Answer to Anonymous Referee #2

We thank Referee #2 for pointing out sections of the paper that require clarification and providing suggestions and constructive criticism for improvements. In the following text, we answer all the points discussed by Referee #2, where Referee comments are written as R: and authors comments are written as A:.

R: The paper by Renoult et al presents a new Bayesian method for dealing with emergent constraints for estimating climate sensitivity from palaeoclimate model simulations. I have little expertise in the use of emergent constraints so I will concentrate my comments on the statistical methodology used. For such a simple approach they have made their technique remarkably opaque. For this reason it is hard to recommend an editorial decision for this paper - I will leave it up to the other reviewers to determine novelty and suitability for this journal. However I think there needs to be a considerable improvement in the explanation of the mathematical approaches.

If we start with the OLS method, we have a data set Si,Ti, for i = 1,...,n simulators where we use the model:
$Si = \alpha * Ti + \beta + \varepsilon$
And obtain estimates of alpha, beta, and the residual standard deviation sigma. A user comes along and provides us with a new value $T*$ and we obtain $S*$ from the fitted model. Uncertainty arises from the potential uncertainties in the estimates of the parameters, and the choice of whether prediction or confidence intervals are used.
So far so good. The authors point out that the Bayesian approach is often superior to these traditional models because of its more sensible handling of uncertainty and the allowance of brining in external information in the form of prior distributions. I agree totally.
Unfortunately here is where things get a little more confusing. The authors then state that the model they want to fit is:
$p(S|T) = p(T|S)p(S)/p(T)$,
i.e. a standard application of Bayes' theorem which provides us with a posterior distribution of S given T. This is where the notation starts to get into a bit of a mess, because now we're not told where the observations fit in to the model. My guess is that what the authors mean in the above equation (using my notation) is actually:
$p(S*|T*) = p(T*|S*)p(S*)/p(T*)$
Where the likelihood p(T*|S*) is actually integrated over the posterior set of parameters
from a new linear regression model
$Ti = \gamma * Si + \delta + \gamma$
Where I've named these new slopes/intercepts differently to highlight the different from
the previous OLS approach.
This is a more complicated model, and most of has come from guesswork because the authors haven't provided enough information for me to work out exactly what is happening. I'd really appreciate the authors doing (the quite large job) of either clearing up their maths or making sure that my incorrect assumptions are not made by others.

A: We agree with Reviewer 2, as it led to some confusions also pointed out by Short Comment #2. The description of the model(s) by Reviewer #2 is, however, completely accurate and we thank them for giving us insights on how to clarify our study.
Indeed, the likelihood is built as an integration over the parameters alpha, beta and epsilon. Therefore, the computation of the likelihood, and the updating of the prior, actually calls two different Bayesian processes.

Following the notations of Reviewer 2, the model could be written as:
- The likelihood is a regression model defined by $T = \gamma S + \delta + \varepsilon$ with the triple of uncertain parameters (γ, δ, ε) which are conditioned on the model ensemble, where T and S are respectively the temperature and sensitivity of a given climate model. Consequently, the likelihood $p(T^* \mid S^*)$ for a given S^* is an integration over the S^* posterior distribution of T^* predicted by this regression relation (convolved with observational uncertainty), (γ, δ, ε), conditioned on the model ensemble, where T^* represents the observed (geological) value.
- The Bayesian updating of S corresponds then to:
  $p(S^* \mid T^*) = p(S^*) * p(T^* \mid S^*)$

As accurately described by Reviewer 2, the computation of the likelihood here does not take place within the computation of the emergent constraints model, but forms part of the Bayesian updating, which uses the emergent constraints model.

The notation in the paper will be revised and more clarifications and description of the two stages of the process will be added.

R: The paragraph in the intro which starts "Two recent papers have also addressed. . .” makes some odd statements about KFs. It points out that everything is Gaussian then states that “it is fairly difficult to generate posterior values which are outside of the prior range”. This seems surprising if everything is Gaussian. I haven’t read the other paper so perhaps explain more clearly?

A: The words we used here were likely too vague. It is true that posterior values could lie outside the prior range. However, it is usually difficult if the observed value is rather uncertain and/or close to the prior mean (which is the case here). We should have rather said that most of the posterior values are generally within the prior range, and this will be clarified in the revised paper.

We suggest the following correction:

“In particular, most of the posterior values would lie in the range covered by the ensemble of models if the observed value is either uncertain and/or close to the prior mean. This is a direct consequence of the joint probability distribution produced by the Kalman filter, which in the case of joint Gaussian distributions, will produce a tighter posterior Gaussian distribution.”

R: The first sentence in the methods section involves, a load of unnecessary commas, which, in my view, makes the sentence, and hence, the definition, of the key concept, of emergent constraints, very hard to understand. There must be a simpler way of writing it.

A: This sentence will be clarified to make it easier to understand.

We suggest the following:

“The general method of “emergent constraints” seeks a physically plausible relationship in the climate system between two model variables in an ensemble of results from different climate models. Consequently, an observation of one measurable variable (such as past tropical temperatures) could be used to better constrain the other investigated variable, usually unobserved and difficult to measure (such as climate sensitivity).”

R: Also in that sentence it says ‘. . .’then an observation . . .’. An observation of what?

A: Here we refer to any kind of measurable quantity in the climate system. For instance, an “observation”, or shall we say a measure of the past temperature of the LGM. This will be made more explicit (see suggestion above).

R: L95 should be N(0, σ2) to match standard notation. This mistake is made throughout. There’s similarly a bizarre use of ∈ from set theory to write ε ∈ N(0,σ) which I think should be ε ~ N(0,σ2) everywhere.

A: We argue here that σ2 could be confusing in some cases, as it implies the user to take the square root to know what σ is, which could make it difficult to estimate 2*σ range, or how many σ lies within a certain range. To match standard notation, we will explicitly write the variance as σ^2, i.e. if N(m, 0.5^2) rather than computing it as N(m, 0.25).
For the use of ∈, this will be changed to ~.

R: There seems to be a kind of deeper issue that perhaps should be mentioned somewhere that these regression approaches really should involve measurement error (separate from model error as in the Williamson/Samson) paper. The literature on this is well-developed and is pretty easy to include in Bayesian models.

A: This is indeed a relevant point that will be mentioned in the text. We have in fact accounted for observational uncertainty in the natural way in the likelihood, although this was not clearly described in the text.
For the linear regression, it is true that uncertainties exist on S and T of the models, and should, in theory, be included in the Bayesian Linear Regression. However, these uncertainties come from computation methods and are in general very small. For instance, we estimate the measurement error on S to be close to 5% of each value and this is small compared to other uncertainties. In particular, observational errors on S and T from models are small compared to the structural differences that are responsible of the dispersion of the points around the regression line. This point was argued in Hargreaves and Annan (2016) regarding the use of S and T of the mid-Pliocene, where the errors were shown to be small enough to be ignored without a significant impact on the outputs.
Additionally, the strength of the Bayesian Linear Regression is that it does not create a single "best" line, but multiple lines that include all these uncertainties. For instance, it is very likely that multiple lines are already drawn within the range of measurement error of each model. Consequently, we expect the posterior outputs to display variations which could be much less than 0.5 K on both percentiles combined.

R: L158 the use of sequential updates appears for the first time but I can't really see why this is relevant or used elsewhere?

A: Removed "sequentially".

R: The Kalman filter method seems like a really important rival approach but despite being given a full subsection 2.3 this only has one paragraph and no mathematical definition. It would be nice to be able to compare the approaches more clearly

A: The paragraph about the Kalman filter will be extended to give more details, as also requested by Reviewer #1. However, we do not consider the Kalman filter presented here, i.e. a one-step Kalman filter, as an important rival approach, as we state using the climate model ensemble as a prior is in our opinion too strongly restrictive for the question of emergent constraints. We will emphasise this latter point.

R: PlioMIP appears in L200 without being mentioned before

A: Corrected.

R: There really is very little need to use a complex Hamiltonian Monte Carlo method like NUTS on a simple linear regression problem. With a small change in prior from Cauchy to Inv-Gamma the whole problem would become conjugate and could be done exactly on a pocket calculator.

A: There are (at least) three reasons behind the choice of NUTS. The first one is that we consider the Python package PyMC3 explicit and well-described, even for users without statistical background. Thus, using the by-default methods of PyMC3 such as NUTS allow future users to have access to a wide range of online help. The second reason is that the method presented in this study aims at being used in a wider context, i.e. more complex regression problems. The third reason is that using NUTS actually avoids using conjugate priors. We believe that adding more mathematical restrictions on the forms of prior would make the problem less relatable to reality where prior choices may be physically motivated.

R: L270 the posterior distributions of what?

A: Posterior distribution of S (sensitivity). Will be corrected (and other mentions thereafter).

R: L273 and elsewhere. There are some weird mentions about Cauchy distributions having a finite integral whilst Uniform distributions do not. This makes no sense to me (the Uniform is only an improper prior if it has infinite limits). None of this is referenced so needs clarifying.

A: This will be removed as it, indeed, makes little sense in the context.