Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2017-48-AC1, 2017

Geoscientific Model Development Discussions

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

Haoyu Xu et al.

ocean920329@gmail.com

Received and published: 3 November 2017

First of all, the authors would like to thank the reviewer for the valuable comments, suggestions as well as generous recognitions, which would greatly improve the clarity of our presentation and help our revision.

Comment: 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.

Response: Thanks. We have tried our best to conduct several rounds of proofreading and substantially improved English presentation in our revision manuscript.

Comment: 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.

Response: We apologize for the confusion. We meant to express the point that optimizing turnover rates and other related parameters with pool-based datasets is computationally too demanding for land surface models. Optimizing flux-related parameters is computationally challenging but possible. The study by Post et al. 2017 was conducted to optimize photosynthesis-related parameters with eddy-flux data. It is a great work. To optimize parameters, such as those we estimated in this study, has to overcome additional computational challenges. For example, it takes a very long spin-up time to run the whole CLM model. Tuning parameters for the whole model requires the extreme computational and temporal cost.

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

Response: Thanks for your suggestion. Now we cite the paper by Huang et al. 2017, which shows that the matrix equation not only can exactly reproduce the original CLM4.5bgc but also offers the simplicity in coding, diagnostic capacity, and computational efficiency. The latter enables optimizing pool-related parameter estimation.

Comment: 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. 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. Some additional detail is required here – for example, what are the meteorological drivers, what are the inputs? “NPP” is mentioned, but never explained.

Response: Thanks for your questions. We use the term CLM to refer Community Land Model in a general term. CLM-CASA’ is a version CLM3.5 of CLM. The CLM-CASA’ C-only version is the same model CLM-CASA’ only when we consider the C processes of that model. We developed SBO to optimize parameter estimation for the CLM-CASA’ C only version using the steady-state approximation. Parameter optimization cannot easily be done at the non-steady state unless time-series data sets are used as did by Zhou et al. 2013 and 2015.

As the matrix equation results from the re-organization of exact equations as in the original model, the parameter values estimated by SBO can be directly transferred to the original model. Moreover, the matrix representation offers the solution to the land carbon cycle modeling, the National Center for Atmospheric Research (NCAR) land modeling team will adopt the matrix equation as the main frame of CLM in the future version. Thus, any parameter estimation with the matrix equation can be directly used to improve the original model.

We added sentences to clarify this point in our revised manuscript.

Comment: 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.

Response: The initial parameter is generated using LHS(Latin Hyper-Cuber Sampling) and we agree that the details of the SBO algorithms should be given. An appendix which includes a detailed description of SBO has been added in the revised manuscript. The reason why SBO outperformed than other global optimization algo-

rithms is that SBO uses a surrogate model to simulate the source model (CLM carbon process) and avoids much bad parameter point (‘bad’ means the high prediction error). By using the surrogate model, the SBO can save many model running times and is appropriate for expensive computation cost models.

Comment: 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.

Response: We agree with the reviewer. To validate this, we carefully choose three different models to evaluate our algorithm. The nonlinearity and complexity of the three models are different. The number of parameters and the equations are different, and the performance of surrogate models are also different. We will add some detailed analysis in the revised version.

Comment: 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?

Response: We greatly appreciate the reviewer pointing out this issue. It is still not satisfactory to explain 40% of observed variation with the optimized model. This is similar to another study by Hararuk et al. 2014 and much better than the model with default parameter values, which only explains 27% of the observed variation. The unexplained variation is partly due to uncertainty in observations. Indeed, the world homogenized soil data is grid-based map of soil carbon content, which was developed from pedon data. The equation used for the homogenization only can explain 27% of the variation in the original pedon data. That means that the homogenization itself generates 73%

variation. To improve the model-observation fitting, we need to understand uncertainty sources from data, model structure, parameters, and forcing.

The edge-hitting of estimated parameters is usually related to correlations among parameters. We need information of covariances among parameters to resolve the edge-hitting issues.

It is not very clear to us what the reviewer tried to ask with the question Are the numbers "Biological feasible"? Is it about the number of parameters that can be constrained by the dataset we used in this study? In general soil carbon content rich information to constrain several parameters related to soil carbon pool turnover as showed in this study.

The improvement in fit with microbial model is largely due to the nonlinearity, which is more flexible to fit data. It is not clear to us whether the improved fit has anything to do with spatial variation in base rates. We may design a different study to explore this issue.

References: Todd-Brown, K. E. O., J. T. Randerson, W. M. Post, F. M. Hoffman, C. Tarnocai, E. A. G. Schuur, and S. D. Allison (2013), Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations, Biogeosciences, 10(3), 1717–1736, doi:10.5194/bg-10-1717-2013. Weng, E., and Y. Luo (2011), Relative information contributions of model vs. data to short- and long-term forecasts of forest carbon dynamics, Ecol. Appl., 21(5), 1490–1505, doi:10.1890/09-1394.1. Luo, Y., L. W. White, J. G. Canadell, E. H. DeLucia, D. S. Ellsworth, A. Finzi, J. Lichter, and W. H. Schlesinger (2003), Sustainability of terrestrial carbon sequestration: A case study in Duke Forest with inversion approach, Global Biogeochem. Cycles, 17(1), 1021, doi:10.1029/2002GB001923. Xia, J., Y. Luo, Y. P. Wang, and O. Hararuk (2013), Traceable components of terrestrial carbon storage capacity in biogeochemical models, Global Change Biol., 19, 2104–2116, doi:10.1111/gcb.12172. Zhou, T., and Y. Luo (2008), Spatial patterns of ecosystem carbon residence time and NPP-

driven carbon uptake in the conterminous United States, Global Biogeochem. Cycles, 22, GB3032, doi:10.1029/2007GB002939. Huang YY, XJ Lu, Z Shi, D Lawrence, C Koven, JY Xia, ZG Du, E Kluzek, YQ Luo. (2017). Matrix approach to land carbon cycle modeling: A case study with Community Land Model. Global Change Biology DOI: 10.1111/gcb.13948.

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