Earth System Dynamics Discussions

Earth Syst. Dynam. Discuss., author comment AC1
https://doi.org/10.5194/esd-2020-85-AC1, 2021

Reply on RC1

Benjamin M. Sanderson et al.

Author 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-AC1, 2021

Many thanks to the reviewer for taking the time to assess the paper - we will respond to comments in line. Following the comments of this review and others, we have restructured the paper as a review article

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.

Thanks the kind comments and thoughtful review.

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.

We argue below why we find this position to be justified.

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.

Thanks for this point. We, in fact, fully agree that the analytical aspects of this paper do not represent sufficient quantitative advances in the understanding of combining constraints, and are rather intended to illustrate the wider conceptual points which are the focus of the paper (which the reviewer clearly appreciates, given the comments which follow). Given this (and similar assessments from the other reviewers), we have requested for the paper to be be classed as a review. In this context, we have removed the bootstrapping analysis - given the point has been largely covered by Caldwell 2018, Schlund 2020 and Brient 2019.

We see the paper's novelty as a conceptual framework to assess how emergent constraints relate to model assumptions. Given this, the toy model analysis does not provide deep insight on the nature of ocean timescale response, nor was it intended to. The point illustrates that if a model ensemble contains common approximations and a small number of degrees of freedom (in this case, represented by a single layer ocean), then strong emergent constraints can emerge which would confidently constrain the future response to a certain value - but the constraint may be broken, or simply not exist if a more complex model with additional degrees of freedom is used (here illustrated by a 2 layer ocean). The actual models used are incidental - they simply illustrate models at two different complexity levels. We go in in the discussion section to consider cases where such common oversimplifications may exist in the CMIP class models (e.g. soil respiration-temperature relationships).

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.

We broadly agree that the key issue at hand is the structural shortcomings of the models. However, we disagree that the ECs can 'only improve' existing model output - because of the implied precision in the value of the true future response which arises from considering only the error implicit in the regression relationship. An emergent constraint, literally, proposes that an unknown quantity is constrained, unlike an ensemble multi-model mean which just a model estimate with no pretense of precision. Thus, the precision of the EC can potentially be over-stated through the lack of consideration of aspects of common, oversimplified structural assumptions which create strong ensemble relationships. As such, although the structural errors are ultimately features of the models themselves, the use of ensemble relationships which are themselves subject to model errors impacts directly on the 'added value' of emergent constraints.

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.

This point is well taken, but contains an implicit assumption that the use of an emergent constraint can \emph{only improve} a projection - where we would argue that ECs have the potential to make an overconfident projection due to a consideration of only the subset of available model validation data which was used in the EC. Fundamentally - the structural errors have the capacity not just to impact the simulated values of A and B, but also the the relationship between A and B - perhaps creating relationships which would not be present in a superset of models. We have already seen that different emergent constraints can emerge from the same ensemble with differing conclusions (e.g. the fluctuation dissipation relationship of Cox et al 2019, and the atmospheric stability relationship of Sherwood 2015). Either of these constraints used alone would constrain ECS to low or high values respectively. These problems are not insurmountable - the Bayesian framework of Williamson 2019 allow for the combination of different lines of evidence, but enacting such a strategy requires the modeling community to engage with (1) the independence of different constraints (e.g. Caldwell 2018) but also (2) the degree to which common structural errors could bias A, B or the relationship between them.

Our position is thus not that the constraints have no value, rather that the common assumption that a constraint used in isolation to constrain a projected ensemble distribution is not justified without considering other relevant aspects of model performance (including other ECs) and the realism of potential common structural assumptions which may have biased or created the emergent relationship. As we argue in our conclusions - we believe that there is a middle ground between standalone emergent constraints, and generic model multi-variate skill scores which could allow a focus on variables which might be of relevance to projected climate without relying exclusively on a single relationship.

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 constraint but the model ensemble. However, it is equally important to mention and assess these uncertainties when presenting mutli-model means and constrained results.

We agree that the structural biases will impact the unconstrained model projection distribution, including the multi-model mean. But there are two further points to consider: (1) the multi-model mean and model distribution is not advertised as a calibrated estimate of a specific value, together with an implicit error and (2) the added value of the emergent constraint is also potentially subject to structural error.

The reviewer's example considers one example of structural error (the omission of N

fertilization), and indeed - this is a case where the structural omission may create an over-confident constrained result if used to calibrate future GPP projections (if present day LAI simulation is a function of N-fertilization). We would argue that this is more problematic for the EC than the multi-model mean. Using the EC to calibrate future GPP is explicitly calibrating GPP to compensate for a missing process. This is an additional source of error because the effect of the bias of the missing process may be different for present day and future GPP, and therefore the constrained result stands to be confident, yet wrong.

The use of a single metric therefore throws away any information which is not in the constraint itself. In the reviewer's example, a model tuned to have the right temperature and LAI may be subject to other biases (in present day GPP or latent heat flux, for example), which would have been considered by the developers calibrating the model. As such, the ensemble distribution of projections and ensemble mean result are of course also subject to biases from missing process, but the models are expert tuned considering a wide range of metrics and therefore, to some degree more robust. The use of the EC to calibrate projections effectively allocates zero weight to any orthogonal aspects of model performance (some of which may themselves be emergent constraints on the same quantity in the ensemble).

To put it in Bayesian terms, the prior ensemble distribution is subject to structural errors, but so are the relationships which provide the basis for the emergent constraint. Therefore, the ECs have the potential to be additionally impacted by structural errors than the unconstrained model distribution.

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.

We agree with the reviewer that the models were not intended to be exchangeable with reality. Model developers are generally keenly aware of the approximations made in their parameterizations. However, we maintain that the exchangeability assumption is implicit in the use of an emergent constraint to estimate the most likely value and uncertainty in that value. This is the view shared by Williamson (2019).

This problem does not apply to the simple estimate of mean and variance of the original model ensemble projections - which may be biased due to missing processes, but they are not, themselves, uncertainty estimates in a projected quantity. ECs, however, frequently use the error in the regression relationship to estimate an uncertainty in the projected quantity (see Cox 2018, Varney 2019 amongst many other examples) - and this additional step is a strong assumption, that the relationship is equally applicable to reality as it is to the models in the ensemble. We have expanded on this point when first introducing the concept of exchangeability in the introduction (line 94).

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.

Agreed, this broadly represents our intended illustration with the simple model - that the 'improvement' in this case would give a confident, but ultimately incorrect, constraint on the future dynamics of the model.

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?

Here we differ from the reviewer. The single layer ocean is useful for some applications. It can capture some of the first order dynamics of the climate system given appropriate parameters. But the nature of that calibration matters. The Emergent Constraint approach takes an extreme position - that only one aspect of the model should be used in tuning (that which is correlated with the future response). This results in a very tight constraint on the single layer model's parameter space, but because the relationship itself is biased by the structural errors in the model - the constrained value is incorrect.

A more robust calibration strategy is to use all available relevant data to calibrate the model (e.g. the entire historical timeseries, paleoclimate records etc). In this case, individual aspects of the model error would still be subject to structural deficiencies, but trade-offs between tuning to different observations would reduce the degree to which the parameters are constrained (see Sanderson 2008 for an example of this in ESMs).

In short, we agree with the reviewer that all model outputs are subject to structural error. But we argue that the sole reliance on projection calibration from a single ensemble relationship in ECs introduces a particularly acute exposure to specific aspects of model error which will invariably lead to overconfident constrained projections.

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.

This is a fair point. Our objective is certainly to talk about the methodology of emergent constraints in general - the key conceptual arguments are not specifically associated with climate sensitivity. We do not seek to exhaustively review or list every published emergent constraint here (this has been done elsewhere, e.g. in Brient 2020, Hall 2019 or Williamson 2018). But - the reviewer is correct that an ocean-specific example would be desirable to illustrate relevant structural assumptions for different broad genres of constraint. We have added a new section 5.3 on constraints on future ocean carbon uptake. Many thanks for the references.

Furthermore, the ocean is ignored in questions about the ECS or atmospheric CO$_2$ (point 6), although the 2 timescale models clearly indicate that the ocean is important.

Point well taken. We have revised the case studies to include ocean processes at relevant points.

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 r2 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 r2 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 a. (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.

The revised paper has removed this section entirely - given the paper was already long, and the assessment of EC correlation has been well discussed by Caldwell (2018)

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.

Thanks for this point. We have removed the specific references to climate policymaking. We would counter, however, that the majority of emergent constraints use probabilistic language in their primary conclusions - but in almost all cases, these probabilities exclude the potential for errors which are the focus of this study (that is, the uncertainty arising from model common simplifications which project onto either simulated quantities or intra-ensemble relationships).

6) Lines 442-446: The authors claim that the differences in atmospheric CO$_2$

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.

Totally agreed. We apologise for the land-centric discussion and have updated the 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.

We agree with the reviewer that ocean circulation biases play a significant role. However, we disagree that this is the sole source of uncertainty in atmospheric CO_2 concentration biases. As we discuss in the revised section - this is a trade-off between numerous factors: ocean productivity, circulation, land carbon and concentration feedbacks and soil temperature response. Systematic common biases in any of these aspects would qualify as a potential structural uncertainty in future CO_2 concentrations.

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; Nijsse 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 + r2 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.

We have removed this figure in the revised version.

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

Table 1,3 now removed. Labelling fixed.

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

Analysis removed in revision.

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?

This section has been removed in revision

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

This section has been removed in revision

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.

Agreed. and again, this section has now been removed.

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

Thanks - agreed. Added.

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 futureresponse.

Thanks - we agree that these are process-based constraints, we've removed the word "feedback" from the title to allow a broader definition.

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

Agreed, thanks.

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.

We disagree with this. If there are a small number of parameters governing a process in similar parameterisation schemes throughout the ensemble - the response of such models to both historical and future forcings will be a function of that small set of parameters, increasing the chance that an emergent relationship might be found.

Soil respiration/temperature relationships are a good example in CMIP, as we point out in the following paragraph. A simple temperature dependency equation has good skill in representing soil respiration in CMIP as a function of temperatures (Todd-Brown et al. 2013), and this enables strong emergent constraints on future soil respiration temperature sensitivity such as Varney (2020). But - this simple temperature relationship also fails to represent a large fraction of spatial variability in observed soil respiration (Todd-Brown et al. 2013) - which is potentially attributable to the common over-simplicity of the representation of the process in the ensemble.

Adding complexity in this case may indeed increase variance in the predictor or predictand, but the increased number of degrees of freedom in the process representation has the potential to add noise to the relationship between the two. However, if the ensemble diversity can be demonstrably reduced to a single process equation with (in the extreme case) one free parameter - responses to different forcings will be correlated by construction.

11) Equations are not numbered}

Fixed

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

Thanks, fixed

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

We've deleted this paragraph.

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.

Thanks for this suggestion - we've incorporated this discussion as suggested.

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

We've deleted this paragraph - as the core of the argument is more clearly repeated in the following paragraph.

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

Thanks - yes. There's also a valid point from this paper on the significance of constraints persisting for multiple generations.

References

Williamson, D. B., & Sansom, P. G. (2019). How are emergent constraints quantifying uncertainty and what do they leave behind?. Bulletin of the American Meteorological Society, 100(12), 2571-2588.

Varney, R. M., Chadburn, S. E., Friedlingstein, P., Burke, E. J., Koven, C. D., Hugelius, G., & Cox, P. M. (2020). A spatial emergent constraint on the sensitivity of soil carbon turnover to global warming. Nature communications, 11(1), 1-8.

Cox, Peter M., Chris Huntingford, and Mark S. Williamson. "Emergent constraint on equilibrium climate sensitivity from global temperature variability." Nature 553.7688 (2018): 319-322.

Sanderson, B. M., Knutti, R., Aina, T., Christensen, C., Faull, N., Frame, D. J., ... & Allen, M. R. (2008). Constraints on model response to greenhouse gas forcing and the role of subgrid-scale processes. Journal of Climate, 21(11), 2384-2400.

Brient, Florent. "Reducing uncertainties in climate projections with emergent constraints: Concepts, examples and prospects." Advances in Atmospheric Sciences 37, no. 1 (2020): 1-15.

Todd-Brown, K. E. O., Randerson, J. T., Post, W. M., Hoffman, F. M., Tarnocai, C., Schuur, E. A. G., & Allison, S. D. (2013). Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations. Biogeosciences, 10(3), 1717-1736.