Answer to Anonymous Referee #1

We thank Referee #1 for their careful review of our paper and for the improvements and revisions suggested in their comments. In the following text, we answer to all points discussed by Referee #1, where Referee comments are written as R: and authors comments are written as A:.

R:
In this paper the authors develop a novel technique to combine emergent constraints. Their main step forward is reconsidering the emergent constraint regression as a likelihood model so that it can be combined with a prior, allowing for Bayesian updating. This is particularly important for estimates of climate sensitivity, whose IPCC range has barely changed since 1990, even though independent lines of evidence have strengthened. The technique is elegant, transparent and I wish I'd come up with it. The accompanying code is also clear.
I suggest the authors clarify some of their text and if available include more PMIP4 models.

A:
More PMIP4 models will be added in the revised paper, as several Pliocene and LGM simulations became available during the reviewing process. So far, we obtained 3 more models (with the possible addition of a fourth model).

R:
Minor comments:

A:
Syntax and choice of words will be changed accordingly to the minor comments. Clarifications and change of jargon will be applied when asked to. Specifically, clarifications will be added following these comments:

> R: 11: it's not a 100% clear whether this is a combination of the restricted ensemble of the nonrestricted ensemble. Either clarify, or remove the unrestricted estimate altogether.
>
> A: We kept the numbers, but will add the clarification on the use of the restricted ensemble here. We remind here that restricted ensemble refers to an ensemble consisting of the latest version of each model.
>
> 16: I don't quite understand the last half of the sentence: "higher bound by construction"
>
> A: Here we meant that the 5th percentile is higher than expected using the OLS method only because of geometrical reasons: If the correlation is weak, then the slope of the regression line is usually reduced and therefore the 5th percentile is higher. This will be made more explicit in this part of the abstract.
>
> We suggest the following correction:
>
> "An interesting implication of this work is that OLS-based emergent constraints on ECS generate tighter uncertainty estimates, in particular at the lower end, suggesting a higher percentile value due to a flatter regression line in case of lack of correlation."
>
> 104: I didn't quite understand what "percentage of intervals to contain .." means. Please clarify.
>
> A: We agree this specific statistical jargon can be clarified. What is meant here is that in frequentist confidence intervals, on 100 random 90% confidence intervals arising from separate experiments, 90 intervals would be expected to contain the true interval bounds, which is different than credible intervals, where we believe with 90% probability that the truth lies in the specific interval obtained.
>
> We suggest the following correction:

“The former is the representation of the number of random intervals to contain the true interval bounds (at 90% confidence, this would lead to 90 out of 100 random intervals to contain the true bounds), while Bayesian credible interval is an interval which we believe (with the given probability) to contain the truth.”

126: typo: roles

139: “observation operator”. Operator is unnecessary jargon.

A: Removed (both).

169: A two line explanation of a (one step) Karman filter might benefit readers.

A: We agree, it will be clarified.

182: Phase 4 of PMIP are used in the study. Please replace explanation by saying not much data is available instead of none.

A: Corrected.

236-237: I don’t think it’s necessary to include this test any more.

A: We assume Reviewer #1 is referring to the test of other MCMC methods (lines 238-239) which, indeed, can be removed from the study as it is quite trivial.

291: I’m quite surprised that OLS is more tight. Could you check code or provide an explanation?

A: We tried to explain this in the following line (L292 - 293), although it seems to be a limited explanation, which will be therefore extended. The reason behind OLS being tighter is similar to the “higher bound by construction” comment. If the correlation is weak, the OLS will create a regression line that usually has a 90% interval expanding on the range of the ensemble of models. This is not the case for Bayesian regression, as in case of low correlation, the influence will not come from the ensemble of the models but from the prior (here, a Cauchy prior), which has a very wide 90% interval. Therefore, when the correlation is so weak, such as with PMIP3, OLS is much tighter than the Bayesian method. However, it does not imply that either range is closer to reality. They are both different representations of the uncertainties behind each method. The code has been checked and it seems the intervals computed for the OLS method were different than what we sought to represent here. The computations of the intervals were changed to follow the same method as introduced in previous studies (Hargreaves et al., 2012; Schmidt et al., 2014, Hargreaves and Annan, 2016), which is to generate a predictive ensemble aimed at representing the uncertainty in the tropical temperature. This method is actually the one used to generate the 5-95% interval of the Bayesian method. Nevertheless, the “new” intervals are only slightly wider, which leads us to exactly the same interpretation of methods comparison as commented above: Bayesian intervals tend to be wider in case of lack of correlation.

We suggest extending the explanation at lines 292-293 with the following:

“As previously argued for the combination of PMIP2 and PMIP3, the OLS produces a tighter posterior range. In the absence of a correlation, the Bayesian method relaxes to the prior, whereas the OLS method is heavily influenced by the range of the ensemble. However, we emphasise that this does not suggest that either range is closer to reality.”

Additionally, the intervals related to the OLS method in Table 1 and in the text will be updated to the new intervals (3 values).

R:

349-350: a logical extension of the methodology is to apply it to CMIP, where we find many emergent constraint on the same models. It would be nice if the authors could comment on whether they see this as a problem, given that these models may have similar systematic biases.

A:
This is an interesting point that will be emphasised in the revised paper. The method was originally designed to be used for different emergent constraints, although we eventually decided to focus on past climates to illustrate the method. There should not be any problem in using CMIP models, as long as the emergent constraints relationship investigated is physically plausible.

337: merely → nearly or almost.

A: Corrected.

374: add 'in a systematic way' or something similar. The principle behind emergent constraints relies on the fact that models deviate from reality, so that's not the problem.

A: Added.

386: pertinent → why not use simpler word such as relevant.

406: ordinary least squares doesn't require capitalization

A: Corrected.

Fig1 caption: what is a 'wide' ensemble proxy?

A: Changed to "multi-proxy ensemble".

R:
Fig2 – Fig9: in the pdf, the colour orange might imply to a tired reader that only PMIP3 is used from the figure on the left. Purple or other dark colour might be more clear. I'm not convinced that all figures are necessary for the paper. The summary in the table may suffice for more regressions, such as the one in Fig 9.

A:
The colours of the figures will be changed accordingly. We agree that not all figures are necessary. Thus, the number of figures will be reduced, as a lot of outputs are summarised in Table 2.