Earth Syst. Dynam. Discuss.,
https://doi.org/10.5194/esd-2019-33-AC2, 2019

Earth System Dynamics
Discussions

ESDD

Interactive comment

Interactive comment on "Bayesian deconstruction of climate sensitivity estimates using simple models: implicit priors, and the confusion of the inverse" *by* James Annan and Julia Hargreaves

James Annan and Julia Hargreaves

jdannan@blueskiesresearch.org.uk

Received and published: 13 September 2019

Thank you for the interesting comments.

There appear to be several distinct issues raised, namely:

(a) definition of likelihood (as discussed on most of p1–2 of the review)
(b) straw-manning (p2)
(c) excessive/indulgent detail (last paragraph p2)

Addressing these in reverse order:

(c) We wish we could agree with the reviewer on this point, but this introduction was recently expanded at the specific request of a reader who found a previous shorter version of the manuscript rather too terse. Reviewer #1 has also asked for some additional explanation. It may seem a rather unsatisfactory state of affairs but the fact is that these sort of calculations are routinely carried out by a wide range of researchers who are not going to go away and take undergraduate statistics classes or read statistics textbooks however desirable this would be. We have presented evidence that misunderstandings associated with the confusion of the inverse are widespread, perhaps even ubiquitous, and it is not uncommon for undergraduate-level teaching material to be misleading on these issues. See eg p116 of the first edition of Wilks 'Statistical methods in the Atmospheric Sciences' where it is said of classical frequentist hypothesis testing: "If the test statistic falls in a sufficiently improbable region of the null distribution, H_0 is rejected as too unlikely to have been true given the observed evidence". As another example, the STEPS glossary at http://www.stats.gla.ac.uk/steps/glossary/confidence_intervals.html says "A confidence interval gives an estimated range of values which is likely to include an unknown population parameter". Since statistical authorities can make such confusing statements we hope we can be indulged with a bit of extra commentary explaining why this is wrong. In our experience, a large majority of scientists (including ourselves) have often misinterpreted frequentist confidence intervals in this way, and it takes quite a bit of time and care to explain why this interpretation is invalid.

(b) On the issue of straw-manning, we broadly agree with the reviewer that the manuscript would read better if we were less dogmatic in imputing motive and/or belief to authors. To that end, we agree that the paper would be better edited along the lines of the reviewer's "if paper X meant Y . . . " construction and plan to revise it throughout along these lines. However, in our defence, we should mention that our manuscript was originally motivated by discussions with some of the cited authors who had observed the discrepancy arising from the two calculation methods using otherwise identical numerical values and wondered if there was a way of reconciling the two approaches and/or a route to interpreting the 'sampling the pdfs' method within the Bayesian paradigm. It is very clear from our discussions with them, that they had indeed been working under the assumptions as we have stated, namely that an uncertain observation can be assumed to be a probabilistic estimate for the measurand in the way described. Moreover, this manuscript has also been widely circulated to a wider range of relevant researchers (including several more of those cited) for their views and none of them complained that we were incorrectly putting words into their mouths. Nevertheless, we acknowledge that it would be better written in a more neutral manner.

(a) The most fundamental criticism appears to be:

However, I believe that the core underpinning assumption that the authors make, namely that interval estimates given by researchers on observations should be treated as likelihoods within a Bayesian framework, cannot be justifiably imposed on a researcher and makes little sense even if it could be.

The reviewer appears to endorse our introduction using the general notation of x_T and x_o in the measurement model, up to and including the resulting definition of likelihood on line 20 of page 3. They then criticise our interpretation of temperature estimates ΔT. Our intention was for this interpretation to be mathematically identical to that of x, merely making the formal substitution of notation in order to pass from the general to the specific case. Therefore, we are unsure what to make of the reviewer's comment "This cannot be true as is, because the likelihood is a function of the unknown ΔT_T, and an estimate is a given interval." When researchers present an uncertain estimate in interval form $\mu \pm \sigma$, in the vast majority of cases they don't simply mean to convey an interval $[\mu - \sigma, \mu + \sigma]$ and nothing more, but almost invariably have a measurement model similar to that of equation 1 in mind and are merely using the interval as a convenient notation to summarise the observed value and magnitude of its uncertainty. This can be seen very clearly in Mauritsen and Pincus (2017) where this interval notation (in their case presented as as 5–95% intervals) is used widely to represent Gaussian distributions. We don't believe this is at all unusual or controversial but will happily note it as an explicit assumption of our analyses. We also admit the existence of some observations which do not fit to our simple model, one example of this could be where the magnitude of observational uncertainty is not assumed to take a constant known value but instead varies with x_T. However we do not believe that this is relevant to the observations discussed in this manuscript.

Therefore, while we do agree that we cannot automatically impose this type of likelihood on researchers in all cases, we don't understand the reviewer's comment (eg in their first paragraph) that "it makes little sense to do so", at least in the specific situations we have described.

A typographical error in our manuscript may have helped to cause confusion. Where we wrote $\Delta T_o \sim N(0.77, 0.08)$ on p11 this would better have been written as $\Delta T_T \sim N(0.77, 0.08)$ or otherwise reworded in such a way as to make the meaning clearer. ΔT_o is simply a fixed known value (once the observation has been made) and has no non-trivial distribution. Thus the reviewer's comments referring to a distribution for ΔT_o have no applicability. There is no such distribution (at least once the observation has been made).

In summary, we would like to submit a revised version of the manuscript in which we make clear that the analyses presented here depend on the assumption of a measurement model of the type presented in Equation 1. Moreover, we will try to avoid any imputation that cited authors have incorrectly interpreted the observational analyses, but rather outline how different interpretations could arise and demonstrate (as we have shown) how they can be reconciled within a Bayesian framework with particular priors.

Interactive comment on Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2019-33, 2019.