Please find our proposed changes to manuscript "Emergent constraints for the climate system as effective parameters of bulk differential equations" in response to this review. Our replies are in blue font and indented.

Review of Huntingford et al 2023:

On one hand, this paper is clear, well-written, and its PDE examples are simple, relevant, and pleasant to work through. On the other hand, I didn't really learn anything from reading this. For example, I can't imagine anyone understanding Cox et al (2018) without having a deep understanding of the notion that the emergent equations governing temperature changes on various timescales are linked via heat capacity. This left me wondering whether the paper is worth publishing. Ultimately, I think the answer is yes because if someone didn't intuitively understand that emergent constraints occur due to links between underlying governing equations, this paper would do a nice job of introducing them to the concept. I doubt the paper will be cited much, but that doesn't mean it shouldn't be published.

> First, we thank the reviewer for their time assessing our manuscript. We are fully aware that this manuscript will only appeal to a subset of those using Emergent Constraints (ECs) to refine the understanding of climate system components. However, we regard that group as important, representing many with a mathematic interest asking: "What underpins the EC method?". With such a high profile of ECs to constrain uncertainty, we hope of paper will trigger new lines of questioning and understanding of the technique.
>
> Our proposed replies are below, indented and in blue.

Minor comments (Note my convention is P2 L1 = Page 2, Line 1):

1. P2 L1: observationalists would disagree that ESMs form the basis of climate research. I tend to say they're a pillar of climate research.

> We will adjust the manuscript to use the word pillar.

2. P2 L6: It's not accurate to say that ESMs are typically forced with historical and scenario GHGs. A lot of time is spent on PI control, abrupt4xCO2, 1%CO2, etc. Minor rewording is needed.

> We will note the extensive and often long PI control simulations in the CMIP databases, as well as factorial experiments such as abrupt4xCO2.

3. P2 L19: "main possibly simplest answer" is awkward grammar

> We will split the sentence and reword as "*…why are ESMs different….. The simplest answer is….*"

4. P3 L30 – P4 L2: The first sentence here isn't very clear. I think you are saying that Schlund and others found that ECs based on CMIP5 were generally worse when applied to CMIP6. As written, it sounds like any EC, including ECs developed from CMIP6 data, would have wider bounds. I also found your wording a bit confusing because wider bounds could come from worse correlations between EC predictor and predictand OR from larger spread in the observations used to constrain. I guess the problem must be the former, but it takes the reader some unnecessary thought to get to that conclusion. Following on this, I think the obvious explanation for larger spread in CMIP6 is that the ECs from CMIP5 were overtrained: they are capturing noise rather than real EC signal. I'm confused how this possibility isn't even in your proposed reasons at all.

We agree that this text was not as well phrased as it should have been. In our revised manuscript, we will therefore replace lines P3 L30 to P4 L3, with:

> "*Schlund et al. (2020) tested the robustness of proposed emergent constraints by out-of-sample testing on a different model ensemble. These researchers found that emergent constraints on ECS, which were developed using the CMIP5 ensemble, do not provide useful constraints on ECS in the CMIP6 models. These ECs, therefore, fail to be "confirmed" (Hall et al., 2019). Recognising the danger of arriving at spurious emergent constraints based on the results of relatively small model ensembles (Caldwell et al., 2010), Williamson et al. (2021) have set the challenge of deriving more robust theory-based emergent constraints. To inform attempts to meet that challenge, here we address the basic, almost philosophical question: "What is an emergent constraint?"*"

5. P5 L15-16: you introduce T* here but don't use it again except P8 L3. I suggest deleting both T* references. In particular, the wording of the first intro to T* was very confusing (and I think, unnecessary).

> Our intention is that "*" represents measurements. We will reword this sentence and make the notation clear. Rather than adjust this notation, we will instead annotate the word "Data" above the vertical uncertainty bars in Figure 1 and Figure 2 to be "Data, ΔT_S*".

6. P5 L26-28: When you say "running mean", I immediately wonder what the averaging period is. I think it would be better to call this statistic the "annual average". Relatedly, the running mean itself isn't a measure of climate change. The time derivative of the running mean is your proxy for climate change. But of course, the annual average isn't special in this regard – the long-term average of the time derivative of the instantaneous T(t) equation would give the same answer because the derivative is a linear operator.

> We will reword this as: "*A second temperature-based statistic we can consider are changes in the annual averages. The time derivative of annual averages is a proxy for global warming. Annual averaging, denoted by an overline, is…*"

7. P6 paragraph starting L8 and P8 paragraph starting L23: I think this discussion can be improved. I think the big point you're trying to make is that while the fact that there exists a predictive relationship between the observable and the future quantity of interest allows you to predict ECS, the slope of that relationship provides interesting information about the physical equations that underpin that relationship. I think you are further pointing out that even though there may be uncertain terms in the equation governing the current-climate variable and in the equation governing future change, those uncertain terms sometimes cancel out when the quantity you're actually interested in is the ratio between predictor and predictand. As it stands, I don't think it is interesting that uncertainty in either of 2 parameters would give rise to the intermodel spread needed to compute an emergent constraint. I also don't think you adequately explained why b_i/H_{0i} would be constant across models.

> We agree that in a standard research paper on ECs, then finding a relationship between the observable and future quantity of interest is sufficient. Here, we try to provide a potential mathematical explanation for why such regressions emerge. We hope our suggestions will open new ways to interpret and understand ECs, which matters given their widespread use in climate science. Based on this reviewer's comment, we will add a sentence to this effect in the Discussion.
>
> We are keen on an example where the aggregation of internal behaviours may involve a single effective parameter that is invariant, inter-ESM. In that case, uncertainty may be in the forcing. To maintain a single degree-of-freedom, we consider this uncertainty to be in the

forcing and identically between seasonal and long-term forcing (i.e. this implies b_i/H_{0i} fixed between ESMs). We accept that this appears slightly contrived, but it does allow us to illustrate a key point in our conceptual structure. We will, though, enhance the sentence to make this much clearer at the point where uncertainty in forcing is introduced.

8. This is a minor point, but the seasonal cycle in eq 4 and eq 11 won't be exactly equal to the observed seasonal cycle in a warming planet (H_0>0) because the planet will have warmed a bit in the 6 months between winter and summer.

We will put in a sentence to this effect.

9. P8 L17: defining your current-climate metric as d/dt(annual-ave T(x=0)) makes sense, but multiplying it by sqrt(t) seems contrived. If you have such an exact understanding of the underlying equations, you'd probably already know what H_0 was, so a regression would be unnecessary!

Process knowledge of any heat conduction (i.e. "parabolic") PDE will indeed suggest that sqrt(t) behaviour is often present. Ideally, the "y"-axis of any emergent constraint is independent of time (or scenario), as shown. But this does raise an interesting yet rarely mentioned point, that in some instances, an EC may be robust but its position on an 'x-y' graph depends on a future time or GHG level. We will add an additional sentence that makes this point, noting that should the 'y'-axis be simply dT/dt, then the EC would move downwards by 1/sqrt(t).

10. P11 L34-P12 L1: I can't imagine how a real emergent constraint wouldn't have a physical underpinning that can be expressed as an equation. We may not know what that equation is, but if there truly isn't an underlying equation behind an empirical relationship, how could that relationship possibly be real?

The original paper version contains the sentence (P11, L34): "We do not propose this as a universal theory of ECS, as some may function for other reasons". Based on this reviewer's comment, we will remove this sentence. The manuscript will then lead directly into the next sentence that discusses the importance of an underlying physical process. We will ensure that sentence is written more strongly, noting we should always be pursuing explanatory physical mechanisms in EC research.

I don't think the emergent relationships that distill into an EC are necessarily (or even typically) PDEs. In your examples, the fine scale governing equations are PDEs, but the equations you derive for seasonal cycle and warming tendency are not. Similarly, concepts like "if you don't have much cloud in the current climate, then you don't have much cloud to lose in the future" are fundamentally connecting model state to model change. This doesn't invalidate any of this work, but a reframing of the title and some rewording in the abs

OK, some large-scale implicit equations may be ODEs (as per our first example) or, indeed, simply algebraic. In the other direction, the overall system may still be relatively simple but involves coupled equations. We will modify the Abstract to remove mention of "PDEs" and replace it simply with "*equations*". An extra sentence will be added to the Discussion to note that the equations may take many formats.