Reply on RC2
Benjamin M. Sanderson et al.

Sanderson et al provide discuss the nature of emergent constraints (ECs), particularly the potential role of structural errors in driving uncertainties in ECs. Overall, I found it difficult to know what to do with this paper. While the material is clearly presented and interesting to read, it feels more like a review article or a perspective, rather than a journal article. The discussion is generally qualitative and/or speculative, and I'm still not quite sure what the main takeaways are.

Thanks to the reviewer for the feedback on the manuscript. The opinion that the paper is better categorised as a review or perspective was shared by the other reviewers, and the paper is now classed as a review by the journal (and we are framing it as such).

This paper was never intended to be a quantitative analysis - rather as a perspective on the application of emergent constraints found in an ensemble of model simulations where we know there exist structural limitations to models which are represented in model simulations.

Our takeaways are that we demonstrate in a simple model ensembles how structural errors can produce overconfident emergent constraints, introduce a classification scheme for emergent constraints, and discuss a large number of case studies on where structural limitations in current models may be an additional source of error in published constraints.

Furthermore, we would argue that this general qualitative discussion is necessary in the literature before specific quantitative advances can be made. The nature of structural errors, by definition, is processes which are not resolved or incorrectly resolved in the model. As such, a quantitative analysis of such errors is not necessarily feasible, but this does not mean that this source of error can be completely disregarded when considering constrained distributions of projected climate quantities produced through consideration of ensemble derived correlations - as is the case in the majority of EC studies published to date.

The quantitative analysis in the paper is limited to Figure 1, which shows the relationships between a number of previously published ECs ... covered in more
Given the classification of the paper as a review, we have removed Figure 1 entirely - which originally served to illustrate inter-constraint relationships, but we agree that this topic is adequately covered by Caldwell, Bretherton and Schlund.

Figure 2, which uses the simple energy balance models to illustrate different kinds of ECs identified in the paper. The 2-layer energy balance model has been extensively discussed by Geoffroy et al, Armour, and Lutsko & Popp (some of these papers are cited in the present manuscript).

We agree that Geoffroy, Armour and Lutsko have produced fine papers illustrating the dynamical assumptions of simple energy balance models, but they do not make the point we are making here - that an overly simple model structure can produce very strong emergent constraints which can be demonstrated to be overconfident in the context of information provided by a more complex class of models.

Without more novelty or substance, it is difficult to recommend publication, although I enjoyed reading the manuscript.

We hope our arguments here persuade the reviewer of the novelty of the article - and the need for a non-quantitative discussion on this topic. We hope further that the restructuring of the manuscript as a review will better reflect the article's purpose.

Moving forward, the authors might want to think of ways to deepen their analysis. One approach might be to develop a mathematical framework or procedure for identifying and speaking about structural errors in emergent constraints.

We believe this paper has already been written, by Williamson (2019) - which motivated us to write the present study. Williamson (2019) demonstrated how structural errors could theoretically be implemented in regression-based constraints in a Bayesian statistical model, but did not discuss any concrete case studies of what such structural errors would look like. In this study, we highlight potential sources of structural error and potential paths forward to future approaches which could better represent these errors in constraints.

Alternatively, they could focus in on a particular kind of emergent constraint and probe the structural assumptions used by this kind of emergent constraint in more depth. For example, they could dig into the cloud schemes responsible for the process-based constraints on ECS (e.g., the Sherwood, Brient and Zhai constraints) to really understand the underlying structures. A template could be the recent paper by Thackeray et al, which investigates the snow albedo feedback over multiple generations of climate models, including its relationship to the well established emergent constraint on the feedback.

Papers focusing on the structural assumptions in individual constraints are enormously valuable, and our section 5 discusses a large number of case studies and how sources of structural error might arise. However, to objectively quantify these errors for any of the studies referenced would be a study in itself, and is far beyond the scope of this article.

In conclusion, we see the necessity for a general article on the potential for structural errors in emergent constraints, rather than a deep dive into a specific process, because there is a general lack of discussion of such errors in papers published to date - with a near blanket assumption that relationships found between predictors and predictands in the multi-model ensemble can be used to reduce uncertainty in projected quantities. In
the study, we have demonstrated a simple model case where this assumption is demonstrably false, and we have highlighted where relevant structural assumptions may exist in the current model archive. We hope, in turn, that this will provoke future study which will work towards building constraints which are more robust to structural errors.

**Technical Corrections:**

*The title is vague: "On Structural Errors in Emergent Constraints", and again makes it hard to know what the main takeaways are.*

Revised to "On the potential for structural errors in emergent constraints" - which well captures our topic.

*In the first sentence of the introduction, I'm not sure it's right to state that higher CO2 concentrations are a "boundary condition which has yet to be realized". Increasing CO2 concentrations doesn't enter the boundary conditions, it adds a forcing term. So I would describe climate forecasting as an initial value problem, rather than a boundary condition problem.*

We have changed "boundary conditions" to "forcings" to remove this ambiguity.

More generally - we agree, that if humans are considered part of the Earth System, then it could be argued that climate projections are an initial condition problem (with boundary conditions being only variations in solar forcing). However, it is generally accepted in the climate literature that the initial conditions refer only to the state of physical climate system at the time of initialization (see, e.g. Hawkins and Sutton 2009, Deser 2012, Hawkins 2016, Sriver 2015).

*The paper claims that the Cox and Sherwood D constraints are well correlated with each other. But in fact the correlation co-efficient is only 0.31 (~10pct of the variance explained).*

This section is entirely deleted in the new version - but we do agree that this sentence was incorrect.

*Typo at L610: "wider constrained range wider"*

Thanks, fixed.

*L654: In terms of multi-metric approaches, the authors may wish to cite the "cloud-controlling factor" approach (see Klein et al for a recent review) which has recently shown promise for constraining cloud feedbacks.*

Thanks for this. Completely agreed that there is an argument for greater robustness through the consideration of "bottom-up" decomposition of net feedbacks, such as that demonstrated in the Klein paper. We've added a paragraph on this topic in the discussion.

**References:**


