https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

Calibrating the GAMIL3-1 °climate model using a derivative-free optimization method

Wenjun Liang[1,2], Simon Frederick Barnard Tett[3], Lijuan Li[4], Coralia Cartis[5], Danya Xu[6], Wenjie Dong[1,2*]

[1] School of Atmospheric Sciences, Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Sun Yat-sen University, Zhuhai, 519082, China
[2] Key Laboratory of Tropical Atmosphere-Ocean System, Ministry of Education, Zhuhai, 519082, China
[3] School of Geosciences, University of Edinburgh, Edinburgh, United Kingdom
[4] Key Laboratory of Earth System Numerical Modeling and Application, Chinese Academy of Sciences, Beijing, China
[5] Mathematical Institute, University of Oxford, Oxford, United Kingdom
[6] Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519082, China

Correspondence to: Wenjie Dong (dongwj3@mail.sysu.edu.cn)

Abstract. Parameterization in climate models often involves parameters that are poorly constrained by observations or theoretical understanding alone. Manual tuning by experts can be time-consuming, subjective, and prone to underestimating uncertainties. Automated tuning methods offer a promising alternative, enabling faster, objective improvements in model performance and better uncertainty quantification. This study presents an automated parameter-tuning framework that employs a derivative-free optimization solver (DFO-LS) to simultaneously perturb and tune multiple convection-related and microphysics parameters. The framework explicitly accounts for observational and initial condition uncertainties (internal variability) to calibrate a 1-degree resolution atmospheric model (GAMIL3). Two experiments, adjusting 10 and 20 parameters, were conducted alongside three sensitivity experiments that varied initial parameter values for a 10-parameter case. Both of the first two experiments showed a rapid decrease in the cost function, with the 10-parameter optimization significantly improving model accuracy in 24 out of 34 variables. Expanding to 20 parameters further enhanced accuracy, with improvement in 25 of 34

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

variables, though some structural model errors emerged. Ten-year AMIP simulations validated the robustness and stability of the tuning results, showing that the improvements persisted over extended simulations. Additionally, evaluations of the coupled model with optimized parameters showed--compare to the default parameter setting--reduced climate drift, a more stable climate system, and more realistic sea surface temperatures, despite a slight energy imbalance and some regional biases. The sensitivity experiments underscored the efficiency of the tuning algorithm and highlight the importance of expert judgment in selecting initial parameter values. This tuning framework is broadly applicable to other general circulation models (GCMs), supporting comprehensive parameter tuning and advancing model development.

1 Introduction

Assessing current and future climate change risks to natural and human systems heavily relies on numerical simulations using advanced climate or Earth System Models (ESMs). In recent decades, significant progress has been made in developing the major components of the Earth system (i.e., atmosphere, ocean, land, human systems, etc.) and in the coupling techniques required to form fully integrated ESMs. However, many unresolved issues remain in the development of ESMs, including but not limited to simulation bias in air-sea interactions (Ham et al., 2014; Bellucci et al., 2021; Wei et al., 2021; Meng et al., 2022), the double Intertropical Convergence Zone (ITCZ) problem (Tian et al., 2020), and the coupling of biogeochemical cycles such as the carbon cycle, nutrient cycles with the physical climate system (Erickson et al., 2008). The complexity of the Earth's climate system and the inherent uncertainties in climate models present significant challenges in achieving reliable projections. One of the key sources of uncertainty arises from the representation of unresolved physical processes through parameterizations (Gentine et al., 2021; Jebeile et al., 2023).

Parameterizations are crucial when accounting for processes that occur at unresolved scales or are missing from the model formulation. Parameterizations provide simplified representations of sub-grid processes like cloud convection and turbulence, which cannot be explicitly resolved at scales smaller than the model's grid

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

resolution due to computational constraints. For example, processes such as atmospheric radiative transfer and cloud microphysics are too complex to be represented in full detail within ESMs, so parameterizations offer simplified approximations to capture their essential effects. Parameterization often involves parameters whose values are frequently not well-constrained by either observations or theory alone (Ludovic, 2021; Jeliele et al., 2023), which can directly affect the performance of the model simulation. Consequently, parameter tuning, the process of estimating these uncertain parameters to minimize the discrepancy between specific observations and model results, becomes a critical step in climate model development (Hourdin et al., 2017).

Appropriate parameter tuning can improve the accuracy and skill of climate model outputs by optimizing parameter values to better match observations or high-resolution simulations used as calibration targets (Mauritsen et al., 2012; Bhouri et al., 2023). For example, parameter tuning allows adjusting the values of parameters in parameterizations that approximate these unresolved processes like cloud convection, turbulence, etc (Golaz et al., 2013; Zou et al., 2014; Mignot et al., 2021; Xie et al., 2023). By tuning parameter values during the model calibration process, modelers can partly compensate for known structural errors, deficiencies, or missing processes in the underlying model formulation itself (Williamson et al., 2015; Hourdin et al., 2017; Tett et al., 2017; Schneider et al., 2024). What's more, exploring the range of plausible parameter values through tuning allows quantifying parametric uncertainties and their impacts on model outputs and projections (Jacksonet al., 2004; Neelin et al, 2010; Williamson et al., 2013; Tett et al., 2013; Qian et al., 2016).

Broadly speaking, parameter tuning methods aim to quickly optimize a cost function that measures the distance between model simulations and a small collection of observations. Applications of such methods in climate science include studies by Bellprat et al. (2012), Tett et al. (2013), Yang et al. (2013), Zou et al. (2014), Zhang et al. (2015), and Tett et al. (2017). For instance, in the experiments conducted by Tett et al. (2017) with an atmospheric GCM, 7 and 14 parameters were estimated using variants of the Gauss-Newton algorithm (Tett et al., 2013) to minimize the difference

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

between simulated and observed large-scale, multi-year averaged net radiative fluxes. These optimized parameters were then applied in a coupled GCM. Zhang et al. (2015) utilized an improved downhill simplex method, focusing on seven parameters, and reported successful optimization of an atmospheric model. This improved method overcomes the limitations of the traditional downhill simplex method and offers better computational efficiency compared to evolutionary optimization algorithms.

Traditionally, uncertain parameters have been tuned manually through extensive comparisons of model simulations with available observations. This approach is subjective, labor-intensive, computationally expensive, and can lead to under-exploration of the parameter space, potentially underestimating uncertainties and leaving model biases unresolved (Allen et al., 2000; Hakkarainen et al., 2012; Hourdin et al. 2017; Hourdin et al., 2023). By contrast, automatic and objective parameter calibration techniques have advanced rapidly due to their efficiency, effectiveness, and wider applicability (Chen et al., 1999; Elkinton et al., 2008; Bardenet et al., 2013; Zhang et al., 2015). Bardenet et al. (2013) combined surrogate-based ranking and optimization techniques for surrogate-based collaborative tuning, proposing a generic method to incorporate knowledge from previous experiments. This approach can effectively improve upon manual hyperparameter tuning. Zhang et al. (2015) proposed a "three-step" methodology for parameters tuning. Before the final step of applying the downhill simplex method, they introduced two preliminary steps: determining the model's sensitivity to the parameters and selecting the optimum initial values for those sensitive parameters. By following this process, they were able to automatically and effectively obtain the optimal combination of key parameters in cloud and convective parameterizations.

However, previous studies were either semi-automatic or lacked sufficient observational constraints, such as the net flux at the top of the atmosphere (TOA). Moreover, earlier objective tuning methods that relied on cost functions often overlooked key sources of uncertainty, including observational uncertainty and the internal variability of variables. To address these limitations, we developed a new objective and automatic parameter tuning framework that is more efficient for tuning

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

parameters in GCMs. Compared to previous automatic tuning efforts, this system operates entirely within a Python environment and includes several new optimization algorithms, including Gauss-Newton (Burke et al., 1995; Kim et al., 2008; Tett et al., 2017), the Python Surrogate Optimization Toolbox (pySOT; Regis and Shoemaker, 2012), and the Derivative-Free Optimizer for Least-Squares (DFO-LS; Cartis et al., 2019; Hough et al., 2022). The DFO-LS package is designed to find local solutions to nonlinear least-squares minimization problems without requiring derivatives of the objective function, and has been numerically tested to be particularly effective in finding global optimization solutions. Our framework supports multiple observations and constraints as optimization targets. Additionally, it considers the internal variability of GCMs and integrates sensitivity analysis with the optimization process, making it a more flexible and efficient model tuning system overall. Moreover, systematically and simultaneously perturbing multiple parameters addresses the concern that optimizing a single objective may lead to suboptimal solutions for other objectives and might overlook the global optimum for the overall tuning metric (Qian et al., 2015; Williamson et al., 2015). We have designed and implemented an automatic workflow to streamline the calibration process, enhancing efficiency. This method and workflow are readily applicable to GCMs, facilitating accelerated model development processes. Using this framework, we tune the latest released version 3 of the Grid-Point Atmospheric Model developed at the State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics (LASG) in the Institute of Atmospheric Physics (IAP), named GAMIL3. The newly released GAMIL3 has a higher resolution (~1°) compared to the previous version 2 (~2.8°), and several parameterization schemes related to cloud processes and microphysics have been updated but not well-tuned. This study demonstrates how the tuning framework can automatically and effectively optimize model parameters to achieve better performance against observations.

Our objectives are as follows:

1. To assess the performance of the tuning algorithm in the GAMIL3 atmospheric model;

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

2. To investigate the impact of various parameters and initial values on the tuning results;

3. To evaluate the performance of the optimized parameters in decadal simulations and long-term coupled model runs.

The paper is organized as follows: Section 2 introduces the proposed automatic framework, the tuning model and experiments, observational data and metrics, and the tuning algorithm. Section 3 presents the evaluation of the tuning results in short- to long-tern simulations, including coupled model runs. This is followed by a discussion in Section 4 and a conclusion in Section 5.

2 Methods

2.1 The automatic tuning framework

Here we present the automatic tuning framework (Fig. 1) we have developed, which includes, but is not limited to, functions such as model compiling, (re)submitting, parameter tuning, results evaluation, and diagnostics. Specifically, the framework comprises three main processing modules that collectively control the entire system: the model preprocessing module (the lower left panel in Fig. 1), the model optimizing module (the middle panel in Fig. 1), and the model post-processing module (the right panel in Fig. 1).

The preprocessing module prepares various input data for the optimization process, with particular focus on model internal variations and observational uncertainties (Tett et al., 2017), which will be further discussed in a later section. The optimizing module, which uses the DFO-LS optimization method, is the core component of this tuning system and is primarily responsible for updating model parameters and running simulations. In the initialization of DFO-LS, the module defines the initial parameters, the initial trust region (which is an algorithm parameter) for these parameters, and any parameter constraints. In this step, the system first automatically conducts perturbed parameter experiments for each parameter individually, resulting in N simulations when N parameters are expected to be tuned. The results from this step can be used to assess the sensitivity of the variables to variations in different parameters. Next, the iteration

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

process of DFO-LS begins, involving several steps essential for determining the optimal parameter values. In each iteration, the algorithm refines the parameter estimates and continues until the termination criteria are met, resulting in optimized parameters. The post-processing module receives the output from the optimization module, including the optimized parameters, the sensitivity of variables to the parameters, and the cost function values from different iterations, and further analyzes these results based on user requirements.

2.2 Model description and experiments

In this study, we utilize the latest version 3 of the Grid-point Atmospheric Model developed at the Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China (IAP LASG GAMIL3). This version represents a significant advancement over its predecessor, GAMIL2 (Li et al., 2013), by introducing a hybrid 2D decomposition that enhances parallel scalability, replacing the one-dimensional parallel decomposition in the meridional direction used in GAMIL2 (Liu et al., 2014). GAMIL3 features a grid structure of 360 × 160 longitude-latitude cells, providing a horizontal resolution of approximately 1°, with 26 vertical σ-layers (pressure normalized by surface pressure) extending to the model top at 2.19 hPa. Significant updates in GAMIL3 include improvements in the two-step shape-preserving advection scheme (TSPAS; Yu, 1994) compared to GAMIL2 (Li et al., 2013), the inclusion of a convective momentum transport scheme (Wu et al., 2007), updates to the planetary boundary layer scheme, and enhancements in the stratocumulus cloud-fraction scheme based on turbulence kinetic energy and estimated inversion strength (Guo and Zhou, 2014; Sun et al., 2016). Additionally, GAMIL3 integrates several parameterizations recommended by CMIP6 to represent anthropogenic aerosol effects (Stevens et al., 2017; Shi et al., 2019).

During optimization, each model simulation is performed for 15 months, forced by observed sea-surface temperature (SST) and sea ice, in an Atmospheric Model Intercomparison Project (AMIP) experiment (Eyring et al., 2016). The period runs from 1 October 2010 to 31 December 2011, with the first 3 months excluded for model spin-

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

up, leaving 12 months for analysis against observations. This method is commonly used for model uncertainty quantification and parameter tuning (Yang et al., 2013; Xie et al., 2023). After optimization, the parameter set that best fits the observations is referred to as the optimized parameter set. We use this to conduct a 10-year AMIP simulation from January 1, 2005, to December 31, 2014, enabling comparison with observed climate data. Additionally, to assess whether tuning atmospheric parameters results in a reasonable coupled model, the GAMIL3 atmospheric model is coupled with land, ocean, and sea ice models, and a 30-year piControl simulation is conducted (Eyring et al., 2016). Lastly, three additional sensitivity experiments, varying the initial values of the first 10 parameters, are carried out to examine the impact of initial parameter selection on the optimized results.

2.3 Observations and parameter selection

To set up our optimization problem, we focus on the large-scale performance of the model and consider the differences between land and ocean, particularly in the tropical region. This region is characterized by distinct air-sea interactions, such as those over the Western Pacific warm pool (Wyrtki, 1975), the Eastern Pacific equatorial cold tongue region (Philander, 1983), and the Indian Ocean Dipole region (Saji et al., 1999). Therefore, following the methods outlined by Tett et al. (2017), we separate the analysis into four regions: the northern hemispheric extra-tropical region (θ > 30° N), the tropical region (30° S $\geq \theta \leq$ 30° N), subdivided into tropical land and ocean, and the southern hemispheric extra-tropical region (θ < 30 °S).

The observational variables used in this study are detailed in Table 1. While most variables are divided into four regions—labeled _TROPICSLAND, _TROPICSOCEAN, _NHX, and _SHX—each with its own target and uncertainty, NETFLUX is averaged over all regions and serves as a global constraint. Specifically, the target values for variables T500, RH500, and MSLP are derived from ECMWF Reanalysis v5 data (ERA5; Hersbach et al., 2020); the radiation variables (OLR, OLRC, RSR, RSRC, and NETFLUX) are sourced from Clouds and the Earth's Radiant Energy System (CERES; Wielicki et al., 1998); and the LAT and PRECIP data come from the

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

Climatic Research Unit (CRU; Jones et al., 2012; Harris et al., 2017). The uncertainties of the variables are derived from the absolute error among different data sources, which will be discussed in a later section. All targets and uncertainties of the variables in Table 1 are for the year 2011, primarily used for model optimization.

The atmospheric model parameters we calibrated are detailed in Table 2, encompassing selections from deep convection, shallow convection, microphysics, cloud fraction, and turbulence schemes. The selection of these parameters, along with their default values and plausible ranges, is based on expert judgment as recommended by the GAMIL3 developers and corresponds to the model configuration used in CMIP6 experiments. For visualization, all parameters are normalized based on their plausible ranges, with 0 representing the minimum value of the range and 1 representing the maximum one. Then two experiments are conducted to assess the impacts of varying the number of parameters on the optimized results:

1. We selected the first 10 parameters (listed in the last column of Table 2) from deep convection, shallow convection, microphysics, and cloud fraction schemes. These parameters are identified as the most sensitive to the model's performance based on Xie et al. (2023), and are therefore chosen for tuning.
2. An additional set of the next 10 parameters (also listed in the last column of Table 2), related to microphysics and turbulence schemes, is included alongside the initial 10 parameters. This approach aims to explore the impact of varying the number of tuning parameters on the optimization results.

2.4 Covariance matrices for observations and model

Two covariance matrices need to be prepared before the optimization process begins. The first matrix assesses the internal variability of the model system (C_i). To derive this, perturbed initial condition experiments are conducted. In this study, these experiments involve running a total of 20 simulations, each with the three-dimensional atmospheric temperature initial state perturbed by increments of +1e-20, while all other settings remain identical to those used in the optimization. The second matrix estimates the uncertainty of observations (C_0), which is generally diagonal, assuming no

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

correlation between different observations, and its values are derived from the difference between two observation datasets. Specifically, data from ERA5 and National Center for Environmental Predictions/Department of Energy (DOE) 2 Reanalysis dataset (NCEP2; Kanamitsu et al., 2002) are used to derive the observation error for variable T500, RH500, and MSLP. Precipitation data from CRU and Global Precipitation Climatology Project (GPCP; Adler et al., 2003) are used for Land Precipitation (Lprecip). Data from CRU and Berkeley Earth Surface Temperature (BEST; Muller et al., 2013) are used for Land Air Temperature (LAT). For the four radiation variables (OLR, OLRC, RSR, and RSRC), uncertainties are based on results from Loeb et al. (2018), giving that a TOA imbalance range of 0-2 W/m^2 is typical for single-year simulations (Mauritsen et al., 2012), we set the uncertainty of NETFLUX at 2 W/m^2. Both matrices contribute to the total uncertainty in the variables relative to the target observations. The total covariance matrix C is composed of the two uncertainties introduced above, calculated as:

$$C = C_0 + 2C_i \quad (1)$$

During optimization, all observation values are standardized using the square root of the diagonal elements of matrix C.

2.5 Evaluation methods

The cost function $F(p)$ is used to measure the difference between the simulated values S and the target observations O based on the parameters p. The cost function is given by:

$$F^2(p) = \frac{1}{N}(S-O)^T C^{-1}(S-O) \quad (2),$$

where S is the simulated values; O is the target (observed) values; C is the covariance matrix (as discussed above); N is the number of observations; $(S-O)^T$ is the transpose of the difference between simulated and observed values; C^{-1} is the inverse of the covariance matrix; N is the number of tuning parameters. This cost function quantifies how far the simulation is from the observations, considering the uncertainty (through C) and correlation between different observations. The cost function can be modified to include additional constraints, such as the net radiation flux

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

at the TOA, along with global averages for surface air temperature and precipitation.

The Jacobian matrix J is the partial derivatives of the simulated results with respect to the parameters being optimized. For each simulated model output S_i and parameter p_j, the Jacobian element J_{ij} is given by:

$$J_{ij} = \frac{\partial S_i(p)}{\partial p_j} \tag{3}$$

This measures how much a small change in the parameter p_j will affect the simulated model outputs $S_i(p)$, revealing the impact of each parameter on the variables and providing insights into their sensitivity. The Jacobians are normalized by the parameter range and internal variability. Further details about the cost function and the Jacobian are available in Tett et al. (2017).

In order to assess the extent to which the optimization has improved the performance of the simulated values, the ratios (Z) of the difference between the optimized and the default one to the standard error was adopted:

$$Z = \frac{|V_{\mathrm{Default}} - V_{\mathrm{Observation}}| - |V_{\mathrm{Optimized}} - V_{\mathrm{Observation}}|}{Standard\ error} \tag{4}$$

The $V_{\mathrm{Observation}}$ V_{Default} , and $V_{\mathrm{Optimized}}$ represent the observation value, simulated values using the default and optimized parameter sets, respectively. The $Standard\ error$ represents the observation error of the corresponding variables. Improvement is expected for the variable if $Z > 0$, while if $Z < 0$, no improvement is anticipated, and performance may even worsen.

2.6 Optimization algorithm

The challenge of optimizing the model parameters numerically lies in the high computational cost and potential noise associated with model evaluations, making traditional derivative-based optimization methods impractical. There are several optimization algorithms the system provides, such as (derivative-free) Gauss-Newton variants, the pySOT algorithm, and the DFO-LS algorithm. We use the DFO-LS algorithm as it appears to have better performance in model calibration (Oliver et al., 2022, 2024; Tett et al., 2022) relative to other algorithms such as Gauss-Newton (Tett et at., 2017) or CMA-ES (Hansen, 2016). This algorithm is a sophisticated optimization method designed to handle nonlinear least-squares problems without requiring

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

derivative information. This algorithm is particularly useful in scenarios where function evaluations are expensive or noisy. Inspired by the Gauss-Newton method, DFO-LS constructs simplified linear regression models for the residuals, allowing it to make progress with a minimal number of objective evaluations (Cartis et al., 2019).

The underlying algorithmic methodology for the DFO-LS algorithm is detailed in Cartis et al. (2019). Here, we provide a brief overview of the algorithm, with a detailed description of its parameter settings available in Supplementary S1. The optimization problem is defined as minimizing the sum of the squared residuals

$$f(p) := \frac{\sum_{i=1}^{N} r_i(p)^2}{N} \quad (5),$$

where $r(p)$ represents the differences between model outputs and observations; in our case, $r_i(p) \coloneqq C^{\frac{1}{2}}(S_i - O_i)$. DFO-LS approximates the residuals without derivatives by creating a linear regression model at the current iteration. DFO-LS employs a trust region framework for stable optimization, which dynamically adjusts the search region to balance exploration and exploitation. After constructing the regression model, the algorithm solves the trust region subproblem to determine the step size and direction for updating parameters. The actual versus predicted reduction in the cost function is calculated to decide whether to accept or reject the step, with adjustments made to the trust region size accordingly. The algorithm follows these steps: initialization of parameters and trust region, model construction at each iteration, solving the trust region subproblem, accepting or rejecting steps, updating the interpolation set, and checking termination criteria. This structured approach ensures robust and efficient optimization in minimizing model discrepancies.

3 Results

3.1 1-year AMIP simulations

3.1.1 GAMIL3 10-parameter case

The first experiment aims to optimize the ten sensitive parameters related to convection and microphysics parameterization schemes (Table 2). In this experiment, several parameters—including *ke* and *captlmt*—were adjusted significantly from their

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

default values, while *cmftau* and *c0* showed only minor adjustments (Fig. 2a). Fig. 2b shows the progression of the cost function over iterations for the 10- and 20-parameter cases. Note that the cost function is divided by the number of observations, and a smaller cost function indicates better simulation accuracy against observations. In the 10-parameter case, the system reaches its lowest cost function value of approximately 3.5 after 19 iterations, excluding the initial 10 runs. The cost function drops rapidly from about 7.5 to 3.5 after the 10 initial runs, followed by a slower decline with some fluctuations.

Fig. 3 shows the reduction or increase in simulation error in terms of the number of standard errors through optimization. In the 10-parameter case (solid dots), 24 out of 34 variables (approximately 71%) show Z values greater than zero, indicating improved performance against the default case. Moreover, for 11 of these 24 variables, the optimization reduced the error by more than 1 standard error, with 5 of these showing improvements greater than 3. This is particularly evident in the RSR, MSLP, and the tropical variables of T500. While most variables can be effectively tuned, several variables, such as OLR, OLRC, and LAT, are worse than the default case. However, except for LAT_NHX, the performance of these variables did not degrade by more than one standard error. The blue dots in Fig. 4 represent the global area-weighted mean of different variables for the tuning year (2011) in the 10-parameter case. Comparing to the observational values, the optimization successfully improved most variables (9 out of 10), bringing them closer to the observations. Although some variables showed slight deviations from the observations after optimization, nearly all remained within their uncertainty range (except for OLRC), which is also reasonable in model tuning.

Since the cost function is a simple statistical indicator of the distance between the area-weighted mean of the simulations and the observations, analyzing the spatial distribution of the variables is crucial when evaluating the performance of the optimized parameter sets. Fig. 5a presents Taylor diagrams for all tuning variables under three parameter cases for the optimized year (2011). The results indicate that, compared to the default case (green patterns), most variables' performance improved to varying degrees in the 10-parameter case (blue patterns). For instance, while the standard deviation (SD) of the MSLP in the default result was much closer to the observations, the 10-parameter case exhibited a larger pattern correlation (PC) coefficient and a smaller root mean square deviation (RMSD). Some variables, including PRECIP, NETFLUX, and T500, showed improvements in all three metrics (SD, PC, and RMSD). However, other variables, such as OLR and RH500, showed slight deterioration after optimization, as partially suggested in Fig. 3.

The "optimized" parameter set referred to in this study is the set where the cost function reaches its lowest value. However, the robustness of this parameter set, compared to others with similar cost function values, remains to be evaluated. To address this, two additional experiments were conducted (Table S1 and Fig. S1), selecting parameter sets with cost function values closest to the optimized one to evaluate the potential impact of this choice. Table S1 shows that the parameter values for the two sets (Experiment1 and Experiment2), which have cost function values close to the minimum (Optimized), are quite similar, particularly for Experiment1, which has the closest cost function value. The results from the 10-year AMIP simulations show that, while most variables exhibit patterns similar to those of the Optimized set, notable differences are observed in T2M and PRECIP. Overall, although differences in model behavior arise from the choice of the optimized parameter set, these differences are not substantial enough to significantly alter the model's performance.

3.1.2 GAMIL3 20-parameter case

To investigate the impact of different numbers of tuning parameters on optimization and the robustness of the tuning results, an additional 10 parameters related to microphysics and turbulence schemes (Table 2) were included alongside the existing 10 parameters. In the 20-parameter case, the initial perturbations for the original 10 parameters were kept the same as in the 10-parameter case to ensure a fair comparison. Comparing the optimal values of the 20-parameter case with the default values shows that several parameters had large changes. Parameters such as *c0_conv*, *ke*, *capelmt*, *dzmin*, *Dcs*, and *ecr* showed significant deviations from their default values (Fig. 2a). Comparing the two sets of optimal parameters reveals both differences and

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

consistencies. While most parameters, such as *capelmt*, *alfa*, and *rhcrit*, change in the same direction and display similar magnitudes, some parameters, like *ke* and *cmftau*, are adjusted in the opposite direction. These differences may be attributed to the compensating errors within in the model. When examining the tuning procedure (Fig. 2b), it is evident that the cost function dropped rapidly to a value very close to the minimum after the initial 20 runs, similar to the 10-parameter case. The system required a total of 31 runs to reach the lowest cost function, just two more than the 10-parameter case. This suggests that adding ten additional parameters increases the total number of evaluations only marginally, indicating that when optimizing with DFOLS, there is no need to be overly selective about parameter choice. The minimum cost achieved is comparable to that of the 10-parameter case, with fewer additional runs required after the initial phase to reach the minimum. This implies that including more tuning parameters has a small impact on the total cost but enhances tuning efficiency. This improvement can be attributed to the inclusion of additional parameters related to other parameterization schemes, which enhances model tuning and yields more realistic results compared to observations.

Comparing the Z values from the 20-parameter case to those from the 10-parameter case (Fig. 3), we find that 25 out of 34 variables (approximately 74%) have Z values greater than zero, slightly higher than in the 10-parameter case. Among these, 11 variables show improvements of more than 1 standard error, with 6 exhibiting significant improvements of over 3 standard errors (notably in T500 and MSLP), which is also better than the 10-parameter case. While most variables in the 20-parameter case demonstrate equal or greater improvements than in the 10-parameter case, some, like OLR and OLRC, perform worse. The global area-weighted mean of all variables (shown by red dots in Fig. 4) indicates that, except for OLR, RH500 and PRECIP, variables improved compared to the default case. Although RH500 shows a greater deviation from observation, it still falls within the uncertainty range. Significant differences between the 20-parameter and 10-parameter cases are observed in the two radiation variables (OLR and RSR) and the two surface-related variables (T2M and PRECIP). These differences may partly result from certain parameters compensating

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

for each other, which will be discussed later. The Taylor diagram in Fig. 5a shows that most variables have improved compared to the default case. Relative to the 10-parameter case, OLR, RSR, RSRC, MSLP, and PRECIP perform better in the 20-parameter case. However, NETFLUX and T2M perform worse.

3.2 10-year AMIP simulations

Tuning and evaluating the model using only a one-year simulation may introduce uncertainties due to the model's capacity to simulate phenomena with significant interannual variability (Bonnet et al., 2024), such as the El Niño-Southern Oscillation (ENSO). Therefore, a longer simulation with adjusted parameter settings using AMIP drivers is necessary to assess the robustness of the tuning. Thus 10-year simulations from 1 January 2005 to 31 December 2014 are conducted for the default and two optimized parameter sets. Compared to the results from 2011, the 10-year average AMIP results (Fig. 3b) show no significant differences between the two cases, as both exhibit similar changes across most variables. For example, T500 and RSR show much improvement in both cases, while OLR and OLRC perform worse. However, several variables show differences between the two conditions. For instance, while the standardized MSLP_TROPICSOCEAN_DGM improved by over 20 in the 2011 simulation with the 10-parameter case, it deviates from the observation by more than 10 standard errors in the 10-year simulation. Additionally, while the 20-parameter case demonstrated improvement in the 2011 simulation, its performance declined in the 10-year simulation.

The time series of the 10-year AMIP simulations in Fig. 4 show that, for the 10-parameter case, 8 out of 10 variables are either much closer to the observations or very similar (OLR, OLRC, and OSRC) to those in the default case. Only two variable, RH500 and PRECIP, are slightly further from the observations but still within uncertainty. The most striking finding is the improvement of the variables related to the equilibrium of the climate system (RSR and NETFLUX). For the default case, due to the large outgoing shortwave radiation, NETFLUX has an error of about 5 W/m^2. In addition, T500 in the default case is too cold by almost 2K. After optimization, while

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

OLR shows little change, RSR decreased by nearly 5 W/m^2, considerably reducing the model bias and leading to smaller biases in NETFLUX and T500. Furthermore, the results suggest that MSLP, RSRC and OLRC are hard to tune. In the 20-parameter case, compared to the default, all variables—except RH500, OLR, T2M and PRECIP—show either reduced biases or biases that are very close (OLRC and OSRC) to those in the default case. This is less successful than the 10 parameter case, where 8 variables exhibit reduced or similar bias relative to the default. However, T500 and the MSLP—two variables that deviated significantly from the observations in the default and 10-parameter cases—have been further tuned and now align more closely with observation. Both the optimized cases show that OLR and PRECIP perform notably worse than in the default case, with both variables being too low compared to the observations.

Similar to the Taylor diagram of the 1-year AMIP results, the 10-year AMIP simulations (Fig. 5b) also demonstrate varying degrees of improvement across the three metrics for most variables in both optimized cases. For instance, both cases improve all three metrics for PRECIP, NETFLUX, and RSRC compared to the default case, consistent with the 1-year AMIP results. While PRECIP, RSRC, T2M, and NETFLUX in both optimized cases exhibit similar behave to the 1-year AMIP results, MSLP, RH500, and RSR behave differently. Comparing this with Figs. 3 and 4, the results suggest that this tuning yields only minor improvements to the spatial patterns of the variables but primarily reduces their biases relative to observations. Examining zonal averages (Fig. 6) reveals more specific details, particularly the differences between tropical and extra-tropical regions. T500 and RSR have large tropical biases which tuning considerably reduces. In contrast, RH500, OLR, RSRC, and MSLP have larger biases in extra-tropical, especially polar regions. These regional biases may come from uncertainties in complex high-latitude processes, such as sea ice and snow cover feedback mechanisms, which are not well represented in the model (Goosse et al., 2018). Across the three cases, average performance is similar to that found earlier, with T500, RH500, OLR, RSR, T2M, and PRECIP most affected by tuning and most sensitive to parameter changes, while OLRC, RSRC, and MSLP are little impacted by optimizing. Additionally, while changing physical parameters generally affects the entire

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

atmosphere, some variables respond differently in specific regions. For example, RH500 shows a more pronounced response in tropical regions, while land T2M responds more noticeably in the extra-tropics.

3.3 Atmospheric model evaluation

What parameters and processes would affect these model tuning behaviors? Analyzing the Jacobian results derived from the perturbation parameter simulations can provide insights into how and to what extent various parameters impact the variables. As shown in Fig. 7, parameters such as *c0_conv*, *cmftau*, *rhcrit*, *rhminl*, *rhminh*, and *Dcs* significantly affect simulated variables, particularly NETFLUX, Lprecip_TROPICSLAND, RSR_TROPICSOCEAN, OLR_TROPICSOCEAN, and TEMP@500. Notably, most of these parameters have also been adjusted significantly in the 10- and 20-parameter cases compared to the default. *rhcrit* defines the RH threshold for triggering deep convection and is a parameter with a strong influence on RH. Fig. 2a shows that *rhcrit* decreased from the default case, whose value is 0.85, to the 10-paramter case and 20-parameter case, whose values are 0.83 and 0.82, respectively. A lower *rhcrit* significantly promotes deep convection by reducing the triggering threshold, which enhances water vapor transport from the lower to the mid and upper atmospheric layers. This could lead to a drop in RH below troposphere and a rise above it (Fig. 8a). This effect is especially pronounced in the tropics, where deep convection dominates vertical moisture transport (Fig. 4b, 6b, and 8b). Additionally, low RH below troposphere can limit moisture availability, weakening updrafts and reducing overall precipitation (blue line in Fig. 4h). This negative impact on precipitation outweighs the positive effect of increased precipitation efficiency (*c0_conv*; Fig. 7).

A deficit in low-level cloud fraction is evident in Fig. 8c-8d, primary due to the increase in *rhminl* from the default value of 0.95 to 0.97 and 0.96 in the 10- and 20-parameter cases, respectively. Although the 10-parameter case has a higher threshold for low level cloud formation than the 20-parameter case, Fig. 8c-8d shows the opposite result, which can be explained by the compensatory effects of other parameters.

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

Optimized results indicate that *cmftau*, another key parameter, has a lower value in the 20-parameter case (~4284) compared to the default (~4800) and the 10-parameter case (~4931). This decrease in *cmftau* likely strengthens shallow convection while weakening deep convection, reducing upward water transport and RH throughout the troposphere, contributing to the decreased low-level cloud fraction (Xie et al., 2018). Consequently, the lower low-level cloud fraction in the 20-parameter case, compared to the 10-parameter case, reflects the compensatory effects of these key parameters, with the influence of the reduced *cmftau* outweighing that of *rhminl*. High-level clouds trap heat by limiting radiation emission into space, thereby warming the atmosphere, while low-level clouds reflect sunlight, producing a cooling effect. Therefore, a reduction in low-level clouds allows more shortwave radiation to penetrate the lower atmosphere, reducing outgoing shortwave radiation to space (blue lines in Fig. 4e and 6e) and warming the region, including near the surface (blue lines in Fig. 4g and 6a; Fig. 8e).

Comparing the 20-parameter case to the default case, the tuning results show that one sensitive parameter, *Dcs*—the autoconversion size threshold for ice to snow—has been significantly increased. This adjustment suggests that a higher *Dcs* leads to increased RSR and T2M, while also resulting in lower OLR and PRECIP (Fig. 7). Specifically, clouds with higher ice content trap more OLR from the Earth's surface, potentially amplifying the greenhouse effect by retaining more infrared radiation (red lines in Fig. 5c and 7c). This results in a warming effect, particularly at lower atmospheric levels and even near the surface, especially during nighttime or in polar regions (red lines in Fig. 4g, 6a, and 6g; Fig. 8f). Additionally, raising the autoconversion threshold from ice to snow is expected to allow more ice to remain in the atmosphere, directly leading to a reduction in precipitation (red line in Fig. 4h). These effects align with the results shown in Fig. 7, with the exception of RSR. Theoretically, increasing *Dcs* (which delays the conversion from ice to snow) could increase ice mass, raising cloud optical thickness and enhancing the cloud's ability to reflect incoming shortwave radiation. However, this expectation contrasts with the findings in Figs. 4e and 6e, which show a slightly lower RSR in the 20-parameter case

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

compared to the default case. This discrepancy can be attributed to compensatory effect among different parameters. As shown in Fig. 7, changes to the parameters *c0_chg*, *rhminl*, and *cmftau* in the 20-parameter case negatively impact RSR, potentially offsetting the positive effect of *Dcs* and resulting in an RSR slightly lower than that of the default case.

3.4 Coupled model results

In order to evaluate the performance of different parameter sets in long-term climate simulations, it is essential to apply them to a coupled model. Here, the GAMIL3 atmospheric model, coupled with land model (CLM2; Bonan et al., 2002) and ocean and sea ice model (LICOM2.0; Liu et al., 2013), was used to assess whether tuning atmospheric parameters leads to a reasonable coupled model.

In the default case the model starts with a large negative NETFLUX of around -4 W/m ^{2}(Fig. 9a), consistent with the results in Fig. 4j, indicating that the climate system is losing energy at this stage. As the model integrates, the NETFLUX increases, approaching zero after approximately five model years, achieving a stable energy budget for the remaining simulation period. This change in NETFLUX is found to be almost equally driven by a ~2 W/m ^{2}reduction in both RSR (Fig. 9b) and OLR (Fig. 9c) simultaneously. However, despite these radiation variables, particularly the NETFLUX, approaching a stable state, the ocean continues to lose energy rapidly (Fig. 9d) with no signs of stabilization by the end of the simulation. For the T2M, the tuning target is 13.6 $\pm$0.5 ℃ (Williamson et al., 2013), which differs significantly from the model's default case results (Fig. 9e). Consequently, although the NETFLUX appears to reach a stable state, the system continues to lose energy and remains far from the tuning target in the default case. Furthermore, the piControl simulation for the default case is notably fragile and prone to crashes due to unstable iterations, particularly in contrast to the two optimized cases. This instability poses a critical challenge, especially for long-term climate simulations.

For both optimized cases, the NETFLUX (Fig. 9a) remains stable throughout the 30-year simulations, with values of about 2 W/m ^{2}. Although slightly further from the

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

target of 0 W/m^2; they are still within the model spread range of -3 to 4 W/m^2 (Mauritsen et al., 2012). Specifically, the change in NETFLUX in the 10-parameter case is primarily driven by a decrease in RSR (Fig. 9b), while in the 20-parameter case, it is mostly due to a reduction in OLR (Fig. 9c), consistent with the results in Figs. 4c and 4e. Both the volume-averaged ocean temperature (Fig. 9d) and the T2M (Fig. 9e) exhibit a slight initial adjustment during the first five years, followed by stabilization.

Results from the simulated SST anomalies in Fig. 10a–10c for the default case show strong cold anomalies relative to observations, with maximum deviations exceeding -4 ℃ over the North of Pacific and Atlantic. The simulated SST anomalies in Fig. 10d–10i indicate that both optimized cases show substantial improvement over the default case in terms of SST patterns and deviations, although some negative deviations in the northern Pacific and Atlantic persist—a common issue for most GCMs (Zhang and Zhao, 2015; Wang et al., 2018). Previous findings suggest that the two optimized cases exhibit cloud fraction significantly different from the default case, with simulated radiation improvements primarily observed in shortwave and longwave radiation in each case, respectively. Therefore, it is necessary to investigate the shortwave and longwave cloud forcing in these two cases (Fig. 11). The results for both cases show that the combined effect of these two cloud forcings acts as a significant positive influence globally, contributing to the ocean surface flux and increasing ocean temperature. Specifically, the shortwave cloud forcing has a greater weight than the longwave in the 10-parameter case, mainly due to the parameters *rhcrit* and *rhminl*, as mentioned earlier. In contrast, the longwave cloud forcing outweighs the shortwave in the 20-parameter case, primarily due to the effects of *Dcs*. While the shortwave cloud forcing exerts a negative effect over the tropical ocean, the longwave cloud forcing provides a significant compensatory effect. A similar behavior is observed in the 20-parameter case.

Overall, the two optimized cases result in a more realistic coupled model, not only maintaining the model's energy balance and reducing climate drift, but also improving the simulated ocean state, such as SST distribution. Although the two optimized cases exhibit different behaviors—with the 10-parameter case showing lower RSR and the

20-parameter case showing lower OLR—tuning has allowed them to achieve stability through distinct mechanisms.

3.5 Sensitivity of initial parameters

As stated in the previous section, the initial parameter values used for tuning are primarily informed by expert judgment, which has been recognized as crucial and necessary in other studies (Hourdin et al., 2017; Williamson et al., 2017; Jebeile et al., 2023; Lguensat et al., 2023). To further investigate the extent to which initial parameter choices influence tuning results, we conducted three additional sensitivity experiments with randomly selected initial parameter values (Table S2), focusing on the first 10 parameters.

The optimized parameter values in these randomized experiments (represented by stars in Fig. 2a) exhibit significantly larger spreads compared to the default and original optimized values (blue dots), particularly for parameters such as *c0_conv*, *capelmt*, and *c0*, which nearly span their entire plausible ranges. This finding indicates that the model could reach entirely different optimized states depending on initial values. During the tuning process, the cost function (Fig. 2c) for these cases exhibited a rapid decrease, stabilizing at similar values across all three experiments after approximately 10 iterations, with an additional 10–20 runs required to reach the optimized state. This pattern further demonstrates the efficiency and robustness of the tuning algorithm.

Given the substantial differences in the optimized parameters, it is worthwhile to further investigate their Jacobian differences to gain a more comprehensive understanding of each parameter's impact on the variables. Fig. 12 shows the Jacobian ranges for four cases (including the original optimized case), with Jacobian calculated around the optimized parameter set for each case. The results generally demonstrate consistency with the parameter sensitivities shown in Fig. 7. Variables sensitive to most parameters exhibit substantial variability, while highly sensitive parameters, such as *c0_conv*, *cmftau*, *rhcrit*, *rhminl*, and *rhminh*, introduce considerable uncertainty across multiple variables, depending on their initial values and interactions with other parameters. Conversely, RSRC and OLRC remain largely insensitive to parameter changes, whereas MSLP, NETFLUX, Lprecip, and TEM@500hPa are influenced by most parameters, also aligning with the findings in Fig. 7.

The performance of these three optimized parameter sets in the 10-year AMIP simulations is shown in Fig. S2. Generally, NETFLUX was most closely aligned with observations across all cases, primarily due to the additional constraint incorporated into the tuning algorithm. However, notable differences across different cases remain, with each case following a distinct optimization pathway, though most results still fall within uncertainty ranges. For example, the third experiment achieved the closest alignment for T500 but at the expense of T2M and PRECIP compared to other cases, highlighting inherent trade-offs and model structural errors that hinder simultaneous optimization of these variables. As seen in prior findings, RSRC and MSLP proved difficult to tune, while OLRC was adjustable but deviated in the opposite direction from observations, accompanied by a discrepancy in RH500 alignment.

Overall, these sensitivity experiments confirm the efficiency of the tuning algorithm and underscore the importance of expert judgment in selecting initial parameter values. Expert selection not only ensures satisfactory model performance at the start of tuning but also enhances tuning effectiveness, even though structural errors in the model remain.

4 Discussion

In this study, we developed an objective and automatic parameter tuning framework using the Derivative-Free Optimizer for Least-Squares (DFO-LS) method to tune the newest version of the Grid-Point Atmospheric Model (GAMIL3). The results highlight the effectiveness of this method in tuning atmospheric parameters, particularly those initially set based on expert judgment, as demonstrated by notable improvements in model accuracy across multiple variables and enhanced climate system stability. However, several aspects of this work require further clarification.

Firstly, as noted earlier, the 'optimized' parameter set in this study refers to the set at which the cost function achieves its minimum value. However, results in Figs. 2b and 2c indicate that, for each case, there are several cost function values close to this minimum. We have shown that these differences are not substantial enough to significantly alter the model's performance. However, this finding suggests that parameter ranges associated with similar cost function values may provide valuable insights into the acceptable parameter space for model optimization. We acknowledge that focusing exclusively on minimizing cost function values to obtain a single optimized parameter set during tuning can increase the risk of overfitting and compensating errors, which is a common challenge in model tuning. Although the results of this study show no clear signs of overfitting—both the 10- and 20-parameter optimized cases, starting from expert-judged initial values, ultimately produce reasonable coupled model results—it remains important to carefully consider potential overfitting impacts.

Secondly, this study shows that tuning either different numbers of parameters or varying initial parameter values can yield diverse optimized results, each improving certain aspects of the model. This suggests that although tuning can lower the cost function to comparable levels, the final tuned state of the model is not necessarily unique—an common issue encountered in model tuning (Hakkarainen et al., 2013; Hourdin et al., 2017; Eidhammer et al., 2024), likely due to the compensating errors within the model and uncertainties in the observational data. On one hand, introducing constraints, such as assigning greater weight in key variables during tuning, could help achieve more realistic results. For instance, applying constraints on NETFLUX during tuning ensures consistently good performance across all the cases in the 10-year AMIP simulations. In the 20-parameter case, adding constraints on OLR and RSR would maintain their performance while also improving T500 and MSLP. On the other hand, while different parameter sets satisfied the lowest cost function in different ways, it is important to remember that the cost function is simply a statistical measure of the distance between the area-weighted mean of the simulations and observations. Therefore, a comprehensive evaluation is essential to identify the most suitable parameter set (Eidhammer et al., 2024). Beyond minimizing cost function values and aligning statistical indicators with observations, it is crucial to evaluate the spatial distributions of variables, the equilibrium state of the climate system in coupled models,

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

and the model's climate sensitivity (Tett et al., 2022; Eidhammer et al., 2024). These aspects should be further evaluated to ensure robust model performance.

Some limitations remain. For instance, although the coupled model simulations show improvements in energy stability and reduced climate drift, certain regional biases in SST persist. These biases suggest that while tuning enhances model performance, there may be systematic issues within the model's physics that cannot be fully addressed through parameter tuning alone. Resolving these regional discrepancies may require further refinement of model physics or additional modifications to the tuning framework. Additionally, the optimized cases show a relatively large energy imbalance at the TOA. Although still within model uncertainty, this issue warrants further investigation. One possible cause could be the non-conservation of energy in the atmospheric model. Preliminary results indicate that the difference between the TOA and Earth's surface energy imbalances in the 1-year AMIP tuning is approximately 1.4 W/m^2, highlighting one of the model's structural errors. This suggests that even in the optimized cases, the atmospheric model may be consuming excess energy, a bias that could carry over to the coupled model. Consequently, one of the lessons from this study is that when tuning the model, attention should also be paid to structural errors, particularly those related to energy conservation.

5 Conclusions

The study focuses on optimizing an atmospheric model by simultaneously perturbing and tuning multiple parameters associated with convection, microphysics, turbulence, and other physical schemes. Two primary experiments were conducted: one involving the adjustment of 10 parameters, and the other with 20 parameters. In the 10-parameter tuning, significant changes were made to several sensitive parameters, resulting in a notable reduction in the cost function and improved model accuracy. Out of 34 variables, 24 showed improved performance, although some remained challenging to optimize due to structure errors in the model. In the 20-parameter tuning, additional parameters related to microphysics and turbulence were introduced, resulting in slight performance improvements for 25 out 34 variables. However, certain variables

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

experienced a decline in performance. While the 20-parameter case achieved a lower cost function more quickly than the 10-parameter case, the increased complexity required careful management of parameter interactions and compensatory effects.

To evaluate the robustness of the tuning results, we conducted 10-year AMIP simulations. The findings showed that the optimized parameter sets maintained their performance improvements over extended simulation periods, though variables like MSLP exhibited variability depending on the specific period analyzed. Time series analyses indicated that the optimized models more accurately captured the equilibrium of the climate system, particularly by improving the balance of outgoing shortwave and longwave radiation and stabilizing surface temperatures. However, some variables remained challenging to optimize consistently across different regions and timescales. The optimized parameter sets were further tested in a coupled model setup that integrated land, ocean, and sea ice components. The results demonstrated improved energy budget stability, reducing climate drift and leading to more realistic SST simulations. Both the 10- and 20-parameter optimizations yielded more reasonable behavior in the coupled model, though persistent regional biases, particularly in the northern Pacific and Atlantic, remained.

Three additional experiments, in which the initial values of the first 10 parameters were randomly selected, were conducted to evaluate its impact on the optimized results. The results further confirm the efficiency and robustness of the algorithm, as it rapidly minimizes the cost function after the first 10 runs, although the optimized parameter values and their performance across different cases show significant variation. Overall, these findings emphasize the importance of expert judgment in parameter selection and its role in enhancing model performance.

In conclusion, the proposed DFO-LS-based tuning framework presents a robust and efficient approach for enhancing climate model performance. This work was primarily conducted by a researcher over 12 months, highlighting the efficiency of the approach in terms of human resources. The adaptability of this methodology to other GCMs holds great potential for accelerating model development and improving the accuracy and reliability of future climate projections. By integrating this framework into broader model tuning efforts, the climate modeling community can make significant strides in addressing parametric uncertainties and advancing the precision of climate forecasts.

Code availability. All codes are available on Zenodo under the DOI 10.5281/zenodo.14772250, with the citation: "Tett, S. (2025). ModelOptimization. Zenodo. https://doi.org/10.5281/zenodo.14772250"

Author contributions. SFBT conceptualized the study, with WJD sponsoring the computational and necessary resources. SFBT and CC developed the optimization algorithm, while SFBT and WJL implemented the optimization framework under the guidance of LJL and CC. WJL conducted all model experiments and performed the necessary analyses. WJL drafted the manuscript, with all co-authors providing feedback and suggestions.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. WJL is funded by the National Natural Science Foundation of China (U21A6001， 42175173), GuangDong Basic and Applied Basic Research Foundation (2023A1515240036), and Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (No. SML2022006). Work partly carried out while WJL was a visiting student at the UoE supervised by SFBT.

References

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P., & Nelkin, E. (2003). The Version-2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979–Present). *Journal of Hydrometeorology*, *4*(6), 1147–1167. https://doi.org/10.1175/1525-7541(2003)004

Allen, M. R., Stott, P. A., Mitchell, J. F. B., Schnur, R., & Delworth, T. L. (2000). Quantifying the uncertainty in forecasts of anthropogenic climate change. *Nature*, *407*(6804), 617–620. https://doi.org/10.1038/35036559

Bellprat, O., Kotlarski, S., Lüthi, D., & Schär, C. (2012). Objective calibration of regional climate models. *Journal of Geophysical Research Atmospheres*, *117*(D23). https://doi.org/10.1029/2012jd018262

Bellucci, A., Athanasiadis, P. J., Scoccimarro, E., Ruggieri, P., Gualdi, S., Fedele, G., Haarsma,

R. J., Garcia-Serrano, J., Castrillo, M., Putrahasan, D., Sanchez-Gomez, E., Moine, M., Roberts, C. D., Roberts, M. J., Seddon, J., & Vidale, P. L. (2021). Air-Sea interaction over the Gulf Stream in an ensemble of HighResMIP present climate simulations. *Climate Dynamics*, *56*(7–8), 2093–2111. https://doi.org/10.1007/s00382-020-05573-z

Bhouri, M. A., Peng, L., Pritchard, M. S., & Gentine, P. (2023). Multi-fidelity climate model parameterization for better generalization and extrapolation. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2309.10231

Bonan, G. B., Oleson, K. W., Vertenstein, M., Levis, S., Zeng, X., Dai, Y., Dickinson, R. E., & Yang, Z. (2002). The Land Surface Climatology of the Community Land Model Coupled to the NCAR Community Climate Model. *Journal of Climate*, 15, 3123–3149. https://doi.org/10.1175/1520-0442(2002)015<3123:TLSCOT>2.0.CO;2.

Bonnet, P., Pastori, L., Schwabe, M., Giorgetta, M. A., Iglesias-Suarez, F., and Eyring, V.: Tuning a Climate Model with Machine-learning based Emulators and History Matching, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2024-2508, 2024.

Burke, J. V., & Ferris, M. C. (1995). A Gauss—Newton method for convex composite optimization. *Mathematical Programming*, *71*(2), 179–194. https://doi.org/10.1007/bf01585997

Cartis, C., Fiala, J., Marteau, B., & Roberts, L. (2019). Improving the flexibility and robustness of model-based derivative-free optimization solvers. *ACM Transactions on Mathematical Software*, *45*(3), 1–41. https://doi.org/10.1145/3338517

Chen, T., Wei, W., & Tsai, J. (1999). Optimum design of headstocks of precision lathes. *International Journal of Machine Tools and Manufacture*, *39*(12), 1961–1977. https://doi.org/10.1016/s0890-6955(99)00034-6

Eidhammer, T., Gettelman, A., Thayer-Calder, K., Watson-Parris, D., Elsaesser, G., Morrison, H., van Lier-Walqui, M., Song, C., and McCoy, D.: An extensible perturbed parameter ensemble for the Community Atmosphere Model version 6, Geosci. Model Dev., 17, 7835–7853, https://doi.org/10.5194/gmd-17-7835-2024, 2024.

Elkinton, C. N., Manwell, J. F., & McGowan, J. G. (2008). Algorithms for offshore wind farm layout optimization. *Wind Engineering*, *32*(1), 67–84. https://doi.org/10.1260/030952408784305877

Erickson, D., Oglesby, R., Elliott, S., Steffen, W., & Brasseur, G. P. (2008). Chapter Seventeen

Challenges in Earth System Modelling: Approaches and Applications. In *Developments in Integrated Environmental Assessment* (pp. 297–306). https://doi.org/10.1016/s1574-101x(08)00617-0

Eyring, V., Bony, S., Meehl, G. A., A, C., Senior, Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Gentine, P., Eyring, V., & Beucler, T. (2021). Deep Learning for the Parametrization of Subgrid Processes in Climate Models. Chap. 21, Pp. 307–314, John Wiley & Sons, Ltd, Chichester, West Sussex, 307–314. https://doi.org/10.1002/9781119646181.ch21

Golaz, J., Horowitz, L. W., & Levy, H. (2013). Cloud tuning in a coupled climate model: Impact on 20th century warming. *Geophysical Research Letters*, *40*(10), 2246–2251. https://doi.org/10.1002/grl.50232

Goosse, H., Kay, J. E., Armour, K. C., Bodas-Salcedo, A., Chepfer, H., Docquier, D., Jonko, A., Kushner, P. J., Lecomte, O., Massonnet, F., Park, H., Pithan, F., Svensson, G., & Vancoppenolle, M. (2018). Quantifying climate feedbacks in polar regions. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-04173-0

Guo, Z., & Zhou, T. (2014). An improved diagnostic stratocumulus scheme based on estimated inversion strength and its performance in GAMIL2. *Science China Earth Sciences*, *57*(11), 2637–2649. https://doi.org/10.1007/s11430-014-4891-7

Hakkarainen, J., Ilin, A., Solonen, A., Laine, M., Haario, H., Tamminen, J., Oja, E., & Järvinen, H. (2012). On closure parameter estimation in chaotic systems. *Nonlinear Processes in Geophysics*, *19*(1), 127–143. https://doi.org/10.5194/npg-19-127-2012

Hakkarainen, J., Solonen, A., Ilin, A., Susiluoto, J., Laine, M., Haario, H., & Järvinen, H. (2013). dilemma of the uniqueness of weather and climate model closure parameters. *Tellus a Dynamic Meteorology and Oceanography*, *65*(1), 20147. https://doi.org/10.3402/tellusa.v65i0.20147

Ham, S., Hong, S., & Park, S. (2013). A study on air–sea interaction on the simulated seasonal climate in an ocean–atmosphere coupled model. *Climate Dynamics*, *42*(5–6), 1175–1187. https://doi.org/10.1007/s00382-013-1847-0

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

Harris, I. C., Jones, P. D., & Osborn, T. CRU TS4. 01. (2017). Climatic Research Unit (CRU) Time-Series (TS) version 4.01 of high-resolution gridded data of month-by-month variation in climate (Jan. 1901–Dec. 2016). *Centre for Environmental Data Analysis*. 25.

Hansen, N. (2016, April 4). The CMA Evolution Strategy: a tutorial. arXiv.org. http://arxiv.org/abs/1604.00772

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., . . . Thépaut, J. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hough, M., & Roberts, L. (2022). Model-Based Derivative-Free methods for Convex-Constrained optimization. *SIAM Journal on Optimization*, *32*(4), 2552–2579. https://doi.org/10.1137/21m1460971

Hourdin, F., Ferster, B., Deshayes, J., Mignot, J., Musat, I., & Williamson, D. (2023). Toward machine-assisted tuning avoiding the underestimation of uncertainty in climate change projections. *Science Advances*, *9*(29). https://doi.org/10.1126/sciadv.adf2758

Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J., Balaji, V., Duan, Q., Folini, D., Ji, D., Klocke, D., Qian, Y., Rauser, F., Rio, C., Tomassini, L., Watanabe, M., & Williamson, D. (2016). The Art and science of climate model tuning. *Bulletin of the American Meteorological Society*, *98*(3), 589–602. https://doi.org/10.1175/bams-d-15-00135.1

Jackson, C., Sen, M. K., & Stoffa, P. L. (2004). An efficient stochastic Bayesian approach to optimal parameter and uncertainty estimation for climate model predictions. *Journal of Climate*, *17*(14), 2828–2841. https://doi.org/10.1175/1520-0442(2004)017

Jakumeit, J., Herdy, M., & Nitsche, M. (2005). Parameter optimization of the sheet metal forming process using an iterative parallel Kriging algorithm. *Structural and Multidisciplinary Optimization*, *29*(6), 498–507. https://doi.org/10.1007/s00158-004-0455-3

Jebeile, J., Lam, V., Majszak, M., & Räz, T. (2023). Machine learning and the quest for objectivity in climate model parameterization. Climatic Change, 176(8). https://doi.org/10.1007/s10584-023-03532-1

Jones, P. D., Lister, D. H., Osborn, T. J., Harpham, C., Salmon, M., & Morice, C. P. (2012).

Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010. *Journal of Geophysical Research Atmospheres*, *117*(D5). https://doi.org/10.1029/2011jd017139

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S., Hnilo, J. J., Fiorino, M., & Potter, G. L. (2002). NCEP–DOE AMIP-II Reanalysis (R-2). *Bulletin of the American Meteorological Society*, *83*(11), 1631–1644. https://doi.org/10.1175/bams-83-11-1631

Kim, N. D. S., & Lee, N. K. (2008). Block-Coordinate Gauss–Newton optimization and constrained monotone regression for image registration in the presence of outlier objects. *IEEE Transactions on Image Processing*, *17*(5), 798–810. https://doi.org/10.1109/tip.2008.920716

Lguensat, R., Deshayes, J., Durand, H., & Balaji, V. (2023). Semi‐Automatic tuning of coupled climate models with multiple intrinsic timescales: Lessons learned from the Lorenz96 Model. Journal of Advances in Modeling Earth Systems, 15(5). https://doi.org/10.1029/2022ms003367

Li, L., Wang, B., Dong, L., Liu, L., Shen, S., Hu, N., Sun, W., Wang, Y., Huang, W., Shi, X., Pu, Y., & Yang, G. (2013). Evaluation of grid-point atmospheric model of IAP LASG version 2 (GAMIL2). *Advances in Atmospheric Sciences*, *30*(3), 855–867. https://doi.org/10.1007/s00376-013-2157-5

Liu, H., Lin, P., Yu, Y., Wang, F., Liu, X., & Zhang, X. (2013). LASG/IAP Climate System Ocean Model Version 2: LICOM2. In Springer eBooks (pp. 15–26). https://doi.org/10.1007/978-3-642-41801-3_3

Liu, L., Yang, G., Wang, B., Zhang, C., Li, R., Zhang, Z., Ji, Y., & Wang, L. (2014). C-Coupler1: a Chinese community coupler for Earth system modeling. *Geoscientific Model Development*, *7*(5), 2281–2302. https://doi.org/10.5194/gmd-7-2281-2014

Loeb, N. G., Doelling, D. R., Wang, H., Su, W., Nguyen, C., Corbett, J. G., Liang, L., Mitrescu, C., Rose, F. G., & Kato, S. (2017). Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) Top-of-Atmosphere (TOA) Edition-4.0 data product. *Journal of Climate*, *31*(2), 895–918. https://doi.org/10.1175/jcli-d-17-0208.1

Ludovic Touzé-Peiffer. (2021). Parameterization of atmospheric convection in numerical climate models-Practices and epistemological challenges. Meteorology. Sorbonne Université English. NNT :2021SORUS539. tel-04215936

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., Haak, H., Jungclaus, J., Klocke, D., Matei, D., Mikolajewicz, U., Notz, D., Pincus, R., Schmidt, H., & Tomassini, L. (2012). Tuning the climate of a global model. *Journal of Advances in Modeling Earth Systems*, *4*(3). https://doi.org/10.1029/2012ms000154

Meng, Z., Zhou, L., Qin, J., & Li, B. (2022). Intraseasonal Air–Sea Interaction Over the Southeastern Indian Ocean and its Impact on Indian Summer Monsoon. *Frontiers in Marine Science*, *9*. https://doi.org/10.3389/fmars.2022.921585

Mignot, J., Hourdin, F., Deshayes, J., Boucher, O., Gastineau, G., Musat, I., Vancoppenolle, M., Servonnat, J., Caubel, A., Chéruy, F., Denvil, S., Dufresne, J., Ethé, C., Fairhead, L., Foujols, M., Grandpeix, J., Levavasseur, G., Marti, O., Menary, M., . . . Silvy, Y. (2021). The tuning strategy of IPSL-CM6A-LR. *Journal of Advances in Modeling Earth Systems*, *13*(5). https://doi.org/10.1029/2020ms002340

Muller, R. A., Rohde, R., Jacobsen, R., Muller, E., & Wickham, C. (2013). A new estimate of the average Earth surface land temperature spanning 1753 to 2011. *Geoinformatics & Geostatistics an Overview*, *01*(01). https://doi.org/10.4172/2327-4581.1000101

Neelin, J. D., Bracco, A., Luo, H., McWilliams, J. C., & Meyerson, J. E. (2010). Considerations for parameter optimization and sensitivity in climate models. *Proceedings of the National Academy of Sciences*, *107*(50), 21349–21354. https://doi.org/10.1073/pnas.1015473107

Oliver, S., Cartis, C., Kriest, I., Tett, S. F. B., & Khatiwala, S. (2022). A derivative-free optimisation method for global ocean biogeochemical models. Geoscientific Model Development, 15(9), 3537–3554. https://doi.org/10.5194/gmd-15-3537-2022

Oliver, S., Khatiwala, S., Cartis, C., Ward, B., & Kriest, I. (2024). Using shortened Spin‐Ups to speed up ocean biogeochemical model optimization. Journal of Advances in Modeling Earth Systems, 16(9). https://doi.org/10.1029/2023ms003941

Philander, S. G. H. (1983). El Niño Southern Oscillation phenomena. *Nature*, *302*(5906), 295–301. https://doi.org/10.1038/302295a0

Qian, Y., Jackson, C., Giorgi, F., Booth, B., Duan, Q., Forest, C., Higdon, D., Hou, Z. J., & Huerta, G. (2016). Uncertainty quantification in climate modeling and projection. *Bulletin of the American Meteorological Society*, *97*(5), 821–824. https://doi.org/10.1175/bams-d-15-00297.1

Qian, Y., Yan, H., Hou, Z., Johannesson, G., Klein, S., Lucas, D., Neale, R., Rasch, P., Swiler, L., Tannahill, J., Wang, H., Wang, M., & Zhao, C. (2015). 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. https://doi.org/10.1002/2014ms000354

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., & Kaplan, A. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research Atmospheres*, *108*(D14). https://doi.org/10.1029/2002jd002670

Regis, R. G., & Shoemaker, C. A. (2012). Combining radial basis function surrogates and dynamic coordinate search in high-dimensional expensive black-box optimization. *Engineering Optimization*, *45*(5), 529–555. https://doi.org/10.1080/0305215x.2012.687731

Rémi Bardenet, Mátyás Brendel, Balázs Kégl, and Michèle Sebag. (2013). Collaborative hyperparameter tuning. In Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28 (ICML'13). JMLR.org, II–199–II–207. https://dl.acm.org/doi/10.5555/3042817.3042916

Saji, N. H., Goswami, B. N., Vinayachandran, P. N., & Yamagata, T. (1999). A dipole mode in the tropical Indian Ocean. *Nature*, *401*(6751), 360–363. https://doi.org/10.1038/43854

Schneider, T., Leung, L. R., & Wills, R. C. J. (2024). Opinion: Optimizing climate models with process knowledge, resolution, and artificial intelligence. *Atmospheric Chemistry and Physics*, *24*(12), 7041–7062. https://doi.org/10.5194/acp-24-7041-2024

Shi, X., Zhang, W., & Liu, J. (2019). Comparison of Anthropogenic Aerosol Climate Effects among Three Climate Models with Reduced Complexity. *Atmosphere*, *10*(8), 456. https://doi.org/10.3390/atmos10080456

Stevens, B., Fiedler, S., Kinne, S., Peters, K., Rast, S., Müsse, J., Smith, S. J., & Mauritsen, T. (2017). MACv2-SP: a parameterization of anthropogenic aerosol optical properties and an associated Twomey effect for use in CMIP6. *Geoscientific Model Development*, *10*(1), 433–452. https://doi.org/10.5194/gmd-10-433-2017

Sun, W., Li, L., & Wang, B. (2016). Reducing the biases in shortwave cloud radiative forcing in tropical and subtropical regions from the perspective of boundary layer processes. *Science China Earth Sciences*, *59*(7), 1427–1439. https://doi.org/10.1007/s11430-016-5290-z

Tett, S. F. B., Gregory, J. M., Freychet, N., Cartis, C., Mineter, M. J., & Roberts, L. (2022).

Does model calibration reduce uncertainty in climate projections? *Journal of Climate*, *35*(8), 2585–2602. https://doi.org/10.1175/jcli-d-21-0434.1

Tett, S. F. B., Mineter, M. J., Cartis, C., Rowlands, D. J., & Liu, P. (2013b). Can Top-of-Atmosphere radiation measurements constrain climate predictions? Part I: Tuning. *Journal of Climate*, *26*(23), 9348–9366. https://doi.org/10.1175/jcli-d-12-00595.1

Tett, S. F. B., Yamazaki, K., Mineter, M. J., Cartis, C., & Eizenberg, N. (2017). Calibrating climate models using inverse methods: case studies with HadAM3, HadAM3P and HadCM3. *Geoscientific Model Development*, *10*(9), 3567–3589. https://doi.org/10.5194/gmd-10-3567-2017

Tian, B., & Dong, X. (2020). The Double-ITCZ bias in CMIP3, CMIP5, and CMIP6 models based on annual mean precipitation. *Geophysical Research Letters*, *47*(8). https://doi.org/10.1029/2020gl087232

Wang, C., Zou, L., & Zhou, T. (2018). SST biases over the Northwest Pacific and possible causes in CMIP5 models. *Science China Earth Sciences*, *61*(6), 792–803. https://doi.org/10.1007/s11430-017-9171-8

Wei, H., Subramanian, A. C., Karnauskas, K. B., DeMott, C. A., Mazloff, M. R., & Balmaseda, M. A. (2021). Tropical Pacific Air-Sea Interaction Processes and Biases in CESM2 and their relation to El Niño Development. *Journal of Geophysical Research Oceans*, *126*(6). https://doi.org/10.1029/2020jc016967

Wielicki, B., Barkstrom, B., Baum, B., Charlock, T., Green, R., Kratz, D., Lee, R., Minnis, P., Smith, G., Wong, N. T., Young, D., Cess, R., Coakley, J., Crommelynck, D., Donner, L., Kandel, R., King, M., Miller, A., Ramanathan, V., . . . Welch, R. (1998). Clouds and the Earth's Radiant Energy System (CERES): algorithm overview. *IEEE Transactions on Geoscience and Remote Sensing*, *36*(4), 1127–1141. https://doi.org/10.1109/36.701020

Williamson, D. B., Blaker, A. T., & Sinha, B. (2017). Tuning without over-tuning: parametric uncertainty quantification for the NEMO ocean model. *Geoscientific Model Development*, *10*(4), 1789–1816. https://doi.org/10.5194/gmd-10-1789-2017

Williamson, D., Blaker, A. T., Hampton, C., & Salter, J. (2014). Identifying and removing structural biases in climate models with history matching. *Climate Dynamics*, *45*(5–6), 1299–1324. https://doi.org/10.1007/s00382-014-2378-z

Williamson, D., Goldstein, M., Allison, L., Blaker, A., Challenor, P., Jackson, L., & Yamazaki,

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

K. (2013). History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble. *Climate Dynamics*, *41*(7–8), 1703–1729. https://doi.org/10.1007/s00382-013-1896-4

Wu, X., Deng, L., Song, X., & Zhang, G. J. (2007). Coupling of Convective Momentum Transport with Convective Heating in Global Climate Simulations. *Journal of the Atmospheric Sciences*, *64*(4), 1334–1349. https://doi.org/10.1175/jas3894.1

Wyrtki, K. (1975). El Niño—The dynamic response of the Equatorial Pacific oceanto atmospheric forcing. *Journal of Physical Oceanography*, *5*(4), 572–584. https://doi.org/10.1175/1520-0485(1975)005

Xie, F., Li, L., Pu, Y., Wang, B., Xue, W., Qiu, X., & Wang, G. (2023). Quantifying parametric uncertainty effects on tropical cloud fraction in an AGCM. *Journal of Advances in Modeling Earth Systems*, *15*(4). https://doi.org/10.1029/2022ms003221

Xie, F., Li, L., Wang, B., & Xue, W. (2017). Impacts of uncertain cloud-related parameters on Pacific Walker circulation simulation in GAMIL2. *Atmospheric and Oceanic Science Letters*, *11*(1), 7–14. https://doi.org/10.1080/16742834.2018.1392228

Yu R. (1994). A two—step shape—preserving advection scheme. *Advances in Atmospheric Sciences*, *11*(4), 479–490. https://doi.org/10.1007/bf02658169

Yang, B., Qian, Y., Lin, G., Leung, L. R., Rasch, P. J., Zhang, G. J., McFarlane, S. A., Zhao, C., Zhang, Y., Wang, H., Wang, M., & Liu, X. (2012). 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. https://doi.org/10.1029/2012jd018213

Zhang, L., & Zhao, C. (2015). Processes and mechanisms for the model SST biases in the North Atlantic and North Pacific: A link with the Atlantic meridional overturning circulation. *Journal of Advances in Modeling Earth Systems*, *7*(2), 739–758. https://doi.org/10.1002/2014ms000415

Zhang, T., Li, L., Lin, Y., Xue, W., Xie, F., Xu, H., & Huang, X. (2015). An automatic and effective parameter optimization method for model tuning. *Geoscientific Model Development*, *8*(11), 3579–3591. https://doi.org/10.5194/gmd-8-3579-2015

Zou, L., Qian, Y., Zhou, T., & Yang, B. (2014). Parameter Tuning and Calibration of RegCM3 with MIT–Emanuel Cumulus Parameterization Scheme over CORDEX East Asia Domain. *Journal of Climate*, *27*(20), 7687–7701. https://doi.org/10.1175/jcli-d-14-00229.1

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

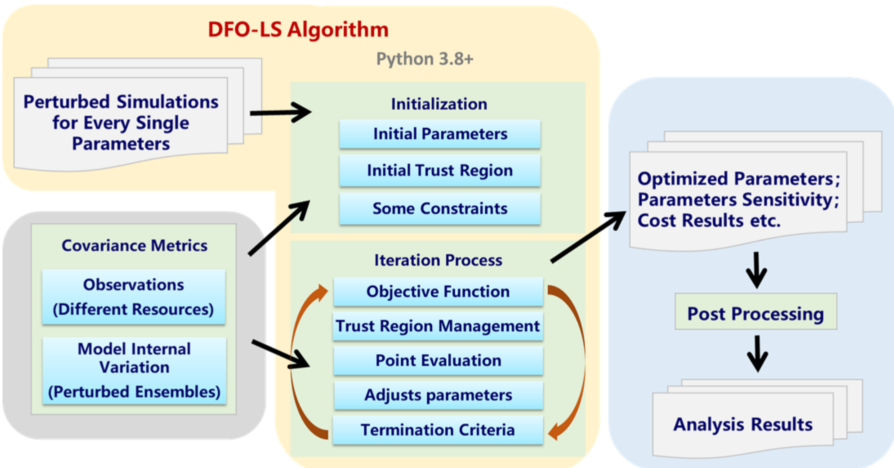


Figure 1. Automatic tuning framework structure. Perturbed simulation results for each parameter are used for sensitivity analysis and determining the trust region size. Two key covariance metrics—observational error and model internal variation—help adjust parameter values in the objective function. The DFO-LS algorithm optimizes the parameters, and the post-processing module analyzes sensitivity, cost function results, and generates visualizations.

Table 1: Observations used for model evaluation, along with their target values and associated uncertainties .

Variables name	Description	Classifications	Target	Uncertainty
MSLP	Mean sea level pressure (hPa);	MSLP_NHX_DGM	277.52	22.85
		MSLP_TROPICSLAND_DGM	35.42	13.69
		MSLP_TROPICSOCEAN_DGM	187.34	1.04
T500	Temperature at 500hPa (K)	TEMP@500_NHX	251.42	0.12
		TEMP@500_SHX	249.38	0.56
		TEMP@500_TROPICSLAND	266.27	0.27
		TEMP@500_TROPICSOCEAN	266.60	0.23
RH500	Relative humidity at 500hPa (%)	RH@500_NHX	52.75	7.04
		RH@500_SHX	51.05	4.79
		RH@500_TROPICSLAND	40.36	6.67
		RH@500_TROPICSOCEAN	32.57	3.01
NETFLUX	Net heat flux at top of atmosphere (W/m^2)	netflux_GLOBAL	0.98	2.0
OLR	Outgoing long wave flux at top of atmosphere (W/m^2)	OLR_NHX	223.57-	2.5
		OLR_SHX	216.86	
		OLR_TROPICSLAND	255.09	
		OLR_TROPICSOCEAN	261.35	

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

OLRC	Outgoing long wave clearsky flux at top of atmosphere (W/m^2)	OLRC_NHX	247.71	4.5
		OLRC_SHX	243.59	
		OLRC_TROPICSLAND	288.64	
		OLRC_TROPICSOCEAN	290.21	
RSR	Outgoing shortwave flux at top of atmosphere (W/m^2)	RSR_NHX	100.91	2.5
		RSR_SHX	107.55	
		RSR_TROPICSLAND	116.04	
		RSR_TROPICSOCEAN	86.92	
RSRC	Outgoing shortwave clearsky flux at top of atmosphere (W/m^2)	RSRC_NHX	57.98	5.0
		RSRC_SHX	53.65	
		RSRC_TROPICSLAND	75.67	
		RSRC_TROPICSOCEAN	42.42	
PRECIP	Total precipitation (m/s)	Lprecip_NHX	1.60e-8	0.35e-9
		Lprecip_SHX	1.42e-8	4.29e-9
		Lprecip_TROPICSLAND	4.47e-8	0.37e-9
T2M	Temperature at 2 meters (K)	LAT_NHX	275.72-	0.06
		LAT_SHX	280.08	0.49
		LAT_TROPICSLAND	297.10	0.31

Table 2: Summary of tunable parameters in GAMIL3, including their default values and plausible ranges.

Parameters	Description	Range	Default Values
c0_conv	Precipitation efficiency for deep convection	1.e-4-5.e-3	1.e-3
rhcrit	Threshold value for RH for deep convection	0.65-0.95	0.85
captlmt	threshold value for cape for deep convection	20-200	70
alfa	Initial deep convection cloud downdraft mass flux	0.05-0.6	0.2
ke	Evaporation efficiency of deep convection precipitation	1.e-6-1.5e-5	9.e-6
c0	rain water autoconversion coefficient	3.e-5-2.e-4	5.e-5
cmftau	characteristic adjustment time scale	1800-14400	4800
rhminl	Threshold RH for low stable clouds	0.8-0.99	0.95
rhminh	Threshold RH for high stable clouds	0.4-0.99	0.5
dthdpmn	Most stable lapse rate below 750hPa, stability trigger for stratus clouds	-0.15- -0.05	-0.08
sh1	Parameters for shallow convection cloud fraction	0.0-1.0	0.04
sh2	—	10-1000	500

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

dp1	Parameters for deep convection cloud fraction	0.0-1.0	0.1
dp2	—	10-1000	500
ccrit	Minimum allowable sqrt(TKE)/wstar	0.0-1.0	0.5
dzmin	minimum cloud depth to precipitate	0.0-100.0	0.0
Dcs	Autoconversion size threshold for ice to snow	1.e-5-1.e-3	2.e-4
ecr	collection efficiency cloud droplets/rain	0.5-2.0	1.0
ai	Fall speed parameter for stratiform cloud ice	500-1500	700
qcvar	Inverse relative variance of subgrid scale cloud water	0.1-2.0	1.0

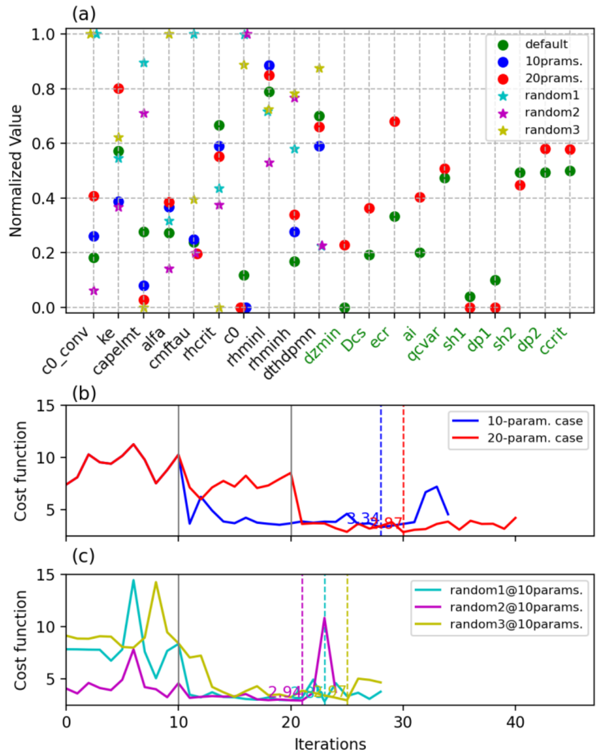
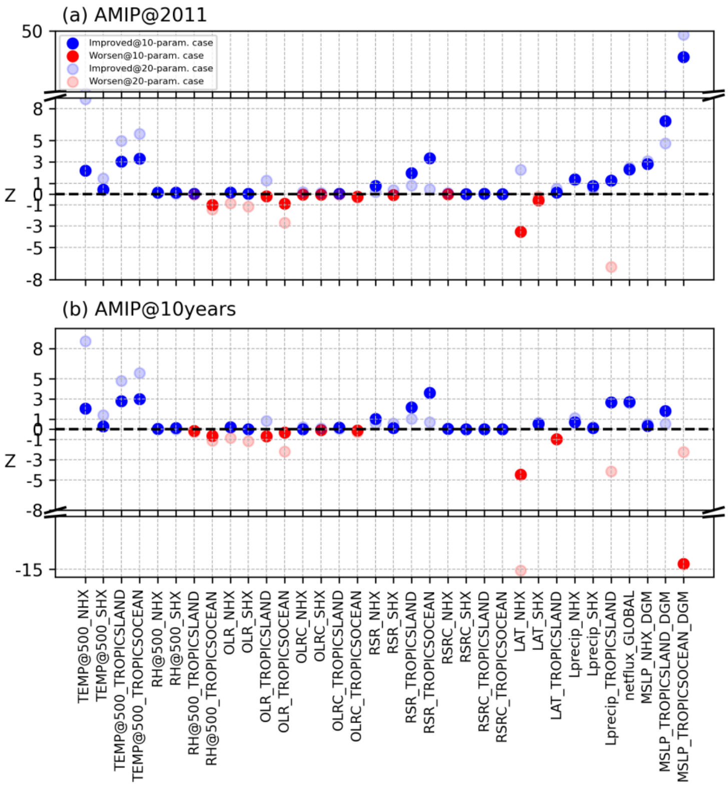


Figure 2. Normalized values of tuning parameters for default and optimized cases (a), along with changes in the cost function value over iterations for the 10- and 20-parameter cases (b) and three sensitivity experiments (c).

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025



1056

1057 **Figure 3**. Z values for the 1-year (a) and 10-year (b) AMIP simulations. Solid and hollow dots
1058 represent tuning with 10 and 20 parameters, respectively. Blue dots indicate improved performance,
1059 while red dots show deterioration. The black dashed line at Z = 0 separates improved from non-
1060 improved variables.

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

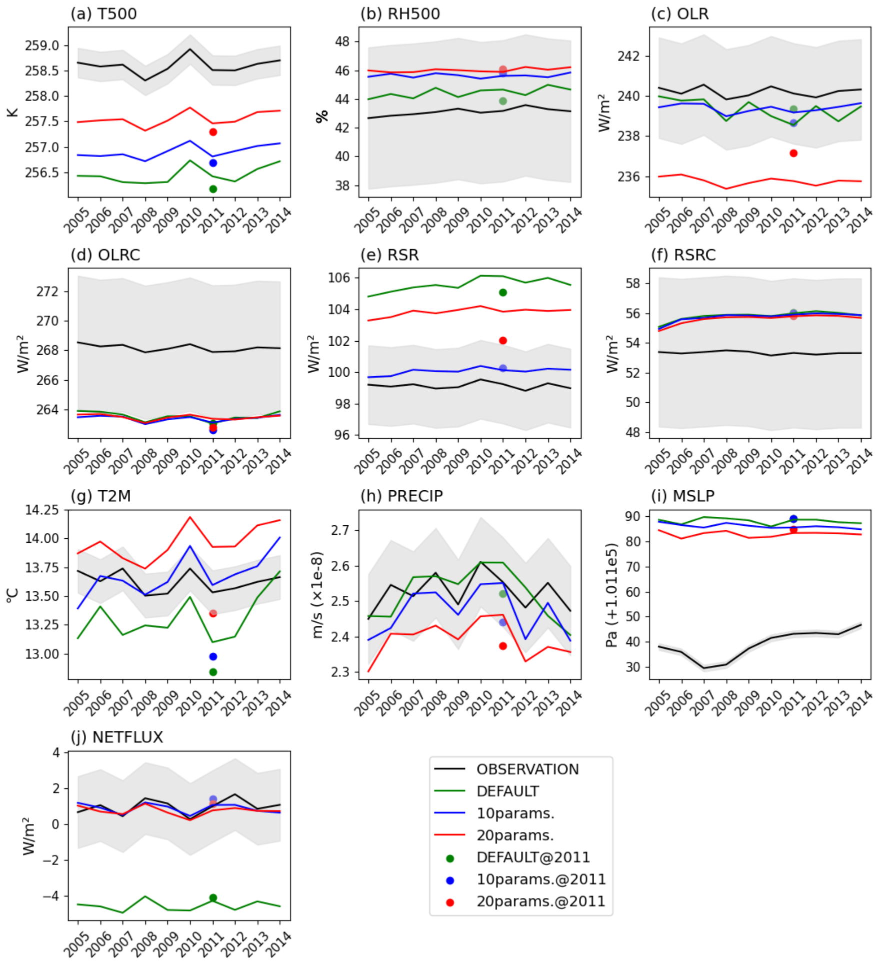


Figure 4. 1-year AMIP results (dots) and time series (lines) for three cases for: T500 (a), RH500 (b), OLR (c), OLRC (d), RSR (e), RSRC (f), T2M (g), PRECIP (h), MSLP (i) and NETFLUX (j). The cases include the default case (green lines and dots), 10-parameter case (blue lines and dots), and 20-parameter case (red lines and dots). The black lines and shadings represent the observations and their associated uncertainties.

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

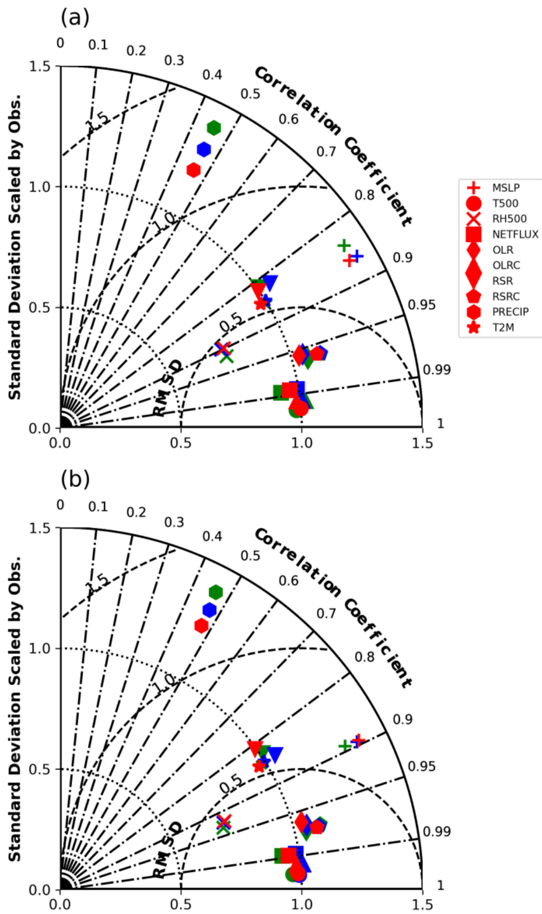


Figure 5. Taylor-diagram showing all variables for three cases in 2011 (a) and the 10-year AMIP simulations (b). Shown are default case (green), 10-parameter case (blue), and 20-parameter case (red).

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

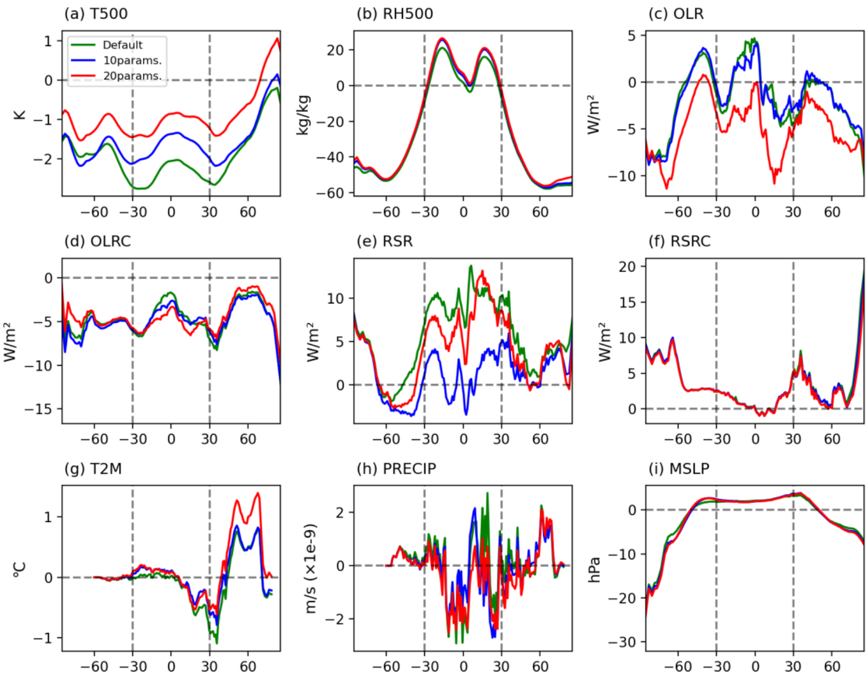


Figure 6. Meridional distributions of the annual mean bias between three cases and observations for: T500 (a), RH500 (b), OLR (c), OLRC (d), RSR (e), RSRC (f), T2M (g), PRECIP (h) and MSLP (i) from the 10-year AMIP simulations. Shown are default case (green), 10-parameter case (blue), and 20-parameter case (red).

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

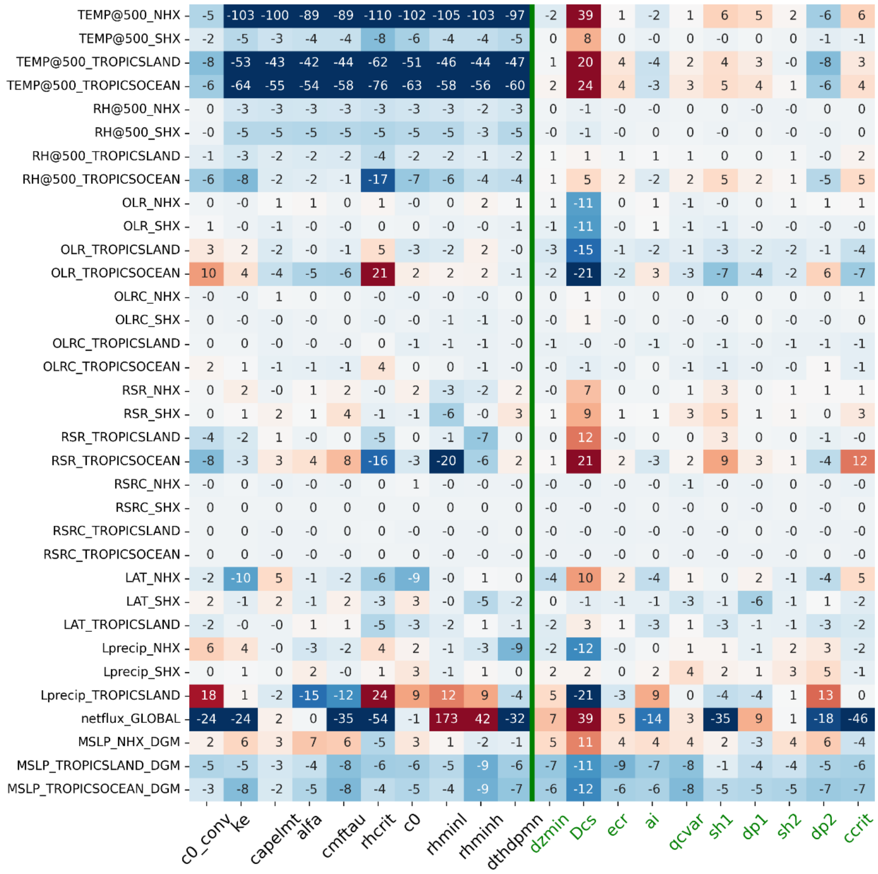
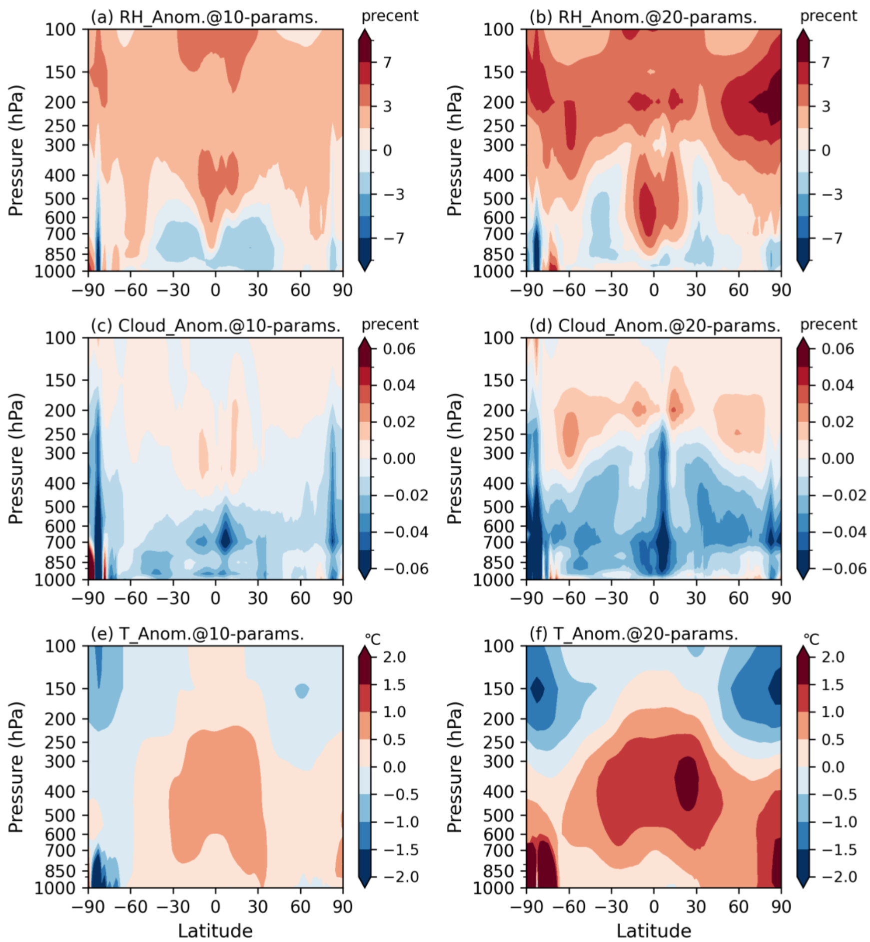


Figure 7. Normalized Jacobian for all 20 parameters, with values normalized by the total covariance metrics. The x-axis shows the parameter names, while the y-axis represents the variables. Black parameters are used in the 10-parameter case, and green ones are added in the 20-parameter case. Red and blue indicate positive and negative effects, respectively, with darker shades showing greater impact.

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025



1083

1084 **Figure 8**. Latitude-pressure anomaly distributions relative to the default case for relative humidity
1085 (a, b), cloud fraction (c, d), and temperature (e, f) from 10-year AMIP simulations: 10-parameter
1086 case (a, c, e) and 20-parameter case (b, d, f).

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

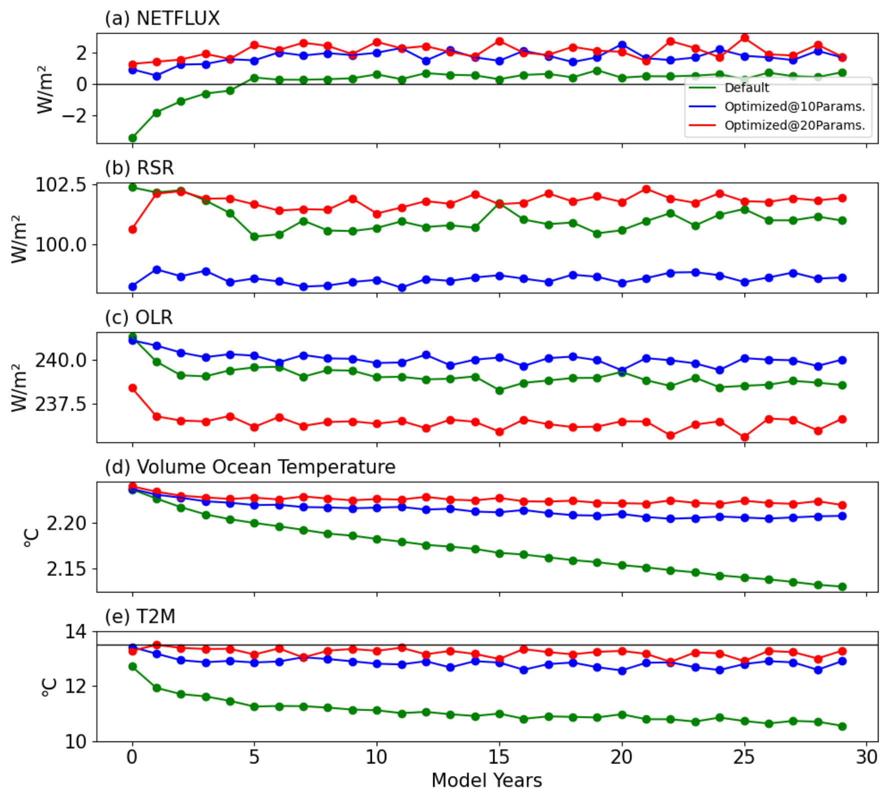
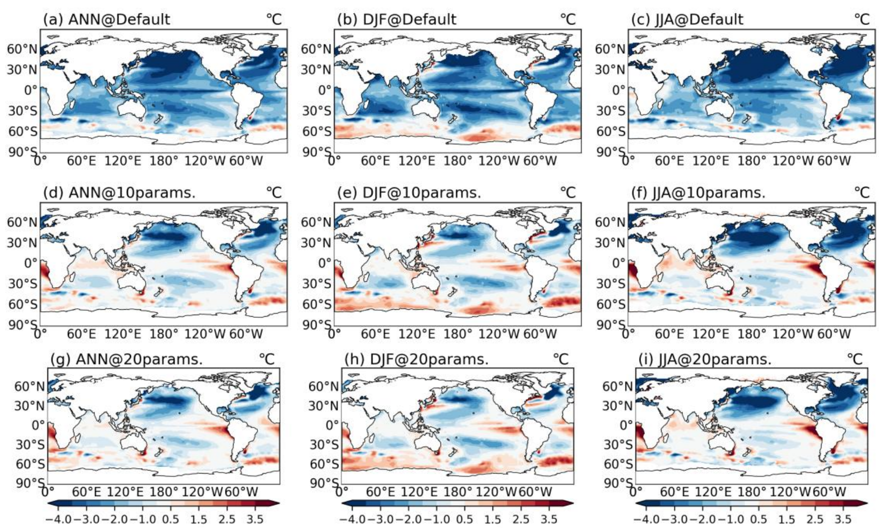
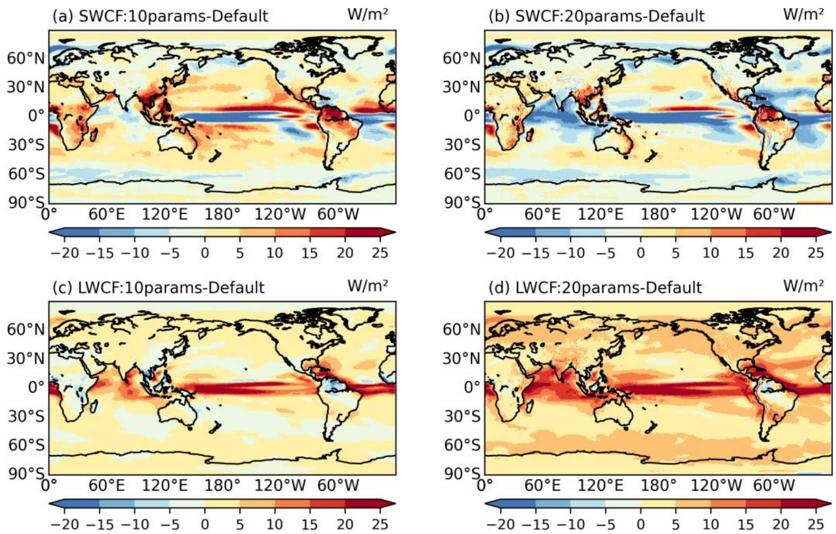


Figure 9. Results from the 30-year piControl simulation for NETFLUX (a), RSR (b) and OLR (c) radiation, mean volume-averaged ocean temperature (d), and T2M in the default (green), 10-parameter (blue), and 20-parameter cases (red) cases.



https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025

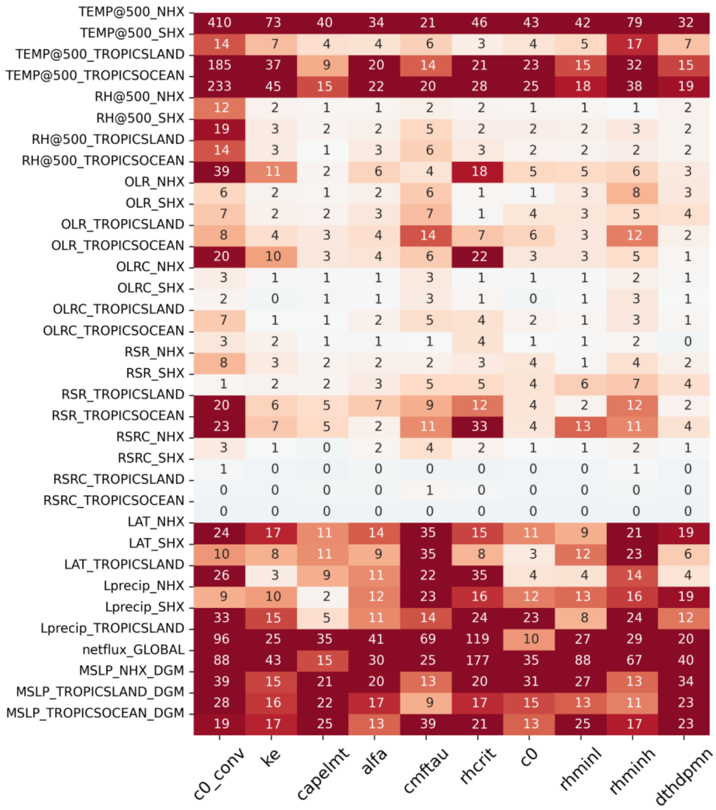
1092 **Figure 10**. Sea surface temperature biases relative to observations (HadISST; Rayner et al., 2003)
1093 from the last 15 years of piControl simulations for the default case (a, b, c) and two optimized
1094 cases (d-i).



1095

1096 **Figure 11**. Distribution of shortwave (a, b) and longwave (c, d) cloud forcing differences between
1097 the two optimized cases and the default case.

https://doi.org/10.5194/egusphere-2024-3770
Preprint. Discussion started: 10 February 2025



1098

1099 **Figure 12**. Similar as Fig. 7, but showing the range of Jacobians calculated from the optimized parameter set across four cases: the original optimized case and three sensitivity cases.

1100