

SANDIA REPORT

SAND2011-8310

Unlimited Release

Printed November 2011

Uncertainty Assessment in Atmospheric Component of Climate Models

Laura P. Swiler, Timothy M. Wildey, and Keith Dalbey

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent 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, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: <http://www.osti.gov/bridge>

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: <http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online>



SAND2011-8310
Unlimited Release
Printed November 2011

Uncertainty Assessment in Atmospheric Component of Climate Models

Laura P. Swiler, Timothy M. Wildey, and Keith Dalbey
Optimization and Uncertainty Estimation Department
Sandia National Laboratories
PO Box 5800
Albuquerque, NM 87185-1318

Abstract

This report summarizes the work focusing on uncertainty analysis in atmosphere models from July-October 2011 under the Climate Science for a Sustainable Energy Future (CSSEF) project. The work had several objectives: the development of surrogate models (including kriging and stochastic expansion), sensitivity analysis and the identification of important input parameters, uncertainty quantification, and some initial calibration. This report documents the progress to date.

Acknowledgments

We thank the Climate Science for a Sustainable Energy Future (CSSEF) for supporting this work. CSSEF is a program in the Department of Energy's Office of Science, under the Biological and Environmental Research Program. We also thank Mark Taylor, Michael Levy, Oksana Guba, Bert Debusschere, Habib Najm, Jaideep Ray, Khachik Sargsyan, Cosmin Safta, and Robert Berry for their helpful comments and interactions.

Contents

1. Introduction.....	6
2. Model Description.....	7
3. Sensitivity Analysis.....	8
3.1 CORRELATION ANALYSIS.....	8
3.1.1 Description.....	8
3.1.2 Results.....	10
3.2 VARIANCE-BASED DECOMPOSITION.....	14
3.2.1 Description.....	14
4. Surrogate Models.....	16
4.1 GAUSSIAN PROCESS MODELS.....	16
4.2 POLYNOMIAL CHAOS EXPANSION.....	17
4.3 STOCHASTIC COLLOCATION.....	18
5. Sparse Grid.....	18
5.1 DESCRIPTION.....	18
5.2 RESULTS.....	19
5.2.1 Comparison of Cumulative Density Functions.....	21
5.2.2 Decay of Polynomial Chaos Coefficients.....	23
5.2.3 Sensitivity Analysis.....	25
6. Calibration.....	27
6.1 PARETO OPTIMIZATION AND THE MOGA ALGORITHM.....	28
6.2 RESULTS.....	31
7. Summary and Next Steps.....	37
References.....	38
DISTRIBUTION.....	40

1. Introduction

The Climate Science for a Sustainable Energy Future (CSSEF) program started in July 2011 as part of a new initiative in the Department of Energy's Office of Science, under the Biological and Environmental Research Program. The program has an overall goal to:

Transform the climate model development and testing process and thereby accelerate the development of the Community Earth System Model's sixth-generation version, CESM3, scheduled to be released for predictive simulation in the 5 to 10 year time frame.

Four research themes are addressed in the project:

1. A focused effort for converting observational data sets into specialized, multi- variable data sets for model testing and improvement.
2. Development of model development test beds in which model components (atmosphere, land, ocean, and sea ice) and sub-models can be rapidly prototyped and evaluated.
3. Research to enhance numerical methods and computational science research focused on enabling climate models that use future computing architecture.
4. Research to enhance efforts in uncertainty quantification for climate model simulations and predictions.[CSSEF Proposal, 2010]

This work focuses on research theme #4 above. With respect to the uncertainty quantification (UQ) thrust, we identified several objectives for the first year:

1. Implement and test production-ready UQ tools in collaboration with test beds
2. Begin initial advancement of adaptive sampling methods for ensemble construction
3. Begin initial advancement of surrogate models for high-dimensional input/output data
4. Research an efficient, scalable Bayesian calibration framework in all test beds
5. Research AD-based optimization for calibration in the land test bed
6. Identification of datasets for climate data UQ and evaluate data UQ methods

The work addressing these objectives is being performed by several DOE laboratories, including Argonne, Lawrence Berkeley, Lawrence Livermore, Pacific Northwest, Los Alamos, and Sandia. The Sandia UQ effort is further decomposed into UQ work supporting the atmosphere component, UQ work supporting the land component, and "cross-cutting" UQ work which supports all of the components.

This report only documents the UQ work at Sandia supporting the atmosphere component. Given the short time-frame of the FY2011 funding, we were asked to develop a set of bi-weekly goals. The July 2011 version of these goals is shown below. Note that CSSEF program is just beginning, and is very multi-disciplinary and multi-laboratory. We are starting to develop collaborations across the laboratories. As we work together, the work plan continues to evolve.

Task Planning July 2011: Explore surrogate models and calibration techniques based on CAM4 ensemble and apply to CAM5 (SNL)

8/1/11: Identify global sensitivity to each parameter based on sensitivity analysis. Identify range of outputs given ranges on inputs.

8/15/11. Complete runs as sparse-grid study for surrogate development.

9/1/11. Complete surrogate models of climate responses as a function of inputs.

9/15/11. Identification of parameters which provide a “good match” to the data according to several metrics.

10/3/11. Perform surrogate model construction and sensitivity analysis based on any CAM5 ensemble data sets available from LLNL, PNNL, and SNL.

10/17/11. Identify differences in sensitivities between CAM4 and CAM5. Set up CAM5 sparse grid study.

10/31/11. Paper documenting the results, evaluation, and comparison of the methods.

The outline of the rest of the report is as follows: Section 2 describes the CAM4 model, Section 3 documents sensitivity analysis methods and results, Section 4 describes surrogate models, Section 5 documents the sparse grid and polynomial chaos results, Section 6 presents some preliminary calibration results, and Section 7 presents a status summary and ideas for next steps.

2. Model Description

We performed sensitivity analysis on CCSM with the CAM4 atmosphere and 2-degree resolution with the F-AMIP configuration. This particular configuration uses the fully active Community Atmosphere Model (CAM), the Community Land Model (CLM), and the CICE model for sea ice. The ocean model is not fully active and uses observed sea surface temperatures. Each simulation runs for 14 years from January 1988 through December 2001, and results were collected from March 1990 through February 2001.

We generated ensembles based on Latin Hypercube sampling (LHS). We identified six input parameters and ten quantities of interest. These were identified in the 2008 paper by Charles Jackson et al. titled "Error Reduction and Convergence in Climate Prediction" in the Journal of Climate.

The input parameters varied for CAM4 are displayed in Table 1:

Parameter	Description	Default Value	Range
RHMINL	Low cloud critical relative humidity	0.91	[0.8, 0.95]
RHMINH	High cloud critical relative humidity	0.8	[0.6, 0.9]
ALFA	Initial cloud downdraft mass flux	0.1	[0.05, 0.6]
TAU	Consumption rate of CAPE	3.6E2	[1.8E2, 2.88E3]
KE	Environmental air entrainment rate	3.5E-3	[3.0E-3, 6.0E-3]
C0	Precipitation efficiency	1.0E-6	[3.0E-6, 10.0E-6]

Table 1: Input parameters examined in CAM4 study

T

The output quantities of interest are shown in Table 2:

Output metric	Description
TREFHT	Reference Height Temperature
T	Temperature
U	Zonal Wind
PS	Surface Pressure
RELHUM	Relative Humidity
LHFLX	Surface latent heat flux
LWCF	Longwave cloud forcing
SWCF	Shortwave Cloud forcing
PRECT	Total precipitation rate
RADBAL	Radiative Balance

Table 2: Output quantities examined in CAM4 study

3. Sensitivity Analysis

To perform sensitivity analysis, we used two approaches: correlation analysis and variance-based decomposition. These are described below along with results.

3.1 Correlation Analysis

3.1.1 Description

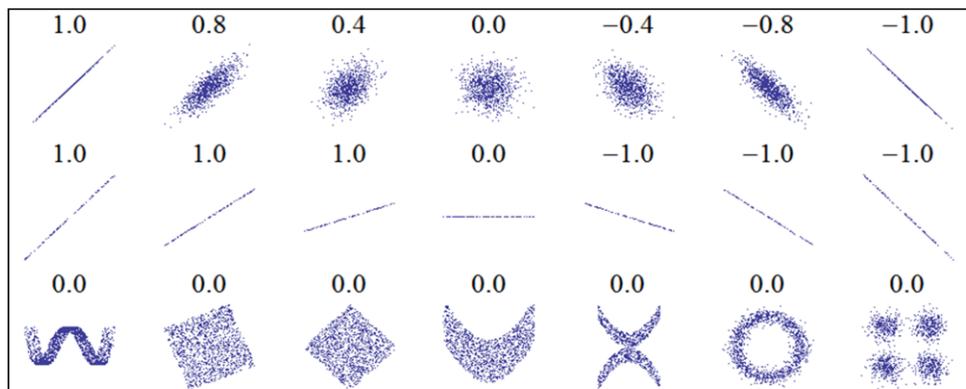
Correlation refers to a statistical relationship between two random variables or two sets of data. In analysis of computer experiments, where an ensemble of simulation runs have been performed according to some type of experimental design, we have a set of results. The convention is to have each “sample” or run of the simulation be written on a separate row. For example, if N simulation runs were performed, with D inputs and P outputs, the resulting ensemble matrix would be of

dimension $N*(D+P)$. In this situation, we can perform a correlation analysis on the entire matrix. However, often the correlations between inputs and inputs are not interesting, especially if the sample design has been constructed so that the inputs are independent and thus the correlations between inputs are near zero. Likewise, the correlations between outputs and outputs may not be interesting, except in the case where some of the outputs are very strongly correlated and thus perhaps one can reduce the analysis by only focusing on a subset of outputs. The main focus of correlation analysis of computer experiments is the correlation between inputs and outputs.

There are several types of correlations that can be calculated: simple, rank, and partial. Simple correlation measures the strength and direction of a linear relationship between variables. Simple correlation refers to correlations performed on the actual input and output data, calculated by the Pearson correlation coefficient. For example, the Pearson correlation between input X and output Y is given by $\rho(X,Y)$ [Larsen and Marx]:

$$\rho_{(X,Y)} = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X) \cdot \text{var}(Y)}} = \frac{E[(X - \bar{x})(Y - \bar{y})]}{\sigma_X \sigma_Y} \approx \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}}$$

The Pearson correlation is +1 in the case of a perfect positive (increasing) linear relationship, -1 in the case of a perfect decreasing (negative) linear relationship, and some value between -1 and 1 in all other cases. A simple correlation near zero means there is less of a relationship between the variables: they are close to being uncorrelated. Figure 3.1 shows some example correlation patterns and corresponding correlation coefficients. Note that if two variables are independent, they will have zero correlation but the converse is not true: they may have zero or near-zero correlation but show a strong type of relationship (e.g. see the last row of Figure 3.1).



Source: <http://en.wikipedia.org/wiki/Correlation>

Figure 3.1: Example Correlation Relationships

Rank correlations refer to correlations performed on the ranks of the data. Ranks are obtained by replacing the actual data by the ranked values, which are obtained by ordering the data in ascending order. For example, the smallest value in a set of input samples would be given a rank 1, the next

smallest value a rank 2, etc. Rank correlations are useful when some of the inputs and outputs differ greatly in magnitude: then it is easier to compare if the smallest ranked input sample is correlated with the smallest ranked output, for example. A rank correlation coefficient is also called a Spearman correlation. Partial correlation coefficients are similar to simple correlations, but a partial correlation coefficient between two variables measures their correlation while adjusting for the effects of the other variables. For example, if one has a problem with two highly correlated inputs and one output, the correlation of the second input and the output may be very low after accounting for the effect of the first input.

3.1.2 Results

We performed simple correlation analysis using Pearson correlation coefficients on 1019 samples generated from CAM4. Note that these samples were generated using a Latin Hypercube sampling strategy called Binning Optimal Symmetric LHS as explained in Section 6.1. The overall correlation table is shown in Table 3. Note that Table 3 presents the correlation results for output averages computed over a band +/- 30° around the equator.

	RHMINL	RHMINH	ALFA	TAU	CZERO	KE
TREFHT	0.33	-0.06	-0.05	0.83	-0.04	0.04
T	0.58	-0.46	-0.35	-0.39	-0.05	0.19
U	-0.17	-0.37	0.07	0.82	-0.01	0.02
PS	0.29	-0.10	0.01	0.62	-0.04	0.03
RELHUM	0.05	0.58	-0.20	-0.74	-0.03	0.15
LHFLX	-0.30	0.31	0.10	0.82	0.01	-0.17
LWCF	-0.23	-0.72	-0.14	-0.59	-0.02	0.12
SWCF	0.92	0.31	0.04	0.21	-0.01	-0.03
PRECT	-0.40	0.38	0.05	0.74	0.03	-0.22
RADBAL	0.97	0.16	-0.03	-0.05	-0.02	0.01

Table 3: Correlation Analysis for CAM4: Results calculated over +/-30° Equatorial Band
Rows are Outputs, Columns are Inputs

In Table 3, a yellow cell represents a correlation coefficient whose absolute value is between 0.2 and 0.5. A red cell represents a correlation coefficient whose absolute value is between 0.5 and 1.0. These correspond to correlations that are considered significant (yellow) and strongly significant (red). To test for significance, we can use the same t-test that is used to detect if the slope coefficient in a simple regression model is nonzero. For this large sample size, one can reject the null hypothesis that the correlation coefficient is zero even for fairly small correlation values because of the large number of samples. A correlation coefficient of 0.2 or greater does lead to a statement that the null hypothesis of zero correlation is rejected with high confidence ($\alpha=0.001$). In this data set, there were very low correlations (near zero) amongst all of the inputs and so we did not show these correlations in Table 3. The low correlation between inputs is to be expected since the samples have been designed so that the inputs are independent. We did see correlations amongst the outputs, but these were not included for space reasons.

Scatterplots of the samples used to create the correlations in Table 3 are shown in Figure 3.2. Note that the scatterplots show the correlation relationships in Table 3. For example, we see the strong positive correlation of RHMIND and RADBAL (lower left cell), with a correlation coefficient of 0.97, and we see the strong correlations between TAU and many of the outputs. We also note that CZERO is not strongly correlated with any output.

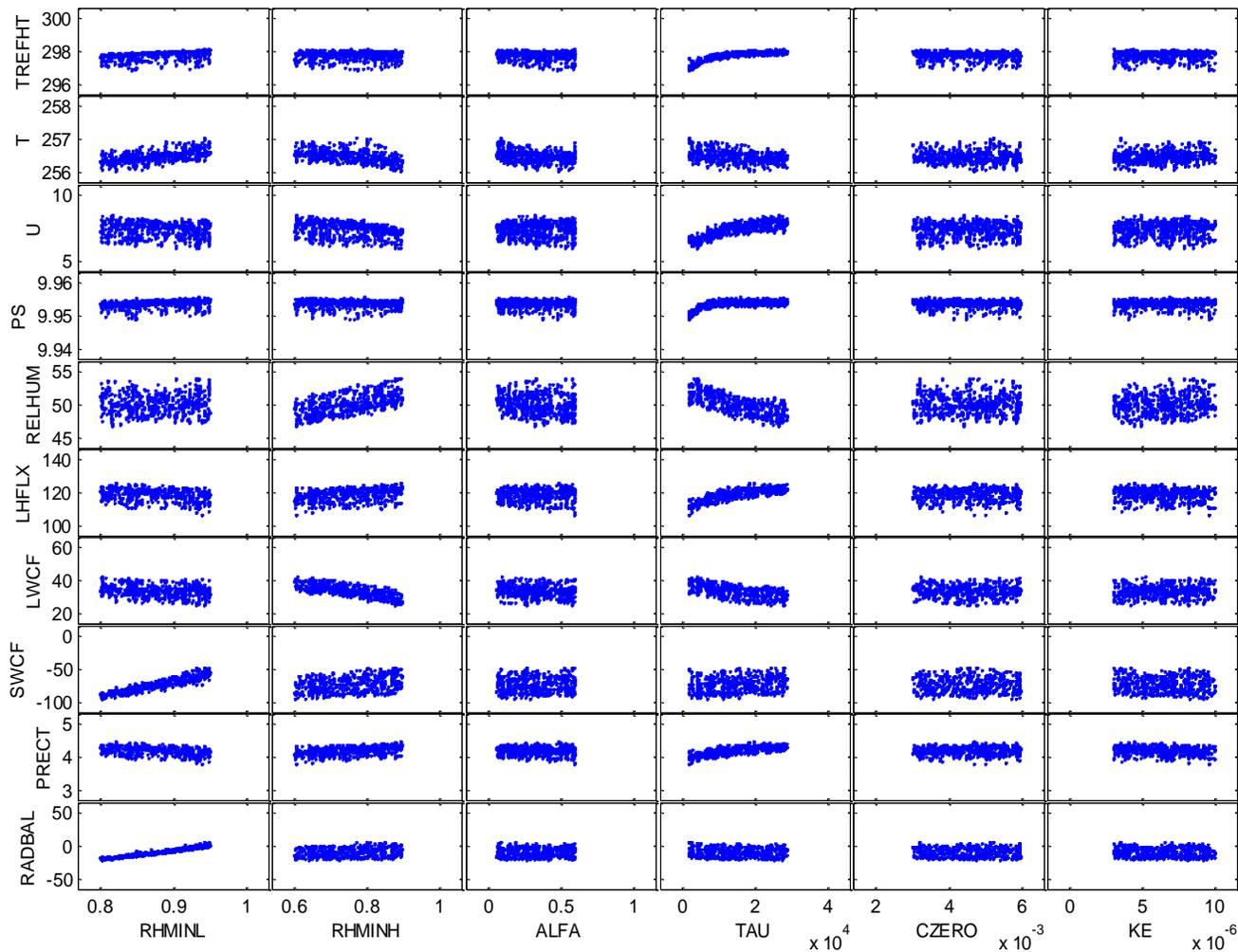


Figure 3.2: Scatterplot of CAM4 Inputs (x-axis) and Outputs (y-axis)

We performed the same analysis, but restricting the annual responses to be calculated over the Southwest region of the United States instead of the $\pm 30^\circ$ equatorial band. The correlation results are shown in Table 4.

	RHMINL	RHMINH	ALFA	TAU	CZERO	KE
TREFHT	0.54	0.25	-0.30	-0.53	-0.01	0.17
T	0.51	-0.10	-0.22	-0.70	-0.04	0.20
U	0.05	0.04	-0.18	-0.25	-0.07	-0.05
PS	-0.03	-0.55	-0.21	-0.61	0.02	0.24
RELHUM	-0.24	0.80	-0.04	0.09	-0.05	0.27
LHFLX	-0.32	0.01	0.18	0.47	0.01	-0.44
LWCF	-0.14	-0.94	-0.03	-0.05	-0.07	0.20
SWCF	0.36	0.86	-0.14	-0.22	0.01	-0.16
PRECT	-0.29	-0.05	0.23	0.58	0.01	-0.33
RADBAL	0.97	0.16	-0.03	-0.05	-0.02	0.01

Table 4: Correlation Analysis for CAM4: Results calculated over the Southwest U.S. Rows are Outputs, Columns are Inputs

Note that many of the correlations are similar between Tables 3 and 4. However, the Southwest results in Table 4 show somewhat stronger correlations of KE with several of the outputs and weaker correlations of TAU with several of the outputs.

Finally, we restricted the Southwest results to only look at these outputs for the summer month average (J-J-A). The correlations in Tables 3 and 4 are calculated over the entire year, but the averages in Table 5 show the correlations for the Southwest summer months:

	RHMINL	RHMINH	ALFA	TAU	CZERO	KE
TREFHT	0.63	0.54	-0.28	-0.23	-0.06	0.03
T	0.66	0.32	-0.24	-0.51	-0.06	0.00
U	-0.33	-0.50	-0.10	-0.64	0.01	0.15
PS	0.04	-0.49	0.15	0.55	0.02	0.25
RELHUM	-0.10	0.47	0.08	0.47	-0.08	0.32
LHFLX	-0.36	-0.03	0.12	0.27	0.04	-0.58
LWCF	-0.04	-0.86	0.03	0.22	-0.16	0.21
SWCF	0.21	0.80	-0.18	-0.40	0.02	-0.20
PRECT	-0.24	-0.03	0.15	0.41	0.03	-0.51
RADBAL	0.97	0.15	-0.03	-0.05	-0.02	0.01

Table 5: Correlation Analysis for CAM4: Results calculated over the Southwest U.S. Summer Average only. Rows are Outputs, Columns are Inputs

Note that the correlations between inputs and summer averages shown in Table 5 are similar to the correlations between inputs and annual averages shown in Table 4, but again there are some differences in the correlations and the importance of some of the input/output relationships. For example, the correlation between RHMINH and Relative Humidity (RELHUM) is significantly smaller in Table 5 (.47) than it is in Table 4 (0.8).

Finally, we looked at the correlations obtained when we ran surrogate models for the CAM4 outputs. There are several types of surrogate models (also called emulators or response surface

models) that can be used: neural networks, splines, polynomial regression, etc. We used a multi-variate adaptive regression spline (MARS) as a surrogate model for each output. Another type of surrogate that we investigated is called a Gaussian process model; this is described in Section 4.1. The MARS implementation we used is documented in the DAKOTA manual (Adams et al.)

For the purposes of this discussion, we just want to demonstrate that the correlations obtained when using surrogate models are similar to the correlations we obtained from the original CAM4 runs as shown in Table 3. Table 6 shows a similar result, but this time the correlations are based on 1000 samples of surrogate models of the outputs. Comparing Table 3 and Table 6, we see that the surrogates generally are able to capture the strong correlations. For example, the correlation between RHMIND and RADBAL is 0.96 in Table 6 vs. 0.97 in Table 3. Similarly, the correlation between TAU and TREFHT is identical (0.83) in both tables. There are some differences, primarily in the variables that are of lesser importance. The MARS surrogate does not pick up any significant correlations between ALFA, CZERO, or KE and any of the outputs. However, Table 3 indicates two: a correlation between ALFA and T of -0.35 and a correlation between KE and PRECT of -0.22. This indicates that the surrogates may not capture the less significant relationships as accurately. One important thing to notice is that the signs are correct: if an input and output is positively correlated in Table 3, it also is in Table 6, and similarly for negative correlations. This behavior is important for surrogates to capture correctly. We will say more about the goodness of surrogates in Sections 4 and 6. For the purposes of this discussion, we wanted to demonstrate that it is possible to perform correlation analysis on surrogates, and the signs of significant correlations are maintained along with a relative ranking.

	RHMIND	RHMINDH	ALFA	TAU	CZERO	KE
TREFHT	0.43	-0.12	0.02	0.83	-0.03	0.01
T	0.80	-0.44	-0.01	-0.35	-0.08	0.00
U	-0.21	-0.30	0.01	0.85	0.02	0.01
PS	0.24	-0.32	0.01	0.73	-0.03	0.02
RELHUM	0.10	0.74	-0.01	-0.56	-0.15	0.02
LHFLX	-0.47	0.45	0.01	0.72	0.02	-0.01
LWCF	-0.21	-0.83	-0.02	-0.44	-0.07	-0.01
SWCF	0.91	0.27	-0.01	0.02	0.19	0.01
PRECT	-0.65	0.55	0.01	0.44	0.07	-0.01
RADBAL	0.96	0.14	0.01	-0.20	-0.02	0.01

**Table 6: Correlation Analysis for CAM4 based on Surrogates:
Results calculated over +/-30° Equatorial Band
Rows are Outputs, Columns are Inputs**

3.2 Variance-based Decomposition

3.2.1 Description

The correlation coefficients described in Section 3.1 only detect linearity or monotonicity. In contrast, the variance-based indices (referred to as Sobol' indices) are not limited in this way. The variance-based indices identify the fraction of the variance in the output that can be attributed to an individual variable alone or with interaction effects [Sobol', Saltelli et al. 2000]. There are two classes of variance-based sensitivity indices: main effects and total effects. The main effects indices, S_i , identify the fraction of uncertainty in the output Y attributed to input X_i alone. The total effects indices, T_i , correspond to the fraction of the uncertainty in output Y attributed to X_i and its interactions with other variables. These sensitivity indices are represented as:

$$S_i = \frac{V(E(Y | X_i))}{V(Y)} \quad (1)$$

$$T_i = \frac{E[\text{Var}(Y | X_{-i})]}{V(Y)} \quad (2)$$

where $\text{Var}(\cdot)$ is the variance, $E(\cdot)$ is the expected value, and $E(Y|X_i)$ is the expected value of Y conditioned on X_i . $\text{Var}(Y|X_{-i})$ is the variance of Y conditioned on all the inputs except X_i . These indices involve multidimensional integrals that, in practice, are evaluated approximately. Note that S_i varies between 0 and 1. Values close to one mean that the uncertainty in variable X_i is very significant in contributing to the uncertainty in output Y . The sum of S_i over all variables i must equal to one. However, there are not the same restrictions on T_i . The values of T_i are greater than or equal to zero, but are not upper-bounded by one and their sum over all variables does not add to one.

The team led by Andrea Saltelli at the European Research Commission is generally credited with popularizing the use of variance-based indices for sensitivity analysis. In the past 10-15 years, several approaches have been developed for calculating the Sobol' sensitivity indices. The recent paper by [Saltelli et al., 2010] provides a detailed comparison of sampling approaches, with some comments about the relationship between the estimators and the sampling methods used.

Ideally, a full factorial sample would be performed with m samples taken in each of d input dimensions. Then, the integrals in the Sobol' formulas can easily be calculated given $n=m^d$ samples. For example, when calculating the numerator in Eq. 1, we calculate the inner expectation term m times, each time averaging over the remaining m^{d-1} points in the other dimensions. We calculate: $E(Y|X_i = x_{im})$ for each of the m points in dimension i , then take the variance of m expected values to obtain the numerator for the main effects indices. The total effects indices are calculated in a similar manner.

The full factorial approach requires $n = m^d$ samples, which may not be practical when each sample is an evaluation of a computationally costly function. Typically, the cost is reduced by sampling the inputs using Latin Hypercube or quasi-Monte Carlo sampling, rather than considering all possible combinations of input values. We generate two independent sets of samples of size n ; in each set all the d inputs are varied. Then, we create d more sets of samples of size n by taking a column from one of the original two sample sets and replacing it by the same column in the other sample set. This column swap-out procedure is described in [Saltelli, 2004]. The total number of samples is $(2+d)n$, which requires far fewer function evaluations than the full factorial approach in most situations.

We use a recent calculation [Saltelli et al. 2010] for the $(2+d)n$ samples that has been improved to remove bias and better capture interaction effects. The actual formulas we used are described in [Weirs et al., 2011]; we describe them here for completeness. Some notation: if we denote the original sample matrices as \mathbf{A} and \mathbf{B} , we denote by $\mathbf{A}_B^{(i)}$ the matrix \mathbf{A} except for the i^{th} column which has been taken from matrix \mathbf{B} . Similarly, $\mathbf{B}_A^{(i)}$ is the matrix \mathbf{B} except for the i^{th} column which has been taken from matrix \mathbf{A} . We define \mathbf{C} as the matrix with $2n$ rows and d columns obtained by appending \mathbf{B} to \mathbf{A} . \mathbf{C} is used in some formulas to estimate the total variance, as all rows of \mathbf{C} are independent. The mean value is denoted by $\langle \cdot \rangle$. The formulas to calculate the indices are given below:

$$S_i = \frac{\frac{1}{n} \sum_{j=1}^n f(\mathbf{A})_j \left(f(\mathbf{B}_A^{(i)})_j - f(\mathbf{B})_j \right)}{\frac{1}{2n} \sum_{j=1}^{2n} f(\mathbf{C})_j f(\mathbf{C})_j - \langle f(\mathbf{C}) \rangle^2},$$

$$T_i = \frac{\frac{1}{2n} \sum_{j=1}^n \left(f(\mathbf{A})_j - f(\mathbf{A}_B^{(i)})_j \right)^2}{\frac{1}{2n} \sum_{j=1}^{2n} f(\mathbf{C})_j f(\mathbf{C})_j - \langle f(\mathbf{C}) \rangle^2}.$$

Finally, we wish to mention that these sensitivity indices may be calculated when stochastic expansion methods such as polynomial chaos or stochastic collocation are used to propagate the uncertainty from inputs to outputs instead of sampling methods. When using stochastic expansion methods, the HDMF (high dimensional model representation) may be exploited to analytically obtain the sensitivity indices. That is, the sensitivity indices S_i and T_i can be calculated as analytic functions of the coefficients of the expansion. This is a very nice property, since one does not have to take additional samples beyond the ones used to construct the expansion initially. The calculations of the sensitivity indices based on polynomial chaos are derived in [Sudret, 2008]; the sensitivity indices based on stochastic collocation are derived in [Tang et al., 2010]. We present the results of variance-based decomposition using polynomial chaos and stochastic collocation in Section 5.

4. Surrogate Models

For this project, we looked at two classes of surrogate models (also referred to as meta-models or response surface models). The first class is typically constructed over a set of random sample points such as a set of Monte Carlo or LHS samples, and includes surrogates such as Gaussian process models, splines, and regression models. The second class is typically constructed over samples constructed using a particular quadrature scheme. This class includes stochastic expansion methods, specifically polynomial chaos expansions and stochastic collocation.

4.1 Gaussian Process Models

Gaussian Process models are used in response surface modeling, especially response surfaces which “emulate” complex computer codes. Gaussian processes have also been widely used for estimation and prediction in geostatistics and similar spatial statistics applications [Cressie]. The recent book by Rasmussen and Williams provides a good overview of Gaussian process models.

A Gaussian process (GP) is defined as follows: A stochastic process is a collection of random variables $\{Y(\mathbf{x}) \mid \mathbf{x} \in \mathbf{X}\}$ indexed by a set \mathbf{X} (in most cases, \mathbf{X} is \mathcal{R}^d , where d is the number of inputs). The stochastic process is defined by giving the joint probability distribution for every finite subset of variables $Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_k)$. A Gaussian process is a stochastic process for which any finite set of Y -variables has a joint multivariate Gaussian distribution. A GP is fully specified by its mean function $\mu(\mathbf{x}) = E[Y(\mathbf{x})]$ and its covariance function $C(\mathbf{x}, \mathbf{x}')$. The basic steps in using a GP are:

1. Define the mean function. The mean function can be any type of function. Often the mean is taken to be zero or a constant, but this is not necessary. A common representation, for example in a regression model, is that $y(x) = \sum_j w_j \phi_j(\mathbf{x}) = \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x})$, where $\{\phi_j\}$ is a set of fixed basis functions and \mathbf{w} is a vector of weights.
2. Define the covariance. There are many different types of covariance functions that can be used (squared exponential, Matern, cubic, etc.). At this stage, we shall focus on stationary covariance functions where $C(\mathbf{x}, \mathbf{x}')$ is a function of the distance $(\mathbf{x} - \mathbf{x}')$ and is invariant to shifts of the origin in the input space. A commonly-used covariance function is:

$$C(\mathbf{x}, \mathbf{x}') = v_o \exp \left\{ - \sum_{u=1}^d \rho_u^2 (\mathbf{x}_u - \mathbf{x}'_u)^2 \right\}$$

This covariance function involves the product of d squared-exponential covariance functions with different length-scales on each dimension. The form of this covariance function captures the idea that nearby inputs have highly correlated outputs.

3. Perform the “prediction” calculations. Given a set of n input data points $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ and a set of associated observed responses or “targets” $\{z_1, z_2, \dots, z_n\}$, we use the GP to predict the target z_{n+1} at a new input point \mathbf{x}_{n+1} . The target is usually represented as the sum of the

“true” response, y , plus an error term: $z_i = y_i + \varepsilon_i$, where ε_i is a zero mean Gaussian random variable with constant variance σ_ε^2 . If C is the $n \times n$ covariance matrix with entries $C(\mathbf{x}_i, \mathbf{x}_j)$, then the prior distribution on the targets z_i is $N(0, C)$. The distribution of the predicted term z_{n+1} is conditional on the data $\{z_1, z_2, \dots, z_n\}$. It is Gaussian with the following mean and variance:

$$E[z_{n+1} | z_1, z_2, \dots, z_n] = \mathbf{k}^T C^{-1} \mathbf{z}$$

$$\text{Var}[z_{n+1} | z_1, \dots, z_n] = C(\mathbf{x}_{n+1}, \mathbf{x}_{n+1}) - \mathbf{k}^T C^{-1} \mathbf{k}$$

where \mathbf{k} is the vector of covariances between the n known targets and the new $n+1$ data point: $\mathbf{k} = (C(\mathbf{x}_1, \mathbf{x}_{n+1}), \dots, C(\mathbf{x}_n, \mathbf{x}_{n+1}))^T$, C is the $n * n$ covariance matrix of the original data, and \mathbf{z} is the $n \times 1$ vector of target values.

The equations for the mean and variance of the predictive distribution for z_{n+1} both require the inversion of C , an $n \times n$ matrix. In general, this is a $O(n^3)$ operation. Also, the covariance matrix may be near singular. Several approaches have been developed to deal both with the ill-conditioning and with large data sets (e.g. greater than 1000 data points). – KEITH – give references.

Steps 1-3 give the general framework for defining a Gaussian process and using it for prediction. However, the length scale parameters in the covariance matrix must be calculated to perform the prediction in equations 2 and 3. There are two main approaches. One is to use maximum likelihood estimation, where one maximizes the likelihood function. This results in point estimates of the covariance parameters. The other approach is to use Monte Carlo Markov Chain (MCMC) sampling to generate posterior distributions on the hyperparameters which govern the covariance function (and the mean function). The assumption of zero mean GPs is often made, so the Bayesian updating only involves hyperparameters governing the covariance function. Since these may be quite complex, one usually still needs a MCMC sampling method to generate the posterior. We use a maximum likelihood method (more details on the correlation length bounding, treatment of the condition number, etc.)

4.2 Polynomial Chaos Expansion

Polynomial chaos is a stochastic expansion method whereby the output response is modeled as a function of the input random variables using a carefully chosen set of polynomials. These polynomials are usually chosen according the Weiner-Askey scheme that provides an orthogonal basis with respect to the probability density function for the input random variables. Orthogonal polynomials can be generated numerically for arbitrary PDF's, but this is beyond the scope of this report.

In general, the polynomial chaos expansion for a response R has the form,

$$R = a_0 B_0 + \sum_{i_1=1}^{\infty} a_{i_1} B_1(\xi_{i_1}) + \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{\infty} a_{i_1, i_2} B_2(\xi_{i_1}, \xi_{i_2}) + \dots$$

where the number of random variables and the order of the expansion are unbounded. This expression is usually written in terms of the order-based indexing,

$$R = \sum_{j=0}^{\infty} \alpha_j \Psi_j(\xi).$$

In practice, both the number of random variables and the order of the expansion are truncated yielding an expansion of the form,

$$R \approx \sum_{j=0}^P \alpha_j \Psi_j(\xi).$$

4.3 Stochastic Collocation

Similar to PCE, stochastic collocation methods construct a polynomial approximation of the output response. The key difference is that the stochastic collocation approximation is a multidimensional Lagrange interpolant based on a chosen set of collocation points. These points may be based on either tensor product grids or on the Smolyak sparse grids discussed in the next section.

5. Sparse Grid

5.1 Description

If the stochastic dimension is larger than 4 or 5, sparse grids are preferable over tensor product grids since sparse grids use a drastically reduced number of evaluation points while maintaining a high level of accuracy [Smolyak 1963, Xiu et al 2005]. Sparse grids use linear combinations of the tensor product rules with the property that only products with a small number of points are retained. An example of the reduction in the number of points versus a tensor product grid is shown in Figure 5.1.

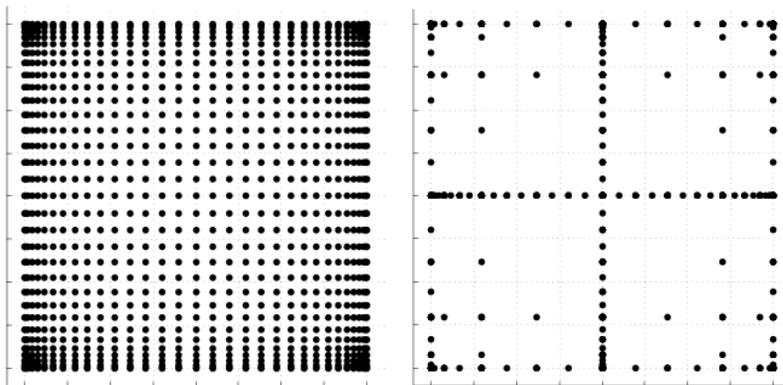


Figure 5.1: Comparison of a tensor product grid in 2D using Clenshaw-Curtis points (left) and a sparse grid (right).

Several variations of sparse grids exist depending on whether the one-dimensional quadrature rules are nested and the growth rate used. Anisotropic sparse grids can also be constructed using either a priori information regarding the significant dimensions, or using a posteriori error indicators [Nobile et al 2008].

The sparse grid is usually used as a collocation method, but the evaluation points can also be used as a quadrature rule to evaluate the integrals in a stochastic spectral construction of a PCE. Unfortunately, this approach performs much worse than stochastic collocation. Subsequently, we use an alternative algorithm to compute separate tensor polynomial chaos expansions for each of the underlying tensor quadrature grids and then sum them using the Smolyak combinatorial coefficient. In this case, the two approaches give identical polynomial representations [Constantine et al 2011].

5.2 Results

We consider the parameters in Table 1 to be uniform random variables and construct a level 2 sparse grid over the 6-dimensional parameter space. This gives a total of 97 evaluation points. We then compare the PCE with the Latin Hypercube study in Section 3.1 using 1147 evaluations. The PCE and stochastic collocation results were nearly identical in all cases, so we only report the PCE results. In Figures 5.2-5.5, we plot the means of the reference temperature and the total precipitation rate computed over the length of the simulation and over the 6-dimensional parameter space using the LHS study and the polynomial chaos expansion.

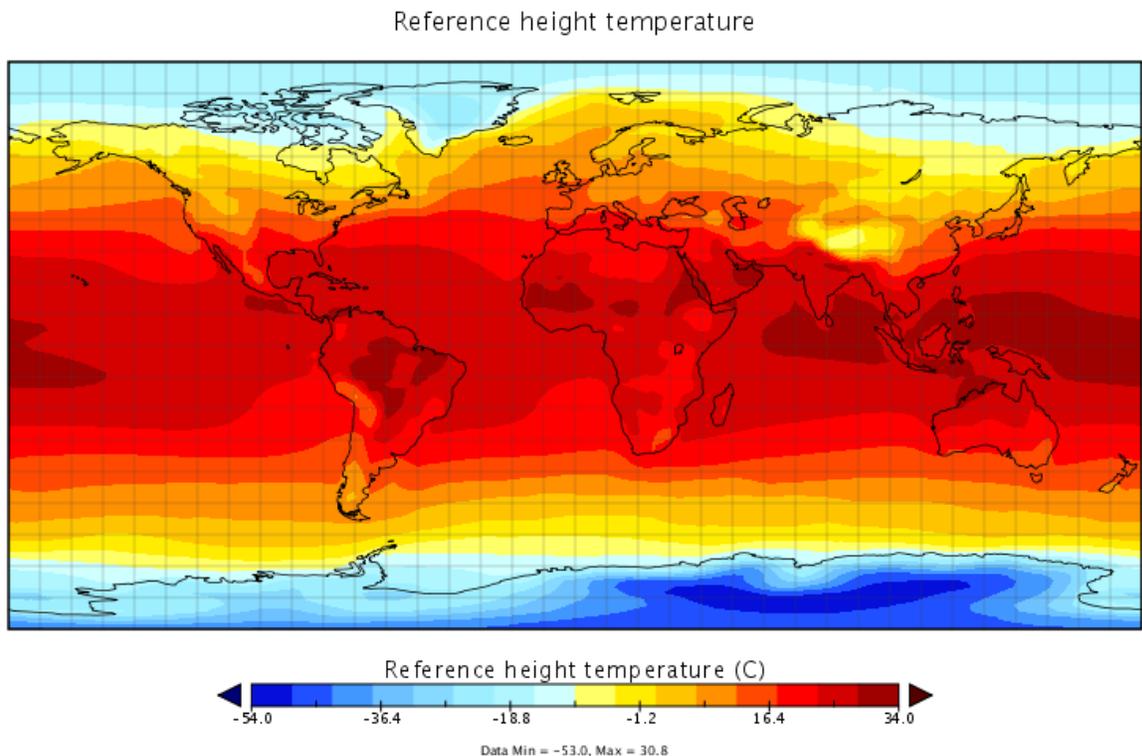


Figure 5.2: Mean of the reference temperature using LHS study.

Reference height temperature

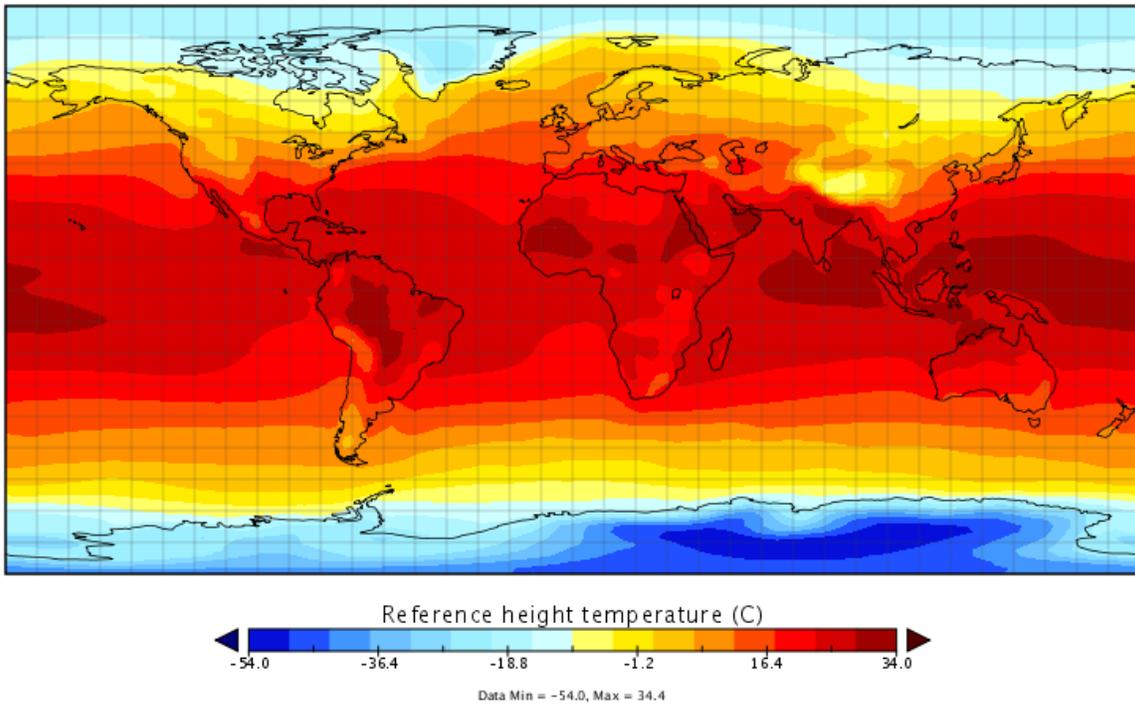


Figure 5.3: Mean of the reference temperature using polynomial chaos expansion.

Total (convective and large-scale) precipitation rate (liq + ice)

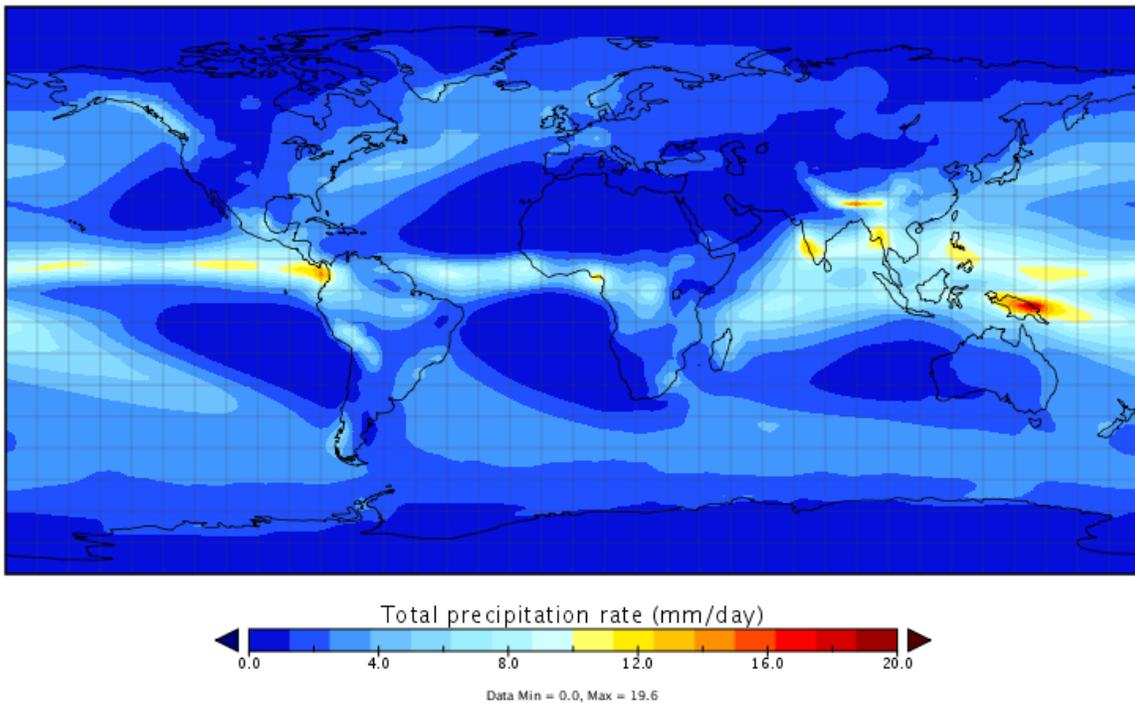


Figure 5.4: Mean of the total precipitation rate using LHS study.

Total (convective and large-scale) precipitation rate (liq + ice)

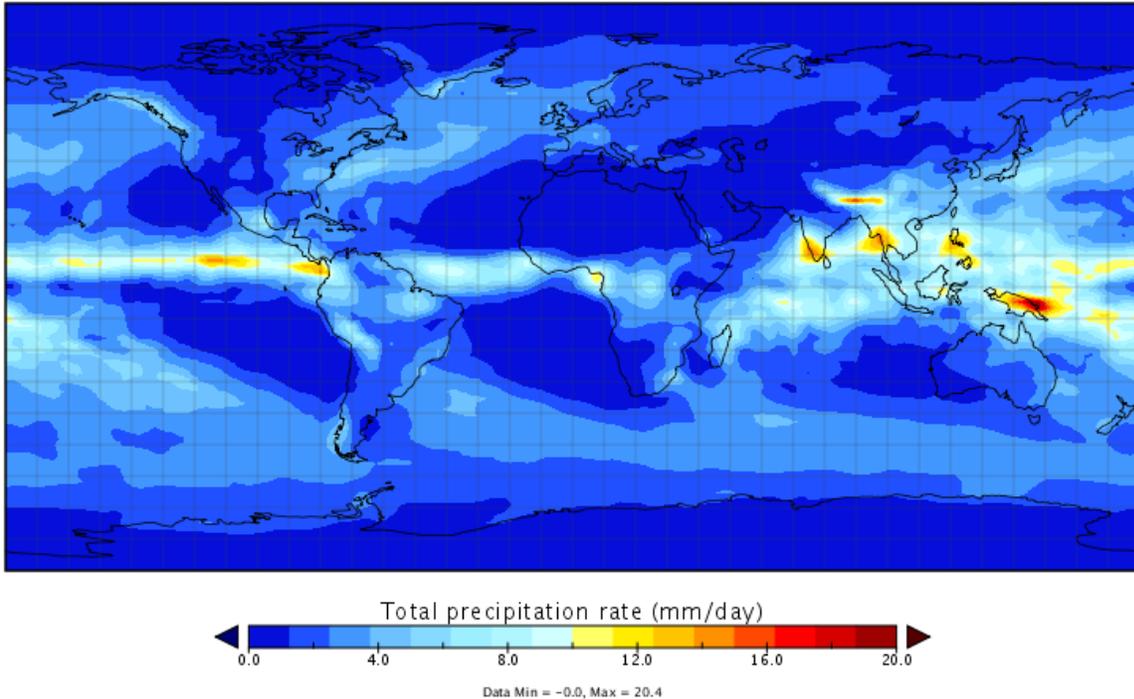


Figure 5.5: Mean of the total precipitation rate using the polynomial chaos expansion.

5.2.1 Comparison of Cumulative Density Functions

For the sake of space, we compare only the reference temperature (TREFHT), the relative humidity (RELHUM), and the precipitation rate (PRECT). We compute a CDF from the polynomial chaos expansion by taking 10,000 samples of the input random variables according to the joint distribution and interpolating the PCE at these sample points. In Figure 5.6, we show the cumulative density functions (CDFs) for each of these quantities averaged the band within 30 degrees of the equator and over the entire simulation time. We note that there is excellent agreement between the CDFs.

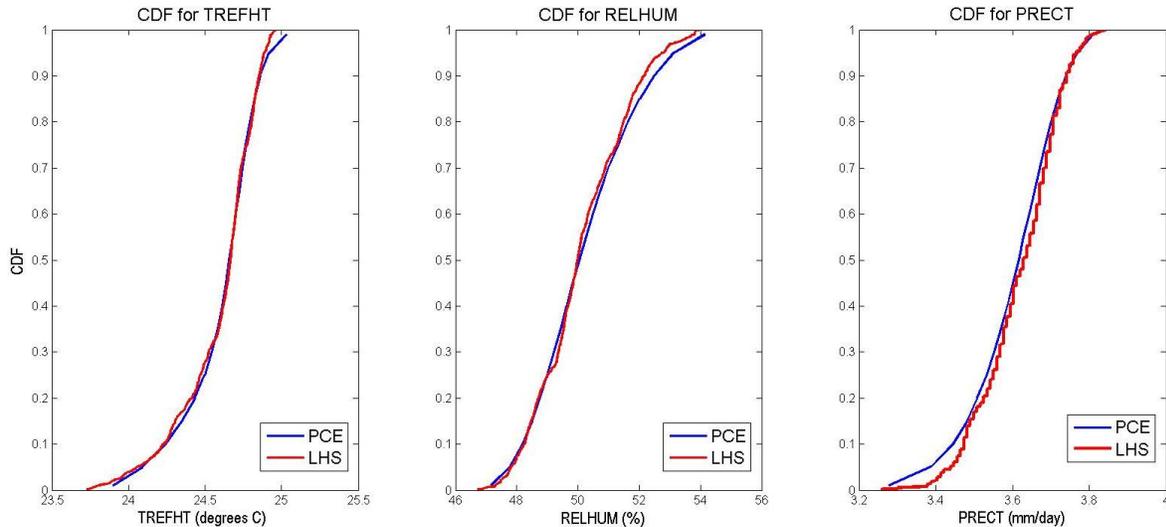


Figure 5.6: Comparison of the cumulative density functions for TREFHT, RELHUM, and PRECT calculated over $\pm 30^\circ$ equatorial band using a PCE expansion and a LHS study.

Next, we compare the CDF's for same three outputs averaged over the latitude range 30:40 and the longitude range 245:265 corresponding to the Southwest United States. In Figure 5.7, we see that the CDFs obtained by sampling the polynomial chaos expansion do not match the CDFs from the LHS study as well as in Figure 5.2. The output ranges and means are in relatively good agreement, but some discrepancy exists between the overall structures of the CDFs.

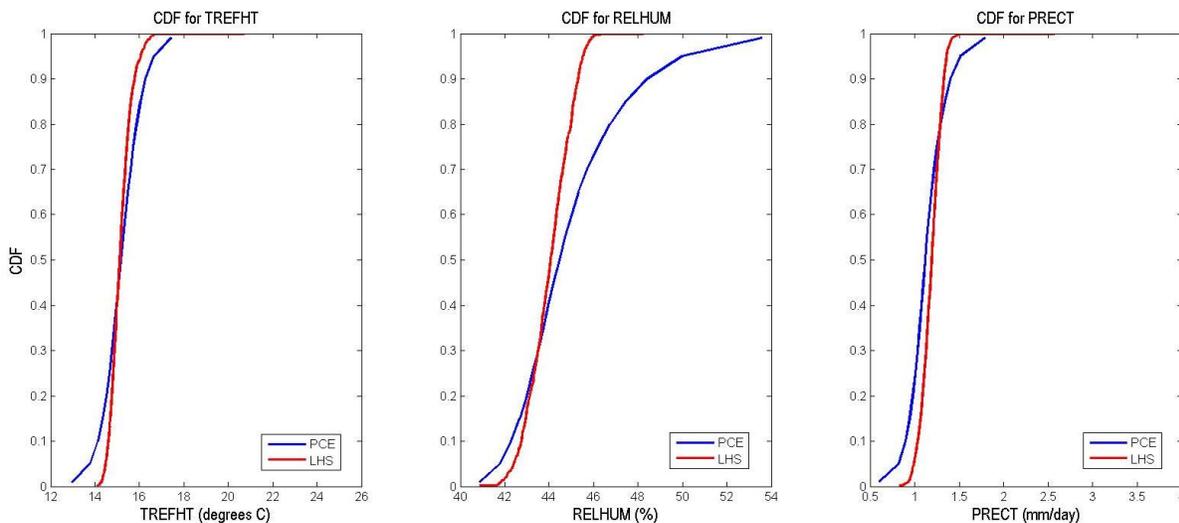


Figure 5.7: Comparison of the cumulative density functions for TREFHT, RELHUM, and PRECT over the Southwest United States using a PCE expansion and a LHS study.

Lastly, we compare the CDFs for each of these quantities averaged over the Southwest United States during only the summer months (June-August). In Figure 5.8, we see that the differences between the CDFs for the relative humidity computed from the polynomial chaos expansion and

the LHS study are comparable to the differences in Figure 5.7. On the other hand, there are significant differences between the CDFs for the reference temperature and the precipitation rate computed from the PCE and the LHS study. This is a clear indication that these particular spatial and temporal averages are more difficult to approximate with a polynomial chaos expansion due to the inherent local variability. This is consistent with the notion that regional climate information is more difficult to predict than global information.

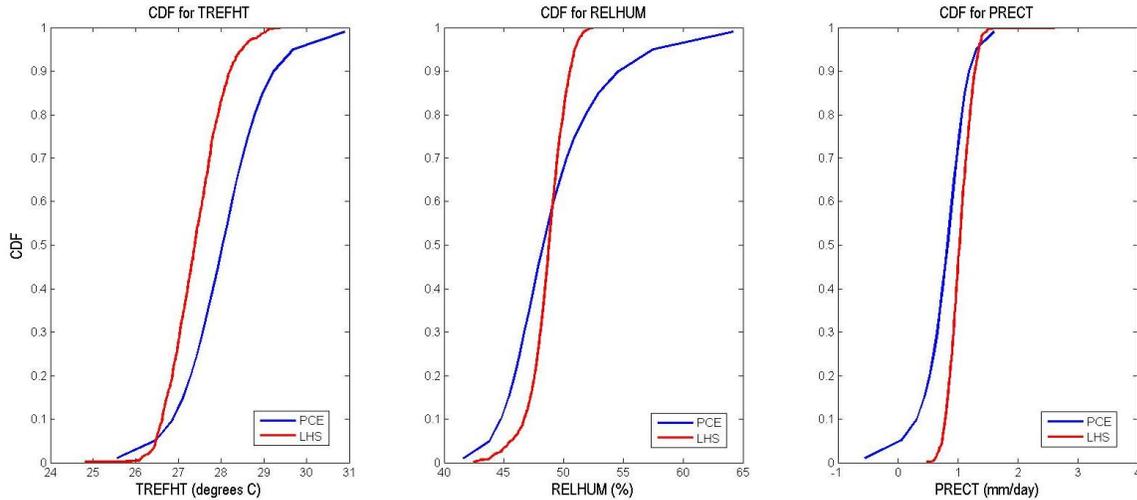


Figure 5.8: Comparison of the cumulative density functions for TREFHT, RELHUM, and PRECT over the summer in the Southwest U.S. using a PCE expansion and a LHS study.

5.2.2 Decay of Polynomial Chaos Coefficients

Section 5.2.1 gives some indication whether the statistical properties of the outputs quantities of interest can be estimated using the polynomial chaos approximation. Another objective of this study is to determine which quantities of interest from climate simulations can be approximated by global polynomials. One indication of this is whether the polynomial chaos coefficients decay as the polynomial order increases. In Figures 5.9-5.11, we plot the magnitude of the polynomial chaos coefficients for each of the three quantities of interest (TREFHT, RELHUM, and PRECT) averaged over the three spatial and temporal regions. In all nine cases, the lowest order coefficient, corresponding to the mean, is much larger than the other coefficients and we omit this term from the plots to more effectively show the decay, or lack thereof, in the coefficients as the polynomial order increases.

In Figure 5.9, we plot the magnitude of the polynomial chaos coefficients for the reference temperature, the relative humidity, and the precipitation rate averaged over the $\pm 30^\circ$ equatorial band and over the year. We see that there is a clear decay in the coefficients for the reference temperature and precipitation rate and very little decay in the coefficients for the relative humidity. This indicates that a low-order polynomial surrogate model may be a sufficiently accurate description of the reference temperature and precipitation rate, but there is much more variability in the relative humidity and higher order polynomials may be required to accurately resolve this field.

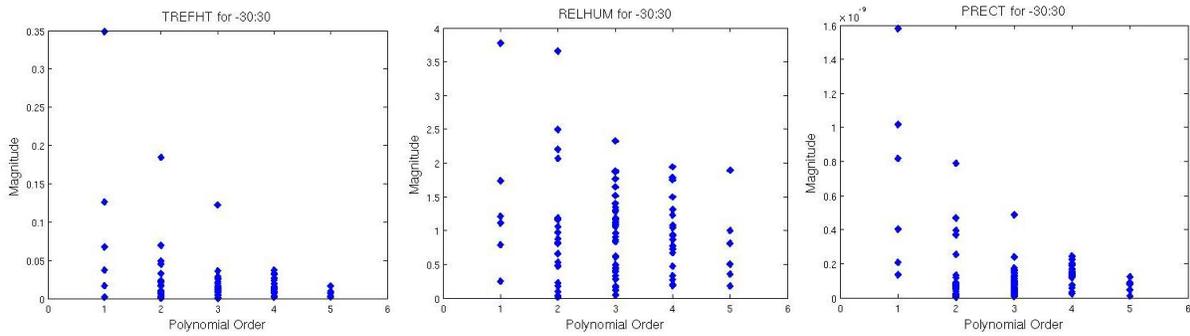


Figure 5.9: Magnitude of the polynomial chaos coefficients for TREFHT (left), RELHUM (middle) and PRECT (right) calculated over $\pm 30^\circ$ equatorial band and averaged over the year.

In Figure 5.10, we plot the magnitude of the polynomial chaos coefficients for the reference temperature, the relative humidity, and the precipitation rate averaged over the southwest United States and over the year. We observe a slight decay in the coefficients for the reference temperature, but less consistent behavior for the relative humidity and precipitation rate. This indicates that higher order polynomials may be required to accurately resolve regional quantities of interest.

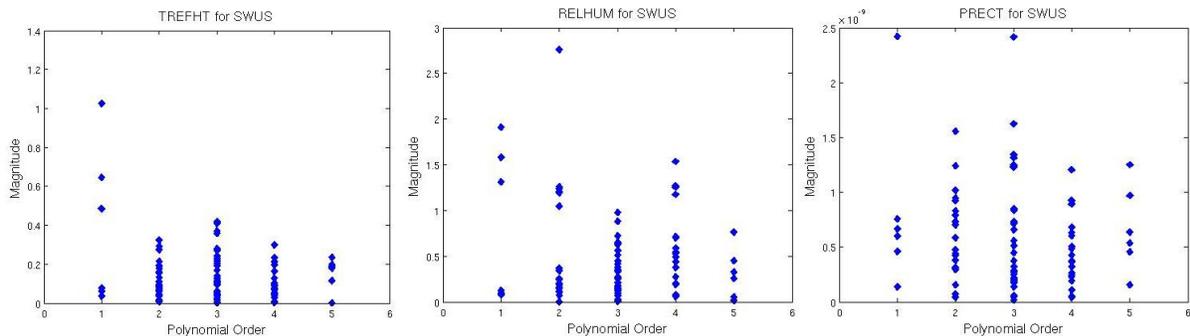


Figure 5.10: Magnitude of the polynomial chaos coefficients for TREFHT (left), RELHUM (middle) and PRECT (right) averaged over the Southwest United States and over the year.

Finally, in Figure 5.11 we plot the magnitude of the polynomial chaos coefficients for the reference temperature, the precipitation rate, and the relative humidity averaged over the southwest United States and over the summer months (June–August). We see almost no decay in the coefficients for the reference temperature and the precipitation rate. On the other hand, there is a clear decay in the coefficients for the relative humidity. This agrees with the observation that the relative humidity during the summer in the southwest is fairly predictable.

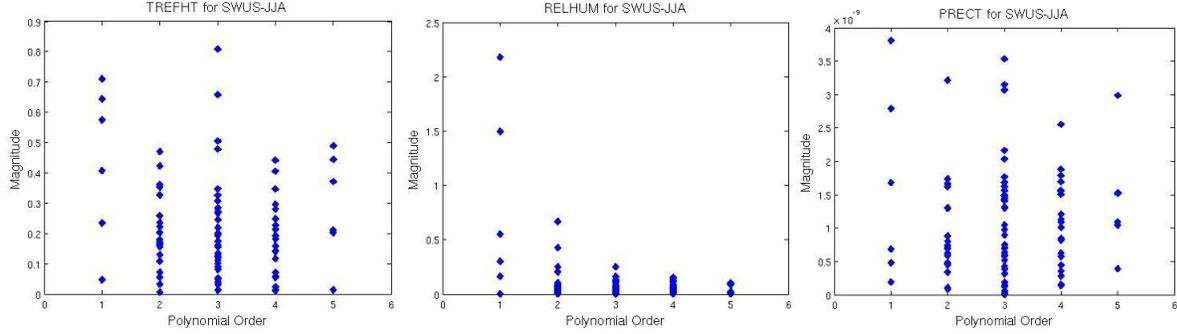


Figure 5.11: Magnitude of the polynomial chaos coefficients for TREFHT (left), RELHUM (middle) and PRECT (right) averaged over the Southwest U.S. and over the summer months.

5.2.3 Sensitivity Analysis

In this section, we use the polynomial chaos expansion to generate analytic approximations of the global sensitivities in terms of the Sobol indices. We compare these results with the sensitivity analysis in Section 3.1.2.

In Table 7, we present the Sobol indices for the averages computed over the $\pm 30^\circ$ equatorial band and over the year. Similar to Section 3.1.2, a yellow cell represents a Sobol coefficient between 0.2 and 0.5 and a red cell represents a Sobol index between 0.5 and 1.0. Comparing Tables 3 and 7, we see that the Sobol indices identify the strong influence of TAU on many of the outputs and some of the dependencies between RHMINL and RHMINH and the outputs.

	<i>RHMINL</i>	<i>RHMINH</i>	<i>ALFA</i>	<i>TAU</i>	<i>CZERO</i>	<i>KE</i>
TREFHT	0.10	0.03	0.01	0.83	0.01	0.00
T	0.23	0.24	0.26	0.15	0.04	0.01
U	0.05	0.17	0.03	0.55	0.04	0.06
PS	0.08	0.12	0.08	0.35	0.07	0.04
RELHUM	0.00	0.28	0.03	0.63	0.00	0.04
LHFLX	0.13	0.12	0.01	0.70	0.00	0.02
LWCF	0.03	0.53	0.03	0.40	0.00	0.01
SWCF	0.85	0.08	0.00	0.04	0.00	0.01
PRECT	0.21	0.13	0.03	0.53	0.02	0.04
RADBAL	0.95	0.02	0.00	0.02	0.00	0.00

Table 7: Sobol Indices Computed from a Polynomial Chaos Expansion for Averages Over $\pm 30^\circ$ Equatorial Band and Over the Year

In Table 8, we present the Sobol indices for the averages computed over the southwest United and State and over the year. We observe that RADBAL still depends strongly on RHMINL, and some of the dependencies are also indicated, but many of the significant and strongly significant correlations in Table 5 are not captured by the Sobol indices.

	<i>RHMINL</i>	<i>RHMINH</i>	<i>ALFA</i>	<i>TAU</i>	<i>CZERO</i>	<i>KE</i>
TREFHT	0.01	0.13	0.23	0.46	0.01	0.00
T	0.01	0.02	0.13	0.63	0.09	0.00
U	0.34	0.03	0.10	0.14	0.08	0.11
PS	0.12	0.09	0.49	0.03	0.05	0.03
RELHUM	0.05	0.18	0.05	0.13	0.24	0.23
LHFLX	0.01	0.01	0.09	0.11	0.07	0.28
LWCF	0.06	0.32	0.03	0.09	0.19	0.21
SWCF	0.03	0.46	0.03	0.05	0.07	0.37
PRECT	0.05	0.02	0.04	0.04	0.06	0.37
RADBAL	0.95	0.02	0.00	0.02	0.00	0.01

Table 8: Sobol Indices Computed from a Polynomial Chaos Expansion for Averages Over the Southwest United States and Over the Year

In Table 9, we present the analogous results for the averages over the southwest United States over the summer months. As in the previous case, the Sobol indices indicate a strong influence of RHMINL on RADBAL. However, most of the other significant correlations in Table 5 are not reflected by the magnitude of the Sobol indices.

	<i>RHMINL</i>	<i>RHMINH</i>	<i>ALFA</i>	<i>TAU</i>	<i>CZERO</i>	<i>KE</i>
TREFHT	0.11	0.11	0.26	0.04	0.01	0.16
T	0.05	0.04	0.13	0.18	0.15	0.18
U	0.02	0.26	0.21	0.18	0.04	0.02
PS	0.31	0.09	0.11	0.06	0.13	0.09
RELHUM	0.08	0.07	0.05	0.08	0.18	0.33
LHFLX	0.20	0.00	0.07	0.18	0.05	0.10
LWCF	0.07	0.42	0.03	0.02	0.14	0.19
SWCF	0.05	0.50	0.02	0.03	0.05	0.21
PRECT	0.34	0.02	0.07	0.17	0.03	0.04
RADBAL	0.94	0.02	0.01	0.02	0.00	0.01

Table 9: Sobol Indices Computed from a Polynomial Chaos Expansion for Averages Over the Southwest United States and Over the Summer

6. Calibration

Calibration goes by several names: data assimilation, parameter estimation, inverse problems, parameter identification. In this work we will use calibration to mean the adjustment of model parameters (denoted by $\boldsymbol{\theta}$) to maximize the agreement of the model predictions with experimental data.

A general formulation of the calibration problem is given by the framework of nonlinear least squares. The nonlinear model of the response y as a function of the n -dimensional inputs \mathbf{x} is given as:

$$y = f(\mathbf{x}; \boldsymbol{\theta}) + \varepsilon$$

where f is the nonlinear model, $\boldsymbol{\theta}$ is a vector of parameters to be calibrated, and ε is a random error term. We assume that $E[\varepsilon]=0$ and $Var[\varepsilon]=\sigma^2$ and the error terms are independent and identically distributed (iid). Usually y is a function of \mathbf{x} but this dependence is often implicit and $y(\mathbf{x})$ simply written as y . Given observations of the response y corresponding to the independent variables \mathbf{x} , the goal of nonlinear regression is to find the optimal values of $\boldsymbol{\theta}$ to minimize the error sum of squares function $S(\boldsymbol{\theta})$, also referred to as SSE:

$$S(\boldsymbol{\theta}) = \sum_{i=1}^n [(y_i - f(\mathbf{x}_i; \boldsymbol{\theta}))^2] = \sum_{i=1}^n [R_i(\boldsymbol{\theta})]^2$$

where $R_i(\boldsymbol{\theta})$ are the residual terms. Nonlinear regression employs an optimization algorithm to find the least squares estimator $\hat{\boldsymbol{\theta}}$ of the true minimum $\boldsymbol{\theta}^*$; a process that is often difficult [Seber and Wild]. Derivative-based nonlinear least squares optimization algorithms exploit the structure of such a sum of squares objective function. If $S(\boldsymbol{\theta})$ is differentiated twice, terms of residual $R_i(\boldsymbol{\theta})$, $R_i''(\boldsymbol{\theta})$, and $[R_i'(\boldsymbol{\theta})]^2$ result. By assuming that the residuals $R_i(\boldsymbol{\theta})$ are close to zero near the solution, the Hessian matrix of second derivatives of $S(\boldsymbol{\theta})$ can be approximated using only first derivatives of $R_i(\boldsymbol{\theta})$.

Cost functionals such as $S(\boldsymbol{\theta})$ are often augmented by adding a regularization term to make the optimization problem better-conditioned (e.g. if the system of residual equations over- or under-determined). Depending on the nature of the problem, the regularization terms can be based on a statistical model or can involve functions of the underlying systems of equations directly. Tikhonov regularization and its variants are a common approach used in this context.

For the climate problem, we had a different issue: it is very difficult to find parameters which result in a good model “match” with respect to the 10 quantities of interest shown in Table 2. It often can be a challenging problem to find parameters which result in calibrated parameters for just one quantity of interest. To address the issue of these disparate responses, we decided not to use a weighted least squares approach. A weighted least squares approach will try to find one solution that minimizes a weighted sum of residuals (e.g. in this case, 10 sets of residuals, one for each objective function). To perform any sensitivity analysis, it is necessary to re-weight the sum of the individual residuals and re-run the optimization to see how strongly the set of optimal parameters depends on the weighting. To avoid this issue, we instead use an approach based on Pareto

optimization, which calculates a Pareto optimal set of solutions all within one optimization procedure. The Pareto optimization yields sets of parameters which explicitly show the tradeoff between matching well on response vs. another. This approach is described below in Section 6.1.

6.1 Pareto Optimization and the MOGA algorithm

Pareto optimization is used for multi-objective problems. These are problems which have objective functions that are vectors, not scalars. Formally, a multi-objective optimization problem can be specified as:

Minimize:

$$\bar{F}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_k(\mathbf{x})]^T$$

Subject To:

$$g_i(\mathbf{x}) \leq 0 \quad i = 1, 2, \dots, m$$

$$h_i(\mathbf{x}) = 0 \quad i = 1, 2, \dots, p$$

$$x_{i_L} \leq x_i \leq x_{i_U} \quad i = 1, 2, \dots, d$$

Where \mathbf{x} is a vector of d input parameters, there are k scalar objectives denoted by $f_j(\mathbf{x})$ where $j = 1 \dots k$, $\bar{F}(\mathbf{x})$ is the overall vector objective, and the problem may have equality constraints $h(\mathbf{x})$ and/or inequality constraints $g(\mathbf{x})$ as well as bound constraints on the parameters.

In a multi-objective problem, there are two or more objectives that you wish to optimize simultaneously. The solution is the set of all points that satisfy the Pareto optimality criterion with respect to the entire decision space. This optimality definition is defined in [Coello Coello et al.]. A feasible vector \mathbf{x}^* is Pareto optimal if there exists no other feasible vector \mathbf{x} which would decrease (improve) some objective without causing a simultaneous increase (worsening) in at least one other objective. The Pareto frontier is composed of all solutions which are Pareto optimal. A typical looking Pareto frontier is shown below in Figure 6.1: In this figure, the blue line shows the Pareto frontier: all points along this curve are Pareto optimal. The goal is to be in the lower left corner (e.g. minimize both Objective 1 and Objective 2). Note that the red circle shows a solution which is NOT Pareto optimal; it is called a dominated solution.

Typical Looking Pareto Frontier

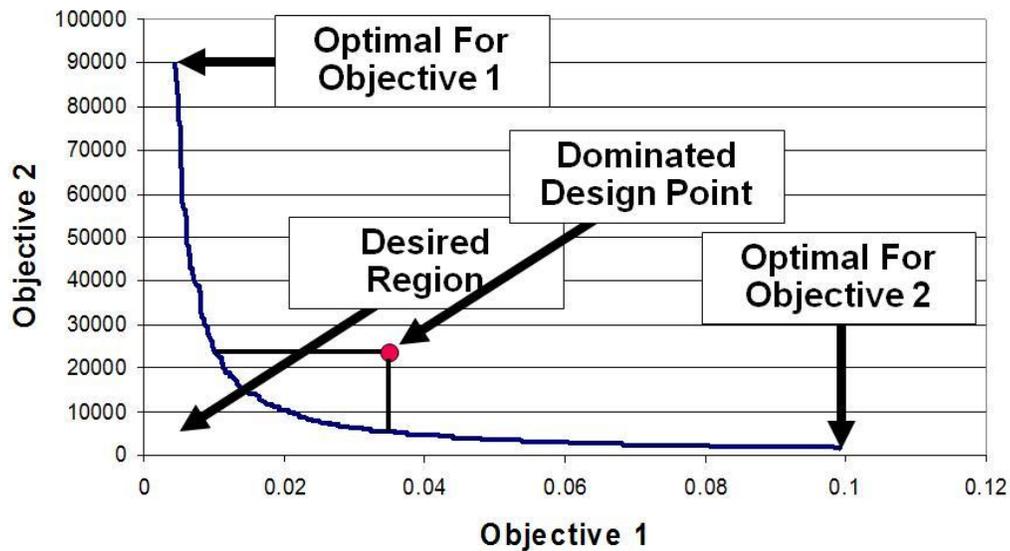


Figure 6.1: Pareto frontier

To solve for the Pareto frontier, we use a multi-objective genetic algorithm (MOGA) that is implemented in the DAKOTA framework [Adams et al. 2010]. MOGA was developed by John Eddy at Sandia National Laboratories. Genetic algorithms are effective at evolving and tracking populations of solutions, so it is easy to adapt these to keep populations of optimal solutions according to the Pareto criterion. Genetic algorithms work by initializing a population of solutions, evaluating their fitness, then selecting “good” members of the population to crossover and mutation and evolve into the next generation [Goldberg]. Over time, genetic algorithms are effective at producing globally optimal solutions.

MOGA was built with some typical genetic algorithm controls: it has a set of initialization, crossover, and mutation controls. There are also aspects of MOGA that have been customized from the single-objective genetic algorithm. For example, the user can specify a fitness type, which can be a “domination count” or a “layered” fitness operator. Both have been specifically designed to avoid problems with aggregating and scaling objective function values and transforming them into a single objective. Instead, the domination count fitness assessor works by ordering population members by the negative of the number of designs that dominate them. The values are negated in keeping with the convention that higher fitness is better. The layered rank fitness assessor works by assigning all non-dominated designs a layer of 0, then from what remains, assigning all the non-dominated designs a layer of 1, and so on until all designs have been assigned a layer. Again, the values are negated for the higher-is-better fitness convention.

MOGA also has some niche pressure operators. The job of a niche pressure operator is to encourage diversity along the Pareto frontier as the algorithm runs. This is typically accomplished by discouraging clustering of design points in the performance space. Currently, the niche pressure

operators available are the radial nicher and the distance nicher. The radial niche pressure applicator works by enforcing a minimum Euclidean distance between designs in the performance space at each generation. The distance nicher enforces a minimum distance in each dimension.

One drawback of the MOGA is that it is computationally expensive. Typically, it is necessary for a genetic algorithm to “evolve” for hundreds to thousands of generations, with hundreds of population members each generation. This means tens of thousands of function evaluations. To overcome this limitation, we use a “surrogate-based MOGA.” The basic idea is to construct a surrogate or meta-model of the expensive simulator, and perform the MOGA on that. However, instead of doing this just once, we do it iteratively. That is, an initial surrogate is built based on a user-specified set of sample points, such as from a Latin Hypercube Sample. This surrogate is then used by MOGA as the function evaluator in generating the Pareto set. After MOGA has finished and identified the Pareto front, selected points along the Pareto front (these are surrogate points) are then evaluated by the “true” function evaluator. These “true” function points are added to the original set of true points, and this “full” set is used to create another surrogate. MOGA is run again, using the surrogate on the “full” set of points. This process is repeated until the Pareto front converges. Note that the surrogate is not updated within MOGA run but between them.

George Box is famous for saying that “All models are wrong but some are useful.” A consequence of this is that no model can perfectly predict all aspects of reality. In the context of calibrating a model with multiple outputs of interest, there typically is no single set of calibration parameters that causes the model to match all outputs better than all other possible calibrations. In other words, model calibration typically involves optimizing a set of competing objectives. In this case, it may be desirable to use an ensemble of models approach when making predictions, where different calibrations of the same simulator may be considered to be different “models.” It is desirable that all calibrations used in the ensemble be Pareto optimal, however, not all Pareto optimal calibrations will be equally “useful” for prediction. The most useful models will be the ones that perform reasonably well in all objectives and do very well in one or more objectives. Additionally, one would like the calibrations to be fairly well spaced, or in the context of MOGA, niched.

Given an infinite computational budget, MOGA could conceptually determine all points or calibrations on the Pareto frontier, i.e. all optimal trade-offs or compromises that could be made. However, most of these wouldn’t be useful, either because they do too poorly in one or more objectives, or because they are too similar to other useful calibrations. After performing the initial 1016 MOGA runs on the simulator we found that many were not useful for the first reason. Since our computational budget was finite, we sought to discourage MOGA from spending effort finding Pareto optimal calibrations that did too poorly in any objective when performing the additional surrogate-based-MOGA runs.

The idea behind our modified approach was to combine physically related objectives (misfits between historical data and predictions of it) into a reduced set of 5 objectives in such a way as to penalize poor performance more than rewarding good performance. The four objectives related to radiation, (LHFLX, LWCF, SWCF, and RADBAL) were normalized by the default values and summed with the largest normalized value being added twice. The two objectives related to precipitation and humidity (PRECT and RELHUM) were likewise normalized and summed with

the larger counting twice. Since the two temperature variables (T and TREFHT) were already had the same units they were not normalized, instead the reference values were subtracted off; they were then summed with the larger relative value being added twice. Since wind speed, U, and sea pressure, PS, are rather different physical quantities their misfits were kept as separate objectives.

The Gaussian Process surrogates predicted the 10 original objective function. If any of the predictions were less than 90% of the lowest simulator output, it was judged to be extrapolation error and the surrogate prediction was increased to the 90% value. These 10 predictions of the objective functions were then combined into 5 objectives, as described in the previous paragraph, which were fed to the MOGA optimizer. A small subset (approximately 8 parameter sets) of the surrogate-based-MOGA Pareto set that were predicted to perform reasonably well in all 10 objectives was selected from each cycle.

In this work, we constructed an initial Gaussian process surrogate based on 1016 samples of CAM4. Then, we supplemented this with another 134 samples, based on 17 surrogate-based-MOGA cycles.

6.2 Results

The goal was to find a small Pareto optimal ensemble of parameter sets that performed well in all 10 outputs. This region is sometimes described as the “knee” of the Pareto frontier because the global shape of the frontier often bends most sharply here. A pictorial example of the knee for a two-dimensional Pareto front is shown in Figure 6.2 . The knee in this schematic is portion of the Pareto frontier closest to the lower left corner because the goal is to minimize all objectives. If the goal was to maximize all objectives, then the knee would be in the upper right corner instead.

Example of "knee region" on a 2D Pareto Front

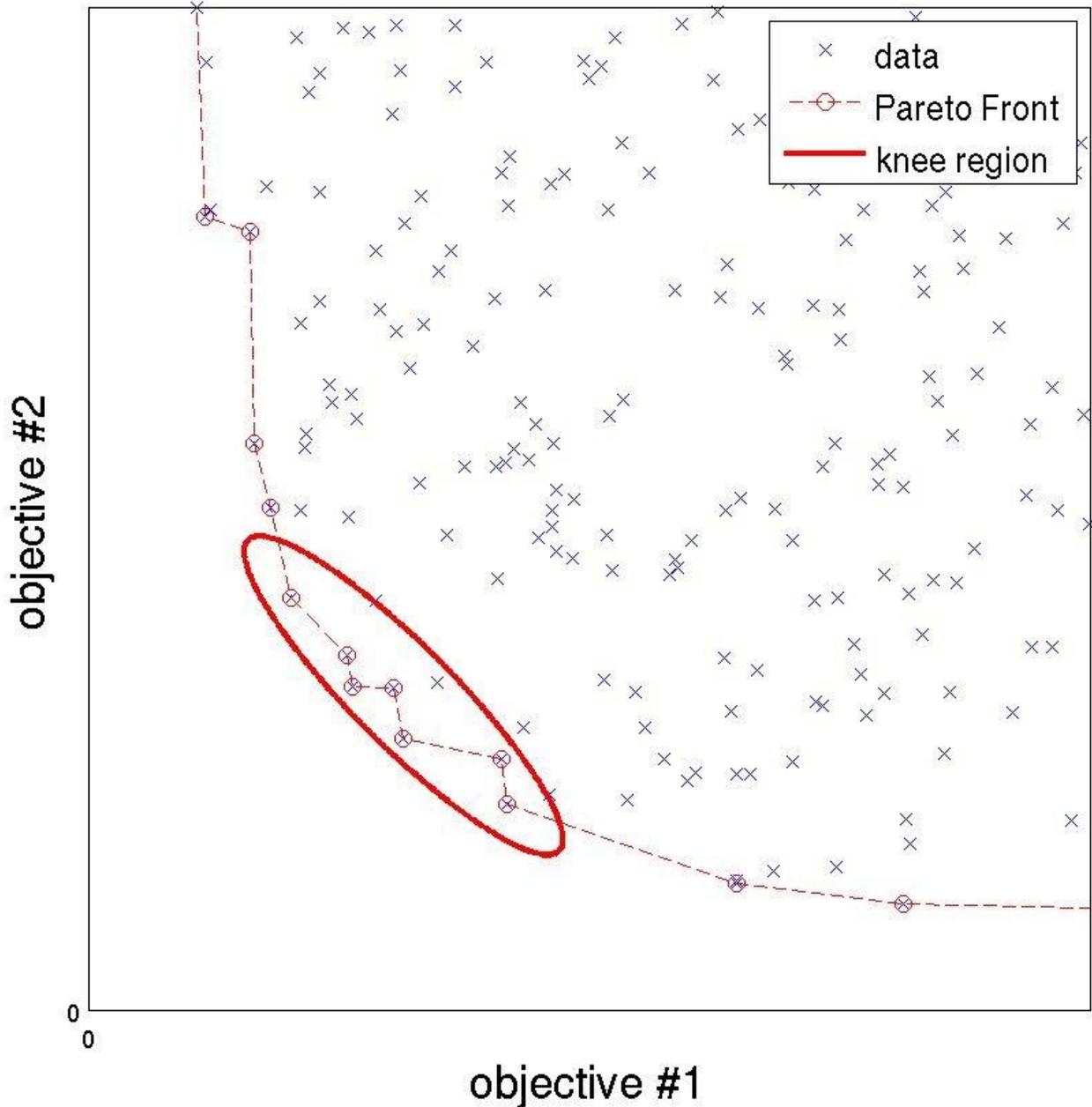


Figure 6.2. Example of the “knee” of the Pareto optimal set of solutions, where the goal is to minimize with respect to both objective #1 and objective #2.

For the climate model, the knee region represents an ensemble of feasible calibrations which could be propagated forward 100 years to estimate the spread/range of possible futures. We have nearly completed the forward propagation of such an ensemble with the atmosphere component of the CCSM4 climate model. The six inputs and the “misfit” between the 10 outputs and calibration data for the knee region of the computed Pareto frontier are plotted in Figure 6.3. This ensemble

contains 18 parameter sets computed by our surrogate-based-MOGA approach, numbered 1 through 18. The reference calibration is also plotted as the red square.

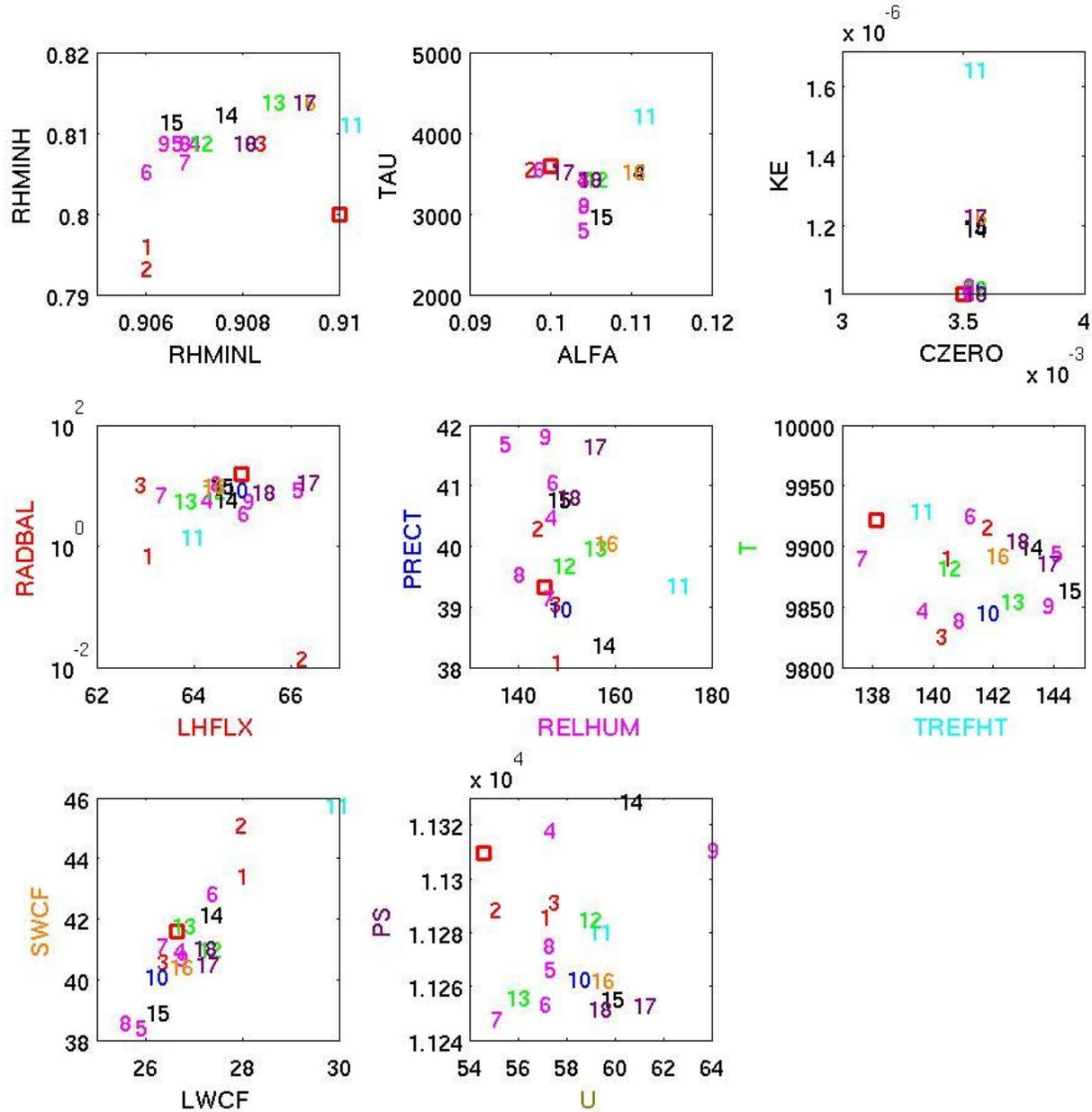


Figure 6.3. Results of Pareto Optimization for the 6 Inputs/10 Output CAM4 Problem. This figure displays the 18 Pareto optimal solutions along the “knee” of the Pareto surface. Top row shows 2D projections of inputs. Second and third row show the placement of the 18 points in 2D projections of the “misfits” of the outputs.

The first row of subplots in Figure 6.3 shows 2D projections of the 6 inputs for the knee ensemble. The second and third rows of subplots show 2D projections of the knee region of the 10D Pareto frontier. This 10D Pareto optimal ensemble was post-processed to determine which parameter sets

would be on the Pareto frontier if only N output dimensions were considered, for when N is increased from 2 to 10 by increment of 1. The order in which outputs dimensions, and parameter sets, were added is indicated by the color coding. The outputs considered for the 2D Pareto ensemble are the misfits in latent heat flu (LHFLX) and radiation balance (RADBAL), which are colored in red (sets 1-3). The misfit in relative humidity (RELHUM, sets 4-9) was added next and colored magenta. This was followed by misfits in precipitation (PRECT, blue, set 10); temperature at the reference height (TREFHT, cyan, set 11); temperature (T, green, sets 12 and 13); long wavelength cloud forcing (LWCF, black, sets 14 and 15); short wavelength cloud forcing (SWCF, orange, set 16); wind speed (U, grayish tan, this did not admit additional parameter sets); and sea pressure (PS, dark purple, sets 17 and 18).

When only misfits in LHFLX and RADBAL are considered, the 2D Pareto frontier (sets 1-3 which are colored red) has the same knee shape as in Figure 6.2. For higher dimensions, the knee shape is harder to discern from the 2D projections. The parameter set indicated by the magenta 7 outperformed the CCSM4 default calibration in 9 out of 10 of the objectives, and was very close in the 10th (U). Note that the numbers are left aligned while the square is center aligned. The 7th set had $RHMINL=0.9067$, $RHMINH=0.8069$, $ALFA=0.10353$, $TAU=3471.0$, $CZERO=3.5e-3$, and $KE=1.0270e-6$. In the following discussion, we compare the performance of the solution of the nominal set of parameters to this MOGA Pareto optimal solution #7.

As a demonstration of what can be done with the Pareto optimal sets, one can take the parameter sets and compare the global results in the future (representing an extrapolation) with the default parameters for CAM4. We ran CAM4 with the nominal and MOGA solution #7 parameters for a 105 year run. We show the averages calculated over the last eleven years of this period, years 95-105. Figure 6.4 shows a comparison of the reference height temperature, averaged over June-July-August (J-J-A) over years 95-105 given the default parameters (top) and the parameters from one of the MOGA solutions (e.g. solution 7, bottom). Figure 6.5 shows a comparison of precipitation, also averaged over J-J-A over years 95-105 given the default parameters (top) and parameters from MOGA solution 7 (bottom). Note in these comparisons that there are many similarities, but there are also differences. The MOGA solution produces results that are closer to the data.

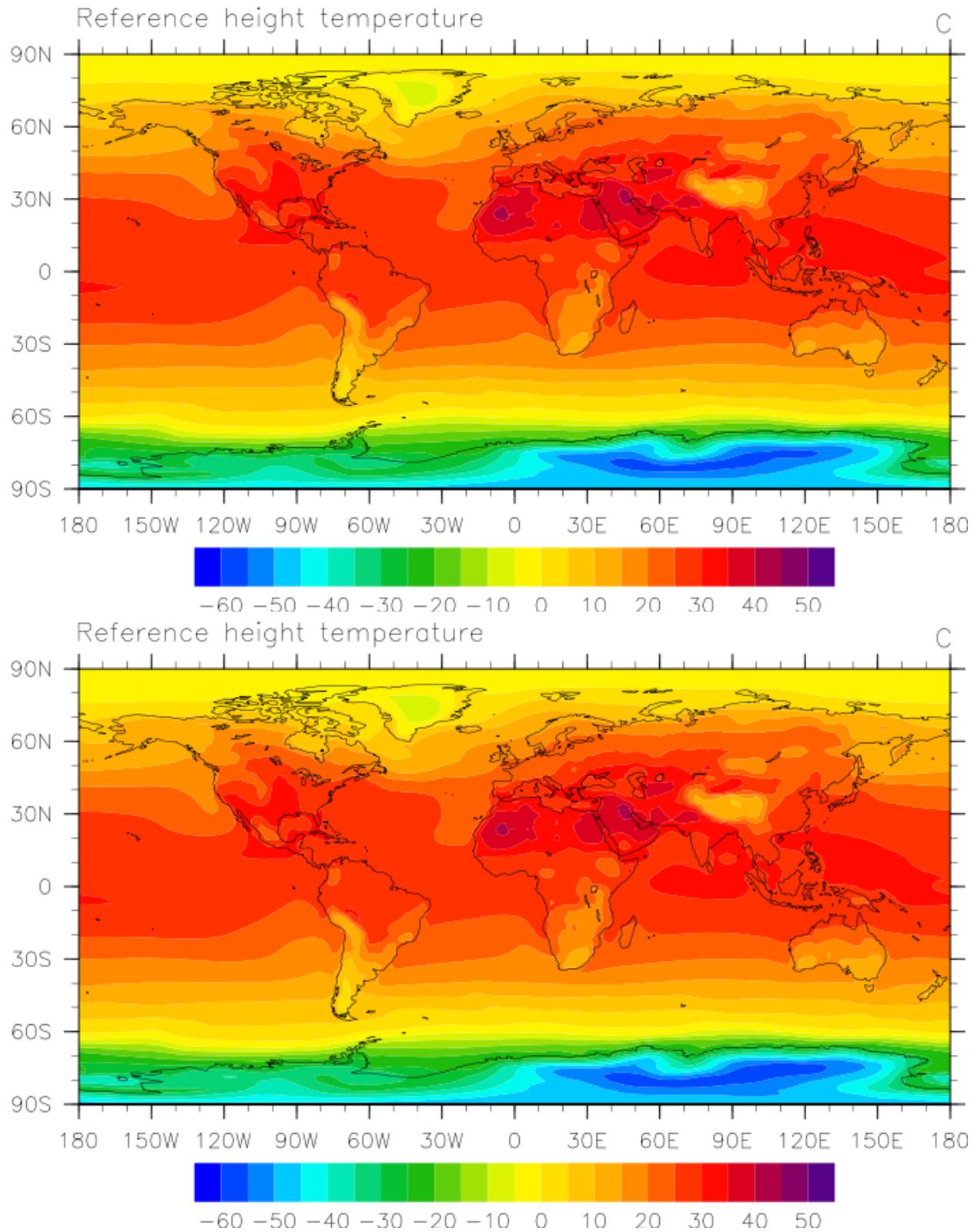


Figure 6.4. Comparison of J-J-A average Reference Height Temperature (in degrees C) over Years 95-105, with Default parameters (top) and MOGA solution 7 (bottom).

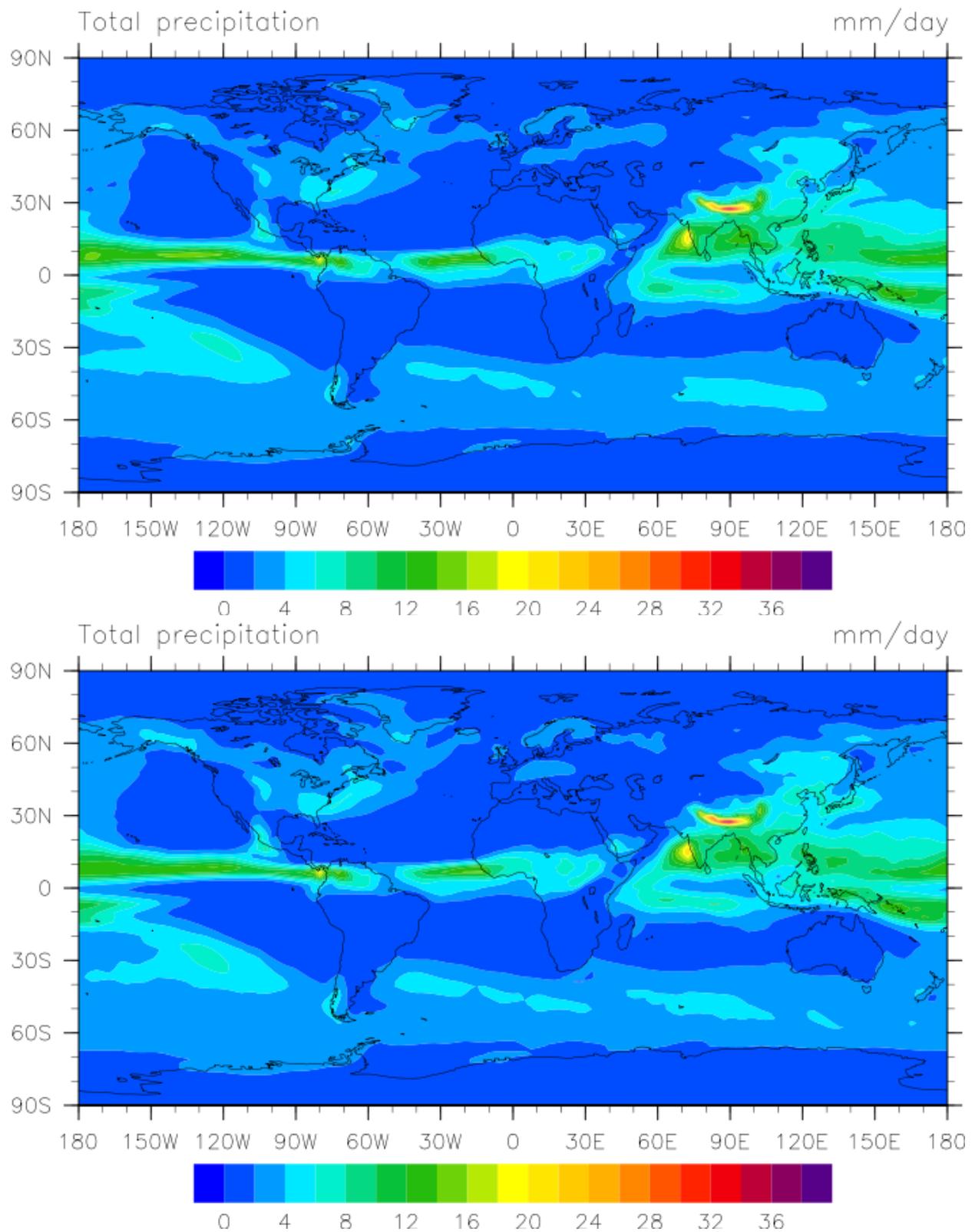


Figure 6.5. Comparison of J-J-A average Precipitation (in mm/day) at Year 75, with Default parameters (top) and MOGA solution 7 (bottom).

7. Summary and Next Steps

We have performed several types of sensitivity and uncertainty analysis on CCSM with the CAM4 atmosphere as a demonstration of methods that may be used on the CESM/CAM5 atmosphere model. Specifically, we performed correlation analysis between inputs and outputs to identify important input parameters (Section 3.1) and we compared that with a more comprehensive sensitivity measure, variance-based decomposition (Section 3.2 and Sections 5.2.3). We saw very consistent results between these methods, although the correlation was based on sampling and the variance-based analysis was based on a stochastic expansion constructed on a sparse grid.

We identified ranges on the outputs given ranges on the inputs (Section 5.2.1). We examined the use of surrogate models, including Gaussian processes, polynomial chaos expansions, and stochastic collocation (Section 4). We discussed the use of sparse grid methods to reduce the number of simulation evaluations (Section 5.1) and we compared the overall uncertainties predicted by Latin Hypercube sampling and stochastic collocation through a comparison of cumulative density functions of the outputs (Section 5.2.1). We demonstrated that these results are similar, especially for globally averaged quantities, and we further demonstrated that sparse grid methods can be used to calculate such CDFs with an order of magnitude reduction in samples (e.g. 97 vs. 1000 for a six dimensional input space). We examined the decay of the coefficients in the stochastic expansion and how these may be used to indicate whether the statistical properties of the outputs quantities of interest can be approximated well by global polynomials. We discussed what polynomial order is required to capture certain effects (Section 5.2.2). We investigated calibration methods, specifically multi-objective methods which aim to find a set of Pareto optimal points that perform well in terms of matching to data from multiple responses (Section 6.1). The MOGA results identified parameters which provide a good match according to several output metrics (Section 6.2).

Overall, this is an initial study that demonstrates methods and tools that are currently available and applicable to climate modeling. This study directly relates to the first objective of the CSSEF UQ area:

1. Implement and test production-ready UQ tools in collaboration with test beds

The study also demonstrates some techniques that are available in surrogate methods, sampling and sparse grid methods, and calibration. We hope to continue and build upon this work demonstrating similar results with CAM5 in FY2012.

References

1. Adams, B.M., Bohnhoff, W.J., Dalbey, K.R., Eddy, J.P., Eldred, M.S., Gay, D.M., Haskell, K., Hough, P.D., and Swiler, L.P., "DAKOTA, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 5.0 User's Manual," Sandia Technical Report SAND2010-2183, December 2009. Updated December 2010 (Version 5.1)
2. Climate Science for a Sustainable Energy Future (CSSEF). Multi-laboratory proposal to the Office of Biological and Environmental Research in the Office of Science of the U.S. Dept. of Energy, July 2010. Principal Investigator: David C. Bader, Oak Ridge. ORNL FWP ERKP791.
3. Coello Coello, C.A., G.B. Lamont, D.A. van Veldhuisen. *Evolutionary Algorithms for Solving Multi-Objective Problems*, 2nd edition. Springer, New York: 2007.
4. Constantine, P.G. and Eldred, M.S. "Sparse polynomial chaos expansions". International Journal for Uncertainty Quantification, in preparation.
5. Cressie, N. A. C. *Statistics for Spatial Data*, Wiley, New York, 1993.
6. Ghanem, R. and Spanos P., *Stochastic Finite Elements: A Spectral Approach*. Springer Verlag, New York, 2002.
7. Goldberg, D.E. *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, 1989.
8. Gerstner, T. and Griebel, M. "Numerical integration using sparse grids." Numer. Algorithms, 18(3-4):209–232, 1998.
9. Jackson, C.S., M.K. Sen, G. Huerta, Y. Deng, and K.P. Bowman. "Error Reduction and Convergence in Climate Prediction." *Journal of Climate*, Vol. 21, pp. 6698-6709, 2008.
10. Larsen, R. J. and M. L. Marx. *An Introduction to Mathematical Statistics and its Applications*, 2nd ed. Prentice-Hall. Edgewood Cliffs, NJ: 1986.
11. Nobile, F., Tempone, R., and Webster, C.G. "An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data." SIAM J. on Num. Anal., 46(5):2411–2442, 2008.
12. Rasmussen, C.E. and C.K.I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.
13. Swiler, L. P. and G. D. Wyss. "A User's Guide to Sandia's Latin Hypercube Sampling Software: LHS Unix Library/Standalone Version." Technical Report SAND2004-2439. Sandia National Laboratories, Albuquerque NM.
14. Sobol', I.M.. *Sensitivity analysis for non-linear mathematical models*. Mathematical Modeling and Computational Experiment 1993;1:407–414.
15. Saltelli, A., Chan, K., Scott, E.M. *Sensitivity Analysis*. New York: Wiley; 2000.
16. Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M. *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. New York: Wiley; 2004.
17. Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., Tarantola, S. "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index." Computer Physics Communication 2010;181:259 – 270.
18. Smolyak, S.A.. "Quadrature and interpolation formulas for tensor products of certain classes of functions". Dokl. Akad. Nauk SSSR, 4:240–243, 1963.

19. Storlie, C.B. and J.C. Helton. "Multiple predictor smoothing methods for sensitivity analysis: Description of techniques." *Reliability Engineering and System Safety*, 93(1):28–54, 2008.
20. Storlie, C.B., L.P. Swiler, J.C. Helton, and C.J. Sallaberry. "Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models." *Reliability Engineering and System Safety*, 94 (2009) 1735–1763
21. Seber, G. A. F. and C. J. Wild, *Nonlinear Regression*, Wiley & Sons, 2003.
22. Sudret, B. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety* 2008;93(7):964 – 979.
23. Tang, G., Iaccarino, G., Eldred, M.S.. Global sensitivity analysis for stochastic collocation expansion. In: Proceedings of the 12th AIAA Non-Deterministic Approaches Conference. AIAA-2010-2922; Orlando, FL; 2010.
24. Weirs, V.G., Kamm, J. R., Swiler, L.P., Ratto, M., Tarantola, S., Adams, B.M., Rider, W.J. and Eldred, M.S., "Sensitivity Analysis Techniques Applied to a System of Hyperbolic Conservation Laws." Accepted by *Reliability Engineering and System Safety* in 2011, publication pending.
25. Xiu, D. and Hesthaven, J.S. "High-order collocation methods for differential equations with random inputs". *SIAM J. Sci. Comput.*, 27(3):1118–1139 (electronic), 2005.
26. Xiu, D. and Karniadakis, G. "The Wiener-Askey polynomial chaos for stochastic differential equations". *SIAM J. Sci. Comput.* 24:619-644, 2002.

DISTRIBUTION

1	MS1318	1440	B. A. Hendrickson
1	MS 1318	1441	J. R. Stewart
1	MS 1318	1441	B. Adams
1	MS1318	1441	K. Dalbey
1	MS 1318	1441	L. P. Swiler
1	MS 1318	1441	T. M. Wildey
1	MS 1318	1441	T. G. Trucano
1	MS1318	1442	O. Guba
1	MS1318	1442	M. N. Levy
1	MS1318	1442	M. A. Taylor
1	MS1325	1460	J. L. Mitchiner
1	MS9051	8351	R. D. Berry
1	MS9051	8351	B. Debusschere
1	MS9051	8351	H. N. Najm
1	MS9051	8351	K. Sargsyan
1	MS9051	8952	C. Safta
1	MS9159	8954	J. Ray
1	MS 0899	9532	RIM-Reports Management (electronic copy)