
Dear Editor Christoph Müller,

We greatly appreciate your granting us another opportunity to revise the manuscript. We are also thankful to the reviewers for their valuable comments and suggestions, which are extremely helpful for us to improve the manuscript. As a consequence, the manuscript has been substantially improved. We hope you will find our revision is satisfactory.

Below please find our responses to comments by the two reviews and the change list in the manuscript

Sincerely,

Wei Xue, Ph.D.

Associate Professor

Department of Computer Science and Technology and Department of Earth System Science, Tsinghua University

Beijing, China, 100084

Letter of Response

Note that comments from referees are in italics and our responses in normal shape.

Referee #1

This paper tests the performance of surrogate modeling approach for estimating parameters in three soil carbon models relative to selection of commonly used direct optimization approaches and an MCMC approach. The results suggest their surrogate modeling approach is amongst the most effective and is the most efficient (by their measure at least).

Response: No response needed

Comment 1.1:

The description of the surrogate modeling and optimization approaches are relatively complete, but details about the models themselves, and how they have been used, are incomplete and confusing. This distracts from the main aspects of the work detailing the optimizations. This confusion is not helped by the authors repeatedly making claims about working with Earth system models and global land surface models, when in reality they working with a matrix approximation of a sub-component of such complex models, or models that are not part of such systems at all. Whilst it is exciting to see such approaches being applied in this way, the usability and applicability with fully functionally ESMs must not be overstated.

It is important to have clarity here, as one of the main arguments you make is how beneficial surrogate models are when working with computationally expensive, highly complex Earth system models. Yet, in reality the surrogate model is being built for the matrix approximation, which in itself was developed to address many of the issues related to parameterization of ESMs. Often in the manuscript you described the days to weeks of wall clock time associated with MCMC model runs in comparison to the efficiency of the surrogate. But is this really the case with the matrix model? Similarly, the 2- and 4-pool models are not highly complex (relative to an ESM), and it is unclear if there are real world benefits from developing a surrogate for such models (that is to say, is the efficiency gain really worth the expense in developing the surrogate for a model of this complexity?)

Response: We greatly appreciate the comment. We clarified this point and made changes in several places of the manuscript. For example, we have revised the second and third paragraphs in the introduction (lines 7-24 on page 2) to clarify those three models for parameter calibration as: (marked as **Change 1.1-1**)

“Considering the high similarity in the structures of the 11 ESMs, the difference in

the SOC simulations mainly results from parameterizations (Todd Brown et al., 2013). Thus, parameter calibration is among the top priorities to improve prediction of global land carbon cycle dynamics (Luo et al., 2016). However, the parameter calibration with global observations has not been widely applied, owing to the high computational cost. A matrix approach has been recently developed to reorganize the carbon balance equations in the original ESMs into one matrix equation without changing any modeled C cycle processes and mechanisms (Luo et al. 2003, 2017; Huang et al. 2018). The matrix land carbon cycle models can be semi-analytically solved to obtain steady-state solutions faster than the original models by tens and hundreds times (Xia et al. 2013). As a consequence, the matrix approach makes parameter estimation and calibration possible. The matrix approach has been successfully used for the parameter calibration to constrain SOC turnover and microbial process with Bayesian Markov chain Monte Carlo (MCMC) algorithm (Harauk et al., 2014, 2015; Shi et al. 2018).

Bayesian MCMC is a sampling-based approach and usually requires a large number of simulations for building an acceptable parameter chain. For instance, over 500,000 simulations are required during the parameter calibration of soil carbon models (Xu et al. 2006). Even using high-performance computers, complex land models, like the latest version of Community Land Model (CLM5.0), require a very long spin-up time for carbon cycle simulation, leading to several hours or days for one simulation. Although the matrix approach has been developed to enable data assimilation of global land carbon cycle models (Harauk et al., 2014 and 2015, Shi et al. 2018), Bayesian MCMC computationally is still very expensive for calibrating global land models. More effective and efficient parameter calibration algorithms are urgently needed.”

We have also revised the second last paragraph in the Introduction (lines 9-17 on page 3) to clarify the purpose of evaluating SBO with the three models as: (marked as **Change 1.1-2**)

“Most studies on both global and surrogate optimizations focus on the mathematical function benchmarks like Comparing Continuous Optimizers, abbreviated as COCO (Hansen et al., 2010; Wang and Duan, 2014). However, the optimization of the mathematical functions may be extremely different from the parameter calibration of complex real-world models. In this paper, we explore the state-of-the-art surrogate optimization method for parameter calibration of three SOC models: CLM coupled with Carnegie-Ames-Stanford Approach biogeochemistry model (CLM-CASA’) and two microbial models, as used in studies by Hararuk et al. (2014, 2015). Although the three models are computationally attainable for parameter calibration, we compare the performance of surrogate-based optimization to advanced global optimization algorithms and the data assimilation method to examine the potential of SBO. The SBO may be extended to other complex global land models.”

We also added one sentence in the Conclusion (Lines 16-17, page 16) to clarify the usefulness of this study as: (marked as **Change 1.1-3**)

“Although the three SOC models used in this analysis are not computationally unattainable for parameter calibration, what we have learned about SBO from this study can be potentially applied to more complex models.”

Moreover, we have thoroughly revised section 2 to clearly describe the three SOC models.

Comment 1.2:

It is unclear how the three soil C models were actually run. What are the inputs? What is the climate forcing? Where do they come from? Are they a time series, or mean values used to estimate a steady state solution? Rather than describing the model equations – which could be moved to appendices – a more holistic description of the modeling approach including details of domain, forcing, NPP inputs would be beneficial.

Related to this is confusion as to how the CLM-CASA’ C-only model is being used. Section 2.1 talks about the matrix approximation in general terms – but not specifically about its use in this case. Whilst acknowledging there are references describing this, some detail is required directly in this manuscript.

Response: We apologize for the confusion of model description. We have thoroughly revised section “2 Global Land Carbon Models, Data, and Cost Function” on pages 3-6 to clarify the descriptions of the three SOC models (marked as **Change 1.2**). The inputs are defined by NPP and the outputs are the carbon content of each pool at a steady state. Hope the reviewers will find our revision satisfactory.

Comment 1.3:

The results themselves are encouraging, and what might be expected, with the exception of the relatively poor performance of the MCMC scheme in the case of the 2- and 4-pool models. The authors suggest this “may be mainly due to the different targets of the parameter selection”. This needs further elaboration because I don’t really understand what that means, and it really is a surprising result.

Response: Same as the response for Comment 2.2.

Comment 1.4:

Figure 2 needs some clarification. Why is the data so “blocky” and covers such a limited extent of the land mass? Is that the domain over which the model was actually run?

Response: Thanks for pointing out this problem. We have replotted Figure 2 (marked as **Change 1.4**). For the previous version, we followed the method using in Hararuk et al. 2014.

Hararuk O, Xia J, Luo Y. Evaluation and improvement of a global land model against

soil carbon data using a Bayesian Markov chain Monte Carlo method. *Journal of Geophysical Research: Biogeosciences*, 2014, 119(3): 403-417.

Comment 1.5:

It seems like much of the appendix replicates Section 3.3 – but is actually written better and is clearer. I would suggest merging them into Section 3.3 alone.

Response: Thanks for the valuable suggestion. We have merged the appendix into Section 3.3 (marked as **Change 1.5**).

Comment 1.6:

The quality of the English is very variable across different sections of the manuscript. The abstract and introduction are very poor with numerous language errors – far too many for this reviewer to list individually and any resubmission will require a thorough proof reading.

Response: Thanks. We have carefully edited the manuscript and improved English presentation in our revised manuscript, especially for the abstract and introduction sections.

Referee #2

Comment 2.1:

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 S_0 (derived in step 1 of the pseudo-code block). I suppose the reason for this is the fact that S_0 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.

Response: Thanks. We shift the curves of SBO to the right by adding the number of initial simulations along with the x-axis (marked as **Change 2.1-1**). In our work, we

used Latin Hypercube Sampling (LHS) method and sampled 18 initial points to start our SBO algorithms.

Moreover, we have also revised the third paragraph of the Section 4.3.1 (lines 6-10 on page 12) to clarify the advantages of our SBO as: (marked as **Change 2.1-2**)

“Compared to the global optimization algorithms, SBO has two main advantages. First, SBO samples an initial parameter set to build a surrogate model, and the building process is a learning process which can better understand the parameter space, thus help conduct better optimization. Second, SBO can avoid some bad parameter points (‘bad’ means high prediction error), which are not supposed to be evaluated with the help of the surrogate model.”

Comment 2.2:

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.

Response: Thanks. This is an interesting argument for our experiments of both SBO and MCMC. We carefully re-analyzed both designs of experiments and try our best to explain the reason of mismatch of results between our SBO and MCMC.

First, the objective function of optimization algorithms (such as DE, PSO or our SBO) is averaged RMSE between the results of model and IGBP-DIS observation data in our case. The target of optimization algorithms is to minimize the averaged RMSE.

Different from optimization algorithms, the Bayesian MCMC approach do not have an optimization objective function or loss function. MCMC is an adaptive sampling method and we use maximum likelihood estimator (MLE) to estimate the parameters from the samples generated by MCMC. MCMC use an acceptance probability to determine whether accept a new parameter point or not. The probability of acceptance can be summarized as follows (Marshall et al., 2004):

$$P(c^{k-1}|c^{new}) = \min\left\{1, \frac{P(O|c^{new})P(c^{new})}{P(O|c^{k-1})P(c^{k-1})}\right\}$$

$P(O|c)$ is a likelihood function of parameter c ; $P(c)$ is prior probability density function of parameter c ; Assuming that the prediction errors are normally distributed and uncorrelated, hence, the $P(O|c)$ can be calculated as follows (Hararuk et al., 2014):

$$P(O|c) \propto \exp\left\{-\sum_{i=1}^N \frac{(X_i - O_i)^2}{2\sigma_i^2}\right\}$$

Where O_i is soil C reported by IGBP-DIS at i th grid cell, X_i is soil C simulated by CLM-CASA' or microbial models at i th grid cell; σ_i^2 is the variance of a measurement at a grid cell. From the two equations above, the MCMC tends to keep the parameter points, whose $\sum_{i=1}^N \frac{(X_i - O_i)^2}{2\sigma_i^2}$ (named as r') are lower, which is different from the objective function or cost function of optimization (averaged RMSE presented in Eq.17 in our manuscript):

$$r = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - O_i)^2}$$

Owing to the difference between r and r' , the final estimation of MCMC could be a little bit different from the solution of parameter calibration with optimization algorithms in our case.

Comment 2.3:

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.

Response: Thanks. We have carefully edited the manuscript and improved English presentation in our revised manuscript.

Comment 2.4:

From the reply it seems that the authors do not 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 optimization 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.

Response: Thanks for pointing out this problem again. We agree with the reviewer and

have modified the corresponding sentences in our revised version to limit the performance comparison.

Comment 2.5:

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?*

Response: Thanks. We have refactored the Section 4.2 and given the detail description of the MCMC for better clarification (marked as **Change 2.5**).

The proposal step generates a new parameter set p^{k+1} from the previously accepted parameter set p^k through a proposal distribution $q(p^{k+1}|p^k)$. The proposal distribution is obtained using 50,000 simulations sampled following a uniform distribution. In the moving step, the probability of acceptance is calculated and used to determine whether accept the new parameter set p^{k+1} . And then the accepted one is added to the final sample set.

As mentioned in the Response of Comment 2.4, the MCMC method tends to keep the parameter points, whose $\sum_{i=1}^N \frac{(X_i - O_i)^2}{2\sigma_i^2}$ are lower. We use MLE estimator to run once on the final sample set generated by the AM algorithm to estimate the best parameter values, that are approximate to the minimum of $\sum_{i=1}^N \frac{(X_i - O_i)^2}{2\sigma_i^2}$.

Comment 2.6:

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.

Response: Thanks for pointing out this issue We have added a new reference of the surrogate model optimization toolbox.

Comment 2.7:

P8L25: it is not clear what is meant with “exploration” and “exploitation”

Response: In our work, “exploration” refers to searching in an unfamiliar area of the parameter space to learn about it and avoid trapping into some local optimum, and “exploitation” means fast convergence in some area. Balancing both the exploration and the exploitation makes our SBO can find global optimum and do not waste more model runs on the meaningless parameter areas.

Comment 2.8:

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

Response: We have improved the presentation for our manuscript, and provided the high quality graphs of Figure 7.

Comment 2.9:

Figure 2: why do the cells near the cost have no data?

Response: Same as the response for Comment 1.4.

Comment 2.10:

Table 1: what is meant with “exit rate”? Turnover/decomposition rate? If so I would suggest to use these terms

Response: Yes. We have replaced “exit rate” with “decomposition rate” in the revised version.

Parameter Calibration in Global Soil Carbon Models Using Surrogate-based Optimization

Haoyu Xu¹, Tao Zhang^{1,2}, Yiqi Luo^{2,3}, Xin Huang^{1,2}, Wei Xue^{1,2}

¹Department of Computer Science and Technology, Tsinghua University, Beijing 100084

²Department of Earth System Science, Ministry of Education Key Laboratory for Earth System Modelling, Tsinghua University, Beijing 100084

³Center for Ecosystem Science and Society, Northern Arizona University, USA

Correspondence to: Wei Xue (xuewei@tsinghua.edu.cn)

Abstract. Soil organic carbon (SOC) has a significant effect on carbon emission and climate change. However, the current SOC prediction accuracy of most models is very low. Most evaluation studies indicate that the prediction error mainly comes from parameter uncertainties, which can be improved by parameter calibration. The data assimilation technique has been successfully employed for the parameter calibration of SOC models. However, data assimilation algorithms, such as sampling-based Bayesian Markov Chain Monte Carlo (MCMC), generally have high computation costs and are not appropriate for complex global land models. This study proposes a new parameter calibration method based on surrogate optimization techniques to improve the prediction accuracy of SOC. Experiments on three types of soil carbon cycle models, including Community Land Model with Carnegie-Ames-Stanford Approach biogeochemistry sub-model (CLM-CASA') and two microbial models show that the surrogate-based optimization method is effective and efficient in terms of both accuracy and cost. Compared to the predictions using the tuned parameter values through Bayesian MCMC, the root mean squared errors (RMSEs) between the predictions using the calibrated parameter values with surrogate-base optimization and the observations could be reduced up to 12% for different SOC models. Meanwhile, the corresponding computation cost is lower than other global optimization algorithms.

1 Introduction

Soil organic carbon (SOC) is the largest pool of global land carbon (Todd Brown et al., 2013; Luo et al., 2015). The emission of CO₂, the most important greenhouse gas, from land ecosystems greatly depends on the amount of carbon stored in soils. Moreover, anthropogenic CO₂ emission leads to climate warming (Houghton et al., 2001), which further stimulates soil carbon release, forming a positive feedback

between the carbon cycle and climate warming (Melillo et al., 2003; Friedingstein et al., 2006; Luo, 2007). In the fifth Coupled Model Intercomparison Project (CMIP5), the outputs of 11 Earth system models (ESMs) show great uncertainty in the SOC predictions. Despite the similarity in model structures (Huang et al., 2017), simulated soil carbon content varies six-fold, ranging from 510 to 3040 PgC, among the models (Todd-Brown et al., 2013). Only half of 11 models have a predicted global total SOC falling within the estimated range of the Harmonized World Soil Database (HWSD). Modelled SOC is hardly corrected with the observation (Luo et al., 2015).

Considering the high similarity in the structures of the 11 ESMs, the difference in the SOC simulations mainly results from parameterizations (Todd Brown et al., 2013). Thus, parameter calibration is among the top priorities to improve prediction of global land carbon cycle dynamics (Luo et al., 2016). However, the parameter calibration with global observations has not been widely applied, owing to the high computational cost. A matrix approach has been recently developed to reorganize the carbon balance equations in the original ESMs into one matrix equation without changing any modeled C cycle processes and mechanisms (Luo et al. 2003, 2017; Huang et al. 2018). The matrix land carbon cycle models can be semi-analytically solved to obtain steady-state solutions faster than the original models by tens and hundreds times (Xia et al. 2013). As a consequence, the matrix approach makes parameter estimation and calibration possible. The matrix approach has been successfully used for the parameter calibration to constrain SOC turnover and microbial process with Bayesian Markov chain Monte Carlo (MCMC) algorithm (Harauk et al., 2014, 2015; Shi et al. 2018).

Bayesian MCMC is a sampling-based approach and usually requires a large number of simulations for building an acceptable parameter chain. For instance, over 500,000 simulations are required during the parameter calibration of soil carbon models (Xu et al. 2006). Even using high-performance computers, complex land models, like the latest version of Community Land Model (CLM5.0), require a very long spin-up time for carbon cycle simulation, leading to several hours or days for one simulation. Although the matrix approach has been developed to enable data assimilation of global land carbon cycle models (Harauk et al., 2014 and 2015, Shi et al. 2018), Bayesian MCMC computationally is still very expensive for calibrating global land models. More effective and efficient parameter calibration algorithms are urgently needed.

The parameter calibration of SOC models can be formulated as an optimization problem that aims to minimize the output of a cost function. This cost function evaluates the difference between the outputs

of model simulation and the corresponding observations and returns a single value (e.g., RMSE) to represent the model error. Global optimization algorithms are introduced to find the minimum value of the non-linear, non-convex, and black-box problems (Hapuarachchi et al., 2001; Ma et al., 2006; Rocha, 2008). Unfortunately, the number of required simulations of most global optimizations is very large.

To reduce the number of simulations and decrease the computational cost, we, for the first time, present a surrogate-based optimization method (SBO) for calibrating the soil carbon models. Surrogate models serve as computationally inexpensive approximations of expensive simulation models (Booker et al., 1999), such as complex geoscientific models. During the optimization process, the surrogate model can be used to determine the new promising point in the parameter space at which the expensive simulation model originally has to be evaluated. With the help of the surrogate model, many unnecessary simulations with bad parameter values, which lead to high prediction errors, are avoided. SBO has been showed to find the near-optimal parameter values within only a few hundred simulations for different problems (Aleman et al., 2009; Giunta et al., 1997; Regis, 2011; Simpson et al., 2001).

Most studies on both global and surrogate optimizations focus on the mathematical function benchmarks like Comparing Continuous Optimizers, abbreviated as COCO (Hansen et al., 2010; Wang and Duan, 2014). However, the optimization of the mathematical functions may be extremely different from the parameter calibration of complex real-world models. In this paper, we explore the state-of-the-art surrogate optimization method for parameter calibration of three SOC models: CLM coupled with Carnegie-Ames-Stanford Approach biogeochemistry model (CLM-CASA') and two microbial models, as used in studies by Hararuk et al. (2014, 2015). Although the three models are computationally attainable for parameter calibration, we compare the performance of surrogate-based optimization to advanced global optimization algorithms and the data assimilation method to examine the potential of SBO. The SBO may be extended to other complex global land models.

In this paper, we present the structure and parameters of three SOC models in Section 2. Section 3 introduces the algorithm design of SBO. The parameter calibration results and the analysis of different parameter calibration algorithms are presented in Section 4. Section 5 discusses the calibrated results using SBO. Finally, we draw conclusions in Section 6.

2 Global Land Carbon Models, Data, and Cost Function

Earth system models (ESMs) are a fundamental tool for simulating climate impacts on the carbon cycle at the global scale. There are many common properties among structures of different global land carbon modules of ESMs (Luo et al. 2017). Almost all models have multiple carbon pools. Carbon is transferred among these pools (Weng and Luo, 2011). In this study, we selected three SOC models, which have been previously calibrated for their parameters with Bayesian MCMC algorithm (Hararuk et al., 2015). The first model is the soil carbon component of CLM coupled with Carnegie-Ames-Stanford Approach biogeochemistry submodel (CLM-CASA') (Oleson et al., 2004, 2008). The CLM is the land model for the Community Earth System Model (CESM). The other two SOC models are microbial models, which consider nonlinear regulation of SOC dynamics with microbial biomass.

2.1 CLM-CASA' C-only Version Model

The CLM-CASA' is embedded in CLM3.5. The latter includes biogeophysics and biogeochemistry submodels. CLM-CASA' inputs carbon through net primary productivity (NPP), which is partitioned to three live biomass pools (wood, leaves and fine roots) (Fig. 1a). Dead plant materials become litter and are transferred separately to four litter pools. Litter decomposition results in part of carbon released to the atmosphere as heterotrophic respiration and part of carbon being stabilized into soil carbon pools. Organic carbon in the soil pools is decomposed partially to be released as CO₂ via microbial respiration and partially to be incorporated into other soil carbon pools. One of the key model outputs to indicate SOC dynamics is the total soil organic carbon content, which is the sum of carbon in soil microbial (or active), slow and passive pools (Fig.1).

The CLM-CASA' model simulates soil carbon decomposition as a first-order decay process (Oleson et al., 2004, 2008). Based on theoretical analysis, the carbon cycle of most ESMs can be summarized with one matrix equation (Luo et al. 2001, 2017; Luo and Weng, 2011; Xia et al., 2013) as.

$$\frac{dX(t)}{dt} = A\xi(t)KX(t) + BU(t) \quad (1)$$

Where $X(t)$ is the carbon content of different pools; $\frac{dX(t)}{dt}$ is the change of the carbon content; A is a matrix of transfer coefficients among different pools; $\xi(t)$ and K are both diagonal matrixes, representing environmental scaling factors and baseline carbon decomposition rates, respectively; $U(t)$

is NPP, the carbon influx into the whole system and B represents the partitioning coefficients of the carbon influx among plant pools. The steady state solution of equation is given by Xia et al. (2012) as:

$$X_{ss} = -(A\underline{\xi}K)^{-1}\underline{B}\underline{U} \quad (2)$$

Where $\underline{\xi}$, \underline{B} , and \underline{U} are the long-term averages of the environmental scalars, C partitioning among the three live pools, and NPP, respectively. The steady state soil C generated by this C-only version is in agreement with that simulated by original CLM-CASA' model (Xia et al., 2012). The structural diagram of the CLM-CASA' C-only model are presented in Fig. 1a and the parameters are described in Table 1.

2.2 The Microbial Models

The microorganisms catalyse various processes of land carbon cycle, such as decomposition and stabilization of SOC (Kuzyakov et al., 2000; Luo et al., 2001; Peng et al., 2009). However, most conventional SOC models, such as CLM-CASA', do not explicitly represent microbial processes. Microbially explicit models usually represent SOC decomposition by considering extracellular enzyme activities rather than simple decay constants as in the CLM-CASA' and other traditional SOC models (Schimel and Weintraub, 2003). In this study, we focused on two enzyme-driven decomposition models; one has two pools (Fig. 1b) introduced by German et al. (2012) and the other has four pools (Fig. 1c) introduced by Allison et al. (2010). We call these two models the two-pool microbial model and the four-pool microbial model, respectively. C inputs for the two models are NPP and the outputs are the carbon content of each pool at a steady state.

The two-pool microbial model is described using the following equations (Hararuk et al., 2015).

$$\frac{dMIC}{dt} = CUE \times V_{max} \times MIC \frac{SOC}{K_m + SOC} - r_d \times MIC \quad (3)$$

$$\frac{dSOC}{dt} = Input_{soil} + r_d \times MIC - V_{max} \times MIC \frac{SOC}{K_m + SOC}$$

(4)

Where

$$CUE = CUE_{slope} \times T_s - CUE_0 \quad (5)$$

$$V_{max} = V_{max_0} \times \exp\left(-\frac{E_a}{R \times (T_s + 273)}\right) \times \exp(-par_{clay} \times clay)$$

(6)

$$Km = Km_{slope} \times T_s + Km_0 \times \exp(par_{lig} \times lignin)$$

(7)

where MIC represents the microbial biomass, V_{max} is the temperature adjusted rate of SOC decomposition; K_m is the half - saturation constant for substrate - limited SOC decomposition rate; r_d is the microbial death rate; CUE is the microbial carbon use efficiency; $Input_{soil}$ is the carbon influx to soil, a 30-year averages of soil C input produced by CLM-CASA' (Hararuk et al., 2015); T_s is soil temperature; R is the gas constant (8.31 J K⁻¹ mol⁻¹); CUE_0 and CUE_{slope} are the baseline microbial carbon use efficiency and its dependency on temperature, respectively; V_{max0} is the maximum rate of microbial carbon uptake; E_a is the activation energy of SOC decomposition; and Km_0 and Km_{slope} are the baseline half-saturation constant and its dependency on temperature, respectively; $lignin$ is lignin content; and par_{lig} is a parameter to regulate the lignin - dependent correction factor. See Table 2 for more description of those parameters.

The four-pool microbial model from Allison et al. (2010) is described as follows:

$$\frac{dMIC}{dt} = V_{maxup} \times MIC \frac{DOC}{Kmup+DOC} \times CUE - r_d \times MIC - r_{EnzProd} \times MIC$$

(8)

$$\frac{dDOC}{dt} = a_{lit-to-DOC} \times Input_{soil} + r_d \times MIC \times (1 - a_{MIC-to-SOC}) + V_{max} \times ENZ \frac{SOC}{Km+SOC} + r_{EnzLoss} \times ENZ - V_{maxup} \times MIC \frac{DOC}{Kmup+DOC}$$

(9)

$$\frac{dSOC}{dt} = a_{lit-to-SOC} \times Input_{soil} + r_d \times MIC \times a_{MIC-to-SOC} - V_{max} \times ENZ \frac{SOC}{Km+SOC}$$

(10)

$$\frac{dENZ}{dt} = r_{EnzProd} \times MIC - r_{EnzLoss} \times ENZ \tag{11}$$

where ENZ and DOC are enzyme and dissolved organic carbon pools, respectively; V_{maxup} is the temperature adjusted rate of DOC uptake by microbes; $Kmup$ is a half - saturation constant limiting microbial uptake of DOC ; $r_{EnzProd}$ is a rate of enzyme production; $Input_{soil}$ is C transferred from litter to soil; $a_{lit-to-DOC}$ is the fraction of $Input_{soil}$ that is transferred to DOC ; $a_{MIC-to-SOC}$ is the fraction of dead

microbes transferred to soil; and $r_{EnzLoss}$ is the rate of enzyme loss. The temperature-dependent functions are:

$$CUE = CUE_{slope} \times T_s - CUE_0 \quad (12)$$

$$V_{maxup} = V_{maxup_0} \times \exp\left(-\frac{E_{aup}}{R \times (T_s + 273)}\right) \quad (13)$$

$$Kmup = Kmup_{slope} \times T_s + Kmup_0$$

(14)

$$V_{max} = V_{max_0} \times \exp\left(-\frac{E_a}{R \times (T_s + 273)}\right) \times \exp(-par_{clay} \times clay)$$

(15)

$$Km = Km_{slope} \times T_s + Km_0 \times \exp(par_{lig} \times lignin)$$

(16)

where V_{maxup_0} is the maximum rate of microbial DOC uptake; E_{aup} is the activation energy of DOC uptake; $Kmup_0$ and $Kmup_{slope}$ are baseline half - saturation constants for substrate limitation of DOC uptake and its dependency on temperature, respectively.

Fifteen parameters of the four-pool microbial model are also described in Table 2.

2.3 Data and Cost Function

Microbial models and CLM-CASA' C-only models divide the world into 64*128 grid cells and output SOC content at each grid (Fig. 2). The observed SOC data for parameter calibration comes from the International Geosphere Biosphere Programme – Data and Information System (IGBP-DIS) dataset (Global Soil Data Task Group, 2000). The IGBP-DIS dataset includes a 1-km resolution global land carbon data set that has been widely used in many studies to evaluate and improve models (Zhou et al., 2009; Smith et al., 2013).

The goal of parameter calibration is to improve the SOC predictions to better fit the observations. Therefore, we use the root mean squared errors (RMSEs) between the model SOC predictions and the observations at all grid cells as the cost function. This cost function can be described as the following formula:

$$r = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - O_i)^2}$$

(17)

Where N denotes the total number of grid cells, X_i and O_i are the SOC of model prediction and IGBP-DIS observation, respectively. To avoid overfitting and evaluate the calibrated parameters more fairly, we separate all grid cells into a training set and a validation set. The training set is used to guide the parameter calibration process and the validation set is used to evaluate the calibrated results. Hararuk et al. (2014 and 2015) also used this method when calibrating SOC parameters with the Bayesian MCMC approach. The experiment results in Sections 3 and 4 refer to the results for the validation set.

3 Surrogate-Based Optimization Algorithm Design

3.1 Introduction to the Surrogate-Based Optimization Algorithm

The parameters of most soil carbon models and land models have been traditionally are tuned manually (Luo et al. 2001, 2016). The manual tuning method might be effective for simple models but still highly depends on expert experience. Complex models may consist of various components from different disciplines and have hundreds or thousands of parameters. It becomes impractical for manual tuning.

Available are different parameter calibration algorithms, which have been developed based on optimization theory. The gradient search algorithms like the quasi-Newton method are introduced to search for a set of parameters with better performance in the parameter domain. These algorithms are usually efficient and fast. However, the gradient search algorithms are designed for finding the local optimum. They cannot be used to solve the multimodal problems derived from complex earth system models. In addition, they are based on the gradient information, which is unavailable for most soil carbon and land models. These models are too complex to obtain the gradient information. Thus, the parameter calibration usually becomes a black-box optimization problem. Global optimization algorithms, such as genetic algorithms and particle swarm optimization algorithms, are based on parameter generation and selection strategies. They basically are still gradient independent but can be easily used for parameter calibration of complex earth system models. Global optimization algorithms are designed to find the global minimum. However, the number of samples (model runs) might be still too large to be applicable to complex models with large number of parameters (Jones et al., 1998). Moreover, complex earth system

models, for example CLM, require several hours over hundreds of cores for only one sample run and pose a special challenge for the feasibility of automatic parameter calibration.

SBO is an efficient and effective automatic parameter calibration framework. It fits a surrogate model (or response surface) based on the previous samples and uses this surrogate model to emulate the output behaviours of original models with an acceptable level of accuracy. The main advantage of SBO is to save computational costs during the global optimization by using the surrogate model instead of the original model. And the surrogate model can be continuously improved by exploiting new sample runs with the original model. With the surrogate model, the algorithm can make full use of previous samples' information and reduce the sample size, time-to-solution, as well as the computation cost. SBO has been successfully used to solve the parameter calibration of computationally expensive black-box problems (Vu et al., 2016).

3.2 Key Components of the Surrogate-Based Optimization Algorithm

The flowchart of the SBO is presented in Fig. 3.

First, initial sets of parameter values are generated using a sampling method. These sets are then used as inputs to run the real simulation model. Second, a surrogate model is constructed by fitting the outputs of these sample runs. The surrogate model serves as a computationally inexpensive approximation of the expensive simulation model (Booker et al., 1999). Then in each iteration, new sample points simulated by the real model are generated according to a specific strategy. This strategy can make use of the information gained from the surrogate model and only exploits the avoidable real model runs to meet the accuracy requirement. The new sample points and their simulation outputs are used to update the surrogate model at the same time. Finally, when some stop criteria (typically the maximum number of simulations allowed) are met, the algorithm returns the optimized parameter values. During the SBO process, quite a few sample runs are generated based on the evaluation of the surrogate model and most meaningless simulations with bad parameter values are avoided. As a result, the computationally expensive model is simulated at only a few selected promising parameter points, and the surrogate model will replace the real model during the calibration process. Thus, the computation cost is reduced substantially.

Different surrogate-based optimization algorithms may have different choices with respect to the following:

-
- ✧ The sampling method to generate the initial set S_0 .
 - ✧ The surrogate model, which predicts the output y using the given data point x . Before prediction, some (x, y) data pairs should be given to train the model and the data are called the training set.
 - ✧ How to decide the new points at which to run the real model in each iteration.

For the initial sampling, Monte Carlo sampling and Latin Hypercube sampling (LHS) are two main sampling methods (McKay et al., 1979; Iman et al., 1981). In Monte Carlo sampling, values are sampled from a probability distribution, which is generally a uniform distribution unless we have additional knowledge about the model and the parameters. During the LHS procedure, the range of each parameter is divided into M equally probable intervals. M sample points are selected to cover all intervals of each parameter. Compared to the random sampling, LHS ensures that the ensemble of random numbers is representative of the real variability of the parameters. As a result, we use LHS to generate the initial set S_0 (Iman et al., 1981).

There are various surrogate models, such as multivariate adaptive regression splines (Mars) (Friedman, 1991), polynomial regression models (Myers and Montgomery, 1995), radial basis functions (RBFs) (Gutmann, 2001; Müller et al., 2014; Powell, 1992; Regis and Shoemaker, 2007, 2009; Wild and Shoemaker, 2013), and kriging (Davis and Lerapetritou, 2009; Forrester et al., 2008; Jones et al., 1998).

The Mars model is an extension of naïve linear models introduced by Friedman J H. (Friedman J H., 1991). The form of Mars is presented as follows:

$$\hat{f}(x) = \sum_{i=1}^m c_i B_i(x) \quad (18)$$

Where $\hat{f}(x)$ represents the prediction of y at the point x , and c_i is a constant coefficient to be trained. The $B_i(x)$ is the basis function which can take one of the following three forms: a constant, a hinge function like $\max(0, x - \text{const})$, and a product of more than one hinge function.

The RBF model is a real-valued function. The prediction at a point x using the RBF model only depends on the distance between x and other points in the training set, whose outputs have been already given. The distance $r = \|x, c\|$ is generally the Euclidean distance. The radial function is the function that satisfies the property $\phi(x, c) = \phi(\|x, c\|) = \phi(r)$. The prediction at point x with the RBF model is formulated as:

$$\hat{f}(x) = \sum_{i=1}^N w_i \phi(\|x, x_i\|) \quad (19)$$

Where the x_i represents the point of the training set which has N points in total. Many different radial functions have been introduced and some commonly-used ones are Gaussian $\phi(r) = e^{-(\epsilon r)^2}$, multiquadric $\phi(r) = \sqrt{1 + (\epsilon r)^2}$, and polyharmonic spline: $\phi(r) = r^k \ln(r)$. In our experiments, we choose the Gaussian radial function.

Both the kriging model and the Gaussian process regression model predict the output using a Gaussian process governed by prior covariance. The x and y should be normalized to satisfy a normalization distribution where the means is 0 and the covariance is 1 before they are used to train the kriging model.

The kriging predictor can be found as follows:

$$\hat{f}(x) = \hat{\mu} + \sum_{i=1}^n c_i r_i(x) \quad (20)$$

Where $\hat{\mu}$ is the estimated mean of the Gaussian process, c_i is a constant representing the weight and $r_i(x) = \text{Corr}(x, x^{(i)})$ is the correlation between the x and the i th point $x^{(i)}$ in the training set. $\hat{\mu}$ and c_i can be trained with the training set.

In addition, many machine learning regression models are also introduced, such as support vector regression (Zhang et al., 2009), artificial neural network (Behzadian et al., 2009), and random forest (Breiman, 2001).

The strategies of parameter point generation are iterative algorithms that use data acquired from previous iterations to guide new parameter point generation. Most strategies convert the parameter point generation to optimization problems using an evaluation criterion (Fig. 3). There are many different generation strategies, including Minimizing an Interpolating Surface (MIS) (Jones, 2001) and Maximizing Expected Improvement (MEI) (Schonlau et al., 1997; Picheny et al., 2013). In MIS, the minimum of the surrogate model response surface is found and treated as the new parameter point to evaluate the real simulation model and then update the surrogate model. MEI introduces the ‘‘expected improvement’’ criterion. This criterion estimates the uncertainty of the surrogate model and balances the exploration and exploitation. Exploration refers to searching in an unfamiliar area of the parameter space to learn about it and avoid trapping into some local optimum. Exploitation means fast convergence in some area. Balancing the exploration and exploitation ensures SBO can find real global optimum and does not waste more simulations on the meaningless parameter sets and areas. Another parameter generation strategy is candidate point approach (CAND) (Regis and Shoemaker, 2007). In the CAND strategy, the criterion for exploitation is MIS and the criterion for exploration is the distance of the

candidate point to the set of sampled parameter points from previous iterations. The previous sampled points represent the explored region and we can estimate the uncertainty with the distance to the explored region. A weighted sum of these two criteria is used to determine the new parameter point during the SBO.

3.3 Design of the Surrogate-Based Optimization Algorithm for Soil Carbon Models

Based on the previous introduction of SBO, the detailed procedure of SBO can be found as in the following box.

- Step 1: Generate an initial sampling set S_0 .
- Step 2: Run the real model and calculate the output error of the parameter points of S_0 .
- Step 3: Build the surrogate model using the parameters and the outputs generated in Step 2.
- Step 4: Predict the output errors of those points that do not belong to S_0 using the surrogate model and determine the points at which to run the real model.
- Step 5: Run the real model again for the new parameter points of Step 4 and calculate the output errors of these selected points.
- Step 6: Update the surrogate model with the new data from Step 5.
- Step 7: Iterate through Steps 4 to 6 until the end condition has been met.

The SBO scheme mentioned in previous sections is a parameter calibration framework. The key components introduced in Section 3.2 must be selected when calibrating the parameters of soil carbon models. The LHS can cover the whole parameter space with a limited number of sample points while Monte Carlo sampling usually requires much a larger number of samples. Therefore, we choose the LHS as the initial sampling strategy.

As mentioned in the previous section, many kinds of surrogate-based models have been introduced and developed. The machine learning regression models do not perform as well as RBF and kriging models, according to the evaluation of similar cases (Wang et al., 2014). In this study, we use the RBF surrogate model (RBF-SBO) as our default choice because it has been showed to perform better than other surrogate models (Müller and Shoemaker, 2014). It is easy to be implemented. Our algorithm framework also includes other surrogate models, such as kriging and Mars, and can introduce others in the future.

The surrogate model are not accurate to represent complex and nonlinear models when the SBO starts. The MIS can be very efficient but easy to trap into local optima, since the strategy does not consider the

uncertainty of the surrogate model and only select the optimum of the surrogate model. The MEI eliminates the disadvantage of MIS but can only be used for the kriging surrogate model because the calculation of the expected improvement requires the standard error at the parameter point and only the kriging (Gaussian Process) surrogate model can provide the standard error (Jones et al., 1995). Finally, we use the CAND strategy as the parameter generation strategy in our algorithm to balance the exploitation and exploration of uncertain region.

4 Parameter Calibration Experiments

4.1 Experiment Configuration

In this study, we select the Bayesian MCMC approach and four advanced global optimization algorithms for comparison with our proposed SBO method. Our SBO algorithm is implemented based on the toolkit “Surrogate Model Optimization Toolbox” (Müller, 2014). Three SOC models and their cost functions are introduced in Section 2. The target of parameter calibration is to find the optimal values of parameters to achieve the minimum value of the cost function (average RMSE). Moreover, we repeat the parameter calibration process of each algorithm 50 times and use the average results for algorithm evaluation. We compare the performance of algorithms in terms of both effectiveness and efficiency. The effectiveness refers to the accuracy of the calibrated results and the efficiency can be evaluated by the required simulation times of the original SOC models.

4.2 Various Global Optimization Algorithms and the Bayesian MCMC Approach

The Bayesian MCMC approach and four advanced global optimization algorithms: differential evolution (DE), particle swarm optimization (PSO), shuffled complex evolution (SCE-UA), and the covariance matrix adaption evolution strategy (CMA-ES), are compared with our RBF SBO.

DE (Storn and Price, 1997) and PSO (Kennedy, 1995; Shi and Eberhart, 2009) are the representative algorithms of the evolution strategy and swarm intelligence, respectively. They both have the ability to converge quickly and outperform many genetic algorithms and simulated annealing algorithms (Price and Storn, 2006; Shi and Eberhart, 2009). SCE-UA is designed for the parameter calibration of hydrologic models and has been successfully applied to various hydrology models such as the TOPMODEL, the Xinanjiang watershed model and short-term load forecasting (Hapuarachchi et al.,

2001; Ma et al., 2006; Li et al., 2007). SCE-UA tries to keep both effectiveness and efficiency by combining the local (the simplex method) and global optimization methods. Despite the difference in algorithm details, DE, PSO, and SCE-UA all generate new parameter points according to some simple mathematical formulas. Unlike these three algorithms, CMA-ES creates new parameter points based on a multivariate normal distribution (Hansen and Ostermeier, 2001; Hansen and Kern, 2004). The dependencies between parameters are represented by the covariance matrix of a normal distribution. CMA-ES has been shown to be the best global optimization algorithm in the BBOB-2009 comparison study (Hansen, 2009).

The Bayesian MCMC approach is typically designed to obtain the posterior distributions of model parameters but it can also be used to calibrate parameters to reduce the prediction error. The Bayesian MCMC consists of two main parts: sampling and parameter estimation. During the sampling part, the adaptive Metropolis (AM) algorithm, a Markov chain Monte Carlo method, is used to conduct sampling from the *prior* parameter distributions and generate a parameter chain (Haario et al., 2001). The AM algorithm has two steps: the proposing step and the moving step. A new parameter set p^{k+1} is generated from the previously accepted parameter set p^k through a proposal distribution $q(p^{k+1}|p^k)$. In the moving step, the probability of acceptance is calculated according to the Metropolis criterion (Xu et al. 2006). The parameter set that is not accepted is discarded. The AM algorithm repeats the proposing step until the new parameter set is accepted. The accepted new parameter set becomes the p^{k+1} set of accepted parameters in the posterior parameter distribution (Marshall et al., 2004). The proposal step is usually repeated for 50,000 to one million times to generate enough accepted parameter sets for the posterior parameter distribution. The posterior distribution is used to estimate the Maximum likelihood estimator (MLE). Hararuk et al. (2014, 2015) applied the Bayesian MCMC approach to the parameter calibration problem of the CLM-CASA' C-only model and microbial models. Hararuk et al. (2014, 2015) conducted experiments in which the proposing step required 50,000 simulations for microbial models and 1,000,000 simulations for the CLM-CASA' model. We used the code from Hararuk et al. (2014, 2015) and repeated the calibration experiments. The detailed calibration results from the Bayesian MCMC approach are presented in Table 3.

4.3 Results and Analysis

4.3.1 Effectiveness and Efficiency

Fig. 4 presents the calibrated results (RMSE) of the different algorithms we applied. For each algorithm, we only perform 100 simulations to compare the effectiveness if the simulation times are limited. As the Bayesian MCMC approach requires a large number of samples to reach a stable distribution, over 500,000 simulations have been conducted for the algorithm evaluation.

Clearly, the average RMSE of the RBF-SBO is the lowest (0.6 kg/m², better than the Bayesian MCMC algorithm) for the two microbial models among all the algorithms (Fig. 4b, c). For the CLM-CASA' model, our RBF-SBO algorithm is still superior to the global optimization algorithms. The Bayesian MCMC approach performs slightly better (about 0.02 kg/m²) but requires many more simulations to achieve the results (Fig. 4a).

The results of RBF-SBO also indicate less variation among the 50 repeated experiments than the global optimization algorithms for the three models. For the same reason mentioned before, the Bayesian MCMC approach has less variation than our RBF-SBO algorithm for the two microbial models. For the CLM-CASA' model, our RBF-SBO is still promising to get stable results. Among the global optimization algorithms, CMA-ES shows a very significant fluctuation (Fig. 4b, c), indicating that it is unreliable when the number of simulations is small as 100. This is because the CMA-ES requires quite a few simulations on the exploration of the parameter domain and the construction of the parameter covariance matrix. Therefore, RBF SBO is the most effective and stable one when the number of simulations is limited.

Fig. 5 shows the results in terms of average validation RMSE. We don't compare the efficiency of Bayesian MCMC since it is in nature a sampling algorithm, not an optimization algorithm. The average validation RMSE of RBF SBO is lower than the four global optimization algorithms before the number of simulations increases to 600 for two microbial models and to 200 for the CLM-CASA' model, respectively. Our RBF-SBO requires fewer simulations than the global optimization algorithms when they reach the same RMSE value and accuracy range. Thus, our RBF surrogate optimization is also the most efficient algorithm that requires the minimum simulation times, as well as computational costs. Compared to the global optimization algorithms, SBO has two main advantages. First, SBO samples an initial parameter set to build a surrogate model, and the building process is a learning process which can

better understand the parameter space, thus help conduct better optimization. Second, SBO can avoid some bad parameter points ('bad' means high prediction error), which are not supposed to be evaluated with the help of the surrogate model.

Another important observation is that the difference between the results of our RBF-SBO and the global optimization algorithms decreases as the number of simulation increases (Fig. 5). Moreover, the CMA-ES outperforms the RBF-SBO when the number of simulations exceeds 200 for the CLM-CASA' model (Fig. 5a). Our SBO can build the surrogate model with relatively good accuracy quickly, which help find a near-optimal solution with lower computation cost. However, the surrogate model is only an approximation of the real model and the accuracy might be limited due to the strong nonlinearity and the high complexity of the real model. After gaining sufficient knowledge of the original model through many simulations, the excellent global optimization algorithms, such as CMA-ES, may achieve a similar performance or even outperform our SBO, which suggests that our SBO is better to use for the parameter calibration problem of cost-expensive models.

4.3.2 Impact of the Model Complexity

Compared to the two-pool and four-pool microbial models, the CLM-CASA' model has 13 carbon pools and 20 parameters. Despite the increasing complexity of the CLM-CASA' model, the SBO obtains better results before conducting 200 simulations of the real model (Fig. 5a). Moreover, our SBO is always the best parameter calibration method for the two-pool and four-pool microbial models before conducting 600 simulations (Fig. 5b, c). In addition, only one global optimization algorithm, CMA-ES, has better performance compared to our SBO on the CLM-CASA' model after 200 simulations. Considering the high variance of CMA-ES on two microbial models (Fig. 4b, c), our SBO is more effective and more reliable on average.

4.3.3 Impact of Different Types of Surrogate Models

We select the RBF as the surrogate model in the experiments because the RBF is the widely-adopted one in many SBO algorithms (Müller and Shoemaker, 2014). In this section, we also test two other typical surrogate models, kriging and Mars. The Mars model is simple and has almost no requirements for the sample quality. Mars is very quick to train and predict. Kriging, also known as Gaussian process regression, is a method of interpolation for which the interpolated values are modelled by a Gaussian

process governed by prior covariance. Kriging provides the best linear unbiased prediction of the intermediate values under suitable assumptions on the priors.

Figure 6 presents the results of kriging, Mars and RBF in terms of average validation RMSE. The performance of the three surrogate models is similar. The three surrogate models all have reasonable performance in the parameter calibration of the three types of SOC models and perform better than global optimization algorithms, indicating that our SBO is robust.

5 Analysis of Parameter Calibration Results

5.1 CLM-CASA' Model

The steady state global SOC simulations (Eq. 2) using CLM-CASA' with the default and calibrated parameter values are presented in Fig. 7a and b, which are also compared to the observed SOC pools provided by the IGBP-DIS dataset. The SOC simulation results using the calibrated parameter values from the SBO matches the observation better than that with the default parameter values (Fig. 7c) with a relatively lower RMSE. By using the calibrated parameter values, the SOC simulations are significantly improved in most parts of the world, except some grid cells in the west of Canada and the east of Russia (Fig. 7a, b and c). As a result, the CLM-CASA' simulation result with the default parameter values can only explain 33% of variation in the observed soil C, whereas that with the calibrated parameter values can explain an improved ratio (42%) of variation in the observed soil C. The unexplained variation is partly due to uncertainty in observations. To further improve the model's accuracy, we need to gain a more in-depth understanding of uncertainty sources from the data, model structure, parameters, and forcing.

Figure 8 presents the frequency distributions of the 20 calibrated parameters based on MCMC and the calibrated parameter values using the proposed SBO (the blue lines in Fig. 8). Narrow posterior distributions indicate highly sensitive parameters, consistent with the conclusions of Hararuk et al. (2014) and Post et al., (2008). The calibrated parameter values of SBO are quite similar to the responding parameter values at the peaks of posterior distributions for most highly sensitive parameters, such as the temperature sensitivity of heterotrophic respiration (Q_{10}) and clay effect on C partitioning from slow to passive pools (t_7). The parameter calibration results (RMSE) of SBO and Bayesian MCMC are similar, consistent with the parameter calibration results listed in Table 3.

Some calibrated parameter values are very close to the assigned bounds of the parameters in Fig. 8, which is usually related to the correlations among parameters. Further investigation on the covariance among parameters is necessary to explain this issue. In addition, the unreasonable setting of those bounds might be another possible reason. For instance, the calibrated $c(12,12)$ value (1.01×10^{-3}) reaches its lower bound, indicating that passive SOC residence time almost approaches 1000 years.

As listed in Table 1, the calibrated temperature sensitivity (Q_{10}) decreases from 2 to 1.74. The size of soil microbial and passive pools increases due to the longer residence time of the passive pool and lower temperature sensitivity (Q_{10}). The size of the slow pool, on the contrary, decreases due to the increase in the decomposition rate from the slow pool or the decrease of its residence time. Comprehensively, the size of the SOC, which is the sum of carbon capacity in passive pools, slow pools and soil microbial pools, increases and closely approximates the observation.

5.2 The Microbial Models

According to the calibrated RMSE and r^2 , the SOC simulation of the two-pool and four-pool microbial models are very similar. Without the loss of generality, we only analyse the parameter calibration results of the four-pool microbial model in this section. After parameter calibration using the SBO, the global SOC produced by the four-pool microbial model is improved, especially in China, Russia, Europe, and North America (as shown in Fig. 9). Overall, the microbial models explain a higher fraction of global variability of the observed SOC data and have lower spatial RMSEs than the CLM-CASA' model (as listed in Table 3).

The microbial models achieve better SOC predictions than that of the calibrated CLM-CASA' model in terms of the prediction of the C capacity in the low-temperature regions (Russia, Europe, North America) and in the regions with small soil C inputs (Fig. 7b and 9). The SOC contents are determined by two main factors: the soil carbon inputs and the SOC residence time (Luo et al., 2003). Considering the same soil carbon inputs of the CLM-CASA' and the microbial models, the improvement is mostly induced by the differences in the SOC residence time. In all three models, the SOC residence time is essentially controlled by temperature (Xia et al. 2013). As a result, the temperature sensitivity (Q_{10}) contributes to the difference across the three models. The temperature sensitivity remains constant in the CLM-CASA'. However, both the microbial models calculate spatially variable Q_{10} with higher values in the low-temperature regions and lower Q_{10} in the high-temperature regions, which reflects the impact

of the temperature on the microbial activity. In addition, the SOC residence time can also be affected by the quality of SOC inputs and is related to the microbial decomposition processes. Fresh C input stimulates the microbial dynamics growth, resulting in an increase in the old SOC decomposition rate (i.e., priming effect) (Kuzyakov et al, 2000; Fontaine et al., 2004, 2007). Therefore, the microbial models simulate lower SOC residence times than the CLM-CASA' in the regions with a high SOC input and a high SOC residence time and the regions with a low SOC input. This is due to the nonlinearity of the substrate limitation in the microbial models (Eqs. 8 and 10), as well as the dependency of the residence time in microbial dynamics. Comprehensively, the introduction of microbial dynamics makes the microbial models predict SOC better than the CLM-CASA' model.

Figure 10 presents the posterior distributions of the parameters calculated by Bayesian MCMC and the parameter values calibrated using our SBO. According to the posterior distribution, r_d , CUE_{slope} , CUE_0 , E_a , par_{lig} , and par_{clay} are the most constrained and sensitive parameters. The calibration results of the SBO are consistent with the posterior distributions of these highly sensitive parameters (Fig. 10) except CUE_{slope} and CUE_0 . CUE_{slope} and CUE_0 are highly sensitive owing to their influence on temperature sensitivity. Due to the difference between CUE_{slope} and CUE_0 , the RMSE of SBO is 1.4 kg/m^2 and 0.8 kg/m^2 lower than those with Bayesian MCMC for four-pool and two-pool microbial models respectively (as listed in Table 3). The mismatch of CUE_{slope} and CUE_0 may be mainly due to the different targets of the parameter selection between the two methods.

6 Conclusions

Parameter calibration is becoming more and more challenging for SOC model development, especially for the computationally-expensive global land models owing to the large number of simulations. In this study, we introduce an SBO algorithm to the parameter calibration of three SOC models. The main findings are:

- 1) Compared to advanced global optimization algorithms, SBO is more effective and more efficient on average. Our RBF SBO outperforms other parameter calibration algorithms when the number of simulations is no more than 200.
- 2) The parameter optimization based on RBF surrogate model gains more accurate calibration results than those of the Bayesian MCMC approach in the three soil carbon models.

-
- 3) The SBO scheme is robust. Various types of surrogate models have similar performance in the parameter calibration tasks of SOC models.
 - 4) Although SBO is only guided by a single cost function, it can still result in better parameter values than the default ones. We carefully analyze the spatial SOC distributions produced by the models with the calibrated parameters using SBO, which indicates that SBO truly improves the model's prediction and simulation capability.

Although the three SOC models used in this analysis are not computationally unattainable for parameter calibration, what we have learned about SBO from this study can be potentially applied to more complex models. Currently, more and more complex simulation models present challenges to the SBO algorithm. To improve the accuracy of SBO, better surrogate models are expected. Current surrogate models including our implementation for soil carbon models mainly employ only one surrogate model, which may limit the successful use for different kinds of models. We will focus on the application of multiple surrogate models using ensemble learning in the future.

7 Code and data availability

The code and data of three models and the related algorithm implementations can be found in the supplement. If you have any problem when using the code and repeating the experiments, please feel free to contact the corresponding author of this paper: Wei Xue (xuewei@tsinghua.edu.cn).

References

- Aleman D M, Romeijn H E, Dempsey J F. A response surface approach to beam orientation optimization in intensity-modulated radiation therapy treatment planning[J]. *INFORMS Journal on Computing*, 2009, 21(1): 62-76.
- Huang YY, XJ Lu, Z Shi, D Lawrence, C Koven, JY Xia, ZG Du, E Kluzek, YQ Luo. Matrix approach to land carbon cycle modeling: A case study with Community Land Model[J]. *Global Change Biology*, 2017.
- Allison S D, Wallenstein M D, Bradford M A. Soil-carbon response to warming dependent on microbial physiology[J]. *Nature Geoscience*, 2010, 3(5): 336-340.

Booker A J, Dennis Jr J E, Frank P D, et al. A rigorous framework for optimization of expensive functions by surrogates[J]. *Structural optimization*, 1999, 17(1): 1-13.

Chen H, Tian H Q. Does a General Temperature-Dependent Q₁₀ Model of Soil Respiration Exist at Biome and Global Scale[J]. *Journal of Integrative Plant Biology*, 2005, 47(11): 1288-1302.

Corzo G, Solomatine D. Special Issue: Knowledge-based modularization and global optimization of artificial neural network models in hydrological forecasting[J]. *Neural Networks the Official Journal of the International Neural Network Society*, 2007, 20(4):528-536.

Duan Q, Sorooshian S, Gupta V K. Optimal use of the SCE-UA global optimization method for calibrating watershed models[J]. *Journal of hydrology*, 1994, 158(3): 265-284.

Duan Q, Sorooshian S, Gupta V. Effective and efficient global optimization for conceptual rainfall-runoff models[J]. *Water resources research*, 1992, 28(4): 1015-1031.

Fontaine S, Bardoux G, Abbadie L, et al. Carbon input to soil may decrease soil carbon content[J]. *Ecology Letters*, 2004, 7(4):314-320.

Fontaine S, Barot S, Barré P, et al. Stability of organic carbon in deep soil layers controlled by fresh carbon supply[J]. *Nature*, 2007, 450(7167):277-80.

Friedlingstein, P., et al., Climate-carbon cycle feedback analysis: Results from the C4MIP model intercomparison[J]. *Climate*, 2006, 19(14), 3337–3353.

Friedman J H. Multivariate adaptive regression splines[J]. *The annals of statistics*, 1991, 19(1): 1-67.

German D P, Marcelo K R B, Stone M M, et al. The Michaelis–Menten kinetics of soil extracellular enzymes in response to temperature: a cross-latitudinal study[J]. *Global Change Biology*, 2012, 18(4): 1468-1479.

Giunta A A. Aircraft multidisciplinary design optimization using design of experiments theory and response surface modeling methods[D]. Virginia polytechnic institute and state university, 1997.

Global Soil Data Task Group. Global Gridded Surfaces of Selected Soil Characteristics (IGBP-DIS). [Global Gridded Surfaces of Selected Soil Characteristics (International Geosphere-Biosphere Programme - Data and Information System)]. Data set, 2000, Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, TN.

Gutmann H M. A radial basis function method for global optimization[J]. *Journal of Global Optimization*, 2001, 19(3): 201-227.

-
- Haario H, Saksman E, Tamminen J. An adaptive Metropolis algorithm[J]. *Bernoulli*, 2001, 7(2): 223-242.
- Hansen N, Kern S. Evaluating the CMA evolution strategy on multimodal test functions[C]. *Proceedings of International Conference on Parallel Problem Solving from Nature*. Springer Berlin Heidelberg, 2004: 282-291.
- Hansen N, Ostermeier A. Completely de-randomized self-adaptation in evolution strategies[J]. *Evolutionary Computation*, 2001, 9(2): 159-195.
- Hansen N. Benchmarking the Nelder-Mead downhill simplex algorithm with many local restarts[C]. *Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers*. ACM, 2009: 2403-2408.
- Hapuarachchi H A P, Li Z, Wang S. Application of SCE-UA method for calibrating the Xinanjiang watershed model[J]. *Journal of lake sciences*, 2001, 13(4): 304-314.
- Hararuk O, Smith M J, Luo Y. Microbial models with data-driven parameters predict stronger soil carbon responses to climate change[J]. *Global change biology*, 2015, 21(6): 2439-2453.
- Hararuk O, Xia J, Luo Y. Evaluation and improvement of a global land model against soil carbon data using a Bayesian Markov chain Monte Carlo method[J]. *Journal of Geophysical Research: Biogeosciences*, 2014, 119(3): 403-417.
- Shi Z., S Crowell, YQ Luo, B Moore III. 2018. Uncertainty in soil carbon projection constrained by data but amplified by model structures[J]. *Nature Communications*.
- Houghton, J. T., Y. Ding, D. J. Griggs, M. Noguera, P. J. van der Linden, X. Dai, K. Maskell, and C. Johnson. *Climate Change 2001: the scientific basis*[B]. Cambridge University Press, Cambridge.
- Iman R L, Campbell J, Helton J. An approach to sensitivity analysis of computer models[J]. *Journal of Quality Technology*, 1981, 13.
- Ise, T., and P. R. Moorcroft. The global-scale temperature and moisture dependencies of soil organic carbon decomposition: An analysis using a mechanistic decomposition model[J]. *Biogeochemistry*, 2006, 80(3), 217-231.
- Jones D R, Schonlau M, Welch W J. Efficient global optimization of expensive black-box functions[J]. *Journal of Global optimization*, 1998, 13(4): 455-492.
- Jones, D.R. A taxonomy of global optimization methods based on response surfaces[J]. *Journal of Global Optimization*, 2001,21(4): 345-383.

-
- Kennedy J. Particle swarm optimization[M]. Encyclopedia of machine learning. Springer US, 2011: 760-766.
- Kim K J. Toward Global Optimization of Case-Based Reasoning Systems for Financial Forecasting[J]. Applied Intelligence, 2004, 21(3):239-249.
- Kuzyakov Y, Friedel J K, Stahr K. Review of mechanisms and quantification of priming effects.[J]. Soil Biology & Biochemistry, 2000, 32(11–12):1485-1498.
- Li G, Cheng C, Lin J, et al. Short-term load forecasting using support vector machine with SCE-UA algorithm[C]. Third International Conference on Natural Computation (ICNC 2007). IEEE, 2007, 1: 290-294.
- Luo Y, A Ahlström, SD Allison, NH Batjes, V Brovkin, N Carvalhais, A Chappell, P Ciais, EA Davidson, A Finzi, K Georgiou, B Guenet, O Hararuk, JW Harden, YJ He, F Hopkins, LF Jiang, C Koven, RB Jackson, CD Jones, MJ Lara, JY Liang, AD McGuire, W Parton, CH Peng, JT Randerson, A Salazar, CA Sierra, MJ Smith, HQ Tian, KEO Todd-Brown, M Torn, KJ van Groenigen, YP Wang, TO West, YX Wei, WR Wieder, JY Xia, X Xu, XF Xu, T Zhou.. Towards More Realistic Projections of Soil Carbon Dynamics by Earth System Models[J]. Global Biogeochemical Cycles, 2016, 30(1): 40-56.
- Luo Y. Terrestrial carbon-cycle feedback to climate warming[J]. Annual Review of Ecology, Evolution, and Systematics, 2007: 683-712.
- Luo YQ, TF Keenan, M Smith, Predictability of the terrestrial carbon cycle[J]. Global Change Biology 2015, 1737-1751.
- Luo, Y., L. W. White, J. G. Canadell, E. H. DeLucia, D. S. Ellsworth, A. Finzi, J. Lichter, and W. H. Schlesinger, Sustainability of terrestrial carbon sequestration: A case study in Duke Forest with inversion approach, [J] Global Biogeochem. Cycles, 2003, 17(1), 1021.
- Luo, Y., L. Wu, J. A. Andrews, L. White, R. Matamala, K. V. R. Schäfer, and W. H. Schlesinger, Elevated CO₂ differentiates ecosystem carbon processes: Deconvolution analysis of Duke Forest FACE data, Ecol[J]. Monogr., 2001, 71(3), 357–376.
- Ma H, Dong Z, Zhang W, et al. Application of SCE-UA algorithm to optimization of TOPMODEL parameters [J]. Journal of Hohai University (Natural Sciences edition), 2006, 4:001.
- Madescu G, Boldea I, Miller T J E. The optimal lamination approach (OLA) to induction machine design global optimization[C]// Ias Meeting, Ias '96. Conference Record of the. IEEE, 1996:574-580 vol.1.

Maranas C D, Androulakis I P, Floudas C A, et al. Solving long-term financial planning problems via global optimization[J]. *Journal of Economic Dynamics & Control*, 1997, 21(8-9):1405-1425.

Marshall, L., D. Nott, and A. Sharma (2004), A comparative study of Markov chain Monte Carlo methods for conceptual rainfall-runoff modeling, *Water Resour. Res.*, 40, W02501, doi:10.1029/2003WR002378.

Mckay M D, Beckman R J, Conover W J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code[M]. American Society for Quality Control and American Statistical Association, 2000.

Melillo, J. M., P. A. Steudler, J. D. Aber, K. Newkirk, H. Lux, F. P. Bowles, C. Catricala, A. Magill, T. Ahrens, and S. Morrisseau, Soil warming and carbon-cycle feedbacks to the climate system [J], *Science*, 2002, 298(5601), 2173–2176.

Müller J. MATSuMoTo: The MATLAB Surrogate Model Toolbox For Computationally Expensive Black-Box Global Optimization Problems . arXiv:1404.4261, 2014.

Müller J, Shoemaker C A. Influence of ensemble surrogate models and sampling strategy on the solution quality of algorithms for computationally expensive black-box global optimization problems[J]. *Journal of Global Optimization*, 2014, 60(2): 123-144.

Myers R H, Montgomery D C, Anderson-Cook C M. Response surface methodology: process and product optimization using designed experiments[M]. John Wiley & Sons, 2016.

Oleson, K. W., G. Niu, Z. Yang, D. Lawrence, P. Thornton, P. Lawrence, R. Stockli, R. Dickinson, G. Bonan, and S. Levis, Improvements to the Community Land Model and their impact on the hydrological cycle [J], *J. Geophys. Res.*, 2008, 113.

Oleson, K. W., Y. Dai, G. Bonan, M. Bosilovich, R. Dickinson, P. Dirmeyer, F. Hoffman, P. Houser, S. Levis, and G.-Y. Niu Technical description of the community land model (CLM) [J]. NCAR Tech. 2004. .

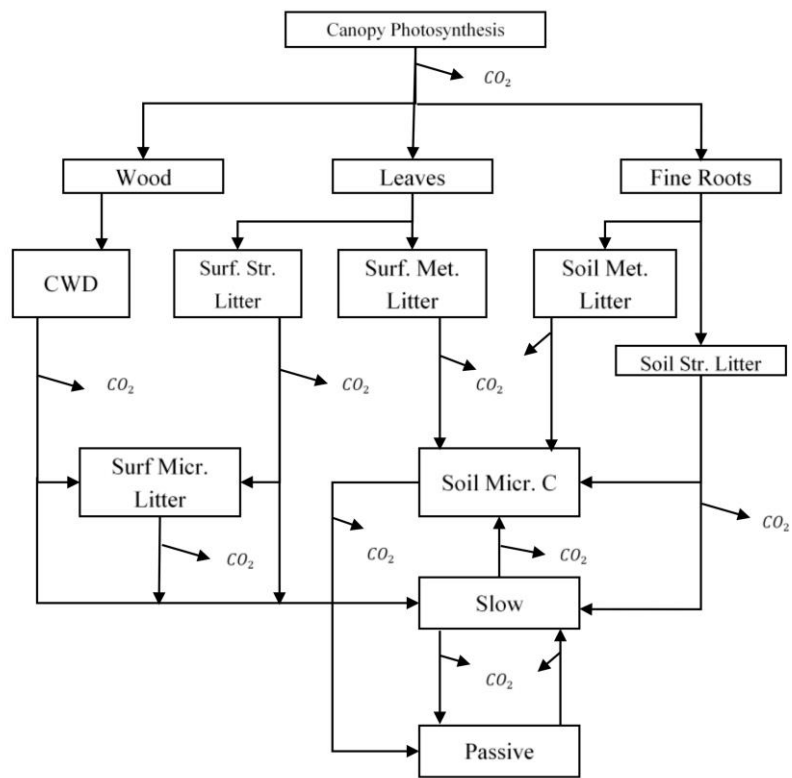
Parton, W. J., et al., Observations and modelling of biomass and soil organic matter dynamics for the grassland biome worldwide [J], *Global Biogeochem. Cycles*, 1993, 7(4), 785–809.

Peng S, Piao S, Wang T, et al. Temperature sensitivity of soil respiration in different ecosystems in China[J]. *Soil Biology and Biochemistry*, 2009, 41(5): 1008-1014.

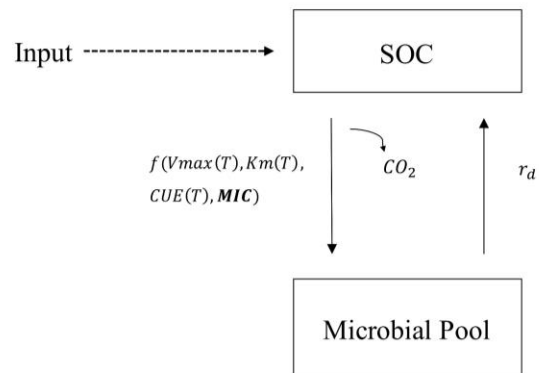
Picheny V, Ginsbourger D, Richet Y, et al. Quantile-Based Optimization of Noisy Computer Experiments With Tunable Precision[J]. *Technometrics*, 2012, 55(1): 2-9.

-
- Price K, Storn R M, Lampinen J A. Differential evolution: a practical approach to global optimization[M]. Springer Science & Business Media, 2006.
- Regis R G, Shoemaker C A. A stochastic radial basis function method for the global optimization of expensive functions[J]. *INFORMS Journal on Computing*, 2007, 19(4): 497-509.
- Regis R G, Shoemaker C A. Parallel stochastic global optimization using radial basis functions[J]. *INFORMS Journal on Computing*, 2009, 21(3): 411-426.
- Regis R G. Stochastic radial basis function algorithms for large-scale optimization involving expensive black-box objective and constraint functions[J]. *Computers & Operations Research*, 2011, 38(5): 837-853.
- Schimel J P, Weintraub M N. The implications of exoenzyme activity on microbial carbon and nitrogen limitation in soil: a theoretical model[J]. *Soil Biology and Biochemistry*, 2003, 35(4): 549-563.
- Schonlau M, Welch W J, Jones D R. Global versus local search in constrained optimization of computer models[J]. *Lecture Notes-Monograph Series*, 1998: 11-25.
- Shi Y, Eberhart R C. Empirical study of particle swarm optimization[J]. *Frontiers of Computer Science in China*, 2009, 3(1):31-37.
- Simpson T W, Mauery T M, Korte J J, et al. Kriging models for global approximation in simulation-based multidisciplinary design optimization[J]. *AIAA journal*, 2001, 39(12): 2233-2241.
- Smith, M. J., D. W. Purves, M. C. Vanderwel, V. Lyutsarev, and S. Emmott, The climate dependence of the terrestrial carbon cycle, including parameter and structural uncertainties [J], *Biogeosciences*, 2013, 10(1), 583–606.
- Sorooshian S, Duan Q, Gupta V K. Calibration of rainfall-runoff models: application of global optimization to the Sacramento soil moisture accounting model[J]. *Water resources research*, 1993, 29(4): 1185-1194.
- Storn R, Price K. Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces[J]. *Journal of global optimization*, 1997, 11(4): 341-359.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, An overview of CMIP5 and the experiment design [J], *Bull. Am. Meteorol. Soc.*, 2011, 93(4), 485–498.
- Todd-Brown, K. E. O., J. T. Randerson, W. M. Post, F. M. Hoffman, C. Tarnocai, E. A. G. Schuur, and S. D. Allison, Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations, *Biogeosciences*, 2013, 10(3), 1717–1736.

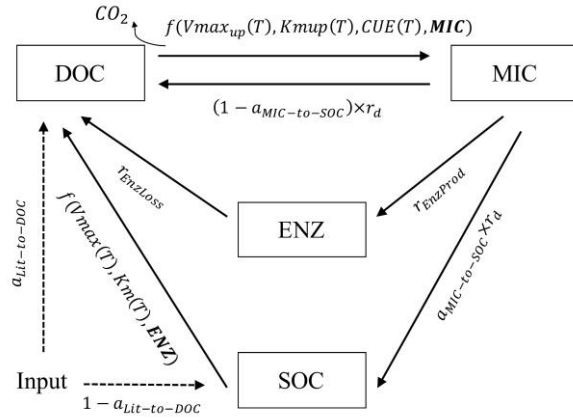
-
- Vu K K, D'Ambrosio C, Hamadi Y, et al. Surrogate-based methods for black-box optimization[J]. *International Transactions in Operational Research*, 2016.
- Wang C, Duan Q, Gong W, et al. An evaluation of adaptive surrogate modelling based optimization with two benchmark problems[J]. *Environmental Modelling & Software*, 2014, 60(76):167-179.
- Weng, E., and Y. Luo, Relative information contributions of model vs. data to short- and long-term forecasts of forest carbon dynamics, *Ecol. Appl.*, 2011, 21(5), 1490–1505.
- Wieder W R, Bonan G B, Allison S D. Global soil carbon projections are improved by modelling microbial processes[J]. *Nature Climate Change*, 2013, 3(10): 909-912.
- Xia, J., Y. Luo, Y. P. Wang, and O. Hararuk, Traceable components of terrestrial carbon storage capacity in biogeochemical models [J], *Global Change Biol.*, 2013, 19, 2104–2116.
- Xia, J., Y. Luo, Y. Wang, E. Weng, and O. Hararuk, A semi-analytical solution to accelerate spin-up of a coupled carbon and nitrogen land model to steady state [J], *Geosci. Model Dev.*, 2012, 5(5), 1259–1271.
- Xu, T., L. White, D. Hui, and Y. Luo, Probabilistic inversion of a terrestrial ecosystem model: Analysis of uncertainty in parameter estimation and model prediction [J], *Global Biogeochem. Cycles*, 2006, 20, GB2007.
- Zhou, T., P. Shi, D. Hui, and Y. Luo, Global pattern of temperature sensitivity of soil heterotrophic respiration (Q₁₀) and its implications for carbon-climate feedback [J], *J. Geophys. Res.*, 2009, 114.
- Friedman J H. Multivariate Adaptive Regression Splines[J]. *Annals of Statistics*, 1991, 19(1): 1-67.



(a) The CLM-CASA' model



(b) Two-pool microbial model



(c) Four-pool microbial model

Figure 1. Schematic representations of (a) CLM-CASA' model, (b) two-pool microbial model and (c) four-pool microbial models

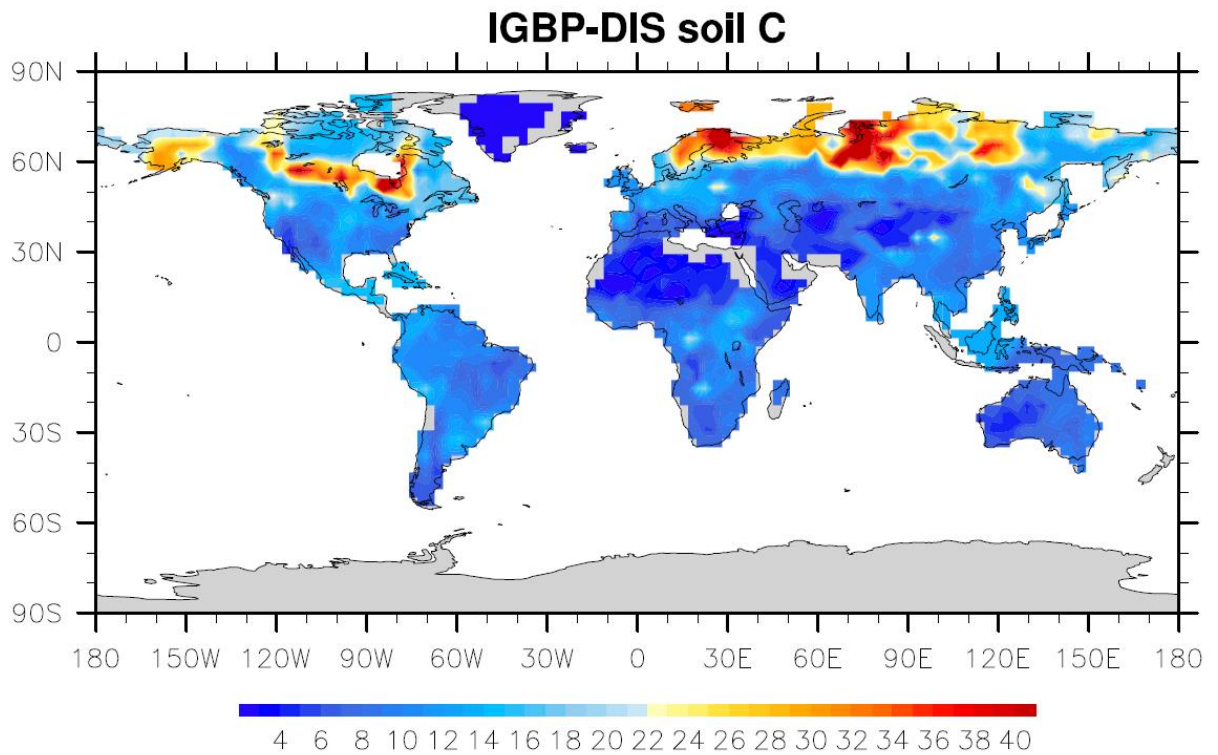


Figure 2. IGBP-DIS soil carbon distribution. Soil carbon varies from 0 kg/m^2 in deserts to 60 kg/m^2 in boreal regions

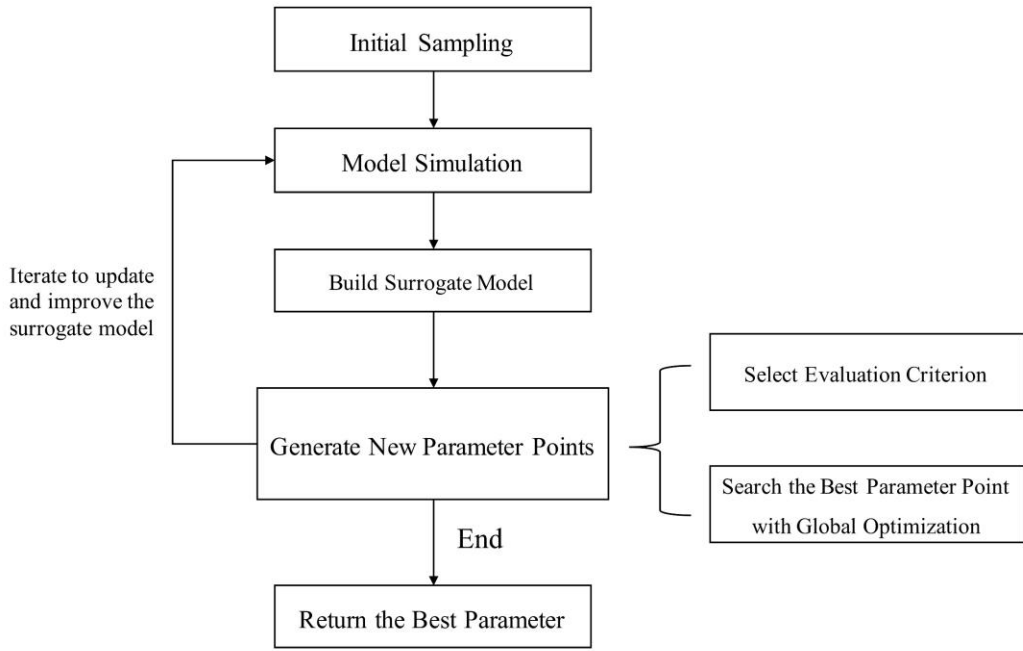
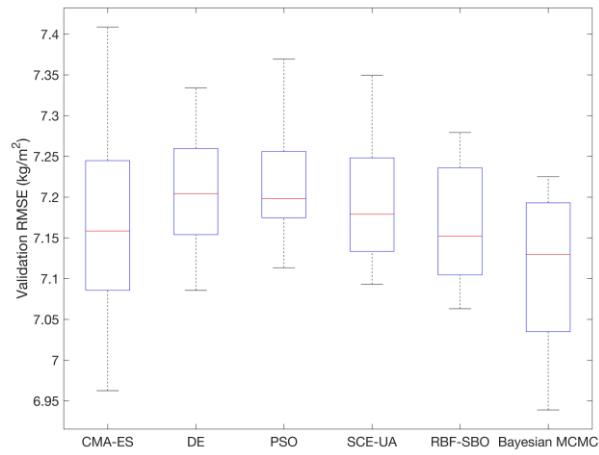
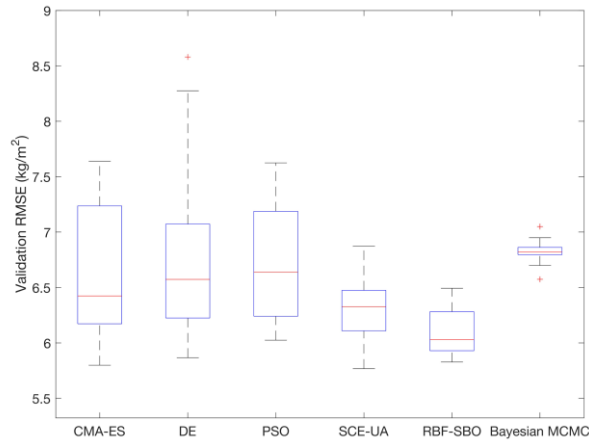


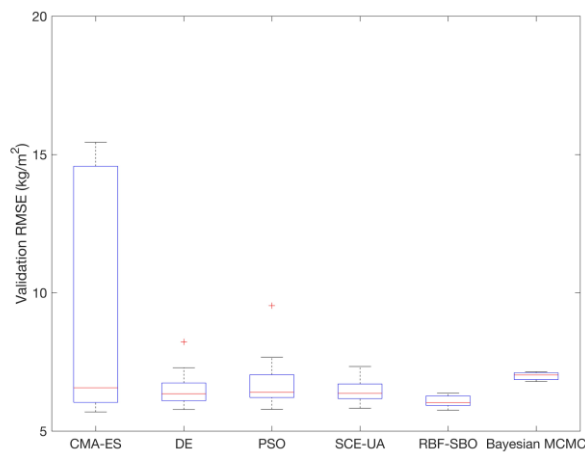
Figure 3. The flowchart of the surrogate-based optimization



(a) CLM-CASA' model

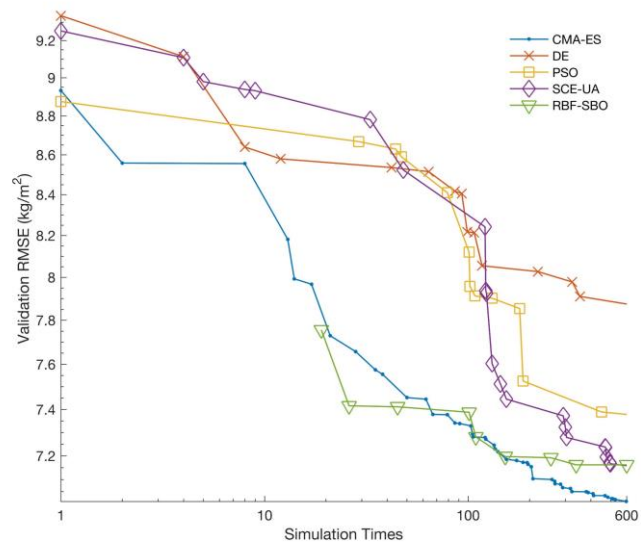


(b) Two-pool microbial model

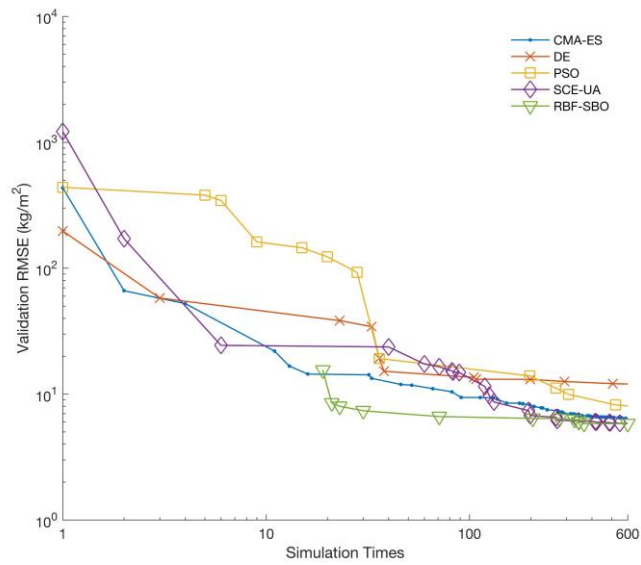


(c) Four-pool microbial model

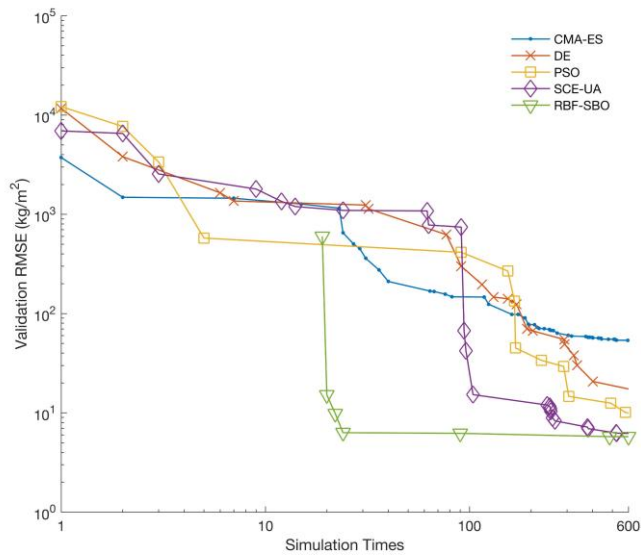
Figure 4. The RMSEs of different optimization algorithms: (a) CLM-CASA' model; (b) two-pool microbial model and (c) four-pool microbial model. The box plots show the means and the quartiles spreading over total 50 calibration runs. The central line indicates the median; the bottom and top of the box are the first and third quartiles; the black bottom and top lines out of the rectangles are the maximum and minimum; the red crosses represent the outliers. The simulation times of former 5 algorithms are 100 and the simulation times of Bayesian MCMC are presented in Table 3.



(a) The RMSEs for CLM-CASA' model

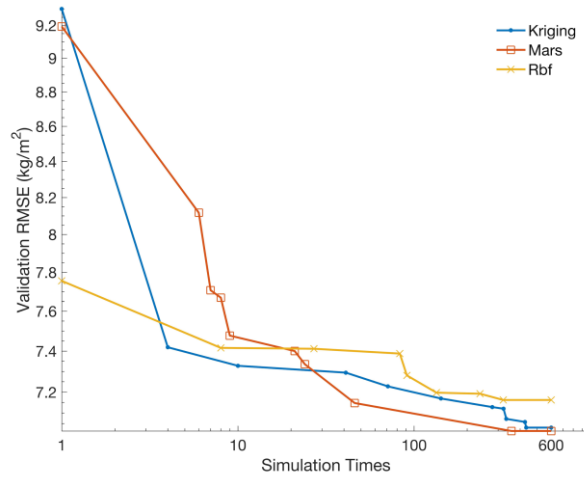


(b) The RMSEs for two-pool microbial model

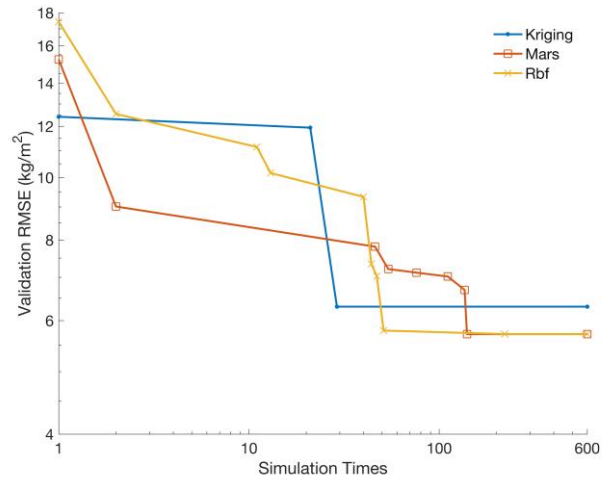


(c) The RMSEs for four-pool microbial model

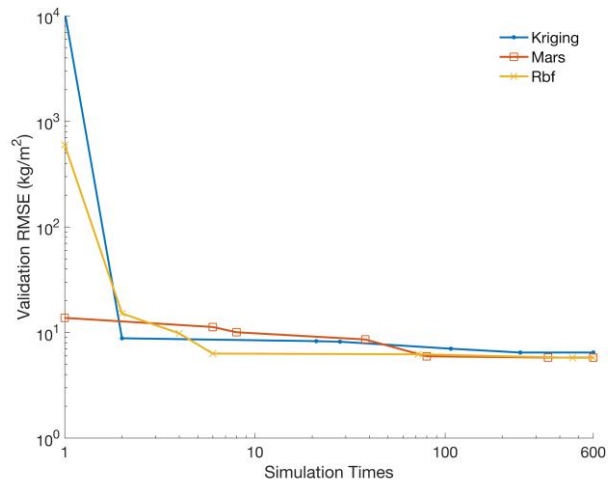
Figure 5. The average RMSEs with the increase of simulation times and different optimization algorithms: (a) the CLM-CASA' model; (b) two-pool microbial model and (c) four-pool microbial model. Since RBF SBO requires some initial simulations to start optimization process, RBF SBO starts when x-axis value is 19.



(a) The RMSEs for CLM-CASA' model

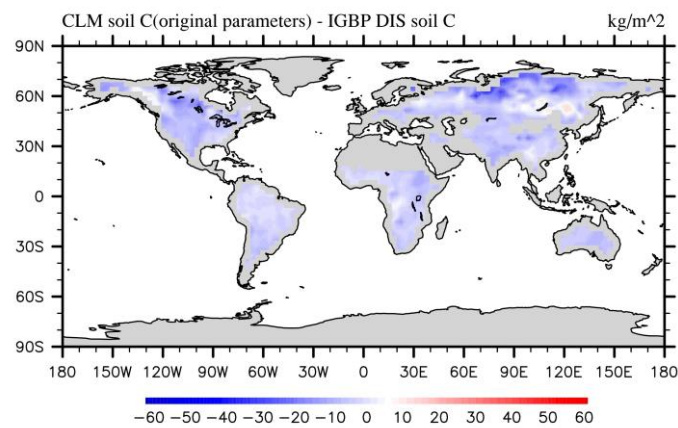


(b) The RMSEs for two-pool microbial model

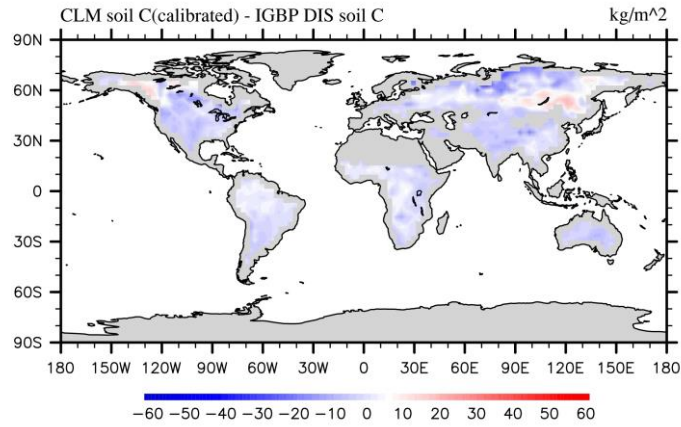


(c) The RMSEs for four-pool microbial model

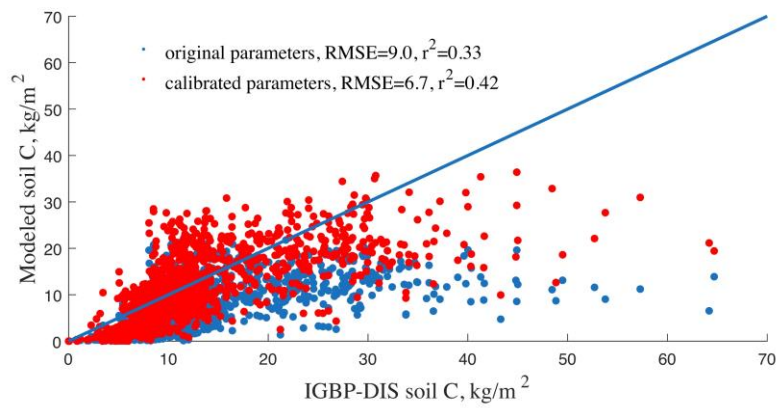
Figure 6. The average RMSEs with the increase of simulation times and different surrogate models.



(a) CLM soil C – IGBP-DIS soil C



(b) CLM soil C (calibrated) – IGBP-DIS soil C



(c) Spatial correspondence between modelled soil and IGBP-DIS soil

Figure 7. Spatial correspondence of SOC produced by CLM-CASA' to SOC reported by IGBP-DIS. The subgraph (a) shows the results using the default parameter values and the subgraph (b) shows the results after parameter calibration using the surrogate-based optimization. The points in Fig. 7c represent the grid cell values (blue ones for the results with default parameter values and red ones for the results after parameter calibration). CLM-CASA' with the default parameter values explains 33% of variation in the observed soil C, while CLM-CASA' with the calibrated parameter values explains 42% of variability in the observed soil C.

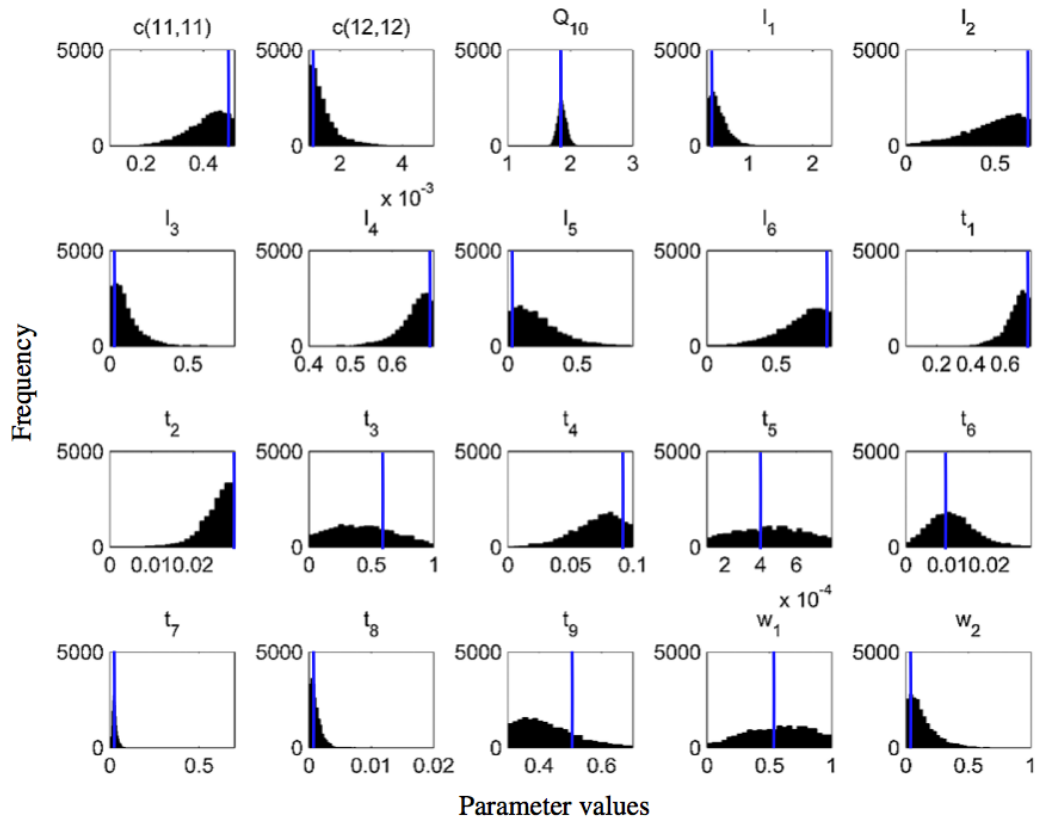


Figure 8. Frequency distributions of 20 calibrated parameters of CLM-CASA' model by Bayesian MCMC approach (Harauk, 2014) and surrogate-based optimization (blue line in each subgraph).

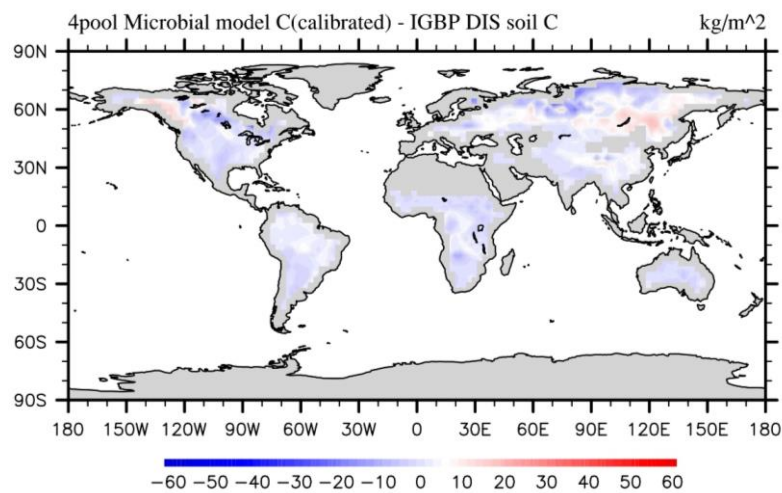


Figure 9. Spatial correspondence of four-pool microbial model produced SOC to the IGBP-DIS reported SOC.

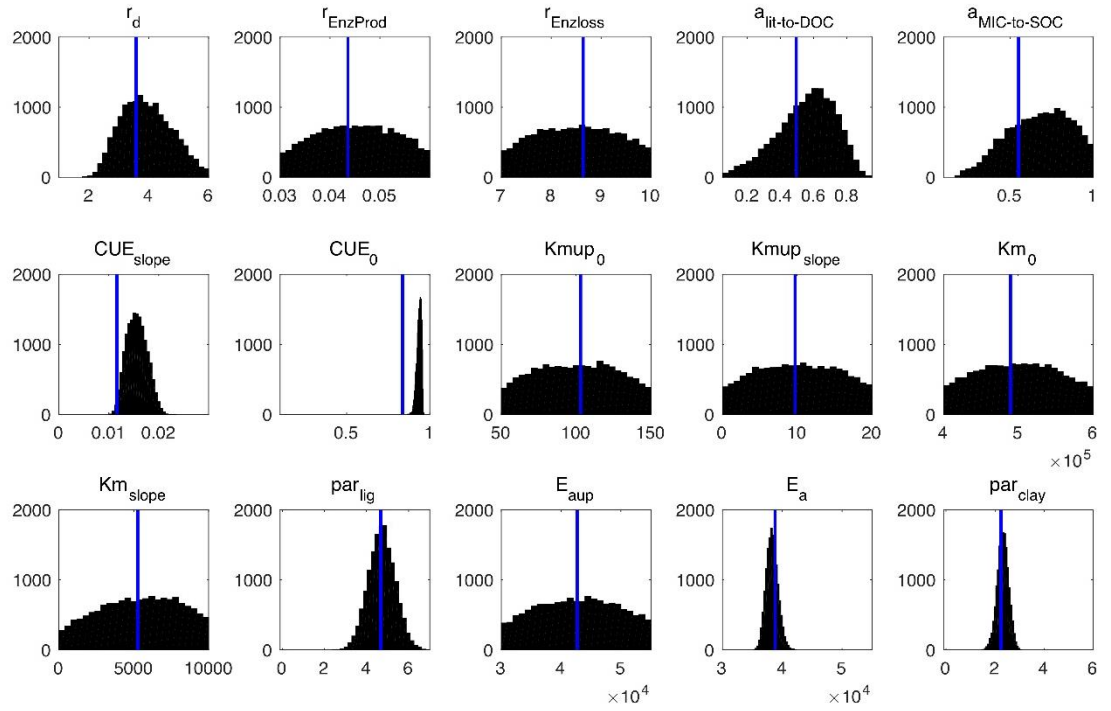


Figure 10. Posterior probability density functions of the four-pool microbial model parameters (generated by Bayesian MCMC). The blue vertical lines are the final calibrated parameter values by our surrogate-based optimization.

Table 1. Parameter description of CLM-CASA' C-only model

Parameter Description	Symbol	Default Value (x0.001)	Calibrated Value by SBO (x0.001)
Decomposition rate from slow pool	$c(11,11)$	200	495.6
Decomposition rate from passive pool	$c(12,12)$	4.5	1.01
Temperature sensitivity of C decomposition	Q_{10}	2000	1737
Labile C fraction effect on C partitioning from leaves to surface metabolic litter	w_1	1000	589.04
Labile C fraction effect on C partitioning from roots to soil metabolic litter	w_2	200	4.52
Partitioning from surface structural to surface microbial pool if no lignin in surface structural litter	l_1	400	384.5
Lignin effect of partitioning from surface structural litter to surface microbial litter	l_2	400	689
Lignin effect on partitioning from surface structural litter to soil slow pool	l_3	700	7.499
Partitioning from soil structural to soil microbial pool if no lignin in soil structural litter	l_4	450	697.7
Lignin effect on partitioning from soil structural litter to soil microbial pool	l_5	450	54.46
Lignin effect on partitioning from soil structural litter to soil slow pool	l_6	700	871.5
C partitioning from soil microbial pool to slow pool if no sand or clay	t_1	169	747.7
Clay effect on C partitioning from soil microbial pool	t_2	5.44	29.6
Sand effect on C partitioning from soil microbial to slow pool	t_3	678	636.8

Combined effect of sand and clay on C partitioning from soil microbial pool	t_4	22	99.5
C partitioning from soil microbial to passive pool if no sand or clay	t_5	0.51	0.152
Sand effect on C partitioning from soil microbial to passive pool	t_6	2.04	12.99
Clay effect on C partitioning from slow pool to passive pool	t_7	4.05	24.2
C partitioning from slow to passive pool if no clay	t_8	14	0.012
C partitioning from slow to soil microbial pool if no clay	t_9	449	368.8

Table 2. Parameter and description of the four-pool microbial models

Parameter Name	Parameter Description	Default Value	Calibrated Value by SBO
r_d	Microbial death rate	4.38	4.89
CUE_0	Baseline microbial carbon use efficiency	0.63	0.965
CUE_{slope}	CUE_0 dependency on temperature	0.016	0.00853
Km_0	Baseline half saturation constant	500000	498467
Km_{slope}	Km_0 dependency on temperature	5000	9751
E_a	Activation energy of SOC decomposition	47000	36669
par_{clay}	Clay limitation	0	2.41
par_{lig}	Lignin-dependent correction factor	0	6.23
$r_{EnzProd}$	Rate of enzyme production	0.0438	0.0361
$r_{EnzLoss}$	Rate of enzyme loss	8.76	8.08
$a_{lit-to-DOC}$	Fraction of $Input_{soil}$ that is transferred to soil	0.3	0.832
$a_{MIC-to-SOC}$	Fraction of dead microbes transferred to soil	0.5	0.716
$Kmup_0$	Baseline half-saturation constants for substrate limitation of DOC uptake	100	134

$Kmup_{slope}$	$Kmup_0$ dependency on temperature	10	4.62
E_{aup}	Activation energy of DOC uptake	47000	34811

Table 3 Calibration results of Bayesian MCMC and our surrogate-based optimization

SOC model	Detail	Method	Lowest RMSE <i>(kg · m⁻²)</i>	Variance Explained	Number of Simulations
Two-pool microbial	8 parameters	Bayesian MCMC	6.609	51.6%	50,000+500,000
	2 carbon pools	RBF-SBO	5.785	51.6%	221
Four-pool microbial	15 parameters	Bayesian MCMC	7.142	51.3%	50,000+500,000
	4 carbon pools	RBF-SBO	5.756	51.4%	199
CLM-CASA'	20 parameters	Bayesian MCMC	7.000	41.0%	50,000+1,000,000
	13 carbon pools	RBF-SBO	7.162	42.8%	321