Earth Syst. Dynam. Discuss.,
https://doi.org/10.5194/esd-2019-77-RC2, 2020

Earth System Dynamics Discussions

Interactive comment on "Relating Climate Sensitivity Indices to projection uncertainty" *by* Benjamin Sanderson

Anonymous Referee #2

Received and published: 23 January 2020

The manuscript illustrates the main differences between metrics that are commonly used to evaluate the response to GHG forcing in climate models. Namely, these are the effective climate sensitivity (EffCS), the transient climate response at CO2 quadrupling (T140), transient climate response at CO2 doubling (TCR), the temperature change after 140 years from CO2 quadrupling (A140). A simple impulse-response model is introducing, separating the response into a fast and a slow component, and whose model parameters are a posterior evaluated through minimization of the cost function with respect to an observational-based dataset. This model is used to consider to what extent the mentioned metrics are able to explain the response to a business-as-usual (RCP8.5) and mitigation (RCP2.6) scenario. It is found that different metrics are able to explain the response to different forcings, and that the simple model that is

here proposed provides different results, compared to state-of-the-art climate models from CMIP5 Project. It is argued that the biases affecting model energy conservation ultimately affect the different explaining capability of the CMIP5 models.

General comment:

Overall, I think that the manuscript is well written, the issue has a great scientific relevance, and the arguments here shown provide significant advancement to the discussion on the topic. Thus, I appreciate that the author addresses them critically, emphasizing that their adoption is conditioned to the problem that one needs to focus on. This is in line with previous works having evidenced the limitations of these metrics for the study of the climate response, especially from a modelling perspective.

I am a bit skeptical about the effectiveness of the impulse-response model, given that it is a purely linear context. The addition of the noise+drift, though, is convincing in explaining part of the discrepancy between the simple model and CMIP5 outputs. The arguments about the applicability of the metrics are thus promising also in a "real-world" context (using the notation adopted by the author), although with some caveats. For this reason, I think it is important that the author puts more emphasis on the nature of the impulse-response model, in the framework of linear response theory (LRT) and Hasselmann-type response (see my specific comments), and evidences its limits.

I think some improvements can be made in terms of how the methodology and results are described. It would be useful to have the "Methods" section in the main part of the manuscript, instead of as an appendix. The notation is not always consistent across the methods. The MCMC procedure should be explicitly described, not only by mentioning the original reference. Some figure captions lack important details.

Specific comments in the following are meant to clarify my general comments and constrain them to the relative sections of the manuscript. I have spotted several typos; these are addressed in the minor comments.

Specific comments:

- ll. 90-92: I do not have clear how the normalized regression coefficients shown in Figure 1f support this argument. Can the author better clarify it?

- ll. 94-96: I do not understand this sentence for a few reasons. Firstly, is the author referring to any specific forcing, when he says that the rate of change in the forcing is approximately constant? In the case of the mitigation scenario (RCP2.6), this is obviously not the case. Secondly, I do not have clear in mind what the author means by "saturation" of the fast feedback response, and if this refers to the whole period 2000 to 2100 or to the end of the period.

- l. 132: the author suggests that the contributions of the two factors are separately addressed in the following, but, in the end, only the overall effect of the bias is taken into account in the following.

- ll. 161-162: the author seems to imply that "real-world applications" are prone to the existence of drifts. But this is rather a model issue, as the unforced "real-world" climate system should not have any drift.

- l. 166: the author did not specify anywhere else in the text what is the length of the abrupt 4xCO2 simulation. As a consequence, "end" of the simulation does not seem to have a specific meaning.

- Appendix B: according to ESD standards, I think that it would be more appropriate if the Methods section are moved in the main text after the Introduction. Moreover, a description of the data that have been used is lacking, especially for what concerns the observational-based datasets used for model optimization.

- Eqs. B1-B2: the impulse response model here adopted requires using only two timescales. Is it sufficient to describe the response? The FAIR impulse-response model here mentioned includes a set of four simple feedback equations (cfr. Hasselmann et al. 1993) differing on the magnitude of the feedback parameter (i.e. on the

timescale of the response). What happens if one includes more than the two timescales considered in this analysis, given that similar strategies applied to geoengineering scenarios have used, for instance, three exponentials (cfr. Aengenheyster et al. 2018)? This is particularly relevant, as the impulse-response model can always be expressed as an infinite sum of exponential behaviors, differing in their timescale, but the response of the real system rarely has the shape of a discrete number of exponential behaviors combined with each other (e.g. Ragone et al. 2016: Lembo et al. 2019). Also, the adoption of the fast-slow scale implies a separation of scale, that is here inferred "a posterior" through heuristic arguments. Nevertheless, there is no reason, in principle, to assume that a scale separation exists, and this problem traces back to the very foundations of the theory about climate response and forced-free fluctuations dichotomy (Lorenz 1979). One way to deal with that would be to evaluate the memory term (cfr. Ghil and Lucarini 2019). I understand that this might go beyond the scope of this work, but I wonder if the author might comment on that in the manuscript.

- Eq. B3: according to the convolution properties, this operation is by all means equivalent to the application of the Ruelle Response Theory (RRT) (Ruelle 1998a; Ruelle 1998b) when a hypothetical impulse perturbation is applied, allowing for a particularly simple derivation of the linear Green function (cfr. Hasselmann et al. 1993). This has found several applications in the context of climate prediction (cfr. Ragone et al. 2016; Lucarini et al. 2017; Ghil and Lucarini, 2019 for a review), not only constraining to the temperature response, but also to a wide range of climatic variables (e.g. Helwegen et al. 2019; Lembo et al. 2019). These arguments provide a rigorous mathematical framework to the experimental protocol here described.

- Sect. B1.1: I believe that a complete description of the model is here lacking and should be included. Referring to the model settings, in particular, it is not clear to me how the ensemble is generated and how many members are taken into account.

- Table B1: is it possible to have a range for rn as well? Also, where does the fr parameter enter the mode? This goes back to my minor comment about consistent notation.

- l. 226: I think that it is important to notice here that in the forcing scenario 1pctCO2 the CO2 concentration reaches doubling after 70 years, as I presume that this motivates the choice of the 61-80 and 131-150 20-years averages.

- Figure A1: the caption does not contain an explanation of the panel b content. Particularly, the author might want to explain the meaning of the red shading, and the range encompassed by the dotted lines.

- Figure A2: the author does not explain why the choice of a single member from each CMIP model ensemble is reasonable in this context.

- Figure A3: it appears that the distributions of fast-scale parameters are much more similar to a Gaussian distribution, compared to the slow-scale parameters. I am surprised that the author does not refer to that explicitly and comments on it. Could it be an evidence that the scale separation that is a priori assumed for parameter model optimization is such that the fast-scale system approaches a stochastic process, in the context of the response of the system to the impulse forcing? This would be certainly reasonable, in an “Hawkins and Sutton, 2009 context" (signal-to-noise ratio approach), but the author might want to justify it in a more rigorous way.

Minor comments:

- l. 19: in this sentence there is a repetition (“range” and “ranging”). Consider rearranging the sentence;

- l. 31-32; I found this part of the sentence a bit difficult to read. A suggestion might be to replace it with “a complication has arisen due to the fact that EffCS seems to be better correlated than TCR with 21st Century warming from present day levels under a business-as-usual scenario.”

- l. 37: replace “have” with “of”.

- l. 60: remove “to”.

- l. 66: it is not clear whether the author refers here to the Appendix A, Appendix B or both.

- l. 69: replace “and” with “to”.

- l. 91: either a sentence breaking is needed here (after the brackets), or “suggest” has to be replaced by “suggesting”.

- l. 125: Replace “of CMIP5” with “for CMIP5”.

- l. 138: if the “Methods” section is in the appendix, they have to be referred to more appropriately as “Appendix (B)”.

- l. 147: replace “the both” with “both”.

- l. 151: replace “Supplemental” with “Supplementary”.

- l. 165: replace “an” with “that a”.

-Eq. B3: I noticed a potential mismatch in the notation, compared to eq. B1. The author may consider adopting the same notation for the temperature evolution in both equations.

- Figure A1: replace “senstivity” with “sensitivity” in the caption.

Interactive comment on Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2019-77, 2019.