Earth Syst. Dynam. Discuss.,
https://doi.org/10.5194/esd-2018-63-RC2, 2018

Earth System Dynamics
Discussions

Interactive comment on "Evaluating Climate Emulation: Unit Testing of Simple Climate Models" *by* Adria K. Schwarber et al.

Anonymous Referee #2

Received and published: 23 December 2018

SCMs are routinely used to emulate state of the art GCMs, and generally display reasonable (though not perfect) agreement when tuned specifically to do so. The authors themselves cite several papers relating to this which discuss strengths and weaknesses of such emulation. While of course SCMs can also be integrated with standard (default) parameter values to provide some guidance as to how the climate system may behave, these simulation will not encapsulate our uncertainty in the best parameter values to use. Furthermore, such simulations will depend greatly on how the default parameter values were chosen, which may differ between SCMs. Given that the GCMs disagree substantially amongst themselves, I do not understand the purpose of this paper in comparing the outputs of standard SCM instances to themselves and GCM output. It is inevitable that these will not match closely when the SCM parame-

ters are set to standard values, and I do not think it is straightforward to attribute such differences to structural limitations of the SCMs without first checking that they cannot be explained by parameter choices. Of course in the simplest of cases one might show that a complex curve output by a sophisticated SCM/GCM simply cannot be explained by a very simple parametric form, but even here it would be appropriate to explore how close a fit could be obtained.

One could reasonably compare SCM responses amongst themselves when tuned to each other or to some common target (either observational or GCM-based). However, this has not been performed here. While in some experiments the sensitivity parameter has been set to a common value of 3, other model parameters appear to differ between the SCMS and were apparently set to standard values which were probably chosen by the SCM authors for a variety of reasons. Thus it is not possible to determine how much of the differences in response are due to model structure, and how much is the result of using different parameter values/tuning strategies.

I would also question whether the relatively unrealistic abrupt tests are a useful diagnostic tool for the model behaviour. While I accept it can be interesting to characterise the response to idealised forcing scenarios, it may be that the differences are much less significant when more realistic scenarios are applied, and the authors acknowledge this point in their conclusions.

Thus, this analysis does not sufficiently advance our understanding of the behaviour of SCMs, and I am sorry to say that I cannot support publication of this manuscript in ESD.

As a minor comment, the "unit testing" terminology seems inappropriate, the test here is rather more comprehensive than such a term usually implies, and furthermore there does not appear to be any clear criteria for success or failure.

Interactive comment on Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2018-63, 2018.