Review of *Towards robust community assessments of the Earth's climate sensitivity*

General Comments

The authors present a nice overview of how Bayesian statistics can be used to make an assessment on climate sensitivity. Importantly, they discuss the various choices taken to make inference on climate sensitivity and how the choices affect the resulting estimate. While the paper is written well and most information is presented well, I feel the key concepts are somewhat obscured by their notation and lack of clarity. Below, I include some specific examples that I feel could be improved upon.

Specific Comments

1. Numerous terms are undefined. For example, in Eq. (1), what is $\mathcal{M}_0$? (It is not defined until two pages later). Same with $\mathcal{M}_\alpha$ and other models. Also, while it may be colloquial for some, $F_{2\times CO2}$ and $\Delta T_{2\times CO2}$ are undefined.

2. Various mathematical terms are undefined.

   - Line 105 - What is $\mathrm{N}(-8.43, 2)$? I assume you mean a normal distribution with mean -8.43 and variance 2. Also, math cal font is used for N in section 7 and not for line 105.
   - Line 201 - What is $\Delta T'(x)$ and what is x?

3. In multiple areas the authors discuss the idea of reducing uncertainties but do not explicitly say which uncertainties are being (or will be) reduced. I think this is a common misnomer when discussing statistical concepts. Some uncertainties are irreducible, such as types of data uncertainty, and some are reducible, such as types of model uncertainty. A key concept in statistics is being able to identify each type and providing the best possible quantification of each - e.g. an appropriate quantification of irreducible uncertainty and reducing all other uncertainty if possible. I feel strongly that the manuscript would benefit if this distinction were made clear.

4. Line 60 - "... update our prior beliefs $P(\Theta)$... " Are you making the distinction that you can update you $P(\Theta|Y)$ if new information becomes available? If so, I think this needs to be reworded. Otherwise, I do not believe Bayes' Theorem says update our prior beliefs. Instead, if you have a prior belief or knowledge, $P(\Theta)$, you can get an estimate of the probability of Θ given data/evidence Y using that prior knowledge. If more data becomes available, you could then refine that belief and use a new prior. This is distinct from updating your prior belief.

5. Section 2 Analysis framework - This is a crucial section for your paper and I would like to see it expanded. Throughout the rest of the paper, terms like *posterior*, *marginal*, *joint probability density*, ... are used but not defined. The general reader of ESD may be unfamiliar with these terms. The latter parts of the paper would be easier to follow if these terms are defined in the context of section 2. Additionally, how does one get $P(\Theta|Y)$? Is it as simple as writing distributions down and using Bayes' Theorem? What if the distribution is not tractable, how would that be handled? Expanding on some of the steps needed to make inference on Θ within

this section will help orient readers as to why this is such a difficult and important problem, and how they can take what you have shown and apply it to their own analysis.

6. Line 120 - The notation surrounding this equation is confusing. It appears as though you are treating $Y = (\Delta T, \Delta F)$ as normally distributed random variables where the mean and standard deviation of each are estimated from experts. You define the joint probability density of ΔT and ΔF as $\mathcal{J}(Y) \equiv \mathcal{J}(\Delta T, \Delta F)$. You then *marginalize* over ΔT and ΔF and somehow get a likelihood of the evidence as $P(\Delta T, \Delta F|\lambda)$. However, this equation (A.3), does not contain either ΔT or ΔF because they have been marginalized out. Instead it contains their mean and standard deviation (assumed fixed?) that are estimated by experts. My confusion is in your definition of evidence and how ΔT and ΔF (or their mean and standard deviation) play a role in that evidence. I think this could be fixed by being more clear on your notation and the steps taken to arrive at the equation on line 120 (and subsequent equations).

7. Model $\mathcal{M}_\alpha$ - This is slightly confusing to me. By definition, $\alpha = \partial\lambda/\partial\Delta T$ is a function of λ. However, you assign independent priors to λ and α when α is constrained by the value of λ. Is there justification for specifying independent priors here? Or is this done for illustrative purposes? If so, I feel it is important to note they are not independent.

8. Eqn. (7) - Same as above comment except now for $\Delta\lambda$.

9. Section 7.1 - I rather like this section and I think it puts a lot of the paper into context. However, I feel as though some terms are not defined and potentially unknown to the general ESD reader. A (Bayesian) hierarchical model is left undefined and for the reader to interpret. Generally, a BHM is defined in terms of data, parameter, and sometimes hyperparameter models. It might help contextualize your message if you state what a hierarchical model is in terms defined from section 2 and then connect it to equations (9)-(12).

10. Line 261 - This is a very bold claim. I would argue it is *one* useful application of hierarchical modeling, but maybe not one of the most.

11. Line 396 - C is already defined as the curve.

Technical Comments

1. A few citations have typos in or surrounding how they are placed in the text

2. Figure 1 - needs labels

3. Line 9 - multiple twice

4. Line 234 - Missing)

5. Line 241 - "... that this is are the ..."

6. Line 243 - Missing)