



An automatic and effective parameter optimization method for model tuning

T. Zhang^{1,2}, L. Li³, Y. Lin², W. Xue^{1,2}, F. Xie³, H. Xu¹, and X. Huang^{1,2}

¹Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

²Center for Earth System Science, Ministry of Education Key Laboratory for Earth System Modeling, Tsinghua University, Beijing 100084, China

³State 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) and L. Li (ljli@lasg.iap.ac.cn)

Received: 7 April 2015 – Published in Geosci. Model Dev. Discuss.: 7 May 2015

Revised: 27 August 2015 – Accepted: 9 October 2015 – Published: 6 November 2015

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 determining the model’s sensitivity to the parameters and the other choosing the optimum initial value for those sensitive parameters, are introduced before the downhill simplex method. This new method reduces the number of parameters to be tuned and accelerates the convergence of the downhill simplex method. 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 parameter 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. Physical parameterizations aim to approximate the overall statistical outcomes of various sub-grid-scale physics (Williams, 2005). 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. Consequently, these parameterizations introduce uncertainties into climate simulations using GCMs (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 (L. 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 extend (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 the 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) approximate the forecast by the generalized likelihood uncertain estimation (Beven and Binley, 1992), a method obtaining parameter uncertainty ranges of a specific confidence level. The Bayesian Markov chain Monte Carlo (MCMC) method (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), the adaptive Metropolis algorithm (Hararuk et al., 2014), and multiple very fast simulated annealing (MVFSa) (Jackson et al., 2008). The MVFSa method is 1 to 2 orders of magnitude faster than the Metropolis–Hasting algorithm (Jackson et al., 2004). However, these methods only attempt to determine the most likely area of uncertain parameters and cannot directly give the best combination of uncertain parameters with an optimum metrics value. Moreover, the PDF heavily depends on the likelihood function assumed, which is usually difficult to determine for climate system model tuning problems.

Optimization algorithms can be used to search the maximum or minimum metrics value in a given parametric space. Severijns and Hazeleger (2005) calibrate parameters of radiation, clouds, and convection in the Speedy model with the 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. The downhill simplex is a fast convergence algorithm when the parametric space is not high-dimensional. However, it is a local optimization algorithm, not aiming to find the global optimal solution. Moreover, the algorithm has convergence issues when the simplex becomes ill-conditioned. Besides the downhill simplex, a few global optimization algorithms are introduced to tune uncertain 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 10 000 steps to get a stable solution (Liang et al., 2013), and MVFSa also requires thousands of steps to get a stable solution (Jackson et al., 2004). MOPSO needs dozens of individual cases in each iteration. All these global optimization algorithms require a large number of model runs and very high computational cost during the model tuning process.

The data assimilation method has been well addressed for state estimation, and can be a potential solution for parameter estimation. Aksoy et al. (2006) estimates the parameter uncertainty in a mesoscale model (Grell et al., 1994) using the ensemble Kalman filter (EnKF). Santitissadeekorn and Jones (2015) presents a two-step filtering for the joint state parameter estimation with a combination method of particle filtering (PF) and EnKF. The EnKF and PF use an ensemble of model simulations to estimate the background error

covariance, which approximate the traditional Kalman filter with a recurrence process (Evensen, 2003; Arulampalam et al., 2002). The accuracy of the error covariance relies on samples. In general, the larger the ensemble size, the more accurate the estimates are. The limitation of ensemble size for practice use and imperfect models make it difficult to select representative samples (Poterjoy et al., 2014). Moreover, as with the MOPSO method, they require a large number of model runs in each iteration with greatly increased computational cost.

A climate system model is a strongly nonlinear system, having a large number of uncertain parameters. As a result, the parametric space of a climate system model is high-dimensional, multi-modal, strongly nonlinear, and inseparable. More seriously, one model run of a climate system model might require tens or even hundreds of 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, the Morris method (Morris, 1991; Campolongo et al., 2007), a global sensitivity analysis method, is chosen to eliminate the insensitive parameters by analyzing the main and interactive effects among parameters. Another global method by Sobol (Sobol, 2001) is used to validate the results of the Morris method. Second, a pre-processing of initial values of selected parameters is presented to accelerate the convergence of optimization algorithm and to resolve the issue of ill-conditioned problems. Finally, the downhill simplex algorithm is used to solve the optimization problem because of its low computational cost and fast convergence for low-dimensional 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. The method and workflow can be easily applied to GCMs to speed up model development processes.

The paper is organized as follows. Section 2 introduces the proposed automatic workflow. 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 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 facilitates model tuning processes with minimal manual management. It effectively manages the dependence and calling sequences of various procedures, including parameter sampling, sensitivity analysis and initial value se-

lection, model configuration and running, and evaluation of model outputs using user-provided metrics. Users only need to specify the model to tune, the parameters to be tuned with their valid ranges, and the calibration method to use.

There are four main modules within the framework as shown in Fig. 1. 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 are included in a post-processing module. The preparation module contains various sensitivity analysis and sampling methods, such as the Morris (Morris, 1991; Campolongo et al., 2007), Sobol (Sobol, 2001), 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) methods. The sensitivity analysis is able to eliminate the duplicated samples to reduce unnecessary model runs. 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) and the Design Analysis Kit for Optimization and Terascale Applications (DAKOTA) (Eldred et al., 2007). Although uncertainty quantification toolkits such as PSUADE and DAKOTA support various calibration and uncertainty analysis methods and pre-defined function interfaces, they cannot organize the above model tuning process as effectively as the proposed model tuning framework.

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 tuning 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 (the fifth phase of the Coupled Model Intercomparison Project) program. The horizontal resolution is $2.8^\circ \times 2.8^\circ$, 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 previous version, GAMIL2 has modifications in cloud-related processes (L. Li et al., 2013), such as

the deep convection parameterization (Zhang and Mu, 2005), the convective cloud fraction (Xu and Krueger, 1991), the cloud microphysics (Morrison and Gettelman, 2008), and the stratiform fractional cloud condensation scheme (Zhang et al., 2003). More details are in L. 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 from the model configuration 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 that a 5-year type of simulation is enough to capture the basic characteristics of simulated mean climate states (Golaz et al., 2011; Lin et al., 2013). The goal of these 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 the GPCP (Global Precipitation Climatology Project, Adler et al., 2003) for precipitation and the 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 reference metrics can be extended, depending on the model performance requirement.

The reference 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 output by that of the control simulation using default parameter values. We introduce an improvement index to evaluate the tuning results, which weight each variable equally and compute the average normalized RMSE. The index indicates an overall improvement of the performance of the tuned simulation relative to the control simulation based on a number of model outputs (Table 2). If the index is less than 1, it means the tuned simulation gets better performance than the control run. The smaller this value, the better the improvement is.

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

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

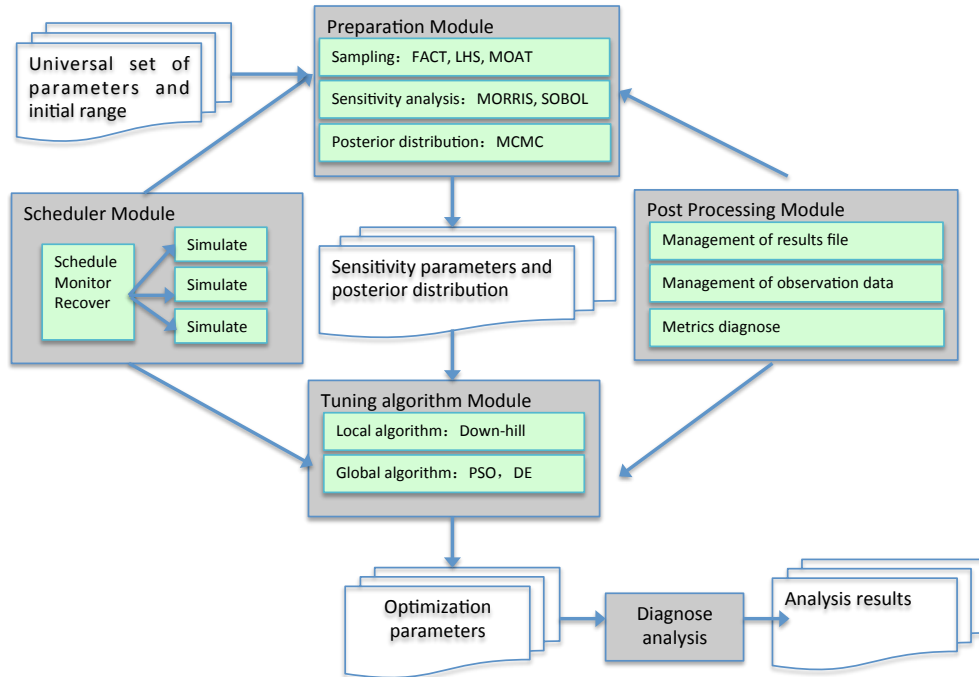


Figure 1. The structure of the automatic calibration workflow. The input of the workflow is the parameter set interest and their initial value ranges. The output is the optimal parameters and their corresponding diagnostic results after calibration. The preparation module provides the parameter sensitivity analysis. The tuning algorithm module offers local and global optimization algorithms including the 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 a parallel system. The post-processing module is responsible for metrics diagnostics, re-analysis and observational data management.

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

$x_m^F(i)$ is the model outputs, and $x_o^F(i)$ is the corresponding observation or reanalysis data. $x_r^F(i)$ is model outputs from the control simulation using the default values for the parameters in Table 1. w is the weight due to the different grid area on a regular latitude–longitude grid on the sphere. I is the total grid number in model. N^F is the number of the chosen variables.

4 Method

4.1 Parameter tuning with global and local optimization methods

Parameter tuning for a climate system model is intended to solve a global optimization problem in theory. As the well-known global optimization algorithms, traditional evolutionary algorithms, such as the genetic algorithm (Goldberg et al., 1989), differential evolutionary (DE) (Storn and Price, 1995), and particle swarm optimization (PSO) (Kennedy, 2010) algorithms, can approach the global optimal solution,

but generally require high computational cost (Hegerty et al., 2009; Shi and Eberhart, 1999). This is because these algorithms are designed following biological evolution of survival of the fittest. In contrast, the local algorithms utilize the greedy strategy, and thus may tap into a locally optimal solution after convergence. The advantage of local algorithms is the low computational cost due to the relatively fewer samples required. In this sense, the 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. The downhill simplex method 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 $N + 1$ vertices 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 improvement index as defined in Eq. (3). The new vertex is determined by expanding or shrinking the vertex with the highest metrics value, leading to a new simplex (Press et al., 1992; Nelder and Mead, 1965).

Two performance criteria are used to evaluate the effectiveness and efficiency of the optimization algorithms in this study. Selection of optimization algorithms for parameter

Table 1. A summary of parameters to be tuned in GAMIL2. The default and final tuned optimum values are also shown. The valid range of each parameter is also included. Note that only four sensitive parameters are tuned and have optimum values.

Parameter	Description	Default	Range	Optimal
c0	Rainwater 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	Rainwater autoconversion coefficient for shallow convection	5×10^{-5}	3×10^{-5} – 2×10^{-4}	–
cmftau	Characteristic adjustment timescale of shallow capes	7200	900–14 400	7198.048

Table 2. Atmospheric fields included in the evaluation metrics and their sources.

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	Clear-sky 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	Clear-sky short-wave net flux at TOA	ERBE

calibration of climate system models is a balance between model improvement (effectiveness) and computational cost (efficiency). In this study, model improvement is measured by an index defined in Eq. (3). The lower this value is, the better the model tuning is. Computational cost is measured by “core hours”, which stands for the computational efficiency. It is computed by $(N_{\text{step}}) \times (N_{\text{size}}) \times$ (the number of processes of a single model run) \times (hours used for a single 5-year model run). N_{step} is the total number of iterations of optimization algorithms for convergence. N_{size} is the number of model runs during each iteration, and it is 1 for the downhill simplex method. In the GAMIL2 case, each model run takes 6 h using 30 processes.

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

To improve the effectiveness of the 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 is achieved by pre-selecting proper initial values for the parameters before using the 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, the downhill simplex method, and the simulated annealing algorithm (Van Laarhoven and Aarts, 1987), are ineffective in high-dimensional problems. Iterations for convergence will increase exponentially with the number of tuning parameters. In addition, climate models generally need a long simulation to have meaningful results. Therefore, the high-dimensional parameter tuning problem suffers from an extremely high computational cost. It is necessary to reduce the parameter dimension before the optimization.

Parameter sensitivity analysis can be divided into local and global methods (Gan et al., 2014). The local method determines the sensitivity of a single parameter by perturbing one parameter with all other parameters fixed. Consequently, it does not consider the combined sensitivity of multiple parameters. On the other hand, the global method perturbs all the parameters to explore the sensitivity of the whole parametric space. In this study, the Morris method (Morris, 1991; Campolongo et al., 2007), a global method, is used to screen out the sensitive parameters. Another global method (Sobol, 2001) is used to validate the results of the Morris method.

Table 3. Effectiveness and efficiency comparison between the original downhill simplex method and the two global methods. N_{step} is the total number of calibrating iterations for convergence. N_{size} is the number of model runs during each iteration. Core hours is computed by $N_{\text{step}} \times N_{\text{size}} \times \{\text{the number of processes of a single model run}\} \times \{\text{hours used for a single 5-year model run}\}$. In the GAMIL2 case, each model run takes 6 h and uses 30 processes.

	Improvement index	N_{step}	N_{size}	Core hours
Downhill_1_step	0.9585	80	1	14 400
PSO	0.9115	24	12	51 840
DE	0.9421	33	12	71 280

The Morris method, based on the MOAT sampling strategy, reduces the number of samples required by other global sensitivity methods (J. Li et al., 2013). Note that a sample is a set of all parameters, not just one parameter. The method is described briefly here, and more details can be found in Morris (1991). Assuming we have k parameters, relative to a random sample $S_1 = \{x_1, x_2, \dots, x_k\}$, another sample $S_2 = \{x_1, x_2, \dots, x_i + \Delta_i, \dots, x_k\}$ can be constructed by perturbing the i th parameter by Δ_i , where Δ_i is a perturbation of this parameter. The elementary effect of the i th parameter x_i is defined as

$$d_i = \frac{f(S_2) - f(S_1)}{\Delta_i}, \tag{4}$$

where f stands for the improvement index as defined in Eq. (3). A third sample $S_3 = \{x_1, x_2, \dots, x_i + \Delta_i, \dots, x_j + \Delta_j, \dots, x_k\}$ can be generated by perturbing another parameter, where j is not i . In doing so k times, we will get $k + 1$ samples $\{S_1, S_2, \dots, S_{k+1}\}$ and k elementary effects $\{d_1, d_2, \dots, d_k\}$ after perturbing all the parameters. The vector of $\{S_1, S_2, \dots, S_{k+1}\}$ is called a trajectory. This procedure is repeated for r iterations, and finally we get r trajectories. The starting point of any trajectory is selected randomly as well as the ordering of the parameters to perturb and the Δ for each perturbation in one trajectory. In practice, a number of 10 to 50 trajectories is enough to determine the feasible sensitivity of parameters (Gan et al., 2014; Morris, 1991). In this study, we have a total of seven parameters, and 80 simulations are conducted.

We define $D = \{D_i(t)\}$, where t is the t th trajectory, and i is the i th elementary effect of the parameter x_i . μ_i , the mean of $|d_i|$, and σ_i , the standard deviation of d_i , are used to measure the parameter sensitivity, defined as

$$\mu_i = \sum_{t=1}^r \frac{|d_i(t)|}{r}; \tag{5}$$

$$\sigma_i = \sum_{t=1}^r \sqrt{(d_i(t) - \mu_i)^2 / r}. \tag{6}$$

μ_i estimates the effect of x_i on the model improvement index as defined in Eq. (3), while σ_i assesses the interactive effect of x_i with other parameters. Those parameters with large μ_i and σ_i are the sensitive parameters. The Morris method results are shown in Fig. 2.

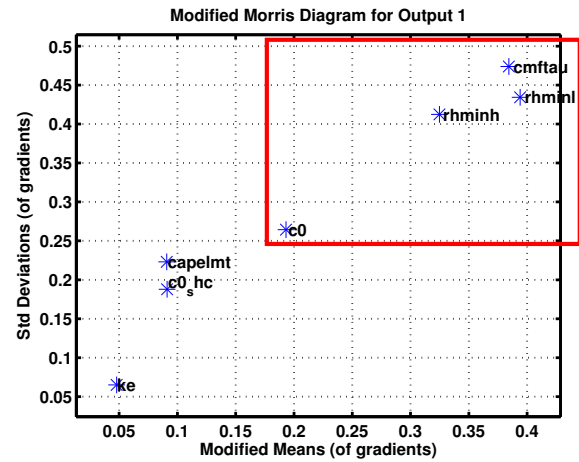


Figure 2. Scatter diagram showing the parameter sensitivity using the Morris sensitivity analysis. The x axis stands for the main effect sensitivity of a single parameter. The y axis stands for the interactive effect sensitivity among multi-parameters. In GAMIL2, c_0 , $rhminl$, $rhminh$, and $cmftau$ have high sensitivity and ke , c_0_shc , and $capelmt$ have low sensitivity.

The parameter elimination step is critical for the final result of model tuning. To validate the results obtained by the Morris method, we compare the results with a benchmark method (Sobol, 2001). Based on variance decomposition, the Sobol method requires more samples than the Morris method, leading to a higher computation cost. The variance of the model output can be decomposed as Eq. (7), where n is the number of parameters, V_i is the variance of the i th parameter, and V_{ij} is the variance of the interactive effect between the i th and j th parameters, and so on. The total sensitivity effect of the i th parameter can be presented as Eq. (8), 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 method.

$$V = \sum_{i=1}^n V_i + \sum_{1 \leq i < j \leq n} V_{ij} + \dots + V_{1,2,\dots,n} \tag{7}$$

$$S_{T_i} = 1 - \frac{V_{-i}}{V} \tag{8}$$

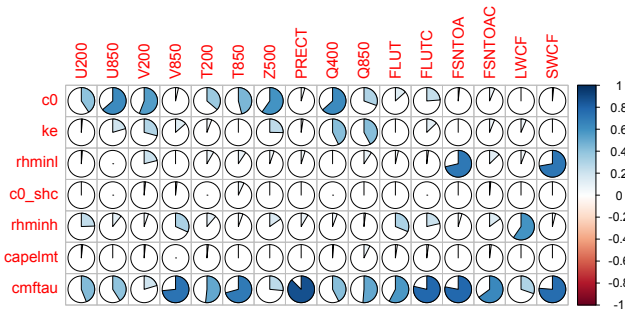


Figure 3. Sensitivity analysis results from the Sobol method. The total sensitivity in Eq. (8) is denoted by the size of the color area. The total sensitivities of ke, c0_shc, and capelmt are less than 0.5 in terms of each variable.

4.3 Proper initial value selection for the downhill simplex method

The downhill simplex method is a local optimization algorithm and its convergence performance strongly depends on the quality of the initial values. We need to find the parameters with the smaller metrics around the final solution. Moreover, we have to finish the searching as fast as possible with minimal overhead. For these two objectives, a hierarchical sampling strategy based on the single parameter perturbation (SPP) sample method is used. The SPP is similar to local sensitivity methods, in which only one parameter is perturbed at one time with other parameters fixed. The perturbation samples are uniformly distributed across parametric space. First, the improvement index as defined in Eq. (3) of each parameter sample is computed. The distance is defined as the difference between the improvement indexes using two adjacent samples, i.e., the model response measured by a certain percentage change of one parameter. We call this step the first-level sampling. The specific perturbation size for one parameter can be set based on user experience. In our implementation, the user needs to set the number of samples. For the first-level sampling, we can use a larger perturbation size to reduce computational cost. If the distance between two adjacent samples is greater than a predefined threshold, more SPP samples between the previous two adjacent samples are conducted, and this is called the second-level sampling. Finally, $k+1$ samples with the best improvement index value are chosen as the candidate initial values for the optimization method. With this hierarchical sampling strategy, we can determine the local parametric space for the final solution and can accelerate the convergence of the following downhill simplex method. This procedure is described in Algorithm 1. It is easy to implement and has lower overhead compared to other complex adaptive sampling methods.

Algorithm 1 Preprocessing the initial values of Downhill Simplex Algorithm.

```

1: //Single parameter perturbation sample(SPP)
2: N=number_of_parameters
3: sampling_sets={ }
4: for each parameter  $P_i$  of  $N$  parameters do
5:   sampling_sets+=SPP( $P_i$ _range, number_of_samples)
6:   //refine sample in the sensitivity range if needed
7:   if metrics of the the adjacent same parameter sampling points
   >= sensitivity_threshold then
8:     sampling_sets+=SPP( $P_i$ _adjacent_parameter_range, re-
       fine_number_of_factors)
9:   end if
10: end for
11:
12: //Initial vertexes with parameters of the  $N+1$  minimum metrics

13: for each initial  $V_i$  of  $N+1$  vertexes do
14:   //get the parameters of the  $i$ th minimum metrics
15:   candidate_init_sets += min( $i$ , sampling_sets)
16: end for
17:
18: //make sure the initial simplex geometry is well-conditioned
19: while one parameter  $k$  have the same values in the  $N+1$  sets
   do
20:    $j = 1$ 
21:   //remove the parameter set with the worst metrics from can-
     didate_init_sets
22:   remove_parameter_set(the parameter set with worse met-
     rics, candidate_init_sets)
23:   //get the parameters of the  $N+1+j$ th minimum metrics
24:   candidate_init_sets += min( $N+1+j$ , sampling_sets)
25:    $j += 1$ 
26: end while

```

At the same time, inappropriate initial values may lead to ill-conditioned simplex geometry, which can be found in the model tuning process. One issue we meet is that some vertexes in the downhill simplex optimization may have the same values for one or more parameters. As a result, these parameters remain invariant during the optimization, and this may degrade the quality of the final solution as well as the convergence speed. A simplex checking is conducted to keep as many different values of parameters as possible during the process of looking for initial values. Well-conditioned simplex geometry will increase the parameter freedom for optimization. In our implementation (Algorithm 1), the vertex leading to the ill-conditioned simplex is replaced by another parameter sample that gives another minimum improvement index value.

These methods mentioned above are summarized as the initial value pre-processing of the downhill simplex algorithm. Sometimes, the samples used during the initial value selection are the same as those in the parameter sensitivity analysis step. In this case, one model run can be used in both steps to further reduce the computational cost.

Table 4. The same as Table 3 but showing the comparison among the three downhill simplex methods.

	Improvement index	N_{step}	N_{size}	Core hours
Downhill_1_step	0.9585	80	1	14 400
Downhill_2_steps	0.9257	25 + 34	1	10 620
Downhill_3_steps	0.9099	80 + 25 + 50	1	27 900

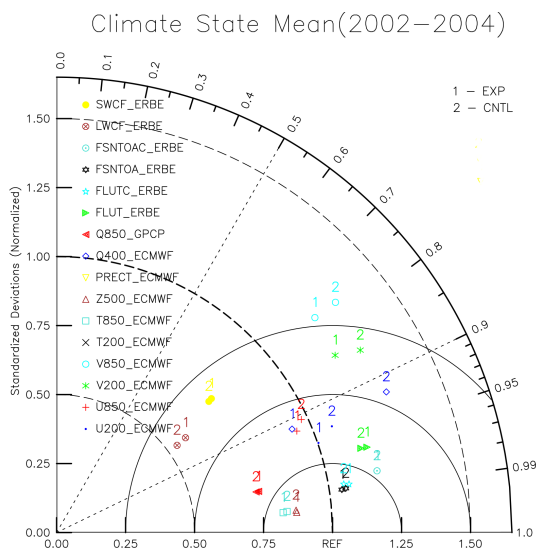


Figure 4. Taylor diagram of the climate mean state of each output variable from 2002 to 2004 of EXP and CNTL.

4.4 Evaluation of the proposed strategy

The effectiveness and efficiency of the three traditional algorithms are compared in Table 3. “Downhill_1_step” represents the original downhill simplex method, which is one of the most widely used local optimization algorithms and has been successfully used in the Speedy model (Severijns and Hazeleger, 2005). PSO and DE are the most widely used global optimization algorithms and are easy to use. Although “Downhill_1_step” achieves slightly worse improvement compared to the two global optimization methods (Table 3), its computation cost is much less (only 20 and 28 % of DE and PSO, respectively).

Two extra steps are included before the original downhill simplex method to overcome its limited effectiveness on model performance improvement. The “Downhill_2_steps” method includes an initial value pre-processing step before the downhill simplex method, and the “Downhill_3_steps” method further introduces another step to eliminate insensitive parameters for tuning by sensitivity analysis. The two steps bring in additional overhead, 80 samples for the parameter sensitivity analysis with the Morris method, and 25 samples for the initial value pre-processing. Tables 3 and 4 show that the proposed “Downhill_3_steps” achieves the best effectiveness, improving the model’s overall performance by

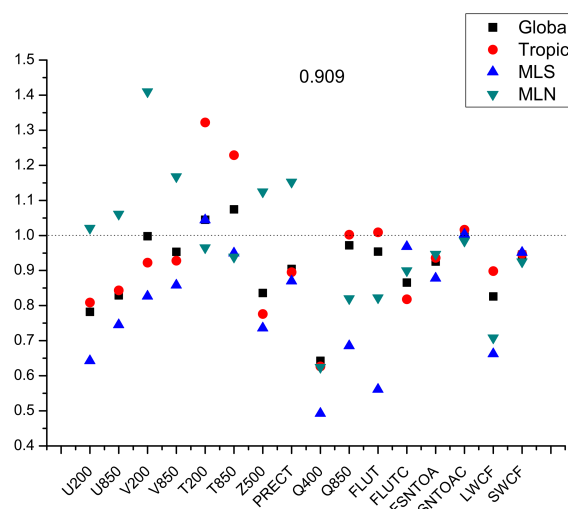


Figure 5. Improvement indices over the global, tropical and mid-high latitudes of the Northern and Southern Hemisphere (MLN and MLS) for each variable of the EXP simulation.

9 %. It overcomes the inherent ineffectiveness of the original downhill simplex method with a much lower computational cost than global methods.

5 Analysis of model optimal results

This section compares the default simulation and the tuned simulation by the 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 control (labeled as CNTL) and optimized (labeled as EXP) simulations. Significant change is found for c_0 , 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 that 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 the other variables are relatively small. The change in terms of the RMSE factor over the globe and

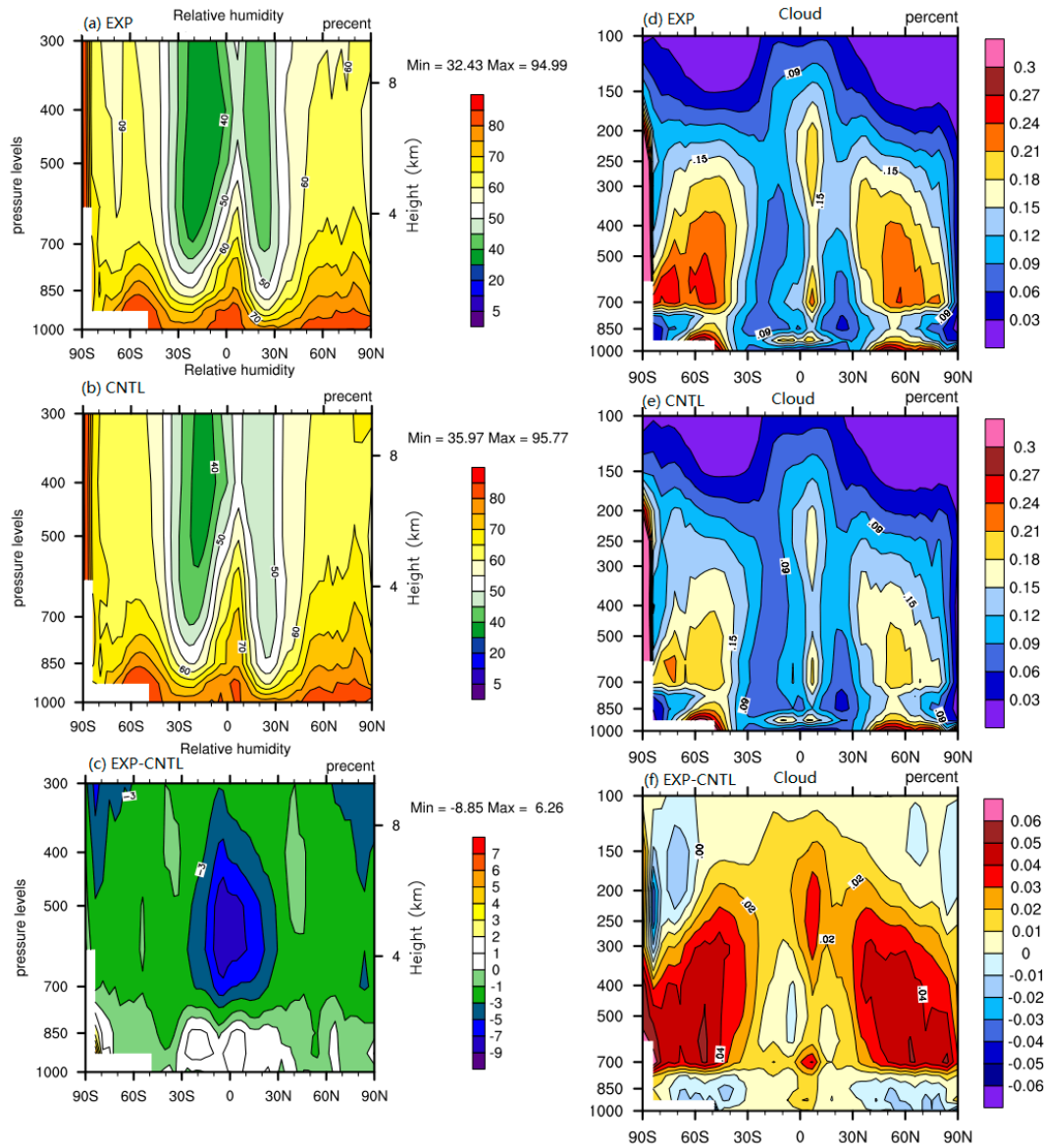


Figure 6. Pressure–latitude distributions of relative humidity and cloud fraction of EXP (a, d), CNTL (b, e), and EXP-CNTL (c, f).

three regions (tropics, SH middle and high latitudes and NH middle and high latitudes) are shown in Fig. 5. First, radiative fields and moisture are improved over all four areas. By contrast, wind and temperature field changes are more diverse among different areas. For example, temperatures over the tropics become worse compared to the control run. There is an overall improvement at the SH middle and high latitudes for all variables except for the 200 hPa temperature. Winds and precipitation at the NH middle and high latitudes become slightly worse in the tuned simulation. Such changes are somewhat intriguing and we attempt to relate these changes to the two parameters significantly tuned.

With a reduced RH threshold for high cloud (from 0.78 in CNTL to 0.63 in EXP, Table 1), the stratiform condensa-

tion rate increases and the atmospheric humidity decreases (Zhang et al., 2003). In addition, with an increased auto-conversion coefficient in the deep convection, less condensate is detrained to the environment. As a result, the middle and upper troposphere is overall drier, especially over the tropics, where deep convection dominates the vertical moisture transport (Fig. 6c). Although the middle and upper troposphere becomes drier over the tropics, the reduced RH threshold for high cloud makes clouds easier to be present. Consequently, middle and high clouds increase over the globe, especially over the middle and high latitudes, with the largest increase up to 4–5 % (Fig. 6f). In the tropics, due to the drier tendency induced by the reduced detrainment, high cloud increase is relatively small (2–3 %) compared to the

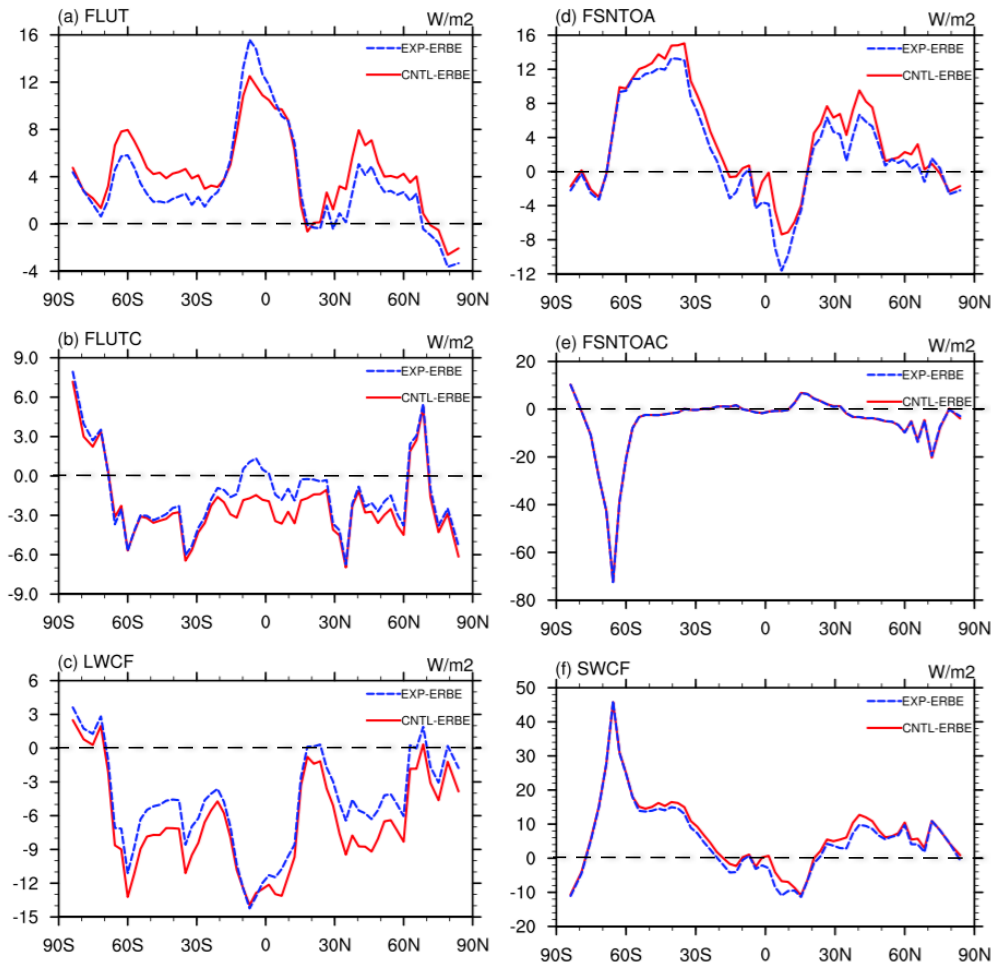


Figure 7. Meridional distributions of the annual mean difference between EXP/CNTL and observations of TOA outgoing long-wave radiation (a), TOA clear-sky outgoing long-wave radiation (b), TOA long-wave cloud forcing (c), TOA net short-wave flux (d), TOA clear-sky net short-wave flux (e), and TOA short-wave cloud forcing (f).

middle and high latitudes. On the contrary, low cloud below 800 hPa decreases by 1–2 % over the middle and high latitudes, with slightly decreased RH (Fig. 6) because of the negligible change of RH threshold for low cloud (Table 1). Overall, the combined effects of all relevant parameterizations lead to the changes in atmospheric humidity and cloud fraction.

Changes in moisture and cloud fields impact radiative fields. With reference to ERBE, TOA outgoing long-wave radiation (OLR) is improved at 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 long-wave upward flux at TOA (FLUT), reducing the FLUT in mid-latitudes of the Southern Hemisphere and Northern Hemisphere (Fig. 7a). 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 out-

going long-wave radiation and reduces the negative bias of clear-sky long-wave upward flux at TOA (FLUTC, Fig. 7b). Long-wave cloud forcing (LWCF) at the middle and high latitudes is improved due to the improvement in the FLUT in these areas (Fig. 7c), but improvement in the tropics is negligible due to the cancellation between the FLUT and FLUTC. Overall, the tuned simulation has a TOA radiation imbalance of 0.08 W m^{-2} , which is better than that of the control run (0.8 W m^{-2}).

TOA clear-sky short-waves are the same between the control and the tuned simulation since both simulations have the same surface albedo. With increased clouds, the tuned simulation has smaller TOA short-waves absorbed than the control. Compared with ERBE, the tuned simulation has better TOA short-waves absorbed at the middle 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, a parameter sensitivity analysis step and a proper initial value selection step are introduced before the low-cost downhill simplex 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 support different uncertainty quantification analysis and calibration strategies. Evaluation of the method and workflow by calibrating the GAMIL2 model indicates that the three-step method 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 original downhill simplex method. The optimal results of the three-step method demonstrate that most of the variables are improved compared with the control simulation, especially for the radiation-related ones. The mechanism analysis is conducted to explain why these radiation-related variables have an overall improvement. In future work, more analyses are needed to better understand the model behavior along with the physical parameter changes. The choice of appropriate reference metrics and related observations are very important for the final tuned model performance. In future studies, we are going to use the more reliable and accurate observations, and add some constraint conditions for parameter tuning to construct a more comprehensive and reliable metrics.

Acknowledgements. The authors are grateful to the reviewers for valuable comments that have greatly improved the paper. This work is partially supported by the Ministry of Science and Technology of China under grant no. 2013CBA01805, the Information Technology Program of the Chinese Academy of Sciences under grant no. XXH12503-02-02-03, the China Special Fund for Meteorological Research in the Public Interest under grant no. GYHY201306062, and the Natural Science Foundation of China under grant nos. 91530103, 61361120098, and 51190101.

Edited by: R. Neale

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. Hydrometeorol.*, 4, 1147–1167, 2003.
- 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.
- 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.
- Arulampalam, M., Maskell, S., Gordon, N., and Clapp, T.: A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, *Signal Processing, IEEE T.*, 50, 174–188, 2002.
- 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.
- Barkstrom, B. R.: The earth radiation budget experiment (ERBE), *B. Am. Meteorol. Soc.*, 65, 1170–1185, 1984.
- Beven, K. and Binley, A.: The future of distributed models: model calibration and uncertainty prediction, *Hydrol. Process.*, 6, 279–298, 1992.
- 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.
- Campolongo, F., Cariboni, J., and Saltelli, A.: An effective screening design for sensitivity analysis of large models, *Environ. Modell. Softw.*, 22, 1509–1518, 2007.
- 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.
- 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.
- Elkinton, C. N., Manwell, J. F., and McGowan, J. G.: Algorithms for offshore wind farm layout optimization, *Wind Engineering*, 32, 67–84, 2008.
- Evensen, G.: The ensemble Kalman filter: Theoretical formulation and practical implementation, *Ocean Dynam.*, 53, 343–367, 2003.
- 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.
- Gilks, W. R.: *Markov Chain Monte Carlo Practice*, Chapman and Hall/CRC, London, UK, 1995.
- 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.
- Gleckler, P. J., Taylor, K. E., and Doutriaux, C.: Performance metrics for climate models, *J. Geophys. Res.-Atmos.*, 113, D06104, doi:10.1029/2007JD008972, 2008.
- Golaz, J., Salzmann, M., Donner, L. J., Horowitz, L. W., Ming, Y., and Zhao, M.: Sensitivity of the aerosol indirect effect to sub-grid variability in the cloud parameterization of the GFDL atmosphere general circulation model AM3, *J. Climate*, 24, 3145–3160, 2011.
- Goldberg, D. E., Korb, B., and Deb, K.: Messy genetic algorithms: motivation, analysis, and first results, *Complex systems*, 3, 493–530, 1989.
- Grell, G. A., Dudhia, J., and Stauffer, D.: A description of the fifth-generation Penn State/NCAR Mesoscale Model (MM5), NCAR Tech. Note NCAR/TN-3981STR, 122 pp., 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, *J. Geophys. Res.-Atmos.*, 99, 20785–20813, 1994.
- Hader, R. and Park, S. H.: Slope-rotatable central composite designs, *Technometrics*, 20, 413–417, 1978.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Kennedy, J.: Particle swarm optimization, in: Encyclopedia of Machine Learning, Springer, New York, USA, 760–766, 2010.
- 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.
- Li, J., Duan, Q. Y., Gong, W., Ye, A., Dai, Y., Miao, C., Di, Z., Tong, C., and Sun, Y.: Assessing parameter importance of the Common Land Model based on qualitative and quantitative sensitivity analysis, *Hydrol. Earth Syst. Sci.*, 17, 3279–3293, doi:10.5194/hess-17-3279-2013, 2013.
- 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.
- Lin, Y. L., Zhao, M., Ming, Y., Golaz, J., Donner, L. J., Klein, S. A., Ramaswamy, V., and Xie, S.: Precipitation partitioning, tropical clouds, and intraseasonal variability in GFDL AM2, *J. Climate*, 26, 5453–5466, 2013.
- 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.
- 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, *J. 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, *Computer J.*, 7, 308–313, 1965.
- Poterjoy, J., Zhang, F., and Weng, Y.: The effects of sampling errors on the EnKF assimilation of inner-core hurricane observations, *Mon. Weather Rev.*, 142, 1609–1630, 2014.
- Press, W., Teukolsky, S., Vetterling, W., and Flannery, B.: Numerical Recipes in Fortran, Cambridge Univ. Press, Cambridge, 70 pp., 1992.
- Raktoe, B. L., Hedayat, A., and Federer, W. T.: Factorial Designs, John Wiley & Sons, Hoboken, New Jersey, USA, 1981.
- Reichler, T. and Kim, J.: How well do coupled models simulate today's climate?, *B. Am. Meteorol. Soc.*, 89, 303–311, 2008.
- Santitissadeekorn, N. and Jones, C.: Two-stage filtering for joint state-parameter estimation, *Mon. Weather Rev.*, 143, 2028–2042, doi:10.1175/MWR-D-14-00176.1, 2015.
- Severijns, C. and Hazeleger, W.: Optimizing parameters in an atmospheric general circulation model, *J. Climate*, 18, 3527–3535, 2005.
- 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.
- Simmons, A., Uppala, S., Dee, D., and Kobayashi, S.: ERA-interim: new ECMWF reanalysis products from 1989 onwards, *ECMWF Newsl.*, 110, 25–35, 2007.
- Sobol, I. M.: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Math. Comput. Simulat.*, 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, Berkeley, California, USA, 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, *Hydrol. Earth Syst. Sci.*, 17, 4995–5011, doi:10.5194/hess-17-4995-2013, 2013.
- Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res.-Atmos.*, 106, 7183–7192, 2001.
- Tong, C.: PSUADE User's Manual, Lawrence Livermore National Laboratory (LLNL), Livermore, CA, 109 pp., 2005.
- Van Laarhoven, P. J. and Aarts, E. H.: Simulated Annealing, Springer, Dordrecht, Netherlands, 1987.
- 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.
- 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.
- Williams, P. D.: Modelling climate change: the role of unresolved processes, *Philos. T. R. Soc. A*, 363, 2931–2946, 2005.
- Xu, K.-M. and Krueger, S. K.: Evaluation of cloudiness parameterizations using a cumulus ensemble model, *Mon. Weather Rev.*, 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., Wang, M., and Liu, X.: Uncertainty quantification and parameter tuning in the

- CAM5 Zhang-McFarlane convection scheme and impact of improved convection on the global circulation and climate, *J. Geophys. Res.-Atmos.*, 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, *Clim. Dynam.*, 1, 1–24, 2014.
- Yu, R.: A two-step shape-preserving advection scheme, *Adv. Atmos. Sci.*, 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, *J. Geophys. Res.-Atmos.*, 110, D09109, doi:10.1029/2004JD005617, 2005.
- Zhang, M., Lin, W., Bretherton, C. S., Hack, J. J., and Rasch, P. J.: A modified formulation of fractional stratiform condensation rate in the NCAR Community Atmospheric Model (CAM2), *J. Geophys. Res.-Atmos.*, 108, 4035, doi:10.1029/2002jd002523, 2003.