# Automatic tuning of the Community Atmospheric Model CAM5.3 by using short-term hindcasts with an improved downhill simplex optimization method

Tao Zhang[1,2], Minghua Zhang[4], Wuyin Lin[2], Yanluan Lin[1], Wei Xue[1,3], Haiyang Yu[4], Juanxiong He[5], Xiaoge Xin[6], Hsi-Yen Ma[7], Shaochen Xie[7], and Weimin Zheng[3]

1. Ministry of Education Key Laboratory for Earth System Modeling, and Department for Earth System Science, Tsinghua University, Beijing, China
2. Brookhaven National Laboratory, New York, USA
3. Department of Computer Science and Technology, Tsinghua University, Beijing, China
4. School of Marine and Atmospheric Sciences, Stony Brook University, New York, USA
5. Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China
6. Beijing Climate Center, China Meteorological Administration, Beijing, China
7. Lawrence Livermore National Laboratory, California, USA

Correspondence: Minghua Zhang (minghua.zhang@stonybrook.edu), Wei Xue (xuewei@mail.tsinghua.edu.cn)

**Abstract.**

Traditional trial-and-error tuning of uncertain parameters in global atmospheric General Circulation Models (GCM) is time consuming and subjective. This study explores the feasibility of automatic optimization of GCM parameters for fast physics by using short-term hindcasts. An automatic workflow is described and applied to the Community Atmospheric Model (CAM5) to optimize several parameters in its cloud and convective parameterizations. We show that the auto-optimization leads to 10% reduction of the overall bias in CAM5, which is already a well calibrated model, based on a pre-defined metric that includes precipitation, temperature, humidity, and longwave/shortwave cloud forcing. The computational cost of the entire optimization

procedure is about equivalent to about a single 12-year atmospheric model simulation. The tuning reduces the large underestimation in the CAM5 longwave cloud forcing by decreasing the threshold relative humidity and the sedimentation velocity of ice crystals in the cloud schemes; it reduces the overestimation of precipitation by increasing the adjustment time in the convection scheme. The physical processes behind the tuned model performance for each targeted field are discussed. Limitations of the automatic tuning are described, including the slight deterioration in some targeted fields that reflect the structural errors of the model. It is pointed out that automatic tuning can be a viable supplement to process-oriented model evaluations and improvement.

## 1 Introduction

In general circulation models (GCMs), physical parameterizations are used to describe the statistical characteristics of various sub-grid-scale physical processes (Hack et al., 1994; Williams, 2005; Qian et al., 2015). These parameterizations contain uncertain parameters because the statistical relationships are often derived from sparse observations or from environmental conditions that differ from what the models are used for. Parameterization schemes that have many uncertain parameters include deep convection, shallow convection, and cloud microphysics/macrophysics. To achieve good performance of the model on some specific metrics, the values of these uncertain parameters are traditionally tuned based on the statistics of the final model performance or insight of the model developers through comprehensive comparisons and theoretical analysis of model simulations against observations (Allen et al., 2000; Hakkarainen et al., 2012; Yang et al., 2013). Generally, the uncertain physical parameters need to be re-tuned when new parameterization schemes are added into the models or used to replace existing one (Li et al., 2013).

Recent studies take advantage of optimization algorithms to automatically and more effectively tune the uncertain parameters (Bardenet et al., 2013; Yang et al., 2013; Zhang et al., 2015). For example, Yang et al. (2013) tuned serval parameters in Zhang-McFarlane convection scheme in Community Atmosphere Model Version 5 (CAM5, Neale et al. (2010)) using the simulated stochastic approximation annealing method. Qian et al. (2015) and Zhao et al. (2013) investigated the parameter sensitivity related to cloud physics, convection, aerosols and cloud microphysics in CAM5 using the generalized linear model. However, optimizations as in these works for GCMs require a long-time spin-up period to attain physically robust and meaningful signals, which is caused by strong nonlinear interactions at multiple scales between relevant processes (Wan et al., 2014). The parametric space of an AGCM is often strongly non-linear, multi-modal, high-dimensional, and inseparable. Therefore, automatically tuning parameters of global climate models requires a lot of model simulations with huge computational cost. This is also true for parameter sensitivity analysis which requires thousands of model runs to attain enough parameter samples.

One approach to reduce the high computational burden is to approximate and replace the expensive model simulations with a cheaper-to-run surrogate model, which uses the regression methods to describe the relationship between input (i.e., the 15 adjustable parameters of a model) and output (i.e., the output variables of a GCM) (Wang and Shan, 2007; Neelin et al., 2010; Wang et al., 2014) to represent a real GCM. However, training an accurate surrogate model requires a large amount of input- output sampling data, which are obtained by running the GCM with different sets of parameters selected in a feasible parameter space. As a result, the total computational cost is still very large. Meanwhile, due to the strongly nonlinear characteristics, the surrogate model of AGCMs often cannot meet the fitting accuracy or can be an overfitting to the model output.

The purpose of this study is to describe a method that combines automatic tuning with short-term hindcasts to optimize physical parameters, and demonstrate its application by using CAM5. The tuning parameters are selected based on previous CAM5 parameter sensitivity analysis works (i.e., Zhang et al., 2015; Qian et al., 2015; and Zhao et al., 2013). A key question is whether the results tuned automatically in hindcasts can truly translate to the model's climate simulation. To our knowledge, this paper is the first to use short-term weather forecasts to self-calibrate a climate model.

The paper is organized as follows. The next section gives the description of the model and experimental design. Section 3 describes the tuning parameters, metrics and the optimization algorithm. The optimized model and results are presented in Section 4. The last section contains the summary and discussion.

## 2 Model and experiments

In this study, we use CAM5 as an example. The dynamical core uses the finite volume method of Lin and Rood (1996) and Lin (2004). Shallow convection is represented as in Park and Bretherton (2009). Deep convection is parameterized by Zhang and McFarlane (1995), which is further modified by Neale et al. (2008) as well as Richter and Rasch (2008). The cloud microphysics is handled by Morrison and Gettelman (2008). Fractional stratiform condensation are calculated by the parameterization of Zhang et al. (2003) and Park et al. (2014). The vertical transport of moisture, momentum, and heat by turbulent eddies are handled by Bretherton and Park (2009). Radiation is calculated by the Rapid Radiative Transfer Model for GCMs (RRTMG, Iacono et al. (2008); Mlawer et al. (1997)). Land surface process are represented by the Community Land Model version 4 (CLM4, Lawrence et al. (2011)). More details are in Neale et al. (2010).

Two types of model experiments are conducted. One is the short-term hindcast simulations for model tuning. The second is AMIP simulation for verification of the tuned model. The hindcasts are initialized by the Year of Tropical Convection (YOTC) from the European Center for Medium-Range Weather Forecasts (ECMWF) re-analysis. The initialization uses the approach described in Xie et al. (2004) in the Cloud-Associated Parameterizations Testbed (CAPT) developed by US Department of Energy (US DOE). Since the objective of the tuning approach presented here is not only for auto-calibration of the model, but also for fast calculations, only one-month hindcasts of July 2009 are~~is~~ used in the tuning process. We carry out the simulations once every 3 days with a 3-day hindcast (labeled as interval-Day3) during the optimization iteration. All of the 3 ~~-~~day simulations for each hindcast run are used to make up the whole monthly data, which constitutes 31 days of model output. The AMIP simulation is conducted for 2000-2004 by using the observed climatological sea ice and sea surface temperature (Rayner et al., 2003). Simulation of the last three years is used for evaluation of the model. All simulations here use 0.9 latitude x 1.25 longitude horizontal resolution, with 30 vertical layers.

The observational data are from the Global Precipitation Climatology Project (GPCP, Huffman et al. (2001)) for precipitation, the International Satellite Cloud Climatology Project (ISCCP)- Flux Data (Trenberth et al., 2009) for radiation fluxes, the CloudSat (Stephens et al., 2002) and the Cloud–Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO, Winker et al. (2009)) for satellite cloud data, and the National Center for Environmental Prediction-National Center for Atmospheric Research (NCEP-NCAR, Kalnay et al. (1996)) reanalysis for humidity and temperature.

For this study, we focus on tuning parameters that are associated with fast physical processes so that short-term hindcasts can be used as an economical way of tuning. The philosophy behind the hindcasts is to keep the model dynamics as close to observation as possible while testing how the model simulates the quantities associated with fast physical processes. In other words, given the correct large-scale atmospheric conditions, errors in the physical variables are used to calibrate the fast physics parameters. This is different from calibration using AMIP simulations in which the circulation responds to the physics. The feasibility of the hindcast approach is based on the fact that errors in atmospheric models show up quickly in initialized experiments (Xie et al., 2004; Klein et al., 2006; Boyle et al., 25 2008; Hannay et al., 2008; Williams and Brooks, 2008; Martin et al., 2010; Xie et al., 2012; Ma et al., 2013, 2014; Wan et al., 2014). This is also found in the present study. Figure 1 shows the characteristics of the main biases in the CAPT and AMIP simulations in the default model for the five fields of long-wave and short-wave cloud forcing (LWCF and SWCF), humidity and temperature at 850 hPa (Q850, T850), and precipitation (PRECT). For the CAPT, the biases are for July 2009, while for AMIP they are for July averaged over three years. It is seen that the CAPT hindcasts capture a great number of the systematic biases in the AMIP simulations.

## 3 Tuning metrics and the optimization method

Parameter estimation for a complex model involves several choices, including (1) what parameters to optimize; and what are the range of uncertainties in the parameters; (2) how to select and construct a performance metric; (3) how to estimate/optimize the parameters in a high-dimensional space; and (4) how to embed the parameter estimation in the process-based evaluation and development of the model. This section describes the first three questions. The last question is left to Section 4.

### 3.1 Model parameters

In our study, the tuning parameters are selected based on the CAM5 sensitivity results of Zhang et al. (2015). They include three parameters from the deep convection scheme and three parameters from the cloud scheme. They are listed in Table 1, along with their default values. The parameters from the convection scheme are the autoconversion efficiency of cloud water to precipitation, separately for land and ocean, and the convective relaxation time scale. The parameters from the cloud scheme are the minimum threshold relative humidity to form clouds, which is an equivalent parameter to the width of the subgrid scale distribution of relative humidity, separately for high and low clouds, and the sedimentation velocity of ice crystals. All these parameters are known to have large uncertainties.

For the uncertainty ranges of the parameters to be used as bounds of optimal tuning, ideally, they should be derived from the development process of the parameterizations as part of the information from the empirically fitting to observations or to process models. In practice, however, most parameterizations do not contain this information. The uncertainty ranges of the parameters in this study are based on ~~on previously published works (~~Covey et al.~~, (~~(2013) and previous CAM5 tuning exercises (Yang et. al., 2013; Qian et. al., 2015). They are listed in Table 1.

### 3.2 The metrics

Several metrics have been used in the literature to quantitatively evaluate and compare the performance of overall simulations of climate models (Murphy et al., 2004; Reichler and Kim, 2008; Gleckler et al., 2008). As a demonstration of the optimization method, in this study we use five fields in Figure 1 (LWCF, SWCF, PRECT, Q850 and T850) to form a metric. The daily observational data sources for these five fields are listed in Table 2. The tuning metric combines the Mean Square Error (MSE) of the five variables into a single target as the improvement index of model simulation, which is regarded as a function of the uncertain parameter values. When calculating the metric, we first compute the MSE of each target variable of the model simulation against the reanalysis/observations as in Equations (1) and (2) for the tuning model and the default model respectively (Taylor, 2001; Yang et al., 2013; Zhang et al., 2015):

$$(\sigma_m^F)^2 = \sum_{i=1}^{I} w(i)\,(E_m^F(i) - E_o^F(i))^2 \qquad (1)$$

$$(\sigma_r^F)^2 = \sum_{i=1}^{I} w(i)\left(E_r^F(i) - E_o^F(i)\right)^2 \qquad (2)$$

where $E_m^F(i)$ is the model output at the $i_{th}$ location, and $E_o^F(i)$ is the corresponding reanalysis or observation data. $E_r^F(i)$ is the model simulated variables using the default parameter values. $I$ is the number of grids. w is the weight value based on grid area. The final target improvement index is calculated by using the average of the MSE normalized by that of the control simulation as defined in Equation (3):

$$\chi^2 = \frac{1}{N^F}\sum_{F=1}^{N^F}\left(\frac{\sigma_m^F}{\sigma_r^F}\right)^2 \qquad (3)$$

where $N^F$ is the number of the variables in Table 2. If the index is less than 1, the tuned simulation is considered as having better performance than the default simulation. The smaller this index value, the better the improvement achieves.

When the differences between model simulation and observation at different grid points are independent of each other and follow normal distributions, minimizing the MSE over all grids would be equivalent to the maximum likelihood estimation of the uncertain parameters. For our experimental design, however, the mismatch between the short-term forecasts and instantaneous observation could be caused by small spatial displacements due to errors in the model initial condition instead of the model parameters. In such cases, errors could be highly correlated between neighboring grids, and the dependence of the metric on the control parameters may be marginalized or obscured. This problem may be lessened in long-term climate simulations, but extra care is needed for short-term forecasts. We therefore choose to use zonally averaged fields from the model and observations in the metric calculation to focus on the effective response at global scale.

### 3.3 The optimization method

The optimization method is based on an improved downhill simplex optimization algorithm to find a local minimum. Zhang et al. (2015) shows this algorithm can find a good local minimum solution based on the better choice of the initial parameter values. Global optimization algorithms that aim to find the true minimum solution always require extreme amount of computational cost compared to the method used here, such as Covariance Matrix Adaptation Evolution Strategy (Hansen et al., 2003), Efficient Global Optimization (Jones et al., 1998) and Genetic Algorithm (Goldberg ,1989), and there is no guarantee they can find one within limited number of iterations which are often invoked for complicated problems. In practice, Zhang et al. (2015) showed the improved downhill simplex method outperformed the global optimization algorithms with the limited optimal iterations.

The optimization procedure takes two steps. First, a pre-processing of selected parameter initial values is carried out to accelerate the convergence of optimization algorithm and to account for the ill-conditioning of the minimization problem. Next, the improved downhill simplex optimization algorithm is utilized to solve the problem due to its fast convergence and low computation for low-dimensional space. Meanwhile, an automatic workflow (Zhang et al., 2015) is used to take care of the complicated configuration process and management of model tuning. In the following, we give a brief description of these two steps. More details can be found in Zhang et al. (2015).

The pre-processing uses a sampling strategy based on the single parameter perturbation (SPP) method, in which at one time just perturbs only one parameter with others fixed. The perturbed samples are uniform distribution across parametric space. Equation (3) defines the improvement index for each parameter samples. The distance of samples, defined as the difference between the indexes from using two adjacent samples, is then calculated. We call this step the first-level sampling. If the distance between two adjacent samples is greater than a predefined threshold, more refined samples between these two adjacent samples are conducted. This is the second-level sampling. Finally, the candidate initial values for the optimization method choose the k+1 samples with the best improvement index values, where k is the number of the parameters. In this study, k is 6. The convergence performance of the traditional downhill simplex heavily relies on the quality of its initial values. Inappropriate ones may give rise to ill-conditioned simplex geometry. Therefore, a simplex checking is carried out to ensure as many distinct values of parameters as possible during the process of looking for initial values to ensure that the simplex is a good-conditioned geometry.

The downhill simplex algorithm calculates the parameter values and the corresponding improvement index as defined in Equation (3) in each step of the iterations. The optimal results are achieved by expanding or shrinking the simplex geometry in each optimal step. In the processes of searching for the minimum index, the best set of tuning parameter values up to the current iteration step is kept to look for the direction and magnitude of the increments. The iteration is terminated when the tuning parameters reach quasi-steady state.

Figure 2 summarizes the workflow of the experiments. The workflow is automated. It has two components: model calibration and verification. The calibration uses the hindcasts, the pre-defined metric and the optimization algorithm to derive the optimal parameter values. The verification uses AMIP climate simulation to check how effective the auto-calibration is for the application goal, which is to improve the metric in the AMIP simulation.

## 4 Results

### 4.1 The optimized model

The change of performance index in the optimization iterations as a function of iteration step is shown in Figure 3. The blue line is the best performance index up to the current step. The red line is the real performance up to the current step. The latter has spikes during the iteration, especially near step 70, suggesting that the performance index in the parameter space has a complex geometry. Each iteration involves 31 days of hindcasts. The iteration is stopped at about the 142th iteration step when the searched parameters reach quasi-steady state. With 180 computing cores on a linux cluster, each iteration takes about 50 minutes. The computational time for an entire optimization is equivalent to about 12 years of an AMIP simulation, which is a tremendous reduction of computing time relative to traditional model tuning.

The tuned values of the parameters are given in the column of "Tuned" in Table 1. In the default model, the autoconversion parameter c0 is smaller over land than over the oceans, reflecting more aerosols and smaller cloud particle sizes over land than over the oceans. When compared with the default values, the tuned c0 value over land is even smaller while the value over the ocean is even larger. The parameter that represents the time scale of the convective adjustment is larger in the tuned model than in the default model. For the three parameters in the cloud scheme, the minimum relative humidity in the tuned model is reduced for high clouds but increased for low clouds in the tuned model. The sedimentation velocity of ice crystals is reduced by over a half in the tuned model. The physical justification of these new parameter values is beyond the scope of this paper, but they are all within the range of known uncertainties by design of the optimal tuning. How the parameter change affects the simulation is discussed in Section 4.2.

The performance index of the tuned model in the hindcasts and the normalized MSE of the individual fields in the metric are given in Table 3 under the "Hindcasts" column. The performance index is reduced by about 10% in the tuned model. This is relatively a significant reduction, considering the fact that CAM5 is already a well-tuned model and a major upgrade of the CAM model from CAM4 to CAM5 also saw the changes in most of the variables are within 10% range in terms of RMSE (Flato et al., 2013). Looking at the MSE of the individual fields in the table, we find that the reduction in the performance index is not evenly distributed across the targeted fields. The largest reduction, at about 40%, is found for the MSE in the longwave cloud forcing LWCF. This is actually not a surprise. Zhang et al. (2015) showed that LWCF is highly sensitive to changes in the CAPE consumption time scale (zmconv_tau) and the minimum rh for high stable clouds (cldfrc_rhminh). Yang et al. (2013) also indicated the zmconv_tau was sensitive for LWCF. The autoconversion efficiency of cloud water to precipitation (zmconv_c0_lnd and zmconv_c0_ocn) and the cloud ice sedimentation velocity (cldsed_ai) were found to be sensitive for LWCF in Qian et al. (2016). That is to say, all the tuning parameters in this study are very sensitive for LWCF, resulting in this field to have the most improvement. There is about 8% reduction of MSE in precipitation PRECT and 4% reduction in 850-hPa temperature T850. However, two fields, the shortwave cloud forcing SWCF and the 850-hPa temperature Q850, are accompanied by 3% and 1% increases of errors respectively. As will be discussed later, this is indication of structural errors in the model whose solution cannot fit to all observations.

The next critical question is whether the optimal results tuned in hindcasts are shown in the AMIP simulation. The last column in Table 3 under the heading of "AMIP" gives the performance index of the tuned model and the normalized MSE of the individual fields from the AMIP simulation. Three things are noted: First, the overall performance index is also improved by about 10% in the AMIP simulation in the tuned model.

Second, as in the hindhcasts, the largest improvement is in the LWCF. Third, the fields that got improved in the AMIP simulations are the same as those in the hindcasts. We therefore conclude that the automatic tuning achieved the design goal of the algorithm.

We also examined a 10-variable metric that is used by the Atmospheric Model Working Group of the Community Earth System Model (CESM) (https://www2.cesm.ucar.edu/working-groups/amwg/metrics). The five variables that we used in the performance index are a subset of these fields, except that precipitation in the AMWG metric is separated into land and ocean components. Therefore, there are six additional fields in the AMWG metric. Table 4 shows the percentage bias of the ten fields between the default/optimized model and the reference observations, which is computed based on 2-dimensional monthly mean fields as the follows:

$$bias||\%|| = |\frac{\overline{EXP}(\overline{CNTL}) - \overline{OBS})}{\overline{OBS}}| \quad (4)$$

It is seen that among the six new variables, surface pressure, oceanic tropical rainfall, Pacific Ocean surface stress, and zonal wind at 300 hPb are all improved in the tuned model. Increased errors are seen surface air temperature and precipitation over land. This evaluation is overall consistent with the improved performance metrics shown in Table 3 in which zonally averaged fields were used. This comparison lends credence to the intended objective of the tuning, with the exception over land for which additional parameters may be included for tuning.

**4.2 Interpretation of the tuned results**

We next examine the physical processes behind the changed performance index in the tuned model. Figures 4a, 4b, and 4d show respectively the annually averaged high cloud amount in the AMIP simulation of the satellite observation from CloudSat and CALIPSO, the default model, and the model bias. It is seen that CAM5 significantly underestimated high clouds in the tropics, including the western Pacific warm pool, and the central Africa and America, except in the narrow zonal band of the Inter-Tropical Convergence Zone (ITCZ) in the Pacific. The model also underestimated high clouds in regions of middle-latitude storm tracks. Since high clouds have large impact on the longwave cloud forcing LWCF, these biases in high clouds would cause underestimation of LWCF. Figures 5a, 5b and 5d show the LWCF in the observation, the default model, and the model bias. The bias field (Figure 5d) clearly shows that the model significantly underestimates the LWCF. Its spatial pattern largely mirrors the bias field in high cloud amount in Figure 4d.

In the model optimization, as described before, a smaller relative humidity threshold value for high clouds in the cloud scheme and a smaller sedimentation velocity of ice crystals were derived. These two parameter adjustments can both act to increase high cloud amount and thus longwave cloud forcing. The simulated high cloud and its bias relative to observation are shown in Figures 4b to 4e. It can be seen that the overall bias in high cloud is significantly reduced in the tuned model. This leads to reduced negative bias in LWCF in the optimal model (Figures 5b to 5e).

Changes in clouds are inevitably accompanied by changes in the shortwave cloud forcing SWCF, which was slightly deteriorated in the tuned model as discussed previously. We find that while high clouds are increased in the tuned model, clouds in the middle troposphere are reduced in middle and high latitudes (Figure 6). This reduction in middle clouds may have compensated the impact of increased high clouds on SWCF since SWCF is also used in the performance metric. This reduction of middle clouds is consistent with the increased precipitation efficiency parameter c0 in the tuned model over the ocean and the reduced convection to be discussed later.

The impact of the tuning on other targeted fields is less dramatic than on LWCF. To see the impact clearly, we show in Figure 7 the zonally averaged biases in the AMIP simulation from the default CAM5 as the blue lines and the optimized model as the red lines. The 2-dimensional map figures are given in the Supplemental Materials. In addition to the large improvement in the LWCF, the overall improvement in PRECT and T850 can be seen. The optimized model simulates slightly smaller precipitation (PRECT) and warmer atmosphere (T850), which are all closer to observations. The reduction in precipitation is consistent with the larger value of the convection adjustment time scale in the tuned model than in the default model. The convection scheme uses a quasi-equilibrium closure based on the Convective Available Potential Energy (CAPE). The adjustment time scale is the denominator in the calculation of the cloud-base convective mass flux. When the time scale is longer, the mass flux is smaller, so is the convective precipitation. This reduction in precipitation is one likely cause of the larger SWCF (less cloud reflection) in the tuned model. In addition to the convection adjustment time scale, other parameters also impact precipitation. In particular, the impact of the increased precipitation efficiency over the ocean in the tuned model should partially offset the impact of the longer convective adjustment time scale. The change of PRECT is the net outcome of the multivariate dependences on all parameters that is found by the automatic optimization algorithm for the overall improvement of the performance index.

The increase in LWCF and the reduced precipitation PRECT in the optimal model are energetically consistent for the atmosphere: There is less atmospheric longwave radiative cooling and less condensational heating in the tuned model. The magnitude of the LWCF increase is large (2.42 W/m2) relative to the change in condensational heating (2.03 W/m2) as derived from the change in global mean precipitation amount. As a result, the atmosphere is slightly warmer, which is also closer to observation (Figure 7e) and this is an improvement to the default model.

While consistent improvements in different fields are desired, this is not always possible. For example, a warmer atmosphere is often accompanied by a moister atmosphere. Since temperature in the tuned model is warmer than that in the default model, there is more moisture in the tuned model. The atmosphere in the default model is already too moist (Figure 7d). As a result, the performance index in Q850 is slightly deteriorated. Since the optimization is based on a single combined metric of several target variables, the algorithm seeks to minimize this combined metric at the expense of the performance of other variables as long as the total metric is reduced. The fact that the default CAM5 overestimated water vapor and underestimated temperature as shown in Figures 7d and 7e indicates structural errors in the model: improving temperature could lead to larger biases in water vapor in the current model.

In summary, the improved performance index in the LWCF is consistent with the dominant impact of the reduced values in the threshold relative humidity for high clouds and the sedimentation velocity of ice crystals. The improvement in PRECT is consistent with the increased convective adjustment time scale. The improvement in T850 is consistent with the large increase in LWCF and reduced radiative cooling of the atmosphere. The deterioration in SWCF is consistent with the impact of increased autoconversion rate, longer convective adjustment time scale, and increased threshold relative humidity of low clouds, all of which can lead to reduction of cloud water. The deterioration in Q850 is likely the result of larger T850 in the tuned model.

These results point to both the benefits and limitation of the described model tuning. The benefit is the improvement in a pre-defined metric, with has led to improvements in several fields. The limitation is that not all fields can be improved. Some fields may get worse as a result of the algorithm in achieving the largest improvement in the total pre-defined metric. One may use different weights for different fields in Equation (1) or impose conditional limits on the normalized MSE for the individual fields. The benefits of such alternative approaches will surely depend on specific applications, but structural errors cannot be eliminated by the tuning.

## 5 Summary and Discussion

We have presented a method of economic automatic tuning by using short-term hindcasts for one month. It is used to optimize CAM5 by adjusting several empirical parameters in its cloud and convection parameterizations. The computational cost of the entire tuning procedure is less 12-years of one single AMIP simulation. We have demonstrated that the tuning accomplished the design goal of the algorithm. We show about 10% improvement in our pre-defined metric for CAM5 that is already a well-calibrated model. Among the five targeted fields of LWCF, SWCF, PRECT, T850 and Q850, the largest improvement is to the longwave cloud radiative forcing LWCF, which has about 40% improvement in the zonal mean MSE. We have shown that while the improvements in LWCF, PRECT and T850 are consistent with the improved atmospheric energy budget, they lead to slight deterioration in the SWCF and Q850 that reflects structural errors of the model. The overall improvement is also seen in the 10-variable AMWG metrics.

The optimized model contains reduced values of the threshold relative humidity for high clouds and sediment velocity of ice crystals, which act to increase the high cloud amount and increase the longwave cloud forcing, thereby reducing its significant underestimation in the default model. The optimization gave increased convection adjustment time that can explain reduced precipitation in the tuned model and the reduction of the precipitation biases. These two changes also help to reduce the temperature bias. The gains in these fields however are accompanied by slight deterioration in shortwave cloud forcing that is consistent with the reduced precipitation, and slight deterioration in humidity that is consistent with the increased temperature. The optimized results can help understand the interactive effect of multiple parameters, discover the systematic and structural errors by exploring the parameter calibration ultimate performance.

While benefits of the automatic tuning are clearly seen, there are several limitations of using the present workflow for automatic tuning of GCMs. First, not all fields can be simultaneously improved since parameter tuning cannot eliminate structural errors in the model. Tuning is not an alternative to improving a model, rather it is an economic way to calibrate some parameters within a candidate parameterization framework. Second, the optimized model may be caused by compensation of errors. Therefore, process-based model evaluation and physical explanation of the model improvements are always necessary. Third, the tuning by using hindcasts is only applicable for parameters affecting fast physics. For model bias that develops over long time scales, such as those from coupled ocean-atmospheric models, this approach cannot be used, although the conceptual approach may be applied with longer integrations. Finally, the choices of the model parameters, uncertain ranges, and metrics are somewhat subjective. It would be much more satisfactory if their selections can be done automatically and more objectively. Several improvements can be made to the presented method. Different weights can be used for the targeted fields. Sensitivity to different target metrics can be studied. Multiple target metrics may be designed to optimize different sets of parameters. Constraints such as energy balance at the top of the atmosphere may be imposed. It is also possible to use time-varying solutions as metrics to target variabilities such as the Madden-Julian Oscillation (MJO) in models. These could be subject for future research.

**Code and data availability.** The source code of CAM5.3 are available from http://www.cesm.ucar.edu/models/cesm1.2/. The downhill simplex algorithm, the scripts of running the model driven by the optimization algorithm, and the scripts of computing metrics can be found at http://everest.msrc.sunysb.edu/tzhang/capt_tune/GCM_paras_tuner/. The observation data which is used to compute the metrics in the short-term hindcast tuning and validate the optimization in AMIP is at http://everest.msrc.sunysb.edu/tzhang/capt_tune/capt_tune_obs/.

**Acknowledgements.** This work is supported by the National Key R&D Program of China (Grant No. 2016YFA0602100 and 2017YFA0604500), the National Major Research High Performance Computing Program of China (Grant 2016YFB02008), and the National Natural Science Foundation of China (Grant No. 91530323 and 41776010). Additional support is provided by the CMDV project of the CESD of the US Department of Energy to the Stony Brook University. T. Zhang partially and W. Lin are supported by the U.S DOE BER's Climate Model Development and Validation project to BNL under contract DE-SC0012704. Hsi- Yen Ma and Shaocheng Xie are funded by the Regional and Global Climate Modeling and Atmospheric System Research programs of the U.S. Department of Energy, Office of Science as part of the CAPT. Hsi-Yen Ma and Shaocheng Xie support under the auspices of the U.S. Department of Energy by LLNL under contract DE-AC52-07NA27344.

## References


Allen, M. R., Stott, P. A., Mitchell, J. F., Schnur, R., and Delworth, T. L.: Quantifying the uncertainty in forecasts of anthropogenic climate change, Nature, 407, 617–620, 2000.

Bardenet, R., Brendel, M., Kégl, B., and Sebag, M.: Collaborative hyperparameter tuning, in: paper presented at the 30th International Conference on Machine Learning (ICML-13), pp. 199–207, ACM, Atlanta, USA, 2013.

Boyle, J., Klein, S., Zhang, G., Xie, S., and Wei, X.: Climate model forecast experiments for TOGA COARE, Monthly Weather Review, 136, 808–832, 2008.

Bretherton, C. S. and Park, S.: A new moist turbulence parameterization in the Community Atmosphere Model, Journal of Climate, 22, 3422–3448, 2009.

Covey, C., Lucas, D. D., Tannahill, J., Garaizar, X., and Klein, R.: Efficient screening of climate model sensitivity to a large number of perturbed input parameters, Journal of Advances in Modeling Earth Systems, 5, 598–610, 2013.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., et al.: Evaluation of climate models, 2013.

Gleckler, P. J., Taylor, K. E., and Doutriaux, C.: Performance metrics for climate models, Journal of Geophysical Research: Atmospheres (1984–2012), 113, 2008.

Goldberg, D. E., Korb, B., and Deb, K.: Messy genetic algorithms: Motivation, analysis, and first results, Complex systems, 3, 493–530, 1989.

Hack, J. J., Boville, B., Kiehl, J., Rasch, P., and Williamson, D.: Climate statistics from the National Center for Atmospheric Research community climate model CCM2, Journal of Geophysical Research: Atmospheres (1984–2012), 99, 20 785–20 813, 1994.

Hakkarainen, J., Ilin, A., Solonen, A., Laine, M., Haario, H., Tamminen, J., Oja, E., and Järvinen, H.: On closure parameter estimation in chaotic systems, Nonlinear Processes in Geophysics, 19, 127–143, 2012.

Hannay, C., Williamson, D., Olson, J., Neale, R., Gettelman, A., Morrison, H., Park, S., and Bretherton, C.: Short Term forecasts along the GCSS Pacific Cross-section: Evaluating new Parameterizations in the Community Atmospheric Model, in: 4th PAN-GCSS Meeting on advances in modeling and observing clouds and convection, Toulouse, France, 2008.

Hansen, N., Müller, S. D., and Koumoutsakos, P.: Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES), Evolutionary computation, 11, 1–18, 2003.

Huffman, G. J., Adler, R. F., Morrissey, M. M., Bolvin, D. T., Curtis, S., Joyce, R., McGavock, B., and Susskind, J.: Global precipitation at one-degree daily resolution from multisatellite observations, Journal of Hydrometeorology, 2, 36–50, 2001.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., and Collins, W. D.: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models, Journal of Geophysical Research: Atmospheres (1984–2012), 113, 2008.

Jones, D. R., Schonlau, M., and Welch, W. J.: Efficient global optimization of expensive black-box functions, Journal of Global optimization, 13, 455–492, 1998.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., et al.: The NCEP/NCAR 40-year reanalysis project, Bulletin of the American meteorological Society, 77, 437–471, 1996.

Klein, S. A., Jiang, X., Boyle, J., Malyshev, S., and Xie, S.: Diagnosis of the summertime warm and dry bias over the US Southern Great Plains in the GFDL climate model using a weather forecasting approach, Geophysical research letters, 33, 2006.

Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J., Zeng, X., Yang, Z.-L., Levis, S., Sakaguchi, K., et al.: Parameterization improvements and functional and structural advances in version 4 of the Community Land Model, Journal of Advances in Modeling Earth Systems, 3, 2011.

Li, L., Wang, B., Dong, L., Liu, L., Shen, S., Hu, N., Sun, W., Wang, Y., Huang, W., Shi, X., Pu, Y., and Yang, G.: Evaluation of grid-point atmospheric model of IAP LASG version 2 (GAMIL2), Advances in Atmospheric Sciences, 30, 855–867, 2013.

Lin, S.-J.: A “vertically Lagrangian” finite-volume dynamical core for global models, Monthly Weather Review, 132, 2293–2307, 2004. Lin, S.-J. and Rood, R. B.: Multidimensional flux-form semi-Lagrangian transport schemes, Monthly Weather Review, 124, 2046–2070,1996.

Ma, H.-Y., Xie, S., Boyle, J., Klein, S., and Zhang, Y.: Metrics and diagnostics for precipitation-related processes in climate model short-range hindcasts, Journal of Climate, 26, 1516–1534, 2013.

Ma, H.-Y., Xie, S., Klein, S., Williams, K., Boyle, J., Bony, S., Douville, H., Fermepin, S., Medeiros, B., Tyteca, S., et al.: On the correspon- dence between mean forecast errors and climate errors in CMIP5 models, Journal of Climate, 27, 1781–1798, 2014.

Martin, G., Milton, S., Senior, C., Brooks, M., Ineson, S., Reichler, T., and Kim, J.: Analysis and reduction of systematic errors through a seamless approach to modeling weather and climate, Journal of Climate, 23, 5933–5957, 2010.

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., and Clough, S. A.: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave, Journal of Geophysical Research: Atmospheres (1984–2012), 102, 16 663–16 682, 1997.

Murphy, J. M., Sexton, D. M., Barnett, D. N., Jones, G. S., Webb, M. J., Collins, M., and Stainforth, D. A.: Quantification of modelling uncertainties in a large ensemble of climate change simulations, Nature, 430, 768–772, 2004.

Neale, R. B., Richter, J. H., and Jochum, M.: The impact of convection on ENSO: From a delayed oscillator to a series of events, Journal of climate, 21, 5904–5924, 2008.

Neale, R. B., Chen, C.-C., Gettelman, A., Lauritzen, P. H., Park, S., Williamson, D. L., Conley, A. J., Garcia, R., Kinnison, D., Lamarque, J.-F., et al.: Description of the NCAR community atmosphere model (CAM 5.0), NCAR Tech. Note NCAR/TN-486+ STR, 2010.

Neelin, J. D., Bracco, A., Luo, H., McWilliams, J. C., and Meyerson, J. E.: Considerations for parameter optimization and sensitivity in climate models, Proceedings of the National Academy of Sciences, 107, 21 349–21 354, 2010.

Park, S. and Bretherton, C. S.: The University of Washington shallow convection and moist turbulence schemes and their impact on climate simulations with the Community Atmosphere Model, Journal of Climate, 22, 3449–3469, 2009.

Park, S., Bretherton, C. S., and Rasch, P. J.: Integrating cloud processes in the Community Atmosphere Model, version 5, Journal of Climate, 27, 6821–6856, 2014.

Qian, Y., Yan, H., Hou, Z., Johannesson, G., Klein, S., Lucas, D., Neale, R., Rasch, P., Swiler, L., Tannahill, J., et al.: Parametric sensitivity analysis of precipitation at global and local scales in the Community Atmosphere Model CAM5, Journal of Advances in Modeling Earth Systems, 7, 382–411, 2015.

Qian, Y., Wan, H., Rasch, P., Zhang, K., Ma, P., Lin, W., Xie, S., Singhy, B., Larson, V., Neale, R., Gettelman, A., Bogenschutz, P., Wang, H., Zhao, C.: Parametric sensitivity in ACME-V1 atmosphere model revealed by short Perturbed Parameters Ensemble (PPE) simulations, https://climatemodeling.science.energy.gov/sites/default/files/presentations/Qian-ShortSimulation-2016SpringMeeting-ACME_Poster.pdf, 2016.

Rayner, N., Parker, D. E., Horton, E., Folland, C., Alexander, L., Rowell, D., Kent, E., and Kaplan, A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, Journal of Geophysical Research: Atmospheres, 108, 1871–2000, 2003.

Reichler, T. and Kim, J.: How well do coupled models simulate today's climate?, Bulletin of the American Meteorological Society, 89, 303–311, 2008.

Richter, J. H. and Rasch, P. J.: Effects of convective momentum transport on the atmospheric circulation in the Community Atmosphere Model, version 3, Journal of Climate, 21, 1487–1499, 2008.

Stephens, G. L., Vane, D. G., Boain, R. J., Mace, G. G., Sassen, K., Wang, Z., Illingworth, A. J., O'Connor, E. J., Rossow, W. B., Durden, S. L., et al.: The CloudSat mission and the A-Train: A new dimension of space-based observations of clouds and precipitation, Bulletin of the American Meteorological Society, 83, 1771–1790, 2002.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, Journal of Geophysical Research: Atmospheres (1984–2012), 106, 7183–7192, 2001.

Trenberth, K. E., Fasullo, J. T., and Kiehl, J.: Earth's global energy budget, Bulletin of the American Meteorological Society, 90, 311–323, 2009.

Wan, H., Rasch, P. J., Zhang, K., Qian, Y., Yan, H., and Zhao, C.: Short ensembles: an efficient method for discerning climate-relevant sensitivities in atmospheric general circulation models, Geoscientific Model Development, 7, 1961–1977, 2014.

Wang, C., Duan, Q., Gong, W., Ye, A., Di, Z., and Miao, C.: An evaluation of adaptive surrogate modeling based optimization with two benchmark problems, Environmental Modelling & Software, 60, 167–179, 2014.

Wang, G. G. and Shan, S.: Review of metamodeling techniques in support of engineering design optimization, Journal of Mechanical Design, 129, 370–380, 2007.

Williams, K. and Brooks, M.: Initial tendencies of cloud regimes in the Met Office Unified Model, Journal of Climate, 21, 833–840, 2008. Williams, P. D.: Modelling climate change: the role of unresolved processes, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 363, 2931–2946, 2005.

Winker, D. M., Vaughan, M. A., Omar, A., Hu, Y., Powell, K. A., Liu, Z., Hunt, W. H., and Young, S. A.: Overview of the CALIPSO mission and CALIOP data processing algorithms, Journal of Atmospheric and Oceanic Technology, 26, 2310–2323, 2009.

Xie, S., Zhang, M., Boyle, J. S., Cederwall, R. T., Potter, G. L., and Lin, W.: Impact of a revised convective triggering mechanism on Community Atmosphere Model, version 2, simulations: Results from short-range weather forecasts, Journal of Geophysical Research:Atmospheres (1984–2012), 109, 2004.

Xie, S., Ma, H.-Y., Boyle, J. S., Klein, S. A., and Zhang, Y.: On the correspondence between short-and long-time-scale systematic errors in CAM4/CAM5 for the year of tropical convection, Journal of Climate, 25, 7937–7955, 2012.

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., and Liu, X.: 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, 395–415, 2013.

Zhang, G. J. and McFarlane, N. A.: Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian Climate Centre general circulation model, Atmosphere-Ocean, 33, 407–446, 1995.

Zhang, M., Lin, W., Bretherton, C. S., Hack, J. J., and Rasch, P. J.: A modified formulation of fractional stratiform condensation rate in the NCAR Community Atmospheric Model (CAM2), Journal of Geophysical Research: Atmospheres (1984–2012), 108, ACL–10, 2003.

Zhang, T., Li, L., Lin, Y., Xue, W., Xie, F., Xu, H., and Huang, X.: An automatic and effective parameter optimization method for model tuning, Geoscientific Model Development, 8, 3579–3591, 2015.

Zhao, C., Liu, X., Qian, Y., Yoon, J., Hou, Z., Lin, G., McFarlane, S., Wang, H., Yang, B., Ma, P.-L., et al.: A sensitivity study of radiative fluxes at the top of atmosphere to cloud-microphysics and aerosol parameters in the community atmosphere model CAM5, Atmospheric Chemistry and Physics, 13, 10 969, 2013.

Table 1. A summary of parameters to be tuned in CAM5. The default and final tuned optimal values are shown, as well as the valid ranges.

| Parameter | Description | Default | Range | Tuned |
|---|---|---|---|---|
| zmconv_c0_lnd | Autoconversion coefficient over land in ZM deep convection scheme | $5.90e^{-3}$ | $2.95e^{-3}$-$8.85e^{-3}$ | $5.35e^{-3}$ |
| zmconv_c0_ocn | Autoconversion coefficient over ocean in ZM deep convection scheme | $4.50e^{-2}$ | $2.25e^{-2}$-$6.75e^{-2}$ | $6.48e^{-2}$ |
| zmconv_tau | Time scale for consumption rate of CAPE for deep convection in ZM deep convection scheme | 3600 | 1800-6400 | 4010 |
| cldfrc_rhminh | Minimum rh for high stable clouds | 0.8 | 0.6-0.9 | 0.661 |
| cldfrc_rhminl | Minimum rh for low stable clouds | 0.896 | 0.8-0.95 | 0.913 |
| cldsed_ai | cloud ice sedimentation velocity | 700 | 300-1100 | 300 |

Table 2. The selected output variables of CAM5 included in the performance metrics and the sources of the corresponding observations

| Variable | Full name | Observation |
|---|---|---|
| LWCF | Longwave cloud forcing | ISCCP |
| SWCF | Shortwave cloud forcing | ISCCP |
| PRECT | Total precipitation rate | GPCP |
| Q850 | Specific Humidity at 850hPa | NCEP |
| T850 | Temperature at 850hPa | NCEP |

Table 3. The optimal improvement index of each variable and total comprehensive metric of CAPT run and AMIP run.

| Variable | Hindcasts | AMIP |
|---|---|---|
| Total metrics | 0.903 | 0.895 |
| LWCF | 0.556 | 0.496 |
| SWCF | 1.069 | 1.004 |
| PRECT | 0.921 | 0.841 |
| Q850 | 1.013 | 1.189 |
| T850 | 0.956 | 0.947 |

Table 4. The percentage biases of the ten fields between the Default/Tuned and their reference observations

| Bias \|\|%\|\| | Default | Tuned |
|---|---|---|
| Sea Level Pressure (ERAI) | 0.007 | 0.004 |
| SW Cloud Forcing (ISCCP) | 3.603 | 5.116 |
| LW Cloud Forcing (ISCCP) | 17.607 | 8.643 |
| Land Rainfall (30N-30S, GPCP) | 7.466 | 7.944 |
| Ocean Rainfall (30N-30S, GPCP) | 30.048 | 25.284 |
| Land 2-m Temperature (Willmott) | 0.128 | 0.175 |
| Pacific Surface Stress (5N-5S, ERS) | 17.866 | 17.295 |
| Zonal Wind (300mb, ERAI) | 7.341 | 7.068 |
| Relative Humidity (ERAI) | 11.383 | 11.610 |
| Temperature (ERAI) | 0.502 | 0.408 |

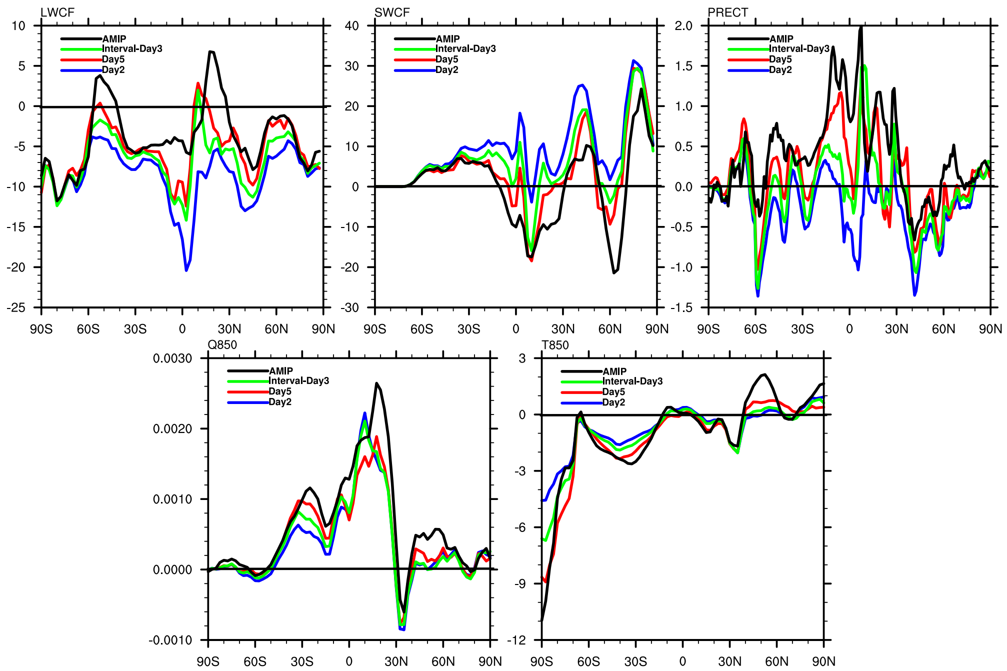


Figure 1. The comparison between short-term hindcasts and long-term AMIP. The Y-axis is bias between the simulations and the observations. The black line is the July mean state from 2000 to 2004 of AMIP simulations. The blue, red, and green lines represent the second day hindcast (labeled as Day2), the fifth day hindcast (labeled as Day5), and the interval-Day3 hindcasts, respectively for July 2009.

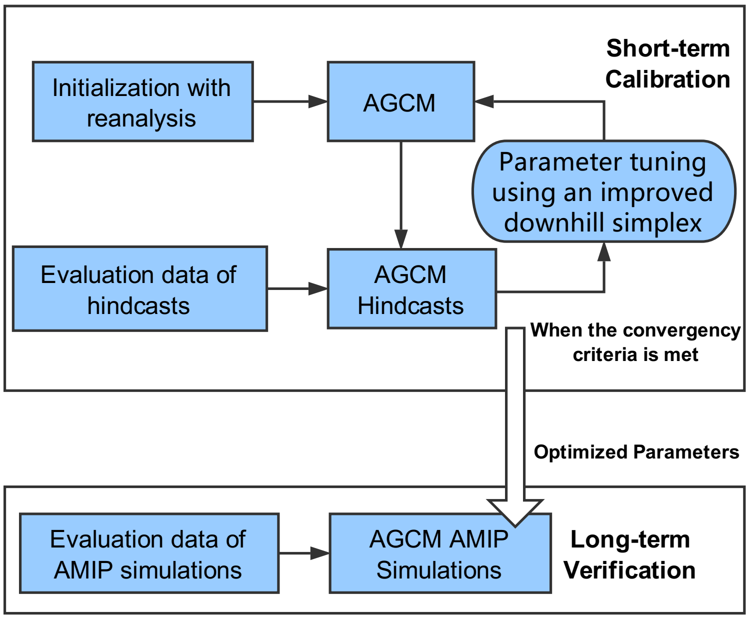


Figure 2. Flow diagram of the automatic calibration of parameters via the short-term CAPT and the verification of optimized parameters through long-term AMIP simulations.

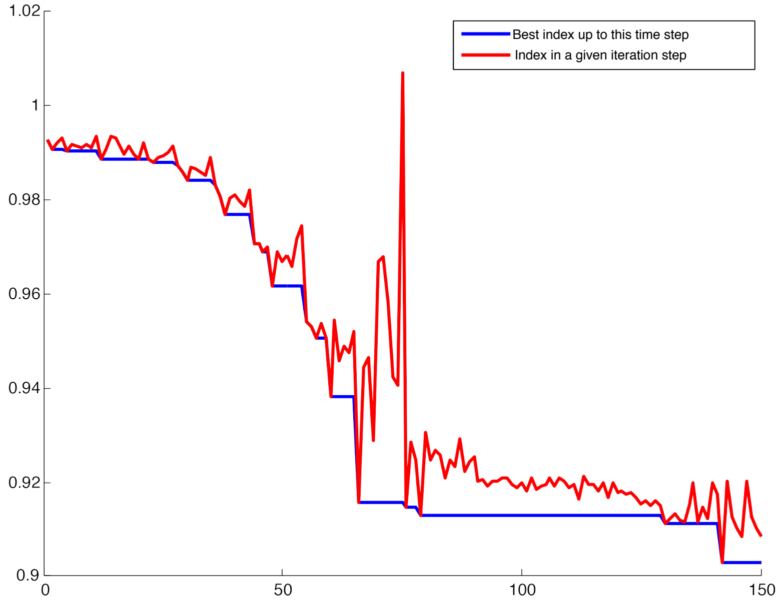


Figure 3. The change of performance index in the optimization iterations. The X-axis is the optimization iterations. The Y-axis is the improvement index in Eq.3. The red line is the index in a given iteration step, while the blue line is the best index up to this time step.

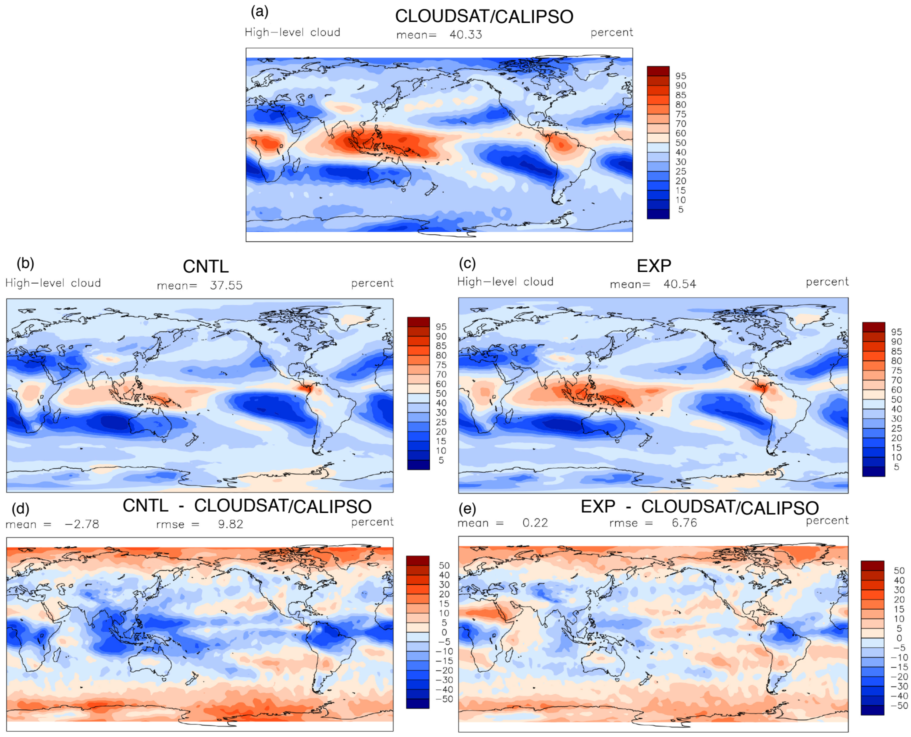


Figure 4. The spatial distribution of high cloud amount in (a) observation, (b) CNTL, (c) EXP, (d) CNTL minus observation, (e) EXP minus observation.

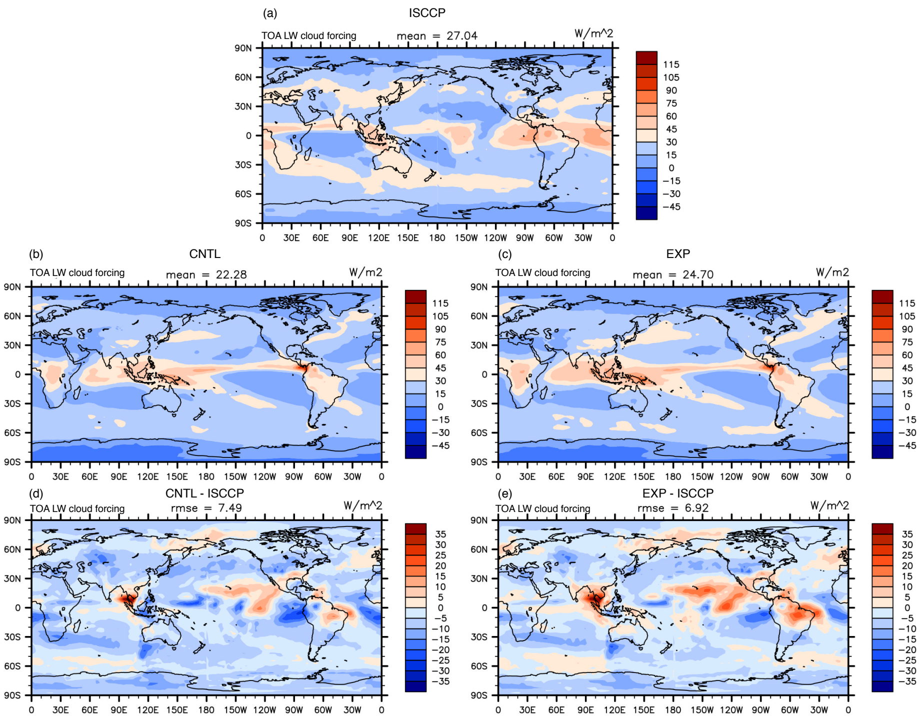


Figure 5. Same as Figure 4 except for LWCF.

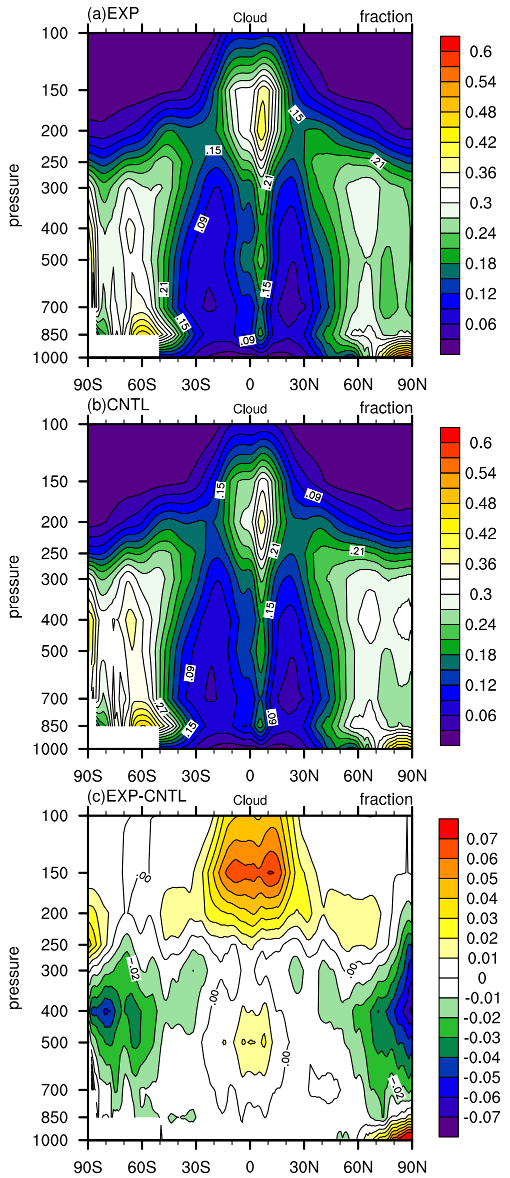


Figure 6. Pressure-latitude distributions of cloud fraction of in (a) EXP, (b) CNTL, and (c) EXP - CNTL.

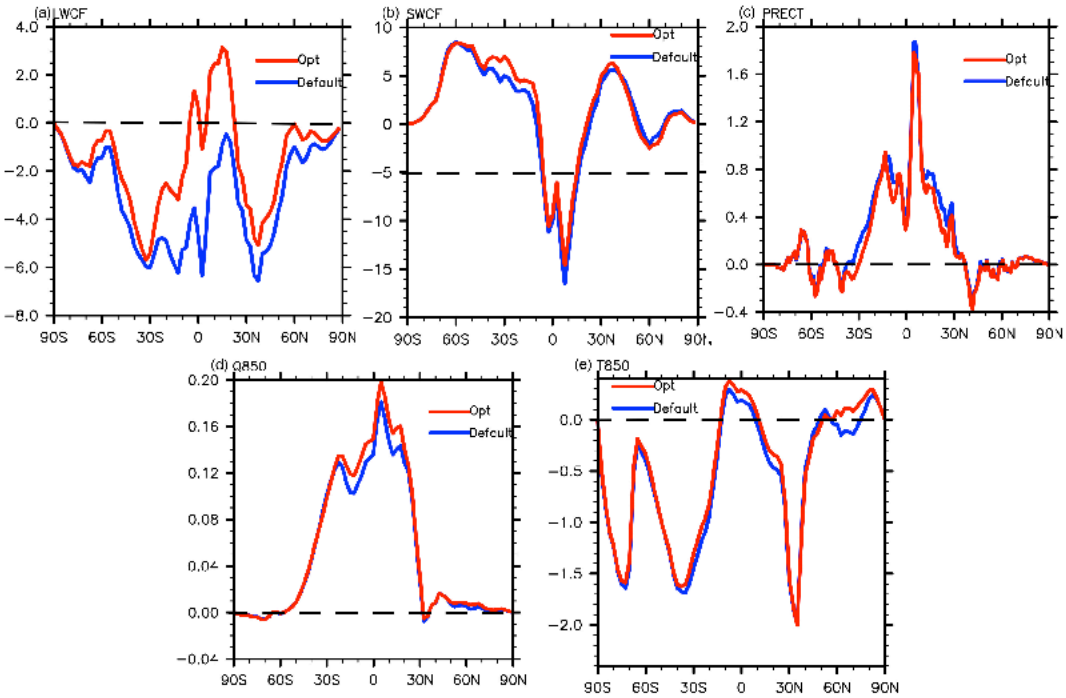


Figure 7. Meridional distribution of the AMIP difference between EXP/CNTL and observations of LWCF (a), SWCF (b), PRECT (c), Q850 (d), and T850 (e). The red line is the output variable of EXP. The blue line is the output variable of CNTL.