

Geosci. Model Dev. Discuss., referee comment RC1  
<https://doi.org/10.5194/gmd-2022-178-RC1>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Comment on gmd-2022-178

Anonymous Referee #1

---

Referee comment on "A new bootstrap technique to quantify uncertainty in estimates of ground surface temperature and ground heat flux histories from geothermal data" by Francisco José Cuesta-Valero et al., Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2022-178-RC1>, 2022

---

Summary: Past global surface temperatures over the past few centuries can be estimated from present borehole temperature profiles applying inversion techniques based on the solution of the heat transfer equation. From a set of sites where temperature profiles have been measured, large-scale temperature reconstructions can be derived by averaging the local retrievals. The manuscript presents a method to estimate the uncertainties in the large-scale average estimations based on a bootstrap approach. The authors conclude that this new method provides better and more realistic uncertainty estimates than previous methods. Those previous methods simply calculated the average of the high and low ends of the local uncertainty ranges.

Recommendation: I think that in general the manuscript is valuable and should be published after some revisions. However, I am afraid that one of the motivations of the present study, namely that the previous estimations of uncertainty was unrealistic, contains a conceptual misconception, although it has been previously published. Therefore, the motivation of the present manuscript should be amended to present a correct statistical case. I explain below in more detail my main concern.

1) The manuscript presents a base method to estimate global or large-scale uncertainties that has been published previously. This method just constructs the high-end (and low-end) uncertain range of the global average by calculating the average of the high-end (or low-end) range of the local estimations. This is, however, not correct, as it can be illustrated in a short counter-example. The interpretation of a 5-95% uncertain range in a frequentist approach is that the range covers the true value with 90% probability (technically, it means that a putative infinite number of realizations of the measurements and their corresponding uncertain estimations will contain the true value 90% of the time). For the sake of this reasoning, we can a bit sloppily say that the probability that the true value is within the estimated uncertainty range is 90%. However, if the uncertainty ranges are constructed by averaging the 5% and the 95% local ranges, this probability is much much larger than 90%. Let us focus on the high end (95%). The

probability for the average to be outside that 95% range is not 0.05, but actually  $0.05^N$  to the Nth power, where N is the number of profiles (sites). This results because each profile from which that average is constructed, has a probability of 0.05. If  $N=100$ , this number is very small, much smaller than 0.05.

The authors realize in the discussion that indeed this estimation is not correct. There, they apply a much more correct estimation assuming that the global profile is the average of N random variables, and therefore, assuming that these N random variables are independent, the error in the average amounts to the sqrt of the average squared error. If all individual errors are equal, this amounts to that individual error divided by the  $\sqrt{N}$ .

There is one important underlying assumption: the errors should be independent across space. But even if this assumption is not completely fulfilled, this estimation is much more realistic than simply the average of the upper and lower local percentiles, which is clearly incorrect.

Thus, to some extent, the manuscript corrects a previous statistical misconception. In this sense, it is useful, but the motivation of the manuscript should be cast differently, as the reader will be really surprised to see, without any caveat, a clearly wrong method as a benchmark.

On the other hand, the question of the spatial correlation of uncertainties, which is critical for the validity of both methods (bootstrap and error propagation) is not mentioned at all.

The bootstrap approach is definitively better - and I could not see any clear error in this application of bootstrapping. However, this approach does not take into account the possible spatial correlation of the local errors. I do not know how significant these correlations might be, but if they are, then the bootstrap estimation of the uncertainty will be too narrow - in the same way as the error propagation proposed by the authors in the discussion - since the effective number of degrees of freedom will not be N, but smaller. If these correlations are relevant, the bootstrap should take it into account, e.g. by block-bootstrap, in which correlated regions are first averaged together, and then bootstrapped. I think that this problem is technically very difficult to solve satisfactorily, but again, I believe that the presented bootstrap approach is indeed useful.

Particular points

2) A definition of the quasi-equilibrium temperature will help some readers

3). Section 3.5, perhaps the most important section, is not very clearly written (I needed

to read it several times). For instance, line 257: 'named Sampling and Bootstrapping ensembles (S and B ensembles in Figure 2). The Sampling ensemble consists ...'. and the reader expects the following sentence to explain what the Bootstrapping ensemble is. However, the text goes on with 'The BTI method considers the uncertainty arising from ...'. This is quite confusing. Actually, the bootstrapping ensemble is a typical bootstrap sampling from the set of individual local profiles, where each profile has been derived from one value of the uncertain parameters ( $T_0$ ,  $\Gamma_0$ , and thermal conductivity). The only restriction is that each site contributes with one member to the ensemble. All in all, I found the technical description unnecessarily too cumbersome.