Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2017-48-RC1, 2017

Geoscientific
Model Development
Discussions

Interactive comment on "Parameter Calibration in Global Land Carbon Models Using Surrogate-based Optimization" *by* Haoyu Xu et al.

Anonymous Referee #1

Received and published: 11 July 2017

This manuscript describes the performance of a surrogate-based approach to calibrating three different soil carbon models relative to three global optimization algorithms and a MCMC algorithm. The results indicate that the surrogate-based optimization employing radial basis functions outperforms the other approaches in nearly all circumstances. Model calibration improves the fit of the models to a global dataset of soil carbon values, with the models incorporating soil microbial dynamics explicitly fitting the data more effectively than a model based on CLM-CASA.

Unfortunately, the quality of the English throughout the manuscript is extremely poor with numerous grammatical errors throughout all the text. Without a great deal of additional editing for language alone this will not be publishable in GMD.

All these English language errors, which are far to numerous to call out individually,

make it very difficult to undertake a review of scientific merit, but there are a number of areas that clearly require further elaboration and clarification.

Whilst it is certainly challenging, the authors assertion that it is not possible to optimize parameters directly in land surface models such as CLM is not true – see for example Post et al., 2017 JGR-B and reference there in.

The assertion that the "structures of land carbon cycle" with ESMs "are almost the same" maybe true but requires evidence and references.

It is unclear what are the differences between CLM, CLM-CASA and CLM-CASA C-only. My interpretation is that CLM-CASA C-only is the steady-state approximation detailed in Xia et al, 2012, and the SBO was developed for this. Some additional detail is required here – for example, what are the meteorological drivers, what are the inputs? "NPP" is mentioned, but never explained. This is important, as the relevance, or otherwise, of this work to informing ESM development can only be understood if the implications of using a surrogate model to parameterize a matrix-based approximation of the steady-state of the simplistic soil component of an old land model are fully articulated.

The description of how the specific SBO algorithm and parameter point generation strategies is unclear – what is about the nature of the algorithms chosen that makes them appropriate for this particularly use case?

Given the code available in the supplementary material, it is apparent that the various optimization algorithms were implemented in Matlab and relies heavily on material from the File Exchange. Details of this implementation need to be in the main text.

As the authors highlight, "sample size, the nonlinearity and complexity of the real model" all impact surrogate performance. This is partially addressed through the use of three models with different numbers of pools/parameters but not well explained, nor is there reference back to the role of surrogates with ESMs of full complexity.

The analysis of the results (Section 5) fails to discuss the implications of the optimizations for CLM-CASA C-only. What does it mean for the model if even when optimized it can only explain 40% of observed variation? Why are so many parameter values right at the edge of their prior range? Are the numbers “biological feasible”? To what extent is the improvement in fit with microbial model due to the inclusion of microbes, or rather due to spatially varying base rates?

Overall, the work described in this manuscript has the potential to inform future land surface model developments, and highlights the possibilities of using surrogate-based optimization at a fraction of the computational cost of MCMC-type approaches. With much improved editing, clarification of the points outlined here, and a more involved discussion of the outcome of optimization exercise – which is the point of the whole exercise after all – hopefully it can be considered more favorably for inclusion in GMD in the future.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-48, 2017.