Authors' Response to Referee 3 (ESDD esd-2018-71)

March 31, 2019

Please accept my apologies for this very slow response to reviewing: "Limitations of Emergent Constraints on Multi-Model Projections: Case Study of Constraining Vegetation Productivity With Observed Greening Sensitivity" Emergent Constraints (ECs) have become a very popular mechanism to collapse inter-GCM differences, and in order to make more refined future projections. It is therefore highly relevant to verify how robust the methodology is, and/or find counter-examples which illustrate potential issues with the technique.

We thank the reviewer for her/his interesting comments on the fundamentals of the EC method and for sharing the opinion, that scrutinizing the EC methodology towards robustness is highly relevant. Apparently, the reviewer also got the notion that the purpose of this study is to question the general validity of the EC approach (see comment 1.7). This is not case. In fact, we illustrate the range of applicability of ECs and elaborate on caveats and potential pitfalls. We revised the manuscript to avoid such misunderstandings.

1 General Comments

1.1 *This is a slightly superficial review, but what I would encourage the authors to do is to focus more tightly on the issue of potential problems with ECs – maybe at the expense of some of the other text describing so fully the particulars of vegetation greening.*

Yes, we agree with the reviewer. In the revised manuscript we bring the general applicability of EC more into focus. We encapsulated the particulars of vegetation greening, but keep a certain detail, so that the reader can easily follow the narrative of the article.

1.2 *While there are some concerns surrounding the EC approach, some of the criticisms levelled by the authors are only valid if the approach is applied carelessly. So I am not convinced these are limitations, and instead, a better title might be "Careful Application needed by ECs. ...".*

As stated in the response to the reviewer's summary comment, the intention of this study is not to question the general EC approach, but rather challenge its robustness and illustrate caveats. But we agree with the reviewer, the manuscript is mostly dealing with the applicability of ECs and that inaccuracies in the methodology can crucially influence conclusions drawn from the constrained estimate. We adjusted the title in the revised manuscript.

1.3 *The Abstract raises two concerns. "The method critically depends on first an accurate estimation of the predictor from observations and models". This is true, but this is not particular to ECs any more than it is for any other environmental science modelling exercise. It is always essential to ensure that measurements align tightly with models to – for instance – allow model calibration. For example, the need for "like-for-like" comparison is routinely addressed when utilising Earth Observing data to constrain terrestrial ecosystem models.*

Yes, observational uncertainty is an important issue in many statistical methods, and in calibrating or benchmarking environmental numerical models. However, we argue, that the EC method is particularly sensitive to observational uncertainty (P15, L5-6), because the single observational estimate essentially determines the EC. On the contrary, the emergent linear relationship is established based on a collection of multi-model estimates, where each model gets 'one vote' (however, some models might be more influential than others; Bracegirdle and Stephenson, 2012). Thus, the observational uncertainty has a much larger bearing on the EC than the uncertainty of each individual model. To overcome this source of uncertainty, various meaningful observations, if applicable, should be taken into consideration.

1.4 *"Second, depends on a robust relationship between inter-model variations in the predictor-predictand space". This is really what lies at the heart of emergent constraints, which by definition is the search for emerging regressions across "X" and "Y"-axis space. However, if no relationship is present, then clearly the method would not be used. An interesting question to ask, however, is if intuitively a relationship is expected, but is not seen inter-GCM, then what does this imply?*

The reviewer raises an interesting issue here. In the 'search' for an emergent linear relationship, the researcher might stumble upon one or two predictor-predictand combinations that do not show a tight connection, albeit it was expected based on the current process understanding. This implies that (some) models miss or misrepresent the process of interest, assuming that indeed a meaningful predictor was chosen. For the sake of model development and advancement, such non-existent, yet expected, Emergent Constraints should also be reported and scrutinized to find out why the individual models deviate. We implement this aspect in the discussion section in the revised manuscript.

1.5 *The Conclusions are much more nicely set out, and I think clearer to understand. However, to just run through the points raised: (*) The paragraph starting "The importance of how the observational predictor....." again raises the need for all EC modellers to ensure a direct 1-1 mapping between modelled "X"-axis quantity and measurements. The next paragraph correctly identifies the importance of accurate spatial aggregation, when the GCMs themselves are predicting bulk quantities, defined as only valid over large regions (e.g. mean "greening" levels"). A large source of uncertainty is associated with temporal variability". The EC method does account for uncertainty in the measured "X"-quantity, which is why the standard diagrams place bounds on that – in addition to uncertainty associated with the model-based regression. If only one measurement is available, based on averaging over multiple years, then standard statistical techniques can be used to build error bounds. These can include, for instance, sampling only subsets of the years. Methods like this can also be applied where there is a mismatch in window length, to ensure larger uncertainty bounds where the measured quantity is over a short period.*

Yes, the EC method accounts for temporal fluctuations, for models and observations alike. However, the signal-to-noise ratio changes with increasing forcing, i.e. the predictor is strongly influenced by temporal fluctuations at a low CO_2 forcing (Figure 3). If these fluctuations are taken into account for the modeled predictors, but not for the observational estimate (due to the lack of certain observations at low CO_2 forcing), predictor comparability is not given, which results in a questionable constrained estimate.

1.6 *The conclusion hints at the issue of the importance of both identical "X"-axis temporal length (both model and measurement), and additionally the need for identical time-periods. In its most extreme for instance, it would not be appropriate to take 30-year segments of GCM period 1850-1889, comparing to 1990-2019 measurements. This is because an EC can change in time. Such variation is sometimes used to question ECs, but as long as the "X" model and "X" data are for the same period, then the method remains valid. Indeed here, the authors acknowledge dGPP v dLAI_max relationships do change for increasing CO_2 levels. These changes are not a failure of the EC method, simply that (i) timescales need to line up correctly for present day (data versus models), and (ii) users need to be aware of what CO_2 level is being considered for the "Y"-axis. I think the authors might miss a trick here, and especially for vegetation analysis. Where the EC approach is at risk of failure is if all GCMs currently miss an important process, and that will only become critical into the future. One prominent example is where, until recently at least, very few GCMs describe possible future down-regulation of fertilisation through geochemical cycles such as that of Nitrogen.*

Yes, the EC approach is prone to fail if the majority of models miss an essential process, that might critically influence the development of predictor and predictand with increasing forcing. We discuss this aspect in Section 3.6 **Uncertainties in the multi-model ensemble** in the manuscript (P14, L3-4). Also, we addressed the prominent case of potential nitrogen limitation in a high CO_2 world in the companion paper by Winkler et al. (2019); discussed in length in the *Peer Review*

File https://doi.org/10.1038/s41467-019-08633-z. In theory, the EC relationship should approximately stay constant if more models included a well-calibrated interactive nitrogen cycle. For example, if the HadGEM2-ES model implemented a reasonable nitrogen limitation, estimates of historical greening sensitvitiy (predictor) as well as of future GPP increase (predictand) would be lower, thus, HadGEM2-ES would move down the EC slope approaching the constraint estimate (see Fig.2c in Winkler et al., 2019).

1.7 *I certainly do not want this review to appear defensive of ECs, and this is indeed a very interesting and thought-provoking manuscript. There are definitely things that require investigation associated with the technique. It is just that most of the points raised do not invalidate the EC approach – the examples are much more a case of "please use ECs carefully"? Sorry, this is a short review, but if another version if generated then I would be happy to see the paper again.*

Please see our response to the reviewer's summary comment and to comment 1.2. We thank the reviewer for acknowledging again the relevance of this manuscript and pointing out its thought-provoking character. We are also thankful that the reviewer is willing to see the revised manuscript again.

References

Bracegirdle, T. J. and Stephenson, D. B. (2012). On the Robustness of Emergent Constraints Used in Multimodel Climate Change Projections of Arctic Warming. *Journal of Climate*, 26:669–678.

Winkler, A. J., Myneni, R. B., Alexandrov, G. A., and Brovkin, V. (2019). Earth system models underestimate carbon fixation by plants in the high latitudes. *Nature Communications*, 10(1):885.