Review of a "Parameter Calibration in Global Land Carbon Models Using Surrogate-based Optimization" by Xu et al.

Maarten Braakhekke

General comments

This manuscript presents a novel approach for calibrating global land carbon cycle models that are computationally costly (i.e. need a long time for a single simulation). The approach, dubbed surrogate-based optimization, uses a uses a computationally cheap surrogate model, which mimics the original model, to generate candidate parameter sets at each iteration. Since the original model is only run for "good" new parameter sets, this approach avoids evaluation of the original model for "bad" parameters, thereby substantially reducing the number of model iterations, and thus computation time. The authors apply the algorithm for the CLM-CASA global land surface model in order to optimize parameters related to soil carbon cycling against global gridded soil carbon stocks from the IGBP-DIS dataset. Additionally, the approach is applied for two other soil carbon models, which explicitly represent microbial dynamics. The calibration results are compared to those of four other optimization schemes and a Bayesian MCMC algorithm.

To my mind the approach is very promising and helps to tackle an important issue with calibrating global models: the high computation cost. I'm not very experienced with optimizing global models and thus I cannot say if there are other techniques that achieve the same thing, and how these compare to the approach presented here. Nevertheless, I think the paper is relevant, and a strong contribution to the field of global modelling. Furthermore, I think the authors were quite thorough in testing the new approach by applying it to three models, and comparing the results to five other optimization/sampling schemes.

However, I do have several criticisms that should be addressed. These relate mainly to the text.

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.
2. The description of the methods need to be considerably expanded since much important information is missing, most importantly, on the algorithm itself. Ideally, one should be able to reproduce the approach from the description in the main text, appendix, or supplemental information. However, in this manuscript not nearly enough information is provided for this. For example, I would guess that the algorithm evaluates and rejects several proposal steps using the surrogate model, before a parameter set is deemed good enough to be evaluated by the true model. However, no information is provided as to how these kinds of choices are made. I would suggest including a pseudo-code block to describe the working of the algorithm.
   Additionally, the surrogate model is constructed based on "radial basis functions" but no additional information is given on how this works. Since the approach for surrogate model is a critical choice (as acknowledged by the authors, P6, L14) this approach needs to be described in much more detail.

There are several other places in the text where more information should be provided. These are given in the specific comments below. Parts of these descriptions may be placed in an appendix or online supplement.

3. I find the paper a bit biased towards a positive assessment of the algorithm and superiority over other algorithms. The paper would benefit from an additional discussion section on the possible limitations of the approach, which I'm sure exist. For example, the limitations of using a surrogate model for mimicking complex models is briefly mentioned (P9, L5-11), but its consequences are not further discussed. 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.
4. 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.
5. From what I can understand from the paper (P3, L5-15) the authors only ran and calibrated soil carbon models, no full land carbon model. Therefore, I find the title somewhat misleading. The approach can probably be used to optimize a full land carbon model, but this has not been shown. I could imagine that the limitations posed by using a surrogate model would become more relevant for a full land carbon model. Hence, I would suggest replacing "land carbon models" with "soil carbon models", or "the soil carbon component of land carbon models".

Specific comments

Abstract

- P1, L11: I suggest to either replace "which can be obviously improved" with "which can obviously be improved", or remove "obviously" altogether

Section 1

- P1, L21: "SOC is the largest pool of global land carbon." please provide a reference for this statement.
- P1, L21: I suggest replacing "a famous" with e.g. "the most important".
- P1, L29: I suggest elucidating "agree with". E.g. "For only half of the 11 models the predicted global total SOC falls within estimated range of the HWSD"
- P1, L29: remove "s" in "coefficients"
- P2, L3: remove "the" before "parameter"
- P2, L4: "replace "expensive" with "high"
- P2, L7: remove "the" before "high"
- P2, L8: put "like CLM" between comma's. Also, CLM has not previously been introduced (it is two lines below)
- P2, L15: remove "also"
- P2, L17: replace "the" at the end of the line with "an"
- P2, L18: add "the" before "surrogate"

- P2, L23-24: “Quite a few…benchmark. This sentence is unclear. Consider revising.
- P2, L25: add a comma after “Here”
- P2, L29-30: “On average…Bayesian MCMC”. This sentence presents results, and should not be in the introduction. However, I admit this may be a matter of style.
- P2, L30: 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.
- P2, L34: Replace “analysis” with “discusses”

Section 2

- P3, L1-2: “…their structures of land carbon cycle are almost the same”. This is statement is a major oversimplification. I would suggest something like “there are many similarities”
- P3, L6: remove the “s” at the end of “carbons”
- P3, L9: “one of the most popular earth system models in the world”. I suggest replacing with “widely used Earth system model”
- P3, L23: add “model” after “CLM-CASA’”
- P3, L24: add “of” after “linear”
- P3, L30: I suggest replacing “The steady solution of equation (1) is solved by Xia et al. (2012):” “the steady state solution of equation is given by (Xia et al. 2012):”
- P4, L1: add a comma after “NPP”
- Section 2.2: The microbial soil carbon models and the corresponding equations (3)-(16) need to be better explained (e.g. what processes do the different terms in the ODEs represent). For someone not experienced with such models it is currently difficult to understand what’s going on.
- P4, L21: add “be” after “to”
- P5, L12: add “The” to the start of the sentence and remove the “s” in the second “models”
- Figure 2: Why do gridcells near coastlines have no data?
- P5, L22: replace “gird” with “grid”
- Tables 1 and 2: please provide the units of the parameters

Section 3

- Section 3: as discussed above the radial basis functions approach needs to be explained, as well as the approach to generate proposal samples
- P6, L9: I suggest adding “surrogate” before “model”
- P6, L12: I suggest replacing “cancelled” with “avoided”
- P6, L133: I suggest replacing “save much” with e.g. “substantially reduce”
- P6, 24-26: This sentence is rather vague. What is meant with “real variability”?
- P6, L24: please provide a reference for Latin hypercube sampling
- P6, L24: add “which” between “for” and “LHS”
- P7, L3: I suggest replacing “optimum” with “optima”
- P7, L7: I suggest replacing “try to present” with “present” or “try”

Section 4

- Section 4.1: The authors state that the calibration process is repeated 50 times. How do you assure that the you don't get the same result every time? Is the algorithm started with different initial values, or are there stochastic parts in the algorithm?
- P7, L21: add a comma after "algorithms"
- P7, L23: add a comma after "(CMA-ES)"
- P7, L24: I suggest removing "the outstanding"
- P7, L26: remove the parenthesis "(" after "SCE-UA"
- P7, L28: the reference "MA H, et al., 2006) is not present in the bibliography
- P7, L31: add "other" before "three"
- P8, L1 & L2: I assume you mean "normal" instead of "norm"
- P8, L2: I suggest replacing "proven" with "shown"
- P8, L3: replace "on" with "in"
- 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
- Figure 4. I assume the box plots show means and spread over the 50 calibration runs. Please indicate this in the caption
- Figure 4: I suggest replacing "exceptions" with "outliers"
- P8 L15: I suggest replacing "measure" with e.g. "applied"
- P8, L16: please revise "As the requirement of Bayesian MCMC…"
- P8, L19: I suggest replacing "On" with "for"
- P8, L20: I suggest removing "only"
- P8, L21: remove "of" before "more"
- P8, L21: consider rephrasing "can exploit better results"
- P8, L22: I suggest replacing "from the aspect" with "with respect to"
- P8, L24: replace "get" with "gets"
- P8, L25: consider rephrasing "promising one"
- P8, L26: consider rephrasing "It is because…"
- P8, L28: I suggest "…matrix and thus that…" with "…matrix, hence…"
- P8, L29: consider rephrasing "extremely critical"
- Figure 5: For all three models, the RBF-SBO algorithm starts out with a considerably lower RMSE at the first iteration, compared to the other algorithms. Please explain this difference. I wonder if the setup of the algorithm somehow gives RBF-SBO an advantage. This would make the comparison unfair.
- P9, L1 & L15: replace "till" with "until"
- P9, L2: add "s" after "simulation"

- P9, L7: consider removing “as we know”
- P9, L7: replace “the” with “an” before “approximation”
- P9, L14: remove “that”
- P9, L14: replace “increasing” with “higher”
- P9, L15: consider rephrasing “keeps ahead”
- P9, L24: consider rephrasing “for the samples”
- P9 L24: “on the other hand” indicates that what follows contradicts what was stated previously. This is not the case here.
- P9, L29: replace “are” with “is” (refers to “performance” not “models”)
- P9, L29: consider rephrasing “no one can dominate other two”
- P9, L29: consider rephrasing “all get success”

Section 5

- P10, L1: add “state” after “steady”
- P10, L1: replace “Equateion” with “Equation”
- P10, L2: add “the” before “IGBP-DIS”
- P10, L2: I suggest removing “obviously”
- P10, L13: I suggest replacing “sharp” with “narrow”
- P10, L13: I suggest removing “that those are”
- P10, L15: replace “in” with “for”
- P10, L16-17: “On the other hand”: see comment for P9 L24, above
- P10, L19: replace “approximate” with “close”
- P10, L19: replace “assigend" with “assigned”
- P10, L20: consider rephrasing “not so reasonable”
- P10, L20: replace “reaches” with “approaches” or “is close to”
- Section 5.2: for CUE_slope and CUE_0 there is considerable mismatch between the mode of the parameter distributions derived by MCMC, and the estimate from the SBO algorithm. But this is only mentioned in passing (P11, L19-21). It needs to be mentioned more explicitly and the potential reasons and consequences should be discussed.
- P11, L5: move “both” from its current location in the sentence to before “CLM-CASA’”
- P11, L14: consider replacing “biomass” with “dynamics” or “processes” since CLM-CASA’ also has a microbial biomass pool

Section 6

- P11, L23: I don’t agree with the statement that “Bayesian MCMC approach has been used to typical SOC models”. To my mind most of these models have been tuned either manually or with gradient search algorithms
- P11, L24-25: “owing to approximate one million simulations”. The number of required iterations is completely dependent on the specific calibration problem so one cannot state a specific number for calibrating SOC models in general
- P11, L27-28: see comment P2, L30
- P11, L30: I suggest replacing “dominates” with “outperforms”
- 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