



LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

A New Ensemble of Perturbed-Input-Parameter Simulations by the Community Atmosphere Model

C. Covey, S. Brandon, P. T. Bremer, D. Domyancis, X.
Garaizar, G. Johannesson, R. Klein, S. A. Klein, D. D.
Lucas, J. Tannahill, Y. Zhang

October 28, 2011

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

**A New Ensemble of Perturbed-Input-Parameter Simulations by the
Community Atmosphere Model**

Curt Covey, Scott Brandon, Peer-Timo Bremer, David Domyancic, Xabier Garaizar,
Gardar Johannesson, Richard Klein, Stephen A. Klein, Donald D. Lucas, John
Tannahill, and Yuying Zhang

Lawrence Livermore National Laboratory, Livermore, CA

27 October 2011

LLNL-TR-509454

CORRESPONDING AUTHOR:

Curt Covey, LLNL Mail Code L-103, 7000 East Avenue, Livermore, CA 94550

E-mail: covey1@llnl.gov

ABSTRACT

Uncertainty quantification (UQ) is a fundamental challenge in the numerical simulation of Earth's weather and climate, and other complex systems. It entails much more than attaching defensible error bars to predictions: in particular it includes assessing low-probability but high-consequence events. To achieve these goals with models containing a large number of uncertain input parameters, structural uncertainties, etc., raw computational power is needed. An automated, self-adapting search of the possible model configurations is also useful.

Our UQ initiative at the Lawrence Livermore National Laboratory has produced the most extensive set to date of simulations from the US Community Atmosphere Model. We are examining output from about 3,000 twelve-year climate simulations generated with a specialized UQ software framework, and assessing the model's accuracy as a function of 21 to 28 uncertain input parameter values. Most of the input parameters we vary are related to the boundary layer, clouds, and other sub-grid scale processes. Our simulations prescribe surface boundary conditions (sea surface temperatures and sea ice amounts) to match recent observations.

Fully searching this 21+ dimensional space is impossible, but sensitivity and ranking algorithms can identify input parameters having relatively little effect on a variety of output fields, either individually or in nonlinear combination. Bayesian statistical constraints, employing a variety of climate observations as metrics, also seem promising. Observational constraints will be important in the next step of our

project, which will compute sea surface temperatures and sea ice interactively, and will study climate change due to increasing atmospheric carbon dioxide.

Modern climate models produce detailed three-dimensional simulations of temperature, wind, humidity, and other meteorological fields in a large domain extending month-by-month over decades to centuries of time. Errors are inevitable given imperfectly known initial conditions, boundary conditions, and the large number of assumptions that must be made about poorly understood small-scale phenomena, like cloud formation and turbulence. Indeed, numerical simulations of many complex systems, from macromolecules to galaxies, are subject to the old saying that “all models are wrong but some models are useful.” Since the first three-dimensional climate models were run on 1970s supercomputers, attempts have been made to attach defensible probabilities to their predictions. This article presents initial climate-related results from an uncertainty quantification (UQ) initiative at the Lawrence Livermore National Laboratory (LLNL). The initiative follows a broad multidisciplinary approach to advancing UQ in climate science and other endeavors of interest to LLNL (e.g., the National Ignition Facility).

A full probabilistic assessment of climate prediction uncertainty must include not only outcomes that are most likely, but also a quantification of low-probability but high consequence events. Even a 1% chance of catastrophic flooding, for example, might be significant in designing a dam. Therefore much climate UQ in recent years has involved probability density functions (PDFs). The canonical PDF describes the idealized but well-defined quantity $\Delta T_{2\times CO_2}$, the equilibrium global mean surface warming due to doubled atmospheric carbon dioxide. Over the past decade many research groups, using different climate models and different methods of analyzing climate model output, have published $\Delta T_{2\times CO_2}$ PDFs. Traditional

climate model evaluation uses multi-model ensembles (MMEs) in which the output of different climate models, developed somewhat independently, is inter-compared (Randall et al. 2007). $\Delta T_{2\times CO_2}$ PDFs can be derived from MMEs, but small sample sizes limit this approach (Raisanen 2005). Beginning about a decade ago, two groups in the United Kingdom have followed a different approach.

The UK groups constructed perturbed parameter ensembles (PPEs) from variants of the UK Meteorological Office Hadley Center climate model HadCM3. A PPE may be constructed from a single climate model by varying the values of selected internal input parameters within “reasonable” limits. Within the UK, the QUMP project (Quantifying Uncertainties in Model Prediction; Murphy et al. 2004) employs Met Office supercomputer facilities, while the CPDN project (climateprediction.net; Piani et al. 2005, Stainforth et al. 2005) employs worldwide “cloud computing” as described below. Collins et al. (2010) update and summarize the UK PPE results. In the United States, Jackson et al. (2008) and Sanderson (2011) have applied the PPE technique to the Community Atmosphere Model (CAM). Other groups are now generating PPEs from other models (Neelin et al. 2010, Yokohata et al. 2010, Klocke et al. 2011).

A striking implication has emerged from the work to date. Although PDFs of $\Delta T_{2\times CO_2}$ peak in the range $\sim 1.5 - 4.5$ K (consistent with MMEs dating back to the two-member “ensemble” assessed by Charney 1979), they are often noticeably skewed toward higher climate sensitivity. This result implies that $\Delta T_{2\times CO_2}$ values ≥ 5 K for the real Earth are plausible, albeit less likely than the canonical 1.5 – 4.5 K range (see Box 10.2 of Meehl et al. 2007a). Very high climate sensitivity implies

potentially catastrophic effects from even modest carbon dioxide emissions. Most PPE simulations to date, however, come from a single source: HadCM. Sanderson (2011) obtained very different PDFs from his limited number of CAM simulations, despite his attempt to construct a PPE with similar uncertain input parameters to the previous HadCM work. Further, the PDFs shown in Box 10.2 of Meehl et al. (2007a)—which are constructed from both PPEs and a variety of other methods—exhibit substantial variation not only in their “tails,” but also in their means and dispersions. These results provide motivation for continued PPE construction and analysis using CAM.

METHODS. The LLNL UQ initiative constructs PPEs from versions of CAM algorithmically close to those employed by Jackson et al. (2008) and Sanderson (2011), but with a greatly expanded list of variable input parameters (see Table 1). As described below, the CAM ensembles are generated using LLNL’s UQ Pipeline software framework. We aim to assess $\Delta T_{2\times CO_2}$ as well as more directly useful predictions of climate change (e.g., regional as opposed to global-mean changes). First however, we created and are now analyzing a large database of simulations in which sea surface temperatures (SSTs) and sea ice amounts are prescribed to match observations over recent years (1993-2004; the first two years of output are omitted from our analysis to avoid spin-up artifacts). By construction, such “AMIP” simulations exclude substantial changes in the global climate, but they provide a useful assessment of climate model behavior (Gates et al. 1999). In the context of climate UQ, comparison of AMIP simulations with real-world observations provides

an important measure of model accuracy as a function of the assumptions embodied in the model.

Expanding the number of variable input parameters entails different scientific and computational strategies. Different approaches are needed because a naïve search of the space of possible model configurations grows exponentially with the number of parameters that are varied. Studies that vary only a few parameters can be undertaken with methods that would not work for studies that vary more parameters. Sanderson (2011) varied four of CAM's input parameters among high, medium, and low discrete values (with medium values chosen to match the model's default input parameter settings) and ran the model for every one of the resulting $3^4 = 81$ possible combinations. Jackson et al. (2008) varied six of CAM's input parameters and ran the model 518 times to study how climate model development efforts might calibrate input parameters to arrive at a simulation in best agreement with observations. For the studies discussed below, we varied 21 – 28 CAM input parameters listed in Table 1.

These are only a fraction of all of CAM's uncertain input parameters. As in previous work, we selected the parameters to vary and chose their limits of variation in consultation with the respective climate process experts. We also ensured that similar CAM input parameters considered by Jackson et al. (2008) and Sanderson (2011) were included in our studies. (A precise superset could not be constructed because we used newer and slightly different versions of CAM.) As in previous work, our list is focused on subgrid-scale parameterizations of clouds and precipitation, convection, and the turbulent near-surface planetary boundary layer

(PBL)—processes long recognized as the most challenging for weather and climate models to represent. Table 1 notes that the first 25 of our 28 chosen input parameters appear in CAM’s modules parameterizing cloud fraction, cloud optics, cloud particle sedimentation, cloud water in its different phases, the PBL and shallow convection (Hack parameterization), and deep convection (Zhang-McFarlane parameterization). See Collins et al. (2006) and Gent et al. (2009) for summaries of these parameterizations. The remaining three parameters involve the turbulent transfer of water mass, energy, and momentum between the surface and the atmosphere.

As noted in Table 2, we used CAM Version 3.6 in our initial studies, later switching to the publicly released Version 4.0 when it became available. Jackson et al. (2008) used Version 3.1 and Sanderson (2011) used Version 3.5. All of our simulations used the finite-volume dynamical core with a horizontal resolution of $1.9^\circ \times 2.5^\circ$ latitude \times longitude resolution, and 26 vertical levels. This formulation is identical to Sanderson’s and similar to Jackson et al.’s spectral T42 resolution ($\sim 2.8^\circ \times 2.8^\circ$ latitude \times longitude). Little difference exists in subgrid-scale parameterizations between Versions 3 – 4 of CAM compared with the difference between Versions 3 – 4 and Version 5 (Neale et al. 2010). We ran each AMIP simulation for 12 years and analyzed output from the final 10 years.

For most of our CAM runs, input parameter values were selected with equal probability from the ranges defined by the high and low values shown in Table 1. In the context of Bayesian statistics, our uncertain input parameters were described using flat or non-informative priors. We explored input parameter space by the

random-walk MOAT method (Morris 1991) and by Latin hypercube selection (e.g., Section 7.8.1 of Press et al. 2007). For initial benchmark simulations, we also varied each parameter individually with all other parameters set to default values (one-at-a-time or OAT method). In the MOAT method, the input parameter space is partitioned into a lattice, a random point on the lattice is chosen for an initial simulation, and then additional simulations are performed by taking large steps through the lattice, one direction at a time. The result is a “MOAT path,” a random walk in the input parameter space. The chief advantage of the MOAT method is that its computational cost scales linearly with the number of uncertain input parameters. Ten to twenty different MOAT paths are typically sufficient to characterize sensitivities over the global parameter space (i.e., 220-440 simulations for 21 uncertain parameters).

As shown in Table 2, we performed a total of 2,894 AMIP runs for a cumulative period of 34,728 simulated years. We also ran 88 simulations with different initial conditions as well as different input parameter values, but a first look at these suggests that varying the internal input parameters has far more influence on climate, as expected (see Fig. 4 below). Archiving the standard monthly-mean output from these runs produced a database exceeding 42 Terabytes in volume. It is the most extensive set of AMIP PPE output from CAM. Even so, our simulations cover an extremely small fraction of the enormous uncertainty space implied by Table 1. To illustrate, if we were to divide each of our parameter ranges into three bins (e.g., low, medium and high), more than ten billion simulations (3^{21}) would be needed to evaluate all of the possible combinations for 21 parameters.

This number grows exponentially with the number of dimensions, and of course more than three values of each parameter are needed to fully characterize the model's response. Potentially an infinite number of choices could be made for each input parameter. These mathematical facts of life—sometimes called “the curse of dimensionality”—make both raw computational power and sophisticated computational methods mandatory in climate UQ.

To achieve the requisite computational power, the CPDN project has employed a volunteer network of desktop computers located around the world: “cloud computing.” In contrast, our project currently makes semi-dedicated use of LLNL's Atlas machine, a high-performance AMD Opteron cluster with about 8,000 computing cores. Such machines are ideal for running an ensemble of many simulations in parallel, including those that require inter-core communication (e.g., runs at high resolution) and those that do not (e.g., Monte Carlo). We completed ~ 35,000 simulated years in a single calendar year.

Although enumeration of our computational methods (either present or planned) is beyond the scope of this article, we briefly note our use of the LLNL UQ Pipeline. The UQ Pipeline is an end-to-end workflow software system providing an interface to multi-dimension, multi-physics simulation programs like CAM. It stages, executes, and potentially monitors concurrent ensemble simulations, permitting “adaptive sampling refinement” (ASR) in which automated output analysis can guide the varying of input parameters. It includes algorithms to sample high-dimensional input parameter spaces, to construct surrogate models (see below), to generate statistical inferences from ensemble output, and to incorporate

observational data in order to constrain input parameters and determine their PDFs (see Fig. 7 below). In addition to enabling the production and analysis of climate simulations discussed in this paper, the UQ Pipeline has been applied to the design of LLNL's laser-fusion project, the National Ignition Facility (Spears et al. 2009, 2010).

RESULTS. Figure 1 maps the Western Hemisphere surface air temperature response in CAM3 to changes in an input parameter varied within a multi-dimensional space as part of our MOAT study: the timescale τ for the consumption of convective available potential energy (CAPE) in the Zhang-McFarlane deep convection module zm_conv . This input parameter strongly influences a variety of output fields (see Fig. 3 below) and exhibits both expected and surprising features. Figure 1 shows that surface air temperature responds least over ocean areas, where AMIP boundary conditions fix SST, and most over the Amazon basin, the land region where deep convection is most prominent in the hemisphere. (Land-use distributions are fixed in this version of CAM, so effects of deforestation are not considered our simulations.) Surprisingly, Central America exhibits less sensitivity to τ than the “Southwestern monsoon” region of the USA, and the Central American response is widely spread over adjacent downwind ocean areas.

An important near-term goal of our work is to assess the relative sensitivity of CAM output to different input parameters. If some are much less influential than others on the output, they could be omitted from future work. Even slightly reducing the number of input parameters would greatly reduce the computational effort and

complexity of a PPE analysis (due to the “curse of dimensionality” discussed above). Figures 2 and 3 show how the MOAT method can be used for this purpose. In Fig. 2, the effects upon global- and annual-mean Outgoing Longwave Radiation (OLR) of varying the 21 input parameters in our CAM3 simulations are represented by points on a two-dimensional plot. The horizontal coordinate is the mean magnitude of OLR changes per fractional change of the input parameters, averaged over 20 MOAT paths. The vertical coordinate is the corresponding standard deviation. Region (i) of the plot identifies input parameters that generate relatively small mean-magnitude changes ($< 0.4 \text{ W m}^{-2}$). These input parameters are relatively unimportant for determining global- and annual-mean OLR. Regions (ii) and (iii) identify input parameters that generate larger mean-magnitude changes. Input parameters in Region (iii) also generate larger standard deviations. Thus, the effect of each Region (iii) parameter on OLR depends strongly on the location in parameter space. These input parameters exhibit relatively important nonlinear interactions either with themselves or with other input parameters.

Ranking the parameter sensitivities by the mean magnitude changes (i.e., the horizontal axis in Fig. 2) indicates the most-to-least important parameters for a given output. By performing the ranking for many outputs of interest, we can use MOAT to start to identify important sets parameters throughout CAM. Figure 3 shows a MOAT ranking analysis of the 21 input parameters varied in CAM3. For this analysis, we create sensitivity maps of 10 different output fields (e.g., as in Fig. 1), perform area-weighted averaging of the maps, and rank the resulting sensitivities. The output fields include near-surface air temperature TREFHT, precipitation rate

PRECT, and several types of energy flux at the surface and the top of the atmosphere. For example, FLUT is the **F**lux of **L**ongwave **U**pward radiation at the **T**op, also called outgoing longwave radiation or OLR elsewhere in this paper. The output shown in Fig. 3 also includes both Longwave and Shortwave Cloud Forcing at the top of the atmosphere: LWCF and SWCF respectively. We have extended the ranking analysis to other fields (see Lucas et al. 2011), including most of those proposed by Gleckler et al. (2008, Table 2) for objective metrics of climate model performance, and those used by Jackson et al. (2008) as observational constraints on their CAM PPE.

All output fields in Fig. 3 are relatively sensitive to deep convection parameter *zmconv_tau*, cloud fraction parameter *cldfrc_rhminl*, and cloud water parameter *cldwat_icritc*. Other input parameters from the cloud-water module exert relatively weak influence on all output fields, as does cloud optics parameter *cldopt_rliqice*. It is also noteworthy that different output fields are most strongly affected by different sets of input parameters. Thus, before dropping any parameters from further study, we are exploring other aspects of the simulated present-day climate (e.g., clouds) and also considering global climate changes that AMIP boundary conditions may suppress.

As noted above, we have devoted about 100 model runs to assessing the effect of different initial conditions on model output. We varied initial conditions in the traditional way, choosing different initial conditions from a present-day climate simulation. Our initial-condition experiments included 16 different sets of the CAM3 input parameters from previous ensemble simulations. Figure 4 compares the effect

of varying initial conditions (colored lines) with the effect of varying input parameters (gray lines). For the two global- and annual-mean output variables shown, initial-condition effects are about 10% of input parameter effects. For surface air temperature the range produced by varying input parameters is about 2 K (rather surprising since AMIP boundary conditions prescribe SST) but the range produced by varying initial conditions is about 0.1 K. Note, however, that the ensemble spread displayed in Fig. 4 does not consider real-world observations. By calculating the likelihood that the ensemble members agree with observations, the output spread would be appreciably narrowed.

One particularly important climate observable is planetary albedo α , the ratio at the top of the atmosphere of global-mean upward solar energy flux to global-mean downward solar energy flux. Planetary albedo is a fundamental ingredient of energy balance at the top of the atmosphere. The equation asserting exact balance is $(S/4)(1-\alpha) = F$, where F = global-mean OLR, and S = the so-called solar constant (so that $\pi a^2 S / 4\pi a^2 = S/4$ = global-mean downward solar energy flux). Figure 5 plots the residual global mean net downward energy flux at the top of the model (i.e., the difference between the left- and right-hand sides of the balance equation) as a function of α for 895 CAM4 simulations in which 27 input parameters are varied. The actual residual is nonzero because the climate system is not in perfect equilibrium. Observations (mainly of ocean heat content) imply that present-day incoming energy flux at the top of the atmosphere exceeds outgoing by $\sim 1 \text{ W m}^{-2}$ or less (Hansen et al. 2005, 2011; Lyman et al. 2010; Trenberth and Fasullo 2010). AMIP simulations in general give somewhat larger residuals than

seen in the real world because SSTs are not allowed to adjust, even by slight amounts (consistent with observational uncertainty) that could adjust top-of-atmosphere fluxes by $\sim 1 \text{ W m}^{-2}$. For example, CAM3 and CAM4 with default input parameter values give net downward energy fluxes of about $+4$ and $+2 \text{ W m}^{-2}$ respectively.

It is clear from Fig. 5, however, that most of our PPE egregiously exceeds observed limits on top-of-atmosphere fluxes. More than two-thirds of the simulations fall outside a generous 95% confidence range from modern satellite observations (Loeb et al. 2009, Trenberth et al. 2009) despite expert opinion providing the range of variation for each input parameter. Simulations that fall far outside observed limits receive very low likelihood scores in our statistical analysis.

Most of our PPE members that fail a top-of-atmosphere flux test do so with excessively large α . This corresponds to excessive reflected solar energy and a net *upward* energy flux at the top of the atmosphere $> 10 \text{ W m}^{-2}$ for most of the PPE. The points in Fig. 5 fall close to a straight line (correlation $\equiv r = -0.96$ with slope $m = -280 \text{ W m}^{-2}$; repeating the exercise for 757 later CAM4 simulations varying 28 input parameters gives $r = -0.95$ and $m = -290 \text{ W m}^{-2}$) with slope magnitude less than the value $S/4 = 340 \text{ W m}^{-2}$ that would occur if F were constant. In our results F is negatively, albeit rather weakly, correlated with α ($r = -0.57$ and -0.48 in the 27- and 28-parameter studies respectively). Apparently more cloudiness in our PPE entails not only a higher planetary albedo, but also more high altitude clouds, which produce OLR at cooler temperatures. This counteracts to some extent the direct effect of α upon net energy input to the climate system. Nevertheless, it is clear

from Fig. 5 that α is the primary controlling factor in our AMIP simulations. This result confirms a longstanding principle of atmospheric model development that reserves planetary albedo for “final tuning” of the model’s energy budget (e.g., Section 2.5.1 of CCSP 2008).

Preliminary work suggests that filtering by the single global-mean constraint $0.275 < \alpha < 0.320$, as suggested by Fig. 5, improves the simulation of energy flow through the climate system not only globally but also as a function of latitude and season (not shown). There is no guarantee, however, that it will select results that prove satisfactory in aspects other than energy flow through the climate system. Therefore we are applying additional constraints based on a subset of the observables listed in Table 2 of Gleckler et al. (2008). We are also quantifying the input parameter values responsible for the most satisfactory simulations. For each of our fields, we crudely estimate observational uncertainty by taking the difference between two datasets, e.g., ERBE vs. CERES for top-of-atmosphere energy fluxes and ERA-40 vs. ERA-Interim for most other fields. To efficiently search the large input parameter space, we construct surrogate models (a.k.a. response surfaces or statistical emulators) in which the output of CAM as a function of input parameters is approximated at drastically reduced computational cost. We have tested several different methods for surrogate model construction and here show preliminary results from two of them: Gaussian process (Sacks et al. 1989, Levy and Steinberg 2010) and multivariate adaptive regression splines (or MARS; see Friedman 1991).

Figure 6 is a square matrix of plots illustrating some results of our Bayesian statistical analysis. We constructed Gaussian process surrogate models for the

global means of five quantities discussed above—OLR (a.k.a. FLUT), FSUTOA [$\equiv (S/4)(\alpha)$], LWCF, SWCF, and PRECT—together with humidity and temperature at the 850 hPa level, and geopotential height at 500 hPa. For each of these eight quantities, we considered two seasonal means (December-January-February and June-July-August) as well as the annual means, for a total of $3 \times 8 = 24$ response surfaces. We applied observational constraints to a subset of five quantities (OLR, LWCF, SWCF, PRECT, 500-hPa geopotential height) for a total of $3 \times 5 = 15$ observational constraints. Figure 6 shows prior and posterior distributions of seven of the 24 outputs. The posterior input-parameter distributions (not shown) are essentially the likelihood that the parameter sets produce agreement with observations.

Diagonal elements in Fig. 6 give the marginal distributions for each of the selected output variables. Off-diagonal elements are density plots giving the joint distributions for each pair of the output variables. In the off-diagonal plots that pair different seasonal averages of the same variable, the points fall very close to straight lines, indicating a strong correlation between the different seasons. The different quantities LWCF and OLR also exhibit a strong correlation. These results are of course expected, and they reduce the effective number of independent observations. Principal component analysis of all observed fields employed in this exercise finds that the first four components explain 99.9% of the variance. In order to more effectively constrain the input-parameter space, future work will involve additional metrics. These will include spatial as well as seasonal variations. Nevertheless, Fig. 6 indicates that Gaussian process surrogate models can be used to find input

parameters that produce model output in simultaneous agreement with a variety of observations, given reasonable observational error bars.

Figure 7 presents a sample result from the alternate MARS technique of surrogate modeling, but using the same observational constraints and model output. The left side of the figure shows the resulting prior and posterior distributions of one input parameter, *cltwat_icritic*. This parameter is among the most influential on model output fields shown in Fig. 3. Its prior in Fig. 7 is a uniform distribution across the full range of variation, corresponding to our initial assumption that all values of uncertain input parameters are equally likely within the ranges specified by expert opinion. The uniform prior distribution gives way to a broad posterior with a maximum somewhat less than the model's default value, in agreement with analogous results from the Gaussian process technique (not shown). The lack of a sharp maximum hints at the difficulty of tightly constraining input parameters using our trial set of observational constraints, a difficulty evident in prior work with PPEs from both CAM (Jackson et al. 2008) and other models (e.g., Neelin et al. 2010). This problem, however, may be lessened by omitting parameters with little influence on model output and incorporating additional observational constraints.

The right side of Fig. 7 shows corresponding results for the output variable FSUTOA, the global mean upward solar energy flux at the top of the atmosphere. In agreement with Fig. 5, the prior distribution indicates that too much solar energy is reflected back to space in most of our PPE simulations. The posterior distribution shows that this problem is largely corrected by MARS filtering—even though

FSUTOA is not directly used as an observational constraint. The constraint occurs through SWCF, which is highly correlated with FSUTOA.

CONCLUSION. The first three-dimensional climate models were run on 1970s “supercomputers.” Due to their limited power, a typical study focused on a single model with one set of initial conditions and with internal input parameters assuming one set of values chosen to optimize output in some way. Comparing different simulations from different research groups led to multi-model ensembles, beginning with the 1979 Charney report on potential global warming due to increasing atmospheric carbon dioxide. The most prominent MMEs today are the CMIP3 / IPCC AR4 database of climate model output (Meehl et al. 2007b) and its nascent successor, the CMIP5 / IPCC AR5 database of climate model output (see <http://cmip-pcmdi.llnl.gov/cmip5>).

Increasingly more powerful computers allowed climate model simulations to be run as an ensemble with different initial conditions (a procedure that has now become routine, e.g., Huebener et al. 2007) and later with many different settings of internal input parameter values (Murphy et al. 2004; Piani et al. 2005). The latter method, perturbed-parameter analysis, is complementary to multi-model analysis. PPE creation may be regarded as “de-tuning” a model from its designed configuration, but it is necessary in order to fully assess the effects of uncertain assumptions on model predictions.

The first step of our climate UQ project at the Lawrence Livermore National Laboratory has built on prior work (Jackson et al. 2008; Sanderson 2011) to create

the most extensive PPE of AMIP simulations from the Community Atmosphere Model. Archived storage volume exceeds that of the current CMIP3 / IPCC AR4 database, which has been used by thousands of researchers around the world to produce hundreds of scientific papers (Meehl et al. 2007b). To benefit climate science research, we intend to make our PPE output available after completing further analyses. We believe that in addition to traditional climate model evaluation, other methods from mathematics, statistics, and machine learning (e.g. Gerber et al. 2010) are required to comprehensively survey our PPE output.

Our initial work focused on AMIP simulations, which prescribe surface boundary conditions to match present-day observations. We are now entraining both simple (“slab”) and complex (three-dimensional) representations of the ocean together with interactive sea ice. The resulting coupled ocean-atmosphere models can simulate climate changes such as global warming. We believe our work and its archived output will provide useful information on model behavior and its implications for real-world climate change.

ACKNOWLEDGMENTS. We thank David C. Bader, Peter J. Gleckler, Karl E. Taylor, and the Community Earth System Model consortium (especially its Atmosphere Model Working Group) for consultation and advice; and the Livermore Computing center for generous allocation of supercomputer resources. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344 and was

funded by the Uncertainty Quantification Strategic Initiative Laboratory Directed
Research and Development Project at LLNL under project tracking code 10-SI-013.

REFERENCES

- CCSP, 2008: *Climate Models: An Assessment of Strengths and Limitations*. A Report by the U.S. Climate Change Science Program and the Subcommittee on Global Change Research [D. C. Bader, C. Covey, W. J. Gutowski Jr., I .M. Held, K. E. Kunkel, R. L. Miller, R. T. Tokmakian and M. H. Zhang (Authors)]. Department of Energy, Office of Biological and Environmental Research, Washington, D.C., USA, 124 pp. [Available online at <http://www.globalchange.gov/publications/reports/scientific-assessments/saps/sap3-1>.]
- Collins, M., and Coauthors, 2010: Climate model errors, feedbacks and forcings: A comparison of perturbed physics and multi-model ensembles, *Climate Dyn.*, **36**, 1737-1766.
- Friedman, J. H., 1991: Multivariate adaptive regression splines (MARS). *Ann. Statistics*, **19**, 1-141.
- Gates, L., and Coauthors, 1999: An overview of the Atmospheric Model Intercomparison Project, *Bull. Amer. Meteor. Soc.*, **80**, 29-55.
- Gerber, S., P.-T. Bremer, V. Pascucci, and R. Whitaker, 2010: Visual exploration of high dimensional scalar functions, *IEEE Transactions on Visualization and Computer Graphics*, **16**, 1271-1280.
- Gleckler, P. J., K. E. Taylor, and C. Doutriaux, 2008: Performance metrics for climate models, *J. Geophysical Res.*, **113**, D06104, doi:10.1029/2007JD008972.
- Hansen, J., and Coauthors, 2005: Earth's energy imbalance: Confirmation and implications, *Science*, **308**, 1431-1435.

- Hansen, J., M. Sato, P. Kharecha, and K. von Schuckmann, 2011: Earth's energy balance and implications, arXiv ePrint 1105:1140. [Available online at <http://arxiv.org/abs/1105.1140>.]
- Huebener, H., and Coauthors, 2007: Ensemble climate simulations using a fully coupled ocean–troposphere–stratosphere general circulation model, *Phil. Trans. R. Soc. A*, **365**, 2089-2101.
- Jackson, C. S., M. K. Sen, G. Huerta, Y. Deng, and K. P. Bowman, Error reduction and convergence in climate prediction, *J. Climate*, **21**, 6698-6709.
- Klocke, D., R. Pincus, and J. Quaas, 2011: On constraining estimates of climate sensitivity with present-day observations through model weighting. *Climate Dynamics*, in revision.
- Loeb, N. G., and Coauthors, 2009: Toward optimal closure of the earth's top-of-atmosphere radiation budget, *J. Climate*, **22**, 748-766.
- Levy, S., and D. M. Steinberg, 2010: Computer experiments: A review. *Adv. Stat. Anal.*, **94**, 311-324.
- Lucas, D. D., and Coauthors, 2011: Scientific aspects of the climate UQ project at LLNL. Presented at the Community Earth System Atmosphere Model Working Group, Boulder, CO, 10 – 12 February. [Available online at http://www.cesm.ucar.edu/working_groups/Atmosphere/Presentations/2011/lucas_amwg11.pdf.]
- Meehl, G. A., and Coauthors, 2007a: Global climate projections. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds., Cambridge University Press, 590-662.

- Meehl, G. A., and Coauthors, 2007b: The WCRP CMIP3 Multimodel dataset: A new era in climate change research, *Bull. Amer. Meteor. Soc.*, **88**, 1383-1394.
- Morris, M. D., 1991: Factorial sampling plans for preliminary computational experiments, *Technometrics*, **33**, 161-174.
- Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth, 2004: Quantification of modeling uncertainties in a large ensemble of climate change simulations, *Nature*, **430**, 768-772.
- Neelin J. D., A. Bracco, H. Luo, J. C. McWilliams, and J. E. Meyerson, 2010: Considerations for parameter optimization and sensitivity in climate models, *PNAS*, **107**, 21349-21354.
- Piani, C., D. J. Frame, D. A. Stainforth, and M. R. Allen, 2005: Constraints on climate change from a multi-thousand member ensemble of simulations, *Geophys. Res. Lett.*, **32**, L23825.
- Raisanen, J., 2005: Probability distributions of CO2-induced global warming as inferred directly from multimodel ensemble simulations, *Geophysica*, **41**, 19-30.
- Randall, D. A., and Coauthors, 2007: Climate models and their evaluation. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds., Cambridge University Press, 590-662.
- Sacks, J., W. J. Welch, T. J. Mitchell, H. P. Wynn, 1989: Design and analysis of computer experiments. *Stat. Sci.*, **4**, 409-435.
- Sanderson, B. 2011: A multi-model study of parametric uncertainty in predictions of climate response to rising greenhouse gas concentrations, *J. Climate*, **24**, 1362-1377.

- Spears, B., S. Brandon, D. Clark, C. Cerjan, J. Edwards, O. Landen, J. Lindl, S. Haan, S. Hatchett, J. Salmonson, P. Springer, S. Weber, and D. Wilson. Measurements of low-deuterium surrogates for the prediction of ignition implosion performance. *Bulletin of the American Physical Society: 51st Annual Meeting of the APS Division of Plasma Physics*, 54(15), 2009.
- Spears, B. K., S. Brandon, D. Clark, C. Cerjan, J. Edwards, O. Landen, J. Lindl, S. Haan, S. Hatchett, J. Salmonson, P. Springer, S. V. Weber, and D. Wilson. Prediction of Ignition Implosion Performance Using Measurements of Low-Deuterium Surrogates. In *Sixth International Conference on Inertial Fusion Sciences and Applications, Parts 1-4*, volume 244 of *Journal of Physics Conference Series*, Dirac House, Temple Back, Bristol BS1 6BE, England, 2010. IOP Publishing Ltd.
- Stainforth, D. A., and Coauthors, 2005: Uncertainty in predictions of the climate response to rising levels of greenhouse gases, *Nature*, **433**, 403-406.
- Trenberth, K. E., Fasullo, J. T., and Kiehl, J., 2009: Earth's global energy budget, *Bull. Amer. Meteor. Soc.*, **90**, 311-323.
- Trenberth, K. E., and J. T. Fasullo, Tracking Earth's energy, *Science*, **238**, 316-317.
- Yokohata, T., M. J. Webb, M. Collins, K. D. Williams, M. Yoshimori, J. C. Hargreaves, and J. D. Annan, 2010: Structural similarities and differences in climate responses to CO₂ increase between two perturbed physics ensembles. *J. Climate*, **23**, 1392-1410.

FIGURE CAPTIONS

Fig. 1. Mean magnitude of the sensitivity of CAM3 near-surface temperature to the timescale parameter for deep convection, `zmconv_tau`, using the MOAT method to vary 21 input parameters. Units are degrees Celsius because `zmconv_tau` is normalized and non-dimensionalized.

Fig. 2. Mean magnitude and standard deviation of the sensitivity of Outgoing Longwave Radiation to 21 parameters varied using the MOAT method and CAM3. Units on both axes are $W\ m^{-2}$ because all input parameters are normalized and non-dimensionalized. Number labels correspond to parameters in Table 1. Regions (i), (ii), and (iii) divide the parameters into areas that have (i) little influence on OLR, (ii) moderate influence, primarily by direct effects of the parameter's variation, and (iii) strong influence, primarily by nonlinear interactions with itself or with other parameters.

Fig. 3. Ranking of grid-point output sensitivity to 21 parameters varied using the MOAT method and CAM3. Rows give input parameters (see Table 1) and columns give output fields (see text).

Fig. 4. Time series of two global- and annual-mean output fields in CAM3 experiments using the same initial conditions but varying 21 input parameters (gray lines), and in 10 additional perturbed input parameter experiments, each

using different initial conditions (colored bands).

Fig. 5. Filtering of model output by comparison with observed planetary albedo in CAM4 experiments varying 27 input parameters. The green line is the least-squares fit to the data. The observed range is a 95% confidence interval inferred from satellite observations between 2000 and 2005.

Fig. 6. Filtering of model output parameter likelihood by Gaussian process surrogates and comparison with observations, in CAM4 experiments varying 28 input parameters. The diagonal shows the marginal prior (gray) and posterior (black) distributions of three selected output parameters: outgoing longwave radiation (FLUT) and both longwave and shortwave cloud forcing (LWCF and SWCF respectively), all in units of $W\ m^{-2}$. The cloud forcings are shown as annual means (ANN), December-January-February means (DJF), and June-July-August means (JJA). Outgoing longwave radiation is shown only as ANN. The red points and lines in the diagonal graphics show observed $\pm 2\sigma$ values. The off-diagonals show prior realizations (gray dots) and posterior realizations (black dots) from the bivariate distributions, with red crosses showing observed $\pm 2\sigma$ values.

Fig. 7. Filtering of (left) model input parameter likelihood and (right) model output by MARS surrogates and comparison with observations, in CAM4 experiments varying 28 input parameters. The marginal posterior distribution of one selected

input parameter and one global- and annual-mean output field are shown. The shaded 95% Confidence Interval corresponds to the observed range of planetary albedo in Fig. 6.

Table 1. CAM input parameters varied in perturbed-parameter AMIP experiments. Parameters 1-21 were varied in CAM3. In addition, Parameters 22-27 were varied in subsequent runs using CAM4 (green color). In a final set of runs using CAM4, Parameter 28 was also varied (pink color).

| Param # | Param Name | Range | | | Description | Namelist Prefix |
|---------|----------------|---------|---------|---------|---|-----------------|
| | | Low | Default | High | | |
| 1 | rhminh | 0.65 | 0.80 | 0.85 | Threshold RH for fraction of high stable clouds | cldfrc_ |
| 2 | rhminl | 0.80 | 0.91 | 0.99 | Threshold RH for fraction of low stable clouds | cldfrc_ |
| 3 | rliqice | 8.4 | 14.0 | 19.6 | Effective radius of liquid cloud droplets over sea ice | cldopt_ |
| 4 | rliqland | 4.8 | 8.0 | 11.2 | Effective radius of liquid cloud droplets over land | cldopt_ |
| 5 | rliqocean | 8.4 | 14.0 | 19.6 | Effective radius of liquid cloud droplets over ocean | cldopt_ |
| 6 | ice_stokes_fac | 0.25 | 0.50 | 1.00 | Scaling factor applied to ice fall velocity | cldsed_ |
| 7 | capnc | 30.0 | 150.0 | 155.0 | Cloud particle number density over cold land/ocean | cldwat_ |
| 8 | capnsi | 10.0 | 75.0 | 100.0 | Cloud particle number density over sea ice | cldwat_ |
| 9 | capnw | 150.0 | 400.0 | 500.0 | Cloud particle number density over warm land | cldwat_ |
| 10 | conke | 2.0e-6 | 5.0e-6 | 10.0e-6 | Evaporation efficiency of stratiform precipitation | cldwat_ |
| 11 | icritc | 2.0e-6 | 9.5e-6 | 18.0e-6 | Threshold for autoconversion of cold ice | cldwat_ |
| 12 | icritw | 1.0e-4 | 2.0e-4 | 10.0e-4 | Threshold for autoconversion of warm ice | cldwat_ |
| 13 | r3lcrit | 5.0e-6 | 10.0e-6 | 14.0e-6 | Critical radius at which autoconversion becomes efficient | cldwat_ |
| 14 | ricr | 0.1 | 0.3 | 1.0 | Critical Richardson number for boundary layer | hbdiff_ |
| 15 | c0 | 0.3e-4 | 1.0e-4 | 2.0e-4 | Shallow convection precipitation efficiency | hkconv_ |
| 16 | cmftau | 900.0 | 1800.0 | 14400.0 | Time scale for consumption rate of shallow CAPE | hkconv_ |
| 17 | alfa | 0.05 | 0.10 | 0.60 | Initial cloud downdraft mass flux | zmconv_ |
| 18 | c0 | 1.0e-3 | 3.5e-3 | 6.0e-3 | Deep convection precipitation efficiency | zmconv_ |
| 19 | dmpdz | -2.0e-3 | -1.0e-3 | -0.2e-3 | Parcel fractional mass entrainment rate | zmconv_ |
| 20 | ke | 0.5e-6 | 1.0e-6 | 10.0e-6 | Environmental air entrainment rate | zmconv_ |
| 21 | tau | 1800.0 | 3600.0 | 28800.0 | Time scale for consumption rate of deep CAPE | zmconv_ |
| 22 | fac | 10.0 | 100.0 | 200.0 | ustar parameter in PBL height diagnosis | hbdiff_ |
| 23 | fak | 4.25 | 8.50 | 17.00 | Constant in surface temperature excess | hbdiff_ |
| 24 | betamn | 0.02 | 0.10 | 0.30 | Minimum overshoot parameter | hkconv_ |
| 25 | capelmt | 20.0 | 70.0 | 200.0 | Threshold value for CAPE for deep convection | zmconv_ |
| 26 | cdn_scal_fac | 0.8 | 1.0 | 1.2 | Ocean roughness scaling factor | (drv_in) |
| 27 | z0m_scal_fac | 0.8 | 1.0 | 1.2 | Moisture & heat resistance to vegetation scaling factor | (lnd_in) |
| 28 | sgl_scal_fac | 0.8 | 1.0 | 1.2 | Land roughness scaling factor | physpkg_ |

Table 2. Summary of CAM 12-year AMIP simulations.

| Study # | Study Name | CAM Version | # UQ Params | Study Runs | Sim Yrs | Stored Size (TB) |
|---------------------|------------|-------------|-------------|--------------|---------------|------------------|
| 1 | moat1 | 3.6 | 21 | 220 | 2,640 | 3.4 |
| 2 | oat4 | 3.6 | 21 | 43 | 516 | 0.7 |
| 3 | moat2 | 3.6 | 21 | 220 | 2,640 | 3.2 |
| 4 | lhs01 | 3.6 | 21 | 110 | 1,320 | 1.6 |
| 5 | vbd01 | 3.6 | 21 | 121 | 1,452 | 1.7 |
| 6 | nond01 | 3.6 | 21 | 440 | 5,280 | 6.6 |
| 7 | ics01 | 3.6 | 21 | 88 | 1,056 | 1.5 |
| 8 | oat6 | 4.0 | 27 | 55 | 660 | 0.8 |
| 9 | moat3 | 4.0 | 27 | 280 | 3,360 | 4.3 |
| 10 | nond02 | 4.0 | 27 | 560 | 6,720 | 8.3 |
| 11 | oat7 | 4.0 | 28 | 57 | 684 | 0.8 |
| 12 | nond03 | 4.0 | 28 | 140 | 1,680 | 1.9 |
| 13 | nond04 | 4.0 | 28 | 280 | 3,360 | 3.8 |
| 14 | nond05 | 4.0 | 28 | 280 | 3,360 | 3.8 |
| Totals => | | | | 2,894 | 34,728 | 42.2 |

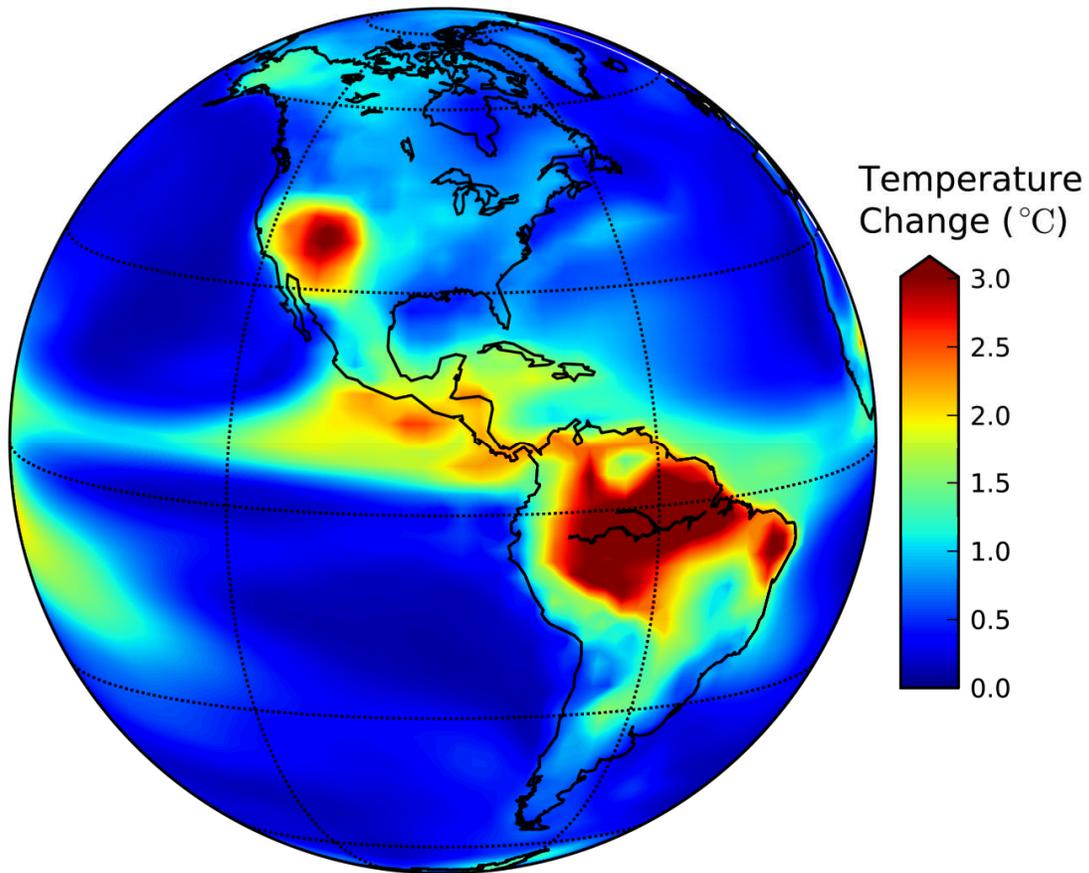


Fig. 1. Mean magnitude of the sensitivity of CAM3 near-surface temperature to the timescale parameter for deep convection, $zmconv_tau$, using the MOAT method to vary 21 input parameters. Units are degrees Celsius because $zmconv_tau$ is normalized and non-dimensionalized.

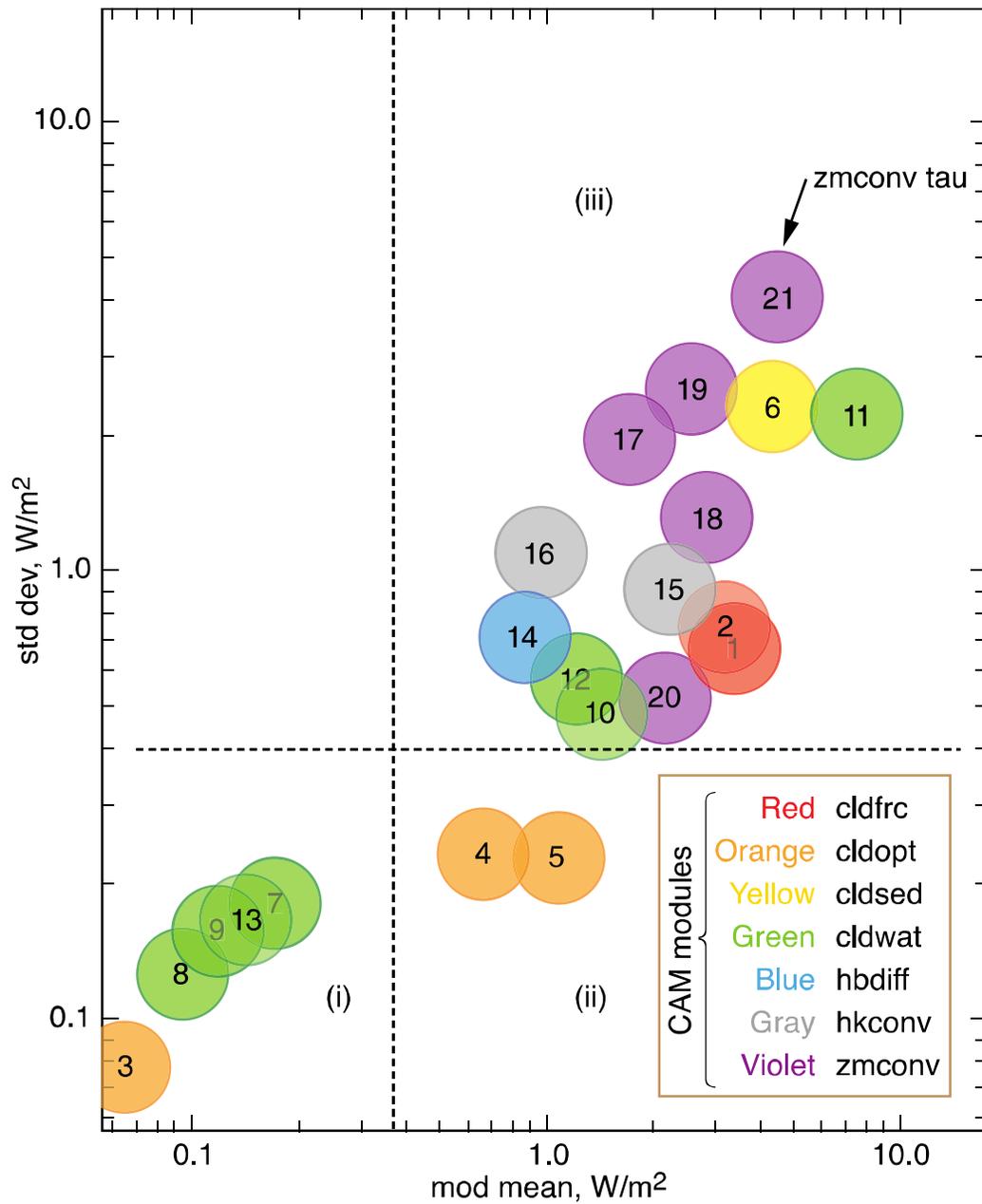


Fig. 2. Mean magnitude and standard deviation of the sensitivity of Outgoing Longwave Radiation to 21 parameters varied using the MOAT method and CAM3. Units on both axes are $W\ m^{-2}$ because all input parameters are normalized and non-dimensionalized. Number labels correspond to parameters in Table 1. Regions (i), (ii), and (iii) divide the parameters into areas that have (i) little influence on OLR, (ii) moderate influence, primarily by direct effects of the parameter's variation, and (iii) strong influence, primarily by nonlinear interactions with itself or with other parameters.

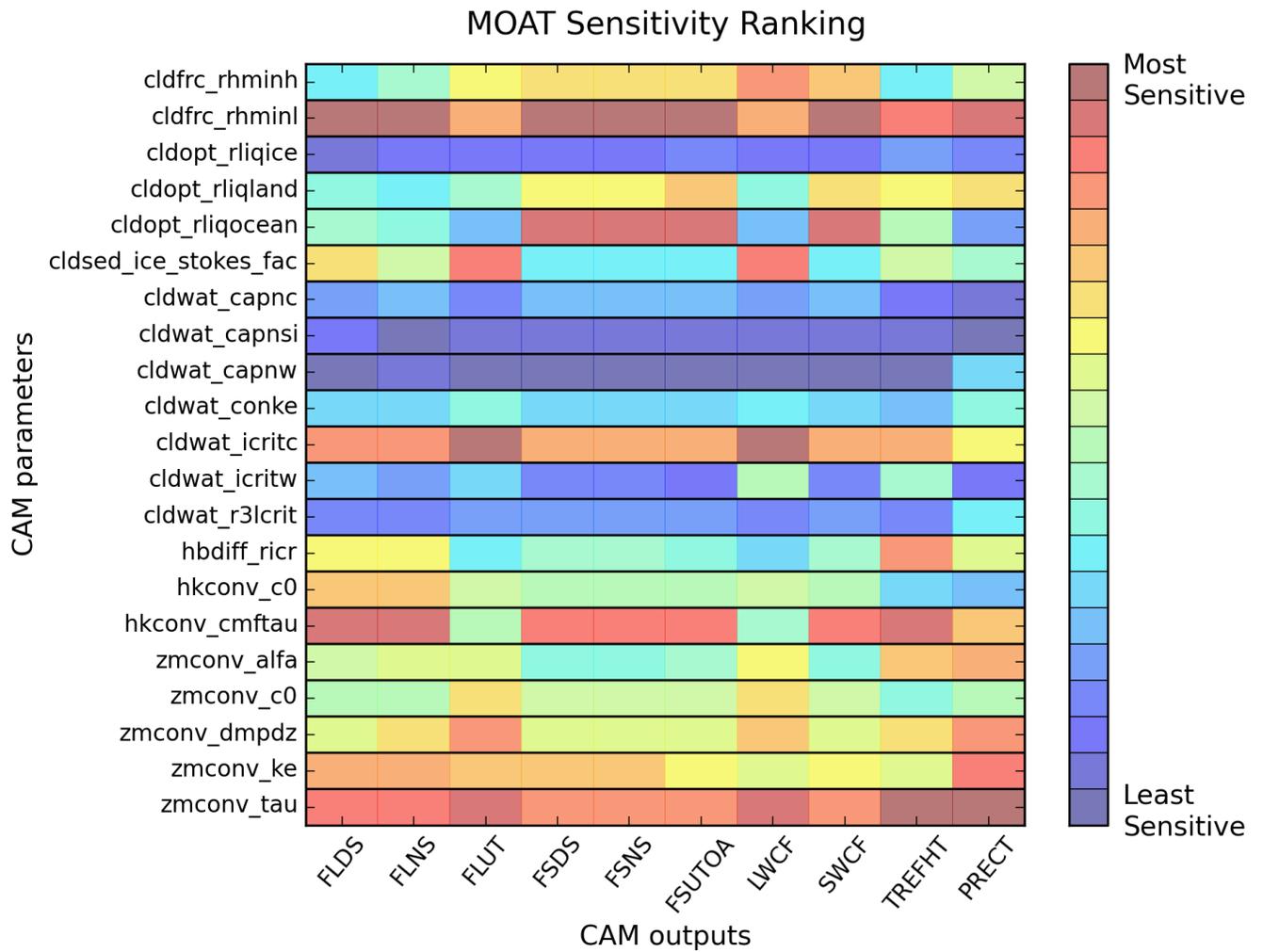


Fig. 3. Ranking of grid-point output sensitivity to 21 parameters varied using the MOAT method and CAM3. Rows give input parameters (see Table 1) and columns give output fields (see text).

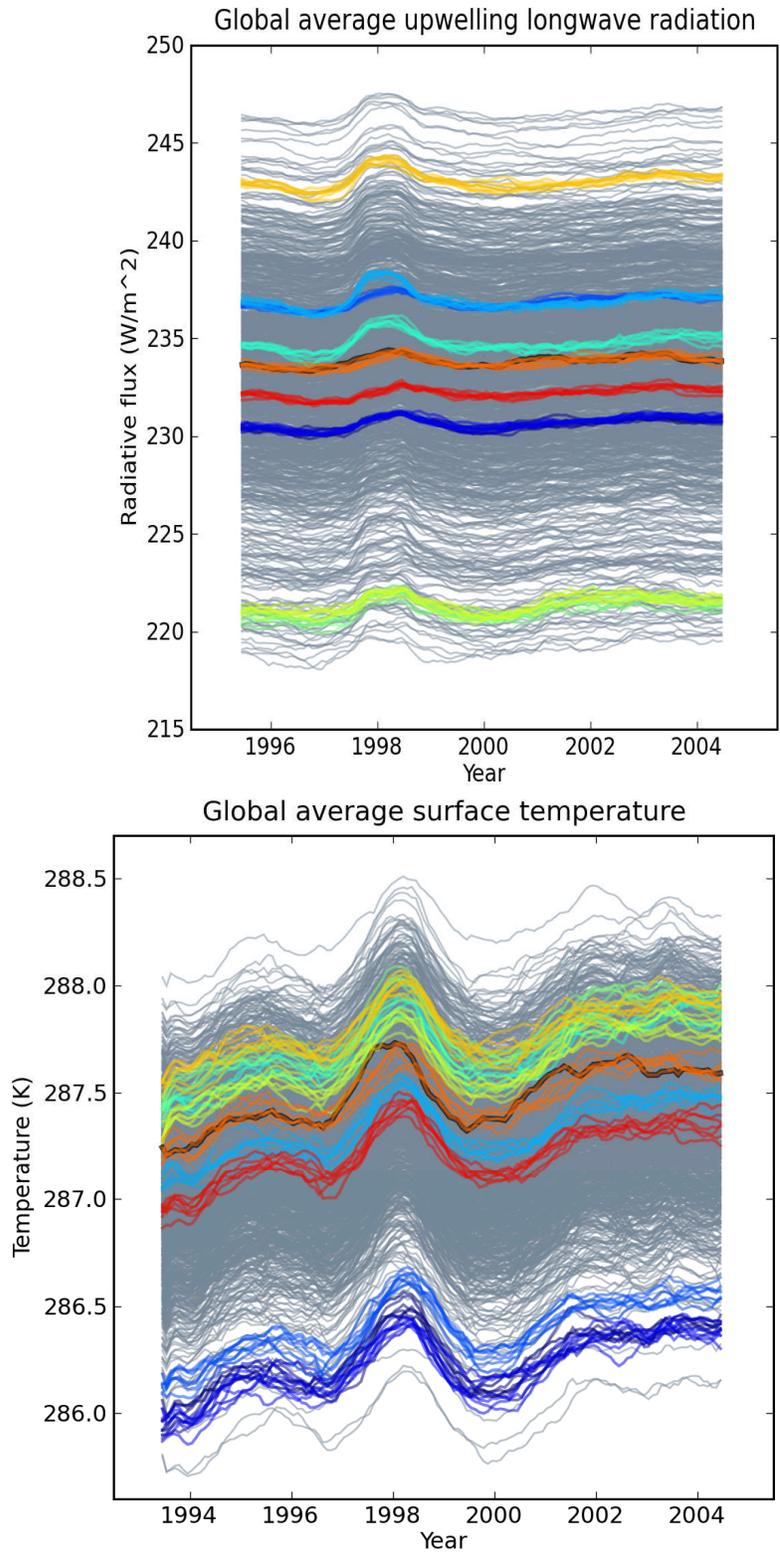


Fig. 4. Time series of two global- and annual-mean output fields in CAM3 experiments using the same initial conditions but varying 21 input parameters (gray lines), and in 10 additional perturbed input parameter experiments, each using different initial conditions (colored bands).

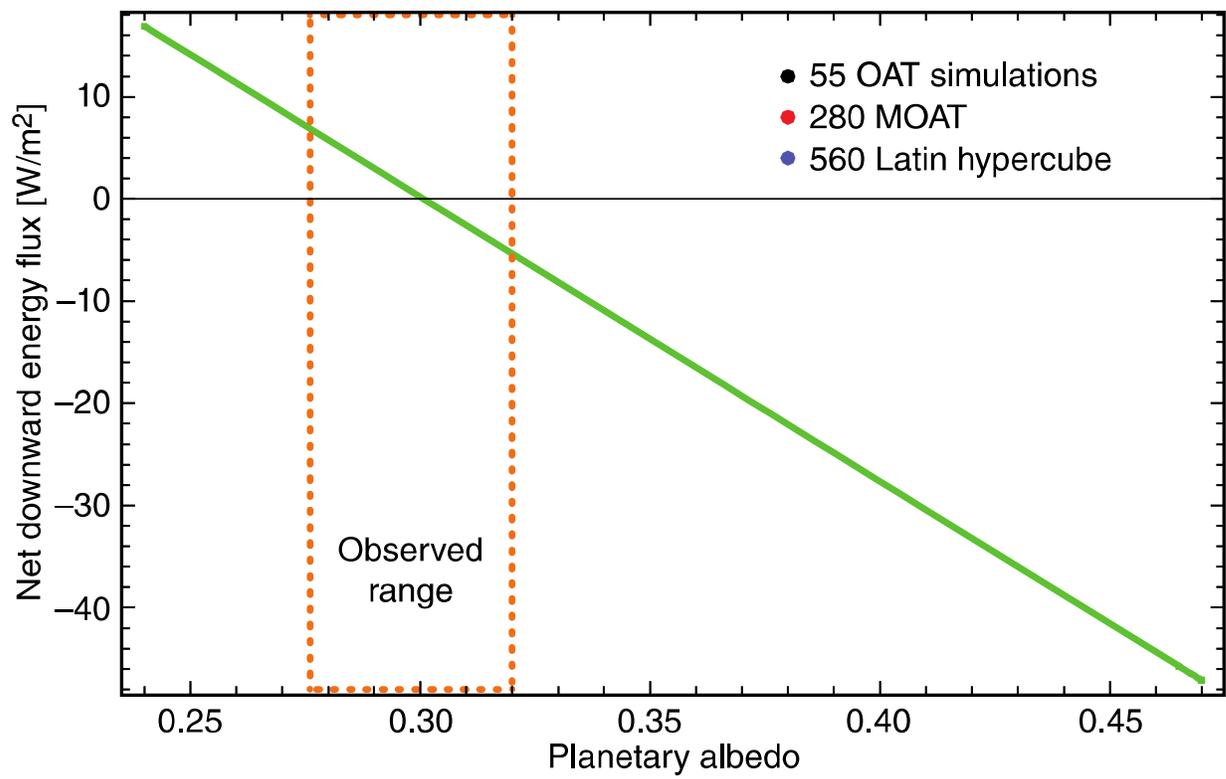


Fig. 5. Filtering of model output by comparison with observed planetary albedo in CAM4 experiments varying 27 input parameters. The green line is the least-squares fit to the data. The observed range is a 95% confidence interval inferred from satellite observations between 2000 and 2005.

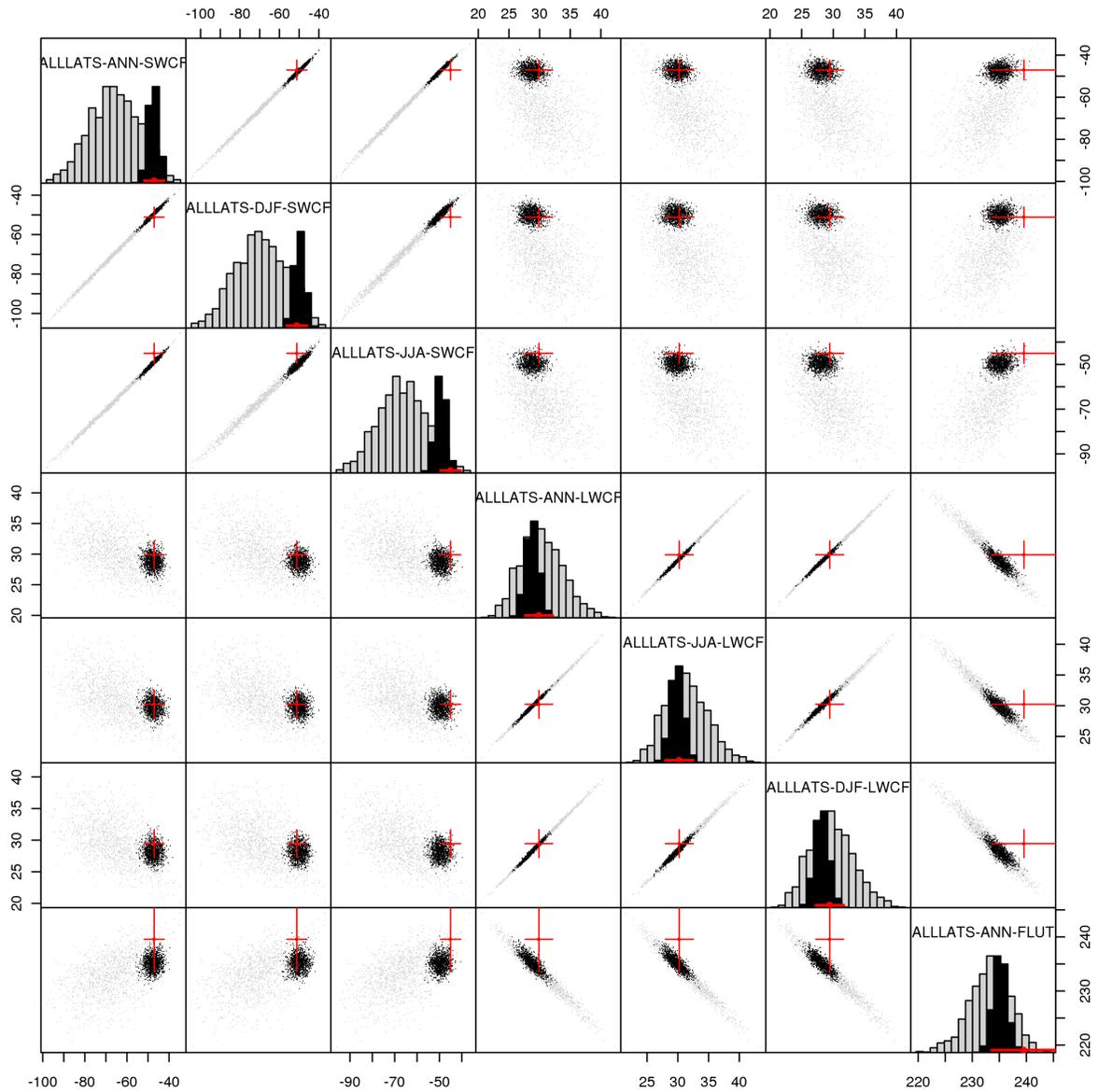


Fig. 6. Filtering of model output parameter likelihood by Gaussian process surrogates and comparison with observations, in CAM4 experiments varying 28 input parameters. The diagonal shows the marginal prior (gray) and posterior (black) distributions of three selected output parameters: outgoing longwave radiation (FLUT) and both longwave and shortwave cloud forcing (LWCF and SWCF respectively), all in units of $W m^{-2}$. The cloud forcings are shown as annual means (ANN), December-January-February means (DJF), and June-July-August means (JJA). Outgoing longwave radiation is shown only as ANN. The red points and lines in the diagonal graphics show observed $\pm 2\sigma$ values. The off-diagonals show prior realizations (gray dots) and posterior realizations (black dots) from the bivariate distributions, with red crosses showing observed $\pm 2\sigma$ values.

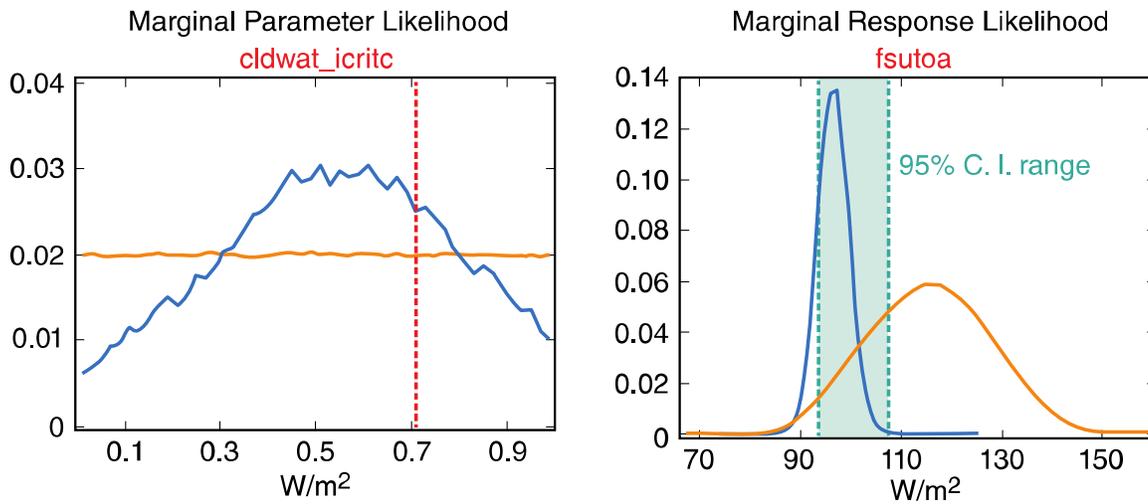


Fig. 7. Filtering of (left) model input parameter likelihood and (right) model output by MARS surrogates and comparison with observations, in CAM4 experiments varying 28 input parameters. The marginal posterior distribution of one selected input parameter and one global- and annual-mean output field are shown. The shaded 95% Confidence Interval corresponds to the observed range of planetary albedo in Fig. 6.