifying this in the figure or caption.

We have added the following text to the caption to clarify the nature of the “Earth” box: *“The “Earth” box in Hector represents long-term carbon storage, including fossil fuels and carbon capture.”*

8. You state on 250 that “...SARF should be assumed whenever radiative forcing is mentioned...”, this is probably worth highlighting in the caption for Table 2 for readers who are looking primarily at the figures/tables to summarize the review.

Thank you for this very useful comment which has forced us to think more deeply about what models are actually representing in terms of their radiative forcing inputs. In essence, where models are calculating a radiative forcing value derived from concentrations or emissions and passing this onto a temperature module, they are implicitly using ERF, since this is the measure that is most closely related to long-term temperature response (e.g. IPCC AR5 WG1 Chapter 8; AR6 WG1 Chapter 7). However, they commonly use expressions that compute SARF instead. We have therefore changed our discussion of radiative forcing here, and updated Table 2 caption as suggested:

*“However, consideration of these nuances between forcing categories is often inconsistent in SCMs. Models may employ SARF expressions to estimate forcing without any further modifications, effectively treating that quantity as ERF. Generally, unless otherwise stated, ERF should be assumed whenever radiative forcing is mentioned in this document, although the underlying formula may have originally been intended to approximate SARF.”*

And in Table 2:

*“Generally, ERF should be assumed, although the underlying mathematical expressions used by the models to compute this quantity may have been originally intended to approximate SARF. More details can be found in section 3.2 and the individual publications.”*

#### Technical Corrections:

1. 16: ‘system’ should be capitalized

Thank you. Now corrected in the text.

2. 37: Typo - ‘downscaling’

Thank you. Now corrected in the text.

3. The signposting is very clear throughout, but inconsistent at times. E.g. The GREB signposting (4.6.X) doesn’t follow the same format as the other models.

1. Some places 4.X.1 is labeled model ‘specification’ and others it’s ‘description’

Thank you. Now changed EM-GC “*model specification*” to “*model description*” for consistency.

4. 241: Missing a second set of closing parentheses after LULCC? Also should the first letters be capitalized when the acronym is spelled out explicitly?

The parenthesis after LULCC are correct (black carbon is part of the albedo changes). However, while checking the paragraph we did detect a missing parentheses after “for a full list” for minor GHG. We have also capitalised the LULCC acronym components.

5. 342: The ‘co2’ in the ‘1pctCO2’ experiment is typically capitalized. Consider italicizing experiment names as well to make it clear they’re separate from the models.

Thank you. Now corrected in the text. We also italicised 1pctCO2 in L342 and R01 in ~1275

6. 680: Typo - ‘continuos’.

Thank you. Now corrected in the text.

7. 833: Same citation listed twice.

Thank you. Now corrected in the text.

8. 941: Units are typically not italicized (should *ppm* be ppm?)

Thank you. Now corrected in the text.

9. Ensure consistent hyphenation of your acronyms, there’s at least one instance of “OML IRM” (no hyphen) on 1019

Thank you. Now corrected in the text to OML-IRM in every instance.

10. 1511: Typo - ‘spacially’ (not sure if this is a valid alternate spelling, but you spell it ‘spatially’ on 1127)

Thank you. Now corrected in the text.

11. 1558: Missing word - “... a decision was made (to) exclude...”

Thank you. Now corrected in the text.

12. 1590: Typo - ‘reviued’

Thank you. Now corrected in the text.

## Other changes

Following a request from the authors, in section 4.9.1 (~L1110) we modified the phrasing regarding MAGICC’s vegetation regrowth parameter. The text now reads:

“To better approximate real-world dynamics where land-use changes usually result in persistent alterations to carbon stocks, the MAGICC carbon cycle module uses a1095 regrowth fraction parameter to adjust the turnover time and thereby allow for partial regrowth (Meinshausen et al., 2011a)”

We also changed incorrectly attributed citations to

Meinshausen, M., Wigley, T. M. L., and Raper, S. C. B.: Emulating atmosphere-ocean and carbon cycle models with a simpler model, MAG-2050 ICC6 – Part 2: Applications, Atmospheric Chemistry and Physics, 11, 1457–1471, <https://doi.org/10.5194/acp-11-1457-2011>, publisher: Copernicus GmbH, 2011c

To the relevant publication:

Meinshausen, M., Raper, S. C. B., and Wigley, T. M. L.: Emulating coupled atmosphere-ocean and carbon cycle models with a simpler model, MAGICC6 – Part 1: Model description and calibration, Atmospheric Chemistry and Physics, 11, 1417–1456, <https://doi.org/10.5194/acp-2045> 11-1417-2011, 2011a.