Geoscientific
Model Development
Discussions

Open Access

An automatic and effective parameter optimization method for model tuning

T. Zhang[1,2], L. Li[3], Y. Lin[2], W. Xue[1,2], F. Xie[3], H. Xu[1], and X. Huang[1]

[1]Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
[2]Center for Earth System Science, Ministry of Education Key Laboratory for Earth System Modeling, Tsinghua University, Beijing 100084, China
[3]State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

Correspondence to: W. Xue (xuewei@mail.tsinghua.edu.cn)

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Abstract

Physical parameterizations in General Circulation Models (GCMs), having various uncertain parameters, greatly impact model performance and model climate sensitivity. Traditional manual and empirical tuning of these parameters is time consuming and ineffective. In this study, a "three-step" methodology is proposed to automatically and effectively obtain the optimum combination of some key parameters in cloud and convective parameterizations according to a comprehensive objective evaluation metrics. Different from the traditional optimization methods, two extra steps, one determines parameter sensitivity and the other chooses the optimum initial value of sensitive parameters, are introduced before the downhill simplex method to reduce the computational cost and improve the tuning performance. Atmospheric GCM simulation results show that the optimum combination of these parameters determined using this method is able to improve the model's overall performance by 9 %. The proposed methodology and software framework can be easily applied to other GCMs to speed up the model development process, especially regarding unavoidable comprehensive parameters tuning during the model development stage.

1 Introduction

Due to their current relatively low model resolutions, General Circulation Models (GCMs) need to parameterize various sub-grid scale processes. However, due to the complexities involved in these processes, parameterizations representing sub-grid scale physical processes unavoidably involve some empirical or statistical parameters (Hack et al., 1994), especially within cloud and convective parameterizations. Physical parameterizations aim to approximate the overall statistical outcomes of various sub-grid scale physics (Williams, 2005). Consequently, these parameterizations introduce uncertainties to climate simulations using climate system models (Warren and Schneider, 1979). In general, these uncertain parameters need to be calibrated or

constrained when new parameterization schemes are developed and integrated into models (Li et al., 2013).

Traditionally, the uncertain parameters are manually tuned by comprehensive comparisons of model simulations with available observations. Such an approach is subjective, labor intensive, and hard to be extended (Hakkarainen et al., 2012; Allen et al., 2000). By contrast, the automatic parameter calibration techniques have progressed quickly because of their efficiency, effectiveness and broader applications (Bardenet et al., 2013; Elkinton et al., 2008; Jakumeit et al., 2005; Chen et al., 1999). In previous studies applying to GCMs, the methods can be categorized into three major types based on probability distribution function (PDF) method, optimization algorithms, and data assimilation techniques.

For the PDF method, the confidence ranges of the optimization parameters are evaluated based on likelihood and Bayesian estimation. Cameron et al. (1999) improves the forecast by the generalized likelihood uncertainty estimation (GLUE) (Beven and Binley, 1992), a method obtaining parameter uncertain ranges of a specific confidence level. The Bayesian Markov Chain Monte Carlo (MCMC) (Gilks, 1995) is widely used to obtain posterior probability distributions from prior knowledge. A couple of specific algorithms based on the MCMC theory are used to calibrate models in the previous literatures, such as Metropolis–Hasting Sun et al. (2013), adaptive Metropolis (AM) algorithm Hararuk et al. (2014), and multiple very fast simulated annealing (MVFSA) Jackson et al. (2008). The MVFSA method is one to two orders of magnitude faster than the Metropolis–Hasting algorithm (Jackson et al., 2004). However, these methods only attempt to determine the most likely area and cannot directly give the best combination of uncertain parameters with a minimum metrics value. Moreover, the posterior distribution heavily depends on the likelihood function assumed, which is usually difficult to determine for climate system model tuning problem.

Optimization algorithms can be used to search the maximum or minimum metrics value in a given parametric space. Severijns and Hazeleger (2005) calibrates parameters of radiation, clouds, and convection in Speedy with downhill simplex (Press et al.,

1992; Nelder and Mead, 1965) to improve the radiation budget at the top of the at-mosphere and at the surface, as well as the large scale circulation. Downhill simplex is a fast convergence algorithm when the parametric space is not high. However, it is a lo-cal optimization algorithm, not aiming to find the global optimal solution. Moreover, the

5 algorithm has convergence issue when the simplex becomes ill-conditioned. Besides downhill simplex, a few global optimization algorithms are introduced to tune uncer-tain parameters of climate system models, such as simulated stochastic approximation annealing (SSRR) Yang et al. (2013), MVFSA Yang et al. (2014), and multi-objective particle swarm optimization (MOPSO) Gill et al. (2006). SSRR requires at least ten

10 thousands of steps to get a stable solution (Liang et al., 2013), and MVFSA also re-quires thousands of steps (Jackson et al., 2004). MOPSO needs dozens of individual cases in each iteration. All these global optimization algorithms lead to large number of model runs and very high computational cost during model tuning process.

Data assimilation method has been well addressed for state estimation, which is

15 also regarded as a potential solution for parameter estimation. Aksoy et al. (2006) esti-mates the parameter uncertainty of the NCAR/PSU Mesoscale Model version 5 (MM5) (Haagenson et al., 1994) using the Ensemble Kalman Filter (ENKF). Santitissadeekorn and Jones (2013) presents a two-step filtering for the joint state-parameter estimation with a combination method of particle filtering (PF) and ENKF. ENKF and PF have the

20 difficulty in looking for the representative samples. Moreover, same as the MOPSO method, they require a large number of individual samples in each iteration with greatly increased computational cost.

Climate system model is a strongly nonlinear system, having large number of un-certain parameters. As a result, the parameter space of a climate system model is

25 high-dimensional, multi-modal, strongly nonlinear, unseparable. The above mentioned methods generally require long iterations for convergence. More seriously, one sample run of a climate system model might require tens or even hundreds years of simulation to get scientifically meaningful results.

To overcome these challenges, we propose a "three-step" strategy to calibrate the uncertain parameters in climate system models effectively and efficiently. First, a global sensitivity analysis method, Morris (Morris, 1991; Campolongo et al., 2007), is chosen to eliminate the insensitive parameters by analyzing the main and interaction effects among parameters. Another global method by Sobol (Sobol, 2001) is used to validate the results of Morris. Second, a pre-processing of initial values of selected parameters is presented to accelerate the optimization and to resolve the issue of ill-conditioned problem. Finally, the downhill simplex algorithm is used to solve the optimization problem because of its low computational cost and fast convergence for low dimension space. Taking into account the complex configuration and manipulation of model tuning, an automatic workflow is designed and implemented to make the calibration process more efficient. This is result already. The method and workflow can be easily applied to GCMs to speed up model development process.

The paper is organized as follows. Section 2 introduces the automatic workflow proposed. Section 3 describes the details of the example model, reference data, and calibration metrics. The three-step calibration strategy is presented in Sect. 4. Section 5 evaluates the calibration results, followed by a summary and discussion in Sect. 6.

2 The end-to-end automatic calibration workflow

We design a software framework for the overall control of the tuning practice. This framework can automatically execute any part of our proposed "three-step" calibration strategy, determine the optimal parameters and produce its corresponding diagnostic results. It incorporates various tuning methods and facilitate model tuning process with minimal manual management. It effectively manages the dependence and calling sequences of various procedures, including parameter sampling, sensitivity analysis and initial value selection, model configuration and running, evaluation of model outputs using user provided reference metrics. Users only need to specify the model to tune, parameters to be tuned with their valid ranges, and the calibration method to use.

There are four main modules within the framework. The scheduler module manages model simulations with the capability for simultaneous runs. It also coordinates different tasks to reduce the contention and improve throughput. Simulation diagnosis and evaluation is included in a post-processing module. The preparation module contains various sensitivity analysis and sampling methods, such as Morris and Sobol, full factorial (FF) (Raktoe et al., 1981), Latin Hypercube (LH) (McKay et al., 1979), Morris one-at-a-time (MOAT) (Morris, 1991), and Central Composite Designs (CCD) (Hader and Park, 1978). The sensitivity analysis is able to eliminate the duplicated samples to reduce unnecessary computing loads. A MCMC method based on adaptive Metropolis–Hastings algorithms is also provided to get the posterior distribution of uncertain parameters. The tuning algorithm module offers various local and global optimization algorithms including the downhill simplex, genetic algorithm, particle swarm optimization, differential evolution and simulated annealing. In addition, all the intermediate metrics and their corresponding parameters within the framework are stored in a MySQL database and can be used for posterior knowledge analysis. More importantly, the workflow is flexible and expandable for easy integration of other advanced algorithms as well as tools like the Problem Solving Environment for Uncertainty Analysis and Design Exploration (PSUADE) (Tong, 2005), Design Analysis Kit for Optimization and Terascale Applications (DAKOTA) (Eldred et al., 2007). Although, uncertainty quantification toolkits, such as PSUADE, DAKOTA, support various calibration and uncertainty analysis methods and pre-defined function interfaces, they cannot organize the above model tuning process effectively.

3 Model description and reference metrics

We use the Grid-point Atmospheric Model of IAP LASG version 2 (GAMIL2) as an example for the demonstration of the workflow and our calibration strategy. GAMIL2 is the atmospheric component of the Flexible Global–Ocean–Atmosphere–Land System Model grid version 2 (FGOALS-g2), which participated in the CMIP5 program.

The horizontal resolution is 2.8° × 2.8°, with 26 vertical levels. GAMIL2 uses a finite difference scheme that conserves mass and energy (Wang et al., 2004). A two-step shape-preserving advection scheme (Yu, 1994) is used for tracer advection. Compared to the pervious version, GAMIL2 has modifications in cloud-related processes (Li et al., 2013), such as the deep convection parameterization (Zhang and Mu, 2005), the convective cloud fraction (Xu and Krueger, 1991), ~~and~~ the cloud microphysics (Morrison and Gettelman, 2008). More details are in Li et al. (2013). Empirical tunable parameters are selected from schemes of deep convection, shallow convection, and cloud fraction schemes (Table 1). Default parameter values are the configuration for the standard version used for CMIP5 experiments.

To save computational cost, atmosphere-only simulations are conducted for 5 years using prescribed seasonal climatology (no interannual variation) of SST and sea ice. Previous studies have shown 5 years of this type of simulation is enough to capture some basic model characteristics. The goal of these sensitivity simulations is not to determine their resemblance to observations, but to compare the results between the control simulation and various tuned simulations.

Model tuning results depend on the reference metrics used. For a simple justification, we use some conventional climate variables for the evaluation. Wind, humidity, and geopotential height are from the European Center for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA) – Interim reanalysis from 1989 to 2004 (Simmons et al., 2007). We use GPCP (Global Precipitation Climatology Project, Adler et al., 2003) for precipitation and ERBE (Earth Radiation Budget Experiment, Barkstrom, 1984) for radiative fields. All observational and reanalysis data are gridded to the same grid as GAMIL2 before the comparison. Note that the evaluation metrics can be extended depending on the model performance requirement.

A comprehensive metrics, including various variables in Table 2, is used to quantitatively evaluate the performance of overall simulation skills (Murphy et al., 2004; Gleckler et al., 2008; Reichler and Kim, 2008). The calibration RMSE is defined as the spatial standard deviation (SD) of the model simulation against observations/re-analysis, as

in Eq. (1) (Taylor, 2001; Yang et al., 2013). For an easy comparison, we normalize the RMSE of each simulation by that of the control simulation. We weight each variable equally and compute the average normalized RMSE, which indicates the overall improvement relative to the control simulation if it is less than 1.

$$(\sigma_m^F)^2 = \sum_{i=1}^{I} w(i)(x_m^F(i) - x_o^F(i))^2 \tag{1}$$

$$(\sigma_r^F)^2 = \sum_{i=1}^{I} w(i)(x_r^F(i) - x_o^F(i))^2 \tag{2}$$

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

$x_m^F(i)$ is the model outputs according to selected ones shown in Table 2. $x_o^F(i)$ is the corresponding observation or reanalysis data. $x_r^F(i)$ is the reference results from CMIP5. w is the weight due to the different grid area. I is the total grid number in model. N^F is the number of the chosen variables.

4   Method

4.1   Global and local optimization method

Parameter tuning for a climate system model is to solve a global optimization problem in theory. However, traditional evolutionary algorithms, such as genetic algorithm (Goldberg et al., 1989), differential evolutionary (DE) (Storn and Price, 1995), and particle swarm optimization (PSO) (Kennedy, 2010), generally require quit a few of iterations to get a stable global solution and need to set a population of individuals in each iteration, leading to high computational cost (Hegerty et al., 2009; Shi and Eberhart,

1999). Model tuning is always a trade-off between performance and computational cost. Therefore, it is critical to get the best possible results with limited numbers of simulations. In this sense, local optimization algorithms are the viable options considering their significantly reduced computational cost.

We choose the downhill simplex method for climate model tuning considering its relatively low computation cost. Downhill simplex searches the optimal solution by changing the shape of a simplex, which represents the optimal direction and step length. A simplex is a geometry, consisting of $N + 1$ vertexes and their interconnecting edges, where N is the number of calibration parameters. One vertex stands for a pair of a set of parameters and their metrics. The new vertex is determined by expanding and shrinking the vertex with the highest metrics value, leading to a new simplex (Press et al., 1992; Nelder and Mead, 1965).

According to tuning GAMIL2, global methods, PSO and DE, give better tuning results compared to the local downhill simplex method, but their computational costs are approximately 4 and 5 times of the downhill simplex method, respectively (Table 3).

To improve the effectiveness and performance of the local downhill simplex method, we propose two important steps to significantly improve its performance. In the first step, the number of tuning parameters is reduced by eliminating the insensitive parameters; In the second step, fast convergence for better solution is achieved by pre-selecting proper initial values before downhill simplex method.

4.2 Parameter sensitivity analysis

The number of uncertain parameters in physical parameterizations of a climate system model is quite large. Most optimization algorithms, such as PSO, downhill simplex, and simulated annealing algorithm (Van Laarhoven and Aarts, 1987), are ineffective in high dimension problems. Iterations for convergence will increase exponentially when tuning more parameters. In addition, climate models generally need a long simulation to have meaningful results. Therefore, solving high dimension parameter tuning problem

suffers from extreme calibration computational cost. Thus, it is necessary to reduce the parameters dimension before the optimization.

The sensitivity analysis can be divided into local and global methods (Gan et al., 2014). The local method only gets the main effect of a parameter by perturbing one parameter value. The linear correlation coefficient can only measure the linear sensitivity, but it cannot present the nonlinear sensitivity. The Morris method (Morris, 1991; Campolongo et al., 2007) is a qualitative global sensitivity method. The advantage of this method is that not only the single parameter sensitivity can be calculated, but also the interactive sensitivity among parameters can be known at the same time.

The sampling strategy is based on MOAT experimental design with relatively less samples required. It only needs $(n + 1) \times M$ samples, where n is the number of calibration parameters and M is the number of trajectories, usually from 10 to 20. Considering the n parameters $x_i (i = 1, \ldots, n)$, normalized to [0,1], the influence of each variable is defined as an elementary effect, shown as Eq. (4), where Δ is the step size for each parameter. The starting point of a trajectory is selected randomly and the next point is chosen by changing one unchanged parameter value at one time in a random order until getting $n + 1$ samples. The mean of $|d_j|$ stands for the main effect of a single parameter, and the standard deviation presents the interactive effect among multiple parameters. Therefore, those parameters with a low mean and low standard deviation is regard as the insensitive ones for the metrics and will be eliminated during the following optimization step.

$$d_{ij} = \frac{y(X_1, \ldots, X_j + \Delta, \ldots, X_N) - y(X_1, \ldots, X_j, \ldots, X_N)}{\Delta} \tag{4}$$

$$\mu_j = \text{AVG}(|d_{i,j}|), \sigma_j = \text{SD}(d_{i,j}) \tag{5}$$

Taking GAMIL2 as an example, tunable parameters in Table 2 are required to perform sensitivity analysis. We perform 80 samples, and the results are shown in Fig. 1. The insensitivity parameters, ke, capelmt, and c0 of shallow convection, will not be taken into consideration in the next step.

The parameter elimination step is critical for the final result of model tuning. To validate the results by Morris, we compare the results with those with Sobol's benchmark method (Sobol, 2001). It is also a quantitative method based on variance decomposition requiring more samples than the Morris, with a higher computation cost. The variance of the model output can be decomposed as Eq. (6), where n is the number of parameters, and V_i is the variance of the ith parameter, and V_{ij} is the variance of the interactive effect between the and jth parameters, and so on. The total sensitivity effect of ith parameter can be presented as Eq. (7), where V_{-i} is the total variance except for the x_i parameter. The Sobol results are shown in Fig. 3. The screened out parameters are the same as those of the Morris.

$$V = \sum_{i=1}^{n} V_i + \sum_{1 \leq i < j \leq n} V_{ij} + \ldots + V_{1,2,\ldots,n} \tag{6}$$

$$S_{T_i} = 1 - \frac{V_{-i}}{V} \tag{7}$$

4.3 Proper initial values selection for downhill simplex

Since the downhill simplex method is a local optimization algorithm, its convergence performance strongly depends on the quality of the initial values. We need to find the

parameter combinations with the smaller metrics around the final solution. Moreover, we have to complete the searching as fast as possible with minimal overhead. For these two objectives, a hierarchical sampling based on the full factor sample method is presented in this paper. The method uses a longer distance to find the candidate re-
gions for the optimal solution first followed by a second round sampling using a smaller distance in the sensitivity range. This simple sampling method is easy to implement and has lower overhead compared to other complex adaptive sampling methods.

At the same time, inappropriate initial values may lead to ill-conditioned simplex geometry, which can be found in model tuning. One issue we meet is that some vertexes in downhill simplex optimization may have the same values on one or more parameters. As a result, these parameters are invariant during the optimization by using the downhill simplex method and this leads to poor performance of optimization. Consequently, simplex checking is conducted to keep as many as different values of parameters during looking for initial values. Well-conditioned simplex geometry will increase the parameter freedom for optimization.

These methods mentioned above are presented as the initial value pre-processing of the downhill simplex algorithm. It is noted that samples for looking for initial values sometimes can be the same ones in dimension reduction step. In this case, one model run can be used in the two steps to further reduce computational cost.

4.4 Evaluation of the proposed strategy

First, we compare the performance of DE, PSO and downhill simplex for GAMIL2 tuning by optimization results, convergence and computational cost. In Table 3, PSO gets the best solution. But this global method spends much more computational cost than the local downhill simplex method.

~~Taking into account the bad effectiveness of downhill simplex, we present two other strategies, the proposed three-step, and a "two-step" method only including the initial value pre-processing and downhill simplex method.~~ The downhill simplex in the three-step tunes the sensitive parameters described in Sect. 2.1. The pre-processing of initial

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

values requires extra 25 samples, and the parameter sensitivity analysis with Morris requires 80 samples. In Table 4, the two-step gets a better solution than the "one-step" downhill simplex. It indicates pre-selecting of the proper initial values can remarkably improve the calibration performance. Although the two-step method has the best efficiency, the solution is worse than the three-step method. Meanwhile, the computational cost of all strategies based on the local algorithm are smaller than those of the global methods. With the results in Tables 3 and 4, we can conclude that the proposed three-step method can achieve the best trade-off between accuracy and computational cost.

5   Analysis of model optimal results

This section compares the default simulation and the tuned simulation by three-step method with a focus on the cloud and TOA radiation changes. Table 1 shows the values of the four pairs of sensitive parameters between the default (labeled as CNTL) and optimized simulation (labeled as EXP). Significant change is found for c0, which represents the auto-conversion coefficient in the deep convection scheme, and rhminh, which represents the threshold relative humidity for high cloud appearance. The other two parameters have negligible change of the values before and after the tuning and thus it is expected their impacts on model performance will be accordingly small.

The overall improvement after the tuning from the control simulation can be found in the Taylor diagram (Fig. 4), with improvement for almost all the variables, especially for the meridional winds and mid-tropospheric (400 hPa) humidity. Improvements for other variables are relatively small. The change in terms of the RMSE factor over the globe and three regions (tropics, SH mid- and high-latitude and NH mid- and high-latitude) are shown in Fig. 5. First, radiative fields and moisture are improved over all the four areas. By contrast, wind and temperature field changes are more diverse among different areas. This is partly due to the fact that the tuned parameters have direct impacts on moisture and cloud fields. While wind and temperature fields are indirectly influenced following the cloud and radiative impacts. For example, temperatures over the tropics

become worse compared to the control run. There is an overall improvement in the SH mid- and high-latitude for all variables except for the 200 hPa temperature. Winds and precipitation in the NH mid- and high-latitude become slightly worse in the tuned simulation. Such changes are kind of intriguing and we attempt to relate these changes to the two parameters significantly tuned.

~~With increased auto-conversion coefficient in the deep convection, less condensate is detrained to the environment. As a result, mid- and upper-troposphere is overall drier, especially over the tropics where deep convection dominates the vertical transport of water vapor (Fig. 6a). Although the mid- and upper-troposphere become drier over the tropics, reduced RH threshold for high cloud makes clouds easier to be present. Consequently, middle and high clouds increase over the globe, especially over the mid- and high-latitudes with the largest increase up to 4–5 %. In the tropics, due to the drier tendency induced by the reduced detrainment, high cloud increase is relatively small (2–3 %) compared to the mid- and high-latitudes. Below 800 hPa, low clouds decrease by 2–3 % over the mid- and high-latitudes. The reason for this low cloud reduction is still under investigation.~~

Changes in moisture and cloud fields impact radiative fields. With reference to ERBE, TOA outgoing longwave radiation (OLR) is improved in the mid-latitudes for EXP, but it is degraded over the tropics (Fig. 7a). Compared with the CNTL, middle and high cloud significantly increase in the EXP (Fig. 6). Consequently, it enhances the blocking effect on the longwave upward flux at TOA (FLUT), reducing the FLUT in mid-latitudes of the southern and Northern Hemisphere (Fig. 8a). Clear sky OLR increases for the EXP and this is due to the drier upper troposphere in the EXP (Fig. 6). The decrease in the atmospheric water vapor reduces the greenhouse effect. Therefore, it emits more outgoing longwave radiation and reduces the negative bias of clear sky long wave upward flux at TOA (FLUTC, Fig. 8b). Longwave cloud forcing (LWCF) in the middle and high latitudes is improved due to the improvement of FLUT in this area (Fig. 8c), but improvement in the tropics is negligible due to the cancellation between the FLUT and FLUTC.

TOA clear sky shortwave are the same between the control and the tuned simulation since both simulation has the same surface albedo. With increased clouds, the tuned simulation has smaller TOA shortwave absorbed than the control. Compared with ERBE, the tuned simulation has better TOA shortwave absorbed in the mid- and high-latitudes, but it slightly degrades over the tropics.

6 Conclusions

An effective and efficient three-step method for GCM physical parameter tuning is proposed. Compared with conventional methods, an insensitive parameter reduction step and a proper initial value selection step are introduced before the low cost local optimization method. This effectively reduces the computational cost with an overall good performance. In addition, an automatic parameter calibration workflow is designed and implemented to enhance operational efficiency and to support multiple uncertainty quantification analysis and calibration strategies. Evaluation of the method and workflow by calibrating GAMIL2 model indicates the three-step outperforms the two global optimization methods (PSO and DE) in both effectiveness and efficiency. A better trade-off between accuracy and computational cost is achieved compared with the two-step method and the downhill simplex method. The optimal results of the three-step method demonstrate that most of the variables are improved compared with the control experiment, especially for the radiation related ones. The mechanism analysis are conducted to explain why these radiation related variables have an overall improvement.

Recently, the surrogate-based optimization method has been an active research area. The idea is to approximate the real models by statistical regression methods which can greatly reduce the computational cost. However, the precision of surrogate models cannot meet the requirement for the strong non-linear climate system model, especially for wind fields. Therefore, it is also worth to continue improving the calibration strategies based on real models targeting these difficulties. We plan to evaluate the computationally-cheap surrogate model, further reducing the computational cost

for calibrating the climate system model. In addition, more analyses are needed to better understand the model behavior along with the physical parameters changes.

Algorithm 1 Preprocessing the initial values of Downhill Simplex Algorithm.

//full factorial sample
N=number_of_parameters
sampling_sets={}
for each parameter P_i of N parameters **do**
    sampling_sets+=full_factorial_sampling(P_i_range, number_of_samples)
    //refine full factorial sample in the sensitivity range if needed
    if metrics of the the adjacent same parameter sampling points >= sensitivity_threshold **then**
        sampling_sets+=full_factorial_sampling(P_i_adjacent_parameter_range, refine_ number_of_factors)
    end if
end for
//Initial vertexes with parameters of the $N + 1$ minimum metrics
for each initial V_i of $N + 1$ vertexes **do**
    //get the parameters of the ith minimum metrics
    candidate_init_sets += min(i, sampling_sets)
end for
//make sure the initial simplex geometry is well-conditioned
while one parameter k have the same values in the $N + 1$ sets **do**
    $j = 1$
    //remove the parameter set with the worst metrics from candidate_init_sets
    remove_parameter_set(the parameter set with worse metrics, candidate_init_sets)
    //get the parameters of the $N + 1 + j$th minimum metrics
    candidate_init_sets += min($N + 1 + j$, sampling_sets)
    $j+ = 1$
end while

References

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P.-P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P., and Nelkin, E.: The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present), J. Hydrometeor., 4, 1147–1167, 2003. 3797

Aksoy, A., Zhang, F., and Nielsen-Gammon, J. W.: Ensemble-based simultaneous state and parameter estimation with MM5, Geophys. Res. Lett., 33, L12801, doi:10.1029/2006GL026186, 2006. 3794

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. 3793

Bardenet, R., Brendel, M., Kégl, B., and Sebag, M.: Collaborative hyperparameter tuning, in: Proceedings of the 30th International Conference on Machine Learning (ICML-13), 16–21 June 2013, Atlanta, Georgia, USA, 199–207, 2013. 3793

Barkstrom, B. R.: The earth radiation budget experiment (ERBE), B. Am. Meteorol. Soc., 65, 1170–1185, 1984. 3797

Beven, K. and Binley, A.: The future of distributed models: model calibration and uncertainty prediction, Hydrol. Process., 6, 279–298, 1992. 3793

Cameron, D., Beven, K. J., Tawn, J., Blazkova, S., and Naden, P.: Flood frequency estimation by continuous simulation for a gauged upland catchment (with uncertainty), J. Hydrol., 219, 169–187, 1999. 3793

Campolongo, F., Cariboni, J., and Saltelli, A.: An effective screening design for sensitivity analysis of large models, Environ. Modell. Softw., 22, 1509–1518, 2007. 3795, 3800

Chen, T.-Y., Wei, W.-J., and Tsai, J.-C.: Optimum design of headstocks of precision lathes, Int. J. Mach. Tool. Manu., 39, 1961–1977, 1999. 3793

Eldred, M., Agarwal, H., Perez, V., Wojtkiewicz Jr., S., and Renaud, J.: Investigation of reliability method formulations in DAKOTA/UQ, Struct. Infrastruct. E., 3, 199–213, 2007. 3796

Elkinton, C. N., Manwell, J. F., and McGowan, J. G.: Algorithms for offshore wind farm layout optimization, Wind Engineering, 32, 67–84, 2008. 3793

Gan, Y., Duan, Q., Gong, W., Tong, C., Sun, Y., Chu, W., Ye, A., Miao, C., and Di, Z.: A comprehensive evaluation of various sensitivity analysis methods: a case study with a hydrological model, Environ. Modell. Softw., 51, 269–285, 2014. 3800

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Gilks, W. R.: Markov Chain Monte Carloin Practice, Chapman and Hall/CRC, London, United Kingdom, 1995. 3793

Gill, M. K., Kaheil, Y. H., Khalil, A., McKee, M., and Bastidas, L.: Multiobjective particle swarm optimization for parameter estimation in hydrology, Water Resour. Res., 42, W07417, doi:10.1029/2005WR004528, 2006. 3794

Gleckler, P. J., Taylor, K. E., and Doutriaux, C.: Performance metrics for climate models, J. Geophys. Res.-Atmos., 113, D6, doi:10.1029/2007JD008972, 2008. 3797

Goldberg, D. E., Korb, B., and Deb, K.: Messy genetic algorithms: motivation, analysis, and first results, Complex systems, 3, 493–530, 1989. 3798

Haagenson, P. L., Dudhia, J., Grell, G., and Stauffer, D.: The Penn State/NCAR Mesoscale Model (MM5) Source Code Documentation, Mesoscale and Microscale Meteorology Division, Boulder, Colorado, USA, 1994. 3794

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, J. Geophys. Res.-Atmos., 99, 20785–20813, 1994. 3792

Hader, R. and Park, S. H.: Slope-rotatable central composite designs, Technometrics, 20, 413–417, 1978. 3796

Hakkarainen, J., Ilin, A., Solonen, A., Laine, M., Haario, H., Tamminen, J., Oja, E., and Järvinen, H.: On closure parameter estimation in chaotic systems, Nonlin. Processes Geophys., 19, 127–143, doi:10.5194/npg-19-127-2012, 2012. 3793

Hararuk, O., Xia, J., and Luo, Y.: Evaluation and improvement of a global land model against soil carbon data using a Bayesian Markov chain Monte Carlo method, J. Geophys. Res.-Biogeo., 119, 403–417, 2014. 3793

Hegerty, B., Hung, C.-C., and Kasprak, K.: A comparative study on differential evolution and genetic algorithms for some combinatorial problems, in: Proceedings of 8th Mexican International Conference on Artificial Intelligence, 9–13 November 2009, Guanajuato, Mexico, 88, 2009. 3798

Jackson, C., Sen, M. K., and Stoffa, P. L.: An efficient stochastic Bayesian approach to optimal parameter and uncertainty estimation for climate model predictions, J. Climate, 17, 2828–2841, 2004. 3793, 3794

Jackson, C. S., Sen, M. K., Huerta, G., Deng, Y., and Bowman, K. P.: Error reduction and convergence in climate prediction, J. Climate, 21, 6698–6709, 2008. 3793

Parameter optimization method for model tuning

T. Zhang et al.

Title Page

Abstract	Introduction
Conclusions	References
Tables	Figures

◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Jakumeit, J., Herdy, M., and Nitsche, M.: Parameter optimization of the sheet metal forming process using an iterative parallel Kriging algorithm, Struct. Multidiscip. O., 29, 498–507, 2005. 3793

Kennedy, J.: Particle swarm optimization, in: Encyclopedia of Machine Learning, Springer, New York, USA, 760–766, 2010. 3798

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), Adv. Atmos. Sci., 30, 855–867, 2013. 3793, 3797

Liang, F., Cheng, Y., and Lin, G.: Simulated stochastic approximation annealing for global optimization with a square-root cooling schedule, J. Am. Stat. Assoc., 109, 847–863, 2013. 3794

McKay, M. D., Beckman, R. J., and Conover, W. J.: Comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics, 21, 239–245, 1979. 3796

Morris, M. D.: Factorial sampling plans for preliminary computational experiments, Technometrics, 33, 161–174, 1991. 3795, 3796, 3800

Morrison, H. and Gettelman, A.: A new two-moment bulk stratiform cloud microphysics scheme in the Community Atmosphere Model, version 3 (CAM3). Part I: Description and numerical tests, J. Climate, 21, 3642–3659, 2008. 3797

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. 3797

Nelder, J. A. and Mead, R.: A simplex method for function minimization, Computer J., 7, 308–313, 1965. 3794, 3799

Press, W., Teukolsky, S., Vetterling, W., and Flannery, B.: Numerical Recipes in Fortran, Cambridge Univ. Press, Cambridge, 70 pp., 1992. 3793, 3799

Raktoe, B. L., Hedayat, A., and Federer, W. T.: Factorial Designs, John Wiley & Sons, Hoboken, New Jersey, USA, 1981. 3796

Reichler, T. and Kim, J.: How well do coupled models simulate today's climate?, B. Am. Meteorol. Soc., 89, 303–311, 2008. 3797

Santitissadeekorn, N. and Jones, C.: Two-stage filtering for joint state-parameter estimation, Mon. Weather Rev., accepted, doi:10.1175/MWR-D-14-00176.1, 2013. 3794

Severijns, C. and Hazeleger, W.: Optimizing parameters in an atmospheric general circulation model, J. Climate, 18, 3527–3535, 2005. 3793

Shi, Y. and Eberhart, R. C.: Empirical study of particle swarm optimization, in: Proceedings of the 1999 Congress on Evolutionary Computation, CEC 99, vol. 3, IEEE, 6–9 July 1999, Washington, DC, USA, 1945–1950, 1999. 3798

Simmons, A., Uppala, S., Dee, D., and Kobayashi, S.: ERA-interim: new ECMWF reanalysis products from 1989 onwards, ECMWF Newsl., 110, 25–35, 2007. 3797

Sobol, I. M.: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Math. Comput. Simulat., 55, 271–280, 2001. 3795, 3801

Storn, R. and Price, K.: Differential Evolution – a Simple and Efficient Adaptive Scheme for Global Optimization over Continuous Spaces, ICSI Berkeley, Berkeley, California, USA, 1995. 3798

Sun, Y., Hou, Z., Huang, M., Tian, F., and Ruby Leung, L.: Inverse modeling of hydrologic parameters using surface flux and runoff observations in the Community Land Model, Hydrol. Earth Syst. Sci., 17, 4995–5011, doi:10.5194/hess-17-4995-2013, 2013. 3793

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res.-Atmos., 106, 7183–7192, 2001. 3798

Tong, C.: PSUADE User's Manual, Lawrence Livermore National Laboratory (LLNL), Livermore, CA, 109 pp., 2005. 3796

Van Laarhoven, P. J. and Aarts, E. H.: Simulated Annealing, Springer, Dordrecht, Netherlands, 1987. 3799

Wang, B., Wan, H., Ji, Z., Zhang, X., Yu, R., Yu, Y., and Liu, H.: Design of a new dynamical core for global atmospheric models based on some efficient numerical methods, S. China Ser. A, 47, 4–21, 2004. 3797

Warren, S. G. and Schneider, S. H.: Seasonal simulation as a test for uncertainties in the parameterizations of a Budyko-Sellers zonal climate model, J. Atmos. Sci., 36, 1377–1391, 1979. 3792

Williams, P. D.: Modelling climate change: the role of unresolved processes, Philos. T. R. Soc. A, 363, 2931–2946, 2005. 3792

Xu, K.-M. and Krueger, S. K.: Evaluation of cloudiness parameterizations using a cumulus ensemble model, Mon. Weather Rev., 119, 342–367, 1991. 3797

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

Parameter optimization method for model tuning

T. Zhang et al.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

in the CAM5 Zhang-McFarlane convection scheme and impact of improved convection on the global circulation and climate, J. Geophys. Res.-Atmos., 118, 395–415, 2013. 3794, 3798

Yang, B., Zhang, Y., Qian, Y., Huang, A., and Yan, H.: Calibration of a convective parameterization scheme in the WRF model and its impact on the simulation of East Asian summer monsoon precipitation, Clim. Dynam., 1, 1–24, 2014. 3794

Yu, R.: A two-step shape-preserving advection scheme, Adv. Atmos. Sci., 11, 479–490, 1994. 3797

Zhang, G. J. and Mu, M.: Effects of modifications to the Zhang-McFarlane convection parameterization on the simulation of the tropical precipitation in the National Center for Atmospheric Research Community Climate Model, version 3, J. Geophys. Res.-Atmos., 110, D09109, 10.1029/2004JD005617, 2005. 3797

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Table 1. Initial selected uncertain parameters in GAMIL2 and their optimal values in EXP.

Parameter	Description	Default	Range	Optimal
c0	rain water autoconversion coefficient for deep convection	3.0×10^{-4}	$1. \times 10^{-4}$–5.4×10^{-3}	5.427294×10^{-4}
ke	evaporation efficiency for deep convection	7.5×10^{-6}	5×10^{-7}–5×10^{-5}	–
capelmt	threshold value for cape for deep convection	80	20–200	–
rhminl	threshold RH for low clouds	0.915	0.8–0.95	0.917661
rhminh	threshold RH for high clouds	0.78	0.6–0.9	0.6289215
c0_shc	rain water autoconversion coefficient for shallow convection	5×10^{-5}	3×10^{-5}–2×10^{-4}	–
cmftau	characteristic adjustment time scale of shallow cape	7200	900–14 400	7198.048

Table 2. Model output variables and evaluation data in the metrics.

Variable	Observation	Variable	Observation
Meridional wind at 850 hPa	ECMWF	Geopotential Z at 500 hPa	ECMWF
Meridional wind at 200 hPa	ECMWF	Total precipitation rate	GPCP
Zonal wind at 850 hPa	ECMWF	Long-wave cloud forcing	ERBE
Zonal wind at 200 hPa	ECMWF	Short-wave cloud forcing	ERBE
Temperature at 850 hPa	ECMWF	Long-wave upward flux at TOA	ERBE
Temperature at 200 hPa	ECMWF	Clearsky long-wave upward flux at TOA	ERBE
Specific Humidity at 850 hPa	ECMWF	Short-wave net flux at TOA	ERBE
Specific Humidity at 400 hPa	ECMWF	Clearsky short-wave net flux at TOA	ERBE

Table 3. Comparison with local and global algorithms. Downhill simplex is a local method. We use "Downhill_1_step" represents the traditional downhill simplex method, distinguished from our proposed optimal strategies based on the downhill simplex. PSO and DE are the global methods. Optimal solution is the final optimal result. N_{step} is the total numbers of calibrating iteration for convergence. N_{size} is the size of population of the global algorithms. Core hours is computed by $N_{step} \times N_{size} \times$ numbers of process × hours of 5 years simulation. In GAMIL2 case, each model run takes 6 h and uses 30 cores.

	Optimal solution	N_{step}	N_{size}	Core hours
Downhill_1_step	0.9585	80	1	14 400
PSO	0.911537	24	12	51 840
DE	0.942148	33	12	71 280

Table 4. Comparison with optimal strategies based on the downhill simplex. The initial values pre-process is applied to Downhill_2_steps and Downhill_3_steps with extra 25 samples. In the Downhill_3_steps, a step of parameter sensitivity process is conducted before the initial values pre-processing with extra 80 samples.

	Optimal solution	N_{step}	N_{size}	Core hours
Downhill_1_step	0.9585	80	1	14 400
Downhill_2_steps	0.9256899	25 + 34	1	10 620
Downhill_3_steps	0.9098545	80 + 25 + 50	1	27 900

Figure 1. The structure of the automatic calibration workflow. The input of the workflow is the parameters of interest and their initial value ranges. The output is the optimal parameters and its corresponding diagnostic results after calibration. The preparation module provides the parameter sensitivity analysis. The tuning algorithm module offers local and global optimization algorithms including downhill simplex, genetic algorithm, particle swarm optimization, differential evolution and simulated annealing. The scheduler module schedules as many as cases to run simultaneously and coordinates different tasks over parallel system. The post-processing module is responsible for metrics diagnostics, re-analysis and observational data management.
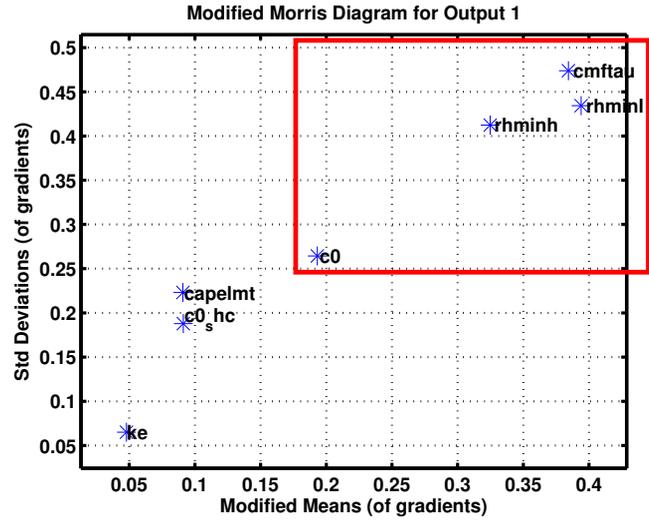
Figure 2. Scatter diagram of Morris sensitivity analysis. The *x* axis stands for the main effect sensitivity of single parameter. The *y* axis stands for the interactive effect among multi-parameters. In GAMIL2 case, c0, rhminl, rhminh, and cmftau have high sensitivity. ke, c0_shc, and capelmt have low sensitivity.
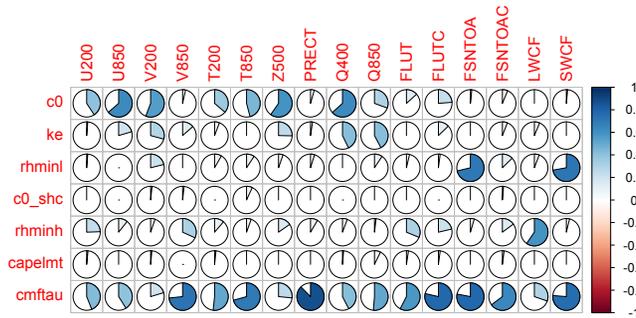
Figure 3. Sobol sensitivity results. The total sensitivity in Eq. (7) is presented by the size of color area. The total sensitivities of ke, c0_shc, and capelmt are not more than 0.5 with regard to each output variable. So they are insensitive.
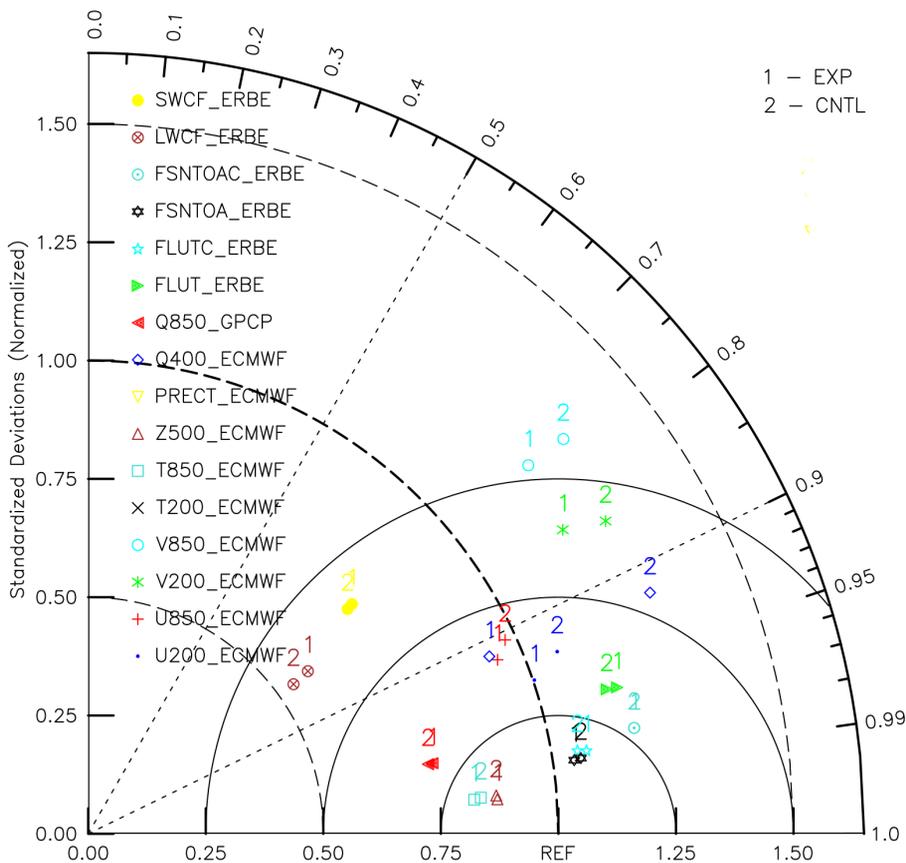
Figure 4. Taylor diagram of the climate mean state of each output variable from 2002 to 2004 of EXP and CNTL.
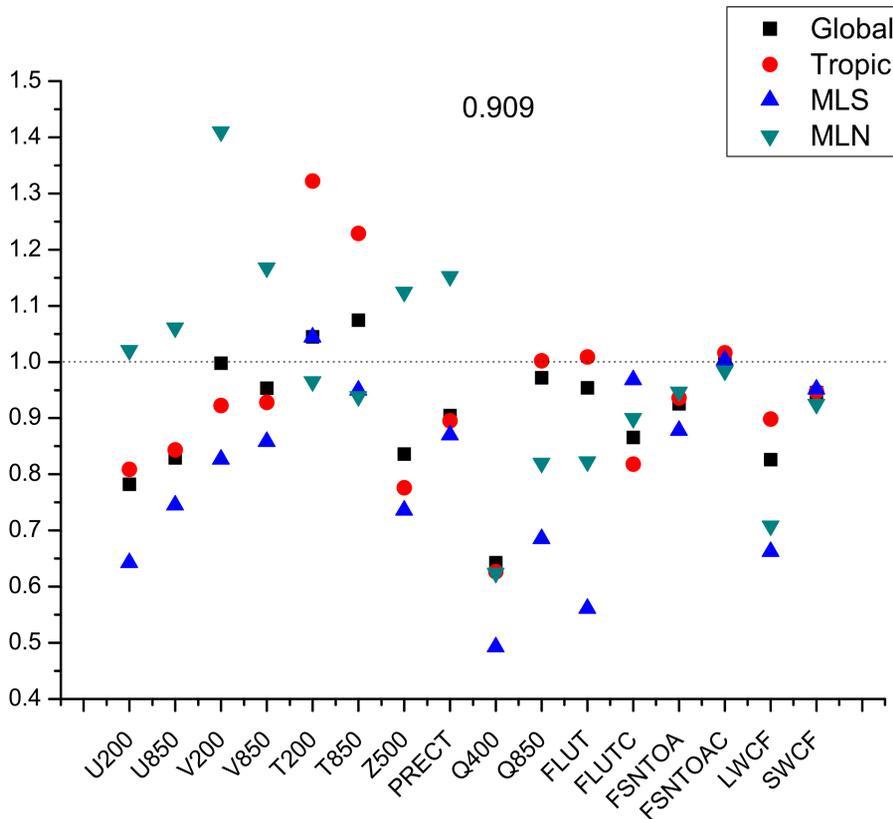
Figure 5. The EXP metrics of each output variable with the global, tropical, and northern/southern mid- and high-latitude areas.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion
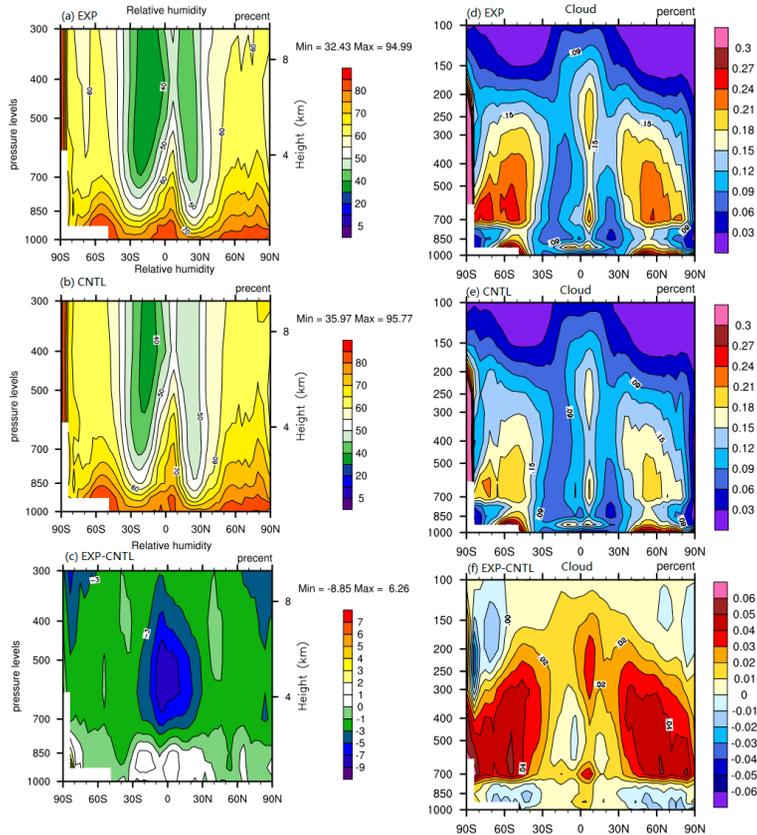
Figure 6. Pressure–latitude distributions of relative humidity and cloud fraction of EXP **(a, d)**, CNTL **(b, e)**, EXP-CNTL **(c, f)**.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion
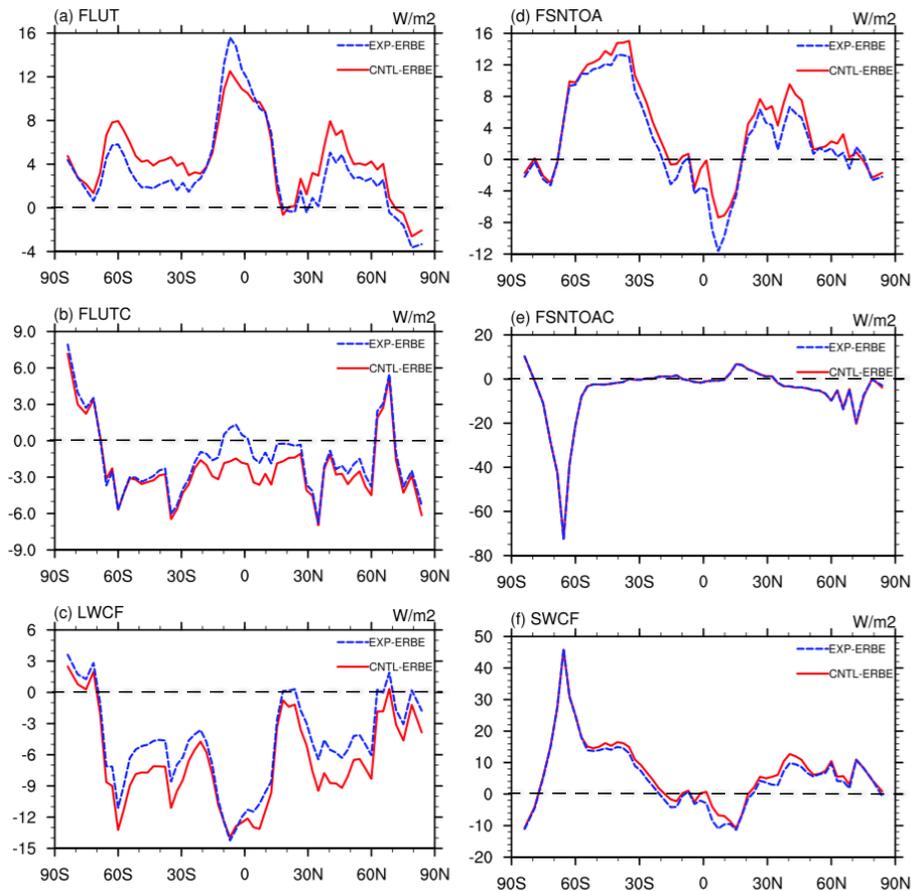
Figure 7. Meridional distributions of the annual mean difference between EXP/CNTL and observations of FLUT **(a)**, FLUTC **(b)**, LWCF **(c)**, FNSTOA **(d)**, FNSTOAC **(e)**, and SWCF **(f)**.

Title Page

Abstract	Introduction
Conclusions	References
Tables	Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion