The paper "Emergent constraints for the climate system as effective parameters of bulk differential equations" by Chris Huntingford et al. provides a formal description of emergent constraints as parameters of large-scale partial differential equations (PDEs). In contrast to small-scale PDEs explicitly coded into Earth system models (ESMs), these large-scale PDEs are not directly included in the models, but emerge across ESMs when aggregated across larger scales. Huntingford et al. provide two example PDEs derived from simple thermal models. By assuming different bulk parameters (e.g., heat capacities) for the different ESMs, they show that these PDEs can be used to derive emergent relationships between short-term and long-term responses of the system, which ultimately can be used as emergent constraints with appropriate measurements of the real Earth system.

> First, we thank the reviewer for their time assessing our manuscript.
>
> We appreciate the reviewer's summary above. As noted, our view of many ECs is that the "emergent" property is the discovery of large-scale differential equations coded implicitly in ESMs (via aggregation of explicit coding at finer scales). To be clear, we will amend the Abstract sentence at line 17 to read: "*We suggest that many ECs link to effective hidden PDEs implicit in ESMs and which aggregate small-scale features*"

General Comments

This paper reads well and provides an interesting approach that allows the derivation of emergent constraints from bulk PDEs. I agree with the authors that an emergent constraint discovery method based on physical reasoning and mathematical models is much more desirable than data mining, and will eventually lead to more credible and robust emergent constraints. However, I have some concerns about the relevance of this study regarding "real" emergent constraints.

> We are grateful that the reviewer thinks our paper reads well and is an interesting approach. We take full account of their concerns listed below, responding in full. Our replies are in blue font and indented.
>
> Concerning some of the more technical points, please note that there was an issue with the diagram .pdfs and the ESD online converter. The correct diagrams are presented below, and if our manuscript is accepted, we will work carefully with ESD to make sure they are reproduced as expected.

Currently, a large part of the argumentation of the paper is based on two very simple PDEs. Especially in the context of a changing climate (which is a necessary condition here), I think the equations are too simplified. Since the PDEs are missing a "loss" term, a constant forcing will lead to an infinitely rising temperature, which is not realistic. For example, what happens if you add linear loss terms (linear feedback) $-\lambda*T$ to your PDEs (e.g., so that your eq. (2) is similar to eq. (1) of Cox et al. 2018)? Could you still derive the emergent relationships from these new equations? I can imagine that there are certain conditions (e.g., small times, small λ, large forcings, …) under which your original equations are good approximations, but it would be good to guide the reader in detail through this process. Additionally, it would be very helpful if you can provide more details on these emerging bulk equations themselves and why they should be present in an ensemble of ESMs. Do you have any recommendations how to find such PDEs? An example with a real emergent constraint would also be incredibly helpful. All this will ultimately help the reader to gain more trust in your framework.

> The reviewer asks some fascinating questions here but answering these is cutting-edge research that is beyond the scope of this initial short perspective paper. Instead, we will add text to acknowledge the nature of the challenge that the reviewer poses. Specifically, we will add additional lines of text as:

> "*PDEs emerge commonly where state variables are globally conserved (i.e. for state variables closely related to energy and momentum). To aid transparency, we have also assumed underlying PDEs that are simple by design. Making these underlying models more relevant to the Earth's climate is an outstanding challenge. For example, in addition to horizontal heat transport, our planet emits longwave radiation to the wider universe. Such radiation provides the restoring force, λ, that ultimately stabilises the near-surface temperature. Including such a restoring force in our simple PDE models is one possible extension of our analysis, although, in tandem with an unknown heat capacity, c_p, this would potentially generate a two-dimensional EC. In practice, fitting a two-dimensional EC may be challenging given the relatively small number of data points (i.e. individual ESMs). Furthermore, analytical solutions may exist that allow for a time-varying value of H that approximates known historical climatic forcing*".

Finally, two technical comments: first, it would be very helpful if you could use continuous line numbers (and not start with "1" on every page) and also add line numbers to figure captions. Second, please consider depositing your code in a publicly accessible repository (e.g., Zenodo) to make your analysis more transparent and reproducible for other researchers.

> Unfortunately, I think the ESD template for submission causes this form of line numbering. Final ESD papers have no line numbers. We are very happy to upload our code to a standard scientific repository.

Specific Comments

1. P.2, l.30: Maybe add a reference here? E.g., Knutti et al. (2017), https://doi.org/10.1002/2016GL072012

> Thank you. We will add this reference.

2. P.3, l.4: It would be more precise to refer to "observational" data here (alternatively "observation-based").

> We will make this suggested wording alternation.

3. P.3, l.12: A better reference for this might be Hall & Qu (2006), https://doi.org/10.1029/2005GL025127. You might also want to cite Allen & Ingram (2002), https://doi.org/10.1038/nature01092 here.

> We will add these two references and associated additional wording around their citation.

4. P.4, l.1-2: It might be helpful for the reader to add the key conclusion(s) of the discussion of Fasullo et al. (2015) you mention here.

> Yes, the Fasullo paper is important and interesting, and we will cite further its key findings.

5. P.4, l.29: I guess technically it's a function of the total noise, so ε **and** η, not only ε.

> Correct. That should read ε+η.

6. P.5, l.18: Required for what?

> We will rewrite this as: "*ECs require a quantity that is both modelled for the contemporary period and is available as a measurement, such as the seasonal range, ΔT_S.*

7. P.6, l.15: It’s not only the data points (I guess by "data points" you are referring to the (x, y) tuples you get from the models?), but also the measurements that constrains the forcing element b.

> Please see our response below, which concerns the same sentence.

8. P.6, l.15-16: I think this sentence is not clear enough: "With the forcing uncertainties common for both short– and long–term drivers". You need to explicitly assume that b_i/H_{0i}=const across models; you should mention that.

> We will rewrite this sentence (and split it into two), as well as adjust the following sentence to: “*In this case, the emergent constraint represents the discovery that there is a single ESM-independent internal bulk parameter (i.e.* c_p*’). Measurements then provide the constraint to remove uncertainty in the forcing element* b_i*. With the forcing uncertainties common for both short-and long-term drivers (i.e. the assumption that* b_i/H_{0i} *is constant), the measurements implicitly constrain* H_{0i}*, and thus the background warming, dT/dt.*”

9. P.6, eq. (8): You might want to refer to Fourier’s law here.

> Yes, we will do that.

10. P.8, l.17: Why don’t you simply divide T(0, t) by sqrt(t) to get a y that is not dependent on t?

> Yes, we do exactly that scaling as the ‘*y*’-axis of the EC (please see Figure 2). Please note that our other reviewer encouraged the opposite, of not normalising by sqrt(t). We hope the current framework of keeping in the sqrt(t) in the text but normalising in the EC diagram (so making the EC time-independent) is a satisfactory presentation.

11. P.12, l.10: I think this classification only applies to linear second-order PDEs, not to every PDE.

> Yes, we will make that point clear.

12. P.12, l.10-12: Can you elaborate what you exactly mean by these "one-to-one mappings" and why this should be the case? This is not clear to me.

> We will rewrite this in more straightforward language. We are trying to say that if our paper encourages discovering EC underpinning beyond physical intuition, instead with a more rigorous mapping to differential equations, then such equations can be characterised by standard mathematic terminology.

Technical Corrections

1. P.3, l.19-20: The second part of this sentence is hard to understand, please rephrase.

> We will write this sentence more simply. We will point out that many ECs relate high-frequency fluctuations for the contemporary period, and for which measurements exist, to slower-changing but important quantities that describe features of future climate change.

2. P.3, l.20-21: This sentence is also not easy to understand, please rephrase.

> Similar to the response above, we will also rewrite this sentence in simpler language, noting that if high-frequency changes in the Earth system are ignored, we may be discarding valuable information that can constrain understanding of longer-term climatological variation.

3. P.5, l.10: I wonder if your notation would be simpler if your variable t represented seconds, not years. Then you could absorb the seconds-per-year factor into the frequency ω and drop all the primes for the heat capacity altogether.

> We will certainly consider this. There is always an attraction, of course, to stay in SI units throughout a manuscript.

4. P.5, l.26: There is a "." missing after “Eq”.

> We will correct this.

5. P.8, l.22: There is a "." missing after the end of the sentence.

> We will correct this.

6. P.11, l.5: It would be good to add a name for the symbol epsilon here, maybe “error term” or similar.

> We will remind the reviewer at this sentence, in words, that this is a “noise term”.

7. P.11, l.16-17: Something is wrong with this sentence.

> We will split this sentence in to two parts, as it carries two messages. Teleconnections can either be constrained by (1) a knowledge of advective winds, or (2) by the differences between two quantities in different locations.

8. P.14, l.17: This reference points to a preprint, please update with the published reference.

> Apologies, we will give the full reference for the Nijsse and Dijkstra paper.

9. Caption of Fig. 1: I think there is a word missing after "This response contains a seasonal (x axis) and long–term (y axis, with seasonality ignored)".

> Yes, the word missing is “variation”. We will correct for this.

10. Caption of Fig. 2: "seasonal" forcing instead of "season" forcing. Second to last line: the "measured" value of ΔT_S.

> Thank you – we will correct both of these typos and with the words suggested.

11. Fig. 2: The argument in the cosine of the response term has a different sign than eq. (10). This does not matter due to the symmetry of the cosine, but should be identical to have a consistent notation.

> Of the two choices, we will change the sign of Eqn (10). Using a plus sign feels more natural for increasing time.

12. Fig. 2: The square root in the denominator of the second part of the response is missing. Same for the x and y axis label in (b).

> This is very unfortunate. We created the .pdfs for the diagrams in python and checked the figures carefully after running our script. I had naively assumed that once a .pdf is built, it is the same on all platforms. Unfortunately, the ESD online submission system removed key

> characters and symbols from the figures (e.g. the explanatory underbraces of equations terms and related text). The correct diagrams are shown below, and this also answers reviewer points 13 and 14.

13. Figs. 1 and 2: The index "p" is missing for the heat capacity. In addition, sometimes the prime is missing.

> Please see the correct diagrams, presented on the two pages below.

14. Fig. 1 and 2: Why are some parts of the formulas underlined?

> Please see the correct diagrams, presented on the two pages below.

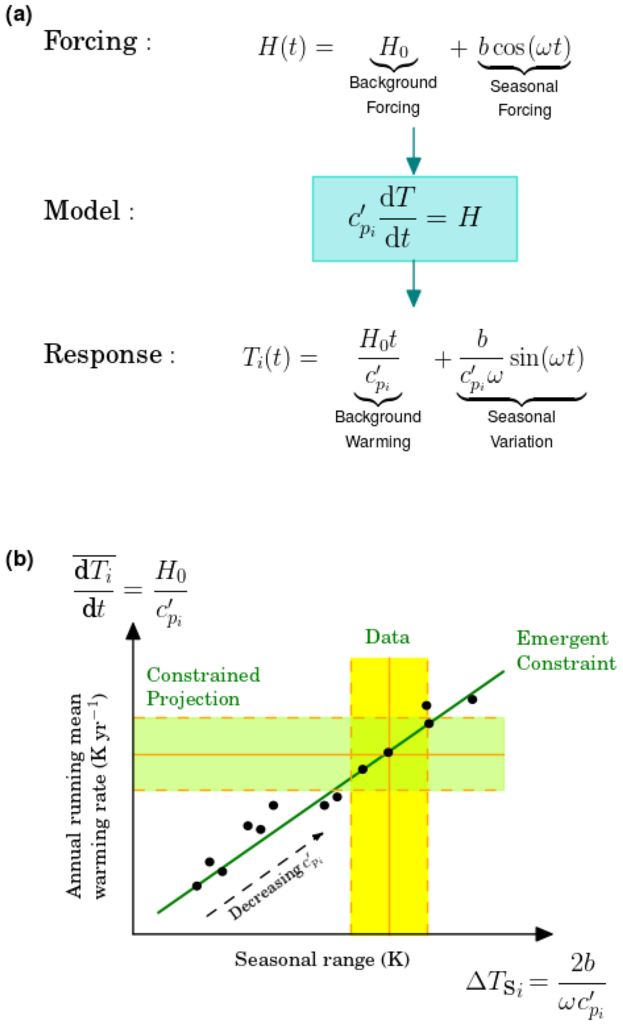

Figure 1. Schematic representation of a simple emergent constraint. Panel (a) (top row) shows the combined equation for long–term and seasonal forcing (so with $\omega = 2\pi$ yr^{-1}) driving the thermal box model given by Eq. (2) (middle row), and the related response to both forcings, which combine additively to give Eq. (5) (bottom row). Panel (b) illustrates a related emergent constraint, based on the response Eq. (5), as also shown in panel (a). This response contains a seasonal (x axis) and long–term (y axis, with seasonality ignored), and the EC links the two. The EC allows the observation of seasonal fluctuations to constrain the long–term rate of change of state variable, T. Each model (black dots, indexed by i) has a different implicit value for c_p' i.e. c'_{p_i}. The EC is assumed to not be exact, with noise causing variation around the regression line (the ϵ_i and η_i terms of Eq. (1)). The vertical yellow band represents uncertainty in the measurement, ΔT_S^*. The constrained projection of the long–term warming rate (based on the EC, the value of ΔT_S^* and its uncertainty) is given by the green horizontal band.

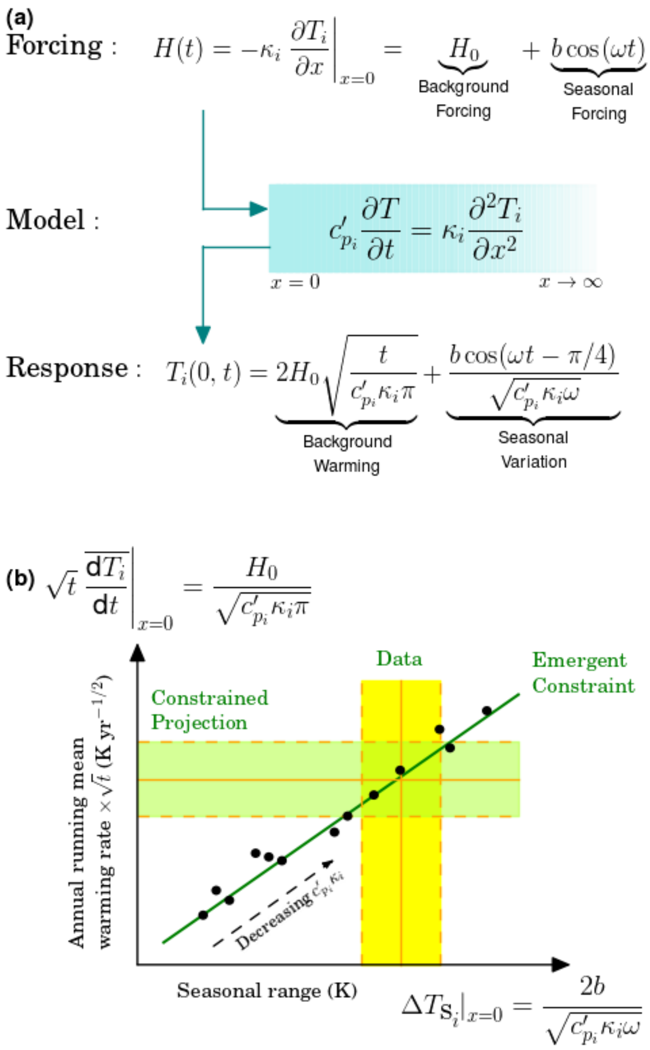

Figure 2. Schematic representation of an emergent constraint with a spatial component. The spatial dimension is defined by x. Panel (a) (top row) shows the combined equation for long–term and season forcing at $x = 0$, driving the diffusive model given by Eq. (7) (middle row), and the related response at $x = 0$ and $t > 0$ given by Eqs. (10) and (14) (bottom row). The seasonal forcing (so with $\omega = 2\pi$ yr^{-1}) is given by Eq. (8) and the long–term forcing to the thermal model given by Eq. (12). These two forcings generate a response in T at $x = 0$ given by Eqs. (10) and (14) respectively, that combine additively and as shown. Panel (b) illustrates the related emergent constraint, based on the response $T_i(0,t)$ shown in panel (a). This response contains a seasonal (x-axis) and long–term (y axis, with seasonality ignored) part, and the EC links the two. The EC allows the observation of seasonal fluctuations to constrain the long–term rate of change. Each model (black dots, indexed by i) has a different implicit value for $c'_{p_i} \times \kappa_i$. As for the example of Fig. 1, the EC is again assumed to not be exact, with noise causing variation around the regression line. The vertical yellow band represents uncertainty in the measurement of ΔT_S. The constrained projection of the long–term warming rate (multiplied by $\sqrt{t}$, and based on the EC, the value of ΔT_S and its uncertainty), is given by the green horizontal band.