

Earth Syst. Dynam. Discuss., referee comment RC1  
<https://doi.org/10.5194/esd-2020-85-RC1>, 2021  
© Author(s) 2021. This work is distributed under  
the Creative Commons Attribution 4.0 License.



## Comment on esd-2020-85

Anonymous Referee #1

---

Referee comment on "The potential for structural errors in emergent constraints" by Benjamin M. Sanderson et al., Earth Syst. Dynam. Discuss., <https://doi.org/10.5194/esd-2020-85-RC1>, 2021

---

The authors have assessed the application of emergent constraints to estimate uncertainties in unknown climate projections. The recent increase in the application of emergent constraints in Earth Science makes this a timely issue, even more in the last year since the new ensemble of Earth system models (CMIP6) became available. Furthermore, the manuscript is very well written and structured, which made it a pleasure to read. However, the main motivation for the paper is that emergent constraints are used too literally (line 681) and "confusing to policymakers" (line 17). I do not fully share this assumption and I argue below why I think this is not the case. Overall, I have major and minor questions that would need to be resolved before a possible publication.

Major comments:

1) The manuscript reads more like a review and less like a research article. The analysis in this paper includes (a) calculating correlations between different published variables, (b) a bootstrapping approach to exploit previously published emergent constraints, and (c) the exploitation of two differential equations with a random number generator to create many ensemble members.

Taken together, I cannot see how this would be a sufficiently novel concept, idea, tool, or data given that was not previously exploited.

Point (b) somehow tests the robustness of the linear relationship. What is the advantage of this method compared to the published measures of uncertainty (e.g. prediction intervals, see minor comment 1). Already the published results of all shown ECS are not different within their still relatively large uncertainties (Table 4 in Schlund et al. (2020)). So, I am wondering what additional information is obtained by performing this bootstrapping.

Point (c) seems to be a more of a fancy way of saying that the Earth System responds at different timescales and not just one, a well-known concept and used to calculate the

temperature response to radiative forcing changes (Stocker et al., 2013; Otto et al., 2015). It is therefore obvious that the 2-timescale model continues to heat although radiative forcing stabilizes, and the 1-timescale model does not.

Taken these results together, I do not see a (novel) advance here although the Discussion and presentation of the results is very well done. Overall a lot of times 'may' and 'might' are used indicating that this is more speculative. However, I hope that the authors can convince me that I am wrong. If not, I would propose to make it a review or perspective paper.

2) Many points that are discussed here as shortcomings of emergent constraints should, in my opinion, be attributed to the models themselves. Emergent constraints can, by definition, only improve the model output. If the models have structural shortcomings, they exist in the multi-model mean and the constrained results. Emergent constraints can thus 'only' improve the existing model output and cannot go further. The *exchangability* argument (lines 66-76, lines 296-304, lines 412-420) is thus wrongly stated in my opinion.

An emergent constraint can only say: models that tend to simulate a large (small) variable A at present (e.g. extent of the Hadley Cell) also simulate a large (small) increase in variable B. If, and only if, a mechanistic relationship can be given and proven to a sufficient level, one could conclude that if models had rightly (as in the observations) simulated variable A, the model result for variable B would be the following.

The constraint can never overcome model shortcomings or biases that exist across the entire model ensemble and is not designed to do that. The *exchangability* argument would thus only hold if the constrained result would be considered the truth, which it is obviously not. It is just an improved projection, but still based on imperfect models.

As an example, the present temperature and the leaf area index are within the current model ensemble good predictors of future GPP (Schlund et al., 2020). By using the present-day temperature and the leaf index, one can thus show how a model of the same ensemble would likely simulate future GPP if it had the right temperature and leaf index. Nevertheless, if the whole ensemble would be systematically biased and missing out an important process, like nitrogen limitation (lines 296-304), the emergent constraint cannot assess this. The systematic bias is present in the multi-model mean results and the constraint results and thus not primarily an issue of the constrained but the model ensemble. However, it is equally important to mention and assess these uncertainties when presenting multi-model means and constrained results.

Overall, I think the *exchangability* argument is wrongly stated as the models in an emergent constraint were never meant to be exchangeable, but observations are only used to inform the likely best guess of a given model ensemble. Having said this, the question remains how systematic model biases should be accounted for in uncertainty range. When calculating the multi-model mean, the standard deviation is often a measure of uncertainty, but this is not taking into account biases or structural errors. So, one can argue that the given uncertainties for a constrained result are equivalent to the uncertainties calculated by the multi-model standard deviation.

Lines 412-420 make the point very clear. The current knowledge of the Earth System is as good as possible implemented in the Earth system models. Some processes known to be missing or wrongly represented, others are probably missing but not known yet (like the ocean mixing in the 1-timescale model). However, these missing processes are strictly speaking a problem of the models. If we had no knowledge of the ocean's importance

(1-timescale model) and only lambda would be of importance, we had implemented no ocean in our models. Thus, the relationship would resemble the red points in figure 2e,f with model lambdas being different because of different atmospheric model components. By applying the hypothetically 'observed' lambda, we would reduce uncertainties related to the atmospheric model. We would, however, not find the right results because all models are missing the ocean component. Nevertheless, given the assumed hypothetical current state of knowledge (ocean is not important), and the consequent hypothetical models (no ocean), our knowledge of lambda and the emergent constraint would still 'improve' model projections under this assumption. Following this argument in its strict sense, would lead to the conclusion that models cannot be used because important steps might be missing and NOT that emergent constraints cannot be used. The 'assessment of underlying model assumptions' (line 436) should always be done if model output of any type is published, constrained or not constrained.

If I understand the conclusion in lines 676-678 right, the authors argue that CMIP models are a comprehensive representation of the Earth System. Following that line of argument, we should not use them either to make projections of climate change at all? Or under which conditions?

3) The manuscripts title, abstract, Conclusion (and partly Introduction) suggests that emergent constraints across the field of Earth Science are addressed. However, throughout the manuscript it becomes clear that the focus of the manuscript is on emergent constraints of the Equilibrium Climate Sensitivity. In addition, other constraints, such as for the Transient Climate Response and the land carbon cycle, are (briefly) discussed. However, constraints for the ocean, a major part of the Earth System, are not discussed at all. Thus, the title and abstract are highly misleading.

I would suggest to either clearly indicate that the manuscript is on emergent constraints on ECS and only discuss ECS constraints or add examples of ocean constraints and largely expand their exposition in section 5.

Examples for emergent constraints in oceanography would be: Kessler et al. (2016), Kiwatkowski et al. (2017), Goris et al. (2018), Terhaar et al. (2020a), Terhaar et al. (2020b). The list is very likely not sufficient.

Furthermore, the ocean is ignored in questions about the ECS or atmospheric CO<sub>2</sub> (point 6), although the 2 timescale models clearly indicate that the ocean is important.

4) I am not a statistician, but I have strong concerns regarding the application of statistics in this study. First, I disagree that the Sherwood "D" and Cox constraints are correlated (Line 141). A r (if it is r) of 0.31 is a r<sup>2</sup> of less than 0.1. A p-value is not given but I do not expect it to be supportive of a correlation. Even for Lipat and Qu, the r<sup>2</sup> is 'only' 0.33. Please do not use the term correlated if the variables are not statistically correlated. Second, the two constrained ECS do not disagree (Line 142). Sherwood et al. (2014) find an ECS likely at 4°C with 3°C as a lower limit. Cox et al. (2018) report 2.8 ± 0.6 °C. Within the uncertainty ranges, they agree with each other. Lipat et al. (2018) and Qu et al. (2014) do not even give a constrained result for ECS as far as I can see this. But for this argument we can use the Schlund et al. (2020) estimate for the EC from Lipat et al., which is 3.0 ± 0.8 °C and try to read the result for Qu et al. (2014) from the corresponding subpanel, leading to 3.5 ± 0.4°C. These two constrains do also agree.

The following paragraph paragraph and conclusions are thus wrong.

5) The authors often use the argument that emergent constraints might be confusing to policy makers or other people. Furthermore, they say speak about their 'literal interpretation (line 196). I can see no evidence supporting this claim. On the contrary most emergent constraints only give a 'likely' estimate (summarized in table 4 in Schlund et al. (2020)) and even if all ECSs were used to give a best estimate with an uncertainty range, all ECS would agree. Thus, these ECS seem to be used to exclude outliers and not give a narrowly constrained result. Given that they all agree, I do not see the possible confusion.

6) Lines 442-446: The authors claim that the differences in atmospheric CO<sub>2</sub> are caused by the land carbon sink, whereas Hoffmann et al. (2014) clearly state that "Weak ocean carbon uptake in many ESMs contributed to this bias, based on comparisons with observations of ocean and atmospheric anthropogenic carbon inventories." While the land carbon sink is very uncertain, the ocean has been found by Hoffmann et al (2014) to cause the bias. Please correct your paragraph accordingly. I would also argue that the bias-persistence in the too small ocean carbon sink (Kessler et al. (2016), Goris et al. (2018)) is caused by the circulation differences and is persistent over large timescales and thus not overconfident. The whole section should hence be replaced.

#### Minor comments:

1) Figure 1: The panels are too small and impossible to read, especially on the diagonal. I suggest keeping y and x labels with the name of the constraint only at the left column and the bottom line. Furthermore, I cannot understand the added value from the bootstrapping algorithm from the manuscript. Often the uncertainty of the fit is estimated by prediction intervals (Bracegirdle et al. 2012; Nijssse et al. 2020; ...). To which degree and why does the bootstrap method improve the results, or the estimated uncertainties compared to these prediction intervals. If no improvement exist, why would you not just show the published results (Schlund et al. 2020)? And if you recompute them, why not showing the mean estimate + uncertainties + r<sup>2</sup> or something similar. At the moment the subpanels do not allow to assess the mean, the uncertainty or anything else because they are too small.

2) Line 102: Table 3 is mentioned in the text before table 2.

3) Is the bootstrap approach general knowledge? If not, please consider telling the reader how it works or give a reference.

4) Lines 114-121: This is very hard to read, and I am not sure that I understand the message. Could you try to rephrase it and make sure that the reader understands when a combination is appropriate and when not and why?

5) In general: How do you define correlation (Pearson's product-moment coefficient r or something else?)

6) Lines 206 to 209: I do not agree with this statement. Let's assume variable A (present) is correlated to variable B (projection) across a model ensemble and the correlation is mechanistically profound and supported by theory and observations. If variable A is now a

very complex interplay of many processes, it could have a large inter-model spread without a lack of diversity. Thus, the presence of an EC can be a lack of diversity or a complex interplay of different processes. The sentence now is rather misleading.

7) Lines 227-230: You should include Kessler et al (2016) and Goris et al. (2018) here.

8) Lines 250-267: You should add Terhaar et al. (2020a,b) here, although it is not strictly a feedback process but identifying the leading order process that describes the future response.

9) Lines 269-285: You should add Kwiatkowski et al. (2017) here.

10) Line 288-290: I again do not agree with this sentence. A set of models with very complicated assumptions in different processes that govern both related variables, variable A (observable) and variable B (projected), would lead to a large spread in A and B and possible to a good correlation and EC.

11) Equations are not numbered

12) Equation on line 341 is difficult to read (latex problem?)

13) Lines 492-499: What is the added information here? It sounds more speculative than informative.

14) Section 5.3: Your two timescale models are constructed to make just this point. Maybe you could use this here and emphasize hence the importance of the ocean for long-scale warming (ECS) and point out that the difference in the ocean may, according to your model, be responsible for the different long-term temperature trajectories.

15) Lines 537-644: I do not see from which results this conclusion is drawn. Could you please just point me to it? And what other metrics are you referring to here?

16) Lines 554-660: Please cite here the multi-variable approach by Schlund et al. (2020)

## References

Bracegirdle, T. J., & Stephenson, D. B. (2012). On the robustness of emergent constraints used in multimodel climate change projections of Arctic warming. *Journal of Climate*, 26(2), 669-678.

Brient, F. (2020). Reducing uncertainties in climate projections with emergent constraints: Concepts, examples and prospects. *Advances in Atmospheric Sciences*, 37(1), 1-15.

Cox, P. M., Huntingford, C., & Williamson, M. S. (2018). Emergent constraint on equilibrium climate sensitivity from global temperature variability. *Nature*, 553(7688), 319-322.

Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., ... & Williamson, M. S. (2019). Taking climate model evaluation to the next level. *Nature Climate Change*, 9(2), 102-110.

Goris, N., Tjiputra, J. F., Olsen, A., Schwinger, J., Lauvset, S. K., & Jeansson, E. (2018).

Constraining projection-based estimates of the future North Atlantic carbon uptake. *Journal of Climate*, 31(10), 3959-3978.

Hall, A., Cox, P., Huntingford, C., & Klein, S. (2019). Progressing emergent constraints on future climate change. *Nature Climate Change*, 9(4), 269-278.

Hoffman, F. M., Randerson, J. T., Arora, V. K., Bao, Q., Cadule, P., Ji, D., ... & Wu, T. (2014). Causes and implications of persistent atmospheric carbon dioxide biases in Earth System Models. *Journal of Geophysical Research: Biogeosciences*, 119(2), 141-162.

Kessler, A., & Tjiputra, J. (2016). The Southern Ocean as a constraint to reduce uncertainty in future ocean carbon sinks. *Earth System Dynamics*, 7(2), 295-312.

Kwiatkowski, L., Bopp, L., Aumont, O., Ciais, P., Cox, P. M., Laufkötter, C., ... & Séférian, R. (2017). Emergent constraints on projections of declining primary production in the tropical oceans. *Nature Climate Change*, 7(5), 355-358.

Lipat, B. R., Tselioudis, G., Grise, K. M., & Polvani, L. M. (2017). CMIP5 models' shortwave cloud radiative response and climate sensitivity linked to the climatological Hadley cell extent. *Geophysical research letters*, 44(11), 5739-5748.

Nijse, F. J., Cox, P. M., & Williamson, M. S. (2020). Emergent constraints on transient climate response (TCR) and equilibrium climate sensitivity (ECS) from historical warming in CMIP5 and CMIP6 models. *Earth System Dynamics*, 11(3), 737-750.

Otto, F. E., Frame, D. J., Otto, A., & Allen, M. R. (2015). Embracing uncertainty in climate change policy. *Nature Climate Change*, 5(10), 917-920.

Schlund, M., Lauer, A., Gentine, P., Sherwood, S. C., & Eyring, V. (2020a). Emergent constraints on equilibrium climate sensitivity in CMIP5: do they hold for CMIP6?. *Earth System Dynamics*, 11(4), 1233-1258.

Schlund, M., Eyring, V., Campsâ□□Valls, G., Friedlingstein, P., Gentine, P., & Reichstein, M. (2020b). Constraining uncertainty in projected gross primary production with machine learning. *Journal of Geophysical Research: Biogeosciences*, 125(11), e2019JG005619.

Sherwood, S. C., Bony, S., & Dufresne, J. L. (2014). Spread in model climate sensitivity traced to atmospheric convective mixing. *Nature*, 505(7481), 37-42.

Stocker, T. (Ed.). (2013). *Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge university press.

Terhaar, J., Kwiatkowski, L., & Bopp, L. (2020a). Emergent constraint on Arctic Ocean acidification in the twenty-first century. *Nature*, 582(7812), 379-383.

Terhaar, J., Torres, O., Bourgeois, T., & Kwiatkowski, L. (2020b). Arctic Ocean acidification over the 21 st century co-driven by anthropogenic carbon increases and freshening in the CMIP6 model ensemble. *Biogeosciences Discussions*, 1-30.