Earth Syst. Dynam. Discuss.,
https://doi.org/10.5194/esd-2020-49-AC2, 2020

Earth System Dynamics
Discussions

Interactive comment on "Emergent constraints on Equilibrium Climate Sensitivity in CMIP5: do they hold for CMIP6?" *by* Manuel Schlund et al.

Manuel Schlund et al.

manuel.schlund@dlr.de

Received and published: 26 September 2020

Reply to Thorsten Mauritsen (Referee)

Reviewer comments are marked in **bold**, our answers in red.

Review of "Emergent constraints on Equilibrium climate sensitivity in CMIP5: do they hold for CMIP6?" by Manuel Schlund and co-authors.

In this study a series of mostly process-oriented emergent constraints that were developed on earlier model ensembles are applied to the latest CMIP6 ensemble. This is a very welcome attempt and in a broad sense testing scientific repro-

Printer-friendly version

Discussion paper

ducibility. My major concern is with the main conclusions drawn, or perhaps not drawn, from the results. The fact that estimated ECS based on these constraints increases roughly in proportion to the mean ECS increase from CMIP5 to CMIP6 suggests that these constraints are in not actually constraints on ECS, rather, at best they are constraints on the feedback processes they target. I develop argumentation this below, along with providing some more technical comments. I sign this review such that should the authors have any issues understanding my point they can contact me directly.

Sincerely,

Thorsten Mauritsen

We thank the reviewer for the helpful and constructive comments. We have revised our manuscript in light of these and the other reviewer's comments we have received. A pointwise reply is given below.

Major comments

Climate sensitivity is inversely proportional to the feedback parameter (lambda in equation 2) which in turn is a sum of a series of processes (sum of lambda_i). My take on the situation is that many of the emergent constraints (except COX) were successful on CMIP5 because inter-model spread in ECS in that ensemble was dominated by spread in low-level cloud feedbacks in the tropics. However, if any other feedback, e.g. water vapour feedback or any other cloud feedback, is biased in the ensemble as a whole then these kinds of process-oriented emergent constraints will necessarily be biased in their estimates of ECS. Probably even collectively since they thrive on the same kind of model spread, and so just because there are many studies that agree doesn't increase our confidence in

Printer-friendly version

Discussion paper

their quantitative outcome. Likewise, if structural commonalities among models cause an unreasonable low inter model spread in some other feedback process then the emergent constraint is going to be over-confident. All in all, the results suggest that the original studies were overly confident and that changes in feedbacks not constrained by these studies cause them to be biased with a sign that cannot be determined (since CMIP6 probably also contains collectively biased feedback processes). Thus, these process-oriented emergent constraints are perhaps best thought of as constraints on the processes that they target, rather than constraints on ECS, and in extension the original studies have been disproven by the results of this study.

There are alternatives to process-oriented emergent constraints, though, one of them which is included in this study (COX, more about this study and why I think there is a shift below). Emergent constraints that use global temperature change as a predictor of ECS do not suffer from the same problem: even if one feedback is biased in a model ensemble the constraint can in principle still work since both global change and ECS are inversely proportional to the sum of feedbacks. Suggestions of emergent constraints of this kind include Last Glacial Maximum (Hargreaves et al. 2012 doi:10.1029/2012GL053872), Pliocene warming (Hargreaves and Annan, 2016 doi:10.5194/cp-12-1591-2016), and post-1970s warming (Jimenez-de-la-Cuesta and Mauritsen 2019 doi:10.1038/s41561-019-0463-y). All of these ideas have been tested across ensembles including CMIP6/PMIP4 (Tokarska et al. 2020 doi:10.1126/sciadv.aaz9549; Renoult et al. 2020 doi:10.5194/cp-2019-162), finding essentially unchanged results between ensembles. Other studies worth mentioning are Bender et al. (2010, doi:10.1007/s00382-010-0777-3) and Dessler and Forster(2018), although I haven't seen tests of these.

I think all of the above is rather straightforward and fairly easy to understand. I think the authors have everything at hand that they need to draw the conclu-

Printer-friendly version

Discussion paper

sion that the process-oriented emergent constraints are not useful for estimating ECS, but rather should be better thought of as ideally constraining part of the cloud feedback. There are several places throughout that needs revising.

We think the reviewer has a very good point. We therefore extended the section 4 adding the following discussion:

"Our findings suggest that the process-oriented emergent constraints (i.e. all of the emergent constraints investigated here except COX) are only successful in constraining ECS as long as the uncertainty in ECS is dominated by the same process or feedback. In the CMIP5 ensemble, cloud feedback is the main contributor to the spread in ECS with low-level clouds in tropical subsidence regions dominating the spread in cloud feedback (e.g. Ceppi et al. (2017)). If any other process or feedback is biased (or missing) in the ensemble as a whole, then these process-oriented emergent constraints will be biased in their estimates of ECS. The appearance of diverse new feedback processes in CMIP6 could explain reduced skill when applied to CMIP6 data, and a tendency for these to be positive would explain the upward shift in the model ECS distribution that is not captures by the CMIP5-trained constraints. Process-oriented emergent constraints are therefore perhaps best thought of as constraints on the processes that they target, rather than constraints on ECS.

Emergent constraints that use global temperature change as a way of constraining ECS could in principle not suffer from the same problem. If one feedback is biased in an ensemble the constraint might still work as both, global temperature change and ECS, might similarly reflect the sum of all feedbacks. Emergent constraints of this kind include e.g. the tropical temperature during the Last Glacial Maximum (Hargreaves et al., 2012), tropical temperature anomalies during the mid-Pliocene Warm Period (Hargreaves and Annan, 2016), and post-1970s warming (Jimenez-de-la-Cuesta and Mauritsen, 2019). This seems to be supported by the findings of Tokarska et al. (2020), who tested an emergent constraint for the transient climate response based on recent global warming trends on the CMIP5 and CMIP6 ensembles with similar results for both model ensembles. However, these temperature-based estimates are sensitive to assumptions about forcings and unforced decadal temperature variations, which could also be systematically wrong, as could model-predicted relationships between feedbacks on short and long time scales that are implicit in most such measures. Indeed the significance of the COX constraint dropped as much from CMIP5 to CMIP6 ($p = 0.0032$ to $p = 0.1689$) as most of the other constraints in this study."

I think it is not reasonable to provide best estimates of ECS in the abstract and summary based on this study for the following reasons:

1) The above issue.

2) Because the study does not apply the latest observations to the constraints, rather opts for using the original observations. This is a perfectly fine choice given the scope of the paper, but it does mean the constraints are not up to date.

3) Because the study uses an implicit flat prior which in case of weak data automatically leads to high-biased results.

I would instead suggest the authors cite percentage increases which is anyway all that is relevant here.

We agree and therefore removed the best estimates of ECS from the abstract and the summary section and replaced them with percentage increases as suggested. The best estimates are now only given in the results section (3).

Regarding the Cox et al. 2018 constraint, section 3.2, this is built on the Hasselmann (1976) single heat capacity model. In this model there is a linear relationship between Psi and ECS, however, and despite what they claim if you add a deep ocean to the model you obtain a non-linear relationship wherein the relationship is weaker for higher ECS, see Annan et al. (2020 doi:10.5194/esd-11-709-2020), their figure 5 (note flipped axes). If you look at how the CMIP6 models

Printer-friendly version

Discussion paper

are distributed they are simply situated in the flatter part of the expected curve, and if you fit a straight line to it you will obtain different slopes than for CMIP5.

The reviewer has a good point. We therefore added the following sentence to section 3.2:

"For example, Annan et al. (2020) showed that the assumed linear relationship between Ψ and ECS does not hold when adding a deep ocean to the model.".

In this regard, and this applies not only to this study but most of these kinds, I am concerned with the general use of linear regression. The most silly example is SU constraint, where despite getting the wrong sign of the slope in CMIP6, you obtain a constraint on ECS. I think studies must be much more smart about their choice of statistical model, and not just use linear regression when non-linear behaviour is expected or other physical constraints can be applied such as a near-zero intercept, examples in Jimenez-de-la-cuesta-Otero and Mauritsen (2019), Annan et al. (2020) and Renoult et al. (2020). In case of process-oriented emergent constraints one could perhaps think of using Equation (2) in the form ECS$\sim$a/(b+x) where x is a process-oriented predictor. I am not saying the authors need to change this, but it would be worthwhile acknowledging that using linear regression, heedlessly, can lead to misleading and over-confident results.

We added the limitations of the approach including the references that you mention to the methods section that introduces emergent constraint methodology (section 2.2) and also briefly refer to this in the summary section:

"A further limitation of our approach is the statistical model itself. Similar to many other emergent constraint studies, we use an ordinary least squares linear regression model for each emergent constraint. However, in some cases this might not be appropriate, e.g. when we expect non-linear behavior or when physical constraints can be used to derive further constraints for the regression model like a zero intercept (Annan et al.,

ESDD

Printer-friendly version

Discussion paper

2020; Jimenez-de-la-Cuesta and Mauritsen, 2019; Renoult et al., 2020)."

I found the discussion of statistical significance somewhat disturbing. The chosen thresholds seem purely subjective, as far as I can tell. I would suggest to delete this whole discussion which seem rather pretentious.

Following this comment and a similar comment by Peter Caldwell (2nd reviewer) we decided to remove the whole bootstrap significance testing from the paper. Following the reviewers' suggestions, we replaced the bootstrapping method with a t-test on the correlation coefficient. The null hypothesis of this t-test is that no correlation exists between the predictor and ECS. In the revised version of the manuscript, we now give p-values of the emergent relationships that correspond to the probability that the absolute correlation is larger than $|r|$ even though the null hypothesis is true, i.e. the true underlying correlation coefficient is zero. Moreover, we do not use the p-values anymore to specify absolute significance (as you noted, our categories "highly significant", "barely significant", etc. were rather subjective), but only use them to specify relative significance, i.e. to indicate whether the statistical significance changes when moving from CMIP5 to CMIP6. Similar to our original bootstrapping approach, the new approach using the t-test shows that except for the ZHA constraint, all emergent relationships show a higher significance for the CMIP5 ensemble than for the CMIP6 ensemble.

I found Sections 4 and 5 rather long and repetitive. I would suggest revising and sharpening.

In order to address the two reviewers' comments we changed sections 4 and 5 substantially. Section 4 (discussion) now discusses possible reasons for the change in skill in the emergent constraints when moving from CMIP5 to CMIP6. Section 5 (summary) now gives a summary and discusses limitations of our study, some of which are now described in the methods section.

Minor comments

19, 'of spread in ECS among models'

Replaced "source of uncertainty in ECS" by "source of spread in ECS among models".

36, 'concentration over pre-industrial levels'

Added suggestion.

69, I am not sure Forster et al. 2020 is correctly cited here

We removed the reference to Forster et al. (2020).

85-92, perhaps drop Delta from F, and when using a specific forcing in equation 2 write F_{4x} or something?

We greatly simplified section 2.1. In the revised version, equation (2) does not appear anymore.

90-93. did you account for model energy leakage and drift? Concerning drift, some do account for this, but is not always obvious if there is a best way, nevertheless you must document what you did.

We accounted for possible model drift when calculating ECS by subtracting a linear fit of the pre-industrial control simulation from the abrupt4xCO2 experiment. Other than that, no explicit corrections for drift or energy leakage have been done. We added this information to section 2.1 of the revised manuscript:

"In this calculation, the linear fit of a corresponding pre-industrial control run is subtracted from the 4xCO2 run to remove any model drift that is present in the control climate without adding noise (Andrews et al., 2012). Other than that, we do not explicitly account for other problems such as energy leakage."

120-123, there are also an intermediate option, e.g. the Cauchy prior used in Annan and Hargreaves (2011, Climatic Change). Regarding the uniform prior, please specify which cut-off you use.

Following this comment and a similar comment by Peter Caldwell (2nd reviewer) we made the following changes to the manuscript:

"In this derivation of the probability $P(y)$ we do not assume any prior knowledge on ECS – in other words, that an ECS near 8 K would be deemed just as probable as one near 4 K if both are equally consistent with the observational best estimate x_0. We do this for simplicity. The PDFs would shift somewhat lower with a broad prior on processes instead (see Sherwood et al. (2020)), but we are concerned here with how outcomes compare using CMIP5 vs. CMIP6 data, rather than the exact ranges obtained. Such comparisons are not sensitive to the prior."

We emphasize that we focus on the relative differences in the constrained ECS distribution between CMIP5 and CMIP6, which does not depend on the choice of the prior.

154-155, this is a misinterpretation, the IPCC 'likely' statements refer to 66-100 percent probability.

Thanks for clarifying this! We replaced "IPCC likely" by "66% confidence interval (17–83%)".

158-159, I felt this statement could be made more informative by explaining that it is the covariance of clouds with surface temperature anomalies.

As suggested, we added this extra information to the beginning of sect 3.1.

180, 'results to choices made in the analysis'

Added suggestion to manuscript.

236-263, these constraints seem to have some legacy with Fasullo and Trenberth (2012, Science), perhaps worth mentioning if the authors agree?

We mentioned the Fasullo and Trenberth (2012) constraint in the sections describing the SU constraint (3.7) and the TIH constraint (3.8).

276-277, it is incorrect that Volodin (2008) was the first emergent constraint on ECS, there is Covey et al. 2000 and Knutti et al. 2006 before then.

We rewrote the sentence and removed the wrong part that stated that Volodin (2008) was the first emergent constraint on ECS. Thank you for spotting this.

320, I would suggest deleting 'describing the real world'

Removed "describing the real world".

354-355, The idea and strength of an emergent constraint is that you use something you can observe to predict ECS. It really shouldn't matter if a process is slightly different in the warmer 2xCO2 world.

We agree with the reviewer that it probably does not matter if a process is different in the future climate as long as the ESMs know about it. But there may be processes that are unimportant in the ESMs and hence not captured by the emergent constraints but that are important in reality. We clarified this by rephrasing the corresponding paragraph in section 4 of the revised manuscript:

"[...] While this assumption seems to make sense, we do not know whether the ESMs cover all relevant processes of the real Earth system. For example, it may be possible that there exist processes that are unimportant in the ESMs (and hence are not captured by the emergent constraints) but are actually important in reality. This lack of relevant processes may lead to an overconfident constraint. Thus, the more complex ESMs of the CMIP6 ensemble are more likely to capture relevant processes of the true climate which leads to weaker emergent relationships. On the other hand, emergent constraints on the less complex CMIP3 and CMIP5 ensemble may be overconfident."

357-359, same applies here

See answer above.

360-361, this statement goes further than Zelinka et al. 2020. I would suggest replacing 'dominated by' -> 'to some extent associated with'

Changed as requested by the reviewer.

402-403, as per my above argumentation, I would be very careful with making this statement.

Since we removed the bootstrapping testing from the paper, we also removed this whole paragraph.

409, this might also have been shown by Klocke et al. (2011), check.

In Klocke et al. (2011), they state: "This suggests that model weighting based on statistical relationships alone is unfounded and perhaps that climate model errors are still large enough that model weighting is not sensible.". In our opinion, this does not really fit into our argumentation, since we do not consider model interdependence and model skill at this point.

421-426. the very same paper also shows that the ECS estimated from 4xCO2

runs is higher than twice that in 2xCO2 runs, and that the bias of the same order of magnitude.

We did not find any reference to the reviewers point in Klocke et al. (2011) (https://doi.org/10.1175/2011JCLI4193.1).

Interactive comment on Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2020-49, 2020.