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 comments

I personally enjoyed reading this comprehensive and ambitious review on SCMs. I believe it will serve as a valuable cornerstone for the field, particularly for newcomers from different disciplines. I especially appreciated the synthesis presented in Tables 1–4 and Figures 1–2, as well as the historical perspective provided on both the general development of SCMs and the evolution of individual models. The authors have reviewed an impressive amount of material with great thoroughness, and I really appreciate this effort. The balance between high-level discussion and technical detail is particularly well thought of – the level of detail is sufficient to provide clear context without overwhelming the reader. I only have a few clarifications/suggestions that I'm outlining here below:

Thanks for your kind words and comments. 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.

1. L37: "Downscalling" might be a typo

Thank you, it was indeed a typo. Now corrected in the text.

2. L98: "rather than approximations"; this is a bit nitpicky, but I would argue that also SCMs are driven by fundamental laws of physics (e.g. mass balance, energy balance), except for perhaps purely regression-based models. The difference is that loss rates for chemical species and radiative transfer for heat are highly parameterized in SCMs (first-order kinetics, linear forcing-feedback for heat) but better resolved in ESMs.

We agree. We have slightly reframed the text to focus instead on the methodological differences: ESMs aiming to resolve processes explicitly (as much as feasible) while SCMs aim to produce accurate parametrisations that emulate the effects of those processes. The text now reads:

"This progress culminated in the development of the first AOGCM by Manabe and Bryan (1969), marking a pivotal shift towards physically-complex climate models aiming to explicitly resolve processes, rather than relying on highly-parametrised approximations."

3. L305: I would stress here that lambda is arguably the most important parameter in this formulation, and that is a strong control on climate sensitivity (the central quantity in climate science). I think that is mentioned later, but I wonder if the casual reader would think at this stage that lambda is basically constrained by 4\sigma*T_0^3 ~ 5 W m-2 K-1, which is actually pretty far from the real feedback (~1 W m-2 K-1 for a climate sensitivity of ~3.5K). In your formulation, that means that tuning k is a really important step for any SCM. In other words,

there are lot of things that happen in SCMs but perhaps the most consequential parameter to set in an EBM is lambda (and I would probably call out here that k would need to take into account all the important climate feedbacks such as water vapor, lapse rate, clouds, albedo… and that all of that complexity is swept under the number assigned to lambda).

We agree with the review that this is a critical quantity in any SCM, and that it would be beneficial to point this out explicitly in the text. We have modified the paper to read as follows:

"… While seemingly simple, Eq. 7 is arguably the single most important parametrisation in most SCMs, as it controls the model's temperature response for a given forcing. The k parameter (or, indirectly, the λ parameter) abstracts away the great complexity of the climate system and its numerous feedbacks effects (e.g., albedo change, aerosol interactions, etc.). In most instances, this is a constant model parameter that can be tuned to emulate results from other, more complex ESMs. However, more complex formulations are possible, like MAGICC's (Sect. 4.9.3) time-varying, and WASP's (Sect. 4.5.3) forcing-agent-specific time-varying λ parameters."

4. L352: not sure if I follow the "reproduce non-linearities" part. Even with n-layers Equation (11) is still a linear system of ODEs. I would also call out here that adding n boxes for the ocean does not alter the equilibrium solution, which is still F/k1 for all temperatures, because at equilibrium T1=T2=…=Tn. In other words, having n boxes allow for extra flexibility in setting the time scales of response (the "transient"), but not the final surface temperature equilibrium, which is set by forcing F and feedback k1 (using the notation of Equation 11)

Thank you for raising this, particularly the insightful point about the difference between transient and equilibrium dynamics. 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, along with the remark regarding the immutability of the equilibrium solution:

"A higher number of layers increases the number of distinct timescales in the system (see τi 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/κ1 when T1 = T2 = … = Tn."

5. Table 4 and beyond: there are a lot of github links in Table 4, and other URLs across the paper that point to model websites. I see why this is very useful for the reader interested in the code, but I worry that in the relative near future (e.g. >5 years) some of those links might be broken since they are not "permanent" repositories. Since the paper is typically a stand alone contribution that lives in eternity (hopefully :)), I wonder if the temporary http links should be added as a separate supplement that says "current links" or something along those lines. On the other hand I do see the value of having the links there in the near-term; I just wanted to flag this as a potential issue thinking longer-term.

Thank you for our comment. While we agree that these links will eventually all be broken, we think that the usefulness from a user/reader perspective to have them easily findable outweighs the eventual redundancy of having a list of broken links. However, to partially mitigate this eventual redundancy, we have added a mention in the table caption to the Zenodo archive associated with the paper that stores a permanent copy of all open-source models reviewed in the text: *"A* *permanent archive of the publicly-available models reviewed in this work can be found in the* Code availability *section".*

6. I really enjoyed the summary info provided by Table 1-4 and Figure 1-2, and I appreciate the authors' efforts to summarize a large amount of literature so effectively. I wonder if maybe having an additional column/table with (1) a typical use case where the SCM has been used and (2) a specific aspect of the model that distinguishes it from others could be useful for the reader to identify straight away what model they might be interested in. Or maybe a combination of (1) and (2). For example for OSCAR, that "peculiar" thing that distinguishes it from other models would be strong focus on LULCC and carbon cycle, and if a reader is interested in studying LULCC effects they would immediately know that OSCAR is probably the first thing to look at. For MAGICC that would be the long history and the strong use in IPCC. For GREB that would be the focus on "understanding" the climate system and university teaching. Just some food for thought – Not sure if that would play out well with all the models so feel free to disregard if not.

We agree that a brief summary concerning 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 RC1), 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 a 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. We didn't put this into a table format because we felt that would convey too much weight to arguably one of the most subjective sections of the review, which may not necessarily align with the opinions of model developers. We have tried to be as free from value judgements as possible in this paper, and we preferred to handle this opinion in text form in the discussion section to allow for a more nuanced treatment. The paper now includes the following text in the "Discussion" section (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 texts 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 CO2 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."*

7. L785: I would point out that GREB has the sensible/latent fluxes as explicit terms because they are looking at the surface energy balance, in contrast with all other models that are looking at the energy balance from the top-of-atmosphere (hence they only have radiative fluxes F and \lambda*T and no turbulent fluxes). In other words, the "flavor" of GREB's EBM is quite different from the typical TOA EBMs that most SCMs use, despite still balancing energy fluxes.

Thank you for your interesting remark, which we had overlooked in the text. We agree that this is relevant for the reader wishing to understand the how GREB works and how it differs from other models, so we have added the following text:

"This EBM possesses two main peculiarities that set it apart from other SCMs in this review. First, this EBM focuses on the surface energy balance, as opposed to other SCMs that focus on the top-of-atmosphere energy balance. This allows GREB to include explicitly turbulent heat fluxes like the sensible and latent heat fluxes"

8. L832: I think Hartin (2015) is cited twice, might be a typo

Thank you. Now corrected in the text.

9. L1015: by "this publication" do you mean Meinshausen's or your own article? Maybe that should be clarified since it could be interpreted as both.

We meant the Meinshausen publication. Now modified to "that publication" instead of "this publication" for clarity.

10. L1472: Not sure if I follow this part (not familiar with ESMICON) – "tracking the distribution of heat across the system" and "heat inventories" sound like an energy balance model to me (i.e., calculate fluxes in and out of predefined boxes…). I would also add that EBMs also

include all the listed feedbacks (increased OLR, increase water vapor, clouds, etc.. they just do it through a lambda parameter). Can this idea of "heat inventory" be clarified if it is completely different than traditional EBMs? In this framework, how is a CO_2 perturbation added if not through extra energy (radiative forcing)?

Thank you for raising this, as the text may have not been clear enough on this regard. Although using a different framework and vocabulary, we agree that the description of the temperature module in ESMICON effectively works as an EBM. The key distinction between ESMICON and other SCMs is the absence of a concept of "radiative forcing", with the various processes such as GHG-induced warming or ice-volume having a direct impact on several feedback loops that ultimately affect the heat distribution in the system. We have updated the text to make this clearer:

"While not using explicitly an EBM framework to determine temperature anomalies, ESMICON uses instead an analogous system dynamics framework to determine those anomalies based on the distribution of heat across the system. In particular, it uses the heat inventories from the atmosphere and surface, along with the associated heat capacities, to estimate GMST. The key distinction between the more common EBM and ESMICON's temperature module is the absence of an explicit1500 radiative forcing concept and the parametrisation of feedbacks through a λ parameter, as in Eq. 8. Instead, state variables in the system, such as CO_2 concentrations and ice volume, impact directly various feedback loops driving the model's dynamics, and, ultimately, the heat distribution."

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.