Replies to Referee #2, GMD-2022-264

Jiaxu Guo on behalf of all authors

April 18, 2023

Thank you very much for your patient and detailed comments on our work[1]. These valuable comments are very helpful for us to improve this paper. After carefully reading all the questions, we have answered each of them and will make appropriate corrections in the revised version of our manuscript.

In this attachment, the red paragraphs represent your refree comments, and the black paragraphs below are our corresponding replies.

1 Replies to major issues

In this paper, although the authors evaluate the accuracy of the NN model in terms of precipitation, it probably exists the inconsistent between ML and real model, which don't be highlighted in this paper.

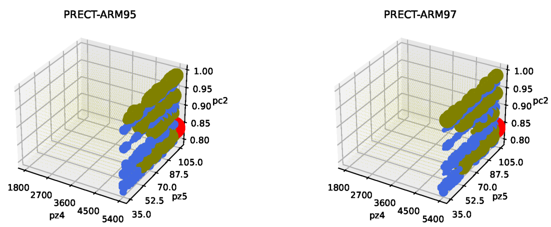
We agree that it may be difficult for a surrogate model to be fully consistent with a real model. We use the RMSE in training as a loss function to verify the correctness of the method, that is, whether the parameter tuning of SCAM can be accelerated by training a surrogate model, and to compare different regression methods. We will add more descriptions about the inconsistency to the manuscript from two perspectives: From a model training perspective, the value of loss function indicates the error in the training. However, the process by which we train the model is also the process by which the error is gradually reduced. From a practical perspective, when we use the surrogate model for optimization, the solutions obtained are also validated in the original SCAM case to ensure as much consistency as possible.

The authors believe that due to the high computational cost of the GCM, the SCAM can be the alternative model for parameter SA and tuning. In reality, the optimal parameters tuned in SCAM could not be suitable for GCM, due to the global regions and more complex large-scale circulation.

We agree that the solution set obtained by parameter tuning on SCAM is not directly applicable to GCM. However, through our attempts at parameter tuning on SCAM cases located in different regions, we can find commonalities and patterns in the parameter response of these cases. We believe such an idea can be applied to the parameterization scheme of the GCM. Our exploration of parameterization solutions using SCAM is mainly from a methodological and ideological point of view. That is, SCAM is a simpler and less costly way to perform numerical simulations. The training of a surrogate model for SCAM is a further extension of this idea.

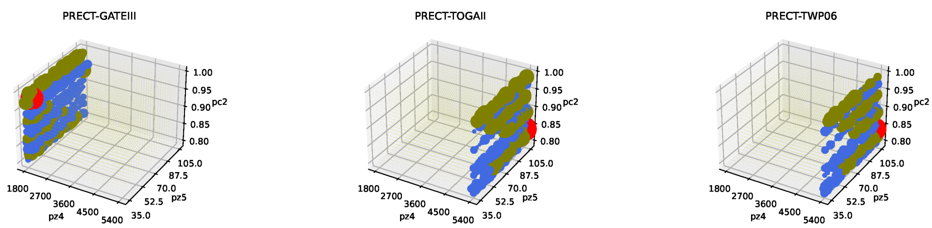
Overall this manuscript, the organization and writing are not clear and should be well structured. There are a very large number of language errors. The English writing should be greatly improved.

We will reorganize the structure of the article in a revised version to rationalize the logic. We will also work on improving the English writing style to make it more fluent.



(a) (b)

AC-Figure 1: Results of a three-parameter full-space grid search for ARM95 and ARM97 using the surrogate model. The points closest to the observed data are shown in red, those ranked 2-64 are shown in olive, and those ranked 65-256 are shown in blue. Where, (a) is ARM95 and (b) is ARM97.



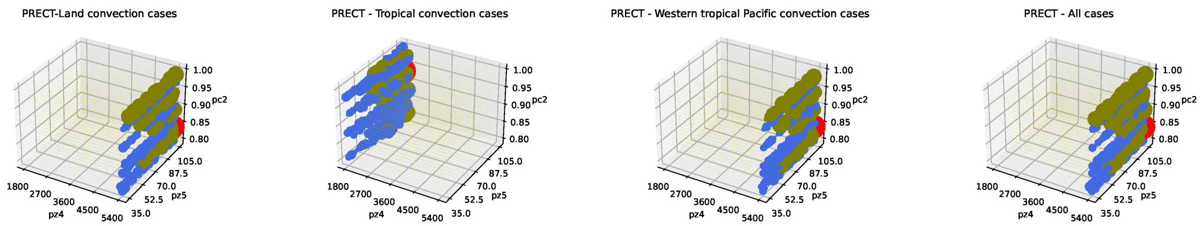
(a) (b) (c)

AC-Figure 2: Results of a three-parameter full-space grid search for GATEIII, TOGAII and TWP06 using the surrogate model. The points closest to the observed data are shown in red, those ranked 2-64 are shown in olive, and those ranked 65-256 are shown in blue. Where, (a) is GATEIII, (b) is TOGAII and (c) is TWP06.

The authors separately tune the parameters in SCAM for each site and get the different sensitive parameters and different optimal values. It is difficult to transfer this information to GCM. If the authors can do the multi-objective tuning for these sites with the same parameters, it could be helpful for global model tuning because these SCAM sites indeed represent the different regimes.

We agree that a combined optimization that tries to minimize the differences against observations across all five cases would be more meaningful. Such tests and results were included in a previous version of the draft. However, we removed the results at certain stage for a more focused description in the result section.

As suggested by the reviewers, we have carried out a careful analysis and done corresponding experiments. After retraining the surrogate model for each of the five cases by using ResNet,



(a) (b) (c) (d)

AC-Figure 3: Results of a three-parameter full-space grid search for the multi-objective scenarios using the surrogate model. The points closest to the observed data are shown in red, those ranked 2-64 are shown in olive, and those ranked 65-256 are shown in blue. Where (a) is the scenario of two land convection cases, (b) is the scenario of three tropical convection cases, (c) is the scenario of two western tropical Pacific cases, and (d) is the scenario of all five cases.

we also carried out multi-objective optimization for each of the four scenarios. Separately, they are: for ARM95 and ARM97 cases, for three tropical convection cases, for TOGAII/TWP06 (they're close in location) and all five cases.

Combined with results shown in AC-Figure 1 and 2 of our grid search for the five cases individually, it can be seen that cases located at the same or similar locations really have similar distributions of precipitation output in response to parameters. Therefore, this kind of joint optimization based on the multi-objective idea is of general interest.

For the workflow, there should be a "metrics" component because it is very important for tuning. No matter SCAM or GCM, the tuning metrics could be the cost function between model simulations and observations. The different designs could affect the optimization. In terms of the metrics, it could consider the 1) different statistic errors between simulation and observation, such as RMSE, performance score like Yang et al. (2013), 2) one objective or multiple objectives, and how to deal with the multiple objectives.

I agree that "metrics" are important factors to consider in the workflow. We will use RMSE as the metric of error in the revised version.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{1}$$

After using RMSE as the metric for training the model and parameter optimization, the corresponding results were recalculated. In particular, the losses during the training of the model are shown in AC-Table 2. The 3D parameter space responses for the five cases themselves and for the four multi-objective joint conditioning scenarios are shown in AC-Figure 3 to 2.

Regarding multi-objective tuning, we also designed corresponding experiments when we originally wrote this paper. Four scenarios are included, separately, they are: combined tuning for ARM95 and ARM97 cases, combined tuning for three tropical convection cases, combined tuning for TOGAII/TWP06 and all five cases. From results we can see that the parameter responses for ARM95 and ARM97 have a similar distribution and thus the results of the multi-objective optimisation for both of them reflect this. The same trend is reflected similarly in the TOGAII and TWP06.

Line 35: The statement that the Morris SA method cannot get the interactive sensitivity could be wrong. Aurally, the standard deviation of MOAT samplings can stand for the interactive effect of one parameter with others (Morris, 1991).

We will refine the description of the Morris SA method in the revised manuscript so as not to introduce ambiguity: "Morris SA can give the individual sensitivity of each parameter, including their interaction sensitivity. However, this is not intuitive enough if the user wants to know directly from a combined perspective which set of parameters has the most significant effect on the results."

Line 45: as the part of introduction, the authors should explain the challenge of the SA methods, why you choose Morris and Sobol, the computational cost issue, surrogate problems using machine learning. If there are previous works, what's your contribution?

We will complete this part based on a further full investigation. In the revised version, we will give a more detailed explanation. Both Morris and Sobol are typical SA methods that have a wide range of applications in many fields. As there are already proven application examples, it makes sense to conduct experiments based on the above methods. In addition to this, we have introduced several new SA methods that have been proposed in recent years. Although SCAM, as a single-column model, already consumes less computational resources than GCM, the computational resources of a system are not infinite. To further explore their parametric features, a study using surrogate model is necessary. Surrogate models [2] can significantly reduce the computation time of individual tasks, thus making it possible to scale up experiments.

Our contribution lies in the fact that we have trained the surrogate models on SCAM using a regression-based approach, and with the help of the surrogate models we have conducted parameter sensitivity tests for combinations, as well as tuning for the most significant parameter combinations.

Line 55: the authors should do comprehensive literature research, even for GCM, there are a large number of work for tuning, such as Yang et al. (2013) and Zou et al. (2014). In addition, the NN surrogate model is used to tune as well. But the authors don't introduce the previous work and challenge in terms of this issue. The introduce section should be more clear.

We will complete the introduction section by conducting a more detailed literature survey of the work related to the content of this paper. This also includes the literature you mention, such as [3] and [4]. Yang et al. [3] analysed the sensitivity of nine parameters in the ZM deep convection scenario for CAM5 and used the SSAA (Simulated stochastic approximation annealing) method to optimize the precipitation performance in different regions by zoning. Zou et al. [4] conducted a sensitivity analysis for seven parameters in the MIT-Emanuel cumulus parameterization scheme in RegCM3. The precipitation optimization process for the CORDEX East Asia domain was carried out using the MVFSA (Multiple very fast simulated annealing) method.

The current challenges lie in the following areas. (1)No similar experiments have been conducted on SCAM. The short run time of the SCAM makes it easier to obtain more samples in a short period of time and thus scale up the experiments. (2)The usual SA methods will give the sensitivity of the individual parameters. But when we look at a set of parameters, is the best combination of N parameters the top N sensitivity of a single parameter? This is a question worth exploring. (3) As various neural network methods are applied to the field of regression, more appropriate regression methods should be applied to the parametric study of Earth system models. Unlike the various public data sets commonly used, ESM-based experiments will also further enrich the practical implications of research in the field of regression analysis.

In our manuscript, we innovatively use a neural network-based agent model for parameter tuning of different cases in SCAM. For each case studied in the paper, the surrogate models are trained separately based on different methods, and the best performing model is selected by their training errors (RMSE).

Line 75, Acutely, there are existing SA and tuning workflow used in climate models, such as PSUADE and DAKOTA, the authors don't compare their workflow to these packages. It's not new for the community.

This is indeed something we need to improve further in the literature survey. We have done a survey of the packages such as PSUADE, DAKOTA and STATA etc., including some studies based on their work in different fields. These packages can indeed implement the functionality of SA and tuning. We also compared the workflow proposed in this paper with the above software packages. From a method perspective, we added the comparison of the new SA methods in recent years, such as RBD-FAST, Delta and HDMR. The above approaches haven't been fully supported by all the packages above. For using neural networks to train surrogate models, DAKOTA currently only supports neural networks with a single hidden layer. Meanwhile, the proposed method uses more types of neural networks and supports the adaptive selection of the best performing network to train the surrogate model.

Line 88: The authors don't mention the 30% improvement for tuning error and computational cost. How do they come from?

The tuning error is compared with control experiments. This is the result of a simulation using the default parameter values of SCAM. The improvement in computational overhead comes from comparing it with a traditional optimization workflow. In the original optimization workflow, the original SCAM is invoked for each calculation of the objective function. In the approach proposed in this paper, the objective function is replaced with a surrogate model. This allows the overall workflow execution time to be compressed, thus the computational overhead could be reduced.

Table 2 is wrong. Each IOP file includes many variables, not just these four variables. Therefore, the statement that you choose precipitation is wrong.

AC-Table 1: Observed variables included in the IOP file of each case.

Variable	Description	ARM95	ARM97	GATEIII	TOGAII	TWP06
Prec	Precipitation rate	✓	✓	✓	✓	✓
totcld	Total cloud	✓	✓	-	-	✓
shflx	Surface sensible heat flux	✓	✓	-	✓	✓
lhflx	Surface latent heat flux	✓	✓	-	✓	✓
U	Eastward wind speed	✓	✓	✓	✓	✓
V	Northward wind speed	✓	✓	✓	✓	✓
Q	W.V. Mixing Ratio	✓	✓	✓	✓	✓
T	Temperature	✓	✓	✓	✓	✓
omega	vertical motion	✓	✓	✓	✓	✓
windsrf	Surface wind speed	✓	✓	-	✓	✓
REHUM	Relative humidity	-	-	✓	-	✓
CAPE	Convective available potential energy	-	-	✓	-	-

We are sorry for the misunderstanding. We agree that there are many variables in the IOP file for each case, not just these four, as can be seen from AC-Table 1. We include Table 2 in our manuscript mainly to illustrate that the variables contained in different IOP files are different. We will add this in the revised manuscript.

Line 120: This issue could not be a significant challenge. Some simple scripts can collect the data.

Yes, scripts are able to do that. Of course, we will also achieve efficient management of experimental data in our method. Due to the amount of data generated during this stage, better data management is necessary. We will adopt a more reasonable description in the revised manuscript.

We will revise these statements as follows: "In this step, we also integrate the collection and processing script for the post-sampling results. As the output of SCAM is stored in binary files in NetCDF format, the precipitation variables we want to study need to be extracted from a large number of output files in order to proceed to the next step. This will further accelerate the degree of automation of scientific workflows and thus accelerate the conduct of research in this area of the earth system models."

The statement about Morris is wrong, see 3.

We plan to correct this issue as follows: "Morris SA can give the individual sensitivity of each parameter, including their interaction sensitivity. However, this is not intuitive enough if the user wants to know directly from a combined perspective which set of parameters has the most significant effect on the results."

In section 3.1: sampling is not equal to SA. In this section, the authors introduce the SA methods. You should consider change the structure or change the title.

We agree that sampling can't be confused with SA. We will refine the article structure of this section in a revised version to make the expression easier to understand. The sampling method and the SA method will be split into two subsections to be described.

AC-Table 2: RMSE of different surrogate models for five cases.

Case	LR	RF	MLP	XGBoost	ResNet
ARM95	0.235	0.197	0.294	0.184	**0.038**
ARM97	0.188	0.158	0.555	0.136	**0.045**
GATEIII	0.646	0.432	0.108	0.538	**0.137**
TOGAII	0.179	0.112	0.245	0.118	**0.041**
TWP06	0.344	0.220	0.304	0.220	**0.040**

Subsection 3.1, entitled "Sampling methods used in this framework", will be devoted to the sampling methods covered in this paper and the rationale for their selection.

Subsection 3.2, entitled "SA methods used in this framework", will provide a more detailed description of the SA methods covered in this paper and the rationale for their selection. At the end of this subsection, our proposed combined SA methods for supplementary validation will be introduced.

Line 177: it could be better to compare NN with other surrogate models, such as xgboost, ResNet.

Yes, the properties of the various surrogate models are something we all care about. Through our research, we learned that both XGBoost[5] and ResNet[6] can be used to perform regression tasks and train surrogate models. Here, we will compare the effectiveness of several methods such as LR (Linear Regression), RF (Random Forest), MLP (Multi-Layer Perceptron), XGBoost and ResNet for training surrogate models. The results are shown in AC-Table 2. Since ResNet has better performance in model training, we will choose ResNet as the network for training the surrogate model. Subsequent experiments will also be supplemented in the revised paper.

Line 184: The 768 samples seem not enough for training NN, do the authors evaluate the performance of NN? In Figure 4, how do you define the accuracy?

Yes, we have evaluated the performance of NN and measured its ability on regression. After our analysis of the sampling results, we found that 768 samples can already cover the range of values of the parameters and output variables. Therefore, the number of parameters selected is sufficient to obtain a good training effect. The results of the training process and the experimental results also support our conclusion. For the second point, The R^2 score is used as the accuracy rate in Figure 4. It is defined as

$$R^2 = \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2} \tag{2}$$

where $\hat{y}$ is the predicted value and $\bar{y}$ is the mean of the test set. In addition to this, RMSE was also used as an important metric to measure the loss during training.

Line 185: Do the authors do the hyper-parameter tuning for NN?

In the initial version of our manuscript, our aim was mainly to verify the feasibility of this approach, so we used a set of empirical parameters to train the NN. Now we will conduct hyperparameter tuning trials in NN training to make the proposed method more solid. We will refine these in a revised version of the manuscript.

Line 195: Due to the uncertainties of each method, ensemble can't guarantee to reduce the error.

We agree that there are errors in out initial manuscript. We feel sorry for the term "integrate" in the original paper may bring some ambiguity. We've used several SA methods to make a more intuitive side-by-side comparison, which allows us to find the best method for each case. It's not necessarily about reducing the error.

Line 210: Equation (1), the left hand of this equation is not the number of processes. It should be the number of simulations. The number of simulations could depend on different sampling method. For Morris, it could not require such number of samples. It is not clear for this description.

Since SCAM is a single-process task, one simulation is equivalent to one process. Of course, as you said, it is necessary to distinguish between simulation and process more accurately. For Morris, 768 was used as the sample number to keep the sample number of various sampling methods consistent. We will add the above contents to refine these issues in a revised version of the manuscript.

For a SCAM simulation, it usually requires 10-20 minutes, why do you require more than one hour?

We agree that for a single SCAM job it only takes one process to run, and it does not take very long for a single SCAM to run. The one hour mentioned here is mainly the time it takes to complete a whole workflow of parameter analysis and tuning. This includes multiple iterations and delays in queuing batch jobs. We will refine these descriptions above in the revised manuscript and try to avoid ambiguities.

It is confusing that you can re-use the sampling from SA to train the surrogate model for tuning, but in section 3.3, you mention that the surrogate model can be also used in SA?

We agree that the trained surrogate model can be also used in SA. Since we want to find the most sensitive set of parameters in combination, this process requires a larger experiment scale. Therefore, to perform large-scale parametric analysis experiments faster, we use surrogate models to speed up the process.

Line 233: how do you get the conclusion? It is not convinced.

Here, the running time is compared to one full simulation of the original SCAM. Here we are trying to make two points. On the one hand, the surrogate model can simulate the output of SCAM more accurately. This makes it an effective alternative to SCAM in terms of parameter response. On the other hand, the execution time of the surrogate model is very short, which saves a lot of time for parameter analysis and optimization tuning. Combining these two reasons above, the use of a surrogate model for parameter tuning of SCAM is a very reasonable strategy. We will add some necessary information in our revised manuscript to make our conclusion more convincing.

It is very confusing for section 3.4 and is difficult to follow your idea. You could consider to re-organize this section.

We're sorry for not being able to make it easy for you to follow our idea. We will reorganize the language to make it easier to understand. In this subsection, we focus on an enhanced SCAM parameter tuning process. There are two main contributions. On the one hand, we use the trained surrogate model as the objective function for optimization, which allows the optimization time to be compressed considerably. On the other hand, we propose an enhanced parameter optimization process based on grid search. We use grid search to reduce the search range in the optimization process and thus obtain better results in fewer iterations. In the revised manuscript we will elaborate further on these two points above.

Line 270: how many samples do you have for the correlation? Do you do the p-value test?

Here, we carried out correlation analysis on the respective optimal solutions of the five SCAM cases. Therefore, a total of five vectors were used to calculate the correlation For the second point, now we will add tests on p-values in the revised manuscript.

Figure 7: it is difficult to evaluate the tuning performance in Figure 7. It could be better to use metrics like Yang et al. (2013).

What we want to express in Figure 7 is the statistics of the sampling results. The phrase "when the parameters are tuned" refers to the fact that the parameters are given different values during the sampling stage, not the final tuning stage. We're sorry for this ambiguity and will respect your suggestions to revise the description appropriately. When revising, we will fully refer to the relevant statements of [3] in order to make the results more clear and easy to understand.

Figure 4, Figure 5 are the results but appear in section 3 (methods). It could be reorganized.

The results described in these two figures belong to the pilot test, which is to prove the rationality of our experimental ideas. We will respect your suggestions and provide appropriate revisions to make the organization of our paper more rigorous.

Line 313: pz2 (c0_ocn) should be high influence on the ocean case. But in Figure 8, it doesn't have the high effect on PRECT at TWP. Could you explain the reason?

It can be seen from Figure 8 that no matter which sensitivity analysis method is used, pz2 (c0_ocn) has a certain influence on the results. It is a deep convection parameter related to the ocean-land intersection, but this does not mean that it must be the one that has the most influence on TWP. The influence of the ocean has been demonstrated here.

Line 317: the reason for the different between ARM 95 and ARM 97 is not convinced.

The difference between the cases is also the focus of our interest and attention. Indeed, the gap between ARM95 and ARM97 indicated by the experimental results in this paper is worthy of attention. We retrained the surrogate model using ResNet and used RMSE as the metric of error. We found that the response of the precipitation output to the parameters in the new surrogate model was very similar for both cases which can be shown in AC-Figure 1. This suggests that in the original version, the difference between the two examples is due to the error in the model. Your advice on the choice of the neural network would also be much appreciated here.

Line 345: are the 16 iterations enough for convergence? Why don't use the general optimization, such as PSO, GA that you mentioned in the introduction section?

In our experiment, 16 rounds of iteration can already make the results converge, as shown in Figure 6 (in our preprint paper[1]). So the selection of the number of iterations is sufficient. WOA[7] is chosen here mainly because it is an optimization algorithm proposed later. Theoretically speaking, in the optimization stage of the workflow proposed in this paper, no matter what kind of optimization algorithm is selected, the purpose of optimization can be achieved, including PSO, GA, etc. With full respect for your suggestions, we can also compare the performance of these optimization algorithms in our framework.

2 Replies to minor issues

In Fig. 1, is there an arrow pointing from "SA methods" to "testing of combinations"?

There isn't an arrow here because only the trained surrogate models are used for testing combinations of parameters.

Caption in Figure 1: The sentence "SCAM launcher, the data collector and the jobs therein represent the batch execution of the SCAM algorithm" should be rephrased. What are the "jobs" and "batch execution"? It is not clear.

There isn't an arrow here because only the method of the surrogate model trained using neural networks will be used for parameter combination SA. Here, a job refers to a single simulation of SCAM. A batch is a collection of jobs that are submitted to the computing queue at once. We will elaborate more in detail in the revised manuscript.

Line 25: should explain "ne30"

This is a description of the spectral element method grids[8]. It refers to a model grid with a $ne30np4$ spectral element (approximately 1-degree) atmosphere and land grids. $ne[X]np[Y]$ are cubed sphere resolutions where X and Y are integers. The short name generally is $ne[X]$. We will add more details in our revised manuscript.

Line 38: the reference of Sobol method is wrong, pls use the original paper.

We will correct it [9] in the revised manuscript.

Line 43: The QMC and LHC are sampling methods, not SA methods.

We will correct this in the revised manuscript. Sampling methods and SA methods will be more clearly distinguished to avoid confusion.

In Table 3, how do you select these parameters? And how do you define the range of each individual?

These parameters were chosen mainly from reference [10]. The parameters range from 50% to 150% of the default value.

Line 94: there should be a reference for SCAM5.

We will add reference [11] in the revised manuscript to refine the introduction of SCAM5.

Line 102: all sites belong to ARM.

We will correct this issue in the revised manuscript. We will also carry out a more detailed study later.

Line 108: "in the code" change to "in the model"

We will correct it in the revised manuscript.

Line 116: "is an important issue" change to "are important issues"

We will correct it in the revised manuscript.

Line 130: It is only suit for SCAM. For GCM, it is impossible.

Indeed, as you say, it is not practical to run large batches of GCM jobs even on HPC. We will refine this description in the revised manuscript.

Line 165: please explain the "second-order sensitivity"

It refers to the mutual influence between two parameters. We will refine this exposition in the revised manuscript.

Table 4: the reference of Sobol is wrong.

We will correct this reference [9] in the revised manuscript.

Line 174: please consider the correct position of this sentence.

We will correct it in the revised manuscript.

Line 193: It is not clear for "not a direct evaluation."

We will add some necessary information and details in our revised manuscript. What we want to express here is the following: using the above methods, the set of M parameters that most influence the result may be difficult to obtain directly. The method proposed in this paper just fills this gap.

Line 195: add "have" before "its"

We will correct it in the revised manuscript.

Line 215: how do you get the number of multiple thousand?

Combining Equation (1) with the practical problem studied in this paper, we can see that when $C = 5, p = 3$ and $L = 10$, 5,000 simulations are needed even if only one parameter combination is considered. This will be astronomical when more combinations are considered. We will also refine the above in the revised manuscript.

References

[1] J. Guo, Y. Xu, H. Fu, W. Xue, L. Wang, L. Gan, X. Wu, L. Hu, G. Xu, and X. Che, "A learning-based method for efficient large-scale sensitivity analysis and tuning of single column atmosphere model (scam)," *Geoscientific Model Development Discussions*, vol. 2022, pp. 1–28, 2022. DOI: 10.5194/gmd-2022-264. [Online]. Available: https://gmd.copernicus.org/preprints/gmd-2022-264/.

[2] K. Cheng, Z. Lu, C. Ling, and S. Zhou, "Surrogate-assisted global sensitivity analysis: An overview," en, *Structural and Multidisciplinary Optimization*, vol. 61, no. 3, pp. 1187–1213, Mar. 2020, ISSN: 1615-147X, 1615-1488. DOI: 10.1007/s00158-019-02413-5. [Online]. Available: http://link.springer.com/10.1007/s00158-019-02413-5 (visited on 04/03/2023).

[3] B. Yang, Y. Qian, G. Lin, L. R. Leung, P. J. Rasch, G. J. Zhang, S. A. Mcfarlane, C. Zhao, Y. Zhang, and H. Wang, "Uncertainty quantification and parameter tuning in the cam5 zhang-mcfarlane convection scheme and impact of improved convection on the global circulation and climate," *Journal of Geophysical Research Atmospheres*, vol. 118, no. 2, pp. 395–415, 2013.

[4] L. Zou, Y. Qian, T. Zhou, and B. Yang, "Parameter tuning and calibration of regcm3 with mit–emanuel cumulus parameterization scheme over cordex east asia domain," *Journal of Climate*, vol. 27, no. 20, pp. 7687–7701, 2014. DOI: https://doi.org/10.1175/JCLI-D-14-00229.1. [Online]. Available: https://journals.ametsoc.org/view/journals/clim/27/20/jcli-d-14-00229.1.xml.

[5] W. XingFen, Y. Xiangbin, and M. Yangchun, "Research on User Consumption Behavior Prediction Based on Improved XGBoost Algorithm," en, in *2018 IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA: IEEE, Dec. 2018, pp. 4169–4175, ISBN: 978-1-5386-5035-6. DOI: 10.1109/BigData.2018.8622235. [Online]. Available: https://ieeexplore.ieee.org/document/8622235/ (visited on 04/03/2023).

[6] L. Shi, C. Copot, and S. Vanlanduit, "Evaluating Dropout Placements in Bayesian Regression Resnet," en, *Journal of Artificial Intelligence and Soft Computing Research*, vol. 12, no. 1, pp. 61–73, Jan. 2022, ISSN: 2449-6499. DOI: 10.2478/jaiscr-2022-0005. [Online]. Available: https://www.sciendo.com/article/10.2478/jaiscr-2022-0005 (visited on 04/03/2023).

[7] S. Mirjalili and A. Lewis, "The Whale Optimization Algorithm," *Advances in Engineering Software*, vol. 95, pp. 51–67, 2016, ISSN: 18735339. DOI: 10.1016/j.advengsoft.2016.01.008.

[8] (). "Model grids – cime master documentation," [Online]. Available: http://esmci.github.io/cime/versions/master/html/users%5C_guide/grids.html.

[9] I. Sobol, "Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates," *Mathematics and Computers in Simulation*, vol. 55, no. 1, pp. 271–280, 2001, The Second IMACS Seminar on Monte Carlo Methods, ISSN: 0378-4754. DOI: https://doi.org/10.1016/S0378-4754(00)00270-6.

[10] Y. Qian, H. Yan, Z. Hou, G. Johannesson, S. Klein, D. Lucas, R. Neale, P. Rasch, L. Swiler, and J. Tannahill, "Parametric sensitivity analysis of precipitation at global and local scales in the community atmosphere model cam5," *Journal of Advances in Modeling Earth Systems*, vol. 7, no. 2, 2015.

[11] P. A. Bogenschutz, A. Gettelman, H. Morrison, V. E. Larson, D. P. Schanen, N. R. Meyer, and C. Craig, "Unified parameterization of the planetary boundary layer and shallow convection with a higher-order turbulence closure in the community atmosphere model: Single-column experiments," *Geoscientific Model Development*, vol. 5, no. 6, pp. 1407–1423, 2012. DOI: 10.5194/gmd-5-1407-2012. [Online]. Available: https://gmd.copernicus.org/articles/5/1407/2012/.