Manuscript prepared for Geosci. Model Dev.
with version 2014/07/29 7.12 Copernicus papers of the LaTeX class copernicus.cls.
Date: 4 March 2015

An automatic and effective parameter optimization method for model tuning

Tao Zhang[1,2], Lijuan Li[3], Yanluan Lin[2], Wei Xue[1,2], Haoyu Xu[1], Feng Xie[3], and Xin Huang[1]

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

Correspondence to: Lijuan Li (ljli@mail.iap.ac.cn)

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. Results show that the optimum combination of these parameters determined using this method is able to improve model's overall performance by 9% in a AGCM. The method can be easily applied to other GCMs to speed up the model development process, especially regarding unavoidable comprehensive parameters tuning in model development.

1 Introduction

Due to current model resolution, General Circulation Models (GCMs) need to parameterize various sub-grid scale processes. However, due to the complexities involved in these processes, sub-grid scale physical processes are presented as empirical or statistical parameters in climate system models (Hack et al., 1994). There are various empirical determined parameters, 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 new uncertain parameters are required to be calibrated or constrained when new parameterization schemes are integrated into models (Li et al., 2013).

Traditionally, the uncertain parameters are manually tuned by a comprehensive comparison 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). Currently, the automatic parameter calibration technique is a hot topic in uncertainty quantification of climate system models. Previous works focus on the methods of posterior distribution and probability, optimization algorithm, and data assimilation technique.

For the first class method, the confidence range of the optimization parameters is 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 parameters uncertain range of a specific confidence level. The Bayesian Markov Chain Monte Carlo (MCMC) (Gilks, 2005) is widely used to obtain posterior probability distributions from prior knowledge. Sun et al. (2013) demonstrates the possibility of calibration of the hydrologic process in the Community Land Model version 4 (CLM4) (Lawrence et al., 2011; Lawrence and Chase, 2007) with a Metropolis-Hasting algorithm, based on the MCMC approach. Hararuk et al. (2014) calibrates soil C data in the Community Land Model coupled with Carnegie-Ames-Stanford Approach biogeochemistry submodel (CLM-CASA) (Oleson et al., 2004, 2008; Parton et al., 1993), a global land model consisting of biogeophysical and biogeochemical processes, by using an adaptive Metropolis (AM) algorithm (Gilks, 2005). Jackson et al. (2008) obtains parameters posterior probability from clouds and convection physical process in the Community Atmosphere Model version 3.1 (CAM3.1) (Collins et al., 2004) by multiple very fast simulated annealing (MVFSA) according a comprehensive metrics. The MVFSA method is one to two orders of magnitude faster than the Metropolis-Hasting algorithm (Jackson et al., 2004). However, these methods only try to get the likely area and cannot directly find the best combination of uncertain parameters with a minimum metrics value. Moreover, the posterior distribution heavily depends on assuming the likelihood function, 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 the 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 atmosphere 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 local optimal algorithm, which is difficult to find the global optimal solution. Meanwhile, the algorithm has convergence issue when simplex becomes ill-conditioned. Yang et al. (2013) uses simulated stochastic approximation annealing (SSRR) (Liang et al., 2013) to tune the parameters of the Zhang-McFarlane convection scheme to improve convection on the global circulation and climate. Yang et al. (2014) uses the MVFSA algorithm to calibrate a convective parameterization scheme in the Weather Research and Forecasting (WRF) model (Michalakes et al., 2001). Gill et al. (2006) calibrates the the Sacramento soil moisture accounting (SACSMA) model

(Burnash et al.) by multi-objective particle swarm optimization (MOPSO). SSRR requires at least ten thousands of steps to get a stable solution (Liang et al., 2013), and MVFSA also requires 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, leading to unacceptable high computation cost during model tuning.

Data assimilation method has been well addressed for state estimation, which is also regarded as a potential solution for parameter estimation. Aksoy et al. (2006) estimates the parameter uncertainty of the NCAR/PSU Mesoscale Model version 5 (MM5) (Haagenson et al., 1994) by the Ensemble Kalman Filter (ENKF). Santitissadeekorn and Jones (2013) presents two-stage filtering for the joint state-parameter estimation with a combination method of particle filtering (PF) and ENKF. ENKF has the difficulty in looking for the representative samples. Moreover, same as the MOPSO method, ENKF and PF require quite a few of individual samples in each iteration, thus require more computation resources.

Climate system model is a strongly nonlinear system, having large number of uncertain parameters. As a result, the parameter space of climate system model is high-dimensional, multi-modal, strongly nonlinear, and not separable. The mentioned methods require long iterations for convergence. More seriously, one sample run of climate system model might require tens or even hundreds years simulation to get the significant results.

According to these challenges, we propose a “three-step” strategy for calibrating the uncertain parameters in climate system models effectively and efficiently in this paper. In the first step, a global sensitivity analysis method, Morris (Morris, 1991; Campolongo et al., 2007), is chosen to eliminate the insensitivity parameters by analysing the main and interaction effects among parameters. And another global method by Sobol (Sobol, 2001) is used to validate the results of Morris. The downhill simplex algorithm is used to solve the optimization problem because of low computation cost and fast convergence for low dimension parameter problem. A step of pre-processing initial values is presented to perform global optimization and to resolve the issue of ill-conditioned simplex. 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. The three-step calibration strategy is applied to GAMIL2, an AMIP/CMIP5 Atmospheric Model. The experiment results show that the model overall performance can be improved by 9% in GAMIL2 according to the climate mean state. The method and workflow can be easily applied to other GCMs to speed up model development process, especially regarding unavoidable comprehensive parameters tuning in model development.

The remaining part of this paper is organized as follows. Section 2 describes the details of the method and the automatic workflow we propose. The evaluation of the method in GAMIL2 is presented in Section 3 followed by a summary and discussion in Section 4.

2 Method

In this paper, we try to propose an effective and efficient calibration strategy for model tuning problem with high dimension of parameter space. In which, the number of tuning parameters is reduced by eliminating the insensitive parameters during optimization; fast convergence for better solution is achieved by pre-selecting the proper initial values and by using the downhill simplex method as the following optimal algorithm.

2.1 Sensitive parameter determination

The number of uncertain parameters in physical parameterizations of the climate system model is quite large. Most optimization algorithms, such as particle swarm optimization (PSO) (Kennedy, 2010), downhill simplex (Press et al., 1992; Nelder and Mead, 1965), and simulated annealing algorithm are ineffective in high dimension problems. More parameters to tune, the iterations for convergence will increase exponentially. Moreover, some components of the climate system model, such as atmosphere and ocean, require a long time to get significant results. Therefore, solving high dimension parameter tuning problem suffers from extreme calibration computational cost. Thus, it is necessary to reduce the parameters dimension before optimization.

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 got at the same time.

The sampling strategy is based on Morris's one-step-at-a-time (MOAT) experimental design for 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 $x_i (i=1,...,n)$, normalized to [0,1], the influence of each variable is defined as an elementary effect, shown as equation (1), 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 regared as the insensitive ones for metrics and will be eliminated during following optimization step.

$$d_{ij} = \frac{y(X_1,...,X_j+\Delta,...,X_N) - y(X_1,...,X_j,...,X_N)}{\Delta} \tag{1}$$

$$\mu_j = avg(|d_{i,j}|), \sigma_j = stddev(d_{i,j}) \tag{2}$$

Taking GAMIL2 as an example, tuneable parameters in Table 2 are required to perform sensitivity analysis. We perform 80 samples, and the results are shown in Figure 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 got 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 equation (3), where n is the number of parameters, and V_i is the variance of the i_{th} parameter, and V_{ij} is the variance of the interactive effect between the and j_{th} parameters, and so on. The total sensitivity effect of i_{th} parameter can be presented as equation (4), where V_{-i} is the total variance except for the x_i parameter. The Sobol results are shown as Figure 2. The screened out parameters are the same ones as those of the Morris.

$$V = \sum_{i=1}^{n} V_i + \sum_{1 \le i < j \le n} V_{ij} + ... + V_{1,2,...,n} \tag{3}$$

$$S_{T_i} = 1 - \frac{V_{-i}}{V} \tag{4}$$

2.2 Optimization method and initial value selection

Parameter tuning for a climate system model is to solve a global optimization problem in theory. But the common-use evolutionary algorithms, such as genetic algorithm (Goldberg et al., 1989), differential evolutionary (DE) (Storn and Price, 1995), and PSO, generally require at least thousands 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). Therefore, it is the most important thing to get the best possible results with limited iterations as well as model runs for model tuning.

In this paper, we choose the downhill simplex method to tune the uncertain parameters for climate system models with 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 screened in Section 2.1. The vertexes stand for the 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.

Mathematically, downhill simplex method is a local optimization algorithm. Its convergence performance is sensitive according to initial values and is highly related to the quality of initial values. For better 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 for less overhead. For these two objectives, a hierarchical sampling based on full factor sample method is presented in this paper. In which, we use full factor sample method with longer distance to find the candidate regions for the optimal solution first, and then perform the second round sampling with smaller distance to shrink the interested regions as well to determine the final initial values for following optimization. The simple sampling method is easy to implement and can be regarded as low overhead approximation of complex adaptive sampling.

At the same time, inappropriate initial values may lead to ill-conditioned simplex geometry, which can be found in model tuning. One of the issues we meet is that some parameters keep the same value in the initial values. As a result, these parameters cannot change during optimization by using downhill simplex method, which always leads to poor performance of optimization. Consequently, simplex checking is conducted to keep as many as different parameters values during looking for initial values. Well-conditioned simplex geometry will increase the parameter freedom for optimization and get better convergence. This procedure is presented as the initial values pre-processing of the downhill simplex algorithm.

It is noted that samples for looking for initial values might be the same ones in dimension reduction step. In this case, one model run can be used in the two steps, saving computational resource.

2.3 The end-to-end automatic calibration workflow

To perform model tuning for today's climate system models, there are many operations need to carry out, including parameter sampling, model configuration with different parameters, model runs during sampling and tuning, metrics diagnostics, parameter sensitivity analysis, initial values selection, and the model evaluation with optimal parameters. It is worth designing and developing an automatic workflow for model tuning.

The common uncertainty quantification toolkits, such as 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), support various uncertainty analysis methods, pre-defined function interfaces and corresponding workflows. However, scheduling the model runs and performing sample deduplication, and model evaluation after tuning are still need to take into account in the workflow of model tuning as an integrated system. More importantly, the workflow need to be open enough and can be easily integrated the advanced algorithms as well as tools like PSUADE, DAKOTA.

Following the design philosophy, we design and implement an end-to-end automatic workflow system, which can carry through the proposed "three-step" method easily and efficiently. Users only require to specify the model to tune, the corresponding tuneable parameters of interest and their ranges, as well as the calibration method. The workflow can automatically execute any part of the "three-step" calibration strategy, and can finally produce the optimal parameters and its corresponding diagnostic results.

The architecture of the automatic workflow is presented in Figure 1. It consists of four components, dimension reduction, calibration algorithm, post-processing, and task scheduler. Currently, the dimension reduction module provides Morris and Sobol sensitivity analysis methods and various sampling method, such as full factor, Latin Hypercube, MOAT, and Central Composite Designs. Meanwhile, this module can eliminate duplicate samples for reducing unnecessary computing loads. Moreover, the MCMC method based on adaptive Metropolis-Hastings algorithms is also provided to get the posterior distribution of uncertain parameters. The calibration algorithm module offers local and global optimization algorithms including downhill simplex, genetic algorithm, particle swarm optimization, differential evolution and simulated annealing. To make full use of the computational resource, the scheduler module schedules as many as cases to run simultaneously and coordinates the different tasks for reducing the contention and improving throughput. The post-processing module is responsible for metrics diagnostics, re-analysis and observational data management. In addition, all the intermediate metrics and their corresponding parameters are stored in a MySQL database, and can be used for posterior knowledge analysis.

3 Evaluation with GAMIL2 model

3.1 Model description

In this paper, the Grid-point Atmospheric Model of IAP LASG version 2 (GAMIL2) is used for evaluating our method and workflow. It takes part in the Atmospheric Model Inter-comparison Project (AMIP) of IPCC AR5, the Cloud Feedback Model Inter-comparison Project (CFMIP) and the Coupled Model Intercomparison Project Phase 5 (CMIP5) as the atmospheric component of the Flexible Global-Ocean-Atmosphere-Land System Model grid version 2 (FGOALS-g2). The horizontal resolution is 2.8° x 2.8°, with 26 vertical levels. The dynamical core of GAMIL2 uses a finite difference scheme that conserves mass and effective energy (Wang et al., 2004). The moisture equation adopts the two-step shape-preserving advection scheme (Yu, 1994). Compared with the pervious version GAMIL1, GAMIL2 has modifications on 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). The tuneable parameters are selected from deep convection, shallow convection, cloud fraction, cloud microphysical processes and boundary layer scheme, as shown in Table 1. The default parameter values come from the configuration of the standard version, which is used for AMIP/CMIP5 experiments and then is called as CNTL experiment.

3.2 Evaluation data and metrics

Wind, humidity, and geopotential height are derived from the European Center for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA) - Interim reanalysis, 1989 to 2004, and 1.5°x1.5°

horizontal resolution (Simmons et al., 2007). The precipitation data come from the Global Precipitation Climatology Project (GPCP), 1989 to 2004, and 2.5°x2.5° horizontal resolution (Adler et al., 2003). The radiation variables come from the Earth Radiation Budget Experiment (ERBE), 1985 to 1989, and 1.875°x1.875° horizontal resolution (Barkstrom, 1984). The climate mean state of the observational data are require to remap to the grid of GAMIL2. It is noted that the duration of the simulation (2000-2004) is inconsistent with the observational data. So we conduct a longer simulation (1989-2004) with the optimal parameters to validate the model performance with tuned parameters. The results show the consistent improvements with the short simulation.

A comprehensive metrics, including wind, temperature, humidity, geopotential height, precipitation, and radiation flux is used to quantitatively evaluate the simulation performance of overall simulation skills (Murphy et al., 2004; Gleckler et al., 2008; Reichler and Kim, 2008). These variables are shown in Table 2. The model starts in the 2000th year and simulates 5 years. The climate mean state of the last three years is used to calculate the metrics. The calibration RMSE is defined as the spatial standard deviation (SD) of the model simulation against observations/re-analysis, as in equation (5) (Taylor, 2001; Yang et al., 2013). To integrate the 16 output variables into an unique metrics, the SD of the default GAMIL2 simulations against observations (equation (6)) is used to scale each variable calibration RMSE. Finally, the metrics is computed as the average value of each scaling variable, as equation (7). As a consequence, if the metrics is lower than one, the model performance gets improved.

$$(\sigma_m^F)^2 = \sum_{i=1}^{l} w(i)(x_m^F(i) - x_o^F(i))^2 \tag{5}$$

$$(\sigma_r^F)^2 = \sum_{i=1}^{l} w(i)(x_r^F(i) - x_o^F(i))^2 \tag{6}$$

$$\chi^2 = \frac{1}{N^F} \sum_{F=1}^{N^F} (\frac{\sigma_m^F}{\sigma_r^F})^2 \tag{7}$$

$x_m^F(i)$ is the model outputs according to selected shown in the 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.

3.3 The tuning results

We compare the tuning results with five different methods in Table 4, which include the proposed three-step in this paper, the "two-step" methoood only having initial values pre-processing and downhill simplex method (Tao Zhang et al., Quantification and optimization of parameter uncertainty in the Grid-point Atmospheric Model GAMIL2, manuscript in review, 2014), and three "one-step" optimization algorithms of downhill simplex, DE, and PSO. Among these five mothods, eighty samples for reducing parameter space by Morris are required for the three-step method. The two-step and three-step methods require extra twenty-five samples for the initial values pre-processing. The population size of DE and PSO is set to twelve. The effectiveness of each strategy is evaluated by the final optimal solution, and the efficiency is evaluated by the core hours consumed, which are computed as $N_{step} \times N_{size} \times 30 \times 6$, where N_{step} is the number of iterations and N_{size} is the population size. Every GAMIL2 experiment runs with thirty cores, taking six hours for five–year simulation. Among the "one-step" strategies, PSO gets the best solution, but spends more computational cost than the downhill simplex method. Our three-step method gets the best solution using only half of core hours than PSO. The "two-step" mothod has the best efficiency, while the solution is worse than the "three-step" method. Meanwhile, for the downhill simplex method in "one-step", both of the optimal solution and the computation efficiency are worse than the downhill simplex in "two-step". It indicates pre-selecting the proper initial values can remarkablely improve the calibration performance. With the results in Table 4, we can conclude that our proposed method can get the best trade-off between accuracy and computaton cost.

3.4 The mechanism analysis

Figure 4 exhibits the Taylor diagram of the 16 variables for three years of the climate mean state from 2002 to 2004. No variable of EXP becomes much less than CNTL. Instead, zonal wind at 850 hPa (U850), zonal wind at 200 hPa (U200), meridional wind at 850 hPa (V850), meridional wind at 200 hPa (V200), and specific humidity at 400 hPa (Q400) have been improved a lot. Figure 5 shows the metrics of each variable with the global, tropic, and northern/southern middle and high latitudes (NMHL/SMHL). Most variables of the global area are improved compared with CNTL. Specific humidity at 400 hPa (Q400) is the most improved. Two variables, meridional wind at 200 hPa (V200) and clearsky short wave net flux at TOA (FSNTOAC) keep the same magnitude. Temperature at 850 hPa and 200 hPa are worse than CNTL. From the spatial distribution perspective, the SMHL contributes the most improvement.

Figure 5 presents the improvement of all the radiation variables. This could be attributed to specific humidity and cloud improvement in the middle and upper troposphere. The EXP consumes the more water vapor and achieves a the better simulation than CNTL above 700 hPa, shown in Figure 6. The decrease in the atmospheric water vapor reduces the greenhouse effect. Therefore, it emits more outgoing long-wave radiation with a clear sky, reducing the negative bias of clear sky long wave upward flux at TOA (FLUTC), as shown in Figure 8(a).

Compared with CNTL, middle and high cloud significantly increase, as a result of the reducing rhminh parameter, shown in Figure 7. Consequently, it enhances the blocking effect on the long wave upward flux at TOA (FLUT), reducing the FLUT in 30°/60° of the southern and northern hemisphere, shown in Figure 8(b). Taking into account the long wave cloud forcing (LWCF) computed as $FLUTC - FLUT$, the improvement of FLUTC in the tropics makes up for the increasing bias of FLUT. Therefore, the LWCF in tropics stays the same as CNTL. But the LWCF in the middle and high latitudes is improved due to the improvement of FLUT in this area, as shown in Figure 8(c).

4 Conclusions

In this paper, we present the “three-step” method for tuning uncertain physical parameters in climate system models effectively and efficiently, including dimension reducation step, step of pre-processing initial values, and optimization step with the downhil simplex method. To improve the convergence and the solution of downhill simplex, an initial values pre-processing algorithm is presented to evaluate the likely area of the optimal solution quickly and to resolve the issue of ill-conditioned simplex. Moreover, an automatic parameters calibration workflow is further designed and implemented to further enhance operation efficiency and to support multi uncertainty quantification analysis and calibration strategy. The experiment results over AGCM model of GAMIL2 show the “three-step” outperforms the PSO and DE in both effectiveness and efficiency and has a good trade-off between accuracy and computation cost compared with “two-step” methods and downhill simplex method. The optimal results of the “three- step” method demonstrate most of the variables are improved compared with the CNTL experiment, especially for the radiation related variables. The mechanism analysis demonstrates that the reduced simulation errors of water vapor have a direct effect on FLUT, and via cloud fraction indirectly influences FLUTC. The LWCF is improved following the improved FLUT and FLUTC.

In the future, we plan to evaluate the computation-cheap surrogate model, further reducingthe computational cost for calibrating the climate system model. In addtion, more analysis need to be conducted to well understand the model behavior.

References

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P.-P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., et al.: The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present), Journal of Hydrometeorology, 4, 1147–1167, 2003.

Aksoy, A., Zhang, F., and Nielsen-Gammon, J. W.: Ensemble-based simultaneous state and parameter estimation with MM5, Geophysical research letters, 33, 2006.

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.

Barkstrom, B. R.: The earth radiation budget experiment (ERBE), Bulletin of the American Meteorological Society, 65, 1170–1185, 1984.

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

Burnash, R., Ferral, R., and R.A., M.: A generalized streamflow simulation system, report, Joint Fed.-State River Forecast Cent, Sacramento,Calif.

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), Journal of Hydrology, 219, 169–187, 1999.

Campolongo, F., Cariboni, J., and Saltelli, A.: An effective screening design for sensitivity analysis of large models, Environmental modelling & software, 22, 1509–1518, 2007.

Collins, W. D., Rasch, P. J., Boville, B. A., Hack, J. J., McCaa, J. R., Williamson, D. L., Kiehl, J. T., Briegleb, B., Bitz, C., Lin, S., et al.: Description of the NCAR community atmosphere model (CAM 3.0), Tech. rep., Tech. Rep. NCAR/TN-464+ STR, 2004.

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

Gilks, W. R.: Markov chain monte carlo, Wiley Online Library, 2005.

Gill, M. K., Kaheil, Y. H., Khalil, A., McKee, M., and Bastidas, L.: Multiobjective particle swarm optimization for parameter estimation in hydrology, Water Resources Research, 42, 2006.

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.

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, 1994.

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.

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, Journal of Geophysical Research: Biogeosciences, 119, 403–417, 2014.

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, 2009.

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

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

Kennedy, J.: Particle swarm optimization, in: Encyclopedia of Machine Learning, pp. 760–766, Springer, 2010.

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.

Lawrence, P. J. and Chase, T. N.: Representing a new MODIS consistent land surface in the Community Land Model (CLM 3.0), Journal of Geophysical Research: Biogeosciences (2005–2012), 112, 2007.

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

Liang, F., Cheng, Y., and Lin, G.: Simulated Stochastic Approximation Annealing for Global Optimization with a Square-Root Cooling Schedule, Journal of the American Statistical Association, 2013.

Michalakes, J., Chen, S., Dudhia, J., Hart, L., Klemp, J., Middlecoff, J., and Skamarock, W.: Development of a next generation regional weather research and forecast model, in: Developments in Teracomputing: Proceedings of the Ninth ECMWF Workshop on the use of high performance computing in meteorology, vol. 1, pp. 269–276, World Scientific, 2001.

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

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, Journal of Climate, 21, 3642–3659, 2008.

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.

Nelder, J. A. and Mead, R.: A simplex method for function minimization, The computer journal, 7, 308–313, 1965.

Oleson, K., Niu, G.-Y., Yang, Z.-L., Lawrence, D., Thornton, P., Lawrence, P., Stöckli, R., Dickinson, R., Bonan, G., Levis, S., et al.: Improvements to the Community Land Model and their impact on the hydrological cycle, Journal of Geophysical Research: Biogeosciences (2005–2012), 113, 2008.

Oleson, K. W., Dai, Y., Bonan, G., Bosilovich, M., Dickinson, R., Dirmeyer, P., Hoffman, F., Houser, P., Levis, S., Niu, G.-Y., et al.: Technical description of the community land model (CLM), Tech. rep., NCAR Technical Note NCAR/TN-461+ STR, National Center for Atmospheric Research, Boulder, CO, 2004.

Parton, W., Scurlock, J., Ojima, D., Gilmanov, T., Scholes, R., Schimel, D. S., Kirchner, T., Menaut, J.-C., Seastedt, T., Garcia Moya, E., et al.: Observations and modeling of biomass and soil organic matter dynamics for the grassland biome worldwide, Global biogeochemical cycles, 7, 785–809, 1993.

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

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

Santitissadeekorn, N. and Jones, C.: Two-stage filtering for joint state-parameter estimation, Journal of the American Statistical Association, 2013.

Severijns, C. and Hazeleger, W.: Optimizing parameters in an atmospheric general circulation model, Journal of climate, 18, 3527–3535, 2005.

Shi, Y. and Eberhart, R. C.: Empirical study of particle swarm optimization, in: Evolutionary Computation, 1999. CEC 99. Proceedings of the 1999 Congress on, vol. 3, IEEE, 1999.

Simmons, A., Uppala, S., Dee, D., and Kobayashi, S.: ERA-Interim: New ECMWF reanalysis products from 1989 onwards, ECMWF newsletter, 110, 25–35, 2007.

Sobol, I. M.: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Mathematics and computers in simulation, 55, 271–280, 2001.

Storn, R. and Price, K.: Differential evolution-a simple and efficient adaptive scheme for global optimization over continuous spaces, ICSI Berkeley, 1995.

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, Hydrology and Earth System Sciences, 17, 4995–5011, 2013.

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

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

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, Science in China Series A: Mathematics, 47, 4–21, 2004.

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

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.

Xu, K.-M. and Krueger, S. K.: Evaluation of cloudiness parameterizations using a cumulus ensemble model, Monthly weather review, 119, 342–367, 1991.

Yang, B., Qian, Y., Lin, G., Leung, L. R., Rasch, P. J., Zhang, G. J., McFarlane, S. A., Zhao, C., Zhang, Y., Wang, H., et al.: 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.

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, Climate Dynamics, pp. 1–24, 2014.

Yu, R.: A two—step shape—preserving advection scheme, Advances in Atmospheric Sciences, 11, 479–490, 1994.

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, Journal of Geophysical Research: Atmospheres (1984–2012), 110, 2005.

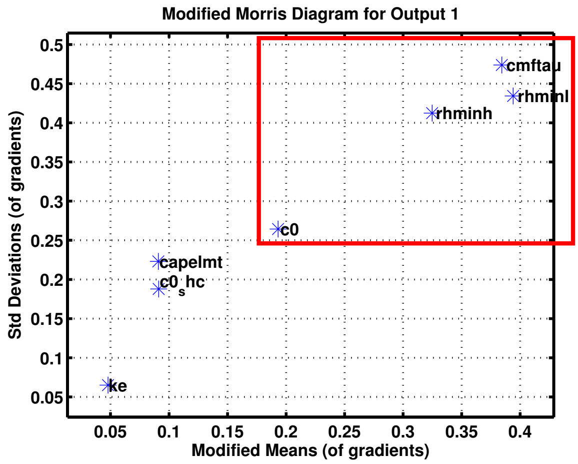

Figure 1. Morris sensitivity scatter diagram. c0, rhminl, rhminh, and cmftau have high sensitivity. ke, c0_shc, and capelmt have low sensitivity.

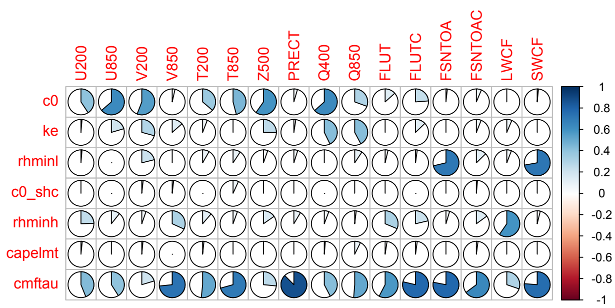

Figure 2. Sobol sensitivity results. The total sensitivity of ke, c0_shc, and capelmt with regard to each variable is not more than 0.5.

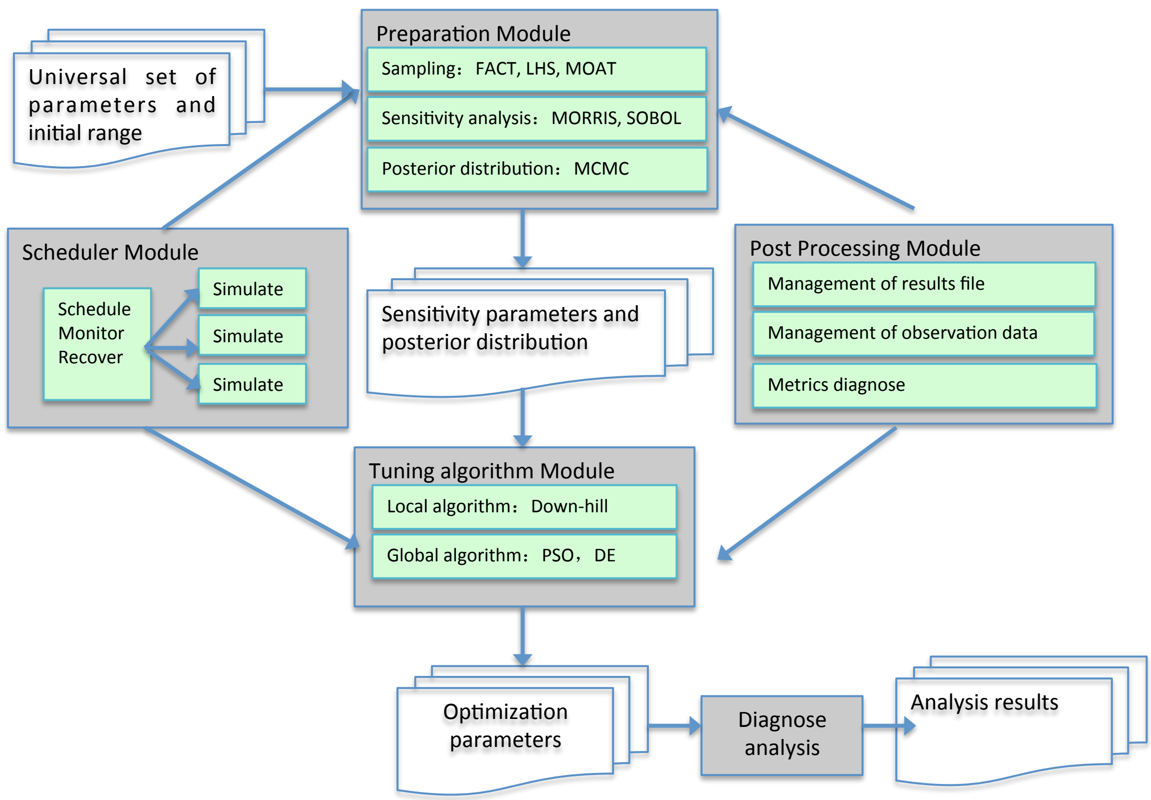

Figure 3. The structure of the automatic calibration workflow.

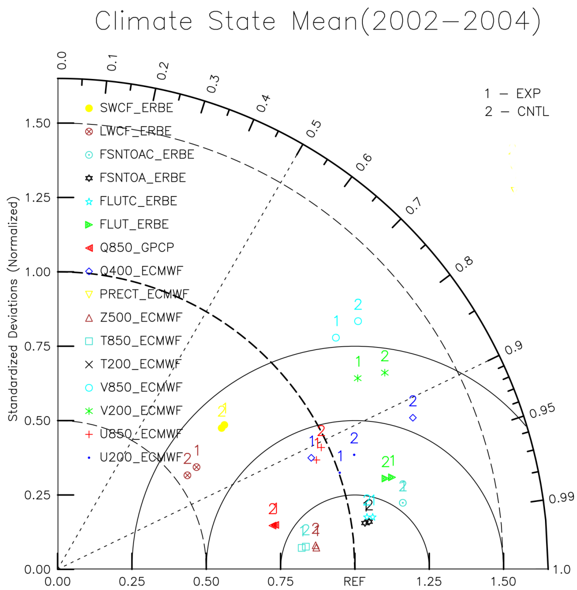

Figure 4. Taylor diagram of the climate mean state of each variable from 2002 to 2004 of EXP and CNTL.

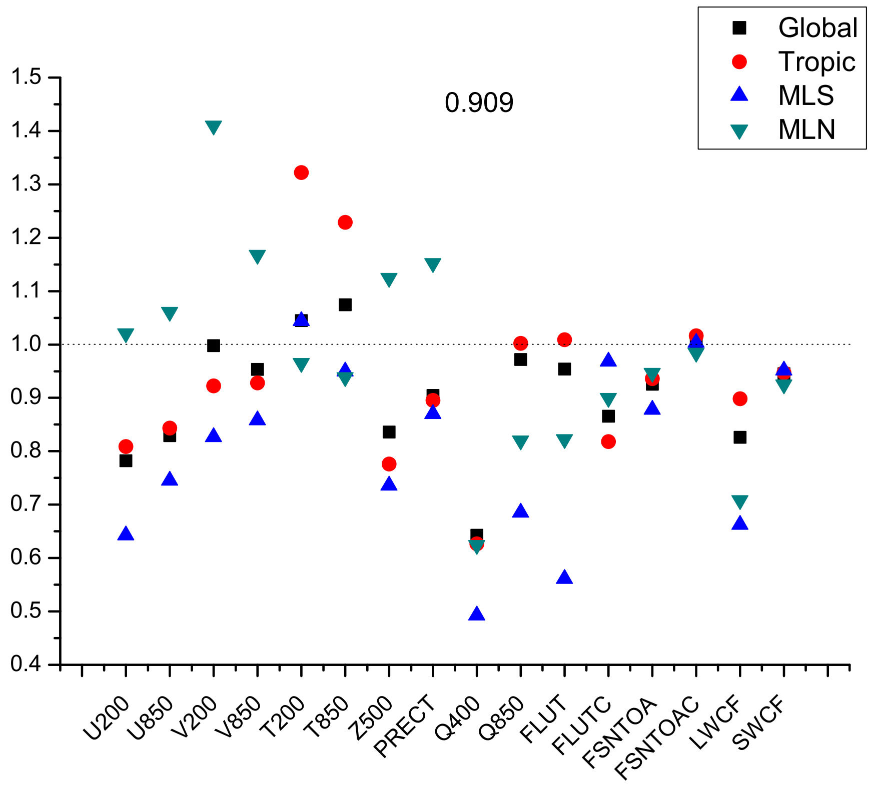

Figure 5. Metrics of each variable with the global, tropical, and northern/southern middle and high latitudes.

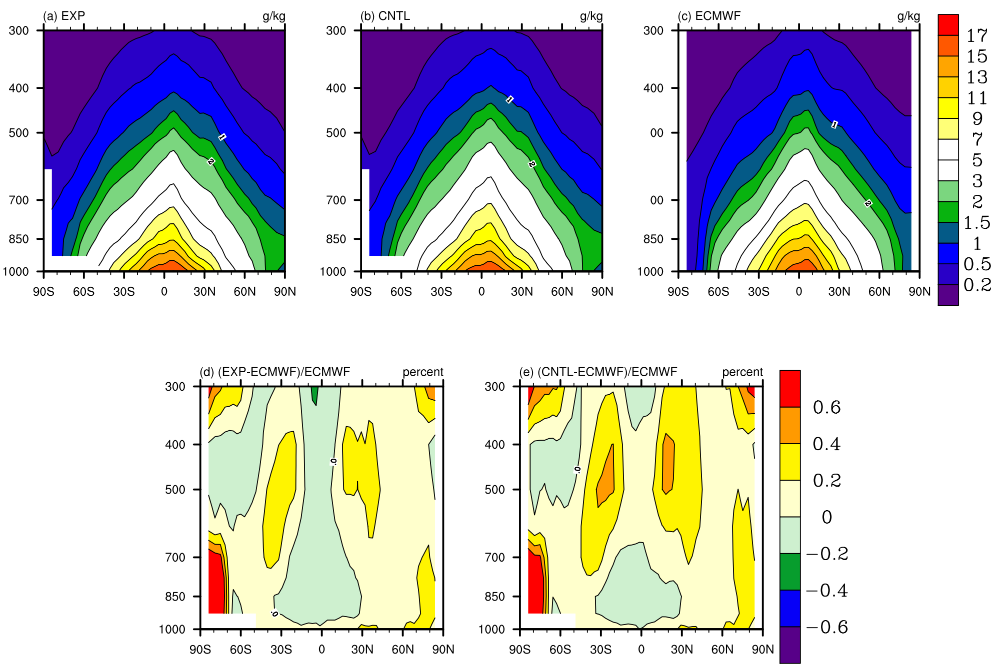

Figure 6. Pressure–latitude distributions of specific humidity of EXP (a), CNTL (b), observations (c), EXP-observations (d), and CNTL-observations (e).

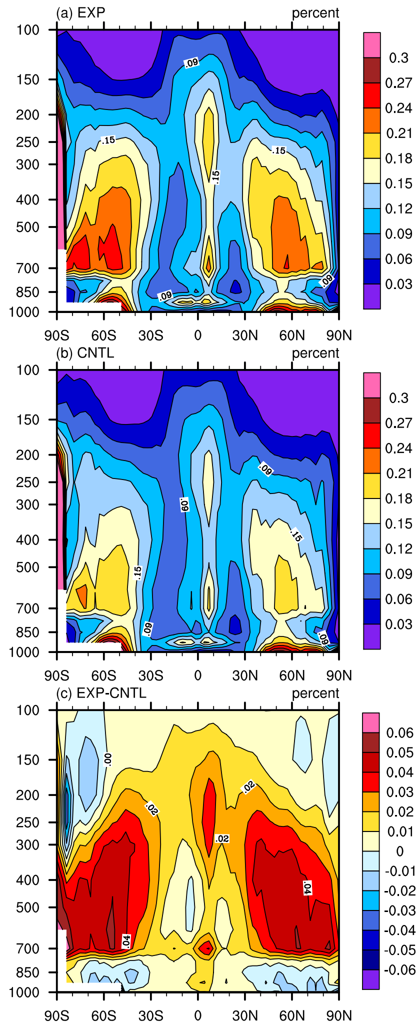

Figure 7. Pressure–latitude distributions of cloud fraction of EXP (a), CNTL (b), and EXP-CNTL (c).

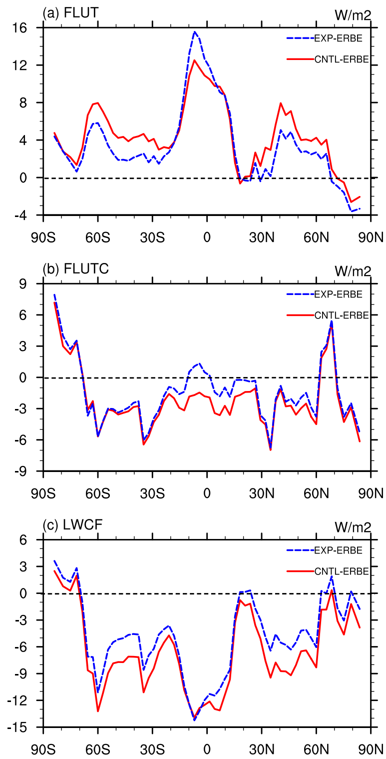

Figure 8. Meridional distributions of the annual mean difference between EXP/CNTL and observations of FLUT (a), FLUTC (b), and LWCF (c).

Table 1. Initial selected uncertain parameters in GAMIL2.

Parameter	Description	Default	Range
c0	rain water autoconversion coefficient for deep convection	3.0e-4	1.e-4 $\sim$ 5.4e-3
ke	evaporation efficiency for deep convection	7.5e-6	5e-7 $\sim$ 5e-5
capelmt	threshold value for cape for deep convection	80	20 $\sim$ 200
rhminl	threshold RH for low clouds	0.915	0.8 $\sim$ 0.95
rhminh	threshold RH for high clouds	0.78	0.6 $\sim$ 0.9
c0_shc	rain water autoconversion coefficient for shallow convection	5e-5	3e-5 $\sim$ 2e-4
cmftau	characteristic adjustment time scale of shallow cape	7200	900 $\sim$ 14400

Table 2. Model output variables in the metrics.

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

Table 3. Fall factor samplings of parameters and metrics.

ID	c0	rhminl	rhminh	cmftau	metrics	ID	c0	rhminl	rhminh	cmftau	metrics
1	1.00E-04	0.915	0.78	7200	1.152	14	3.00E-04	0.875	0.78	7200	1.019
2	3.00E-04	0.915	0.78	7200	1	15	3.00E-04	0.913	0.78	7200	1.007
3	3.04E-04	0.915	0.78	7200	1.054	16	3.00E-04	0.95	0.78	7200	1.094
4	5.08E-04	0.915	0.78	7200	1.017	17	3.00E-04	0.915	0.6	7200	1.00547
5	7.13E-04	0.915	0.78	7200	0.987	18	3.00E-04	0.915	0.675	7200	1.027676
6	9.17E-04	0.915	0.78	7200	1.01	19	3.00E-04	0.915	0.75	7200	1.023358
7	1.12E-03	0.915	0.78	7200	1.04	20	3.00E-04	0.915	0.825	7200	1.028264
8	1.33E-03	0.915	0.78	7200	1.044	21	3.00E-04	0.915	0.9	7200	1.160479
9	2.55E-03	0.915	0.78	7200	1.075	22	3.00E-04	0.915	0.78	900	1.22922
10	3.78E-03	0.915	0.78	7200	1.084	23	3.00E-04	0.915	0.78	4275	1.064064
11	5.00E-03	0.915	0.78	7200	1.09	24	3.00E-04	0.915	0.78	7650	1.004806
12	3.00E-04	0.8	0.78	7200	1.223	25	3.00E-04	0.915	0.78	11025	1.077167
13	3.00E-04	0.838	0.78	7200	1.054	26	3.00E-04	0.915	0.78	14400	1.148265

Table 4. Comparison with effectiveness and efficiency.

	Optimal solution	N_{step}	N_{size}	Core hours
Downhill simplex	0.9585	80	1	14400
PSO	0.911537	24	12	51840
DE	0.942148	33	12	71280
Downhill_2_steps	0.9256899	25+34	1	10620
Downhill_3_steps	0.9098545	80+25+50	1	27900

Algorithm 1 Preprocessing the initial values of Downhill Simplex Algorithm.

```
sampling_sets=full_factor_sampling(parameters_range)
for each initial V_i of N+1 vertexes do
    candidate_init_sets += min(i, sampling_sets)
end for
while one parameter have the same values in the N+1 sets do
    j=1
    remove_parameter_set(the parameter set with higher metrics, candidate_init_sets)
    candidate_init_sets += min(N+1+j, sampling_sets)
    j+=1
end while
```