Replies to Referee #2, GMD-2023-164

Xianwei Wu, Liang Hu, Lanning Wang, Haitian Lu, and Juepeng Zheng

November 15, 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 blue paragraphs represent your comments, and the black paragraphs below are our corresponding replies.

1 Replies to 1-5 questions

Replies to these questions are as follows:

1. The presentation of the algorithm and techniques could be more concise. It would further benefit from clear mathematical notation and equations, a clear nomenclature, and a clearer order. For instance, RMSE is used before properly introduced. The calculation of Vŝ and Vd̂ are hard to follow and not right away clear.

The description of CAND is described in Algorithm 1:

Algorithm 1 Candidate point strategy

Compute $s^{max} \leftarrow \max_{x\in\Omega} s(x)$ and $s^{min} \leftarrow \min_{x\in\Omega} s(s)$
for each $x \in \Omega$ **do**
$$V^S(x) = \begin{cases} \frac{s(x)-s^{min}}{s^{max}-s^{min}} & if\ s^{max} > s^{min} \\ 1 & else \end{cases}$$
Calculate corresponding value of objective function for each sample.
end for
for each $x \in \Omega$ **do**
$\Delta(x) = min_{y\in A} d(x,y)$;
end for
Compute $\Delta^{max} \leftarrow \max_{x\in\Omega} s(x)$ and $\Delta^{min} \leftarrow \min_{x\in\Omega} s(x)$
for each $x \in \Omega$ **do**
$$V^D(x) = \begin{cases} \frac{\Delta(x)-\Delta^{min}}{\Delta^{max}-\Delta^{min}} & if\ \Delta^{max} > \Delta^{min} \\ 1 & else \end{cases}$$
end for
return $argmin_{x\in\Omega} wV^S(x) + (1-w)V^D(x)$

Where, the Ω represents the random samples generated in this iteration process. $S(x)$ represents the predict value of point x generated by surrogate model. $\Delta(x)$ is the distance from point x to the current sampling point set A and y represents each point in set A.

We will add these to the revised manuscript.

We will improve the structure of the study in revised manuscript.

2. The paper lacks describing links between the physics and the choice of parameters. Why do certain parameter combinations perform better (for instance, what do they affect, how does that affect the general performance etc.)

1).The purpose of this work is to improve the CAM5 precipitation simulation result accord to parameter tuning method. rather than analyzing the mechanism. We believe that the goal has been achieved and it is a complete work. In this paper, we propose a surrogate model based method which can quickly calibrate parameters, and improve CAM5 precipitation using the multi-level surrogate model method and non-uniform parameterization schemes. We find a more suitable set of parameters for each region.

2).We do not change the physical processes in the parameterization scheme, we only changed the values of the parameters, or different values in different regions. We believe that these positive improvements achieved by changing the values of the parameters. In [2], there is more introduction to the mechanism, We also refer to this work when selecting parameters and determining the range of these parameters. We will try to explain these effects from the parameter value changes.

3. The paper does not discuss that in general tuning for a single metric is not required as a climate model has many different metrics that need to be fulfilled. Thus, it is required to discuss how the precipitation tuning might degrade other fields. For instance, what is the effect on the global mean temperature from this tuning etc.

In this paper, we propose a parameter tuning method to improve CAM5 precipitation simulation results. The proposed method belongs to a single objective optimization method. This method maybe cannot optimize multiple objectives simultaneously. Multi-objective surrogate model-based parameter tuning method is also one of our future research topics. We will try to discuss that in nonuniform parameter parameterization scheme, how other metrics such as temperature and specific humidity change after precipitation decreases and add these results in revised manuscript.

4. The presentation of introducing the non-uniform parameter values is not entirely clear. Why should that be? What is the physical explanation for using different parameter values in different places? Shouldn't the physics be independent of the location particularly in regions which are relatively similar (South Pacific, Nino?)? Particularly you tune for different ocean regions; it is not clear why different ocean areas should have different parameter tunings. (I could understand a land vs ocean parameter change, however, different oceans or land masses requires more careful introduction and physical justification)

In this paper, the motivation for the nonuniform parameter parameterization scheme is as follows:

1).It is well known that CAM5 is a well tuned model, however globally optimal parameters do not necessarily mean they are the best solutions for every region.

2).Regional optimization experiments demonstrate that some regions have optimal parameters, leading to better results than default parameters.

3).Our experiments show that there is a "rocker effect" in the influence of parameters on precipitation. The same parameter values have different effects on different regions, making it challenging to optimize precipitation for all regions using a single parameter.

In summary, we proposed the nonuniform parameter parameterization scheme.

In section 4.3, we found that these regions have better parameter combinations, even if they are in similar positions.

5. The presentation is unclear as to why is a GP only used for the regional-level surrogate models and not for the global?

Unlike global-level surrogate-model, the local-level surrogate model is required to have a higher accuracy with fewer samples. So that construction method of the loacl model is more important. according to [3, 4]. GP is selected to construct local-level surrogate.

Similar to the global-level surrogate model, we conduct a cross-validation for several method to construct local surrogate model, the results are shown in Figure 1:

We will add these results to the rivised manuscript.

2 Replies to major comments

Replies to major comments are as follows:

1. LL.37-70: it would be worth discussing also the tuning approaches by Hourdin and Williamson in more detail
(http://link.springer.com/10.1007/s00382-013-1896-4 ;
http://link.springer.com/10.1007/s00382-014-2378-z ;
https://gmd.copernicus.org/articles/10/1789/2017/ ;
https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2020MS002423 ;
https://onlinelibrary.wiley.com/doi/10.1029/2020MS002225 ;
https://onlinelibrary.wiley.com/doi/10.1029/2020MS002217)

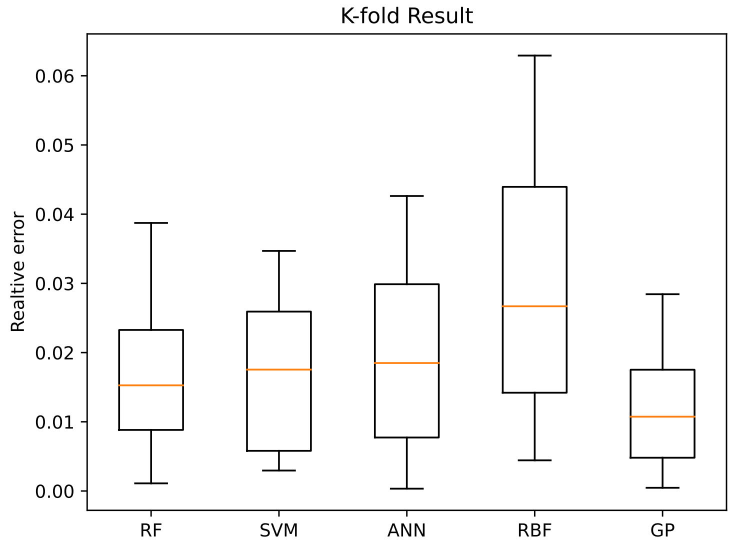

Figure 1: local model surrogate model cross-validation

Thank you for your suggestion. These works have many highlights in the description of the methods and the explanation of the physical mechanisms. For example, In [5, 6], authors discuss machine learning for ESM calibration and use machine learning method on both single-column model and global model. In [7, 8], authors propose a history matching method to analyse the uncertainty of parameters in HadCM3. The author also proposed the concept of "over-tuning" to prevent over fitting in the tuning results.

We believe that our work is also an extension of these tasks and a supplement to certain aspects. They are very helpful in improving our manuscript.

2.Ll.135-138: How do you inform the parameter range of the parameters to tune for? How do you choose exactly those parameters? Generally, there are more parameters in the parameterizations; why exactly those 6?

We know that the sensitivity of parameters has an important impact on pattern tuning. If insensitive parameters are selected for disturbance and adjustment, the simulation results will hardly change significantly. Therefore, the selection of sensitive parameters is one of the important conditions for CAM5 parameter tuning. Many previous studies have conducted sensitivity-related studies on precipitation-related parameters in CAM5. The parameter range and the reason we choose these parameters are according to these studies [2, 9]. These parameters we selected are proved most to sensitive precipitation in these studies.

3. LL. 140-141: Why do you choose particularly those regions? Why are they important?

These regions and there range are selected according to [10]. These regions have a high amount of precipitation value, and tuning these regions can effectively improve the simulation of CAM5 precipitation. More than that, as can be seen in Figure 2, in default experiment, these areas have a certain degree of error compared to observational data. Since CAM5 is a well-calibrated model. Our experiment also proves that the global-oriented tuning effect is not significant. So that we try to research these regions, find a set of parameters to improve the simulation results over these regions.

4. Section 3: I would suggest sticking to terminology! Whenever you talk about an actual model simulation I would suggest using ESM/CAM5 or something like that. Otherwise you start mixing up terms such as global-model, global-level, complex model which does not make it easy to follow which of all the models you refer to or whether it is a new one.

In this paper "global-model" and "global-level" model are equivalent. They represent the global-level surrogate model in the proposed method. "Complex model" represents the optimization problem which the fitness function is hard to calculate. Perhaps the word "model" has different meanings is different terms such as "surrogate model", "complex model" and "earth system model". In order to avoid

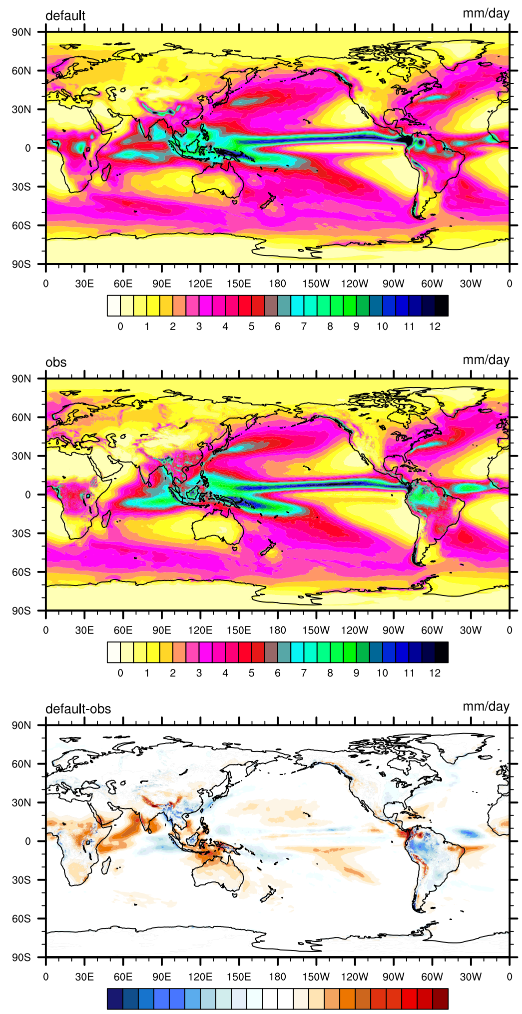

Figure 2: The precipitation distribution of default experiment, there are default experiment, observation data, difference between default experiment and observation data from top to bottom.

ambiguity, "complex model" can be replaced as "optimization problem", "global-level model","global model" can be replaced as "global surrogate model". We will revise these terms.

5. L 213: Why do you use reanalysis data? Why not GPCP for instance?

Both ERA5 and GPCP can be used for precipitation analysis. They both provide global precipitation data. However ERA5 provides higher-resolution data. Data of GPCP are provided on a 2.5 degree grid and ERA5 precipitation data are provided on a 0.25 degree grid. We believe that choosing data with higher resolution can significantly contrast the tuning results, thereby demonstrating the effectiveness of the proposed method. So that we select ERA5 instead of GPCP as the metric for precipitation parameter tuning. The RMSE is calculated between the CAM5 simulation results and ERA5 reanalysis data.

6. Equation 1: Comparing CAM simulations with reanalysis requires regridding data. What technique do you use for that?

The CAM5 case we used in this paper is spectral element dynamical core (SE-dycore) formulation. It need be regridded to compare with reanalysis data. We know that the ERA5 reanalysis data is lat/lon grid. We regrid the simulation result to lat/lon grid. NCL (The NCAR Command Language) is a effective tool to handle CESM simulation results data. So that we use NCL to regrid CAM simulation data to lat/lon grid. Regridding function "ESMF_regrid" is select to complete the regrid operation and "bilinear" is used for the regridding interpolation method as the function input parameter.

7. L. 249: What is the level of the fitness function? Equations for $V\hat{s}$ and $V\hat{d}$ would be very helpful!

"The level of the fitness function" means the quality of the samples, in this paper, it means the fitness values (RMSE values) of these samples. The process of CAND is shown in Section 1 question 1. It can be seen that V^s represents the exploitation mechanism, it describes the simulation result of the sample Ω obtained by the current surrogate model. V^d represents the exploration mechanism, it expresses the exploration for the unknown region of the current surrogate model. V^s and V^d is used to balance exploitation and exploration. We will add the description of CAND to the revised manuscript.

8. L. 267: "..., allowing us to estimate uncertainty from the weight parameter": How do you estimate uncertainty of the weight parameter?

Perhaps there is some ambiguity in our expression, in this sentence, the weight parameter represents V^D in CAND strategy, we use the parameter V^D to represent the exploration mechanism, which means searching a new optimum in uncharted regions.

The sentence does not means that we want to estimate the uncertainty of the parameters on the CAM5 simulation results.

Maybe these terms like "uncertainty", "weight parameter" have other meanings in ESM parameter tuning. We revise the sentence as follows:

"The set of the previously sampled points denote the region which has been explored, allowing us to estimate the uncertainty from the distance between the generated point sets to the explored region."

9. Ll. 222-225: The two sentences appear to have very similar information and should be rewritten

Thanks for your comment. The two sentences both describe that add new samples to update surrogate model. We delete the sentences with similar meanings. The new sentences are as follows:

"Generally, when solving a complex parameter optimization problem by a surrogate model, to improve the accuracy simulation results of the surrogate model, additional new sample points need to be added. Thus reducing the number of simulations of the actual complex model."

We will add these sentences in revised manuscript.

10. Figure 1: Could you put the whole algorithm into the flowchart which indicates which technique is used at which step which would make the whole description much clearer.

Thanks for your comment, We will update the flowchart according to your comment, in each step we will add which technique is used. The new flowchart are shown in Figure 3.

We believe that the new flowchart is much clearer. Compared with the old flowchart, we add the whole process of CAND and trust region method. In order to show the integration of these process, dotted bordered rectangle is used to mark the method which the current process belongs to. In addition, we add the surrogate construction method of each level in the new flowchart.

We will add the new figure in revised manuscript.

11. Figure 3: Why is the relative error of the proposed method increasing with iterations? Shouldn't the surrogate model get more accurate with iteration numbers?

Ideally, the error will gradually decrease, but there will be some inevitable small oscillations during the optimization process, resulting in a slight increase in error.

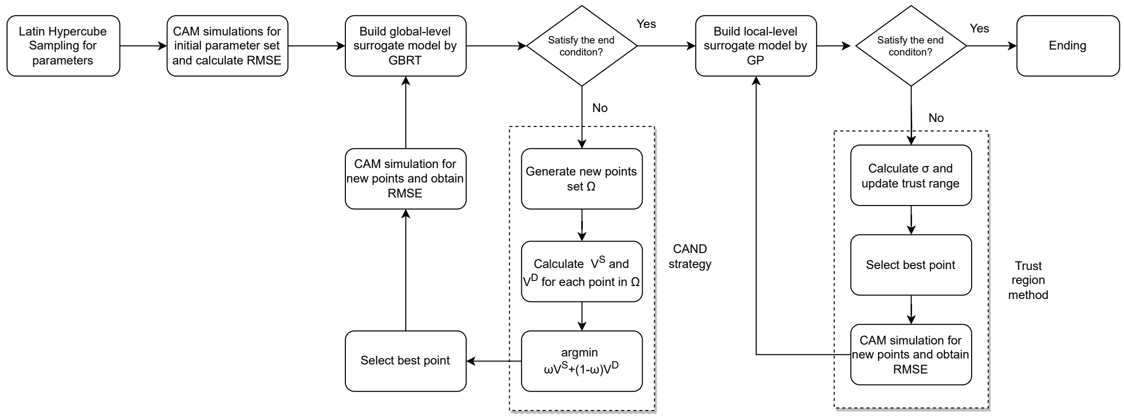

Figure 3: New flowchart

it's possible that our method may indeed have slightly higher errors compared to ASMO in the end. However, our method demonstrates greater stability throughout the entire optimization process, with errors consistently maintained at a lower level. In contrast, AMSO exhibits initial oscillations in errors, indicating that our surrogate model remains stable. While our final error may be slightly higher than that of ASMO, we believe that in cases where the errors are relatively close, the reduction in the number of optimization iterations is a highlight of our method.

12. L 256: "…we only run the real model once, and": But don't you add more than one sample in each iteration? So, how do you need ot run the model only once?

We run the real model (CAM) once in each iteration step, and add the parameters and corresponding RMSE value to sampling set. For example, if the whole tuning process contains 30 steps, in each step one pair of parameters and RMSE value are added to sampling set. There are 30 pairs added to sampling set. We can rewrite the sentence as:

"For the whole tuning process, we only run the CAM5 model once in each iteration step, and all of the samples are predicted using the surrogate model."

We believe that new sentences will express their meanings more clearly, and we will add them in revised manuscript.

13. LL. 377-384: "I think it is important to understand how each parameter influences the model simulations. Why did you pick only one here? This section requires more careful exploration of the parameter itself and the physical mechanisms. What are the physical reasons for the positive and negative correlations described in ll. 377-379

We agree your comment that it is important to understand how each parameter influences the model simulations. The motivation of this work is to propose a tuning method and how to use the mothod for CAM5 precipitation tuning. We analyse the parameter rhminl to prove the "rocker effect": The values of parameters will have different impacts in different regions. Thus introduce a more appropriate way to use the surrogate-based method and the nonuniform parameter parameterization scheme. We are willing to study how each parameter influences the model simulations in future research.

14. Table3/4: Could you also discuss the RMSE of those regions from the optimized parameter set

Table 1: The CAM simulation performance increase for each region.

Region	Default RMSE	Global surrogate RMSE	Optimized RMSE	Reduction Rate
WarmPool	1.985	1.961	1.924	3.07%
South Pacific	0.855	0.788	0.455	46.78%
Niño	0.931	0.855	0.773	17.04%
South America	2.576	2.459	2.371	7.94%
South Asia	1.484	1.352	1.293	12.87%
East Asia	1.213	1.043	0.878	27.68%

from the global-surrogate model? This would clarify what the gain is from the local-level surrogate models.

We agree your comment, discuss the improve from different level of surrogate model will better demonstrate the effectiveness of our method. We will add the tuning result obtained from global-level surrogate. The new table are shown in Table 2

The "Global surrogate RMSE" means the optimization result obtained by global-level surrogate and "Optimized RMSE" represents the final result.

Table 4 in manuscript represents the results of nonuniform parameter parameterization scheme, which does not involve any level surrogate model or new tuning process.

3 Replies to minor comments

Replies to minor comments are as follows:

1. L. 1: "The uncertainty of physical parameters is a major reason for a poor precipitation simulation performance in Earth system models (ESMs), especially over the tropical and Pacific regions.": Is it not only uncertainty of physical parameters but also the microphysics parameterizations itself.

We agree your comment, microphysics parameterization is also one of the main factors influencing the precipitation. Our wording might be too absolute. We should use words like "one of the major reasons".

2. Ll. 14-15: "The results show that the surrogate model-based optimization method can significantly improve the simulation performance of the CAM model.": I would rephrase it stating that the surrogate model-based optimization method allows for better identifying optimal parameter values.

We agree your comment, we will revise the sentence according to your comment.

3. L. 25: "...could lead to huge deviations in the simulations": Deviations from what?

We know that some parameters are sensitive for the simulation result value (eg. temperature, precipitation). A slight change in these parameter values can lead to significant numerical variations in the simulation results. Perhaps our choice of words was not precise. "Error" might be a more accurate term in this sentence.

4. LL. 168-169: "The strategy leverages the information and knowledge obtained from the surrogate model to optimize the run time of the real complex model to fulfill the requirement of accuracy.": How do you optimize for the run time of the real complex model (I guess the ESM)?

The real complex model is the ESM in this paper, we will revise these terms to avoid ambiguity. The appropriate strategy can reduce the number of iteration so as to reduce run time of the real complex model. Perhaps our choice of words was not precise. We will use more accurate term in this sentence.

5. L 171: ".. to update the global-level surrogate model until global-model convergence.": Which global-model do you refer to here, which global-model has to converge?

The global model is the global-level surrogate model. Only when the global-level surrogate model converges, the local-level surrogate model will be created. We will revise these terms to avoid ambiguity.

6. L.172: "... high-waulity CAM" Do you mwan with high-waulity simulations closer to the target value?

Yes, "the high-quality results" means the simulations closer to the target value, in this paper, is the reanalysis data.

7. L. 175-176: "In the parameter tuning process, each surrogate model can fully explore the parameter space to obtain better solutions, generating a large number of samples.": I don't understand

this sentence as earlier (ll 171-172) it is stated that local-level surrogates don't use the whole parameter space?

"Parameter space" means the range of all the parameters, local-level surrogate don't use the whole samples, only high quality samples selected. However the parameter searching in tuning process of local-level surrogate is also over the range of all the parameters (parameter space).

8. LL. 202-211: Why do you talk about 1-D LHS. Usually LHS code can handle several dimensions. LHS code usually makes sure to maximize the minimal distance between all vectors in order to sample the whole space as uniformly as possible.

We agree your comment that LHS code can handle several dimensions. We use a LHS to generate samples, there are 6 parameters selected in this paper. Introduction of 1-D LHS is a example to describe the process of LHS. We use 6-d LHS in this paper.

9. Ll. 222-225: The two sentences appear to have very similar information and should be rewritten.

We will rewrite these sentences in revised manuscript.

10. ll. : 347: Do you compare here the RMSEs of the final optimized parameter set? If so, do they converge to the same parameter set or different ones?

Yes, we compare RMSE of the simulation results of different parameters. However, they converge to different parameter values.

11. Figure 4: It would be good to see what you are tuning for. Can you also add the target value and not only the default and the biases to the target?

We will add these figures according to your comment.

12. L 370: What do you mean with influence mode?

The "influence mode" means the impact of changes in parameter values on the precipitation of each grid, positive correlation or negative correlation.

4 Replies for technical issues

Replies for technical issues are as follows:

1. Labels on contour plots are generally very small.

We will revise these plots according to this comment.

2. L. 63: I might have missed it but "ANNs" acronym was not introduced.

Acronyms will be properly defined in our revised manuscript.

3. L. 78: I might have missed it but "SCA-SMA" acronym was not introduced.

Acronyms will be properly defined in our revised manuscript.

4. L.132: "The compset used in this study is F_2000_CAM5, and the resolution is ne30_g16": To the normal reader these abbreviations don't mean anything. A little bit more explanation would be nice. What is F_2000_CAM5 for instance or what does ne30_g16 mean in the physical world?

We will add more description about CAM5 and the compset used in this study in revised manuscript.

5. Can you use maybe mathematical notation of the original parameters in the paraemeterization instead of the CAM5 parameter naming? For instance line. 148 zmconv_tau is simply tau in the original Zhang McFarlane paper.

We will use mathematical notation of the parameters in revised manuscript.

6. L 275: "actual situation": you probably mean behaviour or something like that.

We agree your comment, it may be more accurate if we use "behaviour" to replace to the "situation".

7. Figure 2: y-axis label missing

We will correct it in revised manuscript.

8. L. 333: What is now X,Y in the S{X,Y} notation?

{X,Y} represents the pairs of parameters and corresponding RMSE value. X represents the parameter combinations of selected 6 parameters and Y represents the RMSE values obtained from CAM simulation results of corresponding parameters.

9. Figure 3: no x-axis label

We will correct it in revised manuscript.

10. L. 398: "precipitation change trend": What trend do you mean here? Time trend?

The "trend" means that South Pacific region generally reveals a ladder-like decline from east to west, which is described in last sentence.

References

[1] X. Wu, L. Hu, L. Wang, H. Lu, and J. Zheng. Surrogate model-based precipitation tuning for cam5. *Geoscientific Model Development Discussions*, 2023:1–30, 2023.

[2] Yun Qian, Huiping Yan, Zhangshuan Hou, Gardar Johannesson, Stephen Klein, Donald Lucas, Richard Neale, Philip Rasch, Laura Swiler, John Tannahill, Hailong Wang, Minghuai Wang, and Chun Zhao. Parametric sensitivity analysis of precipitation at global and local scales in the Community Atmosphere Model CAM5. *Journal of Advances in Modeling Earth Systems*, 7(2):382–411, June 2015.

[3] Reza Alizadeh, Janet K Allen, and Farrokh Mistree. Managing computational complexity using surrogate models: a critical review. *Research in Engineering Design*, 31:275–298, 2020.

[4] Bianca Williams and Selen Cremaschi. Selection of surrogate modeling techniques for surface approximation and surrogate-based optimization. *Chemical Engineering Research and Design*, 170:76–89, 2021.

[5] Frédéric Hourdin, Daniel Williamson, Catherine Rio, Fleur Couvreux, Romain Roehrig, Najda Villefranque, Ionela Musat, Laurent Fairhead, F Binta Diallo, and Victoria Volodina. Process-based climate model development harnessing machine learning: Ii. model calibration from single column to global. *Journal of Advances in Modeling Earth Systems*, 13(6):e2020MS002225, 2021.

[6] Fleur Couvreux, Frédéric Hourdin, Daniel Williamson, Romain Roehrig, Victoria Volodina, Najda Villefranque, Catherine Rio, Olivier Audouin, James Salter, Eric Bazile, et al. Process-based climate model development harnessing machine learning: I. a calibration tool for parameterization improvement. *Journal of Advances in Modeling Earth Systems*, 13(3):e2020MS002217, 2021.

[7] Daniel Williamson, Adam T Blaker, Charlotte Hampton, and James Salter. Identifying and removing structural biases in climate models with history matching. *Climate dynamics*, 45:1299–1324, 2015.

[8] Daniel Williamson, Michael Goldstein, Lesley Allison, Adam Blaker, Peter Challenor, Laura Jackson, and Kuniko Yamazaki. History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble. *Climate dynamics*, 41:1703–1729, 2013.

[9] Raju Pathak, Sandeep Sahany, and Saroj K. Mishra. Uncertainty quantification based cloud parameterization sensitivity analysis in the NCAR community atmosphere model. *Scientific Reports*, 10(1):17499, December 2020.

[10] Ben Yang, Yun Qian, Guang Lin, L. Ruby Leung, Philip J. Rasch, Guang J. Zhang, Sally A. McFarlane, Chun Zhao, Yaocun Zhang, Hailong Wang, Minghuai Wang, and Xiaohong Liu. 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*, 118(2):395–415, 2013. _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2012JD018213.