Anonymous Referee #2

Quilcaille et al describe a large number of simulations conducted with the OSCAR simple climate model, covering some fraction of the CMIP6 experimental design. The paper briefly describes the model and the calibration strategy, and then discusses a number of applied cases.

Clearly a large amount of work has gone into this paper, singlehandedly simulating a large fraction of the CMIP6 experimental design in a single study. However, a side-effect of this is that the paper is somewhat less than rigorous in describing the calibration process and in providing a detailed assessment of the applied experiments in the context of available ESM simulations. It is also sloppily written in places - with a number of spelling errors and imprecise language. Figures require axis labels and subplot labels throughout.

My recommendation is that the scope of the paper is reduced - but greater emphasis is placed on a detailed description of the calibration process, and the sensitivities of simple scenario projections to aspects of that calibration process. Additional experiments (SRM, CDR etc) - are secondary and could be covered in follow-up dedicated studies when the fundamental probabilistic setup is well defined.

Major issues:

1 - More detail is needed on the basic model structure. This version of the model is not documented in the literature, and the first section needs to give a basic overview of the level of complexity being represented. A conceptual figure illustrating the number of domains, and how they interact, would be appreciated.
2 - Perhaps the key aspect of this paper is the effect of observational constraints on projected future climate. However, this is covered quite briefly, and the many degrees of freedom in the calibration process are not comprehensively explored. How do different observations constrain projected warming independently and combined? How are constraints objectively combined? How are prior distributions decided? The paper refers back to Gasser (2017), but the addition of additional constraints in the present paper requires a more detailed description.

3 - The language is often far too vague on key details, like exactly how CMIP5 data were used: (e.g. “The preindustrial state of the land carbon cycle is calibrated against TRENDYv7 and its transient response to CO2 and climate is calibrated against CMIP5 models”). The key thing to communicate in this paper is exactly what data and models were used to calibrate OSCAR, what parameters are being calibrated and what is the sensitivity of the model to each piece of information.

3 - The exclusion of unstable parameter combinations is understandable - but the conditions for instability need to be more objectively quantified. Plots illustrating the instabilities, and conditions for exclusion would be appreciated, together with a process assessment of why they occur and whether the exclusion process might be biasing the observationally constrained distributions.

4 - The joint constraint on cumulative carbon uptake and warming results in a tightening of a prior distribution, which was already narrower than the observational uncertainty (Figure 1 (b?) - sublabels are required!). This tightening occurs due to the effect of the constraint on warming, but means that the model ensemble is not plausible sampling solutions with low carbon uptake. This is a potential bias in the assessment of the model distributions, unless we have perfect confidence in the model structure - which we don’t. The constraining approach would benefit from having a parameter which allowed for model imperfection in calibration (see McNeall 2016 for an example of this calibration problem and Williamson 2019 as an example of a statistical framework to address it).

5 - It would be useful, throughout, for plots to show CMIP5 and CMIP6 distributions where available in addition to the OSCAR distributions.
