Earth Syst. Dynam. Discuss.,
https://doi.org/10.5194/esd-2019-90-RC1, 2020
© Author(s) 2020. This work is distributed under
the Creative Commons Attribution 4.0 License.

Earth System Dynamics Discussions

Interactive comment on "What could we learn about climate sensitivity from variability in the surface temperature record?" *by* James Douglas Annan et al.

Anonymous Referee #1

Received and published: 19 February 2020

1 Summary

Annan et al. (2020) examine the utility of using natural variability of global mean surface temperature (or ocean mixed layer temperature) to constrain equilibrium climate sensitivity. They extend recent work on this topic by using a two layer energy balance model in a "perfect model" framework. They find that the strength of variability-based constraints is substantially weaker when climate sensitivity is large, which is true for both simulations with no external forcing and simulations that use estimates of historical external forcing. For moderate climate sensitivity (2.5 K), the uncertainty in ECS is approximately 4 degrees using information from the entire time series and includ-

ing aerosol forcing uncertainty. For simpler constraints, the uncertainty range is even larger.

This work is a useful expansion of recent literature on this topic. The manuscript is clearly written, though I suggest the authors make a minor modification to the organization, expand their discussion in several places, and consider condensing some figures (or adding a summary figure) to help compare results across the various experiments.

2 General Comments

One suggestion to improve the manuscript is to make it easier for the reader to compare across experiments / figures. For example, Figure 1 and 5 are similar and it would be useful to compare all of these results together (perhaps via plotting them on the same axis with color coding in an additional summary figure or grouping or by putting them on a common figure with different panels). It would be similarly useful to intercompare the various posterior estimates (Fig. 2 and 6 as well as 3, 7, and 8). Perhaps plotting lines corresponding to the 5 - 95% CI and a dot for the most likely value value (which would illustrate the skewness) would help compress the figures (though this may run afoul of the Bayesian framework).

There were a few places (noted below) where it would be helpful to more directly compare and discuss this work in the context of other literature. For example, in some places I thought that Cox et al had commented on some issues (e.g., de-trending, two layer models, etc.) and it wasn't immediately clear how to put that work in the context of this manuscript.

I was confused by the ϵ factor you used as well as the references for the two layer model used here (see below). It would be helpful to clarify some of this in the revised manuscript.

In terms of organization, I thought it would be helpful to include the data (CMIP + HadCRUT) somewhere in the beginning (e.g., a renamed Methods section), rather than introducing with the manuscript's results.

3 Specific Comments

Abstract (line 2) and Page 1 / Line 24: In the abstract I wasn't sure which studies you were referring to that tried to constrain ECS with the trend. This might be an oversimplification of these approaches (the Gregory et al. 2002 and Otto et al. 2013 papers cited), since these publications also considered radiative forcing and ocean heat uptake. I believe more recent work by Jimenez-de-la-Cuesta and Mauritsen (2019; doi: 10.1038/s41561-019-0463-y) and Nijsse et al (2020; doi: 10.5194/esd-2019-86) are consistent the abstract language (with caveats that they focus on TCR and a specific time period). I suggest revising this language to reflect the "energy budget constraint" rather than the trend and/or citing these other relevant publications.

Abstract / Line 15: "observed...observational" consider using "inferred from the detrended observational record"

Page 2 (general comment): Other studies that discuss the utility of variability in understanding the climate response to external forcing include Langen and Alexeev (2005, doi: 10.1029/2005GL024136) and Kirk-Davidoff (2009, doi: 10.5194/acp-9-813-2009). The latter publication seems particularly relevant to the manuscript under review and could be compared to the results here.

Page 2 / Line 21: Cox et al replied (Cox et al, 2018b) to these comments with some analyses relevant to this manuscript. In it, they discuss issues such as the importance of de-trending, their own two layer experiments, and the effect of historic external forcing. Given the relevance to this manuscript (e.g., they two box model results), some of this information could be presented in the introduction or at least compared to the two

layer results shown in this work.

Section 2: Consider describing the CMIP data and HadCRUT observations here

Page 3 / Line 7: In Cox et al (2018b), they did test a two layer model (building off their one-layer model results)

Page 3 / Line 20 and Equation 2: I was confused why ϵ did not appear in Eq. 2 or why it wasn't absorbed into gamma in both Eqs. 1 and 2. In quickly looking at the Winton et al (2010) paper: don't they use this term in part because it is a one layer model (line 10 and line 20 seem to imply this was a two layer model, but I realize now this may not have been intended)? Can ϵ be removed here since you explicitly have deep ocean representation? I would appreciate more text justifying / clarifying the purpose of ϵ. It looks like some of what you attribute to Winton et al (2010) should be attributed to Held and Winton (2010, doi: 10.1175/2009JCLI3466.1)?

Page 3 / Line 26: On first read, I thought you had used a value of the average mixed and deep layer depth based on observations. Suggest making this more clear with something like, "We assume a mixed and deep layer depth of 75 and 1000 m, respectively, which are used to calculate the heat capacities (Cm and Cd, respectively) based on ocean coverage of 70% of the planetary surface area."

Page 3 / Line 27 - 29: It would be useful to provide more information about how you chose your parameter values (and later information about how you get the range of plausible values), citing literature relevant to the selection of these values. It was unclear to me why you didn't simply use the mean or median from Geoffrey et al. (2013), for example.

Page 5 / Line 11 - 12: This suggests that you checked this using the two-layer model, consider putting "(not shown)" to indicate that you checked this.

Page 5 / Line 19 - 20: Does detrending (using 55 year windows as in Cox et al) alter the analysis? It seems like it would be reasonable to linearly detrend (as done in the

Cox et al calculation). In the Cox et al reply, they note that it is still important to de-trend unforced simulations using a two layer model.

Page 5 / Line 23 - 29: This is interesting and useful, though I am not sure how to square this with the results presented in Cox et al. (2018). Could you comment more on this? Is this heteroscedasticity included in their estimate of ECS via linear regression? For example, if you varied S to correspond to the 16 models used in Cox et al. (2018), ran a 150 simulation, and performed linear regression (Ψ versus ECS) would the fit be significantly different from what would be obtained using the 1000-year simulations? Or, in another way, is this issue included in the Cox et al (2018) ECS estimate because the increased variance in high ECS models has the effect of making their linear fit between Ψ and ECS more uncertain (and thus contributes to the uncertainty in GCM Ψ values and, in turn, ECS)? Or is the strength of the relationship in Cox et al fortuitous (as suggested by the importance of what GCM simulations are included in the regression as seen in the Po-Chedley et al comment)? Or perhaps the take home message should be that the observed value of Ψ is uncertain. The Cox et al reply (Extended Data Figure 2, top left) suggests that you would generally expect to get a reasonably strong relationship between Ψ and ECS (even though heteroscedasticity should influence their results, too).

Page 5 (Figure 1 discussion): Note that Po-Chedley et al (2018) found that you only recover a strong Ψ - ECS relationship using all of the piControl data and the relationship is weaker with shorter time series. Kirk-Davidoff (2009) also concludes that shorter time series cannot accurately diagnoses climate sensitivity. Of relevance, Nijjse et al (2019, doi: 10.24433/CO.6887733.v1) show that a metric of decadal surface temperature variability (from piControl data) scales with ECS.

Page 6 / Line 31 - 32: It would be useful to provide a reference regarding this point (since the subsequent linguistic calibration was not immediately intuitive).

Page 6 / Line 32: Should this be “Relative” (where it says “Likelihood values can be

read. . .”?

Figure 1 (and others): Suggest adding a legend with “Single layer (gamma = 0)” and “Two layer”. In general, it would be helpful to the reader to have a legend on all of the figures (with the possible exception of Fig. 4) and there appears to be plenty of white space to do so.

Figure 2: Suggest “dashed” instead of “dotted.” The blue appears purple on my monitor.

Page 8 / Line 2: You could cite Roe and Baker (2007; doi: 10.1126/science.1144735) and perhaps others here.

Page 9 / Line 5: I don’t have intuition for what the relative uncertainty should be, but 20% struck me as small (particularly given the later remark that the observational interannual variability looks relatively large compared to the model simulations).

Page 9 / Line 8: Consider using “purple” instead of “blue” so the reader doesn’t get confused with the cyan line (at least this appears purple on my screen).

Page 11 / Line 26: Does this correspond to any published values of the aerosol forcing uncertainty? Later, you say that a larger aerosol forcing corresponds to larger ECS, but this range suggests that you do not, by default, consider larger aerosol forcing - is that right? If so, why?

Page 11 / Line 33 onwards: Could it also be that the noise term in the two layer models is too small?

Page 12 / Line 2: At first I didn’t understand what 0.13 represented. Consider re-writing as something like: For each of the three simulations, the RMS differences between model output with no internal variability and observations is 0.13 oC.

Page 13 Line 1 (and elsewhere): This record contains more than the 20th century. Consider using a more generic term (like historical) and noting the time period considered. Also consider using “first” instead of “firstly.”

Page 13 / Line 4: In Figure 5, you also only show two layer solutions, which is different from Fig. 1 and could be noted here.

Page 13 / Line 18: Should some of this CMIP information be included in the Methods Section (perhaps revised to "Data and Methods")? What RCP scenario was used to extend the historical time series. This would be a good place to cite Taylor (2011, doi: 10.1175/BAMS-D-11-00094.1). There is also some language that is suggested for acknowledging ESGF and modeling groups (https://pcmdi.llnl.gov/mips/cmip5/citation.html). It would also be helpful to say what variable you are using (tas?).

Page 13 / Line 25 - 27 / Figure 5: It is very useful seeing many CMIP5 realizations on this plot. This is a nice illustration of one of your key points (the range of Ψ values can be quite large for a given model).

4 Grammatical / Other Comments

Page 1 / Line 23: Should this be *mid* 19th century to the early *21st* century

Page 2 / Line 1: focussed -> focused

Page 2 / Line 2: Suggest: "and this topic" -> "which"

Page 2 / Line 31: Remove "to" in "from to"

Page 3 / Line 7: Suggest changing "equilibrium sensitivity" to "equilibrium climate sensitivity"

Page 6 / Line 16: Insert: "so *we* perform. . ."

Page 9 / Line 25: I was not familiar with "i.i.d." - suggest writing this out.

Interactive comment on Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2019-90, 2020.