General Comment:

The paper proposed by Xu et al investigates several issues related to sensitivity analysis (SA) and optimization using the Land Surface Model (CLM). The final objective is to study the effect of using a single or several variables during the optimization process on parameter estimation and the overall performance of CLM applied on an ICOS site located in Russia. Before performing optimization, sensitivity analysis is performed using 4 approaches to identify the parameters that mostly impact the simulated variables. Optimization is then performed in a single or multi-objective mode using the PEM-SMC algorithm that was specifically adapted to reduce the computation burden and make such an optimization possible.

Response:

We appreciate the reviewer's positive and constructive feedback, which will assist us in further improving our work. Below, we outline our planned responses to the issues raised by the reviewer and the specific changes we intend to implement in the revision.

Major Comments:

Comment #1: As mentioned above, the paper is not self-consistent as no information on CLM- equations and parameterization-are provided. In my opinion, the paper should be reshaped to include a part dedicated to the presentation of CLM. Furthermore, the name used in the paper-CoLM-should be changed throughout the paper and turned into CLM to avoid confusion.

Response:

Thank you for your valuable suggestions. We recognize the importance of including equations and parameterization details for latent heat flux (LE) and net ecosystem exchange (NEE) in the Common Land Model to clarify the physical basis of the selected sensitive parameters and the mechanisms underlying their optimized values. Accordingly, We will add descriptions of the LE and NEE equations in Section 2.1 of the revised manuscript. Regarding the abbreviation, while "CLM" was initially used by Dai et al. (2003), subsequent studies have adopted "CoLM" to avoid confusion with the Community Land Model, which is also widely referred to as "CLM." Therefore, we have retained the abbreviation "CoLM" in our manuscript. However, if you believe that

"CLM" should be used, we are open to making this change in the next revision.

Reference:

Dai, Y., Zeng, X., Dickinson, R. E., Baker, I., Bonan, G. B., Bosilovich, M. G., ... & Yang, Z. L. (2003). The common land model. Bulletin of the American Meteorological Society, 84(8), 1013-1024.

Comment #2: I don't get why 3 qualitative sensitivity analysis approaches are used prior to the Sobol's analysis. From Figure 1, it seems that MOAT alone could be sufficient to identify the most sensitive parameters to be used in the following. In my opinion, the need of multiple qualitative approaches, their potential complementarity and what kind of different information they can bring in should be detailed and explained more clearly. The use of 3 methods rather than one makes it more difficult for the reader to understand the overall method. If the use of the 3 approaches is relevant, the description of each approach should be improved to better explain its own interest for sensitivity analysis.

Response:

Qualitative sensitivity analysis methods typically evaluate the importance of input parameters on model output by comparing outputs under different combinations of inputs with relatively few samples. However, complex models often involve characteristics like nonlinearity and parameter interdependence, and different methods may emphasize various aspects of these features. Relying on a single method could overlook or misjudge certain parameters. For instance, in the Opt_LE scenario (Figure 1), MOAT did not identify sensitive parameter P36, while MARS did; both MOAT and MARS identified P8, but DT did not. Combining multiple methods can enhance the robustness of the analysis, reduce errors, and yield more reliable results. Given the previous limitations in our description, we will provide a more detailed explanation of the strengths and differences of these three methods, emphasizing their complementarity, in Section 2.2 of the revised manuscript:

(1) Delta test (DT) method: While it estimates complex nonlinear relationships, it may not identify the most sensitive parameters. Combining DT with MARS and MOAT helps overcome this limitation.

(2) Multivariate adaptive regression splines (MARS) method: Effective in handling nonlinear relationships and interactions, but it can sometimes create overly complex models. Integrating MARS with DT and MOAT helps to clarify the impact of each parameter.

(3) Morris method: Provides a global perspective, offering a comprehensive evaluation of all parameters and reducing biases inherent in local sensitivity analyses.

Comment #3: The size of the different samples seem to be set arbitrarily. Maybe justifications – that are not only related to the computation burden – should be given as it can impact the performance of the sensitivity analysis.

Response:

In this study, the sample size for all qualitative analysis methods was determined based on the findings of Li et al. (2013), who conducted a sensitivity analysis on 40 parameters of the CoLM model. For the DT and MARS methods, sample sizes of 200, 400, and 1000 (i.e., 5, 10, and 25 times the number of parameters, respectively) were evaluated. For the MOAT method, samples were typically set as multiples of n+1, where n is the number of parameters; hence, sample sizes of 205, 410, and 1025 were examined. Their results indicated that a sample size of 400 (10 times the number of parameters) was sufficient for screening the 40 parameters of the CoLM model. Therefore, this study utilizes 400 samples for DT and MARS methods and 410 for MOAT. In the revised manuscript, we will include additional tests and comparisons using different sample sizes (5, 10, and 25 times the number of parameters) to ensure the robustness of the sensitivity analysis results. For the quantitative sensitivity analysis (Sobol'), the sample size was set to 100,000, consistent with the typical range reported in previous studies (10^4 to 10^5). For instance, Rosolem et al. (2012) used 45,000 model runs to evaluate the Sobol' sensitivity indices of 42 parameters in the Simple Biosphere 3 (SiB3) model, while Zhang et al. (2013) employed 60,000 model runs to study the sensitivities of 28 parameters in the Soil and Water Assessment Tool (SWAT) model using the Sobol' method.

Reference:

Li, J., Duan, Q. Y., Gong, W., Ye, A., Dai, Y., Miao, C., ... & Sun, Y. (2013). Assessing parameter importance of the Common Land Model based on qualitative and quantitative sensitivity analysis. *Hydrology and Earth System Sciences*, *17*(8), 3279-3293.

Rosolem, R., Gupta, H. V., Shuttleworth, W. J., Zeng, X., & de Gonçalves, L. G. G. (2012). A fully multiple-criteria implementation of the Sobol′ method for parameter sensitivity analysis. *Journal of Geophysical Research: Atmospheres*, *117*(D7).

Zhang, C., Chu, J., & Fu, G. (2013). Sobol′'s sensitivity analysis for a distributed hydrological model of Yichun River Basin, China. *Journal of Hydrology*, *480*, 58-68.

Comment #4: The description of the overall approach presented from L189 to L212 should be improved. As it stands in this version, sensitivity analysis and optimization are mixed together which is rather hard to catch. I think a scheme is highly needed here. And I also think that the authors should more clearly stands that the target variables are NRMSEs computed with LEE/NEE/both.

Response:

In response to your suggestion, we will include a technical flowchart in the revised manuscript to more clearly illustrate the methods, processes, and metrics involved in the parameter sensitivity analysis. Additionally, we have revised the cost function expressions for both single-objective and multi-objective sensitivity analysis as follows:

Given the varying magnitudes of the target variables (LE/NEE), we employed the normalized root-mean-square error (NRMSE) as the cost function, defined as:

$$NRMSE_i = \frac{\sqrt{\sum_{t=1}^{T} S_i(t) - O_i(t))^2}}{\sum_{t=1}^{T} O_i(t)} \qquad (11)$$

where i *represents the target variable (eg., LE or NEE),* T *is the total number of simulations; and* $S_i(t)$ *and* $O_i(t)$ *represent the simulated and observed values of the target variables, respectively. For single-objective sensitivity analysis, the cost function is expressed as:*

$$F_i = NRMSE_i \qquad (12)$$

where F_i *denotes the error evaluation of the target variable* i *(e.g., LE or NEE). For multi-objective (LE+NEE) sensitivity analysis, the combined objective function can be expressed using a weighted sum of the individual objective functions. The simplest form is:*

$$F_{LE+NEE} = w_{LE} \cdot F_{LE} + w_{NEE} \cdot F_{NEE} \qquad (13)$$

where w_{LE} and w_{NEE} are weights, typically set to 1 to indicate equal weighting.

Comment #5: There are also some discrepancies between what is presented L189 to 212 and what is presented afterwards. It is stated L205 that 10 parameters are selected for SA when less parameter are kept in the application example. It is said that the optimization is guided by Sobol's analysis. Does that mean that some parameters are removed after Sobol's indices are computed?

Response:

We apologize for any confusion. Before addressing your question, I will briefly explain the differences between qualitative and quantitative sensitivity analysis. Qualitative analysis is typically used for initial parameter screening, providing a rough ranking of parameter influence. In contrast, quantitative analysis offers precise quantification of each parameter's contribution to the model output and its interactions with other parameters, though it requires higher computational costs, especially in high-dimensional parameter spaces. Direct application of quantitative sensitivity analysis (e.g., the Sobol' method) to complex models can lead to inefficient use of computational resources and unnecessary complexity. In practice, the number of parameters analyzed using Sobol' methods is usually limited to 10.

Our sensitivity analysis is therefore conducted in two stages. First, qualitative analysis (using DT/MARS/MOAT methods) is performed with a smaller sample size to identify the 10 most sensitive parameters from an initial set of 40 (see Fig.1), thereby making the subsequent quantitative analysis more targeted and effective. In the second stage, Sobol' quantitative analysis is applied to these 10 parameters. Parameters for optimization are selected based on the criterion that their cumulative relative importance exceeds 95%, indicating that these parameters account for 95% of the explained variance (see Fig.2). As a result, some parameters identified in the qualitative analysis are excluded after the Sobol' analysis.

Comment #6: The technique aspects of part 2.3 are very hard to follow. Once again, I feel like a scheme could help understanding what is proposed and done.

Response:

Thank you for your suggestion. Previously, I had attempted to visually represent the process of the PEM-SMC algorithm, as shown in the figure below. In the revised manuscript, I will include a schematic diagram of the improved version of this algorithm to enhance clarity.

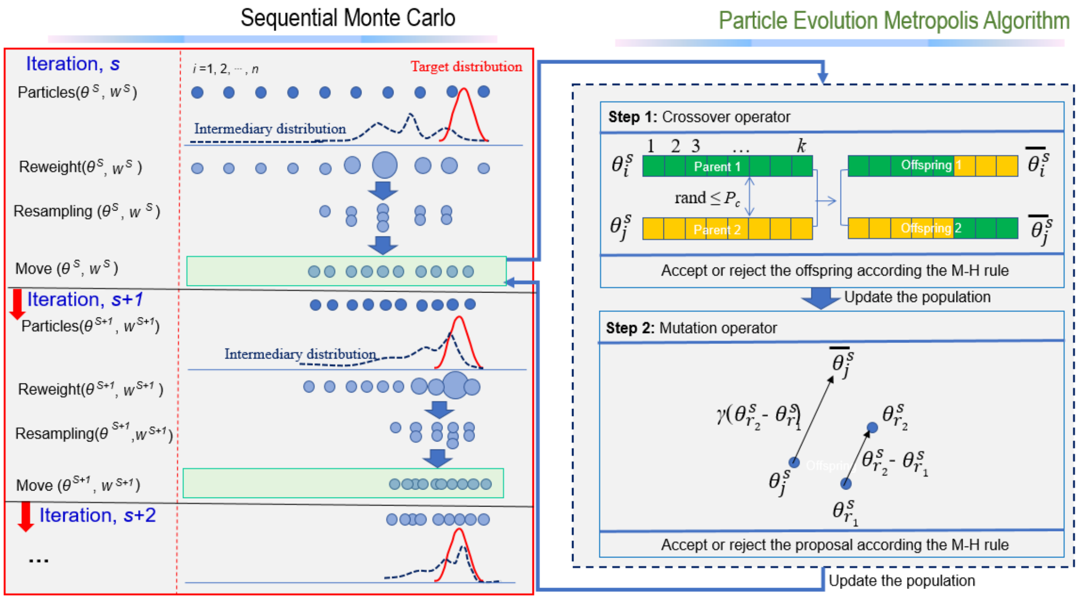

Figure R1 The schematic diagram of the PEM-SMC algorithm process.

Comment #7: It's not clear how many particles/set of parameters are kept during the optimization process. I think this should be clearly specified somewhere. The way the values for non-sensitive parameters are set should also be clearly explained.

Response:

In the PEM-SMC optimization algorithm, the initial number of particles (e.g., N=200) remains fixed throughout the iterative process, while the parameter values represented by each particle are continually updated. In Section 2.3 of the revised manuscript, we will provide a more detailed explanation of the role of particles in PEM-SMC: each particle corresponds to a specific point within a multidimensional parameter space, and through the evolution of these particles and their associated weights, the algorithm progressively converges towards the true posterior distribution of the parameters. For the optimization of each target variable, multiple sensitive parameters are considered, each with a specific range of possible values. Each particle represents a unique combination of these parameter values.

Furthermore, the non-sensitive parameters were assigned based on the model's default settings, which take into account factors such as vegetation type, soil characteristics, or data derived from empirical studies and literature. In the revised manuscript, we will include Table 1 in the Supporting Information, where the default values for the 40 predefined parameters specific to this site will be provided. Additionally, in Section 2.3 of the revised manuscript, we will elaborate on how the non-sensitive parameters were determined using the model's default settings.

Comment #8: After the SA results are presented, I think the physical meaning of the sensitive parameters should be explained. In my opinion, SA brings insights on how a model works. This aspect is rather poorly developed in the paper. This could greatly help for the analysis of the results, especially to understand the different values obtained after single/multiple optimization.

Response:

Thank you for your valuable suggestion. We acknowledge that the sensitivity analysis results are closely tied to the model's representation of relevant processes and its parameterization scheme. For instance, the two most sensitive parameters for calculating LE and NEE, P33 and P34, correspond to the quantum efficiency and the maximum carboxylation rate of vegetation leaves at 25 ℃ in the leaf stomatal photosynthesis-conductance module. These parameters directly affect the calculation of net photosynthetic rate and stomatal conductance. In the original manuscript, our explanation of their physical significance was lacking. In response, we will provide a more detailed exploration of the physical basis of these sensitive parameters, incorporating the CoLM model's parameterization scheme in Section 3.1 of the revised manuscript to improve both the depth and clarity of the analysis.

Comment #9: After the optimization , some optimized parameters – P36 and P3 – reach one of the bounds of its variation interval. In my opinion, this is a bit troublesome and this question the way the bounds of the intervals were chosen.

Response:

The optimized values of P36 and P3 have indeed reached the boundaries of their respective ranges. However, this does not necessarily indicate a flaw in the optimization process or its outcomes. First, the parameter ranges were reasonably established based on experimental data, literature, and field conditions, ensuring compliance with physical constraints rather than being arbitrarily set. Second, the fact that these parameters reached their boundary values may suggest that these values represent the optimal solution, indicating the optimization algorithm identified the best configuration within the given range. Nevertheless, we also recognize that this outcome could be influenced by insufficient sample data or data uncertainty, potentially causing the mode to rely on boundary values for optimal fitting. In response, we will appropriately adjust the parameter ranges - either narrowing or expanding them – in the revised manuscript to explore the impact of these boundaries on the optimization results and to assess the robustness of the outcomes.

Specific comment:

Comment #1: In the abstract and conclusion, the impact of efficiency is sometimes in % and sometimes in raw values. I think it's more convenient and easier to use % everywhere.

Response:

Thank you for your suggestion. We have updated both the abstract and conclusion to consistently use percentages throughout.

Comment #2: L52: what's the difference between LSM and soil-vegetation-atmosphere coupled models?

Response:

Thank you for your question. Land Surface Models (LSMs) focus on simulating energy, water, and carbon exchanges between the land surface (including soil, vegetation, and snow) and the atmosphere. They are typically part of larger climate or weather models. In contrast, soil-vegetation-atmosphere coupled models integrate the interactions between soil, vegetation, and the atmosphere, capturing more complex feedback such as the effects of soil moisture and vegetation changes on atmospheric processes. While LSMs are often part of these coupled models, the latter provides a more comprehensive view of these interactions. If you believe it is better to combine the two model types in the manuscript, we can revise accordingly.

Comment #3: L135: please specify the signification of delta here?

Response:

Thanks for your question. The delta $\delta(\varepsilon)$ denotes the noise variance, which serves to estimate the error in the output Y caused by random noise ε. The Delta Test (DT) method aims to quantify the noise by analyzing the differences between the nearest neighbors in the input space. In this context, δ measures the noise level present in the output. In essence, the DT method calculates the noise variance $\delta(\varepsilon)$ by comparing the output values of nearest neighbors with those of the corresponding original data points. By minimizing $\delta(\varepsilon)$, the method identifies the subset of input parameters that most significantly influences the output.

Comment #4: L187:extensive dataset(104 or 105 or more):I guess it 10^4 and 10^5?

Response:

Thank you for pointing out this mistake. You are correct, it should be 10^4 and 10^5. I have made the necessary corrections in the revised manuscript.

Comment #5: Fig 3: change the values on the x-coordinates. Not easy to read.

Response:

Thank you for your valuable suggestion. In the revised manuscript, we will adjust Figure 3 by plotting the three distributions on a single, linearly-scaled x-axis. This modification will allow for a clearer comparison of the parameter distributions across the different optimization methods.