# Review of "Parameter Calibration in Global Soil Carbon Models Using Surrogate-based Optimization" (version 2)

The authors have made numerous improvements. I also appreciate the clear marking of changes in the PDF which allows cross-referencing with replies to comments. However, for a number of criticisms I am unsatisfied with the reply and/or modifications to the manuscript. I also have a few new comments.

## Replies to specific comments

**Comment 2.1:** The fact that the RBF-SBO starts out with a considerably lower RMSE for all three models (Figure 5) suggests that the calibration setup somehow gives RBF-SBO an unfair advantage over the other algorithms. If this is the case, it would have serious consequences for the paper. Possibly the calibrations would have to be redone in a setup that removes this advantage.

**Response:** The reason that the setup of RBF-SBO is better than other optimization algorithms is that SBO has a sample step and selects the good parameters while other algorithms simply select initial samples randomly. We also repeat the experiments by making the SBO have the same quality setup as other algorithms. The results can be found in the figures below. The results show that RBFSBO has the similar performance on 2-pool microbial model and outperforms other algorithms on 4-pool microbial model and CLM-CASA' Carbon Model with the limit of 200 sample runs.

*If I understand the reply correctly, the left most point in the graphs of Figs 5a-c for the SBO algorithm corresponds to the first iteration of the loop indicated by steps 4-6 in the pseudo-code block depicted on p 9—not to the initial parameter sets S0 (derived in step 1 of the pseudo-code block). I suppose the reason for this is the fact that S0 represents multiple parameter sets needed to construct the initial surrogate model, so one cannot pick a parameter set that is truly the first. It is good that the authors show that also with "the same quality setup as the other algorithms" SBO shows good performance. Nevertheless, in my view, Fig 5 is misleading, because the "real" model has already been run a number of times before the start of the graph. For the other algorithms on the other hand, (I assume) the left-most point of the graph is truly the first time the model is run. For a more honest comparison I would suggest to shifting the curve of SBO to the right by adding the number of initial simulations to the x-axis value; e.g. if there are 50 initial simulations, the curve would shift 50 points to the right. The same can be done for the other algorithms in case they also require a number of initial simulations. I also find the changes made but the authors (change 2.1; marked sentence starting on P11L28) insufficient. In my view this sentence does not explain the apparent advantage of SBO in Fig 5 to the average reader.*

**Comment 2.4:** Furthermore, the SBO based estimates strongly disagree with the MCMC estimates for two of the 4-pool microbial model (CUE_slope, and CUE_0; Figure 10). This is briefly mentioned (P11, L19) but not further discussed. P12, L3-4: "it still can find the true parameter values". The mismatch for CUE_slope and CUE_0 in Figure 10 shows that this is not always the case.

**Response:** The Figure 10 shows that some calibrated values of the SBO are different from the Bayesian MCMC, and these different values make the prediction error of SBO results lower than Bayesian MCMC. According to our understanding, the mismatch of these parameters may be due to the different targets of the parameter selection between SBO and Bayesian MCMC.

**Changes in manuscript:** We briefly discuss the reason of the mismatch issue for 4-pool microbial model in Section 5.2 (marked as Change 2.4-1) and improve the responding statement in Section 6 of the revised version (marked as Change 2.4-2)

*The added/modified section marked with change 2.4-1 (starting P14L26) is unclear. What is meant with "different targets"? If the same cost/objective function is used, the mode of the MCMC derived distribution should match with optimal value derived with the optimization algorithm. Given that the SBO derived value for CUE0 has negligible posterior probability according to the MCMC derived distribution (Fig 10), while it has a lower RMSE than that of the parameter set corresponding to the mode, suggests that something has gone wrong with the MCMC sampling. Most likely the MCMC algorithm has gotten stuck in a local minimum located close to the global minimum.*

**Comment 2.5:** The language in the paper is in general quite poor. There are quite a few spelling and grammar errors, and many sentences are semantically incorrect (e.g. missing or incorrect usage of articles), awkward, or use spoken rather than written English. I've listed a number of them below, but I strongly advise proof-reading by a proficient an editor proficient in the English language. Please check also the citation references, both in the text and in the bibliography. There appear to be quite a few mistakes.

**Response:** Many thanks for your valuable suggestions. We have fixed all the grammatical and formatting issues you pointed out, and tried our best to conduct several rounds of proofreading and substantially improved English presentation in our revised manuscript.

*The English has improved but it is not yet at a sufficient level in my view. There are still quite a few awkward sentences, and some grammatically incorrect ones. Particularly the use of articles and punctuation should be improved.*

**Comment 2.7:** It is rather unfair to compare computational cost of the SBO approach presented here to that of Bayesian MCMC, since the latter is a sampling algorithm, whereas the former is a optimization algorithm. Sampling schemes are intended to obtain a detailed approximation of the posterior/likelihood function whereas optimization schemes only yield an estimate of the maximum likelihood point. Comparing the computational cost to that of the other optimization approaches would make more sense.

**Response:** We agree with the reviewer. The Bayesian MCMC is designed to obtain a posterior likelihood function but it can also be used to calibrate parameters to reduce the prediction error. Moreover, we also compare the SBO with known global optimization algorithms in our manuscript.

*From the reply it seems that the authors do <u>not</u> agree with me, as opposed to their first sentence. I stand by my point that optimization algorithms and sampling algorithms have different purposes and the latter are not a sensible choice when one is calibrating a computationally expensive model. The authors show this themselves because the CLM-CASA model requires 1,000,000 simulations to get convergence (P11L3), while the optimization algorithms (also non-SBO) get a similar result in less than 1000. However, when discussing performance, the authors are still comparing their <u>optimization</u> algorithm to the*

*sampling algorithm MCMC. For example, in the last sentence of the abstract "Meanwhile, the corresponding computation cost required is only one thousandth of that with Bayesian MCMC.", and in the conclusions (point 2). This is misleading. Readers not familiar with MCMC will think that the authors have made an incredible performance improvement, which this not the case. Please limit comparing performances to that of the other optimization algorithms (PSO, SCE-UA, CMA-ES), or at least make it clear that Bayesian MCMC has a different purpose.*

**Comment 2.11:** P8, L4-12 concerning the Bayesian MCMC approach:

- It appears that the authors used the Metropolis algorithm. If so, please state this.
- Have these calibration runs been performed specifically for this study or did the authors use the results from Hararuk et al. (2014, 2015)?
- How is the acceptance probability calculated?
- How was convergence of the MCMC algorithm diagnosed. What criterion was used?
- Please provide more information on how the MLE point is determined
- It is stated that Table 3 provides the detail of the Bayesian MCMC approach. However, other than the number of iterations no information is given

**Response:** The Bayesian MCMC (Hararuk et al, 2015, mentioned before) used the Metropolis algorithm. We have got the code from Hararuk and repeated the calibration experiments. This MCMC approach would run 50, 000 samples before ends.

**Changes in manuscript:** We clarify this point in Section 4.2 of the revised version (marked as Change 2.11).

*I have difficulty understanding the paragraph related to this comment (change 2.7; starting from P10L25):*

- *What is exactly constitutes the proposal step and the moving step?*
- *According to the Metropolis rule the acceptance probability is determined by the ratio of the unnormalized posterior density for the proposal parameter set and that of the current parameter set. Is this what is meant with "a probability of acceptance determined by prediction error is calculated"*
- *Also "The final calibrated parameter set is estimated by Maximum likelihood estimator (MLE) with an accepted parameter chain." Is this an additional calculation at every iteration or done after the calibration?*

## Additional questions/comments

- Briefly inspecting the supplemental information, I found that the authors used the Matlab "Surrogate Model Optimization Toolbox", implemented by Julliane Mueller. This is a substantial amount of work that has been done by others. The method description, however, makes no mention of this—initially I was under the impression that the authors implemented the SBO algorithm themselves. The authors refer to two publications of Mueller et al. in the text but only

for general descriptions of SBO. Please make clear in the main text that you used the implementation of Mueller et al. and refer to the appropriate papers.

- P8L25: it is not clear what is meant with “exploration” and “exploitation”
- P9L2: Sentence starting with “It is worthy noted…”. Not clear what is meant here.
- P11L8: Sentence starting with “For each algorithm”. Not sure what is meant here, That the algorithm is truncated after 100 simulations? If so, why is this done--to limit computation time? Wouldn't the results be strongly dependent on the initial parameter set(s)?
- Figures 7a/b: the quality of the graphs is really low. I would suggest to use a version with a higher resolution
- Figure 2: why do the cells near the cost have no data?
- Table 1: what is meant with “exit rate”? Turnover/decomposition rate? If so I would suggest to use these terms