This paper attempts to introduce members of the paleo-modelling community to uncertainty quantification, in particular using Bayesian statistics. It is in part pedagogy, in part a review, and in part a series of suggestions and recommendations to modellers. I am a Bayesian statistician with extensive experience in the climate sciences and so will be reviewing this paper primarily from a statistical point of view. I agree with the authors on the need for a positional/translational piece of writing to serve as a guide for modellers – statistical analysis isn't easy, especially for those who do not necessarily have formal mathematical training. Despite the want for this paper, I found it long, dense and difficult to follow, largely containing opinions rather than evidence-based guidance, and ignorant of the existing statistical literature. This paper has been previously submitted for publication (a public process in Climate of the Past), and previously reviewed. I have reservations that many of the previous reviewers' comments have not been sufficiently addressed, and find myself agreeing with the previous reviewers on many points. For this review, I have detailed a number of major concerns (in no particular order) and defer any minor editorial concerns for a later submission.

1. **Length.** I struggled to get through the paper (and, in fact, was not able to without multiple sittings). 22 pages are spent introducing facets of the Bayesian framework after which there were a short(er) 4 pages on some useful techniques and then 7 pages of, from what I can tell, opinions with no examples. There are many points which are belaboured (such as the super-strong transponder throwing arm – which does not work as an analogy for me) and unfortunately sections are lost that are valuable to the intended audience (e.g. paragraph starting line 284). I would recommend drastically shortening the paper, particularly section 2, to only retain concepts key to the narrative, and cite the many existing introductory texts for the interested practitioner. For example, I don't think it is necessary to the main text of the paper to include 1.5 pages that build Bayes rule from conditional probability laws. Further, I don't see where MCMC is needed after it is introduced. MCMC is one of the most difficult early conceptual hurdles for the Bayesian statistics practitioner; introducing MCMC and letting it burden the mind when it is not needed is unnecessary. In fact, I struggle to see where any of the mathematics are used in Sections 3 and 4 of the paper. There are statements made in Section 4 such as *"…challenges and fragility of full Bayesian inference…"* where you go on to then recommend history matching, in effect, doing away with pages of statistics that you have tried to introduce. If we're just going to history match why do we need MCMC and posterior predictive distributions? In fact, why do we need Bayes rule at all? All we need is subjective probability and a mild understanding of uncertainty.
2. **Writing and grammar.** As a whole I find the paper to be quite sloppy. To name just some of my concerns, title cases are off, table formatting is inconsistent, there are erroneous parenthesis, the figure fonts are massive, and Tony O'Hagan needs an apostrophe. Some grammatical and editorial errors are inevitable, but I would expect a more thorough edit to be conducted before the submission of a journal article (especially before the second submission). I would also check your usage of colons throughout, the clause preceding the colon should be a complete sentence: see, for example, line 395. Also, the authors' colon spacing (with a space preceding the colon, and only sometimes) is foreign to me and I can't see a similar example in either US or UK style guides. Finally, the writing reads like a series of dot-points that

ramble on rather than with any real narrative or structure. I keep finding myself lost and have to remind myself of what is happening and where am I going. I find this concerning given that I am already familiar with most of the content and applications.

3. **Literature.** There is a lot of literature in statistics on calibration of computer models, prior selection and elicitation, emulation, modelling simulator discrepancies, as well as accessible introductory texts on Bayesian analysis. Not much of this is cited. Some of the more statistical texts are potentially inaccessible to the uninitiated; however, to ignore them does the paper a disservice (I am often surprised by what a motivated student can learn). I also feel the paper would benefit by acknowledging, in text, what applied work is being conducted, what is it doing well, and what is it missing (with appropriate citations). A dense table of citations is hard for the human brain to process, and despite my best efforts I still don't have an appreciation for what the authors are trying to say in the tables.
4. **Lack of examples.** This paper is quite dogmatic about the need to do a seemingly arbitrary sub-set of modelling stages but does not provide an example of it actually being done. Surely the authors have a more relevant toy simulator (that is not a linear model) that can be used to discuss the statistical concepts and also provide a helping hand to the struggling reader. In reference to point 2, adding an example that traverses the paper will add narrative and continuity.
5. **Feasibility.** I have a fundamental problem with the feasibility of some of the methods recommended. In my experience climate models are not computationally cheap to run, although I admit that the authors have more experience than me here and so could perhaps provide run times of certain simulations that are valuable to the community. In the internal discrepancy section the authors recommend that the parameter space is effectively explored and for each of these points the boundary conditions are sufficiently explored to accurately calculate potentially quite large variance covariance matrices. The authors admit that MCMC samples can take 10s of millions of runs, and then further we are required to predict from the simulator to obtain the posterior predictive distribution. The process of (1) exploring the parameter space in the order of millions of times, (2) exploring the spatio-temporal boundary conditions at each of these millions of locations, and (3) predicting and calibrating from these models seems computationally demanding in the extreme. Further, the boundary conditions are generated from *"adding appropriately correlated noise"* – in my personal experience this process is not simple and to cast it as such is wrong. Accurately representing, modelling and quantifying uncertainty for many spatio-temporal processes is exacting on even the most seasoned statistician and I think at least a nod to this should be included. Later in the paper, emulators are thrown into the mix with no explanation or introduction. Do your methodologies need emulators, and if so where and why?
6. **Point of the paper.** I feel the paper tries to do too many things, and so falls short on each of them. It attempts to provide textbook level mathematics on fundamental statistical principles, ground said mathematics in the application, and then provide recommendations and guidance. I know that the authors have tried to write a document that is accessible to a non-mathematical audience, but by taking the middle ground and then trying to teach mathematics it distracts from the rest of the paper. I would recommend removing as many equations as possible and instead

focusing on what the point of the mathematics is. The equations and the rigor can instead be deferred to a supplement for people to read once they've decided that it is worth the time and burden to learn and implement probability theory, MCMC, history matching, and the other number of techniques mentioned.

To reiterate, I agree with the need and usefulness for a document to teach fundamental uncertainty quantification techniques to a modelling audience. In my experience there is much appetite for the adoption of such methodologies in the modelling community. Both authors are senior and respected members of their respective scientific communities with invaluable experience in this space and so are in a good position to write such a document. Unfortunately, I do not think that the current version of this manuscript does a good enough job. I recommend a significant re-write that prioritises the point of the mathematics, and not equations for equations' sake; that has narrative and verve rather than 40+ pages of dense writing; and that leads by example with some demonstration of feasibility so that we are motivated to follow.