

The Pennsylvania State University
The Graduate School

**SENSITIVITY OF DISTRIBUTIONS OF CLIMATE SYSTEM
PROPERTIES TO SURFACE TEMPERATURE DATASETS**

A Thesis in
Meteorology
by
Alex G. Libardoni

© 2011 Alex G. Libardoni

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

August 2011

The thesis of Alex G. Libardoni was reviewed and approved* by the following:

Chris E. Forest
Associate Professor of Climate Dynamics
Thesis Advisor

Klaus Keller
Associate Professor of Geosciences

Raymond G. Najjar
Professor of Oceanography

Johannes Verlinde
Associate Professor of Meteorology
Associate Head, Graduate Program in Meteorology

*Signatures are on file in the Graduate School.

Abstract

Predictions of climate change depend strongly on the accurate implementation and parameterization of climate system properties, processes, and feedbacks. In this study, surface temperature, upper-air temperature, and ocean heat content data are used to constrain the distributions of the parameters that define three climate system properties: climate sensitivity, the rate of ocean heat uptake into the deep ocean, and net anthropogenic aerosol forcing. Climate sensitivity is diagnosed by changing the strength of cloud feedback, the rate of deep-ocean heat uptake is determined by varying the effective vertical diffusivity of heat anomalies in the ocean, and the net anthropogenic aerosol forcing is controlled by scaling the spatial and temporal pattern of sulfate aerosol loadings by the model-defined global value in the 1980s. Running a climate model of intermediate complexity forced by historical forcing patterns over a range of these parameter values allows for the derivation of probability distribution functions for the model parameter values corresponding to the climate system properties. Using five different surface temperature datasets, this study explores the sensitivity of the parameter distributions to the choice of surface temperature data used to evaluate the model output. Differences in estimates of climate sensitivity mode and mean are as great as 1 K between the datasets. Ocean effective diffusivity is poorly constrained using all datasets and the shape of the distribution differs greatly depending on which surface dataset is used. Distributions for anthropogenic aerosol forcing cluster into two groups. While each group has the same general shape, the location of the mode and confidence intervals differ by approximately 0.1 Wm^{-2} between the two clusters. This difference is small compared to other uncertainties in climate forcings. Transient climate response derived from these distributions ranges anywhere between 1 and 3 K and the shape of the distribution of these possible values is surface dataset dependent. Some distributions are tall and narrow, while other distributions are short and broad. Understanding the differences in parameter distributions and predicted warming is critical to understanding the full range of uncertainty involved in climate model calibration studies.

Table of Contents

List of Figures	vi
List of Tables	viii
List of Abbreviations	ix
Acknowledgments	x
Chapter 1	
Introduction	1
1.1 Hierarchy of Climate Models	2
1.2 Past Studies	2
1.3 Optimal Fingerprint Detection	5
1.4 Detection and Attribution	8
1.5 Transient Climate Response	9
Chapter 2	
Surface Datasets	13
Chapter 3	
Methods	17
Chapter 4	
Results	20
4.1 Surface Temperature Time Series	20
4.2 Parameter Diagnostics	22
4.3 Probability Distribution Functions for Climate System Properties .	32
4.4 Transient Climate Response	38
Chapter 5	
Conclusions	41

5.1 Future Work	42
Bibliography	46

List of Figures

4.1	Surface temperature time series from 1896-1995 for each of the individual datasets. Only data beginning in 1906 is used in this study, but the time series begins in 1896 to show data from the same period that was available in Forest et al. (2008). Vertical lines mark the beginning of the period used in the surface diagnostic (1946).	21
4.2	Distributions of r^2 for temperature diagnostics as a function of climate sensitivity and effective ocean diffusivity at an aerosol forcing (labeled as FA) of 0.50 Wm^{-2} . White shading represents the 90-percent confidence region, and light blue shading represents the 99-percent confidence region based on an F-test. Dark blue shading represents regions that are outside of the 99-percent confidence region and are thus rejected at the 1-percent significance level.	24
4.3	As in Fig. 4.2 but at an aerosol forcing of 0.25 Wm^{-2}	25
4.4	As in Fig. 4.2 but at an aerosol forcing of 0.00 Wm^{-2}	26
4.5	As in Fig. 4.2 but at an aerosol forcing of -0.25 Wm^{-2}	27
4.6	As in Fig. 4.2 but at an aerosol forcing of -0.50 Wm^{-2}	28
4.7	As in Fig. 4.2 but at an aerosol forcing of -0.75 Wm^{-2}	29
4.8	As in Fig. 4.2 but at an aerosol forcing of -1.00 Wm^{-2}	30
4.9	As in Fig. 4.2 but at an aerosol forcing of -1.50 Wm^{-2}	31
4.10	Marginal probability distribution functions using each of the surface datasets. Distributions for climate sensitivity (top), effective ocean diffusivity (middle), and net anthropogenic aerosol forcing (bottom) are shown. Box-and-whisker plots show the percentile bounds: 2.5-97.5 (dots), 5-95 (vertical lines at ends), 25-75 (box ends), and 50 (vertical line in box). The distribution mean is represented by a diamond, and the distribution mode is the peak in the distribution	34

4.11	Two-dimensional marginal distribution function for climate sensitivity and effective ocean diffusivity when using HadCRUT2 (black), HadCRUT3 (blue), NCDC (green), GISTEMP 250 (orange), and GISTEMP 1200 (red) surface data. Starting from the outermost moving inward, contours mark the 99-, 90-, and 10-percent confidence regions. The 10-percent confidence region is included to show the modes of the distributions	36
4.12	As in Fig 4.11, but for the two-dimensional marginal distribution function for climate sensitivity and anthropogenic aerosol forcing. .	37
4.13	Histogram of TCR (top) and cumulative density function (bottom) derived from 1000 member Latin Hypercube samples of the joint distribution functions using each surface dataset.	39

List of Tables

2.1	Summary of differences in surface temperature datasets	16
4.1	Percentiles and means for marginal probability distribution functions for each surface dataset	35
4.2	Percentiles for distributions of transient climate response for each surface dataset	39

List of Abbreviations

AOGCM: Atmosphere-Ocean General Circulation Model
CO₂: Carbon dioxide
EMIC: Earth System Model of Intermediate Complexity
ERSST: Extended Reconstructed Sea Surface Temperature dataset
GFDL R30: Geophysical Fluid Dynamics Laboratory, R30 model
GHCN: Global Historical Climatology Network
GISTEMP 250: Goddard Institute of Space Studies surface dataset, 250 km smoothing
GISTEMP 1200: Goddard Institute of Space Studies surface dataset, 1200 km smoothing
HadCM2: Hadley Centre Coupled Model version 2
HadCRUT2: Hadley Centre surface dataset, version 2
HadCRUT3: Hadley Centre surface dataset, version 3
HadISST1: Hadley Centre Sea Ice and Sea Surface Temperature dataset
HadSST2: Hadley Centre Sea Surface Temperature dataset
IPCC: Intergovernmental Panel on Climate Change
IPCC AR4: Intergovernmental Panel on Climate Change Fourth Assessment Report
MIT: Massachusetts Institute of Technology
NASA: National Aeronautics and Space Administration
NCDC: National Climatic Data Center surface dataset
NOAA: National Oceanic and Atmospheric Administration
OISST: Optimum Interpolation Sea Surface Temperature dataset
TCR: Transient Climate Response

Acknowledgments

This study was funded by contributions from the MIT Joint Program on Climate Change, the National Oceanic and Atmospheric Administration, and the United States Department of Energy. The author would like to thank NASA's Goddard Institute of Space Sciences, NOAA's National Climatic Data Center, and the Hadley Centre for Climate Prediction and Research for providing publicly accessible surface temperature data products. Special thanks to my thesis advisor, Chris Forest, and my thesis committee, Klaus Keller and Ray Najjar, for guidance in completion of this study. Thanks to Andrei Sokolov and Jeff Scott of the MIT Joint Program on Climate Change for invaluable discussion and insights regarding this work. Thanks to Tim Hatlee, Anders Jensen, Chris Nowotarski, Andrew Oberthaler, Tiffany Samuelson, Jeff Waters, and especially my family, Alan, Sherryl, and Eric Libardoni, for their contributions and support throughout this project.

To
Mom and Dad
for
their unwavering support in all of my academic endeavors

Chapter 1

Introduction

The climate system is a complex interaction of countless feedbacks and processes, all of which have some degree of uncertainty in their behavior (Bony et al., 2006; Randall et al., 2007). Specific uncertainties include the impacts of clouds and the water vapor, lapse rate, and surface albedo feedbacks. As a whole, these contribute to the long term behavior of the climate system by setting the equilibrium climate sensitivity of the climate system (Randall et al., 2007). Given current concerns over global warming, the need for climate models that yield reliable projections is undeniable. To have a model that yields results in a reasonable amount of time given available computer resources, decisions have to be made as to which climate processes to represent explicitly, parameterize, and omit completely. When parameterizations are required, this reduces complexity in a given process to a set of uncertain parameter values. Parameterization of processes marks a trade-off between completeness in the representation of the climate system and efficiency of the model. This trade-off is minimized when the optimal parameter value is used: the value that yields model output that best matches the historical record. When this parameter value is found, the model best represents the climate system behavior for a specific set of predictions, and as a result will produce the best possible projections of future climate by the given model. The search for these optimal parameter values motivates this study.

1.1 Hierarchy of Climate Models

Climate models exhibit a wide range of complexities. Current models range from fully coupled, three-dimensional Atmosphere-Ocean General Circulation Models (AOGCMs), to Earth System Models of Intermediate Complexity (EMICs), to simple climate models (Randall et al., 2007). AOGCMs represent the most complete climate models through their incorporation of many climate system components in an effort to completely mimic the dynamics of the climate system. Given their complexity, AOGCMs require significant computer time for each model run. As a result, AOGCMs give the most comprehensive climate projections but are not ideal when multiple runs are needed. EMICs include many of the same components as AOGCMs but substitute parameterizations for certain climate system components and include additional processes (e.g., biogeochemistry) not permissible in AOGCMs due to computational costs. Simple climate models include even more parameterized processes and are typically tuned to replicate results from AOGCMs (Randall et al., 2007). Given their simplifications of the climate system, EMICs and simple climate models run much faster than AOGCMs. For example, a single run of an AOGCM may take months to complete, whereas the EMIC used in this study can yield hundreds of runs in a single night. As a result of the fast run times, multiple model runs over a wide range of parameter values can be run in the same time it takes for one run of an AOGCM. This efficiency makes EMICs and simple climate models ideal for probabilistic climate model studies (Randall et al., 2007). This study will make use of the climate model component of the Integrated Global Systems Model, an EMIC developed by the MIT Joint Program on Climate Change (Sokolov and Stone, 1998; Sokolov et al., 2005), to take a probabilistic approach to the parameter calibration problem.

1.2 Past Studies

Numerous studies investigating climate parameter distributions have been conducted (Andronova and Schlesinger, 2001; Knutti et al., 2003; Tomassini et al., 2007; Forest et al., 2002, 2006, 2008; Urban and Keller, 2010). Andronova and Schlesinger (2001), Forest et al. (2002), and Knutti et al. (2003) mark the initial

studies in the field. Each study uses probabilistic methods to determine likely values for model parameters through the comparison of model output to observational data. These methods make use of the likelihood-based approach made possible by the use of EMICs and simplified climate models. Although each study explores the same problem with the same basic approach, each does so using a different methodology. The studies by each group use different approaches to compare model output to observational data, use different diagnostics for evaluation of model performance, and use different estimates of internal climate variability.

Andronova and Schlesinger (2001) estimate probability distributions for climate sensitivity under a range of forcing scenarios by comparing global mean temperature and hemispheric temperature differences. In total, 16 radiative forcing schemes are used by taking combinations of the radiative forcing produced by greenhouse gases, tropospheric ozone, anthropogenic sulfate aerosols, the sun, and volcanic eruptions. These forcing scenarios are used to investigate the impacts that each of the constituent forcings, anthropogenic aerosol forcing in particular, have on model estimated climate sensitivity. To account for uncertainty in the estimation due to internal climate variability and other error sources (i.e., observational errors), bootstrap samples of the correlated observed data are used to create multiple realizations of a natural climate system.

Forest et al. (2002) estimate probability distributions for three climate parameters: climate sensitivity, effective ocean diffusivity, and net anthropogenic aerosol forcing. Distributions of the parameters are derived using statistics from an optimal fingerprint detection algorithm (Allen and Tett, 1999) with changes in surface temperatures, upper-air temperatures, and global ocean heat content used as diagnostics. The optimal fingerprint algorithm acts to maximize the signal-to-noise ratio of the model response through a rotation of the modeled and observed temperature patterns into the coordinate space of the internal variability pattern. This maximization allows small signals in the patterns to be detected as opposed to being interpreted as noise. Rather than use bootstrap samples of the observational dataset as in Andronova and Schlesinger (2001), internal variability is estimated from unforced equilibrium control runs of an AOGCM.

Knutti et al. (2003) use a neural network to estimate probability distributions for climate sensitivity and indirect aerosol forcing in the climate system. Neural

networks use pattern recognition to determine connections between members of the model. In climate studies, these connections represent temperature patterns and trends. A significant benefit of the neural network approach is that once the network has been trained on a set of model runs, no further model runs are necessary. From the training runs, the model response using any combination of the settings can be approximated. Through comparison of the modeled surface temperature and ocean heat uptake trends to the observed trends, the performance of a model run simulated by the neural network is evaluated. One drawback of Knutti et al. (2003) is that the model is too simple to include an estimate of internal variability. It is noted that the results are reliant upon the assumption that external solar, volcanic, and anthropogenic aerosol forcings are sufficient to explain warming on decadal time scales. Furthermore, Knutti et al. (2003) argue that, based on prior studies, it can be expected that the magnitude of internal variability is smaller than the uncertainty in the observations and can thus be ignored.

Parameter calibration studies have been ongoing since the initial work presented above (Forest et al., 2006, 2008; Tomassini et al., 2007; Sansó and Forest, 2009; Urban and Keller, 2010). Forest et al. (2006, 2008) continue working with similar methodology but with updates to both the climate forcings used to drive the model and the climate model itself. Tomassini et al. (2007) and Sansó and Forest (2009) present new approaches to the estimation of climate model parameter distributions. In these studies, Markov Chain Monte Carlo methods are used. These methods use a probabilistic rule to govern the complete sampling of a distribution while spending a majority of the time in high probability regions. Similar to the neural network approach of Knutti et al. (2003), Markov Chain Monte Carlo methods can use training runs to build a statistical emulator of the model. An emulator is built such that the behavior of a model at a set of parameter values where the model has not been run can be approximated based on the behavior of the training runs. The power of an emulator is that all uncertain variables can be built into the model and distribution functions derived from the Markov Chain. Tomassini et al. (2007) build in uncertainty in the external forcings used to drive their model while Sansó and Forest (2009) incorporate uncertainty in the estimate of internal variability into the model and estimate the parameters for the emu-

lator and noise model simultaneously with the climate model parameters within a hierarchical Bayesian framework. Utilizing the efficiency of a simple climate model, Urban and Keller (2010) use Markov Chain Monte Carlo methods to estimate many of the parameters associated with the prediction of the collapse of the Atlantic Meridional Overturning Circulation. In this study, correlations that are present in each observational time series used to evaluate model performance are considered. The estimated autocorrelation is removed from the residuals between model data and observations, and the statistical parameters used to define the autocorrelation structure are estimated. These estimations are included in order to include an evaluation of the structural uncertainties in the study, as well as uncertainties in the parameters.

Building off of the studies just discussed, this study is focused on evaluating how the choice of the observation-based, surface temperature dataset used in the comparison with model output data impacts the parameter distribution functions. In total, five different surface temperature datasets are used to derive the joint probability distribution function for the three climate system parameters evaluated in Forest et al. (2002, 2006, 2008). From this joint distribution function, marginal distribution functions for each parameter, two-dimensional distribution functions for pairs of the parameters, and the distribution for a measure of future climate change are all derived. Specific details on the surface datasets used and the evaluation of future climate change are deferred to later sections.

1.3 Optimal Fingerprint Detection

In order to derive the distribution functions for the model parameters, model performance must be evaluated. A given model run can be evaluated through the use of a goodness-of-fit statistic, r^2 , calculated by

$$r^2 = (\mathbf{x} - \mathbf{y})^T \mathbf{C}_N^{-1} (\mathbf{x} - \mathbf{y}), \quad (1.1)$$

where \mathbf{x} is the vector of model output, \mathbf{y} is the vector of observed data and \mathbf{C}_N^{-1} is the inverse of the noise-covariance matrix. Low values of r^2 represent models that yield temperature patterns that are more consistent with the observations than

models that have higher r^2 values. It should be noted that this definition of r^2 is opposite of the conventional definition.

Model goodness-of-fit statistics in this study are calculated using the same optimal fingerprint detection algorithm initially described in Forest et al. (2001) and used in Forest et al. (2002, 2006, 2008). The methodology represents a variation on the methods of Allen and Tett (1999). Given that the climate varies naturally with time due to non-linear dynamics, both modeled and observed trends will have these natural variability patterns embedded in them. In an effort to estimate this background variability, control run data are obtained by running a climate model for thousands of years with no external forcings. From the control run, spatial and temporal correlations between the regions used in the diagnostics are calculated. For example, for the surface diagnostics which use four zonal bands on a decadal scale, the correlation between the temperatures in one zonal band and itself from decade to decade is calculated, in addition to the correlation within a given decade between the zonal band and the other zonal bands. The complete set of spatial and temporal correlations determines the noise-covariance matrix, $\mathbf{C}_\mathbf{N}$. In total, $\mathbf{C}_\mathbf{N}$ is a square matrix of dimension $N_T \times N_T$, where $N_T = N_{zones} \times N_{decades}$.

Through eigenvalue decomposition, $\mathbf{C}_\mathbf{N}$ is decomposed into a set of orthogonal eigenfunctions that describes the internal climate variability. As the rank of the eigenfunction decreases, so too does the magnitude of its eigenvalue, and thus the fraction of the total variance in the overall pattern described by the individual pattern (Forest et al., 2001). To create an orthonormal set of basis functions, each eigenfunction is normalized by its singular value (i.e., square root of eigenvalue). With the internal variability patterns defined, the model and observed patterns are rotated into the coordinate space defined by the basis functions. This rotation serves to maximize the signal-to-noise ratio in each pattern by transforming temperature patterns along the main components of internal variability into coordinates in the direction of less internal variability. By maximizing the signal-to-noise ratio, small patterns observed in the rotated trends are interpreted as a temperature variation and not erroneously discarded as noise.

The $\mathbf{x} - \mathbf{y}$ pattern in the rotated space is compared to each of the variability patterns and each evaluation contributes individually to the total r^2 value. To avoid rotating into a direction with infinite variance (i.e., dividing by singular values near

zero), the low variance patterns are filtered out of the pattern through a reduction in the number of eigenvalues and eigenfunctions retained in the calculation. The small magnitude singular values associated with the last temperature patterns result in division by a small number and cause the calculation to approach infinity. However, the removal of these patterns can be justified by remembering that they contribute minimally to the total variability and represents a truncation of the empirical orthogonal function decomposition.

Before concluding this section, potential causes of mismatches between models and observations presented in Hegerl et al. (2000) are discussed. Differences can be attributed to both observations and models. As previously mentioned, patterns of internal variability associated with non-linear dynamics are a component of any observed climate record. If the internal variability is large enough, climate change signals are masked by these patterns and may go undetected. A second observational contribution to the mismatch is attributed to the distribution and density of recording stations, as it contributes to errors in the climate record. Given the spatial and temporal inhomogeneities of the observations, regions without data must be filled in through interpolation. Lastly, instrumental errors affect the temperature values obtained from observations and contribute to differences that are present. All three components of observational error are present in all climate records and can contribute to the differences between the observations and climate model simulations.

Given that AOGCMs are approximations of reality, the internal variability derived from them is also an approximation and may not reproduce that of the true system. Any estimation of the internal variability derived from models is therefore likely to be different than that derived from observations. The construction of a climate model requires that assumptions be made in how the system behaves. As a result, the underlying physics and feedbacks that govern the model may not exactly mimic the behavior of the true climate system. However, given the current state of climate models, the approximations are close to reality. Because structural uncertainties still remain, they can lead to differences in temperature patterns produced by models compared to what has been observed. Lastly, uncertainties and errors in the radiative forcing scheme of the model will lead to differences between model output and observations. The climate is driven by these forcings,

and if they do not accurately represent reality, even the best climate models will differ from the historical record. These uncertainties in the model representation of the true climate system will also contribute to the differences between observations and climate model simulations.

1.4 Detection and Attribution

In addition to determining likely values of climate model parameters, the climate science community has actively searched for the causes of temperature change observed in the past century (Tett et al., 1999, 2002; Allen et al., 2000; Stott et al., 2000; Stott and Kettleborough, 2002; Hegerl et al., 2007). In particular, the impacts of anthropogenic sources of climate change are of interest. Tett et al. (1999) identify the main radiative forcings to be those resulting from well-mixed greenhouse gases, tropospheric aerosols, changes in solar irradiance, and stratospheric aerosols due to volcanic activity. Of these, greenhouse gases and tropospheric aerosols are classified as anthropogenic sources and solar irradiance and volcanic aerosols are classified as natural sources. Additional anthropogenic sources of temperature change that have been identified include stratospheric ozone depletion (Stott et al., 2006) and land-use changes (Cox et al., 2000).

In detection and attribution studies, linear combinations of the temperature trends that arise from natural and anthropogenic climate forcings are tested to determine which best match the past climate record. The study of Hegerl et al. (1997) marks the first application of optimal fingerprint techniques to the detection and attribution of climate change using simulated climate change from AOGCMs. Through the use of fingerprint signals, the study investigates the likelihood that forcings due to greenhouse gas emissions, sulfate aerosols, and solar variations will produce temperature patterns consistent with the observed record. The multi-fingerprint method utilized in the study allows for the combination of signals from more than one source to be used in an attempt to explain the recent climate record.

Tett et al. (1999) find that in the early 20th century, natural variability alone could account for the observed temperature change, but a combination of both natural and anthropogenic forcings could also account for the observed change. However, it is also shown that the recent trends in warming cannot be matched

unless the anthropogenic signal is included. This result is in agreement with Hegerl et al. (1997). Tett et al. (2002) also determine that internal variability and natural forcings alone cannot account for the observed temperature change between 1946 and 1996. Furthermore, Stott et al. (2000) show that both the global mean response and some large-scale features are also matched when the anthropogenic forcing pattern is included.

Stott and Kettleborough (2002) use detection and attribution studies to show the effect that mitigation strategies may have on future climate response. It is shown that under the forcing scenarios presented by the Intergovernmental Panel on Climate Change (IPCC), anthropogenic forcings lead unequivocally to a rise in global mean temperature. A summary of detection and attribution studies is presented in Chapter 9 of the IPCC's Fourth Assessment Report (IPCC AR4) (Hegerl et al., 2007). Similar to the studies addressed above, the IPCC AR4 agrees that it is very likely that recent climate change cannot solely be attributed to internal variability. None of the models used in the report have matched the warming trend of the past century when anthropogenic forcings are omitted. Furthermore, detectable anthropogenic signals of climate change have been found on every continent with the exclusion of Antarctica. This exclusion is present only due to the currently poor data coverage over the continent (Hegerl et al., 2007).

The recognition that anthropogenic forcings contribute to observed surface temperature changes has motivated its inclusion in this study. In particular, the anthropogenic aerosol forcing from sulfate aerosols is an adjustable parameter in the model and can be estimated in this study. The magnitude of this forcing plays a critical role in the representation of past and future climate change through its impact on the net radiation balance of the climate system.

1.5 Transient Climate Response

The behavior of the global climate system can be explained with the aid of a simple energy balance model and a discussion is presented in Sokolov et al. (2003). The governing equation of the model is:

$$C \frac{\partial \Delta T}{\partial t} = F(t) - \lambda \Delta T(t) + \phi(t), \quad (1.2)$$

where C is the heat capacity of the climate system, ΔT is the change in global mean surface temperature, $F(t)$ is the net radiative forcing, λ is a feedback parameter, and $\phi(t)$ is the flux of heat into the deep ocean below the oceanic mixed layer. For the climate system, the heat capacity is mainly that of the mixed layer of the ocean and the associated mass defines C . This is due to the relatively large heat capacity of water compared to air and lack of mixing between the mixed layer and deep ocean. The feedback parameter is a single number that is characteristic of the impacts that the active processes in the climate system have on the temperature of the system. Furthermore, the feedback parameter determines the equilibrium sensitivity of the system.

When $F(t)$ is taken to be the additional forcing instantaneously introduced by a doubling of carbon dioxide (CO_2) concentrations, the corresponding temperature change when the system is brought into equilibrium is defined as the equilibrium climate sensitivity. When the system is in equilibrium, the left-hand side of Eq. (1.2) is zero, the heat flux into the ocean is zero, and the corresponding equilibrium temperature change and climate sensitivity of the system is $\Delta T_{eq} = \frac{F_{2XCO_2}}{\lambda_{eq}}$.

If $F(t)$ is taken to be the forcing when CO_2 concentrations are increased at the rate of 1% per year instead of that due to an instantaneous doubling of CO_2 , an effective climate sensitivity, λ_{eff} , can be derived through inversion of Eq. (1.2). At any point in the simulation, all quantities in Eq. (1.2) can be estimated from the model time series. The effective climate sensitivity corresponds to the equilibrium climate sensitivity at any given time in the simulation for the set of active feedbacks in the system. Transient climate simulations include a set of active feedbacks in the climate system that contribute to the change in temperature under idealized or more realistic forcing scenarios (Murphy, 1995). In simulations utilizing this specific forcing scenario, the transient climate response (TCR) is taken to be the temperature change at the time of CO_2 doubling, the average of the 20 year period surrounding year 70.

Through its connection to the feedback parameter, it is clear that the climate sensitivity of the system plays a key role in the transient climate response. What may not be clear is that the rate of ocean heat uptake also plays a large role in transient simulations. The rate of ocean heat uptake is a measure of how well heat advects and diffuses into the ocean below the climatological mixed layer

(Stouffer et al., 2006). If the rate of advection and/or diffusion, and thus rate of heat uptake, is slow, heat is not able to easily cross the interface between the mixed layer and deep ocean. As a result, heat is not transported as efficiently into the deep ocean. This lack of heat removal leads to increasing temperatures in the atmosphere and mixed layer. Because the mixed layer serves as the main component of the heat capacity of the system, changes in mixed layer behavior and interaction with the deep ocean drastically impact the thermal inertia of the climate system and subsequent climate behavior.

Using a box diffusion model, Hansen et al. (1984) first showed that equilibrium climate sensitivity and ocean heat uptake are correlated quantities. The study showed that the ocean relaxation time depended on the feedback factors that define the climate sensitivity. If it assumed that ocean relaxation time is also dependent upon the rate of ocean heat uptake, a positive correlation between climate sensitivity and ocean heat uptake can be inferred. Thus, as the climate sensitivity increases, so too does the rate of ocean heat uptake. The assumed dependence of ocean relaxation time on the rate of ocean heat uptake is justified through the argument that the rate of ocean heat uptake determines the heat capacity of the climate system, and the heat capacity of the climate system determines the longterm behavior of the system. Under the assumption of constant forcings, if the climate sensitivity were to increase, surface and mixed layer temperatures would tend to increase more than if the climate sensitivity were lower. However, in order for temperatures to remain consistent with the observed temperature change, a portion of the additional heat must be removed from the mixed layer so that temperatures do not increase too much and become inconsistent with the observations. As a result, heat must be transported from the mixed layer to the deep ocean. To obtain the necessary heat transfer, the ocean diffusivity, and subsequent ocean heat uptake value, must be greater, showing that the increased climate sensitivity led to the requirement of greater ocean heat uptake. A similar argument can be constructed for when climate sensitivity values are decreased as well.

Andrews and Allen (2008) frame the same argument in terms of effective climate sensitivity by exploring the relationship between effective climate sensitivity, ocean heat uptake, and TCR. In the study, ocean heat uptake is related to the feedback

response time of the system. This definition is based on the argument that the rate of ocean heat uptake defines the heat capacity of the system and thus sets the time it takes for the system to respond to changes in external forcing. In their Figure 2b, contours of constant effective climate sensitivity are drawn as a function of TCR and feedback response time. The contours show that for a fixed TCR, if effective climate sensitivity is to increase, the feedback response time must also increase. Because the feedback response time is proportional to the heat capacity, and the heat capacity is determined by the rate of ocean heat uptake, the study shows that effective climate sensitivity and ocean heat uptake are correlated quantities. Both the physical argument presented above and the study of Andrews and Allen (2008) lead to the conclusion that climate sensitivity and ocean heat uptake must be correlated.

Given that TCR is a function of climate sensitivity and ocean heat uptake, uncertainties in the two properties propagate as uncertainties in TCR. Despite knowing that climate sensitivity and ocean heat uptake are correlated, it has been shown that large uncertainties in TCR still remain due to uncertainties in climate sensitivity (Raper et al., 2002). If an understanding of future climate change is to be gained, more reliable projections of TCR must be obtained. Current estimates from the IPCC AR4 put TCR estimates between 1 and 4 K (Hegerl et al., 2007).

At this point it should be mentioned that surface temperature change is not the only metric of future climate change. Additional metrics of climate change include changes in precipitation, sea-ice coverage, the strength of the Atlantic Meridional Overturning Circulation, and sea-level rise. Although reducing TCR uncertainty may not lead to decreased uncertainties in other climate change metrics, the need for reliable TCR projections motivates the inclusion of estimates for the parameters that define it, climate sensitivity and ocean heat uptake, into this study. Furthermore, this study explores the sensitivity of TCR projections to the surface datasets used to evaluate model performance and derive the parameter distribution functions.

Chapter 2

Surface Datasets

One diagnostic used to estimate distributions of model parameters in this study is the set of surface temperature trends in four equal-area zonal bands. It can be argued that surface temperature change, ΔT , provides the best dataset possible for the comparison of model output and observations. Surface temperature records provide a relatively long dataset, allowing for longer time series to be used in model evaluation studies. Additionally, the large number of stations that have long surface temperature time series allow for greater spatial coverage of the record than other potential datasets. Currently, estimates of past surface temperatures are available from multiple sources. This study will utilize estimates from three research groups that are considered to provide the best datasets available. The high quality of the data products produced by the research groups have led to their inclusion in the IPCC AR4.

In total, five different temperature reconstructions are used. The first two reconstructions are from the Hadley Centre for Climate Prediction and Research and represent their second (Jones and Moberg, 2003) and third (Brohan et al., 2006) versions. These datasets will be referred to hereafter as HadCRUT2 (version 2) and HadCRUT3 (version 3). The third reconstruction is from NOAA's National Climatic Data Center (Smith et al., 2008) and will be referred to hereafter as NCDC. The remaining two reconstructions are from NASA's Goddard Institute for Space Studies (Hansen et al., 2010) and will be referred to hereafter as GISTEMP 250 and GISTEMP 1200. The significance of the 250 and 1200 designations for these datasets will be made clear later in this section. Each dataset reports monthly

surface temperatures as anomalies on a 5x5 degree grid with respect to a given base period. HadCRUT anomalies are with respect to a 1961-1990 base period, NCDC anomalies are with respect to a 1971-2000 base period, and GISTEMP anomalies are with respect to a 1951-1980 base period. While the goal of each group is reconstructing the ΔT time series, the estimations differ from one another. The potential reasons for these differences are discussed below because they impact our ability to estimate climate model parameter values.

The first major difference between the reconstructions is which data records are used in the analysis. All of the reconstructions obtain a majority of their land surface data from the Global Historical Climatology Network (GHCN) from the National Climatic Data Center (Peterson and Vose, 1997). The GHCN is an international effort to provide, maintain, and make available historical surface temperature data. Each of the reconstructions in this study utilize the available data differently. For example, the Hadley Centre requires that stations have sufficient data between 1961 and 1990 in order to be used in the analysis (Jones and Moberg, 2003; Brohan et al., 2006). The choice of ocean surface data also differs between the datasets. Because oceans cover 70% of the Earth's surface, this difference plays a crucial role in causing the differences between the temperature reconstructions (Smith et al., 2008). For HadCRUT2, ocean data from the HadISST1 (Rayner et al., 2003) sea surface temperature reconstruction is used. HadSST2 (Rayner et al., 2006), an update to HadISST1, is used in HadCRUT3 surface temperature estimates. The most recent ocean dataset from the National Climatic Data Center, ERSST version 3b, is used and described in the description of the NCDC surface temperature reconstruction (Smith et al., 2008). GISTEMP temperature reconstructions use HadISST1 data from 1880-1981 and switch to satellite sea surface temperature data (Reynolds et al., 2002) from 1981 to the present. In a test of the sensitivity to ocean data choice, Hansen et al. (2010) show that the GISTEMP temperature reconstructions are affected by the choice of sea surface temperature data. Slightly greater warming trends in recent decades and on the century time scale were observed in the GISTEMP reconstructions when using HadSST2 or ERSST ocean data as opposed to the combined HadISST1 and satellite data currently used.

The second major difference between the surface datasets are the interpolation

and averaging methods used to fill in missing data regions and create the 5x5 degree grid boxes. HadCRUT2 and HadCRUT3 data use the climate anomaly method (Jones et al., 1982) to create the 5x5 degree temperature anomalies. Surface temperature anomalies with respect to the 1960-1990 base period are calculated for each station used in their reconstructions. These stations are carefully selected and include only those with sufficient data in the climatology period. Once the time series of anomalies for each station are calculated, the value for a given grid box is taken to be the average of all of the stations contained in that grid box. NCDC uses the methodology of Van den Dool et al. (2000) to interpolate temperature patterns between observation stations. The interpolation is carried out separately for temperature anomalies over land using land surface data and sea surface temperature anomalies using ocean data. Using the observed data, large and small scale temperature patterns are identified. By weighting these patterns by the amount of variance they explain, regions without data are filled in by taking combinations of the base patterns to identify anomalies for each grid box. The complete reconstruction is determined through a blending of the land and ocean anomalies. GISTEMP calculates temperature anomalies in each of 80 equal-area grid boxes. Each larger box is divided into 100 sub-boxes, and the anomaly for each sub-box is taken to be the weighted average of all stations within a predefined radius of influence (Hansen and Lebedeff, 1987). The weight given to a station in the average is a linear function of the distance between the point and the station. An area-weighted average of the sub-boxes gives the anomaly for each of the larger grid boxes. Data are currently available with 250 and 1200 km radii of influence, and these radii are reflected in the naming of the GISTEMP datasets in this study. Using publicly available FORTRAN codes from their website, the 80 equal-area box reconstruction can be interpolated onto the same 5x5 degree grid used by the other three datasets. A summary of the differences between the reconstructions is given in Table 2.1.

Owing to the different data choices and interpolation methods, the five surface temperature reconstructions differ from each other. Due to the data requirement used in the climate anomaly method, HadCRUT reconstructions have less spatial coverage in the polar regions when compared to the NCDC and GISTEMP reconstructions (Hansen et al., 2010). The polar coverage differences mark the biggest

discrepancies between the datasets. Given these differences in the reconstructions, this study is geared towards understanding the impacts that the choice of dataset has on the parameter calibrations.

Table 2.1. Summary of differences in surface temperature datasets

Dataset	Ocean Data	Land Data	Grid Box Averaging Method
HadCRUT2	HadISST1	GHCN Network - restricted to those with at least 20 years of data in the 1960-1990 climatology period	Climate anomaly method (average of selected stations in each grid box)
HadCRUT3	HadSST2	GHCN Network - restricted to those with at least 20 years of data in the 1960-1990 climatology period	Climate anomaly method (average of selected stations in each grid box)
NCDC	ERSST	GHCN Network - subjective choice of stations	Infilling of data by variance-weighted average of observed temperature patterns
GISTEMP 250	HadISST1 (1880-1981), OISST.V2 (1981-present)	GHCN Network - subjective choice of stations	Weighted average of stations within a 250 km radius of influence in each of 80 equal-area grid boxes
GISTEMP 1200	HadISST1 (1880-1981), OISST.V2 (1981-present)	GHCN Network - subjective choice of stations	Weighted average of stations within a 1200 km radius of influence in each of 80 equal-area grid boxes

Chapter 3

Methods

Following the work of Forest et al. (2002, 2006, 2008), this study estimates the joint distribution function of climate model parameters for climate sensitivity (S), effective ocean diffusivity (K_v), and net anthropogenic aerosol forcing (F_{aer}). Using the climate system portion of MIT's Integrated Global Systems Model (Sokolov and Stone, 1998; Sokolov et al., 2005), the model simulates historical temperature responses given choices of the three climate model parameters. Climate sensitivity in the model is adjusted by changing the strength of the cloud feedback parameter at different levels in the atmosphere. Through these changes, the amount of clouds in the radiative transfer scheme is altered, and the resulting climate system behavior adjusts accordingly. Model effective ocean diffusivity describes the vertical diffusion of heat anomalies below the mixed layer into the deep ocean and tends to be larger than typical ocean diffusivity values which describe the diffusion of heat (Sokolov et al., 2003). The strength of the diffusion coefficient sets the rate of ocean heat uptake. This rate of ocean heat uptake is what was described in connection with the transient climate response. The anthropogenic aerosol forcing used in the model is prescribed spatially by a latitude-dependent pattern and differs over land and ocean. Additionally, the pattern amplitude varies with time based on the estimated anthropogenic emissions of sulfur dioxide, with weaker amplitudes at the beginning of the simulation and progressively larger amplitudes as time progresses. These spatial and temporal patterns can be found in Forest et al. (2001). In the model, the net anthropogenic aerosol forcing parameter sets the strength of the aerosol forcing in 1986, the time when the pattern amplitude

is equal to 1. This is representative of the forcing strength for the decade of the 1980s.

In this study, the parameter space is sampled by varying climate sensitivity between 0 to 8 K, effective ocean diffusivity between 0 to $25 \text{ cm}^2\text{s}^{-1}$ and anthropogenic aerosol forcing between -1.5 and 0.5 Wm^{-2} . Negative anthropogenic aerosol values represent a cooling effect on the climate and positive values represent a warming effect. In total, 640 different parameter combinations are used. Each model run is forced by historical records of greenhouse gas concentrations, sulfate aerosol loadings, tropospheric and stratospheric ozone concentrations, solar irradiance changes, and stratospheric aerosols from volcanic eruptions (Forest et al., 2008).

Model performance under a given set of parameter values is evaluated through comparison of model output to historical data using diagnostics described in Forest et al. (2006, 2008). The first diagnostic used is surface temperature anomaly patterns for four zonal bands on a decadal time scale from 1946-1995 using a reference climatology of 1906-1995. The second diagnostic is the height-latitude pattern of upper-air temperature differences between the periods 1961-1980 and 1986-1995. The third diagnostic used is the trend in ocean heat content for the 0-3 km layer from 1948-1995. Surface temperature diagnostics are calculated using four equal-area zonal bands from 90 to 30°S , 30 to 0°S , 0 to 30°N , and 30 to 90°N . Upper-air temperature diagnostics are calculated using eight vertical layers in the atmosphere on a 5 degree latitude grid (excluding missing values). Ocean heat content diagnostics are calculated on a global scale. The set of diagnostics for this study was chosen to match those used in Forest et al. (2008).

For surface and upper-air temperature diagnostics, an estimate of the unforced variability of the climate system is determined from the control run of the Hadley Centre's coupled atmosphere-ocean global circulation model, HadCM2 (Johns et al., 1997). For the surface temperature and upper-air temperature diagnostics, the internal variability estimate is represented by the noise-covariance matrix. This matrix calculates the spatial and temporal correlations that are present in an unforced model scenario and represents an estimate of the internal variability found in the climate system. Rotating the model and observed temperature patterns into the coordinate space defined by the internal variability serves

to maximize the signal-to-noise ratio in an attempt to allow for the detection of small patterns in the trends. With the rotation, the goodness-of-fit statistic, r^2 , discussed in Section 1.3 is calculated for each model run for each of the diagnostics and gives a measure of how consistent the modeled trends are with observed trends. Because the ocean diagnostic uses the linear trend in a global quantity, there are no spatial correlations between zonal bands to be calculated. As a result, the noise-covariance matrix for the ocean diagnostic reduces to an estimate of the standard deviation of the linear trend values taken from control run data and is a single value. This value is estimated from the control run of the GFDL R30 model (Delworth et al., 2002) and takes the place of \mathbf{C}_N^{-1} in Eq. (1.1). Through an application of Bayes' Theorem (Bayes, 1763), information from the individual diagnostics are combined to give a full evaluation of the model parameters.

Determination of the surface diagnostic requires the derivation of decadal mean temperature anomalies from the monthly observational datasets for each of the four zonal bands. Zonal mean decadal temperatures are derived by first calculating yearly and decadal average anomalies for each of the points on the 5x5 degree grid, then averaging the temporal averages spatially within each latitude band. Due to data coverage issues, coverage thresholds must be placed on the spatial and temporal averages to avoid one data point significantly influencing the time series. When annual averages are taken, eight months within the given year must be defined; otherwise, the grid point is marked as missing data. This threshold is a similar, yet slightly more stringent requirement than that used in Thorne et al. (2002), which requires two months of data in at least three of the four seasons. Once the annual average is calculated, the decadal average for each grid box is calculated with the requirement that at least eight years must contain defined data. With decadal averages for all grid points, the longitudinal average for a given latitude band is calculated with the requirement that at least 20-percent of the zonal band is covered. Finally, the time series for the four zonal bands used in the analysis are calculated by taking a mass-weighted average of the zonal bands contained within the diagnostic band.

Chapter 4

Results

4.1 Surface Temperature Time Series

Using the averaging techniques for deriving decadal mean temperatures described in Chapter 3, time series of the temperature trends used for the surface diagnostics are plotted (Fig. 4.1). It is important to note that the HadCRUT2 data presented here is different than that used in Forest et al. (2008). This study uses the monthly version of the HadCRUT2 dataset to allow for identical treatment of data as for the other four datasets. In general, the patterns in each zonal band are similar, with the sign of the temperature change consistent across a majority of the decades for each dataset. However, the magnitudes of these temperature changes differ amongst each other. In particular, agreement in the trends is weakest in the Southern Hemisphere, particularly in the poleward latitude band. These results follow from the discussion presented in Chapter 2. Because each reconstruction uses different data and interpolation schemes, the temperature patterns cannot be expected to be the same. Furthermore, the observation that Southern Hemisphere patterns show the largest discrepancies is not surprising given that sea surface temperature datasets differ between the reconstructions and a greater fraction of the Earth's surface is ocean in the Southern Hemisphere.

Another significant difference is the overall temperature trend observed during the period used in the surface diagnostic, 1946-1995. In all zonal bands, GISTEMP datasets show either similar or weaker warming trends when compared to the other datasets. This is most evident in the 30-90 °S zonal band, where in the first decade,

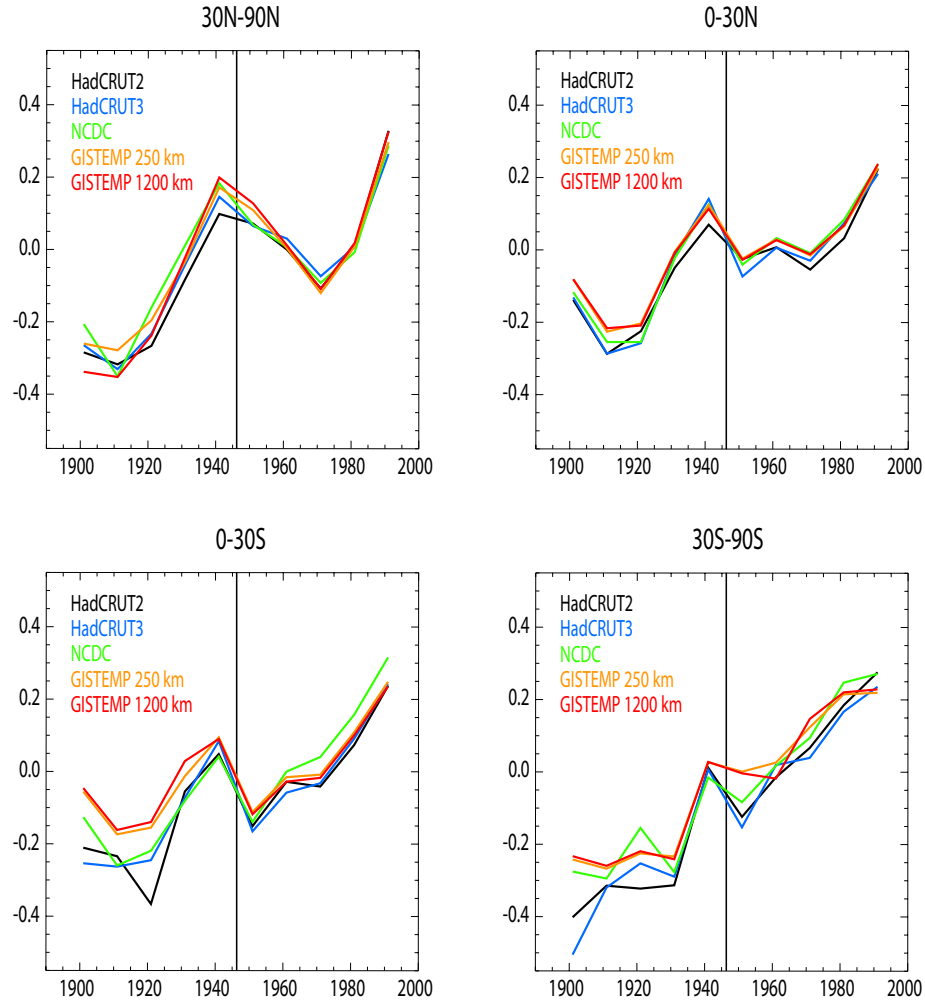


Figure 4.1. Surface temperature time series from 1896-1995 for each of the individual datasets. Only data beginning in 1906 is used in this study, but the time series begins in 1896 to show data from the same period that was available in Forest et al. (2008). Vertical lines mark the beginning of the period used in the surface diagnostic (1946).

GISTEMP data are by far the warmest, yet in the last decade, GISTEMP data are the coldest. Similar, yet weaker, patterns hold in the remaining zonal bands. In general, the NCDC time series yields the next weakest warming trends, followed by the HadCRUT2 and HadCRUT3 datasets, which yield similar overall trends. However, the extent of the differences is much less pronounced than with the GISTEMP datasets, and the rank order of the trends is not consistent across all zonal bands.

4.2 Parameter Diagnostics

With the surface time series determined, the temperature diagnostics for each model run are calculated. Each model run is identified by a unique set of the three parameter values, θ , where $\theta = \{S, K_v, F_{aer}\}$. The model output, $\mathbf{x}(\theta)$ is substituted in for \mathbf{x} in Eq. (1.1) and shows the dependence of the r^2 calculation on θ . At each aerosol forcing level, the r^2 values at a given climate sensitivity value and across all effective ocean diffusivity values are initially smoothed. This smoothing is accomplished by fitting a sixth-order polynomial to the eight data points defined when ocean diffusivity and aerosol forcing are fixed and climate sensitivity is varied over the eight values used when running the model. After the smoothing, the difference between the r^2 value for a given model, $r^2(\theta)$, and the global minimum for the diagnostic, r_{min}^2 , is calculated to give the Δr^2 value for each model run. The r_{min}^2 value for each diagnostic corresponds to the lowest r^2 value calculated across the set of θ s used in this study. Each Δr^2 value is then converted into an estimate of the likelihood that the given model yields results consistent with the observations through evaluation of an F-statistic for each value. The likelihood of a given set of model parameters is calculated by finding the probability that the Δr^2 value from the distribution could be greater than the calculated Δr^2 value for the model. For the surface and ocean diagnostics, the F-distribution used to calculate the statistic has 3 and 24 degrees of freedom, and for the upper-air diagnostic, the F-distribution has 3 and 12 degrees of freedom. The assumptions for using the F-statistic are satisfied because the rotation of the observed and modeled temperature patterns when calculating the r^2 values assures that the errors are identically and independently distributed.

Once the likelihood values have been calculated for each model run, the climate sensitivity-effective ocean diffusivity parameter space is filled by least-squares quadratic interpolation of the smoothed data. Figs. 4.2-4.9 show the resulting likelihood distributions at each of the aerosol levels. At each level the 90- (white) and 99-percent (light blue) confidence intervals are shown. The 90-percent confidence regions mark the strongest acceptance of a given set of parameter values. Regions outside of the 90-percent confidence interval are rejected at the 10-percent significance level for being inconsistent with the observed data, indicating that the

parameter values have a less than 10% chance of yielding results consistent with the data. Similarly, regions outside of the 99-percent confidence region are rejected at the 1-percent significance level for being inconsistent with the observed data, indicating that the parameter values have a less than 1% chance of yielding results consistent with the data. As a result, parameter combinations in dark blue regions (outside of the 99-percent confidence intervals) are rejected for being inconsistent with the data.

Regardless of the observational surface dataset choice, the upper-air and ocean temperature diagnostics are the same in each analysis. These diagnostics are also the same as in Forest et al. (2008). For the strongest positive aerosol forcing that the model was run with, $+0.50 \text{ Wm}^{-2}$, the upper-air diagnostic yields parameter combinations that are accepted at the 90-percent confidence interval for all effective ocean diffusivity values when the climate sensitivity is between 1 and 2.5 K. Additionally, a slight positive trend in the 99-percent confidence interval can be seen, that as the ocean diffusivity increases, so too does the climate sensitivity. As the aerosol forcing becomes less positive, the positive trend between effective ocean diffusivity and climate sensitivity becomes more pronounced and the slope steeper. Additionally, the strong lower bound at 1 K remains intact across all diffusivity values. In the negative aerosol forcing levels, the upper-air diagnostic becomes progressively less informative due to the growth of the 90-percent confidence region. Moving towards more negative aerosol levels brings more regions in the climate sensitivity-effective ocean diffusivity space into the 90-percent confidence region. By the -0.75 Wm^{-2} aerosol level, only extreme combinations of low diffusivity and high climate sensitivity are rejected for being inconsistent with the data at the 1-percent significance level. At the most negative aerosol level, -1.5 Wm^{-2} , almost all parameter combinations fall outside of the 90-percent confidence interval and have a less than 10% likelihood of yielding results consistent with the observations.

The ocean diagnostic shows an initial increase in the size of the 90-percent confidence region through a changing of aerosol values from positive to negative values and rejection of all points at the largest negative aerosol forcing. When moving from the $+0.5$ to -0.5 Wm^{-2} aerosol levels, an increase in the size of the 90-percent confidence region is observed. This acceptance region initially corresponds to low climate sensitivity values and a wide range of effective ocean diffusivity

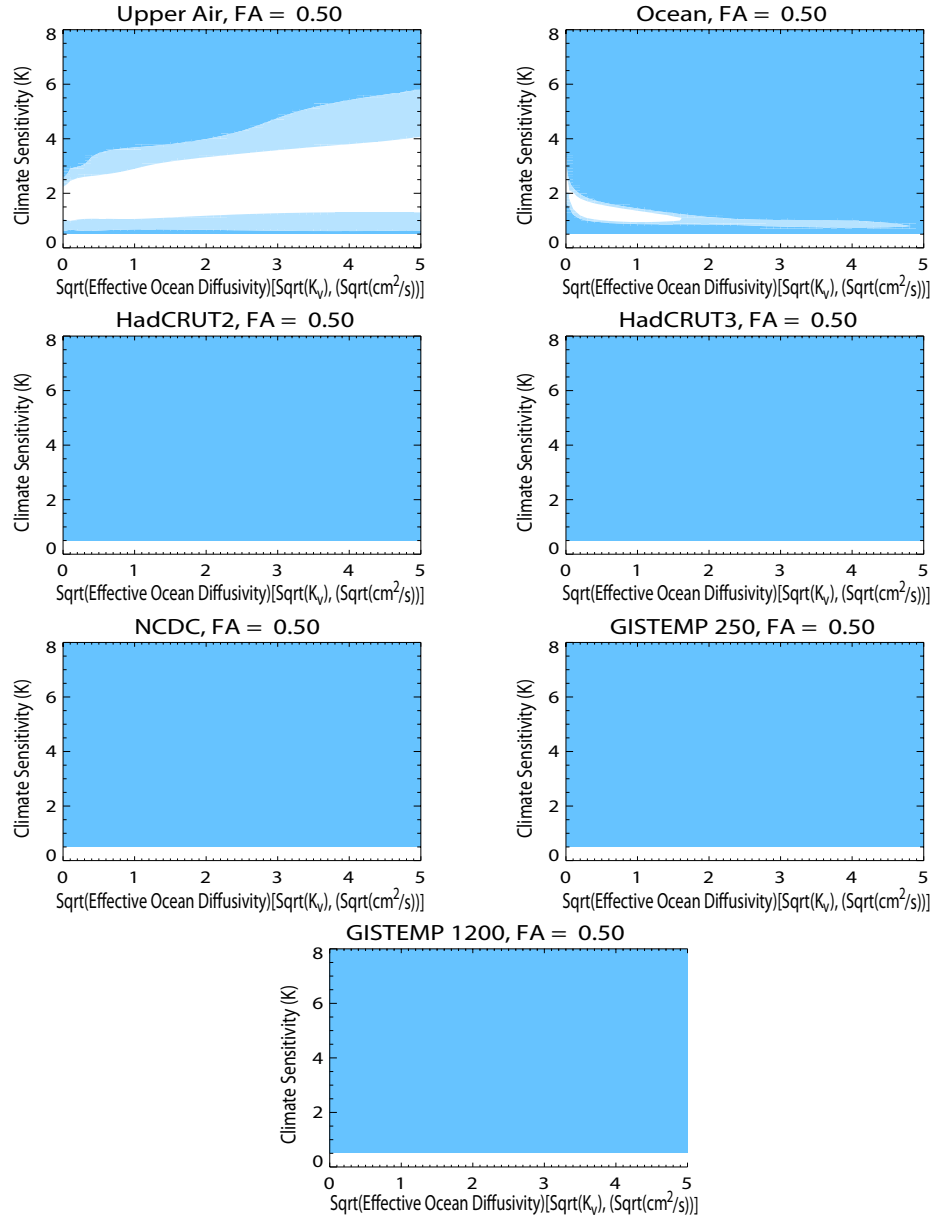


Figure 4.2. Distributions of r^2 for temperature diagnostics as a function of climate sensitivity and effective ocean diffusivity at an aerosol forcing (labeled as FA) of 0.50 Wm^{-2} . White shading represents the 90-percent confidence region, and light blue shading represents the 99-percent confidence region based on an F-test. Dark blue shading represents regions that are outside of the 99-percent confidence region and are thus rejected at the 1-percent significance level.

values. As the aerosol forcing becomes less positive in the $+0.5$ to -0.5 Wm^{-2} range, progressively larger regions of high climate sensitivity and low effective

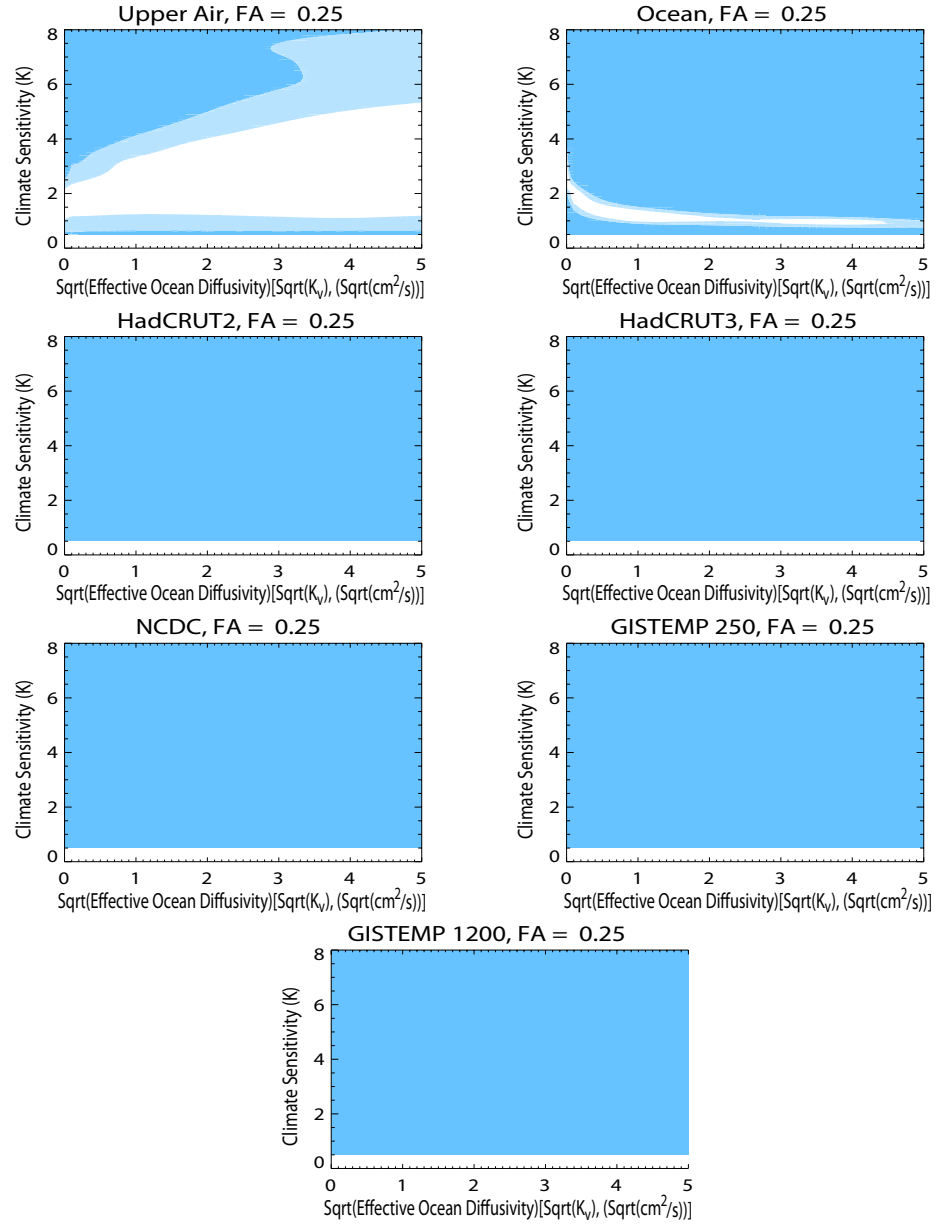


Figure 4.3. As in Fig. 4.2 but at an aerosol forcing of 0.25 Wm^{-2}

ocean diffusivity are consistent at the 99-percent level. The region accepted at the 90-percent level grows and shifts towards regions of higher climate sensitivity and higher effective ocean diffusivity at -0.75 Wm^{-2} aerosol forcing. By the -1.00 Wm^{-2} aerosol level, the acceptance region shifts to include only regions of high climate sensitivity and high effective ocean diffusivity. When aerosol forcing is

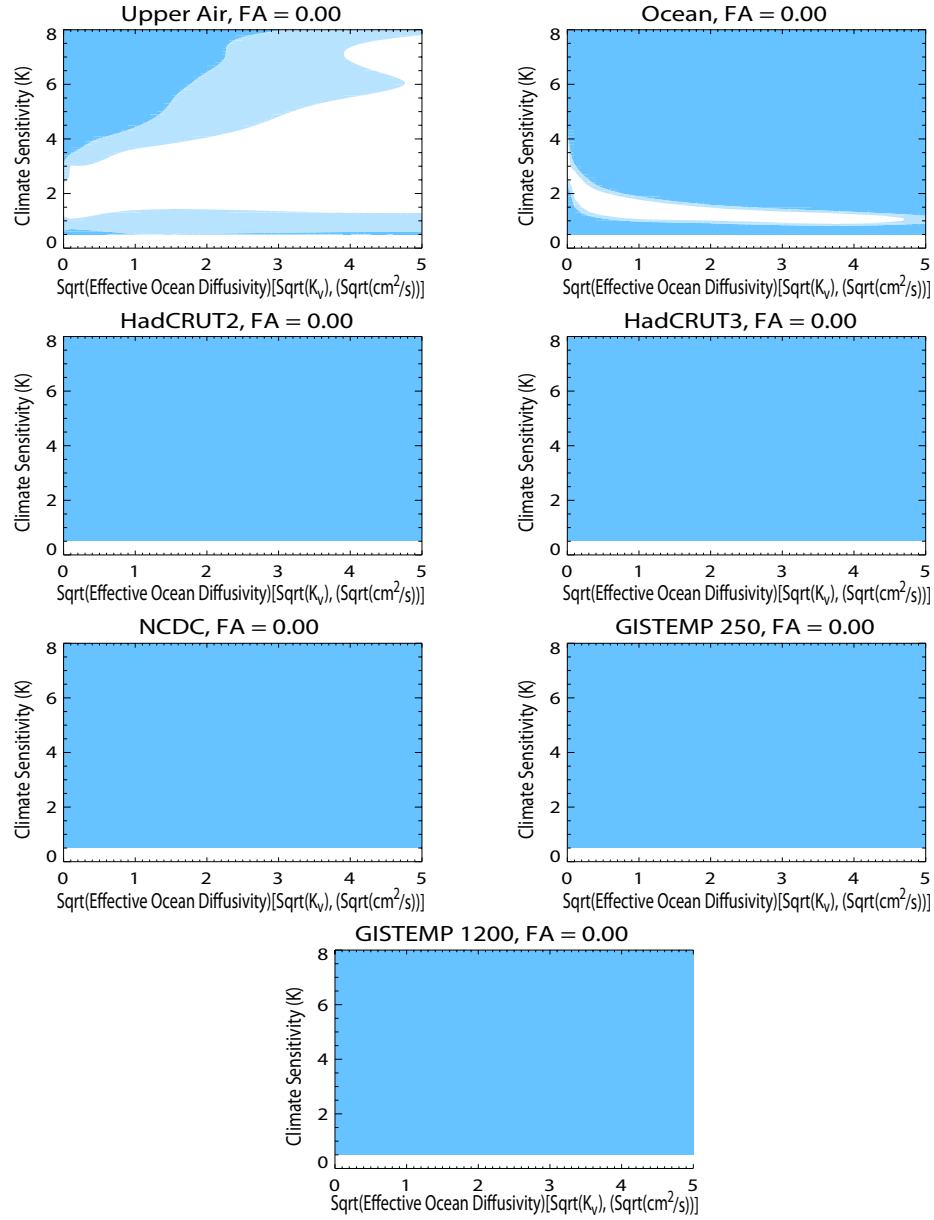


Figure 4.4. As in Fig. 4.2 but at an aerosol forcing of 0.00 Wm^{-2}

-1.50 Wm^{-2} , no regions of the parameter space are consistent with the ocean data.

For positive anthropogenic aerosol forcing values, all values are rejected for being inconsistent with the data by the surface diagnostic, regardless of which observational dataset is used. Additionally, all regions are rejected at zero aerosol forcing. As the aerosol forcing moves to -0.25 Wm^{-2} , regions of the parameter space

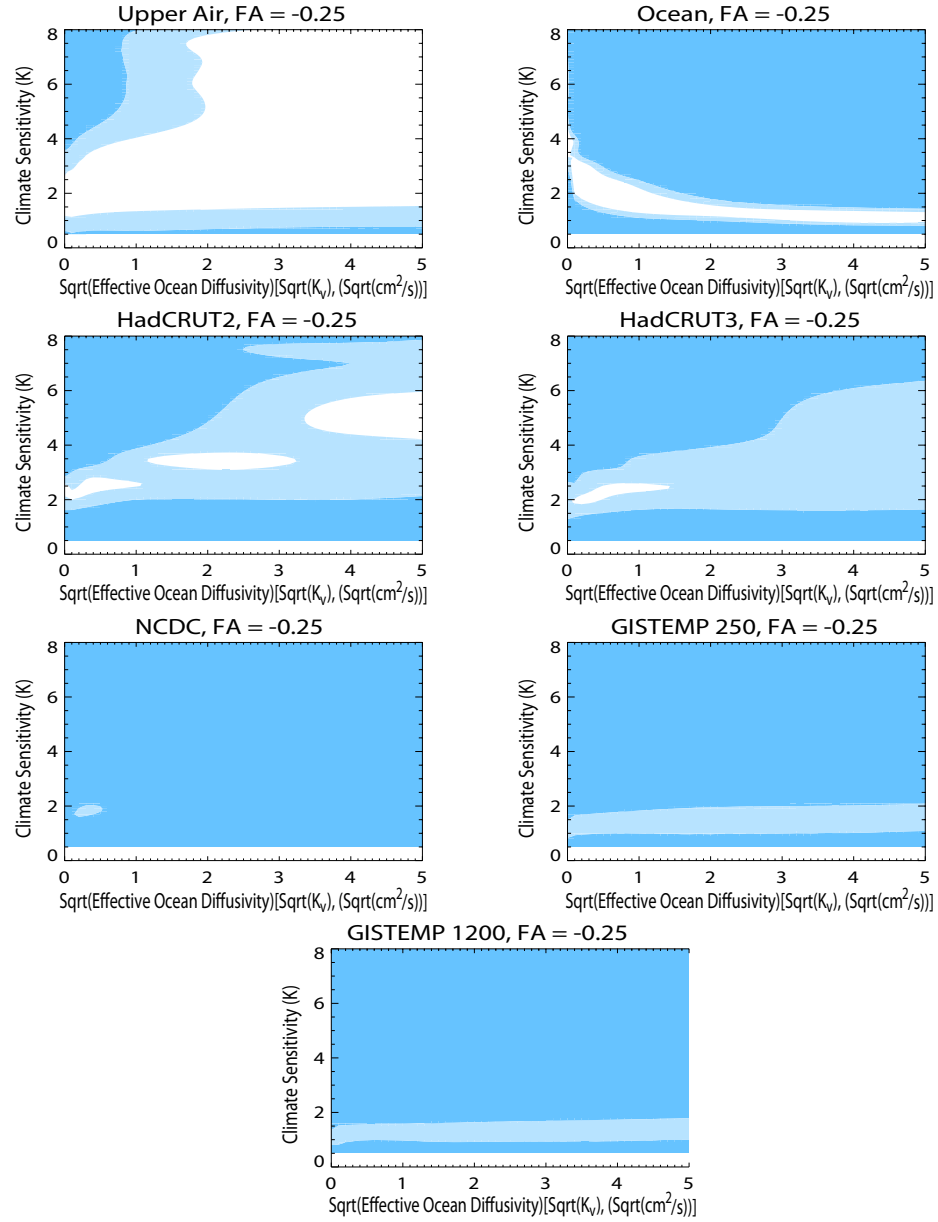


Figure 4.5. As in Fig. 4.2 but at an aerosol forcing of -0.25 Wm^{-2}

using the HadCRUT datasets are accepted at the 90-percent confidence interval, but no points in the NCDC or GISTEMP datasets are consistent with the data at this level. In both HadCRUT cases, a positive correlation between the climate sensitivity and effective ocean diffusivity values in the 99-percent confidence region can be seen.

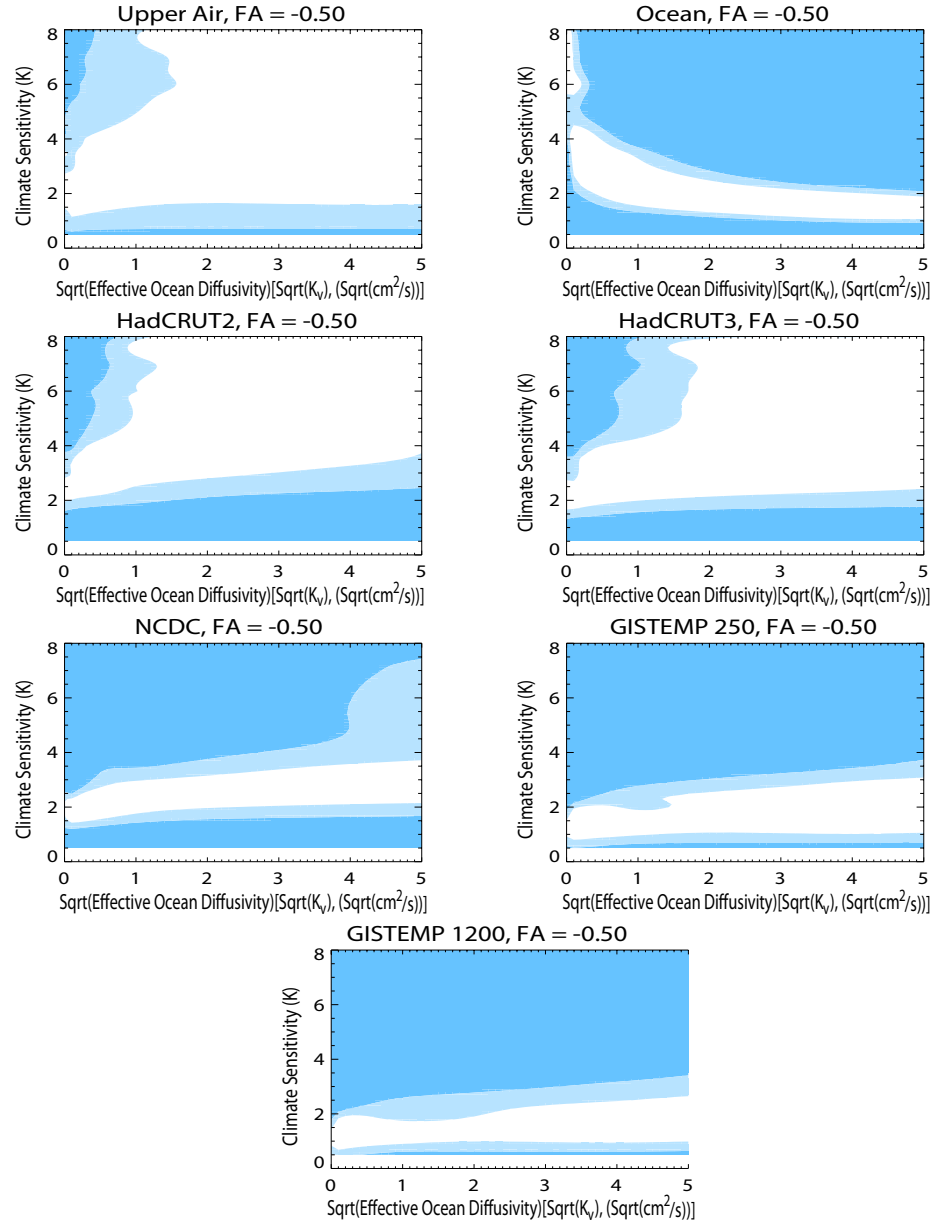


Figure 4.6. As in Fig. 4.2 but at an aerosol forcing of -0.50 Wm^{-2}

At the -0.50 Wm^{-2} level, substantial portions of the parameter space are accepted at the 90-percent confidence interval using the HadCRUT datasets. In each of these cases, only extreme regions of high climate sensitivity and low effective ocean diffusivity are rejected. The same general trend can be seen using the NCDC dataset, but a much larger region of the high sensitivity and low diffusivity quad-

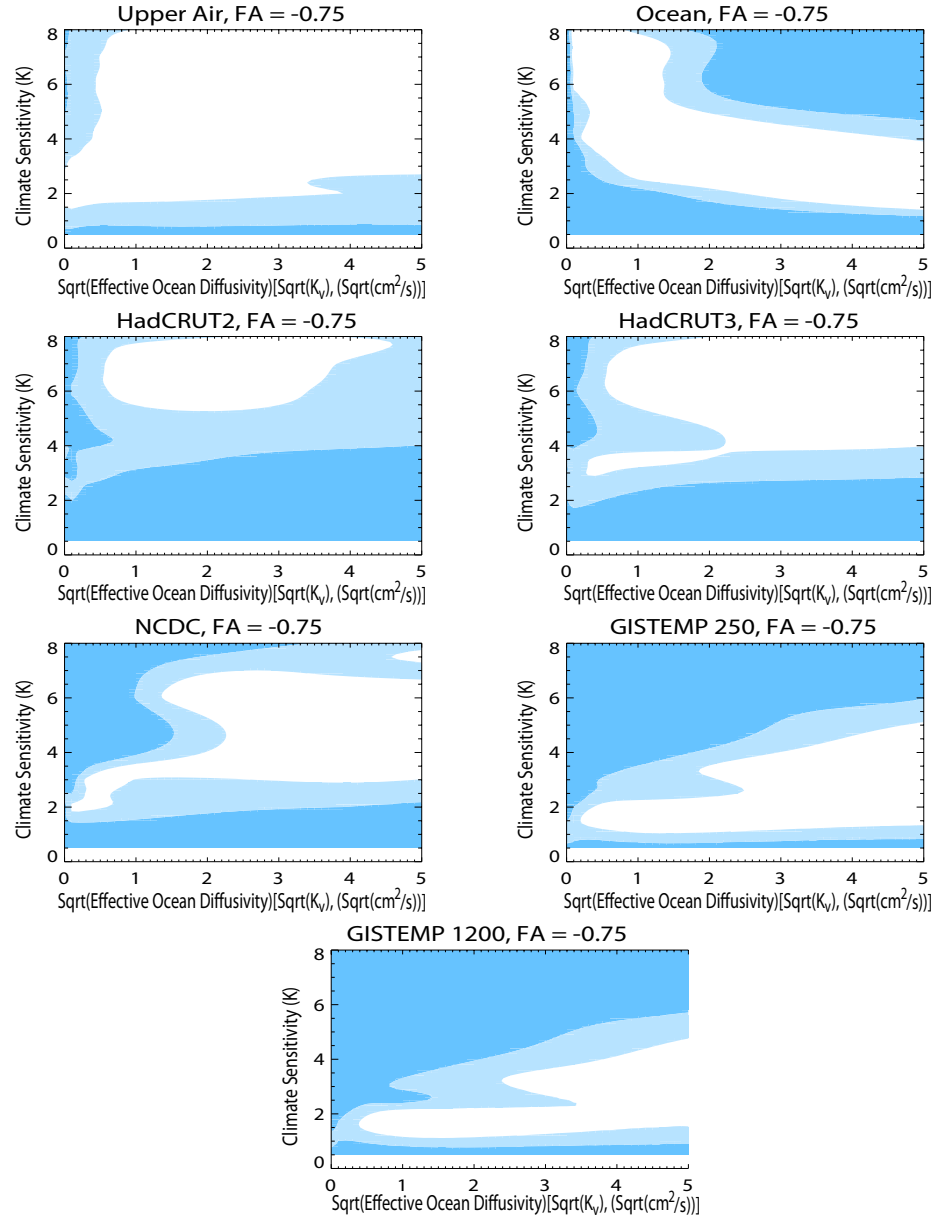


Figure 4.7. As in Fig. 4.2 but at an aerosol forcing of -0.75 Wm^{-2}

rant is rejected. At this level, both GISTEMP datasets reject all regions with climate sensitivities over 3 K, and only small portions of the parameter space are not rejected at the 10-percent level within the same region. Noticeable differences in the surface diagnostics are also present at the -0.75 Wm^{-2} aerosol forcing level. HadCRUT2 data are no longer rejected in the high sensitivity and low diffusivity

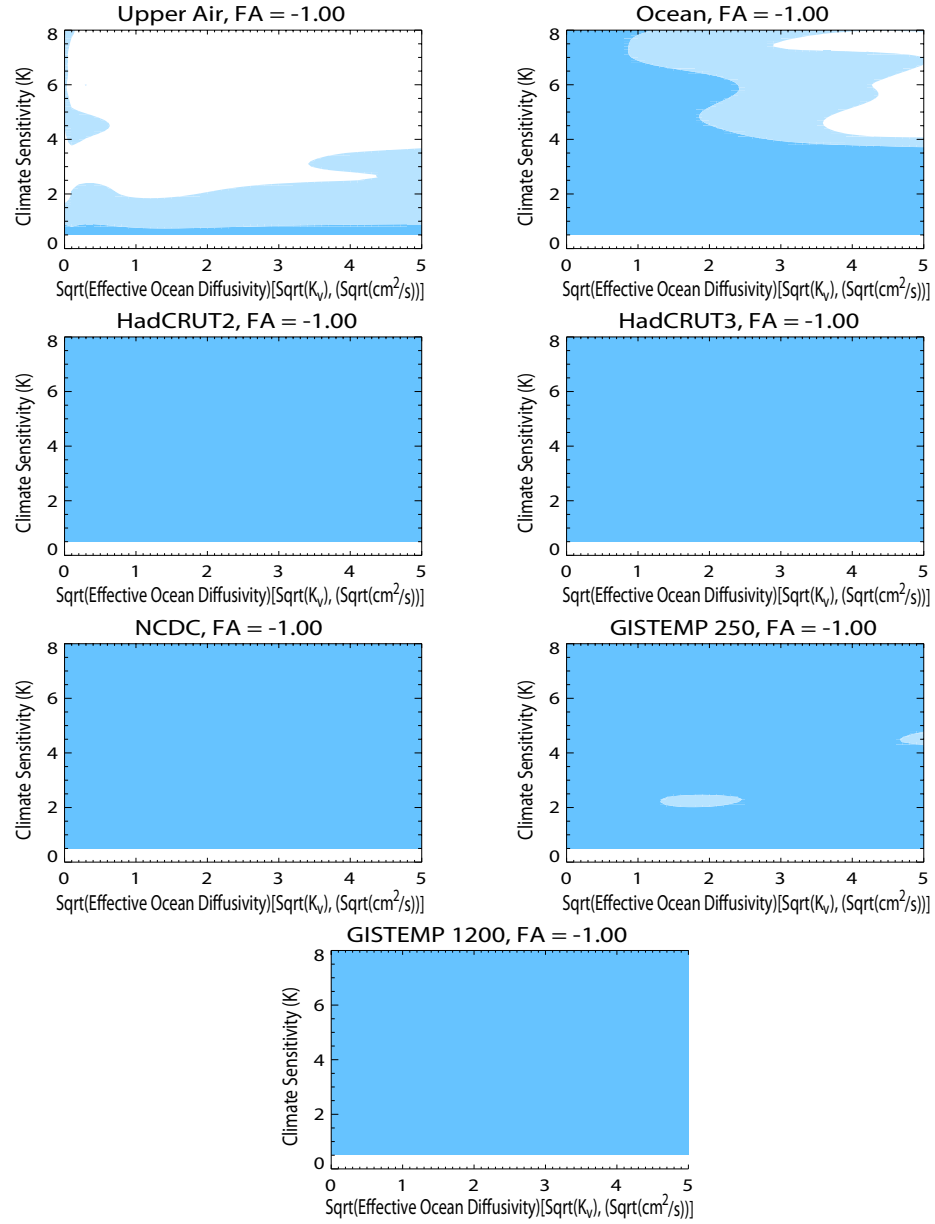


Figure 4.8. As in Fig. 4.2 but at an aerosol forcing of -1.00 Wm^{-2}

regions, and a large portion of the low climate sensitivity region is rejected for being inconsistent with the data. The upper bound of the rejection region for HadCRUT3 is lower than that for HadCRUT2, and a greater portion of the parameter space is still not rejected at the 1-percent significance level. Similar to the HadCRUT3 data, NCDC data accept a large portion of the parameter space at

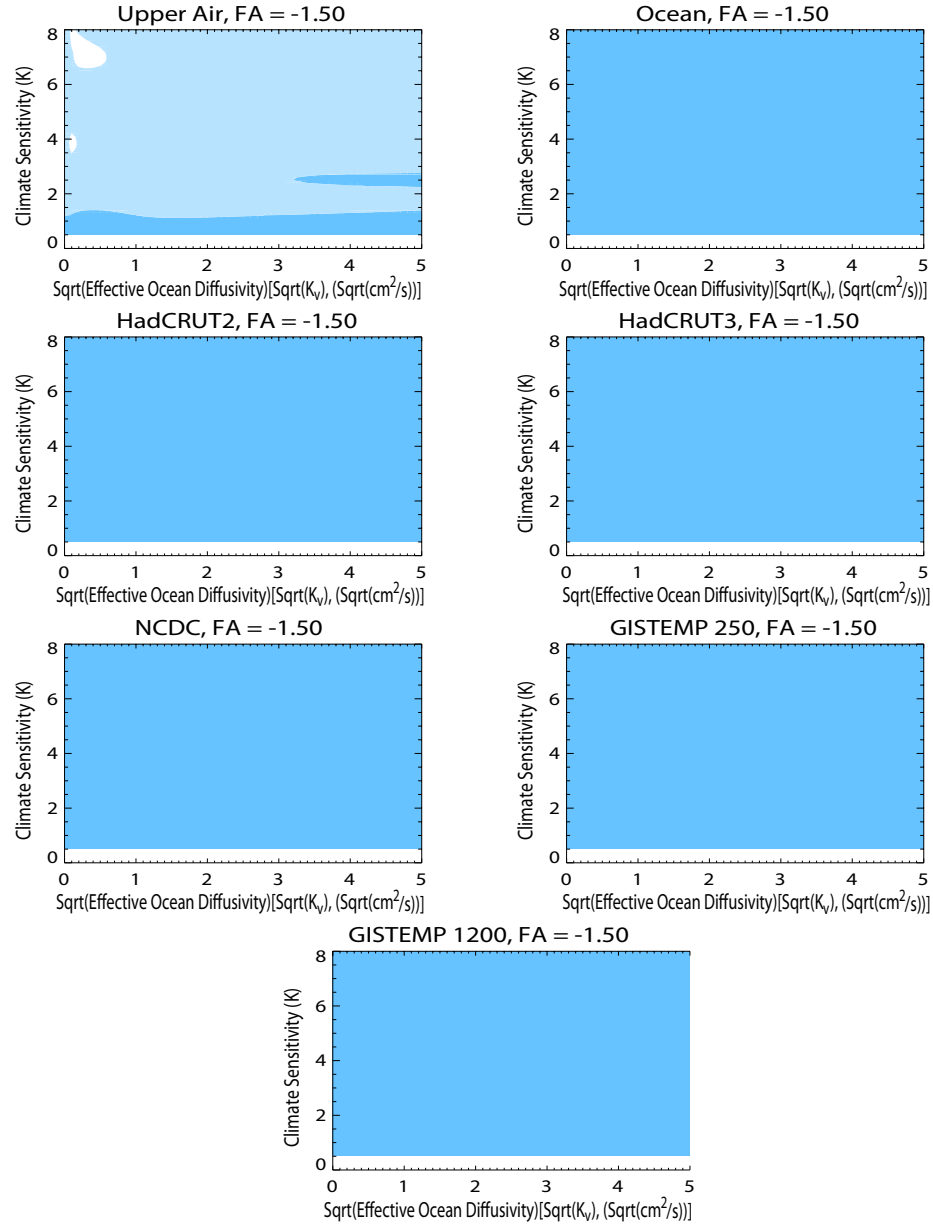


Figure 4.9. As in Fig. 4.2 but at an aerosol forcing of -1.50 Wm^{-2}

this aerosol level. However, a greater portion of the high climate sensitivity and low diffusivity space is rejected. The upper bound of the rejection region for climate sensitivity is also lower. Both GISTEMP datasets continue to reject a larger region of the high sensitivity and low diffusivity space than any of the other surface datasets, with the GISTEMP 1200 data being slightly more restrictive. While

the correlation between climate sensitivity and effective ocean diffusivity values is present in both GISTEMP datasets, the lower bound on climate sensitivity continues to be weak. This is shown by the lack of rejection in low climate sensitivity values across all effective ocean diffusivity values.

At the most extreme negative aerosol values used, -1.0 and -1.5 Wm^{-2} , all five surface datasets strongly reject all of the parameter space. Except for a few regions in the -1.0 Wm^{-2} level, all of the parameter combinations are rejected at the 1-percent significance level. The only deviation from this pattern is the acceptance of a small region of the parameter space in the GISTEMP 250 case falling inside of the 99-percent confidence region at an aerosol value of -1.0 Wm^{-2} . Despite the agreement at these levels and the positive levels, significant differences exist in the surface diagnostics at the levels in between.

4.3 Probability Distribution Functions for Climate System Properties

To generate a full three-dimensional likelihood for each diagnostic, likelihood values between the eight aerosol levels have been interpolated using a thin plate spline to fill the entire parameter space. By applying Bayes' Theorem (Bayes, 1763) to the likelihood distributions derived from each diagnostic, the information is combined to determine the joint probability distribution function for the parameter values. For this study, Bayes' Theorem is written as

$$P(\theta|y) \propto P(\theta) * P(y|\theta), \quad (4.1)$$

where $P(\theta|y)$ is the likelihood of a set of parameter values, given the observed data, $P(\theta)$ represents prior assumptions on the parameters, and $P(y|\theta)$ is the likelihood of the observed data, given that the parameters have been set. For each diagnostic, the likelihood distributions derived in Sec. 4.2 represent the $P(y|\theta)$ term in Eq. (4.1). The information from the three diagnostics is combined in an iterative process. In the first step, an expert prior on climate sensitivity first described and used in Forest et al. (2002) is combined with one of the diagnostics. Uniform priors have been used on effective ocean diffusivity and net aerosol forcing.

The prior on climate sensitivity acts to eliminate large values of climate sensitivity that have been deemed to be inconsistent with temperature changes observed in the paleoclimate records. The resulting likelihood function is then used as the prior for the second application of Eq. (4.1) and incorporates the second temperature diagnostic. To complete the process, the likelihood from the second step is used as the prior for incorporation of the final temperature diagnostic. The resulting likelihood function now incorporates all of the information obtained from the three observational datasets and gives a measure of the likelihood that a given set of parameter values, when used in the model, will yield model output that matches the observed temperature patterns.

From the three-dimensional distribution function, the marginal distributions of each parameter are calculated through integration across the other two parameters. The marginal distributions (Fig. 4.10) show that the constraints on the three parameters are sensitive to the choice of surface dataset. For climate sensitivity, the distribution derived from the GISTEMP datasets give the lowest values for the parameter. This can be traced back to the failure in the surface diagnostic to constrain the lower bound of the distribution. Furthermore, the 95-percent bound has the lowest value due to the rejection of high climate sensitivity values in the surface diagnostic. This rejection can be attributed to the weak warming trends discussed in Section 4.1 because model runs with high climate sensitivity values yield warming which is too strong to be consistent with the GISTEMP datasets. The remaining three datasets yield similar, yet still noticeably different results. HadCRUT datasets show the widest confidence intervals as a result of the large regions of the parameter space found to be consistent in the surface diagnostic. At nearly all levels, these regions were larger than those observed for the NCDC data. NCDC derived climate sensitivity values are only slightly smaller than those derived from the HadCRUT datasets. The similarities in the distributions can be loosely attributed to the similar long-range temperature trends shown in Fig. 4.1 because similar climate sensitivity values would be accepted and rejected based off of the first-order approximation of the temperature change.

Given the wide confidence intervals regardless of which surface dataset is used, effective ocean diffusivity is poorly constrained by the data. Figs. 4.2-4.9 have demonstrated that regions consistent with the data span the entire range of effec-

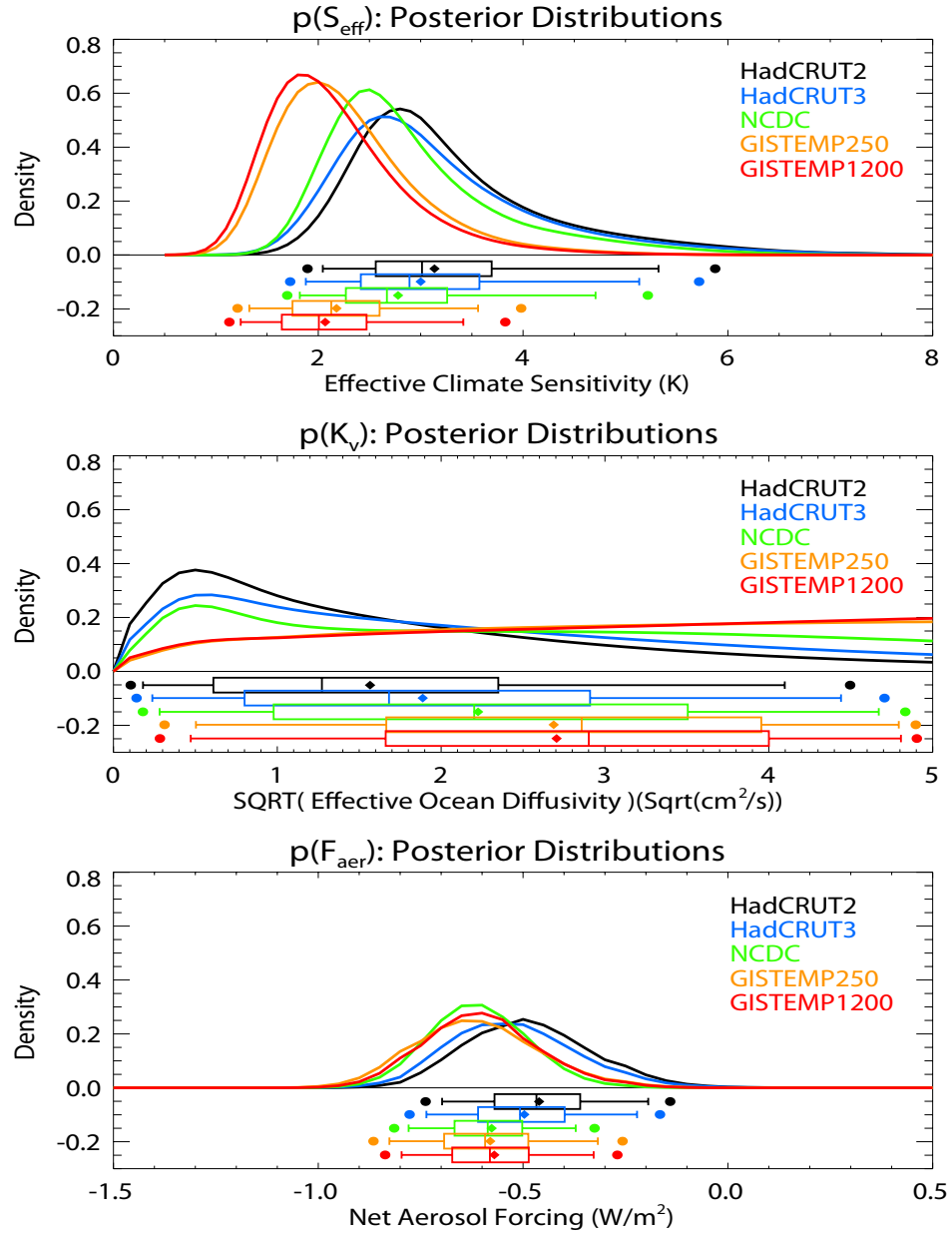


Figure 4.10. Marginal probability distribution functions using each of the surface datasets. Distributions for climate sensitivity (top), effective ocean diffusivity (middle), and net anthropogenic aerosol forcing (bottom) are shown. Box-and-whisker plots show the percentile bounds: 2.5-97.5 (dots), 5-95 (vertical lines at ends), 25-75 (box ends), and 50 (vertical line in box). The distribution mean is represented by a diamond, and the distribution mode is the peak in the distribution

tive ocean diffusivity values. This acceptance across all values leads to the wide parameter distributions seen in Fig. 4.10. With the exclusion of the GISTEMP cases, the mode in the distribution is found for low values of effective ocean diffusivity. This is due to the high climate sensitivity and high effective ocean diffusivity regions being rejected for positive values of aerosol forcing. As mentioned previously, an estimate of natural variability has been included in the ocean heat content diagnostic. This estimation results in a decrease in the significance of the ocean heat content signal and leads to weaker constraints for the effective ocean diffusivity parameter.

The distributions for net aerosol forcing cluster into two groups: those derived using HadCRUT data and those that are not. Weaker aerosol forcing values are estimated when using HadCRUT data than when either of the NCDC or GISTEMP datasets are used. It was noted previously that regions that are consistent with the surface data were first present at the -0.25 Wm^{-2} aerosol level for the HadCRUT datasets. As a result, the marginal distribution for the two datasets is shifted more

Table 4.1. Percentiles and means for marginal probability distribution functions for each surface dataset

	Surface Data	0.05	0.25	0.50	0.75	0.95	Mean
S							
	HadCRUT2	2.0	2.6	3.0	3.7	5.3	3.1
	HadCRUT3	1.9	2.4	2.9	3.6	5.1	3.0
	NCDC	1.8	2.3	2.7	3.3	4.7	2.8
	GISTEMP 250	1.3	1.7	2.1	2.6	3.6	2.2
	GISTEMP 1200	1.2	1.6	2.0	2.5	3.4	2.1
$\sqrt{K_v}$							
	HadCRUT2	0.18	0.61	1.3	2.3	4.1	1.6
	HadCRUT3	0.24	0.80	1.7	2.9	4.4	1.9
	NCDC	0.28	0.98	2.2	3.5	4.7	2.2
	GISTEMP 250	0.50	1.7	2.9	4.0	4.8	2.7
	GISTEMP 1200	0.47	1.7	2.9	4.0	4.8	2.7
F_{aer}							
	HadCRUT2	-0.19	-0.36	-0.47	-0.57	-0.70	-0.46
	HadCRUT3	-0.22	-0.40	-0.51	-0.61	-0.74	-0.50
	NCDC	-0.37	-0.50	-0.59	-0.67	-0.78	-0.58
	GISTEMP 250	-0.32	-0.49	-0.59	-0.69	-0.83	-0.58
	GISTEMP 1200	-0.33	-0.49	-0.58	-0.67	-0.80	-0.57

towards zero than for the other two datasets. Furthermore, the largest regions of consistent results for the NCDC and GISTEMP datasets were found in the -0.75 Wm^{-2} level. This leads to the observed shift in the distribution away from zero. A complete summary of the parameter distribution bounds and modes are presented in Table 4.1.

To gain a better understanding of the interactions between the parameters, marginal two-dimensional distributions are investigated. Like the one-dimensional marginal distributions, these are obtained through integration of the joint distribution over the other parameter values. For all of the datasets, a weak correlation can

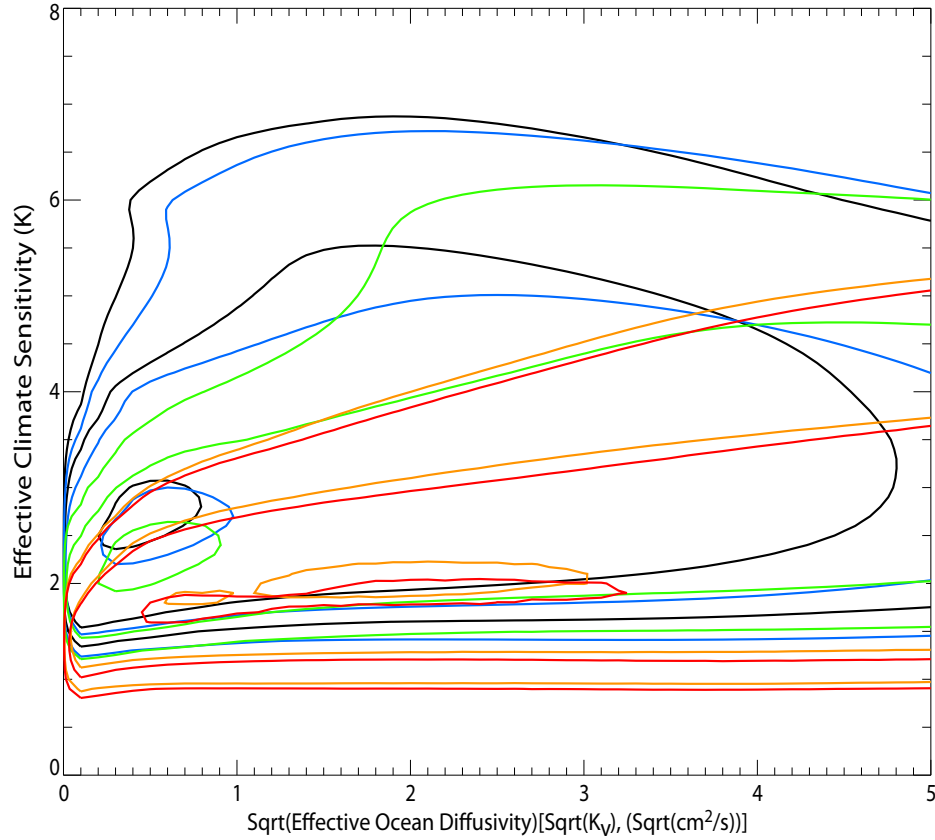


Figure 4.11. Two-dimensional marginal distribution function for climate sensitivity and effective ocean diffusivity when using HadCRUT2 (black), HadCRUT3 (blue), NCDC (green), GISTEMP 250 (orange), and GISTEMP 1200 (red) surface data. Starting from the outermost moving inward, contours mark the 99-, 90-, and 10-percent confidence regions. The 10-percent confidence region is included to show the modes of the distributions

be observed between climate sensitivity and effective ocean diffusivity (Fig. 4.11). As the effective ocean diffusivity increases, the climate sensitivity increases as well. This trend is seen most clearly using the 10-percent confidence region in the HadCRUT2, HadCRUT3, and NCDC sets, but is nearly non-existent using the GISTEMP data. However, in the GISTEMP cases, the upper bounds of the 99-percent confidence region display this trend. These trends are consistent with what has been found in past studies: the two parameters are positively correlated (Hansen et al., 1984).

From the climate sensitivity-aerosol forcing two-dimensional marginal distributions, it can be seen that the two parameters are correlated, regardless of the surface dataset (Fig. 4.12). As the climate sensitivity increases, the aerosol forcing takes on stronger negative values. Similarly, as the climate sensitivity decreases,

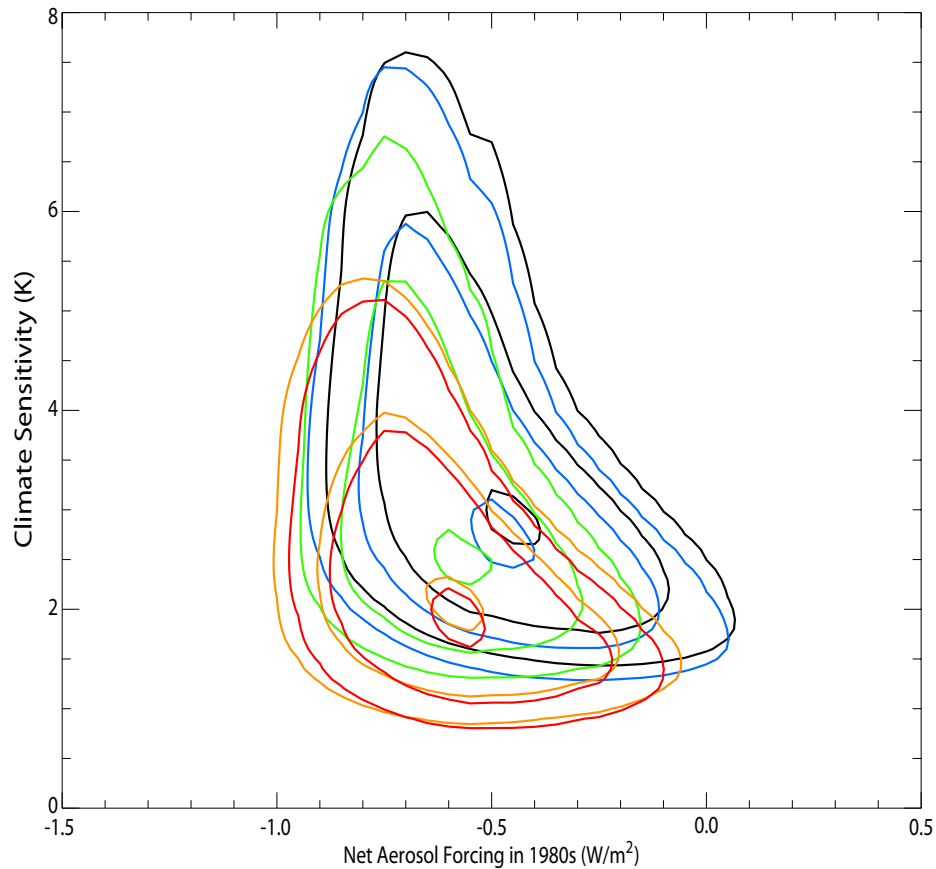


Figure 4.12. As in Fig 4.11, but for the two-dimensional marginal distribution function for climate sensitivity and anthropogenic aerosol forcing.

the aerosol forcing takes on less negative values. When considering the physical processes involved, these results agree with intuition. For all model runs, the climate forcings and surface dataset are fixed. If the climate sensitivity is increased for a model that yielded results consistent with the observed data without any other adjustments, surface warming becomes too high in the model to match the observations. To account for this increased warming, a stronger anthropogenic forcing would need to be used to introduce additional cooling into the system. The same reasoning can be used to argue that a decrease in climate sensitivity should lead to an increase in anthropogenic warming.

4.4 Transient Climate Response

The probability distributions of transient climate response (TCR) using each surface dataset have also been investigated. Similar to Forest et al. (2008), transient climate response in this study is defined as the change in surface air temperature averaged over years 61-80 in simulations that have 1% per year increases in CO₂ concentrations. From each joint distribution, a 1000 member Latin Hypercube sample (McKay et al., 1979) is drawn, whereby climate sensitivity-ocean diffusivity pairs are formed. Each pair of parameter values is then equated to a TCR using a functional fit calibrated by past runs of the MIT climate model. The fit has been derived using many runs of the model and a graphical representation of the derived relationships between climate sensitivity, effective ocean diffusivity, and TCR is shown in Figure 1 of Sokolov et al. (2003). Histograms of the resulting distributions are presented in the top panel of Figure 4.13 with the corresponding cumulative distribution functions in the bottom panel. Table 4.2 presents a summary of the distribution statistics for each of the datasets.

Similar to the parameter distributions, each of the TCR distributions derived from the different datasets takes on a different shape and yields different estimates of future warming. The most striking feature of the distributions is the wide range of possible transient climate responses. Depending on which dataset is used, predictions of future warming range anywhere from less than 1 K to as much as 3 K. Furthermore, the modes of the distributions are all found at different locations, showing that estimates of future climate change are sensitive to which surface

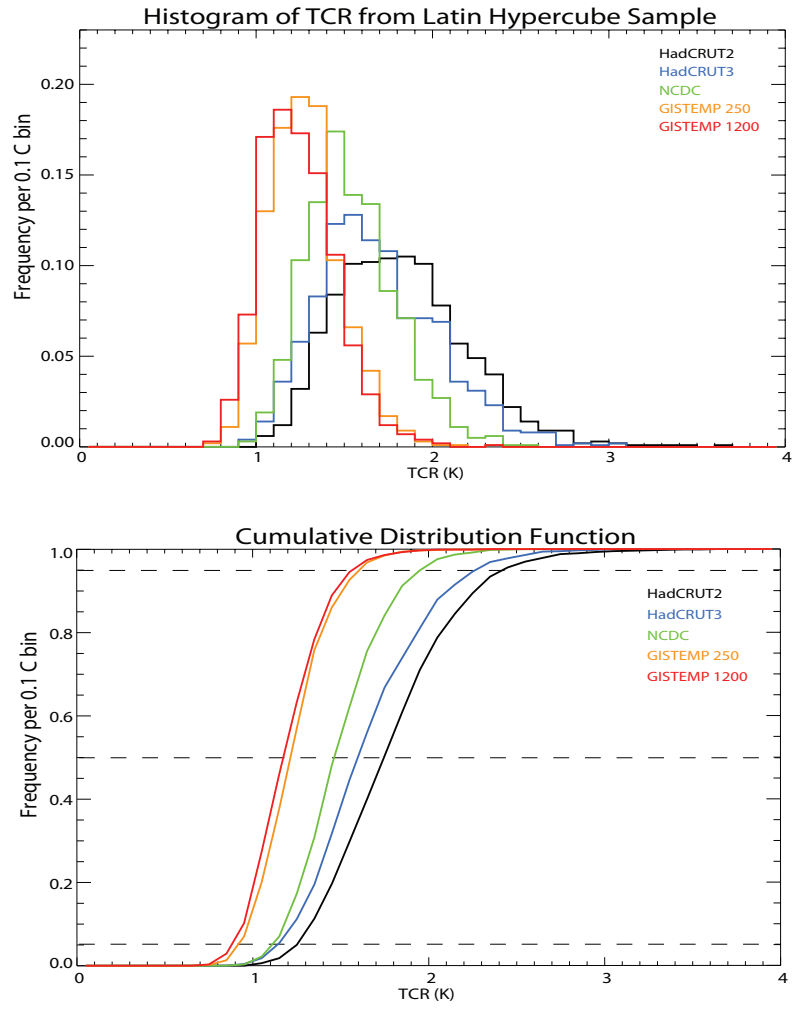


Figure 4.13. Histogram of TCR (top) and cumulative density function (bottom) derived from 1000 member Latin Hypercube samples of the joint distribution functions using each surface dataset.

Table 4.2. Percentiles for distributions of transient climate response for each surface dataset

	Surface Data	0.05	0.25	0.50	0.75	0.95
TCR (K)						
	HadCRUT2	1.24	1.50	1.73	2.00	2.41
	HadCRUT3	1.13	1.40	1.60	1.87	2.28
	NCDC	1.10	1.30	1.46	1.64	1.96
	GISTEMP 250	0.91	1.08	1.21	1.35	1.61
	GISTEMP 1200	0.87	1.02	1.17	1.32	1.58

dataset is used in the analysis.

One of the most noticeable differences is that the GISTEMP datasets produce sharper and narrower distributions and that the TCR is significantly less than the other datasets. Given the sharper marginal distribution for climate sensitivity around lower values found in the GISTEMP datasets, this result is not surprising. When climate sensitivity is lower, surface temperatures will not increase as much when the forcing is held constant. Since a greater number of climate sensitivity values from the GISTEMP datasets are in the low regions, the corresponding transient climate response values are lower as well. The weak constraint on high effective ocean diffusivity values also contributes strongly to the likelihood of lower TCR.

Following the pattern discussed for the GISTEMP datasets, the remaining three datasets follow a similar trend. Given that the effective ocean diffusivity distributions for the HadCRUT and NCDC datasets have similar shapes and comparable confidence intervals, the distribution of effective ocean diffusivity values drawn by the Latin Hypercube sample will be similar. Due to this, the TCR distribution is strongly influenced by the climate sensitivity distributions and weak constraint on effective ocean diffusivity. Both HadCRUT climate sensitivity distributions are wider than the NCDC distribution, and since there are a wider range of possible values to be drawn from in the sampling, the TCR distribution is broader for these datasets. Similarly, since the climate sensitivity distributions are shifted towards higher values for the HadCRUT distributions, the resulting TCR distributions are shifted towards higher values. This feature is in agreement with what was discussed with regard to the GISTEMP datasets, that the transient warming follows shifts in climate sensitivity due to the weak constraints on effective ocean diffusivity. However, differences in the effective ocean diffusivity distribution have also contributed to the differences in TCR for these datasets. Due to the more reasonable constraints on effective ocean diffusivity in the NCDC and HadCRUT datasets (modes near lower values), a significant region of low TCR values is no longer sampled. This contributes to the shifts towards higher TCR observed from these distributions. Given that the aerosol forcing distributions are nearly identical across all datasets, these results also show that TCR follows climate sensitivity and effective ocean diffusivity rather than the aerosol forcing.

Chapter 5

Conclusions

The results presented here show that climate model parameter distributions are sensitive to the surface temperature dataset used to compare model output against. In general, the widths of the climate sensitivity parameter distributions are comparable, but the locations of the means and fractiles are shifted depending on which surface dataset is used. The biggest shift in climate sensitivity is observed when the GISTEMP datasets are used and can be attributed to the weaker warming trends observed in those datasets. Regardless of the surface dataset, effective ocean diffusivity is poorly constrained by the data. This is likely due to the inclusion of the ocean internal variability estimate reducing the detectable signal in the ocean record. Distributions for anthropogenic aerosol forcing cluster into two groups. Distributions derived from the HadCRUT reconstructions indicate a weaker anthropogenic aerosol forcing. Reconstructions from the other groups show that approximately 0.1 Wm^{-2} of additional cooling from anthropogenic sources is needed in order to match the observed temperature records.

In addition to the differences in the one-dimensional marginal distributions, the two-dimensional marginal distributions for each surface dataset are different. However, all five distributions yield results that are consistent with findings from prior studies that investigate the correlations between parameters and are in line with physical reasoning. The positive correlation between climate sensitivity and ocean diffusivity found in past studies is evident in each of the distributions. Stronger correlations are found using the HadCRUT2, HadCRUT3, and NCDC datasets and, although weak, they can also be detected in the GISTEMP cases. All five two-

dimensional distributions for climate sensitivity and anthropogenic aerosol forcing show that if climate sensitivity increases, anthropogenic aerosol cooling must increase, and if climate sensitivity decreases, anthropogenic aerosol cooling must also decrease. These results are in line with physical arguments.

The sensitivity to choice of surface data used to constrain model parameters introduces additional uncertainties into model parameter calibration. When reporting climate system properties and designing climate models, this additional uncertainty must be considered. Failure to do so can result in overconfidence in the results and lead to inaccurate representations of the climate system and predictions of future climate change. The 2 K range in possible transient climate responses found when varying the surface datasets provides an example of the potential pitfalls associated with neglecting this uncertainty.

5.1 Future Work

The work presented here serves as the foundation for a variety of future studies. Thus far, only data from 1906-1995 has been used so that this study exactly matches the work of Forest et al. (2002, 2006, 2008). In those studies, data from 1906-1995 from a version of the HadCRUT2 reconstruction used in Allen et al. (2000) is used. The Allen et al. (2000) version of the data reports decadal mean temperatures from 1896-1995 on a 5x5 degree grid. Because the datasets used in this study report monthly temperature anomalies from pre-1900 and extend to the present, the methods developed here are not subject to the same restrictions as Forest et al. (2002, 2006, 2008). As a result, the period of comparison between model and observed surface temperatures can be modified to include more recent temperature trends. Using the model runs utilized in this study, nine additional years of data can be incorporated into the surface diagnostics. This extension of the data is due to the monthly surface datasets extending at minimum to 2005, but limited by the model runs which end in December of 2004.

Additional surface data can be utilized in one of two ways. The first way to incorporate additional data would be to shift the window for the period of comparison forward nine years. This method would replace the oldest nine years with the newest nine years. As a result of this shift, data which is sparse and less

reliable from the early 1900's will be replaced by more densely distributed and more reliable data from the early 2000's. The increased data coverage would reduce the effects that the differing interpolation schemes have on the surface datasets and lead to a reduction in the differences between the surface dataset time series. Thus, it is hypothesized that this would lead to convergence of the resulting parameter distribution functions.

The second way to utilize the additional data would be to use the additional nine years of more reliable data to add a sixth decade into the surface diagnostic. The addition of data would require the start date of the surface diagnostic to be shifted back one year from 1906 to 1905 in order to fill six complete decades. Shifting back the start year introduces one additional year of data from the less reliable portion of the temperature record, but the addition of this year is offset by the inclusion of nine years of updated, more reliable data. It is hypothesized that the inclusion of the additional decade of surface data will lead to a tightening of the parameter distribution functions. This hypothesis is based on the notion that the additional decade will include recent warming trends and thus lead to the rejection of parameter combinations that do not produce the warming found in the observations.

Much like the surface datasets, additional work to extend the upper-air and ocean temperature datasets is planned. For the upper-air temperature data, new time ranges need to be determined in order to keep the diagnostics similar to what is presented in this study. Choosing which two periods to use to calculate the differences used in the diagnostic would need to both utilize additional data and avoid periods which contain significant volcanic eruptions. For the ocean diagnostic, additional ocean data can be used in much the same way that additional surface temperature data is used. New data can be used to replace old data or new data can be used to extend the period of comparison used in the diagnostic. Similar to the surface temperature data, the use of extended upper-air and ocean data is expected to further tighten the parameter distribution functions.

When extended data are available, care needs to be taken as to how much data to add to the diagnostics. This caveat applies most readily to the surface diagnostic. If the number of zones or decades increases, the dimension of the noise-covariance matrix also increases. When the dimension becomes too large,

the noise-covariance matrix cannot be estimated due to the limited amount of control run data available. Potential methods to reduce the dimensionality of the matrix include moving from a four zonal band pattern to a hemispheric pattern or using 20-year means as opposed to decadal means. In both cases, the amount of control run data required for calculation of the noise-covariance is reduced by a factor of two.

The use of monthly surface temperature datasets greatly increases the flexibility of the method employed in this study. Since yearly and decadal temperatures can be derived, any comparison period and climatology can be used to derive parameter distribution functions. Using this flexibility, the impact that the choice of climatology has on the distribution functions can be further investigated. Initial work on this matter has shown that if a short, 50-year climatology consisting of the most recent data is used, the time series using the different datasets begin to converge. Differences still exist, and although not yet run, it is assumed that these remaining differences will still lead to differences in the parameter distribution functions depending on which dataset is used. However, signs are pointing in a positive direction that the issues addressed in this work may be eliminated in the future due to the increased quality and coverage of observed surface temperatures. Since it may be tens to hundreds of years until full convergence is realized, the issues presented in this study are still relevant, and future work investigating how to best reduce their impacts through climatology choices will help minimize the differences.

One last variation of future interest is altering the scheme used when calculating the surface diagnostic. In the current study, temperature anomalies for a given zonal band are calculated by subtracting the climatology of the band off of the decadal average temperatures for that band. As a result, all temperature anomalies calculated are typically less than 1 °C warmer or colder from the climatology for each zonal band and the equator-to-pole temperature gradient observed on the planet is masked. A diagnostic where the global mean climatology is subtracted from each zonal band would introduce the equator-to-pole temperature gradient into the analysis. This gradient is introduced because equatorward zonal bands would show positive anomalies from the global mean, whereas poleward zonal bands would show negative anomalies from the global mean. As it stands, this

characteristic of the climate record is masked by subtracting a warmer climatology from the equatorward bands and a colder climatology from the poleward bands. Under the proposed modification, the same value would be subtracted off of each of the four zonal bands. By unmasking the equator-to-pole temperature gradient, the observed and modeled temperature trends would be different than what is presented in this study. This variation of the treatment of the surface diagnostic and accounting for spatial differences may lead to significantly different parameter distribution functions.

Bibliography

- Allen, M. R., P. A. Stott, R. Schnur, T. Delworth, and J. F. B. Mitchell, 2000: Quantifying the uncertainty in forecasts of anthropogenic climate change. *Nature*, **407**, 617–620.
- Allen, M. R. and S. F. B. Tett, 1999: Checking for model consistency in optimal fingerprinting. *Clim. Dyn.*, **15**, 419–434.
- Andrews, D. G. and M. R. Allen, 2008: Diagnosis of climate models in terms of transient climate response and feedback response time. *Atm. Sci. Letters*, **9**, 7–12.
- Andronova, N. G. and M. E. Schlesinger, 2001: Objective estimation of the probability density function for climate sensitivity. *J. Geophys. Res.*, **106** (D19), 22,605–22,612.
- Bayes, T., 1763: An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society*, **53**, 370–418.
- Bony, S., et al., 2006: How well do we understand and evaluate climate change feedback processes? *J. Climate*, **19**, 3445–3482.
- Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones, 2006: Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *J. Geophys. Res.*, **111** (D12106), doi:10.1029/2005JD006548.
- Cox, P. M., R. A. Betts, C. D. Jones, S. A. Spall, and I. J. Totterdell, 2000: Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature*, **408**, 184–187.
- Delworth, T. L., R. J. Stouffer, K. W. Dixon, M. J. Spelman, T. R. Knutson, A. J. Broccoli, P. J. Kushner, and R. T. Wetherland, 2002: Review of simulations of climate variability and change with the GFDL R30 coupled climate model. *Clim. Dyn.*, **19** (7), 555–574.

- Forest, C. E., M. R. Allen, A. P. Sokolov, and P. H. Stone, 2001: Constraining climate model properties using optimal fingerprint detection methods. *Clim. Dynamics*, **18**, 277–295.
- Forest, C. E., P. H. Stone, and A. P. Sokolov, 2006: Estimated PDFs of climate system properties including natural and anthropogenic forcings. *Geophys. Res. Lett.*, **33** (L01705), doi:10.1029/2005GL023977.
- Forest, C. E., P. H. Stone, and A. P. Sokolov, 2008: Constraining climate model parameters from observed 20th century changes. *Tellus*, **60A** (5), 911–920.
- Forest, C. E., P. H. Stone, A. P. Sokolov, M. R. Allen, and M. D. Webster, 2002: Quantifying uncertainties in climate system properties with the use of recent climate observations. *Science*, **295**, 113–117.
- Hansen, J., A. Lacis, D. Rind, G. Russell, P. Stone, I. Fung, R. Ruedy, and J. Lerner, 1984: Climate sensitivity: Analysis of feedback mechanisms. *Climate Processes and Climate Sensitivity, Geophysical Monograph*, J. E. Hansen and T. Takahashi, Eds., American Geophysical Union, Washington, D.C., Vol. 29, 130–163.
- Hansen, J. and S. Lebedeff, 1987: Global trends of measured surface air temperature. *J. Geophys. Res.*, **92** (D11), 13,345–13,372.
- Hansen, J., R. Ruedy, M. Sato, and K. Lo, 2010: Global surface temperature change. *Rev. Geophys.*, **48** (RG4004), doi:10.1029/2010RG000345.
- Hegerl, G. C., K. Hasselmann, U. Cubasch, J. F. B. Mitchell, E. Roeckner, R. Voss, and J. Waszkewitz, 1997: Multi-fingerprint detection and attribution analysis of greenhouse gas, greenhouse gas-plus-aerosol and solar forced climate change. *Clim. Dyn.*, **13** (9), 613–634.
- Hegerl, G. C., P. A. Stott, M. R. Allen, J. F. B. Mitchell, S. F. B. Tett, and U. Cubasch, 2000: Optimal detection and attribution of climate change: Sensitivity of results to climate model differences. *Clim. Dyn.*, **16**, 737–754.
- Hegerl, G. C., et al., 2007: Understanding and attributing climate change. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller, Eds., Cambridge University Press, Cambridge, UK and New York, NY, USA, 663–746.
- Johns, T. C., R. E. Carnell, J. F. Crossley, J. M. Gregory, J. F. B. Mitchell, C. A. Senior, S. F. B. Tett, and R. A. Wood, 1997: The second Hadley Centre coupled

- ocean-atmosphere GCM: Model description, spinup and validation. *Clim. Dyn.*, **13**, 103–134.
- Jones, P. and A. Moberg, 2003: Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001. *J. Climate*, **16**, 206–223.
- Jones, P. D., T. M. L. Wigley, and P. M. Kelly, 1982: Variations in surface air temperatures. Part 1: Northern Hemisphere, 1881–1980. *Mon. Weath. Rev.*, **110** (2), 59–70.
- Knutti, R., T. F. Stocker, F. Joos, and G. Plattner, 2003: Probabilistic climate change projections using neural networks. *Clim. Dyn.*, **21**, 257–272.
- McKay, M. D., R. J. Beckman, and W. J. Conover, 1979: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, **21**, 239–245.
- Murphy, J., 1995: Transient response of the Hadley Centre coupled ocean-atmosphere model to increasing carbon dioxide. Part III: Analysis of global-mean response using simple models. *J. Climate*, **8**, 496–514.
- Peterson, T. C. and R. S. Vose, 1997: An overview of the Global Historical Climatology Network temperature database. *Bull. American Meteorological Society*, **78** (12), 2837–2849.
- Randall, D., et al., 2007: Climate models and their evaluation. *Climate Change 2007, The Physical Science Basis*, S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M. Tignor, and H. Miller, Eds., Cambridge University Press, Cambridge, UK, 589–662.
- Raper, S. C. B., J. M. Gregory, and R. S. Stouffer, 2002: The role of climate sensitivity and ocean heat uptake on AOGCM transient temperature response. *J. Climate*, **15**, 124–130.
- Rayner, N. A., P. Brohan, D. E. Parker, C. K. Folland, J. J. Kennedy, M. Vanicek, T. J. Ansell, and S. F. B. Tett, 2006: Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the mid-nineteenth century: The HadSST2 dataset. *J. Climate*, **19**, 446–469.
- Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperatures, sea ice, and night marine air temperatures since the late nineteenth century. *J. Geophys. Res.*, **108** (D14), doi:10.1029/2002JD002670.

- Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, 2002: An improved in situ and satellite SST analysis for climate. *J. Climate*, **15**, 1609–1625.
- Sansó, B. and C. Forest, 2009: Statistical calibration of climate system properties. *Appl. Statist.*, **58**, 485–503.
- Smith, T. M., R. W. Reynolds, T. C. Peterson, and J. Lawrimore, 2008: Improvements to NOAA’s historical merged land-ocean surface temperature analysis (1880–2006). *J. Climate*, **21**, 2283–2296.
- Sokolov, A., et al., 2005: The MIT integrated global system model (IGSM) version 2: Model description and baseline evaluation, MIT JP Report 124. Tech. Rep. http://web.mit.edu/globalchange/www/MITJPSPGC_Rpt124.pdf, MIT, Joint Program on the Science and Policy of Global Change, Room E40-428, 77 Massachusetts Ave., Cambridge, MA 02139.
- Sokolov, A. P., C. E. Forest, and P. H. Stone, 2003: Comparing oceanic heat uptake in AOGCM transient climate change experiments. *J. Climate*, **16**, 1573–1582.
- Sokolov, A. P. and P. H. Stone, 1998: A flexible climate model for use in integrated assessments. *Clim. Dyn.*, **14**, 291–303.
- Stott, P. A., J. F. B. M. M. R. Allen, T. L. Delworth, J. M. Gregory, G. A. Meehl, and B. D. Santer, 2006: Observational constraints on past attributable warming and predictions of future global warming. *J. Climate*, **19**, 3055–3069.
- Stott, P. A. and J. A. Kettleborough, 2002: Origins and estimates of uncertainty in predictions of twenty-first century temperature rise. *Nature*, **416**, 723–726.
- Stott, P. A., S. F. B. Tett, G. S. Jones, M. R. Allen, J. F. B. Mitchell, and G. J. Jenkins, 2000: External control of 20th century temperature by natural and anthropogenic forcings. *Science*, **290**, 2133–2137.
- Stouffer, R. J., J. Russell, and M. J. Spelman, 2006: Importance of oceanic heat uptake in transient climate change. *Geophys. Res. Lett.*, **33** (L17704), doi:10.1029/2006GL027242.
- Tett, S. F. B., P. A. Stott, M. R. Allen, W. J. Ingram, and J. F. B. Mitchell, 1999: Causes of twentieth-century temperature change near the Earth’s surface. *Nature*, **399**, 569–572.
- Tett, S. F. B., et al., 2002: Estimation of natural and anthropogenic contributions to twentieth century temperature change. *J. Geophys. Res.*, **107** (D16), doi:10.1029/2000JD000028.

- Thorne, P. W., et al., 2002: Assessing the robustness of zonal mean climate change detection. *Geophys. Res. Lett.*, **29** (19), doi:10.1029/2002GL015717.
- Tomassini, L., P. Reichert, R. Knutti, T. F. Stocker, and M. E. Borsuk, 2007: Robust Bayesian uncertainty analysis of climate system properties using Markov Chain Monte Carlo estimates. *J. Climate*, **20**, 1239–1254.
- Urban, N. M. and K. Keller, 2010: Probabilistic hindcasts and projections of the coupled climate, carbon cycle and Atlantic meridional overturning circulation system: A Bayesian fusion of century-scale observations with a simple model. *Tellus*, **62A** (5), 737–750.
- Van den Dool, H. M., S. Saha, and Å. Johansson, 2000: Empirical orthogonal teleconnections. *J. Climate*, **13**, 1421–1435.