Scientific Aspects of the Climate UQ Project @ LLNL

Donald D. Lucas + Climate & UQ Teams

Discussion Topics

- **Overview of climate UQ project and methods**
- **Parameter sensitivity update**
- ***Filtering UQ ensembles through observations***
- **Future directions (UQ simulations and observations)**

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

LLNL-PRES-470077

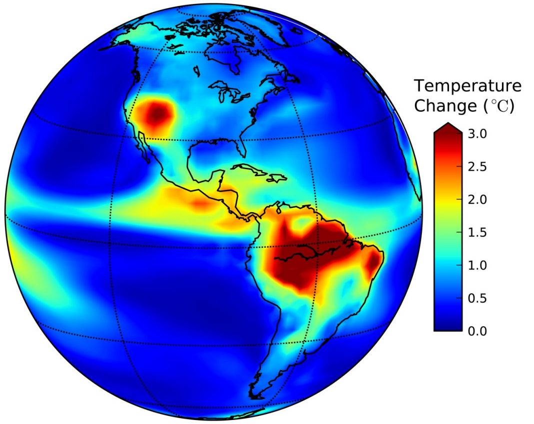
Climate and UQ Teams

"The Advance of UQ Science with Applications to Climate Modeling, Inertial Confinement Fusion (ICF), and Stockpile Stewardship (SSS)," A three-year Laboratory Directed Research & Development Strategic Initiative

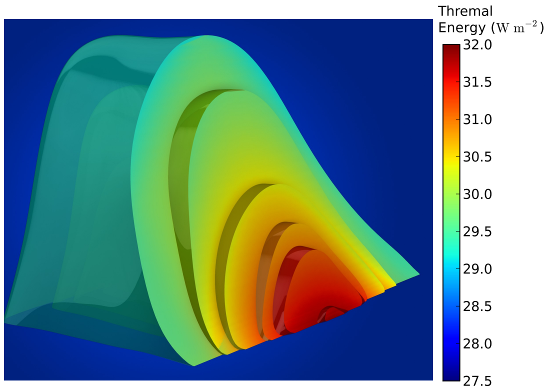
- **Project Leads**
  - Richard Klein (PI)
  - Xabier Garaizar (co-PI)
- **Climate Team**
  - Curt Covey (climate lead)
  - Donald Lucas (modeling and analysis)
  - John Tannahill (software architecture and development)
  - Yuying Zhang (observations and analysis)
- **UQ and Computation Teams**
  - David Domyancic, Scott Brandon (LLNL UQ Pipeline)
  - Gardar Johannesson (curse of dimensionality, adaptive sampling)
  - And others

Overview of Climate UQ @ LLNL

- Adaptively build an ensemble of climate simulations for present-day climate by perturbing uncertain input parameters of the Community Atmospheric Model (CAM)
- Carry out sensitivity and uncertainty analysis of the climate simulations
- Collect a comprehensive set of observations to use for UQ (emphasis on cloud-related observations)
- Calibrate input parameters using observations
- Calculate PDF of climate sensitivity
- Perform UQ analysis of climate change using coupled models and adaptive sampling refinement in LLNL's *UQ Pipeline*



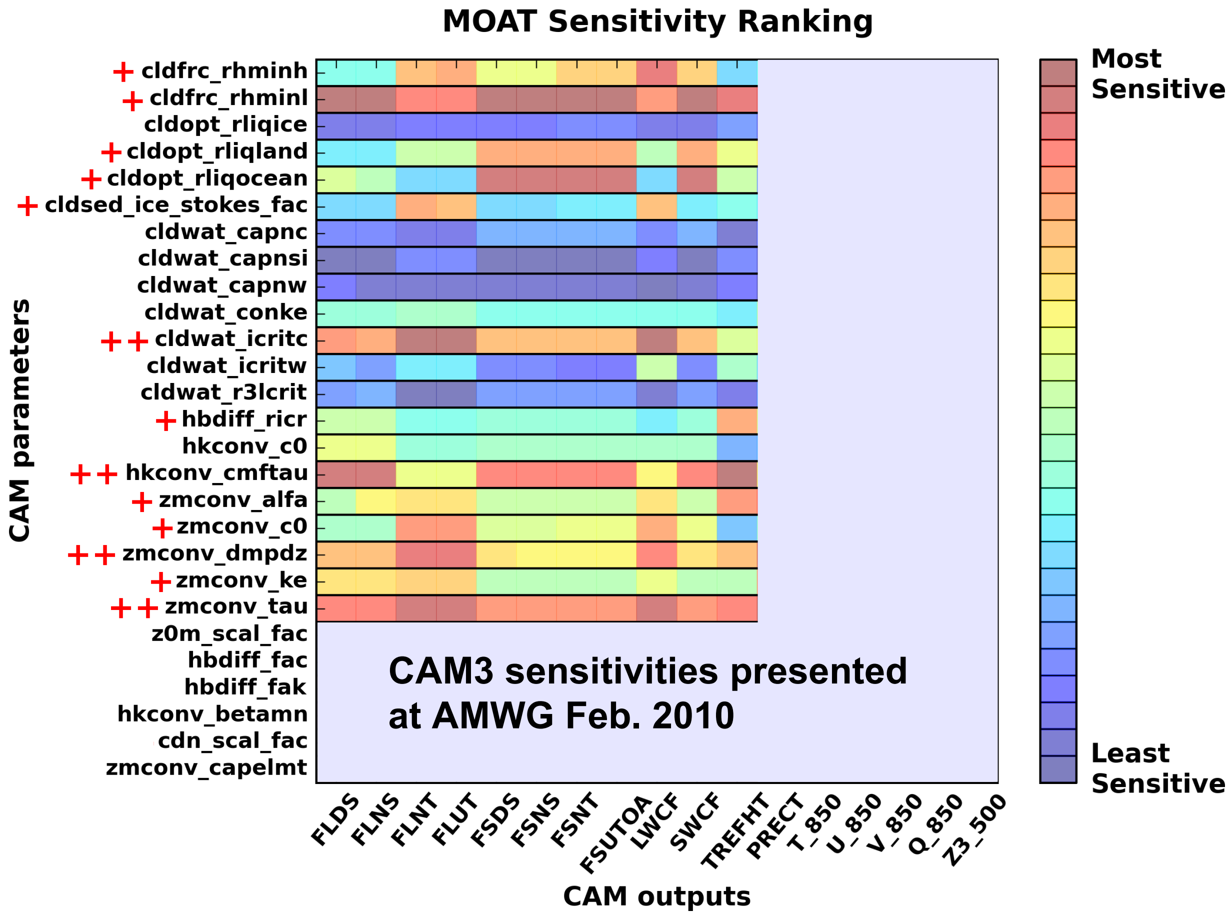
Example of a sensitivity map calculated using the Morris method on CAM3 in a high dimensional parameter space.



Example of a response surface generated using polynomial chaos expansions on CAM3 ensembles. (rendering by Kwei-Yu Chu)

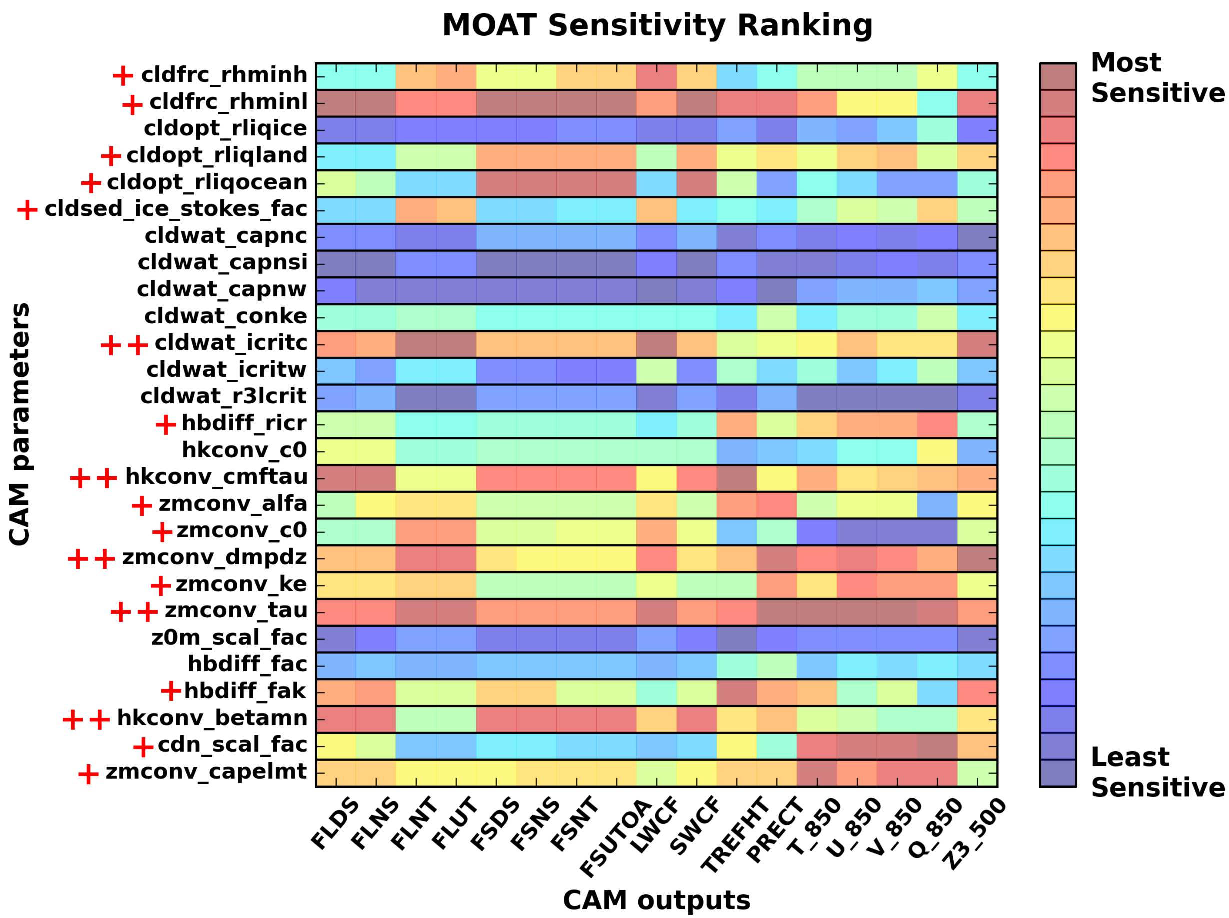
Update on CAM4 parameter sensitivities

- Multiple *global* sensitivity methods were applied to CAM. These methods are *global* in parameter space, accounting for nonlinear parameter interactions.
- Global sensitivity measures are used to identify important parameters and categorize linear and nonlinear effects.
- Highly ranked parameters are targets for calibration.
- A sensitivity ranking for CAM4 using the Morris screening method is shown on the right [Morris, Technometrics (1991)].
    - 27 parameters are ranked across 17 outputs
    - A handful of parameters are important to many outputs (++)
    - Many parameters are important to at least one output (+)

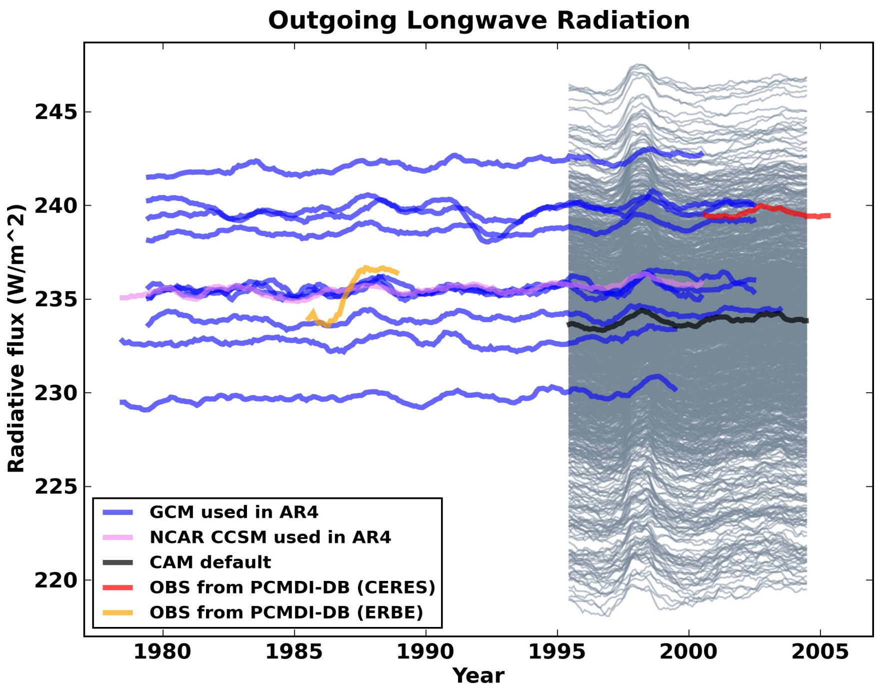


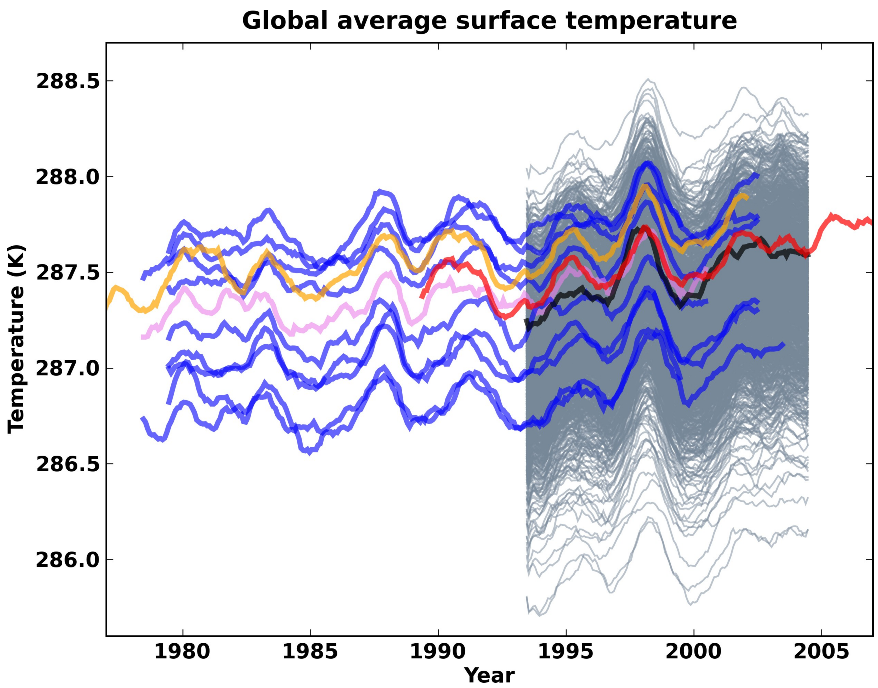
Update on CAM4 parameter sensitivities

- Multiple *global* sensitivity methods were applied to CAM. These methods are *global* in parameter space, accounting for nonlinear parameter interactions.
- Global sensitivity measures are used to identify important parameters and categorize linear and nonlinear effects.
- Highly ranked parameters are targets for calibration.
- A sensitivity ranking for CAM4 using the Morris screening method is shown on the right [Morris, Technometrics (1991)].
  - 27 parameters are ranked across 17 outputs
  - A handful of parameters are important to many outputs (++)
  - Many parameters are important to at least one output (+)



Examples of Unfiltered Ensembles



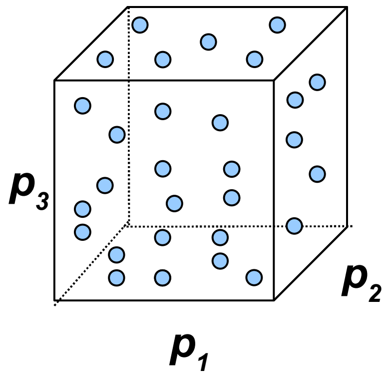


- Unfiltered ensembles consider only the *prior* parameter uncertainties
- *Filtering* is the process of constraining the ensembles with observations
- Having a large unfiltered ensemble spread facilitates the filtering process (i.e. it's easier to interpolate than extrapolate!)

Climate UQ Machinery

Hypercube Analysis

(global sensitivities, unfiltered uncertainties)



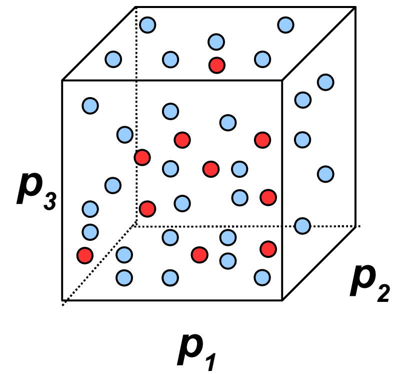
○ = CAM AMIP simulations at sample points

(LHS & MOAT sampling)

$R = f(p_1, p_2, p_3, \ldots)$

Surrogate Models

- Gaussian process models
- Polynomial chaos expansions
- Support Vector Regression
- Multivariate Adaptive Regression Splines (MARS)



● = Surrogate predictions at new sample points

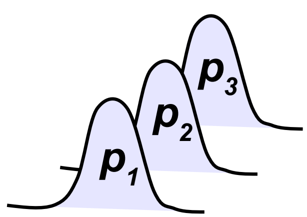
Observational constraint filter

Filtering Methods

- Maximum likelihood parameter estimation
- Statistical filtering
  - sample R using LHS
  - calculate likelihoods
- Bayesian calibration

Filtering Analysis

(parameter PDFs, response PDFs)

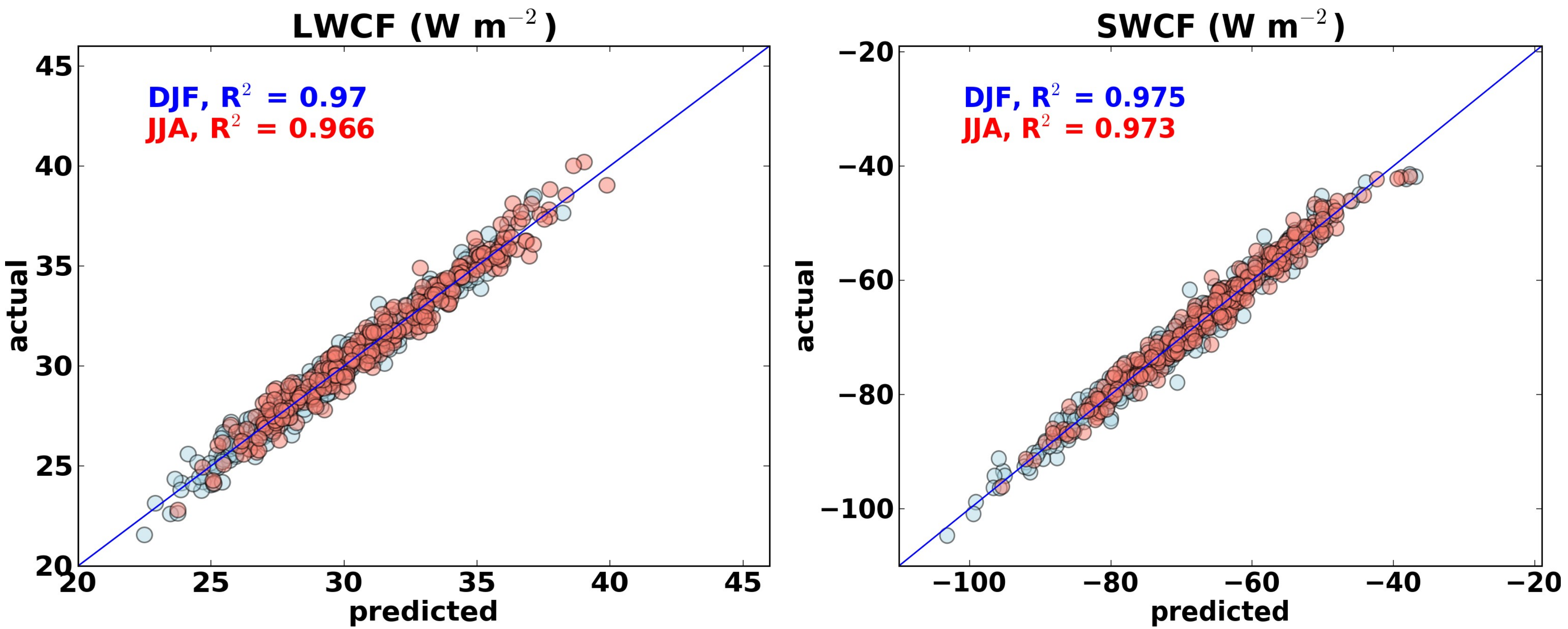


Uncertainty Propagation

PDFs of present day climate quantities of interest

PDFs of future climate quantities of interest (*climate sensitivity*)

Independent Validation of Surrogate Models



- **Surrogate models are validated using independent data.**
- **Examples of the actual and predicted `LWCF` and `SWCF` responses are displayed above.**
  - **surrogates were derived using *Support Vector Regression* trained on over 1,000 CAM4 runs and tested on 300 independent runs.**
- **Surrogate model errors can be important and should be factored in the UQ analysis.**

Estimating CAM4 Parameters Using Surrogate-Based Optimization

Goal: **determine "unknown" parameter values in default version of CAM4 using climate "observations"**

Step 1: Used output fields generated from *CAM4 default* as the target "observations" (*CAM4 default* = oat7, run0001)

Step 2: Used output fields generated from over 1,300 other CAM4 runs to build and validate surrogate models (mainly LHS runs)

- **+ Built 18 surrogates using SVM regression (`CLDTOT`, `FLUT`, `FSUTOA`, `PRECT`, `LWCF`, `T_850`, `TREFHT`, `SWCF`, and `Z3_500`; DJF and JJA global averages)**
- **+ Five-fold cross validation used to tune SVM-R hyperparameters ($R^2 > 0.9$, # SV's ~ ¼ training data size)**
- **+ Held out 300 runs for independent validation (previous slide)**

Step 3: Used inexpensive surrogates to search the parameter space for optimal match with "observations"

Cast as a bound-constrained optimization problem

$$\text{minimize: } \{f_1(p), f_2(p), \ldots\},\ p \in \mathbb{R}^n,\ f_i = |mod_i - obs_i|/scale$$

$$\text{subject to: } p_L \le p \le p_U$$

Estimating CAM4 Parameters Using Surrogate-Based Optimization

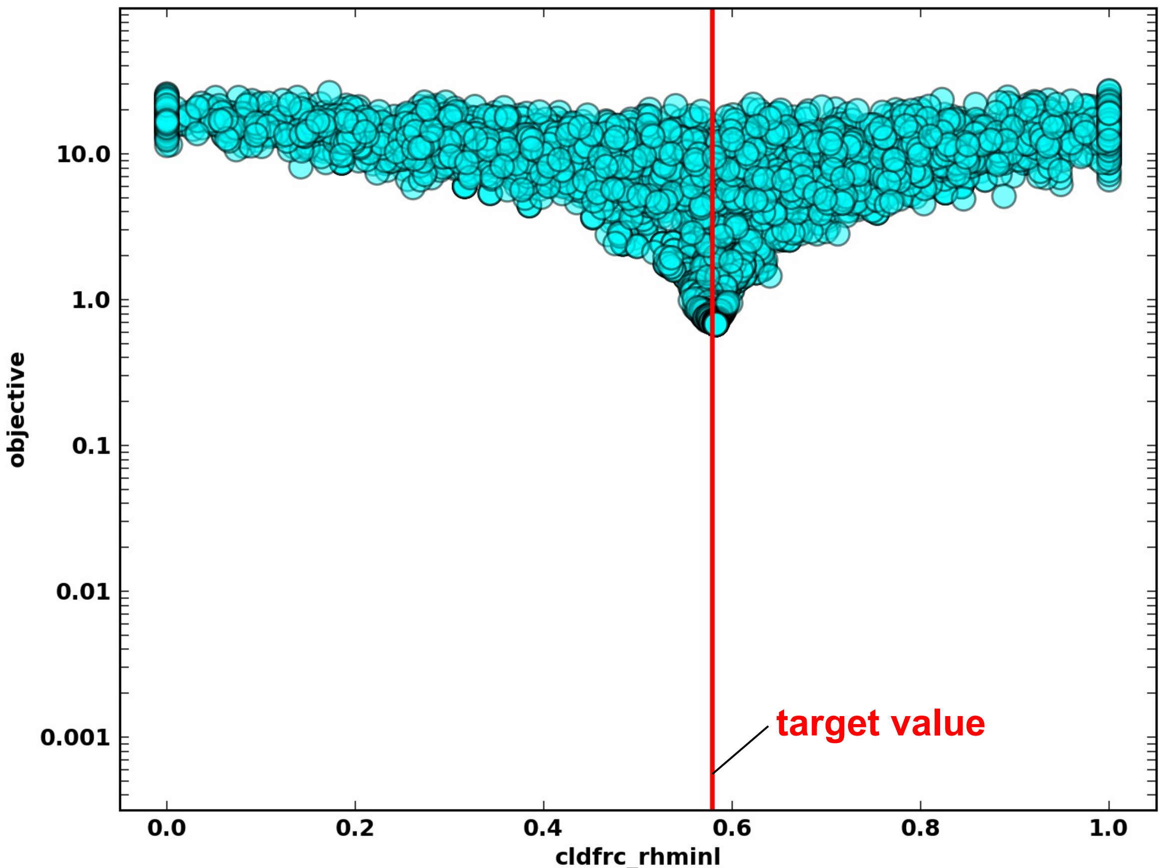
Use multi-stage hybrid sequential optimization

On right: single parameter slice of of a multi-dimensional search. The target value is known in advance (red line). Lower objectives are better.

stage 1: **global, gradient-free search using a genetic algorithm** (7,600 samples, top 25 samples passed to next stage)

stage 2: **local, gradient-free refinement using pattern search algorithm** (30,000 samples, top 25 samples passed to next stage)

stage 3: **local, gradient-based refinement using Newton-type method**



Estimating CAM4 Parameters Using Surrogate-Based Optimization

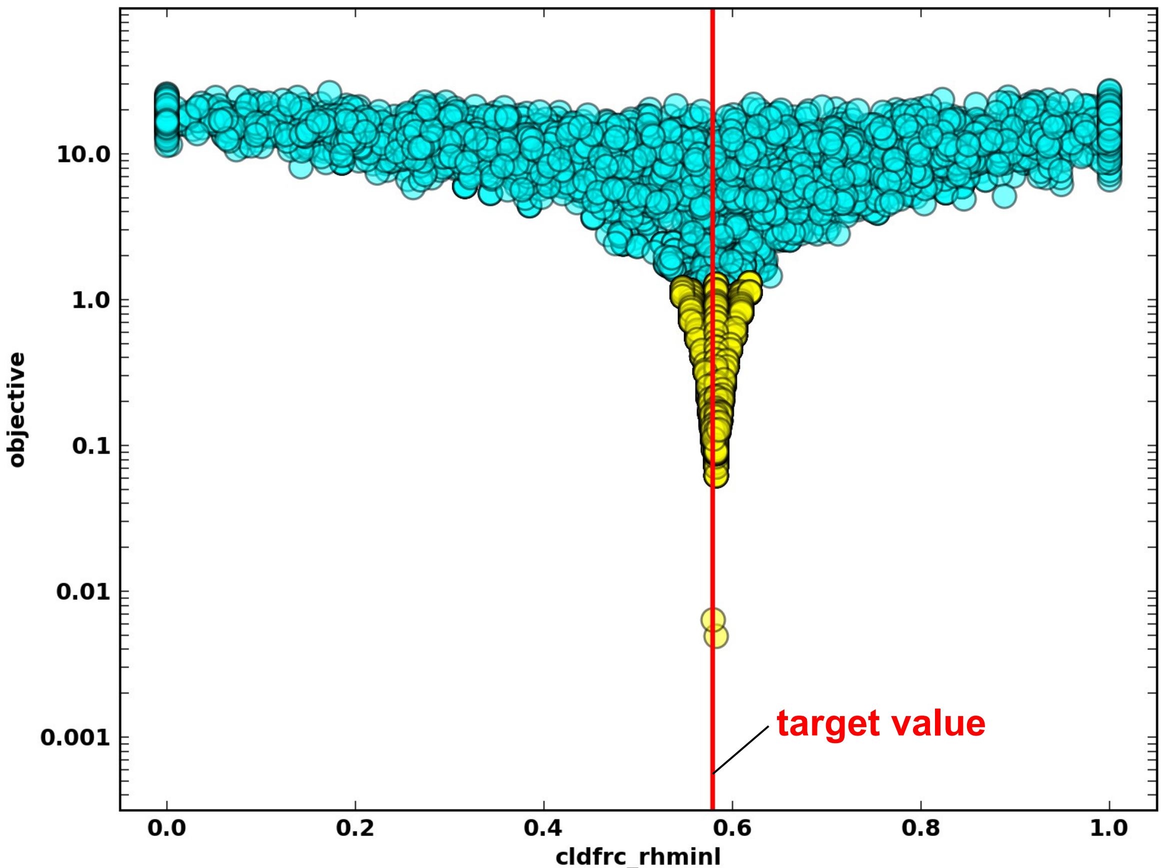
Use multi-stage hybrid sequential optimization

On right: single parameter slice of of a multi-dimensional search. The target value is known in advance (red line). Lower objectives are better.

stage 1: **global, gradient-free search using a genetic algorithm (7,600 samples, top 25 samples passed to next stage)**

stage 2: **local, gradient-free refinement using pattern search algorithm (30,000 samples, top 25 samples passed to next stage)**

stage 3: **local, gradient-based refinement using Newton-type method**



Estimating CAM4 Parameters Using Surrogate-Based Optimization

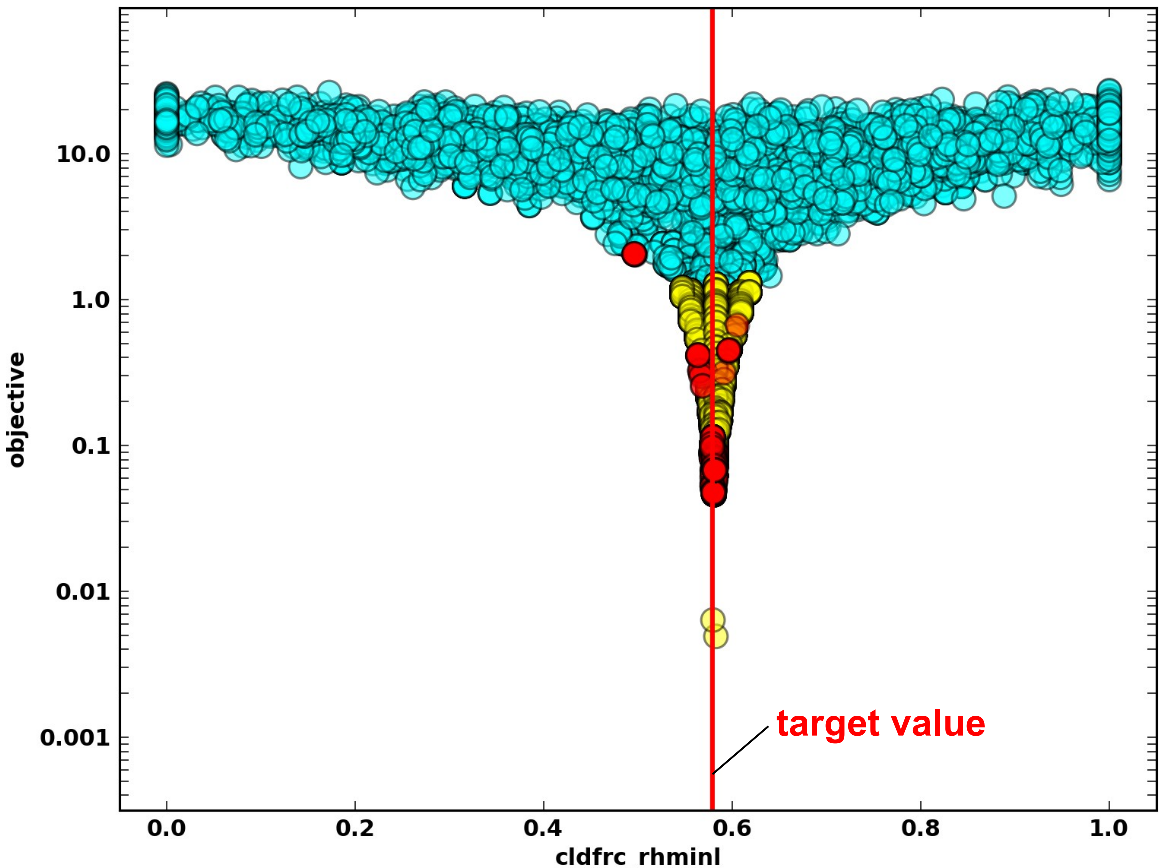
Use multi-stage hybrid sequential optimization

On right: single parameter slice of of a multi-dimensional search. The target value is known in advance (red line). Lower objectives are better.

stage 1: **global, gradient-free search using a genetic algorithm** **(7,600 samples, top 25 samples passed to next stage)**

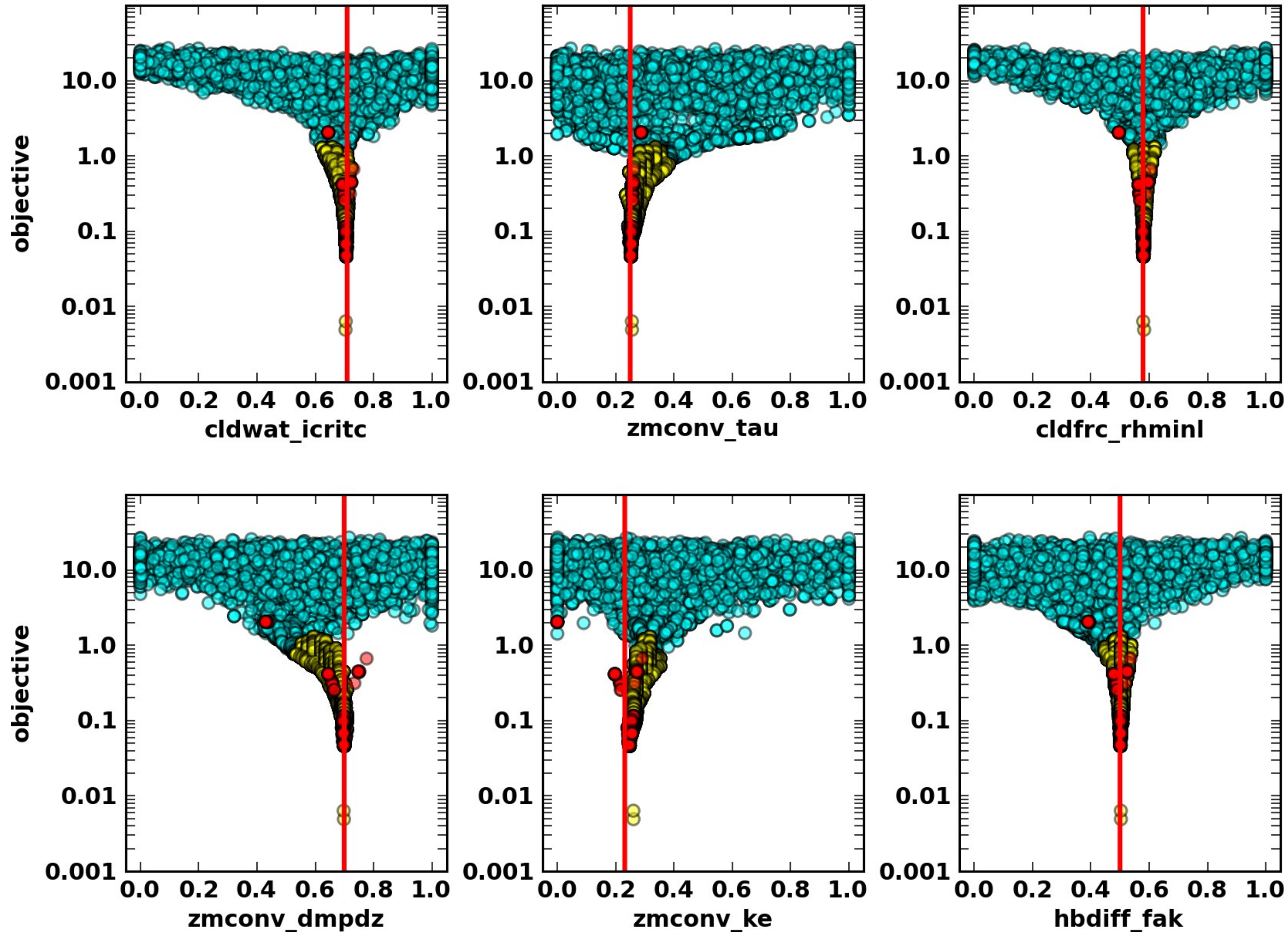
stage 2: **local, gradient-free refinement using pattern search algorithm** **(30,000 samples, top 25 samples passed to next stage)**

stage 3: **local, gradient-based refinement using Newton-type method**



Estimating CAM4 Parameters Using Surrogate-Based Optimization

- Surrogate-based optimization estimates maximum likelihood values for multiple parameters.
- Cheap to execute (runs on a workstation).
- Can be used to refine the search space for high dimensional systems.
- Need to add model and data uncertainties for UQ. *Optimization Under Uncertainty* (OUU) provides a framework for doing this.



Statistical Filtering Example (analysis by S. Brandon)

Approach used by PI's V&V group.

Trained and validated 24 MARS surrogate models on ~1,300 LHS CAM4 simulations: [`FLUT, FSUTOA, LWCF, PRECT, Q_850, SWCF, T_850, Z3_500`] x [ANN, DJF, JJA]

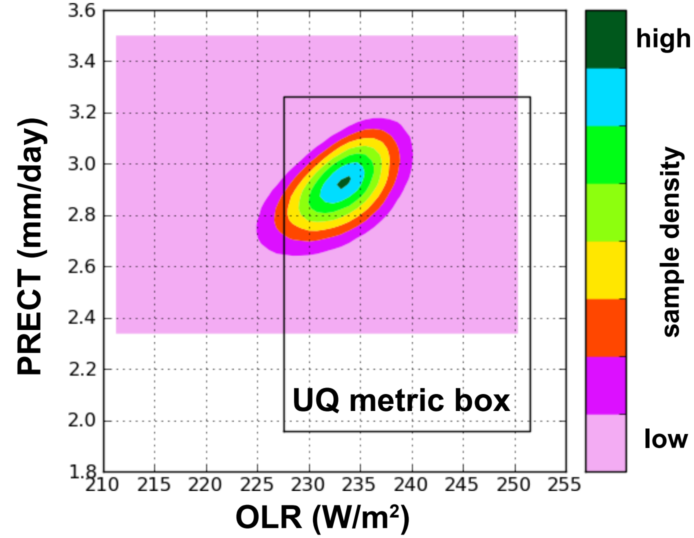
Observational constraints (w/ "loose" uncertainties): CERES (`FLUT, LWCF, SWCF`), GPCP (`PRECT`), NCEP (`Z3_500`)

Brute-force sampling (LHS) of MARS surrogates to generate likelihoods.

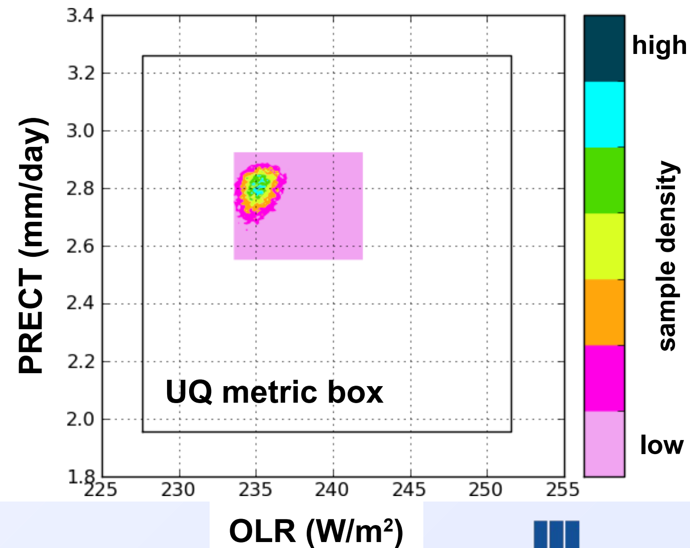
Likelihoods computed with various filters (e.g. top-hat (1 or 0) or Gaussian).

Application of the filters collapses the UQ space; about 10% of the samples satisfy the filters.

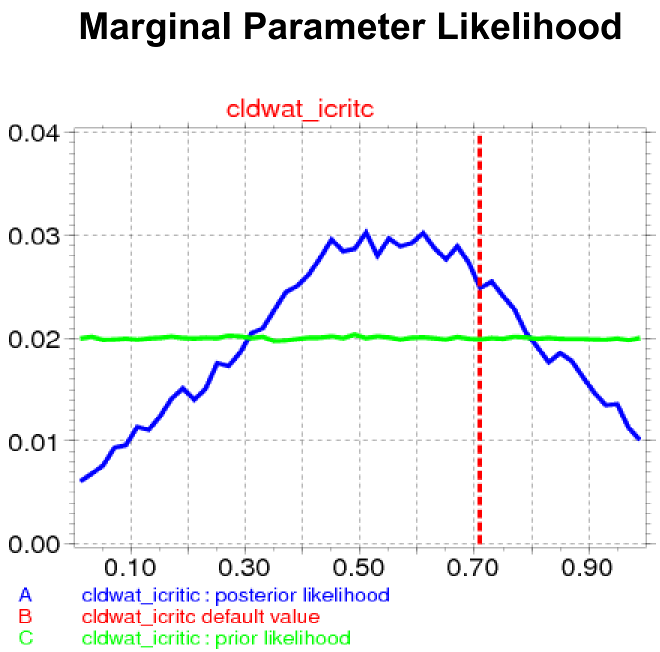
Unfiltered UQ space (projection)

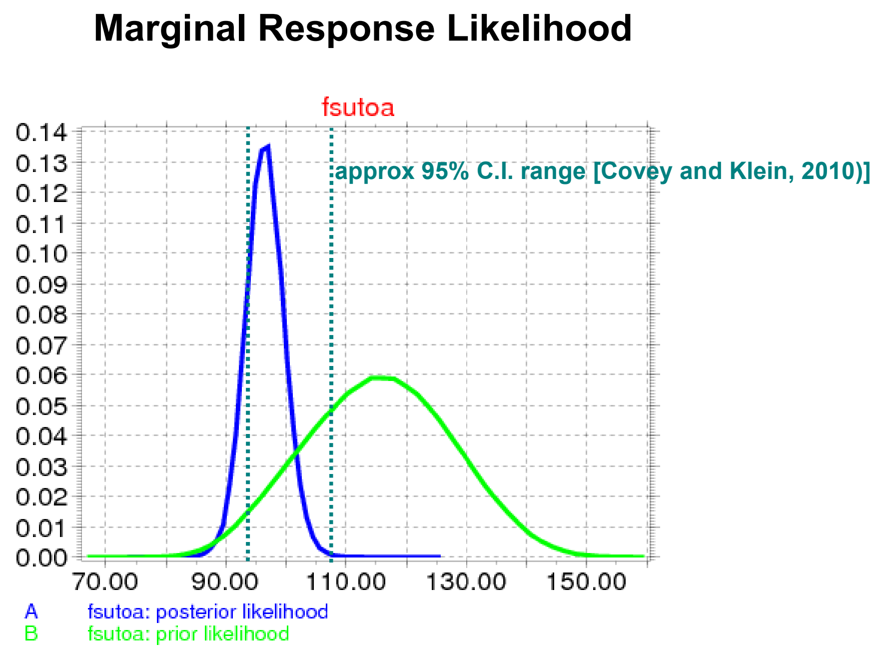


Filtered UQ space (projection)



Statistical Filtering Example (analysis by S. Brandon)





- **Most parameters are not constrained very much by the "loose" filter assumptions**
- **Posterior PDFs generated by normalizing the likelihoods**
- **`FSUTOA` is constrained even though an observational constraint for `FSUTOA` is not applied**

Bayesian Calibration Example (analysis by G. Johannesson)

Sample joint posterior distribution given prior information (uniform PDFs) and observational constraints (likelihoods).

Trained and validated 24 Gaussian Process surrogate models on ~1,300 LHS CAM4 simulations: [`FLUT`, `FSUTOA`, `LWCF`, `PRECT`, `Q_850`, `SWCF`, `T_850`, `Z3_500`] x [ANN, DJF, JJA]

Observational constraints (w/ "loose" uncertainties): CERES (`FLUT`, `LWCF`, `SWCF`), GPCP (`PRECT`), NCEP (`Z3_500`)

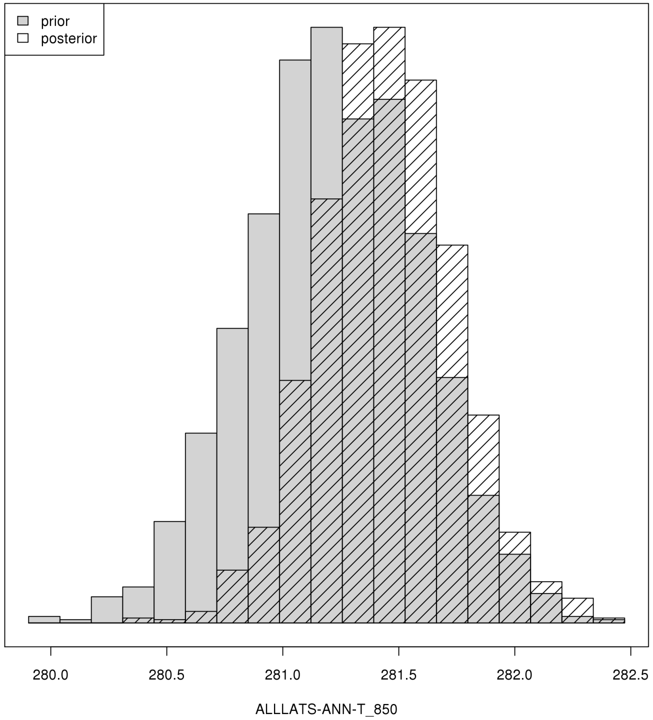
Use a hierarchical Bayesian model

OBS = SYS + OBS_err
SYS = CAM4(p) + MOD_err
CAM4(p) = SURR(p) + SURR_err

MCMC used to sample the joint posterior distribution.

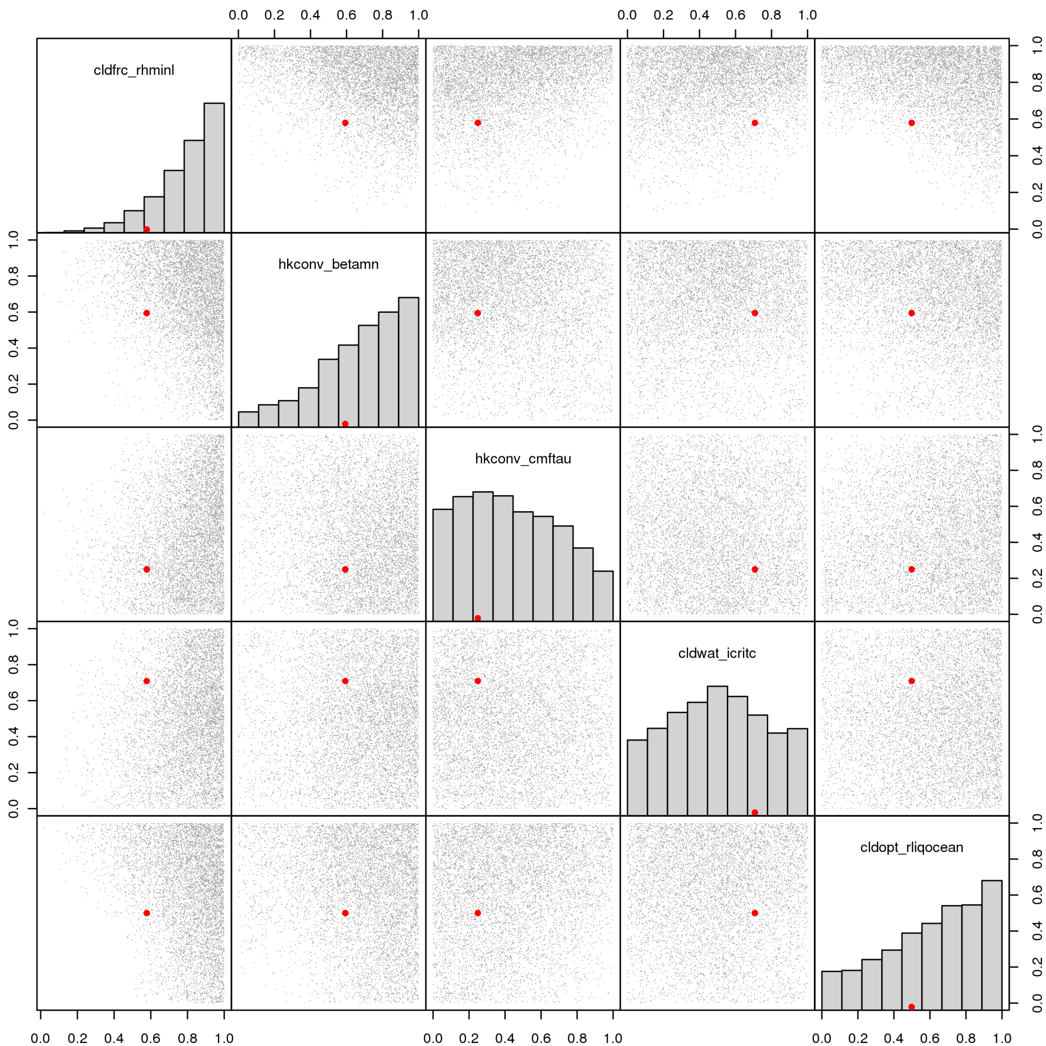
$$P(\text{params} \mid \text{obs}) \propto P(\text{obs} \mid \text{params})\, P(\text{params})$$

posterior **likelihood** **flat priors**



Above: prior and posterior PDFs for a response to which observational constraints were not applied

Bayesian Calibration Example (analysis by G. Johannesson)



Posterior Parameter PDF

Diagonal shows the marginal posterior distribution of 5 selected input parameters (those most constrained by the observations)

Off-diagonal shows posterior realizations (dots) from the bivariate distributions

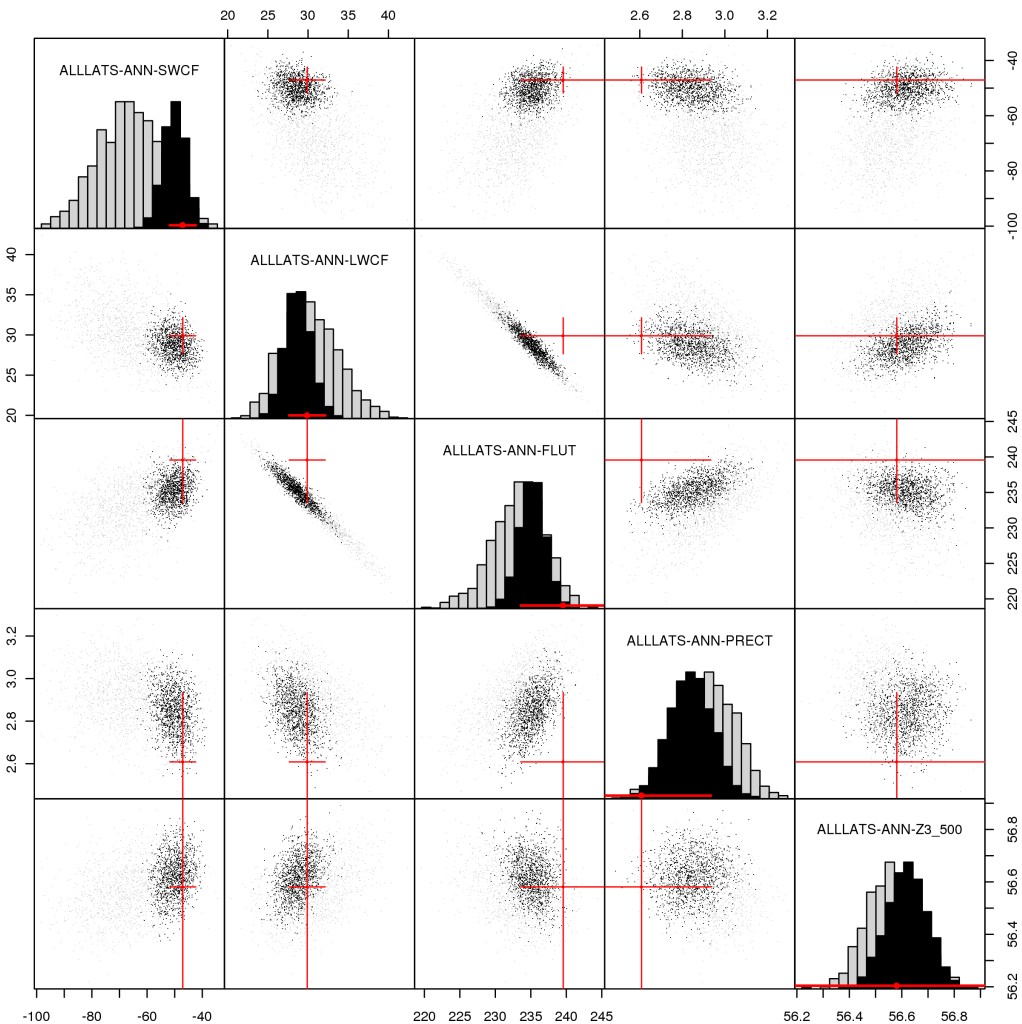
Red dots show the default values

Bayesian Calibration Example (analysis by G. Johannesson)

Posterior distribution of selected output variables

- **Diagonal (marginal)**
  - **light-gray histograms show the prior (unfiltered) distributions**
  - **black histogram the posterior (filtered) distributions**
  - **red dots/bars show the observational constraints**

- **Off-diagonal (bivariate)**
  - **light-gray scatter plots show prior distributions**
  - **black scatter plots show posterior distributions**
  - **along with observations and error bars**



Summary and Next Steps

- **We have developed and demonstrated techniques for performing full UQ analysis on CAM using:**
  - surrogate models as inexpensive proxies for CAM
  - multiple methods for combining observations and ensembles

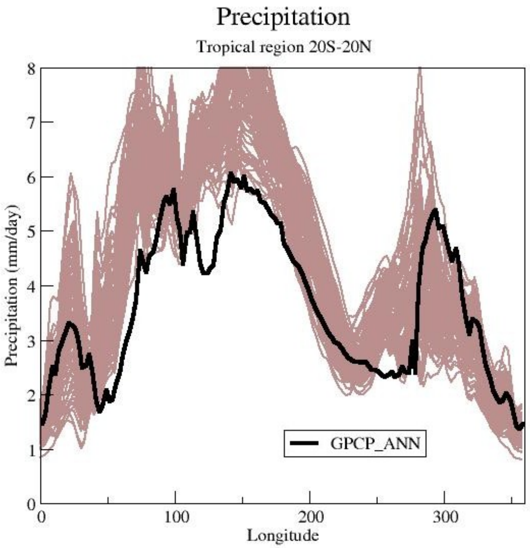
- **Calibrating CAM depends critically on the observations and metrics used to filter UQ ensembles.**
  - Using surrogate models provides an efficient way to quantify the assumptions made during ensemble filtering. (e.g. What observations should we use? How should we combine the data and ensembles?)

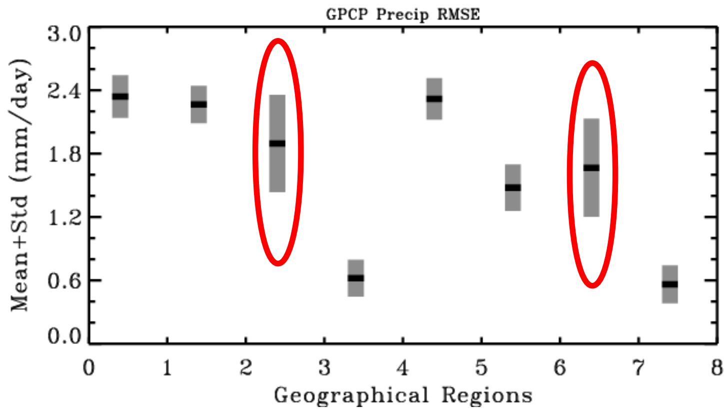
- **Right now, we are performing:**
  - exploratory UQ studies of CAM + CICE + SOM
  - extensive calibration studies with the CAM AMIP ensembles

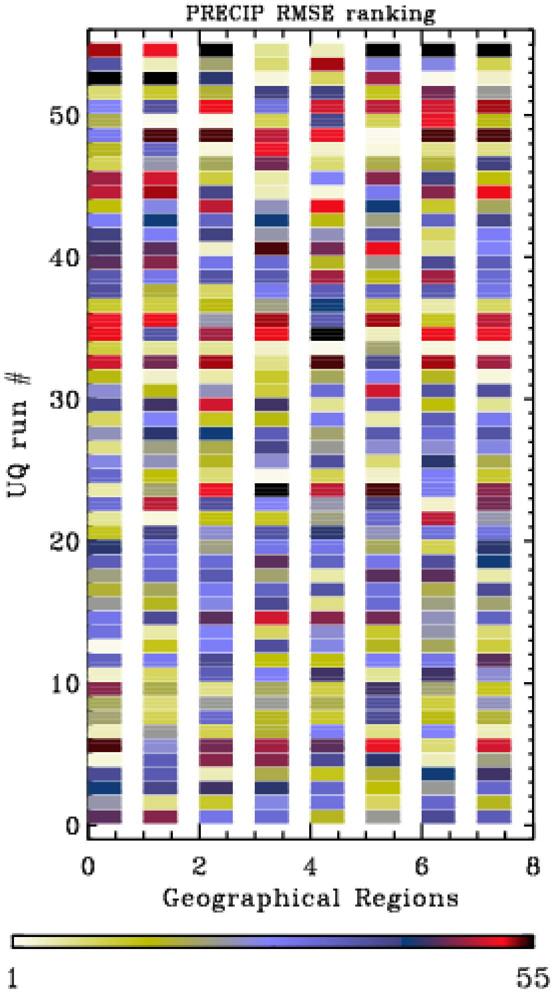
- **Soon we will combine the above for forward UQ propagation for equilibrium climate sensitivity**

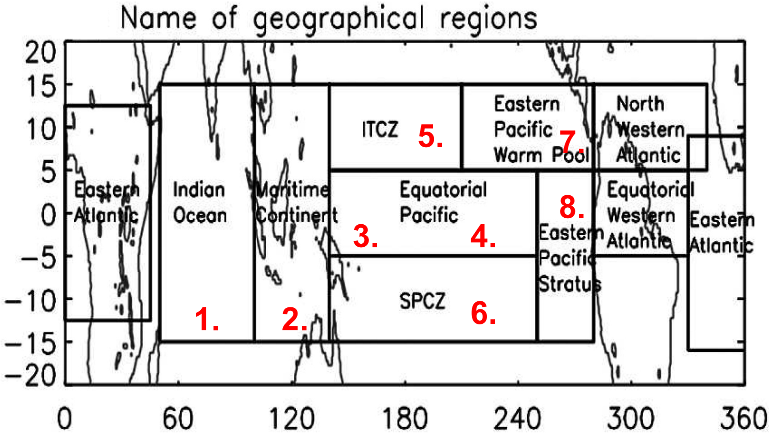
Extra slides from YZ

Regional Analysis of Tropical Precipitation





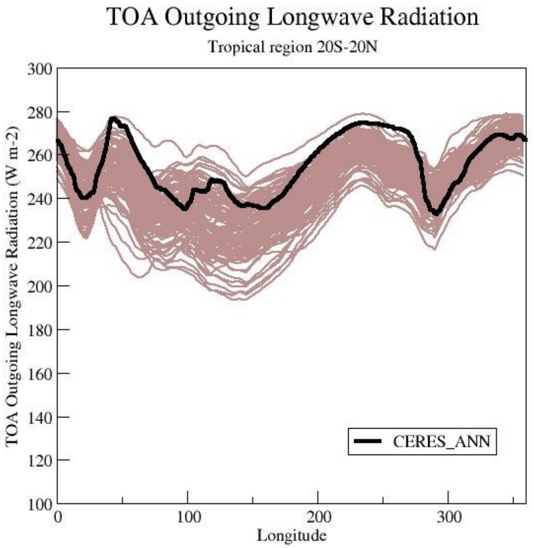


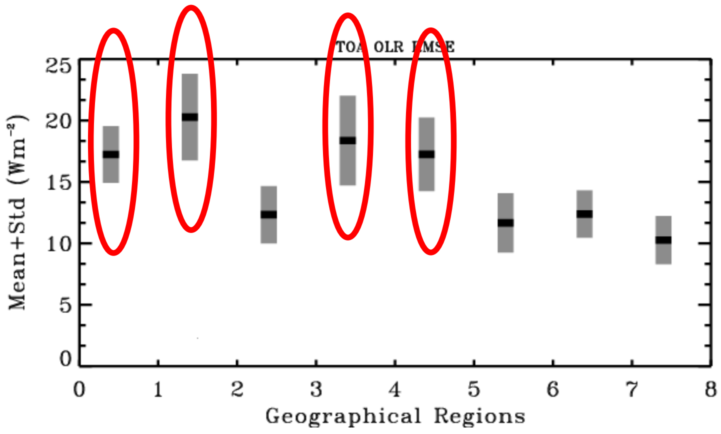


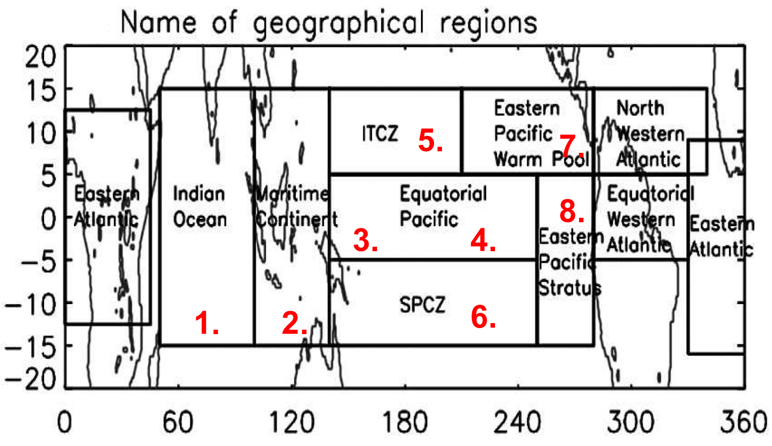
Beginning to conduct regional UQ analysis

- **identify regions important for different physical processes**
- **use PCMDI metrics to better constrain CAM ensembles**

Regional Analysis of Tropical OLR

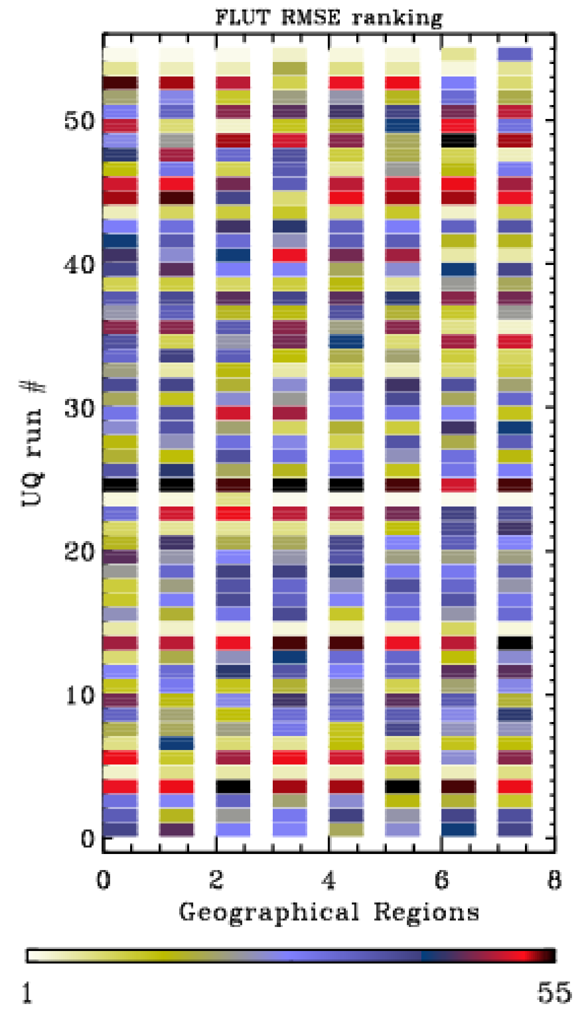






Beginning to conduct regional UQ analysis

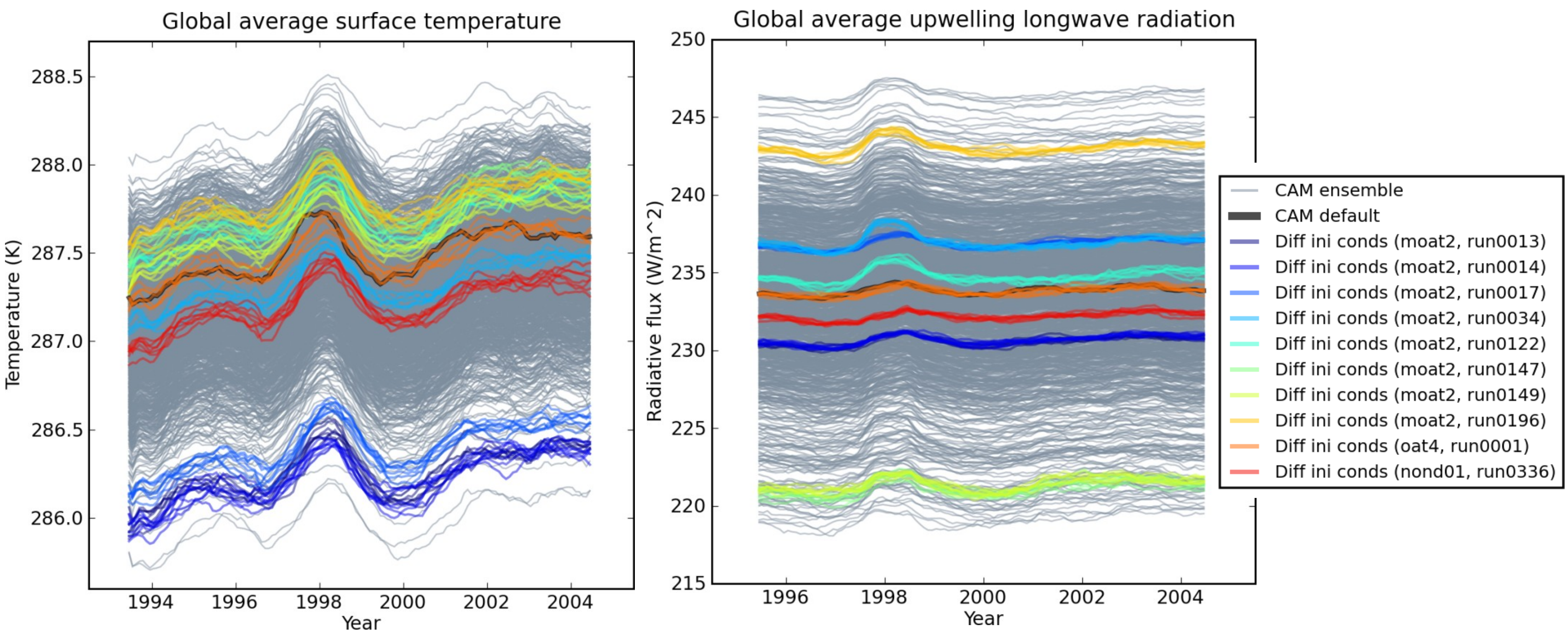
- **identify regions important for different physical processes**
- **use PCMDI metrics to better constrain CAM ensembles**



Extra slides for Node 587

Initial Condition Study (uq_ics01, pre-node 587)

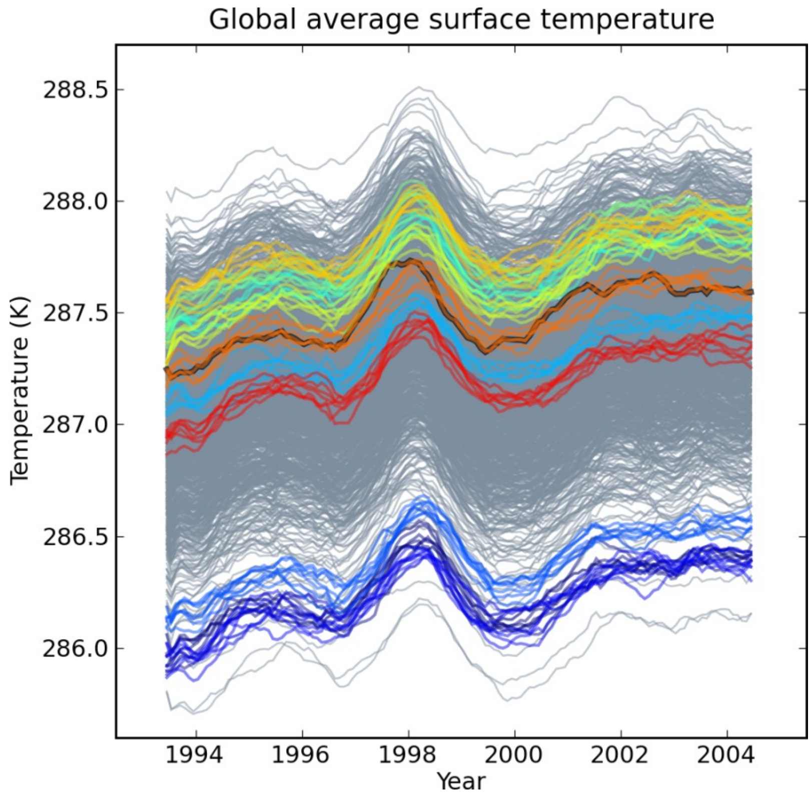
Time series of monthly mean, global averages

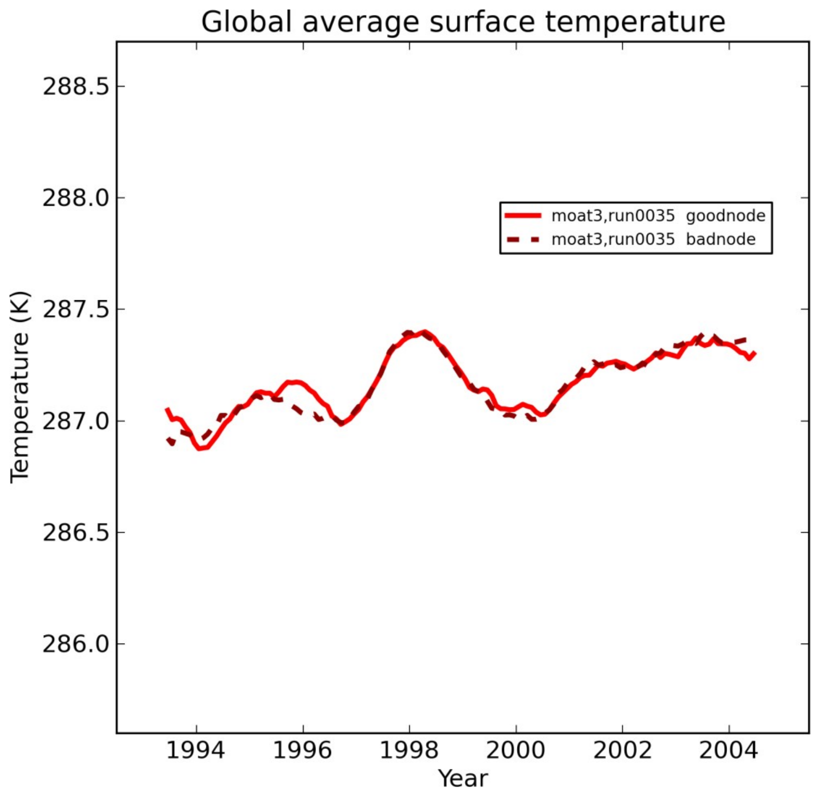


- Initial condition variations are substantially smaller than parameter variations
- Initial condition variability does not increase with time

Node 587 – Changes in time

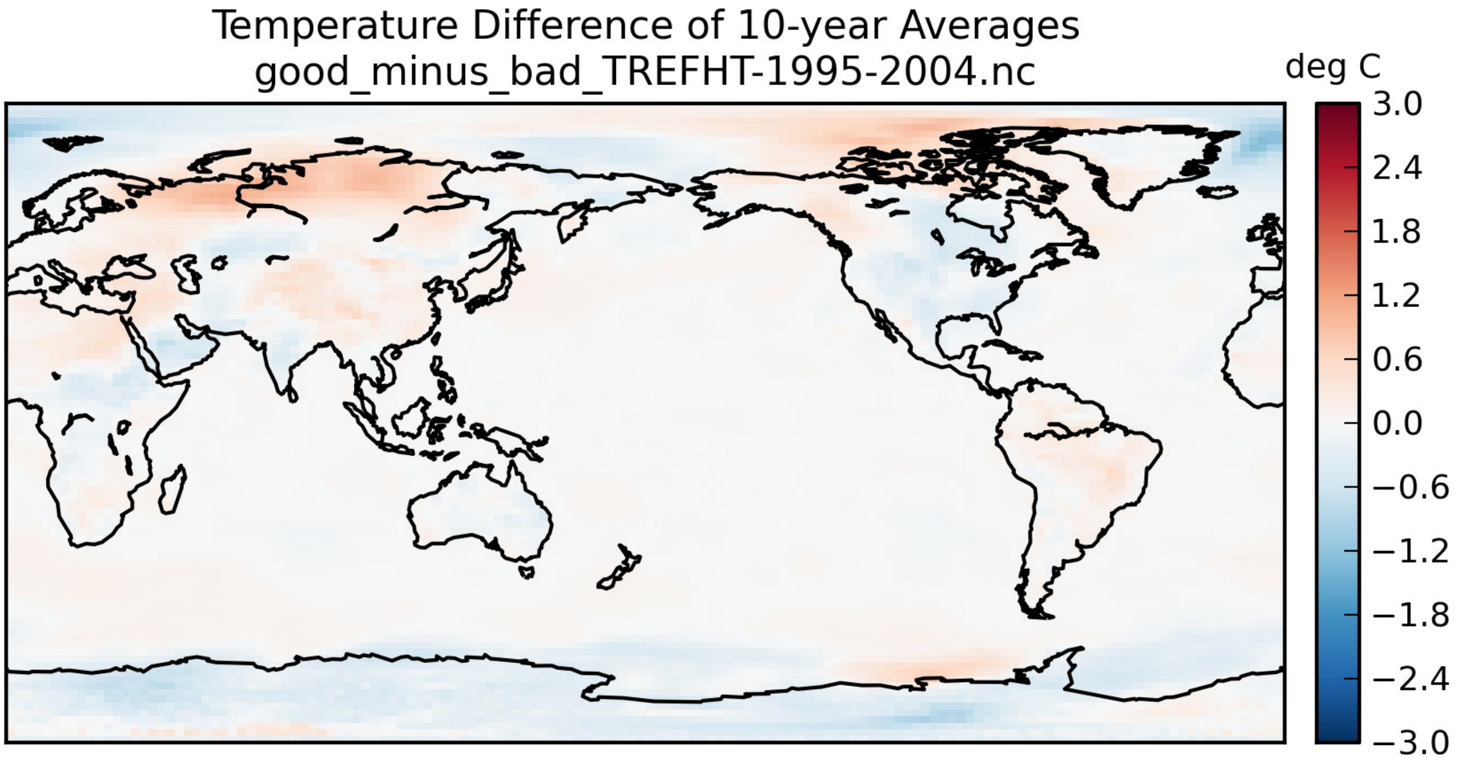
Time series of monthly mean, global averages





- **Node 587 difference is substantially smaller than parameter variations**
- **Node 587 difference does not increase with time**

Node 587 – Changes in space



- **Extreme values of +1 deg C and -1 deg C**
- **Centered at 0 deg C**
- **Global average is approximately zero (consistent with previous slide, even though the order of averaging operations is different)**

Compiler Optimization Study (uq_prec03)

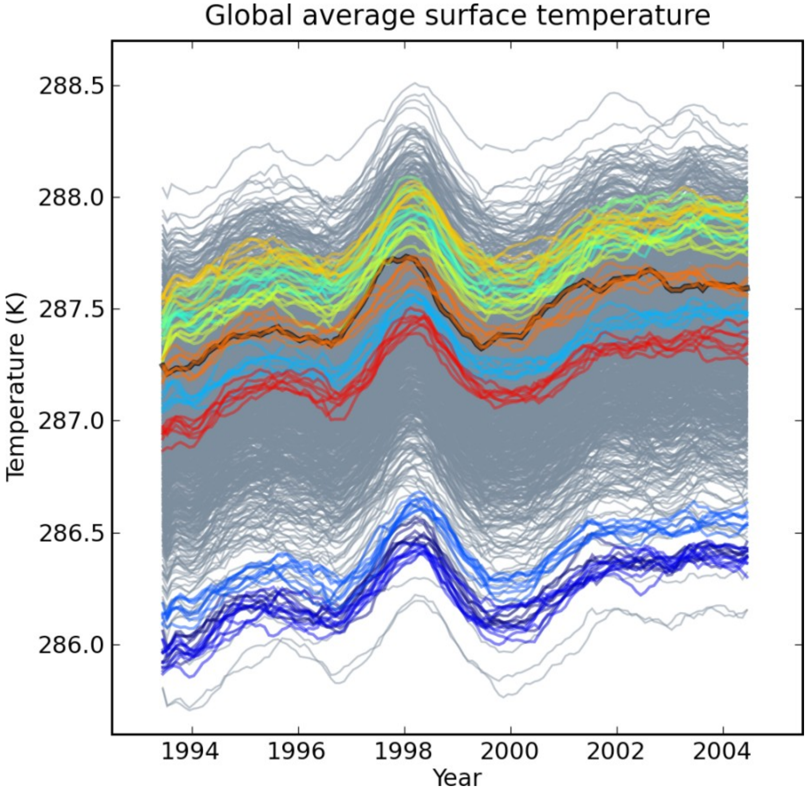
- **CAM was compiled with the pgi compiler at two optimization levels (`-O1` and `-O2`)**
- **Four cases were selected**
  - **`oat6_run0001, moat3_run0013, moat3_run0042, moat3_run0139`**
  - **Did not match cases used in `uq_ics01` study because different versions of CAM and different numbers of parameters were used**

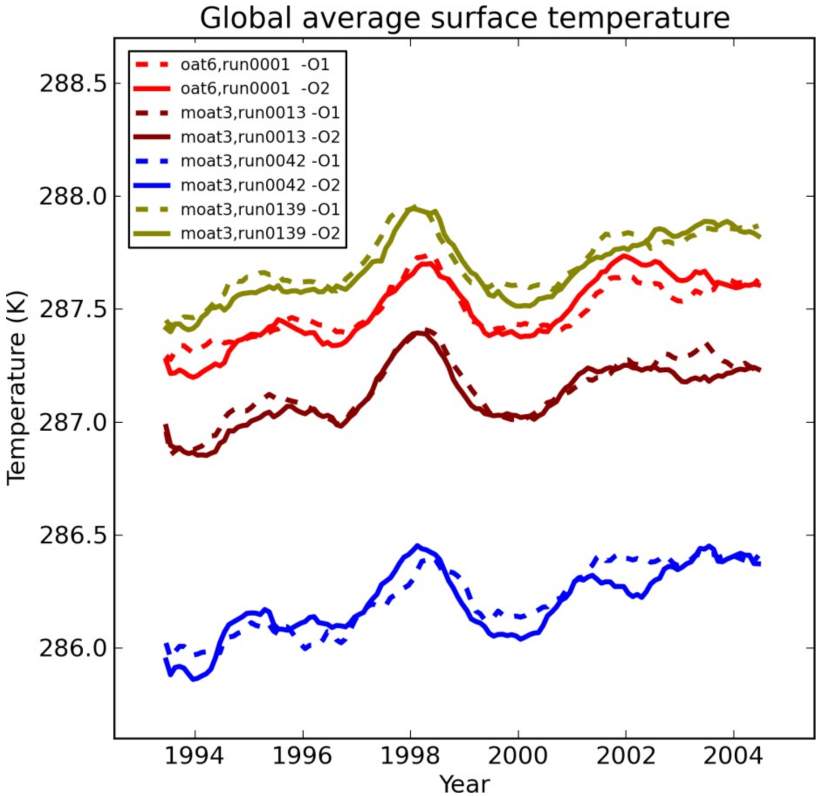
Case	Optimization	Run
oat6, run0001	-O1	1
moat3, run0013	-O1	2
moat3, run0042	-O1	3
moat3, run0139	-O1	4
oat6, run0001	-O2	5
moat3, run0013	-O2	6
moat3, run0042	-O2	7
moat3, run0139	-O2	8

1. **Sensitivity to optimization changes: R6 – R2, R7 – R3, R8 – R4**
2. **Sensitivity to parameter changes: R5 – R6, R5 – R7, R5 – R8**

Compiler Optimization Study (uq_prec03)

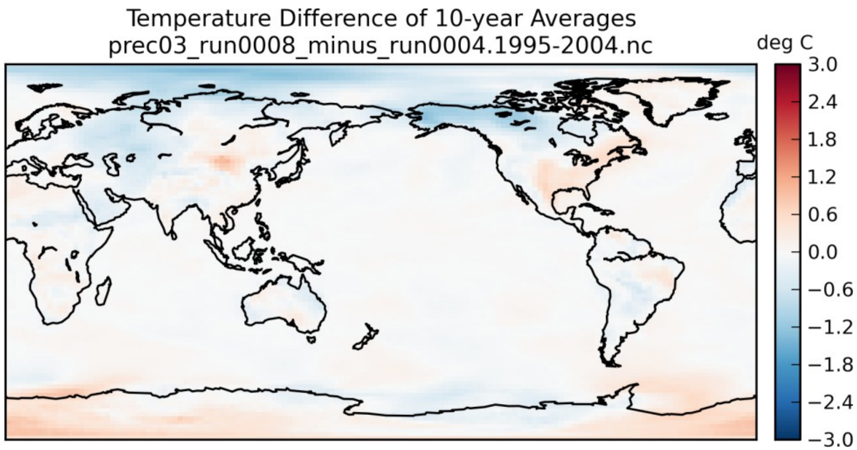
Time series of monthly mean, global averages

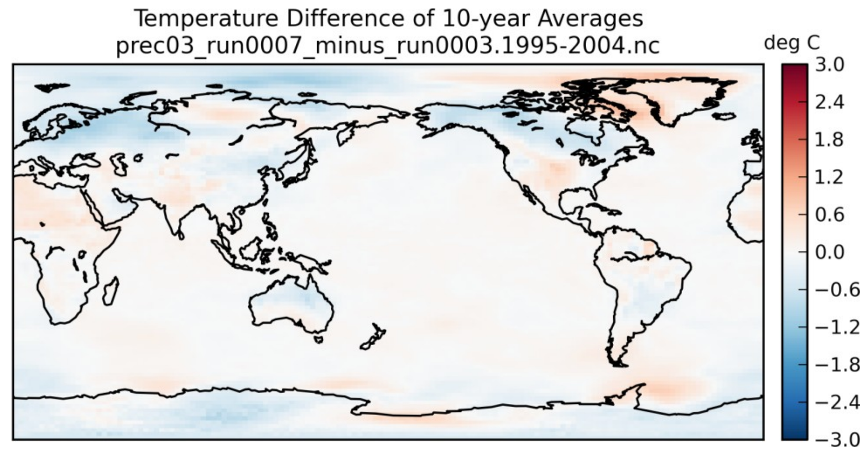


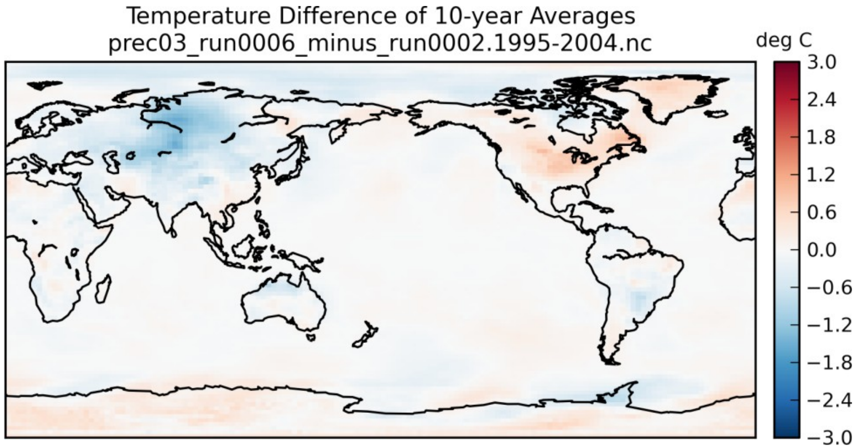


- **Optimization differences are substantially smaller than parameter variations**
- **Optimization differences do not increase with time**
- **Optimization differences are on par with initial condition differences**

Compiler Optimization Study (uq_prec03)



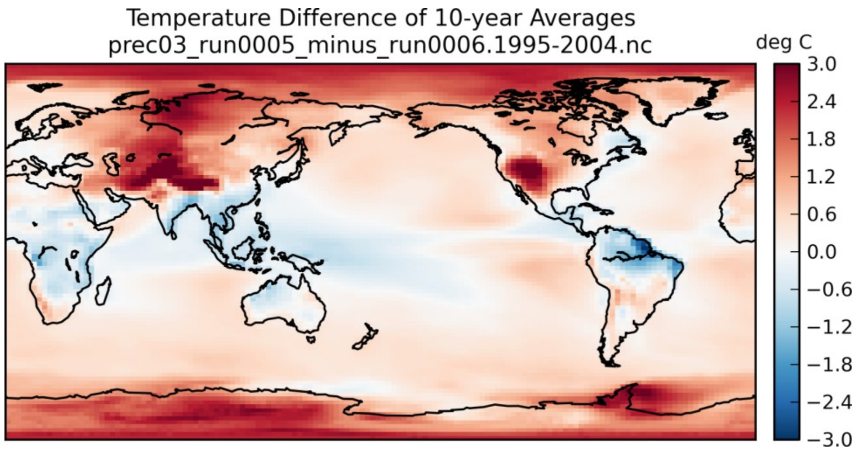


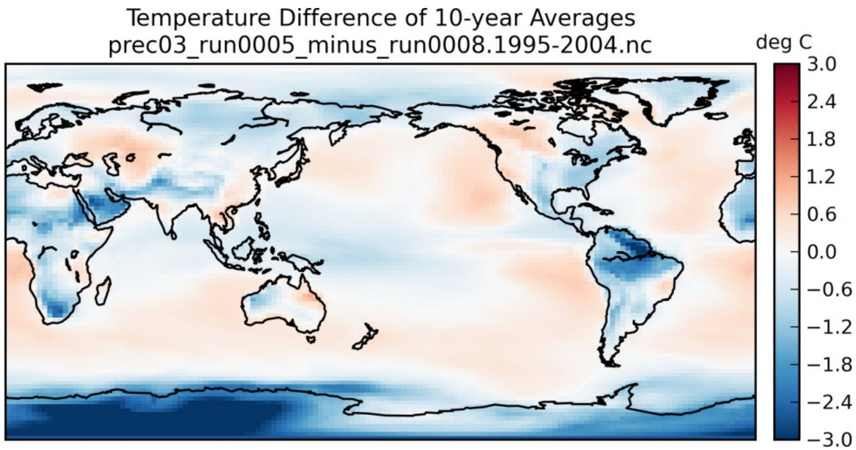


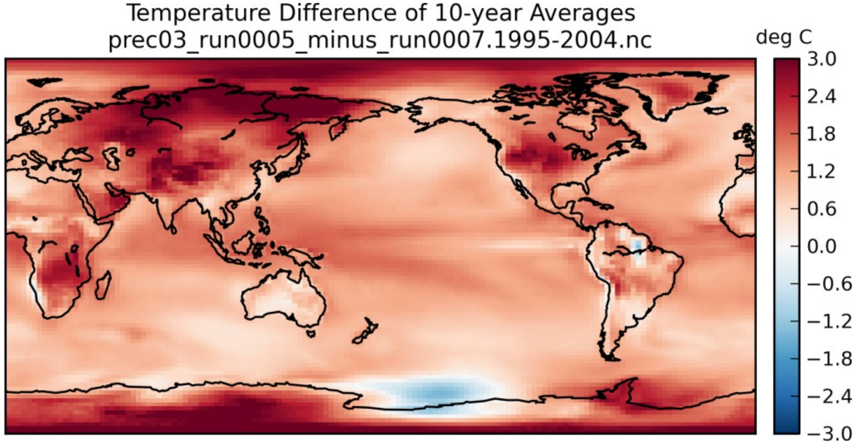
1. **Sensitivity to optimization changes: R6 – R2, R7 – R3, R8 – R4**
2. **Sensitivity to parameter changes: R5 – R6, R5 – R7, R5 – R8**

- **Qualitatively similar to Node 587 differences (extreme values of about +/- 1 deg C, centered at 0 deg C, global average is approximately zero)**

Compiler Optimization Study (uq_prec03)







1. Sensitivity to optimization changes: R6 – R2, R7 – R3, R8 – R4

2. Sensitivity to parameter changes: R5 – R6, R5 – R7, R5 – R8

- **Not similar to Node 587 and Optimization differences (larger extreme values, global average is not approximately zero)**