

Short ensembles: An efficient method for discerning climate-relevant sensitivities in atmospheric general circulation models

H. Wan¹, P. J. Rasch¹, K. Zhang¹, Y. Qian¹, H. Yan^{2,1}, and C. Zhao¹

¹Pacific Northwest National Laboratory, Richland, WA, USA

²School of Atmospheric Sciences, Lanzhou University, Lanzhou, China

Abstract. This paper explores the feasibility of an experimentation strategy for investigating sensitivities in fast components of atmospheric general circulation models. The basic idea is to replace the traditional serial-in-time long-term climate integrations by representative ensembles of shorter simulations. The key advantage of the proposed method lies in its efficiency: since fewer days of simulation are needed, the computational cost is less, and because individual realizations are independent and can be integrated simultaneously, the new dimension of parallelism can dramatically reduce the turnaround time in benchmark tests, sensitivities studies, and model tuning exercises. The strategy is not appropriate for exploring sensitivity of all model features, but it is very effective in many situations.

Two examples are presented using the Community Atmosphere Model version 5. In the first example the method is used to characterize sensitivities of the simulated clouds to time step length. Results show that 3-day ensembles of 20 to 50 members are sufficient to reproduce the main signals revealed by traditional 5 yr simulations. A nudging technique is applied to an additional set of simulations to help understand the contribution of physics-dynamics interaction to the detected time step sensitivity. In the second example, multiple empirical parameters related to cloud microphysics and aerosol lifecycle are perturbed simultaneously in order to find out which parameters have the largest impact on the simulated global mean top-of-atmosphere radiation balance. It turns out that 12-member ensembles of 10-day simulations are able to reveal the same sensitivities as seen in 4 yr simulations performed in a previous study. In both cases, the ensemble method reduces the total computational time by a factor of about 15, and the turnaround time by a factor of several hundred. The efficiency of the method makes it particularly useful for the development of high-resolution, costly and complex climate models.

1 Introduction

Climate, by definition, is the statistical characterization of the state of the Earth's atmosphere, land and ocean on time scales longer than a few months (e.g., IPCC, 2013). Because of the strong natural variabilities resulting from nonlinear interactions between relevant processes, atmospheric general circulation models (AGCMs) used in sensitivities studies need to be integrated for multiple years, usually decades, in order to obtain statistically meaningful and robust signals. However, state-of-the-art AGCMs are computationally expensive to integrate when resolution is high, or when a large number of simulations are needed. Recent examples of such studies include those of Wehner et al. (2013); Zhao et al. (2013); Yang et al. (2012, 2013) and Qian et al. (2014), to name but a few.

The high computational costs have motivated researchers to look for alternative methods to facilitate extracting signals from noise in climate models. For example, Kooperman et al. (2012) showed that anthropogenic aerosol indirect effects could be estimated from substantially shorter simulations if temperature and horizontal winds in the AGCM are relaxed (nudged) towards prescribed conditions to reduce variability in those fields, while allowing the model to calculate the responses to aerosol emissions in cloud, water, and aerosol fields. For more general applications, however, nudging can hide sensitivities in the constrained fields, as well as in feedbacks that involve these quantities.

In the climate modeling community, it has been widely recognized that fast processes (those that produce a model response to a perturbation on a timescale of days when sea surface temperature is fixed, such as those related to clouds) are important sources of discrepancies between the observed and simulated climate, and between the future climate projections provided by different models (Cess et al., 1990; Colman, 2003; Soden and Held, 2006; Ringer et al., 2006;

Dufresne and Bony, 2008). In addition, it has been noticed that when climate models are used in short-range weather prediction experiments starting from realistic initial conditions, many of the key model biases form within a very short time period (hours to a few days, see, e.g., Phillips et al., 2004; Williamson et al., 2005; Boyle et al., 2005; Rodwell and Palmer, 2007; Martin et al., 2010; Xie et al., 2012; Ma et al., 2013, 2014; Klocke and Rodwell, 2013). There has been increasing interests in running climate models in weather prediction mode to diagnose model errors. The most well-known examples include the Climate Change Prediction Program – Atmospheric Radiation Measurement (CCPP-ARM) Parameterization Testbed (CAPT) initiative of the US Department of Energy (Phillips et al., 2004), and the phase II experiment of the Transpose-Atmospheric Model Intercomparison Project (Transpose-AMIP II, Williams et al., 2013) that was run alongside phase 5 of the Coupled Model Intercomparison Project (CMIP5, Taylor et al., 2012).

In this study we demonstrate that the important role of fast processes in the climate system can be exploited in more general ways to provide an alternate strategy to efficiently carry out model sensitivity experiments and tuning exercises. The basic idea is to replace the traditional serial-in-time long-term climate integrations by generating representative ensembles of shorter simulations. (Details are discussed in later sections). Significant gain in computational efficiency can be expected for two reasons: first, unlike a serial-in-time multi-year simulation, the ensemble of realizations can be integrated simultaneously. This introduces an additional dimension of parallelism to better exploit modern supercomputer systems that consist of order 10^5 – 10^6 cores, leading to substantial reduction of the turnaround time in sensitivity experiments. Second, in comparison to a long-term integration which can be understood as an ensemble with auto-correlated realizations, the use of independent members increases the effective sample size. One can thus expect equally robust statistics to be obtained from a smaller number of simulation days, resulting in an reduction of total CPU time.

While the Transpose-AMIP-type evaluation focuses on comparison against observations to understand the initial development of model biases, in this study we are interested in model sensitivities to parametric and structural changes near the model's equilibrium climate. Using the Community Atmosphere Model version 5 (CAM5, Neale et al., 2010, cf. Sect. 2), we present two examples to elaborate the ensemble strategy and evaluate its effectiveness (Sects. 3 and 4). Further discussions and conclusions are given in Sect. 5.

2 Model and initial conditions

The climate model used here is CAM5.1 (Neale et al., 2010) with a finite volume dynamical core that use the numerical schemes of Lin and Rood (1996) and Lin (2004) to represent the hydrostatic adiabatic fluid dynamics and large-scale

tracer transport. Deep convection is treated with the mass-flux-type parameterization of Zhang and McFarlane (1995), with further modifications by Richter and Rasch (2008) and Neale et al. (2008). Shallow convection is parameterized as in Park and Bretherton (2009). Large-scale condensation and stratiform cloud fraction are handled by the parameterization of Park et al. (2014). The stratiform cloud microphysics is represented by a two-moment scheme that explicitly calculates the mass and number concentrations of cloud liquid, cloud ice, rain and snow (Morrison and Gettelman, 2008; Gettelman et al., 2008, 2010). The vertical transport of heat, momentum, and moisture by turbulent eddies is represented following the work of Bretherton and Park (2009). Solar and terrestrial radiation calculations are performed using the Rapid Radiative Transfer Model for GCMs (RRTMG, Iacono et al., 2008; Mlawer et al., 1997). The lifecycle of aerosols is represented with a comprehensive module that describes the aerosol size distribution with 3 log-normal modes (MAM3, Liu et al., 2012). Land surface processes, including hydrological and biogeochemical processes, dynamical vegetation and biogeophysics, are handled by the Community Land Model version 4 (CLM4, Lawrence et al., 2011). A detailed description of the CAM5 model can be found in Neale et al. (2010). All simulations in the present paper used the tropospheric version of CAM5 with 30 vertical layers, at a horizontal resolution of 1.9° latitude \times 2.5° longitude. The default model time step for this configuration is 30 min.

As mentioned in the introduction, the motivation for exploring a new experimentation strategy is to reduce the wall-clock time and CPU time spent on model integration. We thus intend to perform as few as possible simulations that are as short as possible. This requires the ensemble members to be appropriately sampled so that the ensemble average is representative of the long-term climate. Based on the viewpoint that climate is the “average weather”, we initialize individual realizations using atmospheric states representing different synoptic patterns of the large-scale circulation. The source of such initial conditions could be global weather analyses, as done in CAPT and Transpose-AMIP, which would require interpolation and adjustments to take into account the different grids and topography used for the analysis data and by the CAM5 model. Initialization of the aerosol module would remain an issue because detailed information about aerosol concentrations in different size ranges is not normally provided by the analyses. Considering that our focus here is not to compare with observation, it is not necessary to have a realistic initialization of model state variables that matches particular meteorological events. We therefore chose to use initial conditions generated by the GCM itself using an inexpensive model configuration. For the application examples discussed in the present paper, the CAM5 model was integrated for 20 years at 1.9° latitude \times 2.5° longitude resolution using the default choices for model parameters and model time step, driven by annually cycled monthly mean climatological sea surface temperature distributions and sea

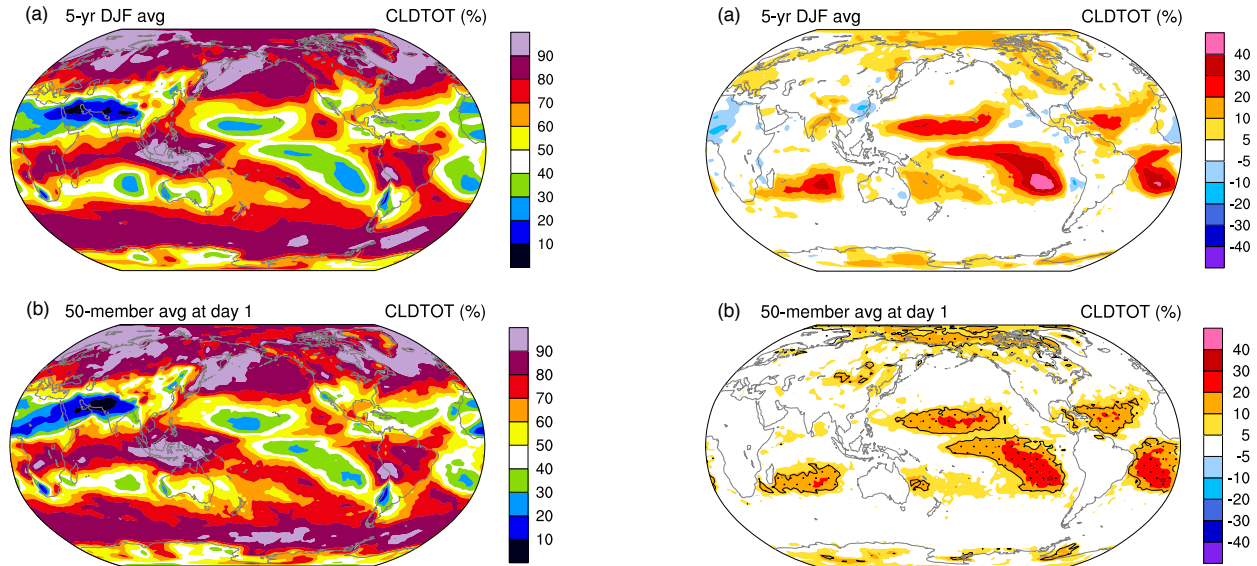


Fig. 1. Total cloud cover (unit: %) in CAM5 simulations using the default model time step (30 min). **(a)** 5 yr December–January–February (DJF) average from a long-term climate simulation. **(b)** 50-member ensemble average of the first model day in a set of short simulations. Further details are explained in Sect. 3.1.

ice concentrations. Emissions of aerosols and reactive gases are specified at their year 2000 values following Lamarque et al. (2010). Model state variables, including the meteorological fields, aerosol concentrations, and land surface variables, are archived at 5 day intervals in the “native” format of the initial condition files.

This initialization procedure requires minimal effort because output from a prior simulation can be used directly in the ensemble simulations, or conveniently interpolated for studies that involve different spatial resolutions. Also, the same archive can be used in different sensitivity studies. When the model configuration (e.g., parameters, resolution, or time step) changes, the simulated climate can change accordingly, in which case the integrations starting from the above-mentioned initial conditions will need some time to adjust before entering the new quasi-equilibrium. Identification of the spin-up phase is one of the issues we attempt to address in the following sections, and we will demonstrate that the initial adjustment is indeed short in the examples shown in Sects. 3 and 4.

3 Example I: time step sensitivity of clouds

In this section the utility of the ensemble approach is demonstrated using simulations in which the model time step used in CAM5 is reduced from the default value of 30 min to 4 min. We focus on cloud and precipitation related model variables. Our example is motivated by the desire (in a separate study with a focus on scientific issues) to characterize

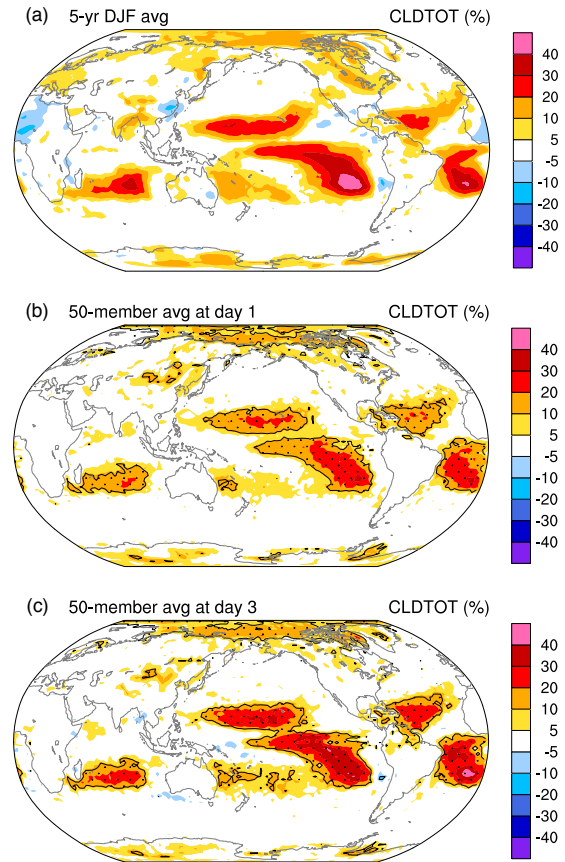


Fig. 2. Differences in total cloud cover (unit: %) between simulations using 4 min and 30 min time steps. **(a)** 5 yr DJF average from a climate simulation performed in the traditional way; **(b)** 50-member ensemble average of the first simulation day; **(c)** as in **(b)** but for the third simulation day. Stippling in panels **(b)** and **(c)** indicates where the differences are statistically significant at the 95 % confidence level. Further details are explained in Sect. 3.2.

the time step sensitivities of the atmospheric water cycle in CAM5 and assess numerical convergence. Since we also want to distinguish different climate regimes, the analysis here focuses on a particular season (boreal winter) to avoid the additional complexity introduced by seasonal variations in geographic locations.

For evaluation purposes, two simulations (with 30 and 4 min time steps respectively) were first performed for 5 years (plus 1 yr spin-up) in the conventional way. Sensitivities in the simulated climate were identified by comparing fields from the multi-year December–January–February (DJF) averages of the two simulations. This pair of simulations are regarded hereafter as the *reference* simulations.

We also performed ensembles of short simulations with the two time step lengths. Within an ensemble, all realizations were assigned a start time of 1 January using different initial conditions drawn from the 20 yr archive (cf. Sect. 2) of dates in the DJF season that were at least 10 days apart,

Table 1. Global mean values and pattern correlations of the atmospheric mean state in a multi-year climate integration and an ensemble of short simulations. The 5 yr mean climatology of December–January–February (DJF) is compared to the 50-member mean of day one average. All simulations are performed with a 30 min model time step. Further details are explained in Sect. 3.1.

	Global Mean			Pattern Correlation
	5-yr DJF Avg.	Day 1 Ensemble Avg.	Relative Difference	
Sea level pressure (hPa)	1011.77	1011.76	−0.0 %	0.99
Planetary boundary layer height (m)	486.57	478.58	−1.6 %	0.98
10 m wind speed (m s^{-1})	5.98	5.96	−0.3 %	0.99
Temperature at reference height (K)	285.02	285.05	0.0 %	1.00
Relative humidity at reference height (%)	79.59	80.13	0.7 %	0.98
Total precipitable water (kg m^{-2})	24.34	24.30	−0.1 %	1.00
Vertically integrated total cloud fraction	0.65	0.66	1.3 %	0.96
Low-level cloud fraction	0.46	0.47	2.3 %	0.98
Mid-level cloud fraction	0.28	0.28	0.0 %	0.95
High-level cloud fraction	0.38	0.38	0.7 %	0.94
Liquid water path (g m^{-2})	42.24	41.92	−0.8 %	0.98
Ice water path (g m^{-2})	17.62	17.66	0.2 %	0.97
Shortwave cloud forcing (W m^{-2})	−54.00	−53.91	−0.2 %	0.98
Longwave cloud forcing (W m^{-2})	23.08	23.06	−0.1 %	0.98
Large-scale precipitation rate (mm day^{-1})	0.89	0.88	−0.9 %	0.93
Convective precipitation rate (mm day^{-1})	2.06	2.07	0.2 %	0.98

in order to ensure independence and representativeness. The same set of initial conditions were used for the 30 min and 4 min ensembles.

Other aspects of the simulation set-up were identical for the long-term and short ensemble simulations. For example, both were forced by yearly cycled climatological sea surface temperature and sea ice concentrations, as well as the year 2000 emissions for aerosols and their precursors (Lamarque et al., 2010).

In the analysis we concentrate on the geographical distribution and radiative properties of clouds. We focus on whether:

- I. ensemble averages of a moderate number of realizations can reasonably represent the long-term climate;
- II. integrations of a few days are sufficient to get rid of the spin-up phase;
- III. responses to time step change detected with the ensemble approach agree quantitatively well with those revealed by the conventional long-term simulations;
- IV. there is a clear gain of computational efficiency.

For evaluation purposes, the ensemble simulations were integrated for a period of 20 days.

3.1 Representing the mean state

Our evaluation of the ensemble strategy starts with question (I) by examining the mean state simulated with the default time step (30 min). Because the initial conditions were generated using the same model configuration and experimental

setup (in other words, sampled from the same climate), there is no spin-up issue here. The question is how many realizations do we need to average out the “weather noise” and obtain the “climate signal”.

In Fig. 1, the vertically integrated total cloud amount is shown for the two different experimentation methods, where the 5 yr DJF average in the long simulation is compared with the day 1 average of a 50-member ensemble. The agreement with the reference simulation is remarkable. High cloud fractions associated with the Intertropical Convergence Zone (ITCZ), the South Pacific Convergence Zone (SPCZ), mid-latitude storm tracks, and high latitude regions in the winter hemisphere are well captured. Less frequent occurrences of clouds over the subtropical ocean high pressure systems and desert areas are also well represented. It is worth noting that the moderately sized ensemble not only reproduces these basic features of the geographical distribution, but also captures the magnitude of total cloud amount quite well at most grid points.

The same can be said for other key aspects of the model climate, as can be seen in Table 1. Global mean values in the 5 yr simulation and the one-day ensemble differ only by a few percent at most, and the pattern correlations are high (> 0.9). This suggests that, at least for the default model configuration, the ensemble average of very short integrations is a good representation of the long-term climatology. In the following, we demonstrate that the ensemble simulations are also able to accurately reproduce the response of cloud-related fields to parameterization changes (in this case time step length).

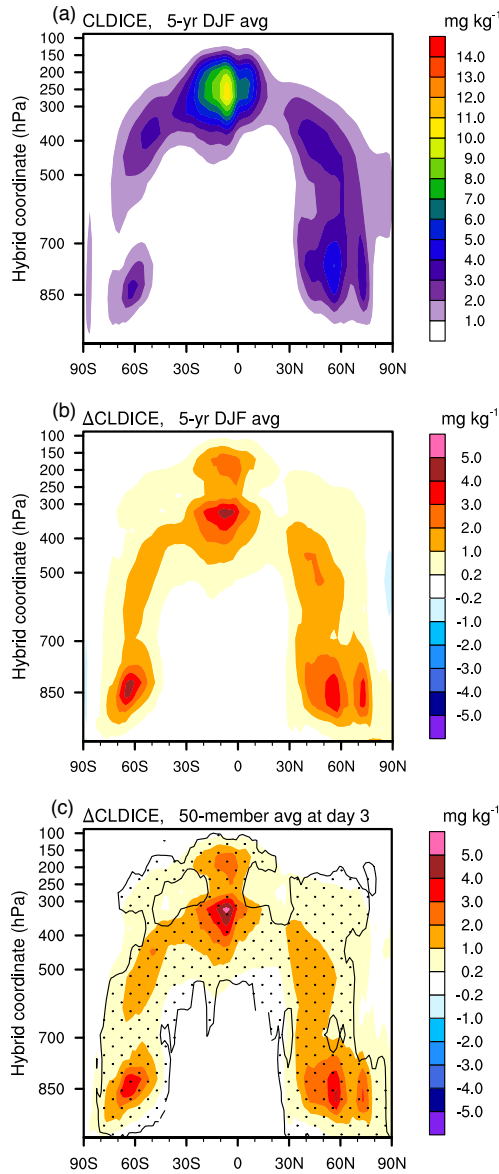


Fig. 3. (a) Zonally averaged, 5 yr DJF mean mass concentration of stratiform cloud ice (unit: mg kg^{-1}) simulated by CAM5 using a 30 min time step. (b) 5 yr DJF mean cloud ice mass concentration differences between simulations using 4 min and 30 min time steps (unit: mg kg^{-1}). (c) As in panel (b) but showing the 50-member ensemble mean at day 3. Stippling in panel (c) indicates locations where the differences are statistically significant at 95 % confidence level. Further details are explained in Sect. 3.2.

3.2 Fast response of clouds

Our ensemble simulations with 4 min time step are initialized using snapshots of atmospheric and land-surface conditions sampled from the model climate resulting from a 30 min time step. Because cloud processes operate on short timescales, we expect quick responses to changes in model time step. This is indeed observed in the simulations.

The 5 yr simulations indicate that a reduction from 30 to 4 min time step leads to an overall increase of total cloud cover. The most prominent signals occur in the shallow cumulus regions where the absolute changes range from about 10 % to more than 40 % in boreal winter (Fig. 2a), corresponding to typical relative changes from 20 % to more than 100 % (not shown). Such characteristic patterns are apparent in the ensemble simulations in the first model day (Fig. 2b). Although the differences are somewhat smaller than the 5 yr DJF averages, statistical tests suggest they are significant at the 95 % confidence level. By the third day, the magnitudes of the differences between 4 and 30 min simulations are close to those seen in the 5 yr average.

The ensemble simulations can also capture changes in vertical structures. Figure 3 shows the zonally averaged stratiform cloud ice mass concentration as an example. According to the 5 yr simulations (Fig. 3b), shorter time step leads to higher ice concentrations throughout the troposphere. The largest increases occur in the lower levels over the storm tracks, and in the tropical upper troposphere between 300 and 400 hPa where deep convection detrains condensate into the environment. There is a secondary center of large increase near 150 hPa, corresponding to frequent homogeneous ice nucleation. The close resemblance between the 5 yr DJF average (Fig. 3b) and the day 3 average of the ensemble results (Fig. 3c) indicates that the characteristic distributions of cloud ice are well established within a couple of model days.

We reinforce the conclusion of fast spin-up by showing the day-to-day variation of global mean cloud cover (Fig. 4a), vertically integrated liquid and ice water path (LWP and IWP, Fig. 4b–c), longwave and shortwave cloud forcing (LWCF and SWCF, Fig. 4d–e) and large-scale precipitation rate (PRECL, Fig. 4f) from the ensemble simulations. The 95 % confidence intervals of the ensemble averages are shown by the filled boxes, with the ensemble mean indicated by the horizontal bar in each box. Lower and upper ends of the whiskers, corresponding to the 10th and 90th percentiles, depict the ensemble spread. The 5 yr DJF and January averages, as well as the year-to-year variations, are shown in the right part of each panel for comparison. Figure 4 indicates that during the 20 day integration period of the ensemble simulations, there is no obvious trend either in the 50-member averages of the 30 min and 4 min ensembles or in their differences; in addition, the ensemble averages agree reasonably well with the 5 yr averages. Therefore, for detecting fast changes in cloud properties and distribution, it is sufficient to perform simulations that are only a few days in length. The additional computing time spent on longer integration does not provide significantly more information.

Figure 4 also provides a quantification of the time step sensitivity in the depicted variables with respect to their natural variability. Comparing differences between the 4 min and 30 min ensemble averages with the ensemble spreads, one can conclude that the total cloud fraction, LWP and

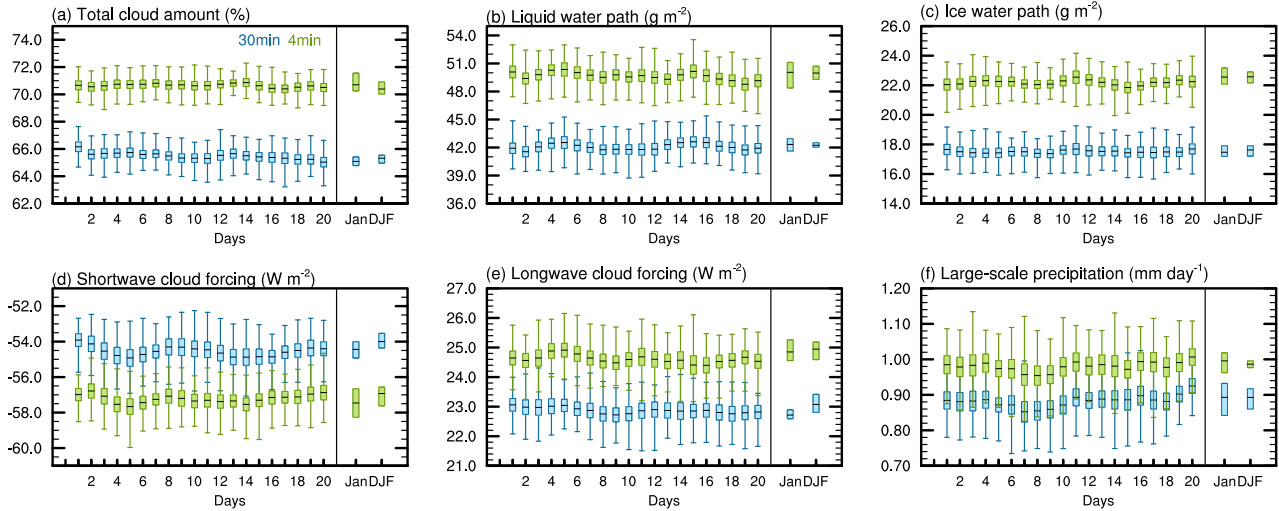


Fig. 4. Global mean values of some cloud-related variables from the 50-member ensembles and from the 5 yr climate simulations. Blue and green indicate simulations performed with 30 min and 4 min time steps, respectively. Left part of each panel show the daily mean global averages in the first 20 days of the ensemble simulations. Lower and upper ends of the whiskers denote the 10th and 90th percentiles. The hinges in the middle indicate ensemble mean. The filled boxes show the 95 % confidence interval of the mean. In the right part of each panel, the January and DJF averages of the 5 yr climate simulations are shown. The bottom and top of each box correspond to the minimum and maximum January or DJF averages in the simulation period. Hinges in the middle indicate the 5 yr average. Further details are explained in Sect. 3.2.

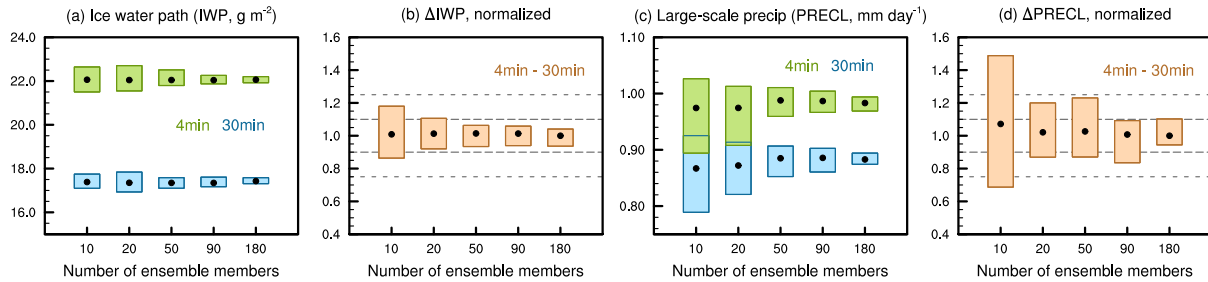


Fig. 5. Impact of ensemble size on the estimated (a) global mean ice water path, (c) global mean large-scale precipitation rate, and (b, d) their sensitivity to model time step. In panels (a) and (c), blue and green correspond to simulations performed with 30 min and 4 min time steps, respectively. The dots inside filled boxes are ensemble mean values averaged over the entire integration period (20 days). Top and bottom of the boxes denote the maximum and minimum daily averages. Similarly, the 4 min/30 min differences are shown in panels (b) and (d), except that all values are normalized by the 20 day average of the 180-member ensemble mean, in order to show the relative differences among the estimates associated with different ensemble sizes. Further details are explained in Sect. 3.3.

IWP (Fig. 4a–c) are more sensitive to the time step change ³⁵⁰ than SWCF, LWCF and PRECL are (Fig. 4d–f). On the other hand, although individual members from the 30 min and 4 min ensembles can have the same globally averaged SWCF, LWCF, or PRECL, the 95 % confidence intervals of the ensemble mean do not overlap on any day during the integration period, suggesting that the time step sensitivities are nevertheless statistically significant. ³⁵⁵

3.3 Ensemble size

So far we have shown the effectiveness of the short simulations using results from 50-member ensembles. We now demonstrate the robustness of the method and discuss the choice of ensemble size. The essence of the experimentation method we propose in this paper is to approximate the long-term temporal average by the ensemble average over a short period. The accuracy of this approximation naturally depends on the ensemble size and properties of the state variable in question.

3.3.1 Global averages

In Fig. 5 the accuracy of estimated global averages is analyzed for the ice water path and large-scale precipitation rate. Results are shown for 10, 20, 50, 90, and 180 ensemble members. At each ensemble size, the day-to-day variability of the ensemble mean daily average is indicated by the vertical extent of a filled box, with its top and bottom showing the maximum and minimum values during the 20 day simulation period. The 20 day averages are denoted by the black dot in each box. Based on the conclusion drawn from Fig. 4 about fast spin-up, it is reasonable to assume when a sufficiently large number of realizations are obtained, the ensemble mean values averaged over 20 days will indicate the long-term climatological mean within a small uncertainty induced by natural variability. Thus the vertical size of a colored box in Fig. 5 can be used as a measure of approximation error in the global averages estimated from a single-day simulation at the corresponding ensemble size. In panels Fig. 5b and d, the 4 min/30 min differences are normalized by the 20 day average of the 180-member ensemble mean, in order to indicate the uncertainties in relative terms.

Figure 5 conveys several messages. First, the 20 day averages change very little with ensemble size (Fig. 5a and c), suggesting that the sampling method is representative in capturing the impact of time step change. Second, as the number of independent realizations increases, variances in the daily average decrease, producing more accurate approximations of the long-term climatology (Fig. 5a and c). Third, different model variables are associated with different variability, thus require different numbers of realizations. Last but not least, good estimates of the time step sensitivities can be obtained with rather small ensembles. With 20 independent members, the global mean ΔIWP ($\Delta PRECL$) calculated from a one-day simulation agrees within 10% (25%) with the 180-member 20 day average (Fig. 5b and d). Similar accuracies are obtained in our experiments for the other variables shown in Fig. 4 for global averages. To capture regional differences, some highly variable fields may need more realizations, as discussed below.

3.3.2 Climate regimes

When assessing model sensitivities, it is often necessary to examine not only global averages, but also regional features and climate regimes. Because clouds are highly variable in their occurrence and properties, regional patterns are sometimes difficult to detect due to the low signal-to-noise ratio. For example, Kooperman et al. (2012) showed that to get a clear signature of the anthropogenic aerosol indirect effect, it is necessary to run CAM5 for multiple decades in conventional climate simulations.

Figure 6 shows the SWCF changes ($\Delta SWCF$) caused by a reduction of model time step in the 5 yr integrations (Fig. 6a) and at day 3 in the ensemble simulations (Fig. 6b

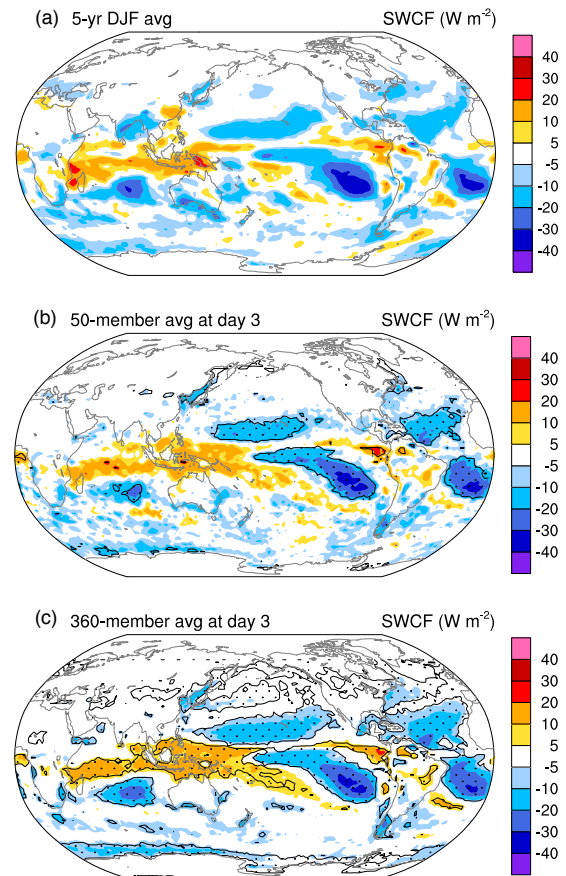


Fig. 6. Shortwave cloud forcing differences (unit: $W m^{-2}$) between simulations using 4 min and 30 min time steps. (a) 5 yr DJF average from a climate simulation performed in the traditional way; (b) day 3 average from ensemble simulations with 50 independent members; (c) day 3 average from ensemble simulations with 360 independent members. In panels (b) and (c), stippled regions are associated with differences significant at the 95% confidence level.

and c). Both methods reveal a systematic increase of SWCF in the trade cumulus regions, while the reduction of cloud forcing in the ITCZ and SPCZ are more clearly seen in the ensemble results. It is remarkable that the convergence zones emerge clearly in Fig. 6b with only 50 ensemble members, a result attributable to our initialization method that uses the same set of initial conditions for the 4 min and 30 min simulations. In the early stage of the integration (first ~ 5 days), the large-scale environmental conditions remain similar in each pair of ensemble members, resulting in the synoptic systems and convective activities occurring at similar locations, thus avoiding strong noise in $\Delta SWCF$ associated with synoptic scale variabilities in the circulation.

In Fig. 6b, $\Delta SWCF$ in the shallow cumulus and deep convection regions have similar magnitudes, but those in the latter regime do not pass the statistical test because of the

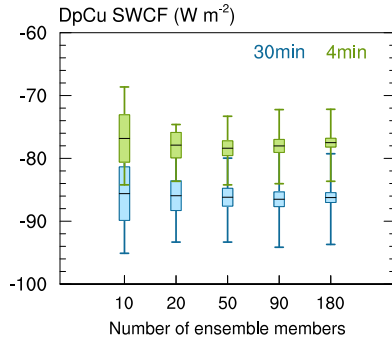


Fig. 7. Shortwave cloud forcing (SWCF) associated with tropical deep convection in simulations performed with 30 min (blue) and 4 min (green) time steps. The SWCF is averaged over tropical ocean grid points where the daily mean convective precipitation rate exceeds 3 mm day^{-1} . Like in Fig. 4, horizontal bars in the middle of filled boxes indicate the mean value of each ensemble. Lower and upper ends of the whiskers correspond to the 10th and 90th percentiles, respectively. Filled boxes show the 95 % confidence interval of the ensemble mean. Further details are explained in Sect. 3.3.

large natural variability of deep convection in the convergence zones. The evaluation procedure can be made more robust using regime compositing, e.g., by assessing the SWCF changes over tropical ($20^\circ \text{ S} - 20^\circ \text{ N}$) ocean grid points where the convective precipitation is important, in this case defined to be where the convective precipitation rate exceeds 3 mm day^{-1} . Figure 7 indicates that for such a “deep convection SWCF”, ensembles of 20 members are sufficient to distinguish the difference between the 4 min and 30 min simulations. For the purpose of verification, we performed additional simulations and present them in Fig. 6c. The 360-member results confirm that the ΔSWCF patterns and magnitudes detected by the 50-member ensembles are not incidental.

3.4 Combining ensembles with nudging

The nudging technique has been repeatedly used in model evaluation and intercomparison studies as a method for constraining model meteorology, reducing uncertainties induced by natural variability, and facilitating comparison with observations (e.g., Jeuken et al., 1996; Feichter and Lohmann, 1999; Machenhauer and Kirchner, 2000; Ghan et al., 2001; Kooperman et al., 2012). Here we briefly show that nudging can be applied in combination with ensembles to assess the contribution of physics-dynamics interaction to the model’s time step sensitivity.

Two sets of ensemble simulations, each with 50 members, were performed with the 30 min and 4 min time steps, respectively, with the horizontal wind and temperature relaxed towards those from the unconstrained 30 min time step simulations. Each pair of control (30 min time step) and sensitiv-

ity (4 min) experiments starting from the same initial conditions were nudged to the same temperature and wind fields, while different pairs were relaxed towards different large-scale conditions. As in Kooperman et al. (2012), a 6 h relaxation time was used.

Figure 8 compares the globally/regionally averaged total cloud fraction, IWP and SWCF in the free-running and nudged simulations. SWCF in the deep and shallow convection regions are presented separately because the two regimes are associated with opposite sensitivities to time step (cf. Fig. 6). As expected, the unconstrained and nudged 30 min simulations give very similar results. The ensemble averages are not distinguishable in a statistical sense, the 95 % confidence intervals are comparable, and the ensemble spreads are also similar. The 4 min simulations, on the other hand, are significantly different. When wind and temperature are constrained, the differences between 4 min and 30 min simulations are reduced by about 30 % for the variables shown in the figure, suggesting that fast interactions (feedbacks) between resolved dynamics and parameterized physics increase the time step sensitivity of the CAM5 model.

3.5 Computational efficiency

The results presented above provide clear answers to the questions posed at the beginning of this section. To detect time step sensitivities in cloud-related fields, it is sufficient to perform 20 to 50 independent 3 day simulations. The ensemble method reveals signals that agree well with those detected by 5 yr simulations performed in the traditional way, but costs substantially less total CPU time, and dramatically less in terms of the experiment “completion time” in situations where there are more processors available than a single job can use effectively, or is allowed to use without a long queuing time, and many realizations can be run simultaneously. Our experience showed that on the Yellowstone supercomputer (Computational and Information Systems Laboratory, 2012) at the National Center for Atmospheric Research (NCAR) Computational and Information System Lab (CISL), a 5 yr simulation with 4 min time step typically takes about 4 to 7 days of wall-clock time to finish with 64 processes running in parallel. (The actual duration depends on the amount of model output as well as traffic in the queuing system.) For the ensemble simulations, a set of fifty 3 day simulations usually takes less than 20 min to finish, counting from the instant when the jobs are submitted till the point when the last job is completed, resulting in a reduction of turnaround time by a factor of several hundred. Such fast turnaround will be particularly helpful when additional simulations are conducted with varied model configurations to identify the source of the time step sensitivity, and when even smaller time steps are used to assess the convergence properties of the model behavior.

From the results above, we conclude that the ensemble method as applied is both effective and efficient for the in-

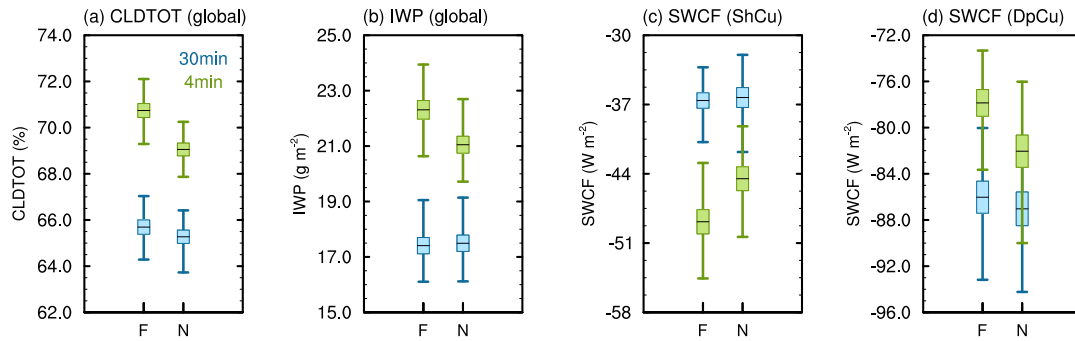


Fig. 8. Comparison of the free-running (“F”) and nudged (“N”) simulations performed with 30 min (blue) and 4 min (green) time steps. Meanings of the whiskers, boxes and hinges are the same as in Figs. 4 and 7. Each ensemble consists of 50 independent members. In the nudged simulations, temperature and horizontal wind are relaxed towards those from the 30 min time step unconstrained simulations, using a nudging time scale of 6 h. Panels (a) and (b) show the globally averaged, vertically integrated total cloud fraction and ice water path, respectively. Panels (c) and (d) show the shortwave cloud forcing (SWCF) averaged over the shallow and deep convection regions. SWCF associated with shallow convection is the average over ocean grid points between 30° N and 30° S where the frequency of occurrence of shallow convection is larger than 0.5, and the daily mean convective precipitation rate is lower than 1 mm day⁻¹. SWCF associated with deep convection is calculated in the same way as in Fig. 7 (cf. Sect. 3.3.2). All results are shown for the third simulation day.

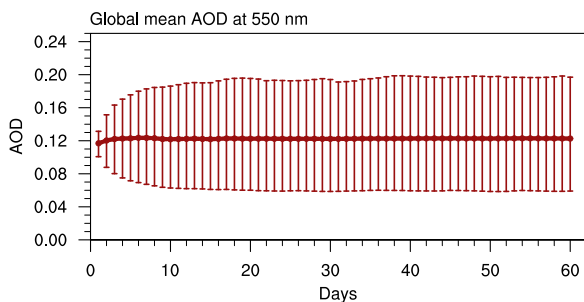


Fig. 9. Time evolution of the global mean aerosol optical depth (AOD, at 550 nm wavelength) in the ensemble simulations described in Sect. 4. The thick curve shows the AOD averaged over 256 ensembles that used different values for 16 empirical parameters in the CAM5 model (cf. Table 2). Vertical bars indicate the spread (minimum to maximum) among the 256 ensembles.

investigation of time step sensitivity. In the next section, we use an additional example to show that the method can also be very useful in other sensitivity studies.

4 Example II: parametric sensitivity of the global mean top-of-atmosphere radiation balance

The parameterization schemes of sub-grid scale processes in AGCMs include various empirical, uncertain constants whose values are often adjusted to obtain desired radiation balance at the top of the model atmosphere (TOA), and to achieve good fidelity when evaluated against observations (e.g., Mauritsen et al., 2012; Golaz and Levy, 2013). There is a large volume of literature discussing the sensitivities of

model behavior to empirical parameters. In the context of global climate change, there is also increasing interest in assessing the impact of such parameters on the uncertainties in future climate projections (e.g., Murphy et al., 2004; Stainforth et al., 2005; Collins et al., 2006).

Because there are a large number of adjustable parameters in AGCMs, and many of them have wide ranges of possible values, systematic investigations of model sensitivity inevitably require numerous simulations. Earlier studies that varied the value of one parameter at a time (e.g., Lohmann and Ferrachat, 2010) only covered very small portions of the full parameter space. In recent years, the use of advanced sampling approaches such as Latin hypercube (McKay et al., 1979) and quasi-Monte Carlo method (Caflich, 1998) have allowed more extensive explorations of the parameter space (e.g., Lee et al., 2012, 2013; Zhao et al., 2013). Perturbing multiple parameters simultaneously not only allows for a dramatic reduction of the number of simulations needed for the sensitivity study, but also provides the opportunity to investigate parameter interactions, leading to a more comprehensive understanding of model sensitivity.

On the other hand, even with efficient sampling approaches applied, systematic investigations of parametric sensitivity are still inherently expensive because of the high dimensionality of the parameter space. For instance, to simultaneously perturb $\mathcal{O}(10^1)$ parameters, one needs to sample $\mathcal{O}(10^2 \sim 10^3)$ points from the parameter space to ensure sufficient coverage. Performing long-term climate simulations with this many model configurations requires a substantial amount of computer time. In this section, we demonstrate that there are circumstances in which a very good characterization of the parametric sensitivity can be obtained with small ensembles of short integrations, resulting in a signifi-

Table 2. Empirical parameters in the cloud microphysics and aerosol lifecycle parameterizations in CAM5 that are perturbed in the sensitivity analysis described in Sect. 4. Adapted from Table 1 in Zhao et al. (2013).

Index	Parameter	Description	Unit	Default value	Investigated range
1	ai	Fall-speed parameter for cloud ice	s^{-1}	700	350–1400
2	as	Fall-speed parameter for snow	$\text{m}^{0.59} \text{s}^{-1}$	11.72	5.86–23.44
3	cdnl	Lower bound of grid-box mean cloud droplet number concentration	m^{-3}	0	0–1e7
4	dcs	Size threshold assumed for the autoconversion of cloud ice to snow	μm	400	100–500
5	wsubmin	Minimum sub-grid vertical velocity assumed for cloud droplet activation	m s^{-1}	0.2	0–1
6	facti	Solubility factor for the removal of cloud-borne aerosols in stratiform clouds	–	1.0	0.5–1
7	factic	Solubility factor for the removal of interstitial aerosols in convective clouds	–	0.4	0.2–0.8
8	ref_dust	Imaginary part of the refractive index of dust in visible bands of the solar radiation	–	0.005	0.001–0.01
9	e_dust	Emission tuning factor for dust	–	0.35	0.21–0.86
10	e_sst	Emission tuning factor for sea salt	–	1.0	0.5–2.0
11	e_soag	Emission tuning factor for secondary organic aerosols	–	1.5	0.5–2.0
12	e_so2	Emission tuning factor for anthropogenic SO_2	–	1.0	0–2
13	e_bc	Emission tuning factor for anthropogenic black carbon	–	1.0	0–3
14	e_pom	Emission tuning factor for anthropogenic particulate organic matter	–	1.0	0–3
15	e_acnum	Emission tuning factor for aerosol number concentration in the accumulation mode	–	1.0	0.3–5.0
16	e_so4f	Tuning factor for the fraction of SO_2 emitted as sulfate	–	0.025	0–0.05

cant reduction in the computational cost.

4.1 Reference simulations

A recent study by Zhao et al. (2013) investigated the sensitivity of TOA radiative fluxes in present-day climate simulations to the values of 16 parameters in CAM5. The 16 parameters included 5 adjustable constants related to stratiform cloud microphysics (indices 1–5 in Table 2), 3 parameters related to the physical properties of aerosols (indices 6–8 in Table 2), and 8 scale factors for aerosol emissions (indices 9–16 in Table 2). To efficiently explore the high-dimensional parameter space, the quasi-Monte Carlo sampling method (Cafisch, 1998) was chosen for its good performance in terms of sample dispersion. From the 16-dimensional parameter space, 256 sample points were drawn. Each sample point corresponds to one set of values for the 16 parameters, which we refer to as a “parameter combination” in the following. For each parameter combination, an AMIP (Atmospheric Model Intercomparison Project, Gates et al., 1998) simulation was conducted for the years 2000 to 2004. The average of the last 4 years (2001–2004) was used in their sensitivity analysis to identify which parameters have the largest impact on the model’s radiation budget.

4.2 Short ensembles

In this study we demonstrate that it is possible to use short ensembles to reproduce the results of Zhao et al. (2013). The same 256 parameter combinations were used in our simulations, while each of their 4+1-year AMIP runs were replaced by an ensemble of short simulations started in each month of the year, so that the ensemble averages characterize the annual averages examined in the reference study. As in Sect. 3, the initial conditions were taken from a prior long-term simulation. The same set of 12 initial conditions were used for all 256 ensembles.

4.3 Spin-up time

It is worth noting that 11 out of the 16 perturbed parameters (indices 6 through 16 in Table 2) directly affect the concentrations of aerosols. How these aerosol-related parameters affect the TOA radiative fluxes is a key question to be answered by the sensitivity analysis. The AMIP simulations of Zhao et al. (2013) were initialized with zero aerosol mass and number concentrations. Such an initialization in CAM5 usually requires a spin-up of several months (or longer) before the aerosol concentrations have evolved and approach the climatological values. Therefore the first simulation year was discarded in the study of Zhao et al. (2013).

Table 3. Global mean TOA net radiative flux (FNET), total cloud forcing (CF), shortwave cloud forcing (SWCF), and longwave cloud forcing (LWCF) in the parametric sensitivity simulations described in Sect. 4. The numbers given are the average of 256 simulations/ensembles \pm one standard deviation (unit: W m^{-2}).

Variable	4 yr average	12-member average at day 10
FNET	-3.73 ± 2.36	-3.84 ± 2.63
CF	-33.70 ± 5.64	-34.07 ± 7.07
SWCF	-56.20 ± 2.13	-56.42 ± 2.12
LWCF	$+22.49 \pm 2.14$	$+22.35 \pm 2.56$

For our ensembles, all simulations were started with aerosol concentrations that were spun-up under the default model configuration and were consistent with the corresponding meteorological fields. This set-up is expected to require shorter spin-up than the zero-aerosol conditions. On the other hand, after the aerosol emissions, solubility factors, and cloud parameters were perturbed (Table 2), we expect an initial adjustment by at least a few days, considering that the global mean aerosol lifetime is about 4 days in MAM3 (cf. Tables 3, 5–8 in Liu et al., 2012). To get a quantitative assessment of the spin-up time, we monitored the time evolution of the aerosol optical depth (AOD) in the ensemble simulations. In Fig. 9 the global mean AOD is shown for the first 60 days. The daily mean values averaged over the 256 ensembles are indicated by the thick curve. Variations among the ensemble averages are shown by the vertical bars, with the lower and upper ends indicating the minimum and maximum, respectively. As expected, the globally averaged AOD of different ensembles are similar at the beginning of the integration. They quickly diverge in the next few days and then stabilize. After about 10 days, there is no substantial changes in either the average or the spread of the 256 ensembles. We thus use the day-10 average for the sensitivity analysis below. In other words, we compare the parametric sensitivities derived from the 12-member ensemble averages at day 10 with the results in Zhao et al. (2013) which were based on 4 yr averages.

4.4 Global mean radiation budget

Our analysis starts with the TOA net radiative flux (FNET). To give a first sense of the model's response to the parameter perturbation, Table 3 lists the mean and standard deviation of the 256 simulations/ensembles. Similar statistics are presented in the same table for the total cloud forcing (CF), as well as for the shortwave and longwave cloud forcing (SWCF and LWCF). The mean FNET obtained with the two methods differ by about 3 % (0.11 W m^{-2}), while the discrepancies in CF, SWCF and LWCF are smaller in terms of relative differences. Variations among the 256 experiments tend to be somewhat smaller in the 4 yr AMIP simulations, probably because the substantially larger number of days involved in the temporal average leads to a stronger smoothing effect.

The sensitivities of global mean FNET to individual parameters are shown in Fig. 10. In each panel, the global mean 4 yr averages (Fig. 10, upper row) or day-10 ensemble averages (Fig. 10, lower row) corresponding to the 256 parameter combinations are sorted into 8 bins according to the value of one perturbed parameter. The square mark associated to each bin indicates the FNET anomaly (relative to the average of 256 experiments) averaged within that bin. The vertical bars depict the spread (minimum to maximum) caused by the perturbation of other parameters.

The 4 yr AMIP simulations (Fig. 10, upper row) indicate that the global mean FNET increases with `dcs` (the size threshold between cloud ice and snow) and `factic` (solubility factor of aerosols in convective clouds). It decreases with `ai` (a fall-speed parameter for cloud ice), `cdnl` (the minimum concentration of cloud droplet number), `wsubmin` (the minimum sub-grid vertical velocity for droplet activation), and `e_sst` (tuning factor for sea salt emission). FNET is less sensitive to the other parameters.

As discussed in Zhao et al. (2013), the detected sensitivities in FNET are mainly attributable to clouds, while the contribution of clear-sky areas is relatively small. Therefore, in Fig. 11 we present the responses of SWCF and LWCF to the parameter perturbation. The longwave cloud forcing is primarily affected by cloud microphysics parameters `ai`, `as`, `cdnl`, and `dcs`. The shortwave cloud forcing is additionally affected by `wsubmin`, and the aerosol-related parameters `factic` (solubility factor of aerosols in convective clouds) and `e_so2` (tuning factor for the emission of anthropogenic SO_2).

Comparing the upper and lower rows of Figs. 10 and 11, we see that not only the qualitative conclusions drawn in the previous paragraphs, but also the quantitative details of the functional relationships between FNET/SWCF/LWCF and the perturbed parameters, are correctly reproduced by the short ensembles. Considering that the ensemble results used to derive these relationships were averaged over only 12 realizations and one model day, the agreement with the 4 yr climate simulations is rather remarkable.

In Figs. 10 and 11, the relative contributions of individual parameters to the total variation of FNET, SWCF and LWCF are noted above the corresponding panels. These numbers were obtained by applying the generalized linear model (GLM) which assumes the relationships between the output variables (i.e., FNET, SWCF and LWCF) and input parameters are polynomial functions that include linear, quadratic, and interaction terms. Percentages given in red in the figures are statistically significant at the 95 % confidence level. Details of the GLM fitting are described in Sect. 2.3.3 in Zhao et al. (2013), and are not repeated in this paper. Here we only point out that the GLM provides a quantitative way to rank the relative importance of the empirical parameters in determining the total variation in the output variables. The rankings derived from the 4 yr simulations and the day-10 ensembles agree quite well. For example, `dcs`,

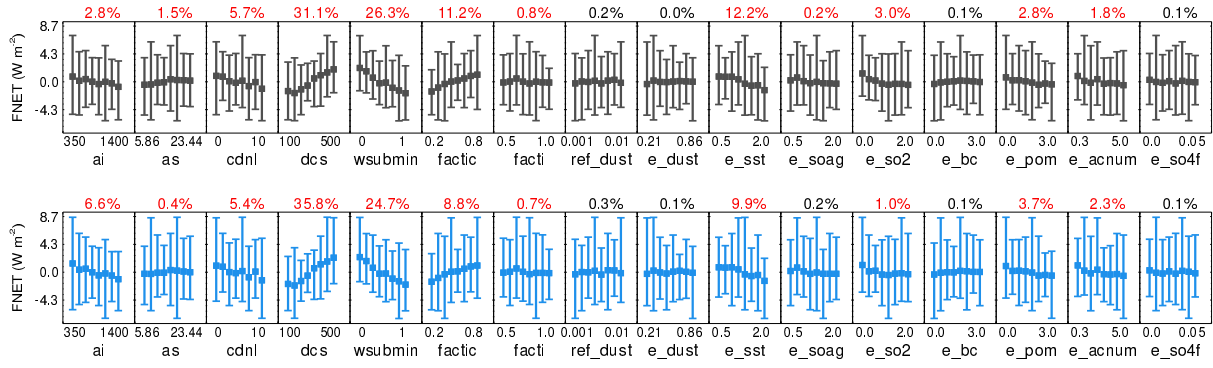
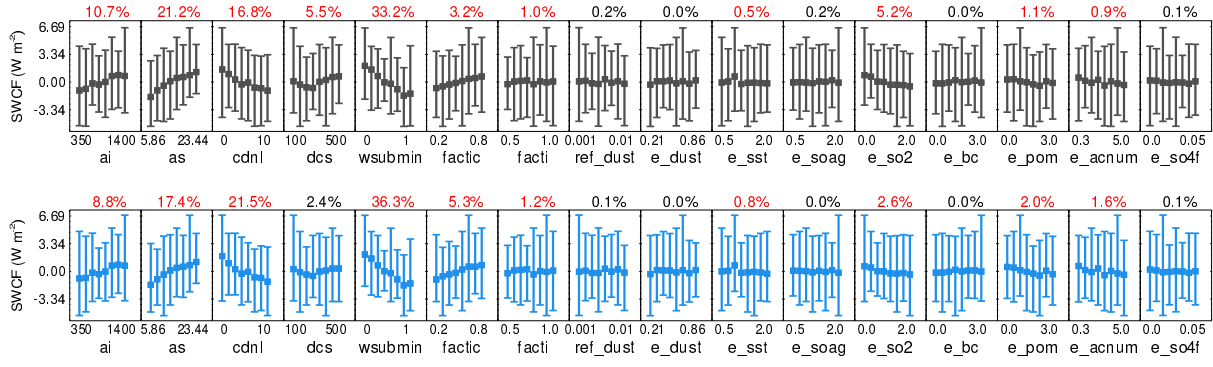


Fig. 10. Sensitivities of the global mean top-of-atmosphere net radiative flux (FNET, unit: W m^{-2}) to the empirical parameters listed in Table 2. CAM5 simulations were carried out using 256 different model configurations corresponding to 256 sampling points drawn from the 16-dimensional parameter space (cf. Table 2 and Sect. 4.1). In each panel, the global mean FNET corresponding to the 256 model configurations are sorted into 8 bins according to the values of one perturbed parameter. The spread (minimum to maximum) of FNET within a bin is shown by a vertical bar, while the mean value is indicated by a square mark. Note that the FNET shown here is the anomaly relative to the mean of the 256 simulations/ensembles. The mean values that have been subtracted are given in Table 3. Numbers noted above the panels are the relative contribution of individual parameters to the total variation of FNET, estimated using a generalized linear model (cf. Sect. 4.4). Red font means the contribution is statistically significant at the 95 % confidence level. The upper row shows results obtained from the AMIP simulations of Zhao et al. (2013) (cf. Sect. 4.1). The lower row shows results from the ensemble simulations performed in this study.

(a) Shortwave Cloud Forcing



(b) Longwave Cloud Forcing

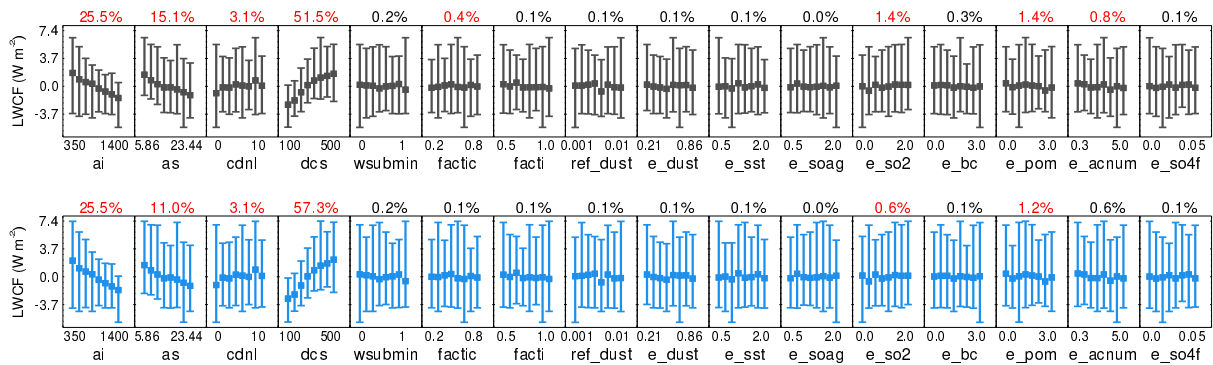


Fig. 11. As in Fig. 10 but showing parametric sensitivities of the global mean (a) shortwave and (b) longwave cloud forcing (unit: W m^{-2}).

`wsubmin`, `e_sst` and `factic` are identified by both methods as the most influential parameters for FNET. In terms of the percentage contribution of individual parameters to the total variance, the results derived from the 4 yr simulations and the short ensembles are also similar. There are a few cases in which the percentage is regarded as significant in the 4 yr simulations but insignificant in the ensembles (e.g., `e_soag` for FNET, `facti` and `e_acnum` for LWCF), but these are typically associated with small contributions, thus should not be considered as large discrepancies.

4.5 Computational efficiency

The twelve 10 day simulations cost about 1/15 of the total CPU time in comparison to the original 5 yr (4 years plus 1 year spin-up) simulations, a substantial reduction in computational cost. As for the turnaround time, on Yellowstone at NCAR CISL, the 256×12 simulations submitted as separate jobs finished within 8 h of wall-clock time. Typically the queuing system allowed 50 to 100 jobs running in parallel. These numbers were smaller than the total number of parameter combinations (256), therefore in this case, the reduction of turnaround time was mainly achieved from the smaller number of simulation days required by the ensemble strategy. On larger computing facilities that could allow more than 256 simultaneous jobs from a single user, it would be possible to make fuller use of the available resources using the ensemble strategy, but not with the long-term simulations. On a dedicated system that could accommodate $\mathcal{O}(10^3)$ concurrent simulations, it would be possible to complete all our ensemble simulations within an hour. Such fast turnaround can be very useful in systematic sensitivity studies, where influential parameters can be identified from a large number of candidates within a reasonable time period, and more attention can subsequently be paid to the most important parameters.

Furthermore, during the development of climate models, it is often necessary to adjust empirical parameters after major updates of model components, so that the long-term global mean TOA radiative flux stays close to zero. Since the global mean FNET, SWCF and LWCF are among the most important metrics for model tuning, our results suggest that short ensembles can be useful in such exercises as well.

5 Conclusions and discussion

We have demonstrated that ensembles of short simulations can be used to estimate the fast responses of a climate model to perturbations. The strategy can produce signatures that agree quantitatively and qualitatively with those produced by traditional multi-year brute force simulation strategies, at a fraction of the computational and wall-clock cost.

Our first example explored the response of simulated clouds to a change in model time step. The results sug-

gest that 3 day integrations are sufficient to reproduce the time step sensitivities seen in the commonly used 5 yr climate simulations due to the rapid response in cloud fields. For the global mean total cloud fraction, liquid water path and ice water path, the time step induced changes can be clearly detected with 20 ensemble members. For the global mean large-scale precipitation rate which has higher natural variability, and for the regional features of cloud forcing, robust signals can be detected from ensembles of 50 members. A combined use of ensemble and nudging led to the finding that interactions between the resolved dynamics and parameterized physics provide positive feedbacks that enhances the model's time step sensitivity.

The second example demonstrated that the strategy is capable of characterizing sensitivities of the global mean TOA radiation budget to 16 empirical parameters related to stratiform cloud microphysics and aerosol lifecycle. This type of investigation is inherently expensive in terms of computational cost, because a large number of simulations are needed to sufficiently sample the high-dimensional parameter space. Following a previous study by Zhao et al. (2013), we used the quasi-Monte Carlo method to obtain 256 sample points (parameter combinations) from the 16-dimensional parameter space. For each parameter combination, ensemble simulations were conducted with one realization starting from each month of the year 2001. We showed that parametric sensitivities of the global mean TOA net radiative flux and cloud forcing derived from 12-member ensemble averages at day 10 agree very well with results obtained by Zhao et al. (2013) who used 4 yr AMIP simulations in their analysis. The short ensembles correctly identified the most influential parameters for the net radiative flux and cloud forcing, and successfully reproduced the functional relationships between these quantities and the perturbed parameters.

These results indicates that although climate is by definition a long-term average, and is associated with strong natural variability, there exist fast processes and robust features that do not need very long simulations to characterize. This fact is already widely known, and has formed the foundation for the CAPT and Transpose-AMIP activities in which climate models are run in weather forecast mode to reveal the biases with respect to observations. Here we have shown that the philosophy behind the Transpose-AMIP-type evaluation can be applied in more general ways to carry out sensitivity studies. Using short ensembles instead of traditional multi-year climate simulations, sensitivity studies can be carried out more efficiently, benefitting from a substantial reduction of the total CPU time spent on numerical integration, as well as much faster turnaround in the investigation because the independent ensemble members introduce an additional dimension of parallelism that can be exploited with current flagship supercomputers.

The strategy discussed in this paper using few day simulations certainly has limitations. It cannot be used as formulated here to investigate modes of climate variability or

feedback mechanisms that operate on time scales of months to years, thus could not replace long-term simulations when long time scales are important. For example, in the time step sensitivity experiments discussed in Sect. 3, while the 5 yr simulations reveal an increase of DJF precipitation in the South Pacific Convergence Zone (SPCZ) when time step is shortened (not shown), the ensemble simulations do not indicate statistically significant differences in this region. This is probably because systematic changes in the SPCZ involve feedbacks from the large-scale circulation that can not sufficiently spin-up in just a few days.

Nevertheless, since fast processes are important contributors to the sensitivities and uncertainties in current climate models, short ensembles can help to obtain a first-order estimates of rapid responses in the climate system rather quickly. Such economical, approximate answers can be useful in various situations. For example, in systematic studies of parametric uncertainties, short ensembles can be used in preliminary investigations to pre-select influential parameters from a large number of candidates, and to narrow down possible ranges of parameter values. In convergence studies, short ensembles may be the only way to conduct simulations at ultra-high spatial and/or temporal resolutions that would otherwise be impractical to complete. As the climate modeling community actively pursues higher resolutions, more physically based parameterizations, and inclusion of new, highly sophisticated processes, wide applications can be anticipated for the experimentation method discussed here.

It should be noted that in this paper we are advocating the ensemble method as a general strategy, not a recipe. As can be seen from the two examples, for different variables and physical processes, one must generate ensembles differently, and may need different spin-up time and/or ensemble sizes. The most beneficial experimentation design for a particular research question needs to be figured out on a case-by-case basis. Whenever affordable, one should first evaluate the short ensembles against traditional climate simulations. If it is impractical to do so, we recommend testing the experimental design using a range of integration lengths and ensemble sizes, so as to obtain a better understanding of the robustness of the results.

An additional remark worth making here is that the definitions of fast and slow processes need to be understood in relative terms. In this paper where an atmosphere-only GCM was used, we considered time scales of a few days as “short”, and simulations of multiple years as “long”. In other situations, fast and slow processes can be reclassified. For example, if one were interested in identifying how seasonal features such as the Asian summer monsoon responded to anthropogenic and natural forcings (e.g., Ganguly et al., 2012; Vinoj et al., 2014; Song et al., 2014), or to changes in model formulation (e.g., Zhou and Li, 2002; Chen et al., 2010), it might be possible to generate realizations of few month simulations, and use ensemble averages to remove multi-year and multi-decade scale noise that would otherwise require hun-

dreds of years of simulations. As such, the ensemble strategy may have much wider applications than demonstrated in the present paper.

Acknowledgements. The authors thank Maoyi Huang, Qing Yang, Steve Ghan, Peter Caldwell, Heng Xiao and Ben Yang for valuable comments and suggestions. We also thank Cecile Hannay for technical help with the CESM model, and Gabriel Kooperman for sharing the implementation of nudging in CAM5. H. Wan is grateful for the support of the Linus Pauling Distinguished Postdoctoral Fellowship of the Pacific Northwest National Laboratory (PNNL). The study described in this paper was conducted under the PNNL Laboratory Directed Research and Development Program. P. J. Rasch was partially supported by the US Department of Energy (DOE) Office of Science as part of the Earth System Modeling Program. P. J. Rasch, K. Zhang, Y. Qian and C. Zhao acknowledge support from the Scientific Discovery through Advanced Computing (SciDAC) Program. Computing resources (ark:/85065/d7wd3xhc) were provided by the Climate Simulation Laboratory at NCAR’s Computational and Information Systems Laboratory, sponsored by the National Science Foundation and other agencies. PNNL is operated by Battelle Memorial Institute for DOE under contract DE-AC05-76RL01830.

References

- Boyle, J. S., Williamson, D., Cederwall, R., Fiorino, M., Hnilo, J., Olson, J., Phillips, T., Potter, G., and Xie, S.: Diagnosis of Community Atmospheric Model 2 (CAM2) in numerical weather forecast configuration at Atmospheric Radiation Measurement sites, *J. Geophys. Res.*, 110, D15S15, 10.1029/2004JD005042, 2005.
- Bretherton, C. S. and Park, S.: A New Moist Turbulence Parameterization in the Community Atmosphere Model, *J. Climate*, 22, 3422–3448, 10.1175/2008JCLI2556.1, 2009.
- Caflich, R. E.: Monte Carlo and quasi-Monte Carlo methods, *Acta Numerica*, 7, 1–49, 10.1017/S0962492900002804, 1998.
- Cess, R. D., Potter, G. L., Blanchet, J. P., Boer, G. J., Del Genio, A. D., Déqué, M., Dymnikov, V., Galin, V., Gates, W. L., Ghan, S. J., Kiehl, J. T., Lacis, A. A., Le Treut, H., Li, Z.-X., Liang, X.-Z., McAvaney, B. J., Meleshko, V. P., Mitchell, J. F. B., Morcrette, J.-J., Randall, D. A., Rikus, L., Roeckner, E., Royer, J. F., Schlese, U., Sheinin, D. A., Slingo, A., Sokolov, A. P., Taylor, K. E., Washington, W. M., Wetherald, R. T., Yagai, I., and Zhang, M.-H.: Intercomparison and interpretation of climate feedback processes in 19 atmospheric general circulation models, *J. Geophys. Res.*, 95, 16601–16615, 10.1029/JD095iD10p16601, 1990.
- Chen, H., Zhou T., Neale, R. B., Wu, X., and Zhang, G.: Performance of the new NCAR CAM3.5 in East Asian summer monsoon simulations: sensitivity to modifications of the convection scheme, *J. Climate*, 23, 3657–3675, 10.1175/2010JCLI3022.1, 2010.
- Collins, M., Booth, B. B., Harris, G. R., Murphy, J. M., Sexton, D. M., and Webb, M. J.: Towards quantifying uncertainty in transient climate change, *Clim. Dynam.*, 27, 127–147, 10.1007/s00382-006-0121-0, 2006.

- Colman, R.: A comparison of climate feedbacks in general circulation models, *Clim. Dynam.*, 20, 865–873, 10.1007/s00382-003-970-0310-z, 2003.
- Computational and Information Systems Laboratory: Yellowstone: IBM iDataPlex System (Climate Simulation Laboratory), national Center for Atmospheric Research, Boulder, CO, available at: <http://n2t.net/ark:/85065/d7wd3xhc> (last access: 1 April 2014), 2012.
- Dufresne, J.-L. and Bony, S.: An assessment of the primary sources of spread of global warming estimates from coupled atmosphere-ocean models, *J. Climate*, 21, 5135–5144, 10.1175/2008JCLI2239.1, 2008.
- Feichter, J. and Lohmann, U.: Can a relaxation technique be used to validate clouds and sulphur species in a GCM?, *Q. J. Roy. Meteor. Soc.*, 125, 1277–1294, 10.1002/qj.1999.49712555609, 1999.
- Ganguly, D., Rasch, P. J., Wang, H., and Yoon, J.-H.: Climate response of the South Asian monsoon system to anthropogenic aerosols, *J. Geophys. Res.*, 117, D13209, 10.1029/2012JD017508, 2012.
- Gates, W. L., Boyle, J., Covey, C., Dease, C., Doutriaux, C., Drach, R., Fiorino, M., Gleckler, P., Hnilo, J., Marlais, S., Phillips, T., Potter, G., Santer, B., Sperber, K., Taylor, K., and Williams, D.: An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I), *B. Am. Meteorol. Soc.*, 73, 1962–1970, 1998.
- Gottelman, A., Morrison, H., and Ghan, S. J.: A new two-moment bulk stratiform cloud microphysics scheme in the Community Atmosphere Model, Version 3 (CAM3). Part II: Single-column and global results, *J. Climate*, 21, 3660–3679, 10.1175/2008JCLI2116.1, 2008.
- Gottelman, A., Liu, X., Ghan, S. J., Morrison, H., Park, S., Conley, A. J., Klein, S. A., Boyle, J., Mitchell, D. L., and Li, J.-F. L.: Global simulations of ice nucleation and ice supersaturation with an improved cloud scheme in the Community Atmosphere Model, *J. Geophys. Res.*, 115, D18216, 10.1029/2009JD013797, 2010.
- Ghan, S., Laulainen, N., Easter, R., Wagener, R., Nemesure, S., Chapman, E., Zhang, Y., and Leung, R.: Evaluation of aerosol direct radiative forcing in MIRAGE, *J. Geophys. Res.*, 106, 5295–5316, 10.1029/2000JD900502, 2001.
- Golaz, J.-C. and Levy, H.: Cloud tuning in a coupled climate model: impact on 20th century warming, *Geophys. Res. Lett.*, 40, 2246–2251, 10.1002/grl.50232, 2013.
- Iacono, M., Delamere, J., Mlawer, E., Shephard, M., Clough, S., and Collins, W.: Radiative forcing by long-lived greenhouse gases: calculations with the AER radiative transfer models, *J. Geophys. Res.*, 113, D13103, 10.1029/2008JD009944, 2008.
- IPCC: Climate Change 2013: the Physical Science Basis. Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, 2013.
- Jeuken, A. B. M., Siegmund, P. C., Heijboer, L. C., Feichter, J., and Bengtsson, L.: On the potential of assimilating meteorological analyses in a global climate model for the purpose of model validation, *J. Geophys. Res.*, 101, 16939–16950, 10.1029/96JD01218, 1996.
- Klocke, D. and Rodwell, M. J.: A comparison of two numerical weather prediction methods for diagnosing fast-physics errors in climate models, *Q. J. Roy. Meteor. Soc.*, 140, 517–524, 10.1002/qj.2172, 2013.
- Kooperman, G. J., Pritchard, M. S., Ghan, S. J., Wang, M., Somerville, R. C. J., and Russell, L. M.: Constraining the influence of natural variability to improve estimates of global aerosol indirect effects in a nudged version of the Community Atmosphere Model 5, *J. Geophys. Res.*, 117, D23204, 10.1029/2012JD018588, 2012.
- Lamarque, J.-F., Bond, T. C., Eyring, V., Granier, C., Heil, A., Klimont, Z., Lee, D., Liousse, C., Mieville, A., Owen, B., Schultz, M. G., Shindell, D., Smith, S. J., Stehfest, E., Van Aardenne, J., Cooper, O. R., Kainuma, M., Mahowald, N., McConnell, J. R., Naik, V., Riahi, K., and van Vuuren, D. P.: Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application, *Atmos. Chem. Phys.*, 10, 7017–7039, 10.5194/acp-10-7017-2010, 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., Bonan, G. B., and Slater, A. G.: Parameterization improvements and functional and structural advances in Version 4 of the Community Land Model, *J. of Adv. Model. Earth Syst.*, 3, M03001, 10.1029/2011MS000045, 2011.
- Lee, L. A., Carslaw, K. S., Pringle, K. J., and Mann, G. W.: Mapping the uncertainty in global CCN using emulation, *Atmos. Chem. Phys.*, 12, 9739–9751, 10.5194/acp-12-9739-2012, 2012.
- Lee, L. A., Pringle, K. J., Reddington, C. L., Mann, G. W., Stier, P., Spracklen, D. V., Pierce, J. R., and Carslaw, K. S.: The magnitude and causes of uncertainty in global model simulations of cloud condensation nuclei, *Atmos. Chem. Phys.*, 13, 8879–8914, 10.5194/acp-13-8879-2013, 2013.
- Lin, S.-J.: A “Vertically Lagrangian” finite-volume dynamical core for global models, *Mon. Weather Rev.*, 132, 2293–2307, 10.1175/1520-0493(2004)132<2293:AVLFDC>2.0.CO;2, 2004.
- Lin, S. J. and Rood, R. B.: Multidimensional flux-form semi-Lagrangian transport schemes, *Mon. Weather Rev.*, 124, 2046–2070, 1996.
- Liu, X., Easter, R. C., Ghan, S. J., Zaveri, R., Rasch, P., Shi, X., Lamarque, J.-F., Gottelman, A., Morrison, H., Vitt, F., Conley, A., Park, S., Neale, R., Hannay, C., Ekman, A. M. L., Hess, P., Mahowald, N., Collins, W., Iacono, M. J., Bretherton, C. S., Flanner, M. G., and Mitchell, D.: Toward a minimal representation of aerosols in climate models: description and evaluation in the Community Atmosphere Model CAM5, *Geosci. Model Dev.*, 5, 709–739, 10.5194/gmd-5-709-2012, 2012.
- Lohmann, U. and Ferrachat, S.: Impact of parametric uncertainties on the present-day climate and on the anthropogenic aerosol effect, *Atmos. Chem. Phys.*, 10, 11373–11383, 10.5194/acp-10-11373-2010, 2010.
- Ma, H.-Y., Xie, S., Boyle, J. S., Klein, S. A., and Zhang, Y.: Metrics and diagnostics for precipitation-related processes in climate model short-range hindcasts, *J. Climate*, 26, 1516–1534, 10.1175/JCLI-D-12-00235.1, 2013.
- Ma, H.-Y., Xie, S., Klein, S., Williams, K., Boyle, J., Bony, S., Douville, H., Fermepin, S., Medeiros, B., Tyteca, S., Watanabe, M., and Williamson, D.: On the correspondence between mean forecast errors and climate errors in CMIP5 models, *J. Climate*, 27, 1781–1798, 10.1175/JCLI-D-13-00474.1, 2014.

- Machenhauer, B. and Kirchner, I.: Diagnosis of systematic initial tendency errors in the ECHAM AGCM using slow normal mode data assimilation of ECMWF reanalysis data, *CLIVAR Exchanges*, 5, 9–10, 2000.
- Martin, G., Milton, S., Senior, C., Brooks, M., Ineson, S., Reichler, T., and Kim, J.: Analysis and reduction of systematic errors through a seamless approach to modeling weather and climate, *J. Climate*, 23, 5933–5957, 10.1175/2010JCLI3541.1, 2010.
- Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., Haak, H., Jungclaus, J., Klocke, D., Matei, D., Mikolajewicz, U., Notz, D., Pincus, R., Schmidt, H., and Tomassini, L.: Tuning the climate of a global model, *J. Adv. Model. Earth Syst.*, 4, M00A01, 10.1029/2012MS000154, 2012.
- McKay, M. D., Beckman, R. J., and Conover, W. J.: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics*, 21, 239–245, 1979.
- Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., and Clough, S. A.: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave, *J. Geophys. Res.*, 102, 16663–16682, 1997.
- Morrison, H. and Gettelman, A.: A new two-moment bulk stratiform cloud microphysics scheme in the NCAR Community Atmosphere Model (CAM3), Part I: Description and numerical tests, *J. Climate*, 21, 3642–3659, 10.1175/2008JCLI2105.1, 2008.
- Murphy, J. M., Sexton, D. M. H., Barnett, D. N., Jones, G. S., Webb, M. J., Collins, M., and Stainforth, D. A.: Quantification of modelling uncertainties in a large ensemble of climate change simulations, *Nature*, 430, 768–772, 2004.
- Neale, R. B., Chen, C. C., Gettelman, A., Lauritzen, P. H., Park, S., Williamson, D. L., Conley, A. J., Garcia, R., Kinnison, D., Lamarque, J. F., Marsh, D., Mills, M., Smith, A. K., Tilmes, S., Vitt, F., Morrison, H., Cameron-Smith, P., Collins, W. D., Iacono, M. J., Easter, R. C., Ghan, S. J., Liu, X. H., Rasch, P. J., and Taylor, M. A.: Description of the NCAR Community Atmosphere Model (CAM5.0), Tech. Rep. NCAR/TN-486-STR, NCAR, available at: <http://www.cesm.ucar.edu/models/cesm1.0/cam/> (last access: 1 April 2014), 2010.
- Neale, R. B., Richter, J. H., and Jochum, M.: The impact of convection on ENSO: from a delayed oscillator to a series of events, *J. Climate*, 21, 5904–5924, 10.1175/2008JCLI2244.1, 2008.
- Park, S. and Bretherton, C. S.: The University of Washington shallow convection and moist turbulence schemes and their impact on climate simulations with the Community Atmosphere Model, *J. Climate*, 22, 3449–3469, 10.1175/2008JCLI2557.1, 2009.
- Park, S., Bretherton, C. S., and Rasch, P. J.: Integrating Cloud Processes in the Community Atmosphere Model, Version 5, 2014.
- Phillips, T. J., Potter, G. L., Williamson, D. L., Cederwall, R. T., Boyle, J. S., Fiorino, M., Hnilo, J. J., Olson, J. G., Xie, S., and Yio, J. J.: Evaluating parameterizations in general circulation models: climate simulation meets weather prediction, *B. Am. Meteorol. Soc.*, 85, 1903–1915, 10.1175/BAMS-85-12-1903, 2004.
- Qian, Y., Yan, H., Zhao, C., Hou, Z., Wang, H., Wang, M., Rasch, P. J., Klein, S., Lucas, D., Johannesson, G., Tannahill, J., and Swiler, L.: Parametric sensitivity analysis of precipitation at global and local scales in the Community Atmosphere Model CAM5, in preparation, 2014.
- Richter, J. H. and Rasch, P. J.: Effects of convective momentum transport on the atmospheric circulation in the Community Atmosphere Model, version 3, *J. Climate*, 21, 1487–1499, 2008.
- Ringer, M. A., McAvaney, B. J., Andronova, N., Buja, L. E., Esch, M., Ingram, W. J., Li, B., Quaas, J., Roeckner, E., Senior, C. A., Soden, B. J., Volodin, E. M., Webb, M. J., and Williams, K. D.: Global mean cloud feedbacks in idealized climate change experiments, *Geophys. Res. Lett.*, 33, L07718, 10.1029/2005GL025370, 2006.
- Rodwell, M. J. and Palmer, T. N.: Using numerical weather prediction to assess climate models, *Q. J. Roy. Meteor. Soc.*, 133, 129–146, 10.1002/qj.23, 2007.
- Soden, B. and Held, I.: An assessment of climate feedbacks in coupled ocean–atmosphere models, *J. Climate*, 19, 3354–3360, 10.1175/JCLI3799.1, 2006.
- Song, F., Zhou, T., and Qian, Y.: Responses of East Asian summer monsoon to natural and anthropogenic forcings in the 17 latest CMIP5 models, *Geophys. Res. Lett.*, 41, 10.1002/2013GL058705, 2014.
- Stainforth, D. A., Aina, T., Christensen, C., Collins, M., Faull, N., Frame, D. J., Kettleborough, J. A., Knight, S., Martin, A., Murphy, J. M., Piani, C., Sexton, D., Smith, L. A., Spicer, R. A., Thorpe, A. J., and Allen, M. R.: Uncertainty in predictions of the climate response to rising levels of greenhouse gases, *Nature*, 433, 403–406, 2005.
- Taylor, K., Stouffer, R., and Meehl, G.: An overview of CMIP5 and the experiment design, *B. Am. Meteorol. Soc.*, 93, 485–498, 10.1175/BAMS-D-11-00094.1, 2012.
- Vinoj, V., Rasch, P. J., Wang, H., Yoon, J.-H., Ma, P.-L., Landu, K., Singh, B.: Short-term modulation of Indian summer monsoon rainfall by West Asian dust, *Nature Geosci.*, 7, 308–313, 10.1038/ngeo2107, 2014.
- Wehner, M. F., Reed, K. A., Li, F., Prabhat, Bacmeister, J., Chen, C.-T., Paciorek, C., Gleckler, P. J., Sperber, K. R., Collins, W. D., Gettelman, A., Jablonowski, C., and Algieri, C.: The effect of horizontal resolution on simulation quality in the Community Atmospheric Model, CAM5.1, *J. Adv. Model. Earth Syst.*, in revision, 2014.
- Williams, K., Bodas-Salcedo, A., Déqué, M., Fermepin, S., Medeiros, B., Watanabe, M., Jakob, C., Klein, S., Senior, C., and Williamson, D. L.: The Transpose-AMIP II experiment and its application to the understanding of Southern Ocean cloud biases in climate models, *J. Climate*, 26, 3258–3274, 10.1175/JCLI-D-12-00429.1, 2013.
- Williamson, D. L., Boyle, J., Cederwall, R., Fiofrino, M., Hnilo, J. J., Olson, J. G., Phillips, T., Potter, G. L., and Xie, S. C.: Moisture and temperature balances at the atmospheric radiation measurement Southern Great Plains site in forecasts with the Community Atmosphere Model (CAM2), *J. Geophys. Res.*, 110, D15S16, 10.1029/2004JD005109, 2005.
- Xie, S., Ma, H.-Y., Boyle, J. S., Klein, S. A., and Zhang, Y.: On the correspondence between short- and long-time-scale systematic errors in CAM4/CAM5 for the Year of Tropical Convection, *J. Climate*, 25, 7937–7955, 10.1175/JCLI-D-12-00134.1, 2012.
- Yang, B., Qian, Y., Lin, G., Leung, R., and Zhang, Y.: Some issues in uncertainty quantification and parameter tuning: a case study of convective parameterization scheme in the WRF regional climate model, *Atmos. Chem. Phys.*, 12, 2409–2427, 10.5194/acp-

12-2409-2012, 2012.

1150 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 im-
proved convection on the global circulation and climate, *J. Geo-
phys. Res.*, 118, 395–415, 10.1029/2012JD018213, 2013.

1155 Zhang, G. J. and McFarlane, N. A.: Sensitivity of climate simula-
tions to the parameterization of cumulus convection in the Cana-
dian Climate Centre general circulation model, *Atmos. Ocean*,
33, 407–446, 1995.

1160 Zhao, C., Liu, X., Qian, Y., Yoon, J., Hou, Z., Lin, G., McFar-
lane, S., Wang, H., Yang, B., Ma, P.-L., Yan, H., and Bao, J.:
A sensitivity study of radiative fluxes at the top of atmosphere
to cloud-microphysics and aerosol parameters in the commu-
nity atmosphere model CAM5, *Atmos. Chem. Phys.*, 13, 10969–
10987, 10.5194/acp-13-10969-2013, 2013.

1165 Zhou, T.-J. and Li, Z.-X.: Simulation of the east asian summer mon-
soon using a variable resolution atmospheric GCM, *Clim. Dy-
nam.*, 19, 167–180, 10.1007/s00382-001-0214-8, 2002.