The Pennsylvania State University
The Graduate School
College of Earth and Mineral Science

IMPROVING CONSTRAINTS ON CLIMATE SYSTEM PROPERTIES WITH ADDITIONAL DATA AND NEW METHODS

A Dissertation in
Meteorology
by
Alex Gordon Libardoni

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2017

The dissertation of Alex Gordon Libardoni was reviewed and approved* by the following:

Chris E. Forest
Associate Professor of Climate Dynamics
Dissertation Advisor, Chair of Committee

Kenneth J. Davis
Professor of Meteorology

Klaus Keller
Professor of Geosciences

Raymond G. Najjar
Professor of Oceanography

Andrei P. Sokolov
Research Scientist, Massachusetts Institute of Technology Joint Program on the Science and Policy of Global Change
Special Committee Member

David J. Stensrud
Department Head and Professor of Meteorology

*Signatures are on file in the Graduate School.

Abstract

Future climate change depends on properties of the climate system and the external forcing factors that drive the global energy budget. Among those properties are climate sensitivity, the rate at which heat is mixed into the deep ocean, and the aerosol forcing on the planet. In this dissertation, we use the newly updated Massachusetts Institute of Technology Earth System Model (MESM) to derive the joint probability distribution function (PDF) for model parameters that represent the aforementioned climate system properties. Climate sensitivity (ECS) in the model is set through an adjustment to the cloud feedback parameter. The vertical diffusion coefficient, K_v, represents the mixing of heat into the deep ocean by all mixing processes. The net anthropogenic aerosol forcing parameter, F_{aer}, estimates the contribution of aerosol cooling to the global energy budget. Using an 1800-member ensemble of MESM runs where the model parameters have been systematically varied, we derive PDFs for the model parameters by comparing the model output against historical observations of surface temperature and global mean ocean heat content. In particular, we answer four main research questions: (1) How are the parameter PDFs derived using the MESM ensemble different from those using a previous version of the model?, (2) How do the estimates change when recent surface temperature and ocean heat content observations are included in the model diagnostics used to evaluate model performance?, (3) How does internal climate variability lead to uncertainty in the parameter estimates?, and (4) What impact do the changes in PDFs have on estimates of future warming, namely estimates of transient climate response (TCR)? We show that estimates of climate sensitivity increase and the aerosol forcing is less negative when using MESM. These shifts are the result of a new forcing suite used to drive the model. By extending the length of the model diagnostics one decade at a time, we show that recent temperature patterns impact our estimates of the climate system properties. The continued rise in surface temperature leads to higher values of ECS, while the increased rate of heat storage in the ocean leads to lower estimates of ECS and higher estimates of K_v. We show that the parameter distributions are sensitive to the internal variability in the climate system and that using a single variability estimate can lead to PDFs that are too narrow. Throughout the dissertation, we show that estimates of transient climate response are correlated with ECS and K_v. Namely, higher ECS and weaker K_v lead to higher values of TCR. When considering all of these factors, we arrive at our best estimate for the climate system properties. We estimate the 90-percent confidence interval for climate sensitivity to be 2.7 to 5.4 °C with a mode of 3.5 °C. Our estimate for K_v is 1.9 to 23.0 cm^{2}s^{-1} with a mode of 4.41 cm^{2}s^{-1}. F_{aer} is estimated to be between -0.4 and -0.04 Wm^{-2} with a mode of -0.25 Wm^{-2}. Lastly, we estimate TCR to be between 1.4 and 2.1 °C with a mode of 1.8 °C.

Table of Contents

List of Figures

List of Tables

Acknowledgments

This study was funded by contributions from the MIT Joint Program on the Science and Policy of Global Change, the National Science Foundation, and the United States Department of Energy. The author would like to thank Berkeley Earth, 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 and NOAA's National Centers for Environmental Information for providing publicly accessible ocean heat content data. The author acknowledges the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for the CMIP5 archive used for control run data. The author would also like to thank the climate modeling groups for producing and making available their model output.

Special thanks to my dissertation advisor, Chris Forest, and my dissertation committee, Ken Davis, Klaus Keller, Ray Najjar, and Andrei Sokolov, for guidance in completion of this study. Thanks to Erwan Monier, Andrei Sokolov, Jeffery Scott, and the rest of the Uncertainty Quantification team at the MIT Joint Program for invaluable discussion and insights regarding this work. Thank you to the Forest Research Team: Alexis Hoffman, Rob Ceres, Judy Tsai, Kristina Rolph, Perry Oddo, and Kelsey Ruckert for invaluable feedback at intermediate stages of this study.

Last, but not least, thank you to my parents, Sherryl and Alan Libardoni, my brother, Eric Libardoni, and my girlfriend, Tiffany Samuelson, for their unwavering support throughout this process. Their love and support carried me through the good times and the bad.

Chapter 1 Introduction

Reliable projections of future climate change strongly depend on models that accurately represent the behavior of the climate system. In this dissertation, we aim to improve our current understanding of the climate system by deriving probability distributions for three properties key to the evolution of future climate: climate sensitivity, the rate of ocean heat uptake, and the net anthropogenic aerosol forcing. Better estimates of these properties will improve projections of future climate change and the associated risks. In this introduction, we provide a broad overview of the topics relevant to the dissertation. Each individual chapter has its own introduction that focuses on the specific research topic being explored. Here, we first discuss the hierarchy of model complexity and present where the model used in this dissertation fits within the hierarchy. We then present a discussion of the energy balance of the climate system and how the climate system properties we estimate in this study fit within that framework. A discussion of the estimates of the climate system properties we explore follows. We conclude the chapter with a discussion of the key questions and an outline for the remainder of the dissertation.

1.1 Hierarchy of Climate Models

Climate models exhibit a wide range of complexities. Current models range from Earth System Models (ESMs), to fully coupled, three-dimensional atmosphere-ocean general circulation models (AOGCMs), to Earth system models of intermediate complexity (EMICs), to simple climate models (Flato et al., 2013). Until recently, AOGCMs represented the most complete climate models in the field. In an effort to fully understand the physical dynamics of the climate system, AOGCMs were developed to couple many components of the climate system into a single model capable of simulating the complex interactions between the components as observed in the real world. ESMs have expanded upon AOGCMs by adding biological and chemical components into the modeling framework. Examples of added components include models of the carbon cycle, ocean biogeochemistry, atmospheric chemistry, and land use/land cover change. Given their complexity, ESMs and AOGCMs require significant computational resources for a single model run.

EMICs include many of the same components as AOGCMs but substitute parameterizations for some climate system components and include additional processes (e.g., biogeochemistry) excluded from AOGCMs due to the computational cost. Simple climate models include even more parameterized processes and are typically tuned to replicate results from AOGCMs. Global energy balance models are one example of a simple climate model (see Section 1.3). Given their simplifications of the climate system, EMICs and simple climate models run much faster than AOGCMs and ESMs. For example, a single multi-century run of an ESM or AOGCM may take months to complete, whereas simple climate models and EMICs can yield hundreds to millions of runs in a single night using fewer computational resources. A schematic of the trade-offs between model complexity and computational requirements is shown in Figure 1.1. 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 and studies that require long simulations (Flato et al., 2013).

1.2 The MIT Earth System Model

The Massachusetts Institute of Technology (MIT) Earth System Model (MESM) is the Earth system component of the MIT Integrated Global Systems Model (IGSM, Sokolov et al., 2005) and is the primary model used in this dissertation. In the past, the Earth system component of IGSM has also been referred to as IGSM, and we will continue with this convention throughout the dissertation. For our work, we use an improved and updated version of the Earth System Model, which we will refer to as MESM hereafter. In particular, we use a version of MESM that turns off sub-models for atmospheric chemistry, ocean biogeochemistry, and the terrestrial ecosystem. The resulting model is an EMIC of the climate system that couples models of the atmosphere, ocean, and land,

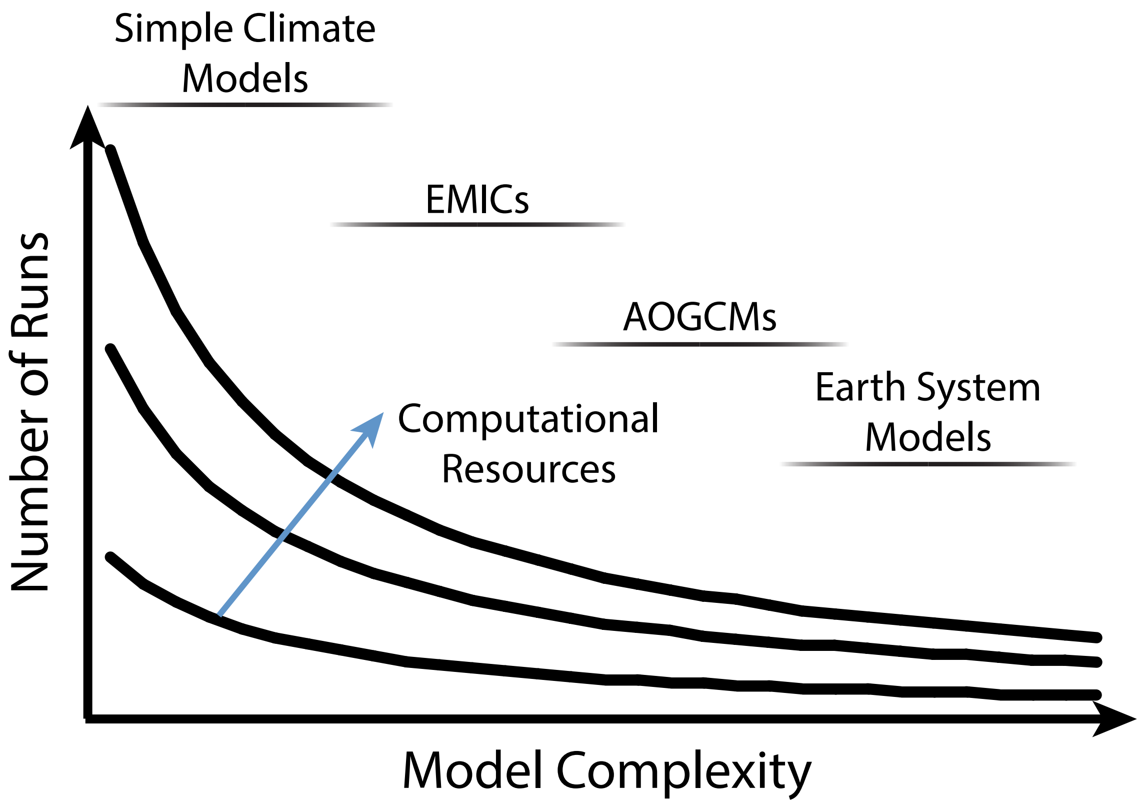

Figure 1.1: Representation of the trade-off between model complexity and the number of runs of a given model. Model complexity takes many different forms, ranging from model resolution to the number of climate components coupled together. Approximate locations of the types of models discussed in the text are indicated on the complexity spectrum. Contours are theoretical lines of constant computational resources.

and allows us to study their evolution and interactions under different forcing scenarios (Figure 1.2). The atmospheric model is a dynamical model derived from the Goddard Institute for Space Studies two-dimensional climate model that was first described in Yao and Stone (1987) and Stone and Yao (1987, 1990) and in final form in Sokolov and Stone (1998). Temperatures over a given latitude band are averaged, leaving only latitude and height as coordinates. The model is run at 4° latitudinal resolution for a total of 46 zonal bands and has a total of 11 vertical layers in the dynamic model and 14 vertical layers in the radiation model. Each latitude band has its surface divided into fractions of land, ocean, sea ice, and land ice, all with their own energy flux and temperature calculations. Zonal values for surface conditions are calculated as the weighted average of the four land surface types. The atmosphere is coupled to a mixed-layer ocean model with the two systems interacting through the diffusion of heat anomalies across the

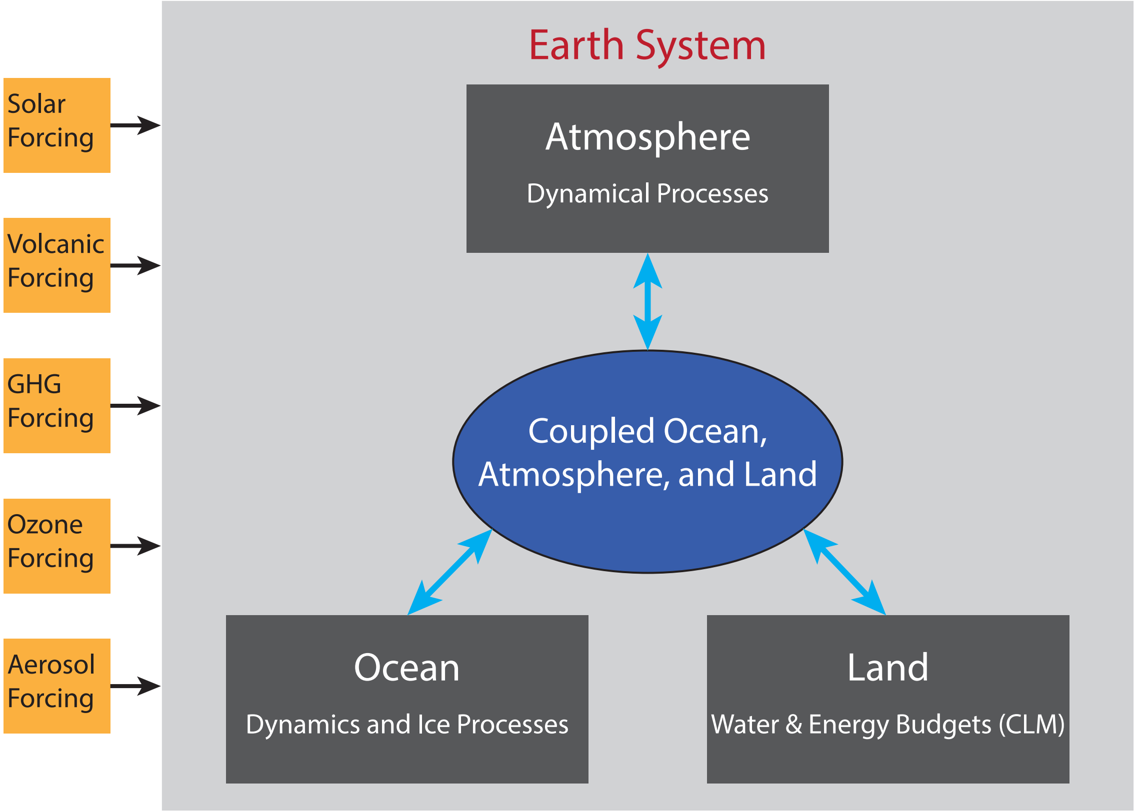

Figure 1.2: Schematic of the components of the MIT Earth System Model used in this dissertation. The coupled atmosphere, ocean, and land system is driven by natural and anthropogenic forcings. Figure adapted from the full diagram of the Integrated Global System Model accessed on the Joint Program on the Science and Policy of Global Change website at https://globalchange.mit.edu/research/research-tools/global-framework.

interface. The mixed-layer ocean consists of an upper layer with horizontal resolution of 4° in latitude and 5° in longitude and a 3000 m deep anomaly diffusing ocean beneath (Sokolov et al., 2007). Sokolov et al. (2007) show that for historical and short-term future climate change scenarios, the mixed-layer ocean model matches the behavior of a three-dimensional version of the model quite well. This coupled system is driven by the natural and anthropogenic climate forcings first described in Forest et al. (2006). Natural sources include solar irradiance changes and stratospheric aerosols from volcanic eruptions. Anthropogenic sources include greenhouse gas concentrations, sulfate aerosol loadings, and stratospheric and tropospheric ozone concentrations. While the forcing suite remains the same, the forcing records have been updated and extended to allow for model simulations that run to near present time.

One benefit of the model is the ease with which three characteristics of the climate system can be modified. These characteristics are climate sensitivity, the rate of ocean heat uptake, and the net anthropogenic aerosol forcing. Climate sensitivity measures the global mean surface temperature response to a change in forcing and is modified by changing the strength of the cloud feedback at different levels in the model. In particular, the cloud fraction predicted by the model, C^{MOD}, is related to the cloud fraction used in the radiation calculations, C^{RAD}, by

$$C^{RAD} = C^{MOD}(1 \pm \kappa \Delta T_{sfc}), \tag{1.1}$$

where κ is the model cloud feedback parameter and ΔT_{sfc} is the difference between the model global mean surface temperature and the global mean surface temperature from a control run simulation. Sokolov (2006) showed that for the range of climate sensitivities found in global climate models, adjusting the cloud feedback simulates well the sensitivity to other forcings. Because high and low clouds have opposite effects on global temperature (warming and cooling, respectively), the sign of the adjustment is different for high (-) and low (+) clouds. A specific value of climate sensitivity is obtained by choosing the appropriate value of κ in Eq. (1.1). By changing the fraction of high and low clouds in opposite directions, a smaller value of κ is needed for a given value of climate sensitivity, thus reducing the artificial change in cloud cover as a result of the adjustment. The utility of the adjustment method is not restricted to MESM alone and has been shown to be effective with a three-dimensional climate model as well (Sokolov and Monier, 2012).

The rate of ocean heat uptake is controlled by the effective ocean diffusivity in the model. In MESM, ocean diffusivity represents the vertical mixing of heat below the mixed layer by all processes and varies as a function of latitude (Sokolov and Stone, 1998). The spatial distribution of the diffusion coefficient is based on measurements of tritium mixing into the deep ocean reported in Hansen et al. (1984), with stronger mixing in polar regions and weaker mixing in equatorial regions. To obtain a specified global mean diffusivity, the pattern is uniformly scaled to preserve the spatial structure of the mixing. Smaller diffusivities result in a slower mixing of heat into the deep ocean, whereas larger diffusivities result in a more rapid mixing of heat into the deep ocean. Because all mixing processes are approximated and aggregated by the diffusivity parameter, values are larger than those for molecular diffusion found in subgrid-scale parameterizations of other models (Sokolov et al., 2003).

The anthropogenic aerosol forcing used in the model is prescribed spatially by a latitude-dependent pattern that differs over land and the ocean. 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 increasingly larger amplitudes as time progresses (Forest et al., 2001). Adjustment of the net anthropogenic aerosol forcing parameter sets the pattern amplitude in the 1980s. As the only adjustable forcing pattern in the model, the aerosol forcing also represents an estimate of all other unmodeled forcings with a global pattern similar to the aerosol loading.

1.3 Global Energy Budget Equation

The first-order behavior of the global climate system can be explained with the aid of the simple two-layer ocean model described in Gregory (2000). In this model, the ocean is divided into two layers: an upper layer that interacts with the atmosphere and a deep layer that interacts with the upper layer. Following the notation of Gregory et al. (2015), the two-layer model is written as a coupled set of two equations,

$$C_u \frac{dT_u}{dt} = F - \alpha T_u - \gamma (T_u - T_d) \qquad \textit{and} \tag{1.2}$$

$$C_d \frac{dT_d}{dt} = \gamma (T_u - T_d), \tag{1.3}$$

where C_u and C_d are the heat capacities of the upper and lower layers, respectively, T_u and T_d are the temperatures in the upper and lower layers relative to the initial equilibrium state, respectively, F is the external forcing on the climate system, α is the global feedback parameter, and γ is a transfer coefficient between the upper and lower layers. In this model, heat stored in land and heat that goes into melting ice are ignored. Adding Equations 1.2 and 1.3 yields the net global energy budget:

$$N = C_u \frac{dT_u}{dt} + C_d \frac{dT_d}{dt} = F - \alpha T_u \tag{1.4}$$

where N is the rate of heat stored in the climate system per square meter. Taken as a whole, the heat stored in the climate system is the difference between the net forcing, F, and the energy required to warm the system above equilibrium.

The model parameters described in Section 1.2— climate sensitivity, ocean diffusivity, and the anthropogenic aerosol forcing scaling factor —tie directly into Eqs. 1.2-1.4.

Climate sensitivity in MESM is related to the inverse of α. In the energy balance model, α is the net feedback of the climate system and represents the sum of all feedbacks in the climate system. α also represents the change in forcing required to change the global mean temperature by 1 °C. In MESM, α is also the sum of individual feedbacks, but can only be adjusted through the value of κ in the cloud adjustment scheme. The ocean diffusivity parameter sets the mixing rate of heat anomalies between the mixed layer and the deep ocean. Thus, there is a direct relationship between the ocean diffusivity in MESM and the transfer coefficient, γ, in the energy balance model. Any forcings that are not explicitly modeled in MESM are approximated using the aerosol forcing, linking this parameter to the net forcing, F, in Eq. 1.4.

1.3.1 Temperature Metrics from the Energy Balance Model

When F in Eqs. 1.2 and 1.4 is taken to be the additional forcing introduced by an instantaneous 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. (1.4) is zero as the net forcing introduced by the CO_2 doubling exactly matches the outgoing radiation due to the increase in global mean temperature. The temperature reached in this new equilibrium state is the climate sensitivity of the system and calculated as $T_{u,eq} = \frac{F_{2XCO_2}}{\alpha_{eq}}$.

Estimating climate sensitivity from the energy balance model is not limited to equilibrium forcing scenarios or equilibrium conditions. Under any forcing scenario, F, an effective climate sensitivity can be estimated from the model time series and is indicative of the strength of the feedbacks currently active in the system. These transient climate simulations allow feedbacks in the climate system to be active under idealized or more realistic forcing scenarios, as opposed to equilibrium runs where the forcing change is abrupt and held constant (Murphy, 1995). In such transient simulations, the rate of heat storage in Eq. (1.4) is non-zero and the system is adjusting to the external forcing. Eq (1.4) can be solved for α and estimated from the time series of N, F, and T to calculate the effective net feedback, α_{eff}. Inversion of the effective net feedback yields the effective climate sensitivity, the non-equilibrium measure of the global mean temperature response to a unit change in forcing in the current state.

An idealized forcing scenario is used to define a second temperature metric related to a doubling of CO_2 concentrations. In this transient climate simulation, F is taken to be the forcing when CO_2 concentrations are increased at the rate of 1% per year instead of instantaneously doubling the CO_2 concentration. In this specific forcing scenario, the transient climate response (TCR) is taken to be the temperature change at the time of CO_2 doubling, which occurs in year 70 of the simulation. Typically, the 20-year average of global mean surface temperature centered on year 70 is taken as the transient climate response. As a shorter time-scale measure of climate change, TCR has more of an impact on climate change over the next century than does equilibrium climate sensitivity (Millar et al., 2015). Because TCR is conducted under a more realistic forcing scenario and does not rely on equilibrium conditions, it is a much more policy-relevant measure of climate change.

1.4 Climate Sensitivity and Methods for Estimation

Climate sensitivity has been given a lot of attention due its impact on the climate system and because it summarizes the impact of CO_2 on climate change in a single number (Charney, 1979; Mitchell et al., 1990; Kattenberg et al., 1996; Cubasch et al., 2001; Meehl et al., 2007; Andronova et al., 2007; Knutti and Hegerl, 2008; Collins et al., 2013). Not only has climate sensitivity played a role in the evolution of surface temperature patterns in the past, it also plays a key role in determining future climate change (Andrews and Allen, 2008). In particular, two climate change metrics, the aforementioned transient climate response (Section 1.3.1) and sea level rise, have been shown to depend on climate sensitivity (Webster et al., 2008; Sokolov et al., 2009; Sriver et al., 2012). These two metrics are relevant because their impacts may require the implementation of adaptation strategies or serve as drivers for policy change.

Estimates of climate sensitivity have been derived using a number of different methods. We summarize four such methods here.

1.4.1 Inference from Feedback Analysis

One of the complexities of climate sensitivity is that it is a function of many feedbacks and processes in the climate system (Flato et al., 2013). In particular, the lapse rate, water vapor, snow-albedo, and cloud feedbacks play critical roles in determining climate sensitivity (Bony et al., 2006). Although there is a general understanding of how each feedback mechanism works, the overall impact of each on other feedbacks and overall climate sensitivity is uncertain. Uncertainties in the individual components lead to uncertainty in the total global feedback and this feedback uncertainty contributes to uncertainty in climate sensitivity estimates.

One method for estimating climate sensitivity is to estimate feedbacks individually and combine them into a total feedback parameter. Climate sensitivity is calculated as the inverse of the total feedback parameter. Methods for estimating feedback strengths from the change in top-of-the-atmosphere radiative flux measurements are summarized in Klocke et al. (2013). These methods include systematically varying individual variables associated with specific feedbacks from a control run to isolate their impacts, calculating radiative kernels for each variable, determining the change in cloud radiative forcing (cloud feedback only), and regression of top-of-the-atmosphere radiative forcing versus surface temperature change. All four methods calculate feedbacks using a linear approximation, and therefore do not allow for non-linearities or interactions between different feedbacks. Further strengths and weaknesses of the methods are more thoroughly discussed in Klocke et al. (2013).

1.4.2 Inference from Perturbed Physics Ensembles

Another method for understanding climate sensitivity is to determine how microphysical properties in a given model affect its behavior. Perturbed physics ensembles (PPEs) are designed to test how climate sensitivity and other properties depend on the small-scale behavior of the model. Parameters such as the fall speed of ice, ice particle size, the critical relative humidity for cloud formation, and gravity wave drag, to name a few, are systematically varied and the sensitivity of the model to these changes is evaluated to determine their impacts on model behavior. Emergent properties of the model, including climate sensitivity, can then be diagnosed for each configuration of the model. From this evaluation, the range of possible climate sensitivity values in the model, and the processes governing these properties, can be better understood.

Multiple studies have used PPEs to evaluate climate sensitivity for individual models. Examples of PPE studies include the Quantifying Uncertainties in Model Predictions project (Murphy et al., 2004; Webb et al., 2006), the *climateprediction.net* experiments (Stainforth et al., 2005; Piani et al., 2005), the Japan Uncertainty Modeling Project (Yokohata et al., 2010), and the National Center for Atmospheric Research CAMcube ensemble (Sanderson, 2011). Each study uses a different model and the number of parameters that are perturbed differs amongst them. One drawback of the PPE approach is that they are typically run using computationally expensive AOGCMs. With the exception of the *climateprediction.net* studies which utilize idle computer time from volunteers, the model complexity significantly limits the number of model runs that can be performed (see Figure 1.1). The limitation on model runs in turn limits the number of parameters that can be varied.

1.4.3 Inference from the Global Energy Budget

A number of studies have estimated climate sensitivity using the global energy budget approach (Gregory et al., 2002; Forster and Gregory, 2006; Otto et al., 2013; Lewis and Curry, 2014; Masters, 2014). When deriving estimates of climate sensitivity, these studies estimate the terms in Eq. 1.4 as the difference between two time periods and solve for climate sensitivity. While the same general method is applied in each of these studies, groups take different approaches to estimating the terms in the energy balance equation. All of the studies estimate global mean surface temperature from observations. The net energy imbalance is estimated either from satellite observations (Forster and Gregory, 2006) or the time series of global mean ocean heat content (Gregory et al., 2002; Otto et al., 2013; Lewis and Curry, 2014; Masters, 2014). Time series of the forcings are typically taken directly from observations (Forster and Gregory, 2006), simulations of fully-coupled AOGCMs (Otto et al., 2013; Masters, 2014), or radiative forcing estimates from the Intergovernmental Panel on Climate Change (IPCC) reports (Gregory et al., 2002; Lewis and Curry, 2014). A more in-depth summary of the state of energy budget approaches to estimating climate sensitivity is presented in Forster (2016).

1.4.4 Inference from Models and Observations

Climate models allow for climate sensitivity to be estimated directly rather than calculating individual feedback strengths or varying sub-grid scale processes. Many model-driven studies use simple models such as the EMICs and energy balance models described in Section 1.1. The benefits of these models are that run times are significantly shorter and model parameters, including climate sensitivity, can be changed easily. When combined, the preceding factors allow for multiple parameters to be adjusted jointly over a wide range values. By sampling different values of multiple parameters, a large number of possible states of the climate system can be simulated. Climate parameters that are typically adjusted include climate sensitivity, an estimate of heat uptake in the ocean, and an estimate of the net anthropogenic aerosol forcing. Observations are then used to constrain the model parameters by evaluating how well model output matches the observations.

Multiple studies have utilized this strategy for estimating climate sensitivity but using different methodologies. One method is to systematically vary the parameters on a grid of model runs and interpolate the results onto a finer grid to estimate parameter distributions (Forest et al., 2002, 2006, 2008; Libardoni and Forest, 2011, 2013; Lewis, 2013; Johansson et al., 2015). A second approach is to build a statistical approximation of the model based on a set of training runs at known parameter settings (Knutti et al., 2003; Tomassini et al., 2007; Drignei et al., 2008; Sansó and Forest, 2009; Olson et al., 2012). This emulator is then used to estimate model behavior at any setting of the parameters, allowing for Markov Chain Monte Carlo methods to be used to estimate the parameter distributions.

1.5 Estimates of Effective Climate Sensitivity and Transient Climate Response

In both Charney (1979) and the IPCC Fifth Assessment Report (IPCC AR5, Bindoff et al., 2013), climate sensitivity was reported to fall between 1.5 and 4.5 °C with a central estimate of 3 °C. Thus, despite the large number of studies aimed at estimating climate sensitivity and an increased understanding of the climate system, the uncertainty in the estimates has not changed. One of these reasons is the mismatch in climate sensitivity estimates derived from energy budget constraints and estimates derived from models. Estimates of climate sensitivity from studies using the global energy budget tend to be lower than those from models. Unless otherwise noted, all ranges presented are 90% confidence intervals. Otto et al. (2013) estimate climate sensitivity in the range 1.2-3.9 °C. Using similar averaging periods to Otto et al. (2013) but including additional uncertainty in the forcing time series, Lewis and Curry (2014) estimate climate sensitivity in the range 1.05-4.05 °C. Masters (2014) found a best estimate of 1.98 °C with a range of 1.19-5.15 °C.

In contrast, climate sensitivity estimated from models tends to be higher. Diagnosing climate sensitivity from the radiative forcing, surface temperature, and ocean heat content time series in the current generation of AOGCMs used in the Coupled Model Intercomparison Project Phase 5 (CMIP5, Taylor et al., 2012) yields an estimate of 3.2 ± 1.3 °C (Forster et al., 2013). Inverse estimation studies using EMICs and other simple models also tend to estimate higher climate sensitivity values and do not rule out the possibility of high climate sensitivity. Using statistical emulators of different EMICs, Sansó and Forest (2009) and Olson et al. (2012) estimate climate sensitivity in the ranges 1.1-5.2 °C and 1.8-4.9 °C, respectively. Using an approach similar to what is presented in this dissertation, Libardoni and Forest (2011, 2013) estimate climate sensitivity in the range 1.2-5.2 °C. One study based on EMIC simulations that estimates low climate sensitivity is Lewis (2013). Using the model simulations of Forest et al. (2006), Lewis (2013) estimates climate sensitivity in the range 1.2-2.2 °C, rather than the 1.9-4.7 °C range published in Forest et al. (2006). The differences arise due to changes in methodology implemented in Lewis (2013). These include a reduction in the number of model diagnostics used to evaluate model performance, the observational dataset used for surface temperature, the likelihood function used to relate model fit to a probability distribution, and the choice of prior on the model parameters.

Unlike with climate sensitivity, there is little separation between CMIP5 models, observationally based energy budget approaches, and model-based estimates of transient climate response. The range from IPCC AR5 estimates TCR between 1.0 and 2.5 °C (Bindoff et al., 2013). Estimates from global energy budget approaches estimate TCR in the ranges 0.9-2.0 °C (Otto et al., 2013) and 0.9-2.5 °C (Lewis and Curry, 2014). Forster et al. (2013) estimates TCR from the CMIP5 models to lie between 1.2 and 2.4 °C. Using an EMIC, Libardoni and Forest (2011, 2013) estimate TCR between 0.9 and 2.3 °C.

1.6 Estimates of Vertical Ocean Diffusivity

Many processes contribute to the turbulent mixing of heat in the ocean. These include wind-driven mixing, the propagation and breaking of internal waves, and flow over rough topography (MacKinnon et al., 2013). The aggregated effect of these mixing process can be summarized by the vertical diffusion coefficient, κ_v. Munk (1966) first estimated $\kappa_v = 1\ \mathrm{cm}^2\mathrm{s}^{-1}$ for depths below 1 km. The estimate was later confirmed by Munk and Wunsch (1998). Many studies have estimated κ_v by following the vertical motion of passive tracers injected into the ocean. Due to the spatial distribution of mixing mechanisms, κ_v is a function of latitude and depth, with weaker mixing above the thermocline and stronger mixing in deeper waters. Further, diffusivity in the thermocline is typically an order of magnitude less than the Munk (1966) 1 cm^2s^{-1} estimate (Ledwell et al., 1993, 1998; Kunze et al., 2006). Gregg (1987) found large spatial variation with a best guess of $\kappa_v \leq$ 0.1 cm^2s^{-1} in the upper thermocline. Polzin et al. (1997), St. Laurent and Simmons (2006), and Waterhouse et al. (2014) all show similar variation in vertical diffusivity with depth. In particular, St. Laurent and Simmons (2006) find κ_v on the order of 0.1 cm^2s^{-1} in the thermocline and 1-10 cm^2s^{-1} in deep and bottom waters and Waterhouse et al. (2014) estimate $\kappa_v = 0.3$ cm^2s^{-1} above 1000 m, 4.3 cm^2s^{-1} below 1000m and 3.3 cm^2s^{-1} for the full depth of the ocean.

In addition to direct calculation from observations, κ_v can be estimated using models. In particular, many EMICs have a parameter that sets the vertical diffusion coefficient. By sampling over a large range of κ_v values in the model and comparing simulations to past climate, probability distributions for κ_v are estimated. As noted in Olson et al. (2012), κ_v is a model-specific parameter, making direct comparison between models difficult. For similar reasons, comparisons between model parameterizations and observations are non-trivial. Using a one-dimensional model, Hansen et al. (1984) found that a global mean κ_v of 1 cm^2s^{-1}, the estimate from Munk (1966), matched well the average mean perturbation of mixed-layer temperature. Kriegler (2005), Tomassini et al. (2007), Forest et al. (2008), Lorenz et al. (2010), and Olson et al. (2012) all use models with an adjustable vertical ocean diffusivity parameter. In general, the estimated values are similar to those found in the thermocline, with estimates typically in the 0.1-0.5 cm^2s^{-1} range. However, Kriegler (2005) and Forest et al. (2008) do not rule out $\kappa_v >$ 1 cm^2s^{-1} in their 95- and 90-percent confidence intervals, respectively. Using the same model as Forest et al. (2008) but different observational constraints, Libardoni and Forest (2011) does not rule out κ_v above 25 cm^2s^{-1}.

1.7 Estimates of Aerosol Forcing

Aerosols in the atmosphere reduce the net radiative forcing on the climate system (Shindell et al., 2013; Boucher et al., 2013). Estimates of the forcing have been made using AOGCMs, observations, and simpler climate models. Boucher et al. (2013) discusses the difference between the radiative forcing and effective radiative forcing due to aerosols. Radiative forcing (RF) is the instantaneous change in forcing due to the direct interaction between radiation and aerosols. The effective radiative forcing (ERF) is the change in forcing after the atmosphere and clouds have adjusted to the presence of aerosols. RF and ERF due to aerosols in AOGCMs are estimated as -0.35 ± 0.5 and -0.45 ± 0.5 Wm^{-2}, respectively (Boucher et al., 2013). For models in the Atmospheric Chemistry and Climate Model Intercomparison Project, Shindell et al. (2013) calculates the spread in model aerosol RF and ERF to be -0.49 to -0.06 and -1.5 to -0.8 Wm^{-2}, respectively.

Satellite measurements of aerosol optical depth typically inform observational estimates of the aerosol radiative forcing. Bellouin et al. (2005) estimate an all sky direct radiative forcing of -0.8 ± 0.1 Wm^{-2} and a clear sky direct radiative forcing of -1.9 ± 0.3 Wm^{-2}. They note that these values are typically greater than model estimates of radiative forcing. In a synthesis of satellite-based estimates, Yu et al. (2009) found the direct radiative effect of aerosols in the range -1.9 to -0.9 Wm^{-2}. Su et al. (2013) estimates all sky and clear sky global mean aerosol forcings of -0.51 or -0.17 Wm^{-2} and -1.55 or -1.19 when combining different satellite aerosol optical depth datasets with cloud observations. Using reanalysis datasets for aerosol optical depth, Bellouin et al. (2013) estimates a global average aerosol radiative forcing of -0.7 ± 0.3 Wm^{-2} between 2003 and 2010. Murphy et al. (2009) use observed forcing patterns to estimate the aerosol forcing as the residual of the observed net forcing and all other measured forcings. An aerosol forcing of -1.1 ± 0.4 Wm^{-2} is estimated with this method.

Energy balance models and EMICs typically include a time series for the aerosol forcing that is scaled by changing the amplitude of the pattern with time. Similar to estimates of vertical ocean diffusivity, the aerosol forcing parameter in these models is model-specific and does not lend itself to clear comparisons amongst models and with observations. This is because the aerosol scaling pattern represents different forcings and the amplitude of the pattern is set for different time periods. For example, while Andronova and Schlesinger (2001) scale the natural and anthropogenic aerosol direct and indirect forcings by adjusting the amplitude in 1990, the aerosol parameter in Knutti et al. (2002) is scaled in 2000 and represents the indirect aerosol effect and any other forcing not explicitly represented in the model. With these differences in mind, estimates of aerosol forcing from energy balance models and EMICs fall in the ranges -1.3 to -0.54 Wm^{-2} (Andronova and Schlesinger, 2001), -1.2 to 0 Wm^{-2} (Knutti et al., 2002), -1.53 to -0.33 Wm^{-2} (Kriegler, 2005), -0.83 to -0.19 Wm^{-2} (Libardoni and Forest, 2011), and -1.7 to -0.4 Wm^{-2} (Skeie et al., 2014). Despite the difficulty in comparing across models, accounting for the uncertainty in the radiative forcing is vital, as under sampling the range can cut off the upper tail of climate sensitivity estimates (Tanaka et al., 2009).

1.8 Evaluation of Model Performance

Throughout the dissertation, we derive probability distribution functions for model parameters. To derive these functions, model output for each run is evaluated against time series of observed climate change. A given model run can be evaluated through the use of a goodness-of-fit statistic

$$r^2 = (\mathbf{x} - \mathbf{y})^T \mathbf{C}_\mathbf{N}^{-\mathbf{1}} (\mathbf{x} - \mathbf{y}), \tag{1.5}$$

where $\mathbf{x}$ and $\mathbf{y}$ are n-length vectors of model output and observed data, respectively, and $\mathbf{C}_N^{-1}$ is the inverse of the noise-covariance matrix. In its simplest form, the r^2 statistic is the sum of n terms. 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 different than the coefficient of determination for the goodness-of-fit of a linear model. In a linear model, high values of r^2 indicate a good fit to the model. As will be made clear later in this section, low values of r^2 indicate a good fit between the model output and the observations.

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 have these natural variability patterns (also called internal variability) embedded in them. In an effort to estimate this background internal variability, control run data are obtained by running a climate model for thousands of years with no external forcings. Estimates of the unforced temperature patterns used as model diagnostics are then estimated by extracting samples of the patterns from the control simulation. Spatial and temporal correlations in the temperature patterns from these samples are then used to estimate the noise-covariance matrix, $\mathbf{C}_\mathbf{N}$.

Through eigenvalue decomposition, $\mathbf{C}_\mathbf{N}$ is decomposed into a set of orthogonal eigenfunctions that describe the internal climate variability. The eigenfunctions are also called empirical orthogonal functions (EOFs) and are directly related to the singular value decomposition of the matrix. 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). 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. The removal of these patterns is justified by noting that they contribute minimally to the total variability. This process represents a truncation of the EOF decomposition and reduces the number of terms in the goodness-of-fit sum.

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. Furthermore, the rotation and normalization ensure that each of the elements in the $\mathbf{x} - \mathbf{y}$ vector are independent and identically distributed random variables from a normal distribution. This leads to each element of the $\mathbf{x} - \mathbf{y}$ pattern in the rotated space having equal weight in the goodness-of-fit calculation. Models that match the observations well have small residuals in each term, resulting in low values of r^2. Thus, low r^2 values indicate a good model fit and a high likelihood that the simulated climate matches the climate system. Similarly, higher r^2 values indicate poor model fits and translate to low likelihoods that the model simulates the climate system.

1.9 Outline of the Dissertation

In this dissertation, we take a systematic and in depth look at each of the terms in Equation 1.5 to address four main research questions and three sub questions:

1. What is the joint probability distribution function for climate sensitivity, effective ocean diffusivity, and the net aerosol forcing when MESM is constrained by surface temperature patterns and global mean ocean heat content?

    (a) How do new forcing time series and an updated land surface model change the probability distributions?
    (b) How does the emergent behavior of MESM compare to the emergent behavior of IGSM? Namely, how do estimates of TCR and sea level rise differ between the models?

2. How do the probability distributions change when recent temperature trends are added to the model diagnostics?
    (a) Do the distributions differ depending upon whether recent trends extend the diagnostics or whether recent trends replace data at the beginning of the diagnostics?
3. Are the results sensitive to the internal variability estimates used to calculate the noise-covariance matrix?
    (a) How can variability estimates from multiple sources be combined to yield more robust probability distributions?
4. How do estimates of transient climate response change in response to changes in the joint probability distributions?

Chapters 2-4 in the dissertation are intended to be stand-alone manuscripts and will be submitted to peer-reviewed journals. The dissertation author, Alex Libardoni, will be the lead author on all three papers, has contributed most of the ideas and experimental designs, and prepared all of the original manuscripts. In Chapter 2, we perform a baseline evaluation of the change in the probability distribution due to switching from IGSM to MESM. We run a large ensemble using MESM and compare the joint probability distribution derived from IGSM using the same model runs and model diagnostics. We also run an ensemble of transient climate simulations to derive a response surface for transient climate response using MESM. This response surface is used throughout the dissertation.

In Chapter 3, we propose new methodology to improve the estimates of the model parameters. In particular, we use radial basis functions to interpolate goodness-of-fit statistics to regions of the parameter space where the model has not been run. Second, we account for uncertainty in surface temperature observations by merging the probability distributions derived from four different datasets into a single estimate. Using the new methods, we show how the distributions change when more recent temperature changes are systematically added to the diagnostics. We show that the results are sensitive to whether the model diagnostics are lengthened while including recent changes or whether the length of the diagnostic is kept constant. Lastly, we assess the impact of including spatial patterns of surface temperature change in the model diagnostics by deriving distributions using global mean surface temperature, hemispheric mean temperatures, and the temperature pattern in four equal-area zonal bands.

In Chapter 4, we show the sensitivity of the distributions to the natural variability estimate used to derive the noise-covariance matrix. Using the pre-industrial control run simulations of 25 different AOGCMs, we estimate many realizations of unforced climate variability and use the noise-covariance matrices to derive new distributions. We merge the variability estimates across models with similar characteristics to test whether more stable distributions are derived when the sample size of variability estimates increases.

Throughout the dissertation, we derive TCR distributions from the joint probability distributions estimated in each study using the functional fit derived in Chapter 2. In Chapter 5, we summarize our findings and suggest avenues of future research.

Chapter 2

Baseline Evaluation of Model Parameter Estimates in the Updated MIT Earth Systems Model

2.1 Introduction

Equilibrium climate sensitivity (ECS), the equilibrium global-mean surface temperature change due to a doubling of atmospheric carbon dioxide concentrations, is a climate system property that has been widely studied and strongly influences future climate projections. One of the complexities of ECS is that it is a function of many feedbacks and processes that act on different spatial and temporal scales. In particular, the lapse rate, water vapor, snow-albedo, and cloud feedbacks play especially critical roles (Bony et al., 2006). Given its influence on future climate change, many studies using a range of methods have attempted to estimate ECS.

One class of studies estimates ECS directly from observations using a global energy budget approach (Gregory et al., 2002; Otto et al., 2013; Lewis and Curry, 2014; Masters, 2014). These studies calculate probability distributions of ECS from estimates of global mean surface temperature change, heat stored in the ocean, and changes in radiative forcing and the uncertainties in their measurements. A second class of studies uses simplified climate models such as Earth System Models of Intermediate Complexity (EMICs) or energy balance models (e.g., Forest et al., 2002, 2008; Libardoni and Forest, 2013; Olson et al., 2013). Taking advantage of the computational efficiency of the simplified models, these studies run large ensembles with a range of climate sensitivity values in addition to other relevant factors, such as the ocean diffusivity and a measure of the net aerosol forcing. By comparing model runs to observations and evaluating how well individual model runs match the past, estimates of ECS and other parameters are then presented as probability distributions.

Transient climate response (TCR) provides a second metric for future climate change and is defined as the global mean surface temperature change at the time of CO_2 doubling in response to CO_2 concentrations increasing at the rate of 1% per year. CO_2 doubling occurs in year 70 of this scenario, making TCR a shorter-term assessment of climate change than ECS. Unlike ECS, which requires reaching an equilibrium state, TCR is estimated while the climate system is still adjusting to a time-dependent forcing. There is a constant evolution in the strength and activity of processes and feedbacks in both the atmosphere and the ocean as the climate system adjusts to reach equilibrium. Due to the long time scales required to reach equilibrium, Allen and Frame (2007) argue that we should focus on estimating TCR, which is more policy-relevant than ECS. Estimates of TCR can be made from current historical observations and are more meaningful on the decadal time scale, whereas even if the equilibrium response is known, it may never be reached. However, even if more focus is placed on TCR than ECS, the two are still closely linked. Warming on time scales relevant to estimating TCR is related to the sensitivity of the climate system to external forcings and the coupling between the atmosphere and the ocean. When considering atmosphere-ocean interactions, we know that TCR depends on both climate sensitivity and the rate at which heat is mixed into the deep ocean (Sokolov et al., 2003; Andrews and Allen, 2008).

One EMIC that has been extensively used in studies estimating ECS and TCR is the Earth system component of the Massachusetts Institute of Technology (MIT) Integrated Global Systems Model (IGSM, Sokolov et al., 2005). Forest et al. (2002, 2006, 2008) and Libardoni and Forest (2011, 2013) estimated the joint probability distribution for climate sensitivity and other model parameters in IGSM. Each study used similar, but not identical, versions of IGSM with changes both to key components of the model and to the input data used to force the model. Climate change diagnostics were also modified in the studies. The Earth system component of IGSM has undergone further development and a new, updated version incorporated into the integrated framework. This study serves as a baseline evaluation of how probability distributions for the model parameters change as a result of updating the Earth system component. In the past, "IGSM" has been used to reference both the fully integrated model as well as the standalone Earth system component. We follow this convention and refer to the older version of the Earth system model as IGSM, and we refer to the updated version of the model as the MIT Earth System Model (MESM). We derive a new joint probability distribution by closely following the methods of Libardoni and Forest (2011) to show the impact that the new version of the model has on the parameter estimates and find that the new version of the model leads to higher climate sensitivity estimates in addition to shifts in the distributions of the other model parameters. The effects on the parameter distributions due to changing observations and temperature metrics are addressed in Chapter 3. We also show here how the emergent behavior of MESM compares to the older IGSM by running a new set of transient simulations and calculating how the response surfaces for TCR and sea level rise depend on ECS and ocean diffusivity.

In Section 2.2, we give a brief description of the MIT modeling framework and the differences between IGSM and MESM. We describe the process for deriving the joint probability distribution function used in Libardoni and Forest (2011) and the modifications implemented in this study in Section 2.3. Parameter distributions and response surfaces are presented in Section 2.4, and conclusions in Section 2.5.

2.2 Model

The coupled atmosphere-ocean-land model of the updated MIT Earth System Model replaces the version described in Sokolov et al. (2005). The first update to the model was the incorporation of a new land surface model. The Community Land Model (CLM) version 3.5 (Oleson et al., 2008) replaced CLM version 2.1 to improve estimates of the surface heat balance in the model. Additionally, the forcings used to drive the model until now (Forest et al., 2006) were extended and, in some cases, new data sources were used. Greenhouse gas concentrations and stratospheric aerosols from volcanic eruptions were obtained from the National Aeronautics and Space Administration Goddard Institute for Space Studies modeling group forcing suite. The procedure for updating the greenhouse gas emissions from Hansen et al. (2007) and the volcanic aerosol forcing from Sato et al. (1993) was described in Miller et al. (2014). Updates included incorporating data from more observational sources and extending the length of the datasets. Sulfate aerosol loading from Smith et al. (2011) was extended to 2011 by Klimont et al. (2013). The Kopp and Lean (2011) solar irradiance dataset replaced the Lean (2000) dataset. Lastly, the ozone concentration database developed by the Atmospheric Chemistry and Climate initiative (AC&C) and Stratospheric Processes and their Role in Climate project (SPARC) ozone concentration database (Cionni et al., 2011) that was developed in support of the Coupled Model Intercomparison Project phase 5 (CMIP5) replaced the concentration data used in Forest et al. (2006). The concentrations in the dataset, hereafter referred to as AC&C/SPARC, drive the tropospheric and stratospheric ozone forcing in the radiation code. In Section 2.4, we show the differences between the old and new datasets for those forcings where the data sources have changed, namely solar and ozone.

Three model parameters that impact the climate system response are easily modified in MESM. These parameters are the effective climate sensitivity (ECS), the effective ocean diffusivitiy (K_v), and the net aerosol scaling factor (F_{aer}). ECS is changed by adjusting the strength of the cloud feedback at different levels in the model (Sokolov, 2006; Sokolov and Monier, 2012). K_v represents the vertical diffusion of heat anomalies into the deep ocean by all mixing processes and tends to be larger than typical ocean diffusivity values which represent the diffusion of heat alone (Sokolov et al., 2003). The mixing pattern is prescribed spatially with stronger mixing in the polar regions and weaker mixing near the equator. K_v represents the global mean diffusion rate and the spatial pattern is scaled to obtain the desired value. The anthropogenic aerosol forcing used in the model is prescribed by a latitude-dependent pattern that differs over land and ocean and is used as an estimate of all unmodeled forcings in the simulations (Forest et al., 2001). This pattern is held fixed spatially but scaled temporally by estimated emissions of sulfur dioxide. F_{aer} sets the amplitude of the pattern in the 1980s.

2.3 Methods

We follow closely the methods of Libardoni and Forest (2011) with two notable changes. First, we run the model over a wider range of parameter values and on a more regular grid. Climate sensitivity is sampled from 0.5 to 10.0 °C in increments of 0.5 °C through the adjustment of the cloud feedback, the square root of ocean diffusivity is sampled from 0 to 8 cm $s^{-1/2}$ in increments of 1 cm $s^{-1/2}$, and the aerosol forcing amplitude is sampled from -1.75 to 0.5 Wm^{-2} in increments 0.25 Wm^{-2}. By choosing this sampling strategy, we have increased the number of runs from 640 with IGSM to 1800 runs with MESM, widened the range of parameter values sampled, and increased the density of model runs within the parameter space (Figure 2.1.).

As a second change, we reduce the number of diagnostics used to evaluate model

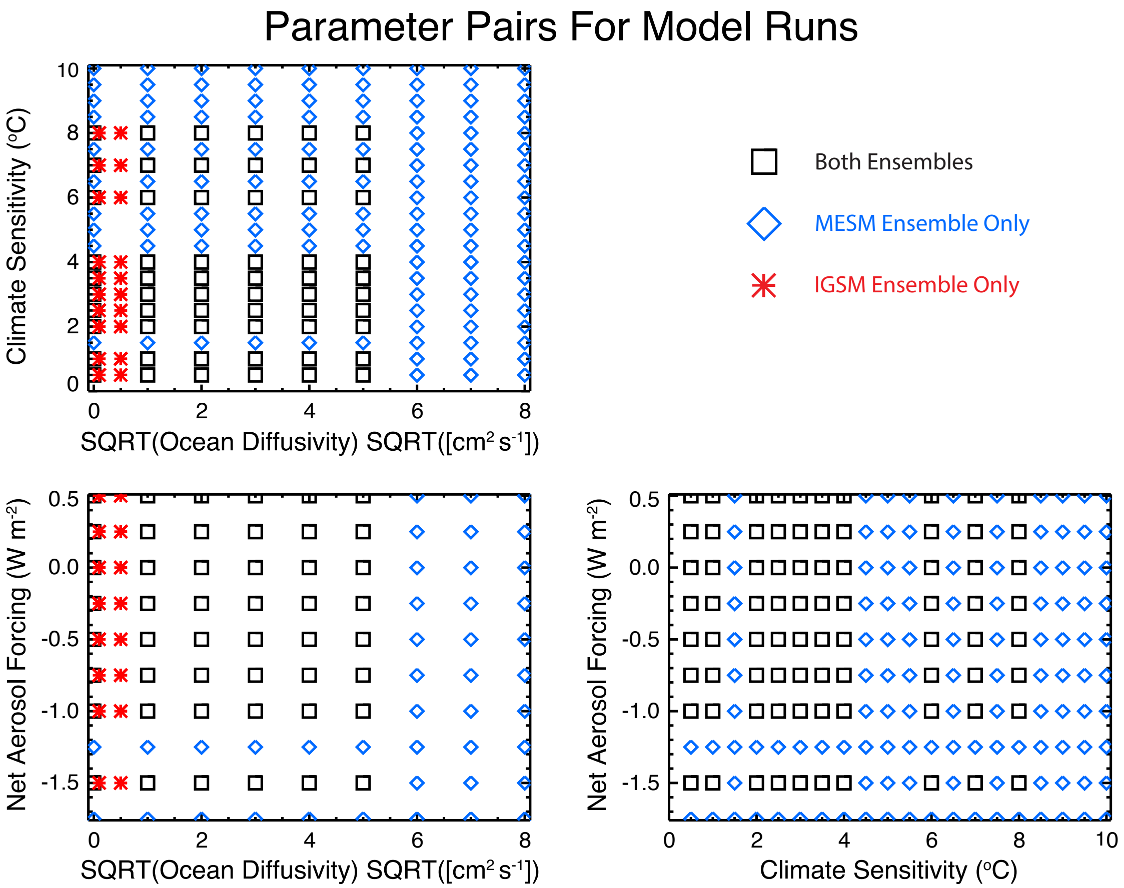

Figure 2.1: Parameter pairings where the models have been run. Points in black are common to both the IGSM and MESM ensembles. Blue points are unique to the IGSM ensemble and red points are unique to the MESM ensemble.

performance. We omitted the upper-air temperature diagnostic because it is highly correlated with the surface temperature diagnostic (Lewis, 2013). This leaves two temperature diagnostics for evaluating model performance: (1) decadal mean surface air temperature anomalies from 1946-1995 with respect to a 1906-1995 climatology in four equal-area zonal bands, and (2) the linear trend in global mean ocean heat content from 1955-1995 in the 0-3 km layer. As in Libardoni and Forest (2011), we use five surface temperature datasets (Jones and Moberg, 2003; Brohan et al., 2006; Smith et al., 2008; Hansen et al., 2010) and one ocean heat content dataset (Levitus et al., 2005) as observations.

2.4 Results

To identifty changes in the forcing time series used to drive the model, we compare the input forcings for the two components for which we have changed datasets. In Figure 2.2, we show the old and new solar forcing time series. We see that the biggest difference observed in the solar irradiance time series is a bias towards lower values when using the Kopp and Lean (2011) data. The bias is relatively constant at approximately 4.5 Wm^{-2} until 1920, then increases towards 5.0 Wm^{-2} moving forward in time. The mean bias is accounted for in the Q-flux adjustment in the mixed-layer ocean model which specifies the vertically-integrated horizontal heat transport in the mixed layer required to maintain historical sea surface temperatures (Sokolov et al., 2005). However, because the Q-flux is calculated offline from control simulations, the pattern is fixed throughout the run. Any time-varying change to an input forcing cannot be accounted for in the Q-flux calculation. Thus, the growth of the low bias means that the solar forcing weakens with time beginning in 1920 in the new suite of forcings.

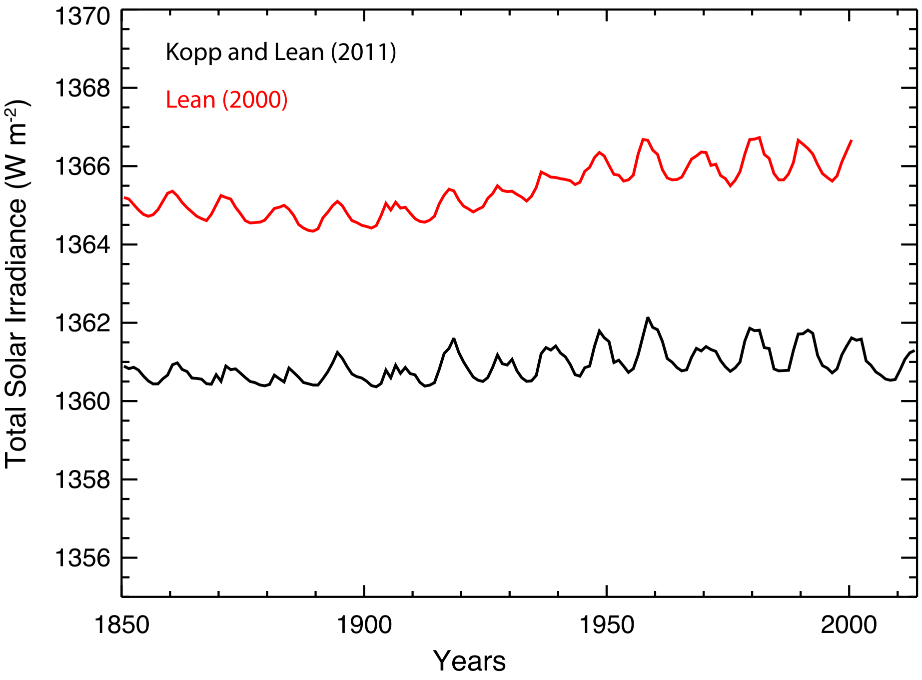

Figure 2.2: Annual mean total solar irradiance. The bias between the Lean (2000) and Kopp and Lean (2011) datasets leads to a reduction in radiative forcing in the new forcing suite.

We observe that the ozone concentrations estimated from the AC&C/SPARC dataset differ in both space and time when compared to the previous concentrations used with IGSM (Figure 2.3). One clear difference is that the AC&C/SPARC dataset introduces more temporal variability in stratospheric ozone concentrations (which we approximate as pressure levels above 200 mb) prior to 1950. Post-1950, AC&C/SPARC tends to have lower ozone concentrations in the stratosphere and slightly greater concentrations in the troposphere (levels below 200 mb). However, similar to with the solar forcing, we are concerned with the temporal change in the forcing imposed by the ozone concentrations, rather than the relative magnitude of the concentrations across datasets. Beginning in 1900, tropospheric ozone concentrations increase less rapidly in the AC&C/SPARC dataset when compared to the IGSM dataset. Differences in stratospheric ozone con-

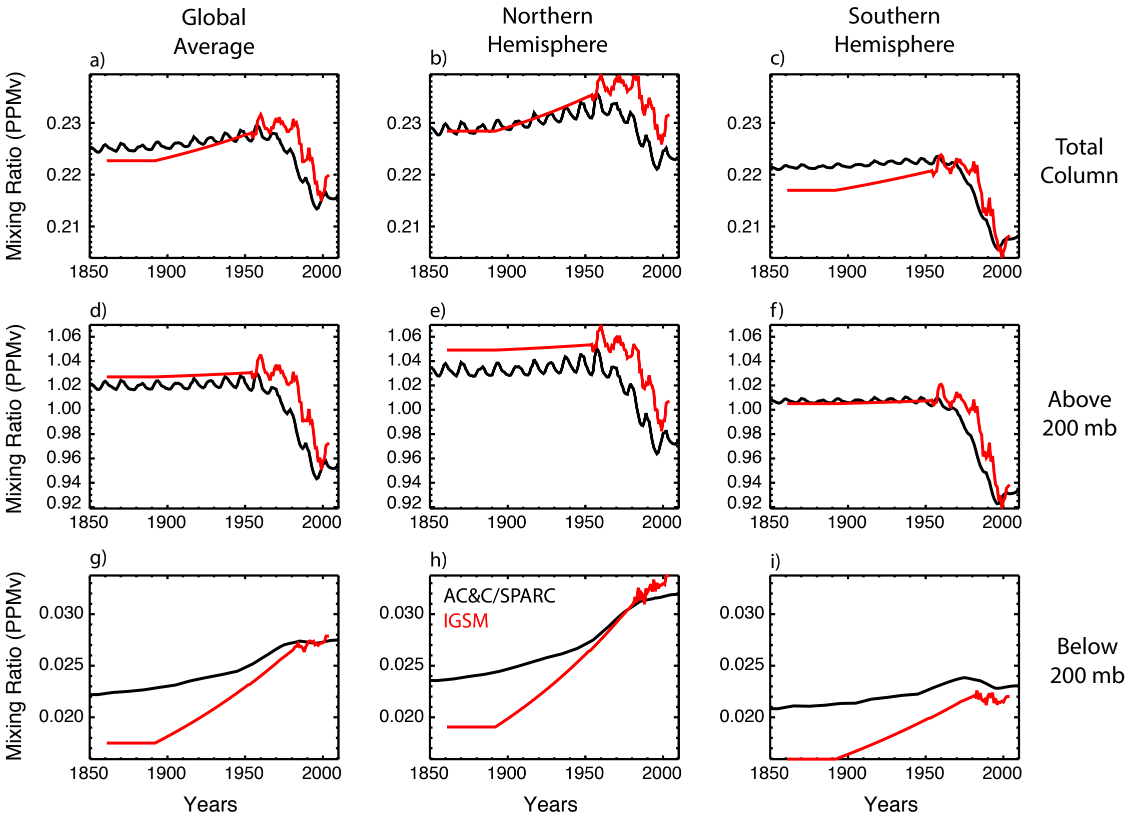

Figure 2.3: Ozone concentration in the old IGSM time series (red) and the Cionni et al. (2011) AC&C/SPARC concentrations (black). (a-c) Annual mean ozone mixing ratio in the total column in the global average (a), northern hemisphere (b), southern hemisphere (c). (d-f) As in (a-c) but for the average above 200 mb. (g-i) As in (a-c) but for the average below 200 mb.

centrations remain relatively constant until 1950, but then decrease at a slower rate in the AC&C/SPARC time series. These patterns are generally consistent in the global and hemispheric means. When considered separately, increased tropospheric ozone concentrations tend to increase radiative forcing (Stevenson et al., 2013) and decreased stratospheric concentrations tend to increase radiative forcing (Conley et al., 2013). Thus, the less rapid increase in tropospheric ozone concentration and less rapid decrease in stratospheric ozone concentration in the AC&C/SPARC dataset both contribute to a weaker radiative forcing over the historical period in the new suite of forcings.

With the input forcings documented, we focus on deriving probability distributions for the model parameters. We first test the impact of omitting the upper-air diagnostic. Starting from the distributions calculated in Libardoni and Forest (2011), we derive new distributions based only on the surface temperature and ocean heat content diagnostics presented in Section 2.3. We show that reducing the number of diagnostics from three to two has little impact on the parameter estimates (Table 2.1). We only present comparisons for ECS and F_{aer} because distributions of K_v were poorly constrained in Libardoni and Forest (2011) and no uncertainty bounds were given. In general, ECS estimates tend to be slightly lower when using only two diagnostics and aerosol estimates are nearly unchanged. Further, the relationships between the distributions with respect to surface dataset are unchanged. Because the changes using only two diagnostics are minimal and do not change any conclusions from the original study, we justify the removal of the upper-air diagnostic.

We next evaluate the impacts that changing the model from IGSM to MESM and updating the forcing suite have on the parameter distributions by comparing model output from each ensemble member against the temperature diagnostics discussed in Section 2.3. Following the methods outlined in Libardoni and Forest (2011), we calculate goodness-of-fit statistics across all runs for each diagnostic and convert them to a joint probability distribution function for the model parameters. Marginal probability distributions for individual parameters are then calculated by integrating the joint distribution over the other two parameters. We present the new distributions in Figure 2.4 and observe significant differences between distributions derived using IGSM and those derived using MESM with the updated forcings (Table 2.2). Across all datasets, climate sensitivity distributions shift towards higher values and the uncertainty bounds encompass a wider range. When considering the 90-percent confidence intervals from across the distributions derived from each surface dataset, we find climate sensitivity now lies between 1.3 and

Table 2.1: 90-percent confidence intervals for ECS and F_{aer}. Distributions that include the upper-air diagnostic are from Libardoni and Forest (2011) and distributions with two diagnostics exclude the upper-air diagnostic.

Surface Temperature Dataset	# Diagnostics	ECS (°C)		F_{aer} (Wm^{-2})	
		5%	95%	5%	95%
HadCRUT2[1]	3	2.0	5.3	-0.19	-0.70
	2	1.9	5.2	-0.19	-0.71
HadCRUT3[2]	3	1.9	5.1	-0.22	-0.74
	2	1.7	5.0	-0.38	-0.79
NCDC[3]	3	1.8	4.7	-0.37	-0.78
	2	1.6	4.8	-0.38	-0.79
GISTEMP250[4]	3	1.3	3.6	-0.32	-0.83
	2	1.1	4.0	-0.35	-0.83
GISTEMP1200[5]	3	1.2	3.4	-0.33	-0.80
	2	1.0	3.7	-0.35	-0.83

[1] Hadley Centre Climatic Research Unit Temperature version 2

[2] Hadley Centre Climatic Research Unit Temperature version 3

[3] National Climatic Data Center merged land-ocean dataset

[4] GISS Surface Temperature Analysis with 250 km smoothing

[5] GISS Surface Temperature Analysis with 1200 km smoothing

5.7 °C, as opposed to the estimated interval of 1.2 to 5.3 °C from Libardoni and Forest (2011). While the uncertainty bounds are still wide compared to other parameters, we observe that K_v is now better constrained with MESM. The distributions of K_v derived using the GISTEMP datasets are still unconstrained with upper tails extending to the edge of the parameter domain, but all other datasets now show an upper bound well within the ensemble range. We also observe a marked shift in the aerosol estimates. When MESM is used with the updated forcing suite, there is a sizeable shift towards weaker aerosol forcing across all datasets. Whereas past estimates put net aerosol forcing between -0.83 and -0.19 Wm^{-2}, our new estimate of aerosol forcing is between -0.53 and -0.03 Wm^{-2}.

The shifts we observe in the parameter estimates are consistent with the changes in the input forcings. Both the solar and ozone forcing patterns lead to a reduction in their contribution to the global radiation budget and decrease the net radiatiave forcing on the planet. Because the diagnostics do not change, model runs with a weaker external forcing are compared against the same observed temperature patterns. Weaker increases

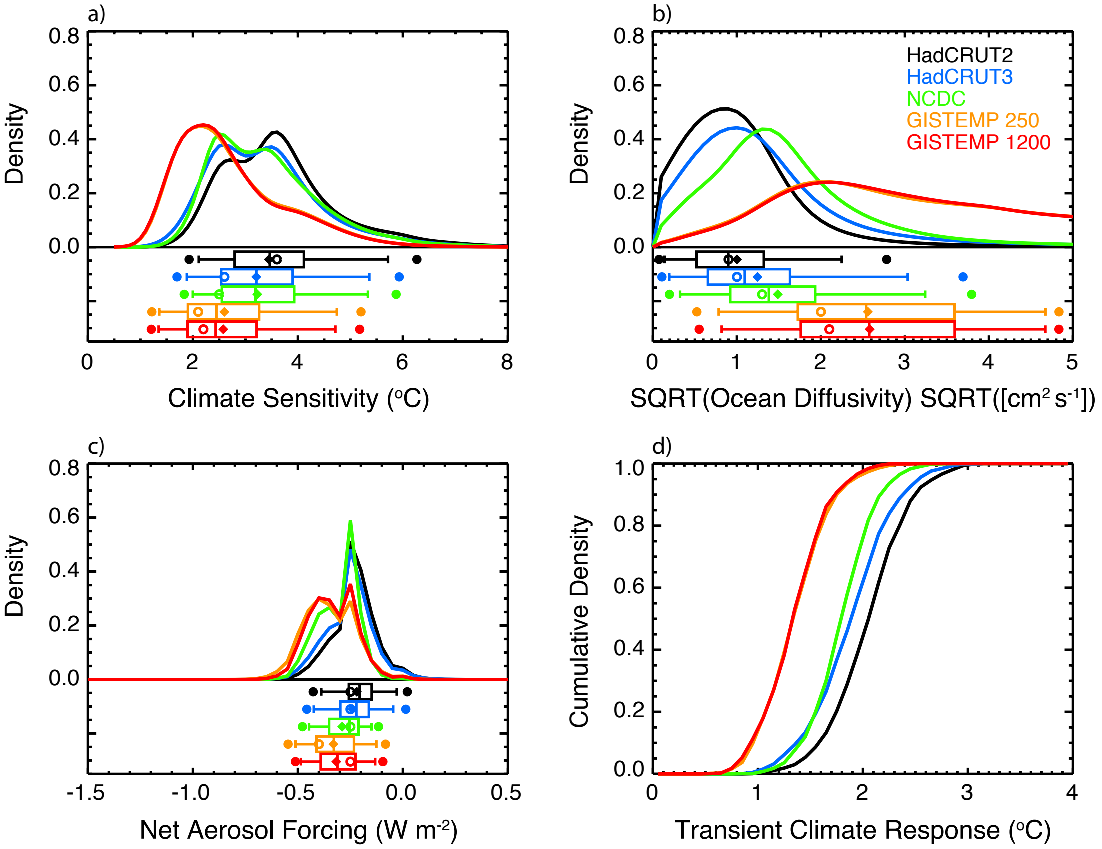

Figure 2.4: Marginal probability distribution functions and TCR cumulative distribution functions derived using the HadCRUT2, HadCRUT3, NCDC, GISTEMP 250, and GISTEMP 1200 surface temperature datasets. (a) ECS, (b) K_v, and (c) F_{aer}. Whisker plots indicate boundaries for the 2.5-97.5 (dots), 5-95 (vertical bar ends), 25-75 (box ends), and 50 (vertical bar in box) percentiles. Distribution means are represented by diamonds and modes are represented by open circles. (d) TCR CDFs are derived from 1000 member Latin Hypercube samples from the joint parameter distributions and the TCR(ECS, $\sqrt{K_v}$) functional fit.

in external forcing require higher climate sensitivity to match the same warming trend. In the model, the aerosol forcing pattern is a negative term in the global energy budget. It should follow that if, as noted previously, there is weaker net forcing due to the changes in forcing datasets, our estimates of the aerosol amplitude should become less negative. From these arguments, combinations of higher climate sensitivity and weaker aerosol forcing offset the impact of a weaker external forcing. We observe these shifts in the distributions when comparing the distributions derived from the old and new models

Table 2.2: 90-percent confidence intervals and means for climate sensitivity (ECS), ocean diffusivity (K_v), and net aerosol forcing (F_{aer}). Surface temperature datasets are the same as in Table 2.1.

Surface Temperature Dataset	Model and Runs	ECS (°C)			$\sqrt{K_v}$ (cm $s^{-1/2}$)			F_{aer} (Wm^{-2})		
		5%	95%	Mean	5%	95%	Mean	5%	95%	Mean
HadCRUT2	Full IGSM	1.9	5.2	3.0	0.1	2.1	0.9	-0.19	-0.71	-0.46
	Subsampled IGSM	1.9	5.2	3.0	0.1	2.1	0.9	-0.16	-0.71	-0.45
	Full MESM	2.1	5.7	3.5	0.1	2.3	1.0	-0.03	-0.39	-0.22
	Subsampled MESM	2.1	5.7	3.4	0.1	2.2	1.0	-0.03	-0.39	-0.22
HadCRUT3	Full IGSM	1.7	4.0	2.8	0.2	2.9	1.2	-0.22	-0.75	-0.50
	Subsampled IGSM	1.7	4.0	2.8	0.2	2.9	1.2	-0.20	-0.75	-0.49
	Full MESM	1.9	5.4	3.2	0.2	3.6	1.3	-0.05	-0.43	-0.24
	Subsampled MESM	1.9	5.4	3.2	0.2	3.0	1.2	-0.05	-0.42	-0.24
NCDC	Full IGSM	1.6	4.8	2.7	0.3	3.7	1.6	-0.38	-0.79	-0.59
	Subsampled IGSM	1.6	4.8	2.7	0.3	3.7	1.6	-0.36	-0.79	-0.58
	Full MESM	2.0	5.4	3.2	0.3	3.7	1.6	-0.15	-0.45	-0.29
	Subsampled MESM	2.0	5.3	3.2	0.3	3.2	1.5	-0.15	-0.45	-0.29
GISTEMP 250	Full IGSM	1.1	4.0	2.1	0.7	4.8	2.7	-0.35	-0.86	-0.61
	Subsampled IGSM	1.1	4.0	2.1	0.6	4.8	2.7	-0.35	-0.86	-0.60
	Full MESM	1.3	4.8	2.6	0.8	7.3	3.5	-0.13	-0.53	-0.34
	Subsampled MESM	1.4	4.7	2.6	0.8	4.7	2.6	-0.13	-0.51	-0.33
GISTEMP 1200	Full IGSM	1.0	3.7	1.9	0.8	4.9	3.1	-0.35	-0.83	-0.56
	Subsampled IGSM	1.0	3.7	1.9	0.7	4.9	3.1	-0.35	-0.82	-0.56
	Full MESM	1.3	4.8	2.6	0.8	7.3	3.5	-0.14	-0.49	-0.33
	Subsampled MESM	1.3	4.7	2.6	0.8	4.7	2.6	-0.14	-0.49	-0.32

(Table 2.2).

To test whether the differences observed in the parameter estimates were due to the model update, rather than the increased density of model runs, we subsampled each ensemble at the 480 parameter settings where they overlap (see Figure 2.1). We summarize the distributions in Table 2.2 and see that there is very little sensitivity when the ensembles are subsampled. Across all datasets, the distributions we derive using the full 640-member IGSM ensemble and those we derive using the 480-member IGSM ensemble are nearly identical for all three parameters. The same is true for the MESM ensemble, except for the distributions we derive for K_v. We consistently estimate a smaller upper bound for K_v in the subsampled MESM ensemble compared to when the full MESM ensemble is used. This arises because the wider range of K_v sampled in the MESM ensemble does not artificially cut off the distribution for values of $\sqrt{K_v}$ greater than 5 cm $s^{-1/2}$. Thus, the differences we observe between the old and new ensembles are due to the differences between the model and forcing themselves and not the increased density of model runs.

To estimate TCR in MESM, we run a 372-member ensemble where all forcings are held fixed and carbon dioxide concentrations are increased by 1% per year. We calculate TCR by estimating the global mean temperature change from the beginning of the simulations at the time of CO_2 doubling. Concentrations double in year 70 and we estimate TCR as the average global mean temperature change in years 60-80 of the simulation. Temperature changes are calculated with respect to a control simulation with the same model parameters and all forcings held fixed. In a similar manner, we also estimate thermosteric sea level rise (SLR) at the time of doubling. Because all forcings except those attributed to CO_2 are fixed, each ECS-$\sqrt{K_v}$ pair yields a single TCR value and a single SLR value, independent of F_{aer}. We fit a third-order polynomial in ECS and $\sqrt{K_v}$ to the TCR and SLR values calculated from each run to derive a functional fit for all parameter pairs within the domain. From these fits, we derive response surfaces for each of the transient properties (Figure 2.5). For comparison, we also show the fit derived using the IGSM and its corresponding 1% per year runs, in addition to the differences between the two. Outside of the region where ECS is greater than 4 °C and $\sqrt{K_v}$ is less than about 0.5 cm $s^{-1/2}$, and away from the edges of the domain, TCR values from IGSM and MESM agree quite well. There is a similar pattern of agreement in the SLR response surface, with the biggest discrepancies occurring in the high ECS-high $\sqrt{K_v}$ region and near the edges of the parameter domain.

We use the response surface to derive probability distributions for TCR. From each of the joint probability distributions derived from the subsampled MESM ensemble, we draw a 1000-member Latin Hypercube Sample (McKay et al., 1979) of model parameters. The subsampled distributions are chosen so that we restrict the domain to that of the IGSM ensemble, allowing for a more direct comparison of the distributions. Otherwise, high K_v values that are within the domain of the functional fit to the MESM runs would be selected, for which there is no fit using the IGSM function. We map each of the ECS-$\sqrt{K_v}$ pairs onto the response surface to provide an estimate of TCR values. Binning the responses in a histogram with bin size = 0.1 °C allows a PDF to be calculated, and the resulting cumulative density functions are displayed in Figure 2.4d. Comparing TCR distributions for the IGSM and MESM ensembles shows a shift towards higher TCR with the latest results. When comparing the range of 90-percent confidence intervals derived using MESM to those from Libardoni and Forest (2011), we find that TCR estimates increase from 0.87-2.31 °C using IGSM to 0.90-2.72 °C using MESM. We have shown previously that the marginal distributions of $\sqrt{K_v}$ are similar between the two models, indicating that this shift towards higher TCR is driven by the higher ECS estimates derived from MESM.

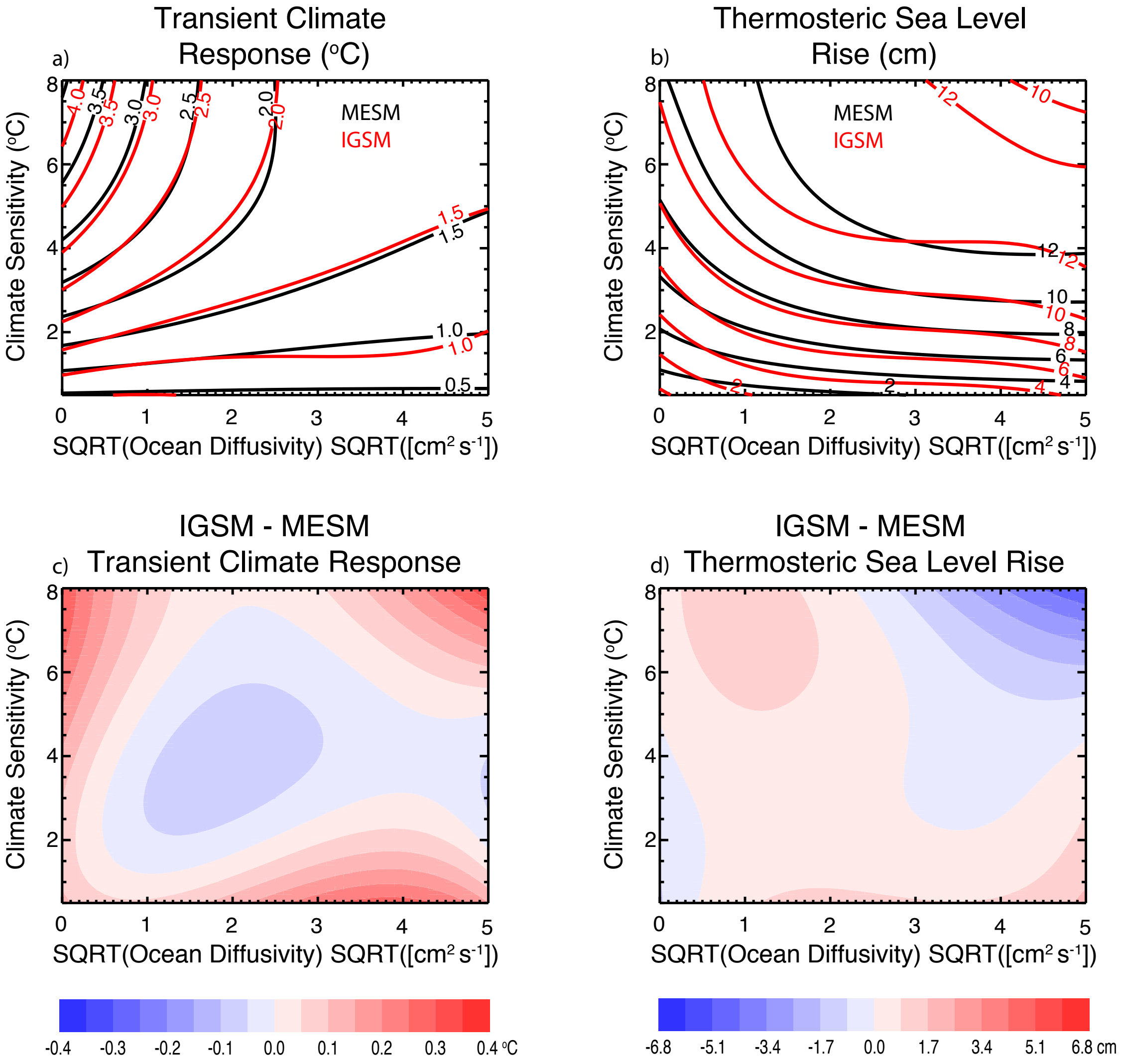

Figure 2.5: Model response surfaces for (a) TCR and (b) thermosteric sea level rise. Contours for the MESM response surfaces are shown in black and contours for the IGSM surfaces are shown in red. Differences between the fits are also shown (c and d).

2.5 Conclusions

By updating the model forcings and replacing CLM2.1 with CLM3.5, we identify the impact that the switch from the MIT Integrated Global Systems Model to the MIT Earth System Model has on the probability distributions of model parameters. The decreases in radiative forcing due to the new solar radiation data and the new ozone concentrations used to estimate the ozone forcing lead to a net energy deficit when compared to the replaced forcings. This drives an upward shift in our estimates of the 90-percent confidence interval for climate sensitivity from 1.2 to 5.3 °C to 1.3 and 5.7 °C, a better constraint on ocean diffusivity, and a decrease in the 90-percent confidence interval for the net anthropogenic aerosol forcing from between -0.83 and -0.19 Wm^{-2} to between -0.53 and -0.03 Wm^{-2}. One caveat of our analysis is that because we changed the forcings and CLM simultaneously, we cannot fully attribute the parameter shifts to the model forcings alone. We have shown the total effect of changing both the model and forcings on the parameter distributions, not the effects of the changes individually

Because TCR is independent of the input forcings, the only difference between the IGSM and MESM configurations in the transient simulations is the land surface model. By showing that the transient climate response surfaces derived from the two models differ only slightly, we provide evidence that the switch to CLM3.5 does not greatly impact the temperature evolution in the model. We have drawn Latin Hypercube Samples from the parameter distributions to provide estimates of TCR from the new response surface. Due to the shift towards higher climate sensitivity and slightly weaker ocean diffusivity, we observe an increase in our 90-percent confidence interval of transient climate response from 0.87-2.31 °C to 0.85-2.73 °C. By investigating the impact that the new forcings and a newer version of CLM have on the estimates of model parameters and TCR, we provide the inherent differences that are present when comparing distributions derived using IGSM and those derived from MESM.

Chapter 3

Estimates of Climate System Properties Incorporating Recent Climate Change

3.1 Introduction

Scientists, policy makers, and the general public are concerned with how surface temperature will change in the coming decades and further into the future. These changes depend on many aspects of the climate system. Among them are climate sensitivity and the rate at which heat is mixed into the deep ocean. Equilibrium climate sensitivity (ECS) represents the global mean surface temperature change that would be realized due to a doubling of CO_2 concentrations after equilibrium is reached. A shorter-term measure of climate sensitivity to greenhouse gas forcing is transient climate response (TCR), defined as the global mean surface temperature change at the time of CO_2 doubling in response to CO_2 concentrations increasing at the rate of 1% per year (Bindoff et al., 2013). Due to the climate system not being in equilibrium, interactions between the surface and the ocean lead to an exchange of energy. In such a scenario, TCR is a function of both the climate sensitivity and ocean circulation and mixing (Sokolov et al., 2003; Andrews and Allen, 2008).

The value of climate sensitivity is uncertain but the processes and feedbacks which set it must be accurately modeled to reliably predict the future. To this end, a number of studies have used Earth System Models of Intermediate Complexity (EMICs) to estimate probability distribution functions (PDFs) for the values of these climate system properties, in particular ECS, ocean diffusivity, and an estimate of the anthropogenic aerosol forcing (Forest et al., 2002; Knutti et al., 2003; Tomassini et al., 2007; Forest et al., 2008; Olson et al., 2012; Aldrin et al., 2012; Libardoni and Forest, 2013, and others). In these studies, EMICs are run for many combinations of the model parameters that set the climate system properties. Model output is then compared to historical temperature change to determine which states of the model climate best match the past.

Time series of surface temperature and ocean heat content are commonly used temperature diagnostics in the evaluation of model performance because they rule out different combinations of the parameters for being inconsistent with past climate (Urban and Keller, 2009). This helps to narrow the estimate of the parameters because only certain combinations lead to accurate representations of the past. Recent observations have shown that the rate of increase of global mean surface temperature has slowed despite the continued rise of global CO_2 concentrations (Trenberth and Fasullo, 2013). This warming hiatus has led some to claim that climate change is not a significant threat and that mitigative action is unnecessary, but has also led scientists to search for the reasons behind the slowdown. Cowtan and Way (2014) and Karl et al. (2015) argue that the hiatus is merely an artifact of the global observing system and that the perceived slowdown is due to incomplete coverage in the polar regions where temperatures have increased most rapidly. The slowdown has also been attributed to changes in the radiative forcing. In particular, its been argued that the forcing due to the sun, anthropogenic aerosols, and volcanoes all contributed to reduce global mean temperature in the 2000s (Huber and Knutti, 2014; Schmidt et al., 2014). Natural variability in the ocean has also been noted as a potential cause for the slowdown (Meehl et al., 2011; Huber and Knutti, 2014; Schmidt et al., 2014). In particular, Meehl et al. (2011) show that in a fully-coupled, three-dimensional climate model, hiatus periods in surface temperature are associated with enhanced mixing of heat below 300 m in the ocean. This finding is supported by recent observations showing that heat is accumulating more rapidly in the deep ocean (Levitus et al., 2012; Gleckler et al., 2016).

In this study, we first seek to improve the methods used in previous work (Forest et al., 2008; Libardoni and Forest, 2013, Chapter 2). Until now, ensembles from different versions of the MIT Integrated Global Systems Model (IGSM, Sokolov et al., 2005) have been used to vary model parameters for ECS, ocean diffusivity, and the net anthropogenic aerosol scaling factor using a gridded sampling strategy. To derive PDFs for the model parameters, metrics of model performance at parameter settings in between those where the model was run are estimated using two-dimensional interpolation algorithms. These algorithms are restricted to gridded samples and at times have led to PDFs that are not smooth. We propose here replacing spline interpolations with a radial basis function interpolation algorithm that leads to PDFs that are both true to the data and smooth. We show this by using the 1800-member ensemble of the MIT Earth System Model (MESM; Sokolov et al., in prep) described in Chapter 2 to derive PDFs for the three model parameters.

Using the updated methodology and the 1800 MESM runs, we answer the following questions: (1) how does the inclusion of more recent data change the PDFs of model parameters? and (2) what do we learn by including spatial information in the surface diagnostic? The inclusion of recent temperature trends can have a significant impact on the estimates of climate system properties (Urban et al., 2014; Johansson et al., 2015). The temperature pattern that the model output is compared against becomes more detailed as data is added and leads to the rejection of more model runs as being inconsistent with the observed records. This generally leads to both a shift in the estimation of a given property and a reduction in the uncertainty in the estimate. Urban et al. (2014) also showed that the ability to distinguish between different states of the climate increases as the length of the model diagnostic increases. Similar to Johansson et al. (2015), we identify the influence of including more recent data by systematically adding data to the time series.

Second, we show how including spatial variability in the surface temperature diagnostic can influence the parameter distributions. In almost all parameter estimation studies, global mean ocean heat content is used as one metric to evaluate model performance and is paired with a surface temperature diagnostic to further test the model runs. Typically, groups use time series of either global mean surface temperature (Knutti et al., 2002; Tomassini et al., 2007; Knutti and Tomassini, 2008; Urban and Keller, 2009; Olson et al., 2012) or hemispheric mean surface temperatures (Andronova and Schlesinger, 2001; Meinshausen et al., 2009; Aldrin et al., 2012; Skeie et al., 2014) as the surface diagnostic. Given the latitudinal resolution of MESM, we can estimate zonal temperature patterns beyond global and hemispheric means. In particular, we use a surface temperature diagnostic that consists of four equal-area zonal bands, allowing the observed amplification of polar warming to be included in the evaluation of model performance. We show the impact of the spatial structure of the surface diagnostic by deriving PDFs using global mean, hemispheric mean, and four zonal mean temperature diagnostics.

In Section 3.2, we introduce the general method for estimating probability distributions for the model parameters, describe the temperature diagnostics, and introduce an interpolation method for the likelihood function using radial basis functions. We present our main findings in Section 3.3 and finish with a summary and conclusions in Section 3.4.

3.2 Methods

As outlined in Section 3.1, we propose and implement a number of methodological changes designed to improve our estimates of the probability distributions of the model parameters. Here, we first provide a general overview of our method for deriving the distributions, including a description of the model diagnostics and their derivation. We follow with a discussion of the new methods used in this study and how they are applied to deriving the new distributions.

Following a standard methodology (Forest et al., 2006, 2008; Libardoni and Forest, 2011; Olson et al., 2012), we derive probability distributions for the model parameters. In this method, EMICs are used to run simulations of historical climate change. By comparing model output to observations, the likelihood that a run with a given set of parameters represents the climate system is determined by how well it simulates the past climate. In this study, we use MESM, which includes three adjustable parameters that set properties that strongly influence the behavior of the climate system. These model parameters are the cloud feedback parameter, which sets the effective climate sensitivity (ECS), the effective ocean diffusivity of heat anomalies by all mixing processes (K_v), and the net anthropogenic aerosol forcing scaling factor (F_{aer}). We identify each run by a unique combination of the model parameters, θ, where $\theta =$ (ECS, K_v, F_{aer}). In this study, we take the 1800-member ensemble described in Chapter 2, spanning a wide range of θ's, as our model output.

We evaluate model performance by comparing each model run to two temperature diagnostics. The first diagnostic is the time series of decadal mean surface temperature anomalies in four equal-area zonal bands spanning from 0-30 and 30-90 degrees latitude in each hemisphere. Temperature anomalies are calculated with respect to a chosen base period. The second diagnostic is the linear trend in global mean ocean heat content in the 0-2000 m layer. For each diagnostic, we now describe the data used for observations and the methods to derive the diagnostics from the observations.

For surface observations, we use datasets from four different research centers. The datasets we use include the median of the 100-member HadCRUT4 ensemble from the Hadley Centre Climatic Research Unit (Morice et al., 2012), the Merged Land-Ocean Temperature (MLOST) dataset from NOAA (Vose et al., 2012), the Berkeley Earth Surface Temperature (BEST) dataset (Rohde et al., 2013), and the GISTEMP dataset with 250 km smoothing (GISTEMP250) from the NASA Goddard Institute for Space Studies (Hansen et al., 2010). All datasets are given as monthly temperature anomalies on a 5x5 degree latitude-longitude grid. The datasets use similar station data over land but differ on which sea surface temperature (SST) dataset is used for the ocean. In particular, the HadCRUT4 and BEST datasets use the Hadley Centre SST (HadSST) dataset (Kennedy et al., 2011a,b) and the MLOST and GISTEMP250 datasets use the Extended Reconstruction Sea Surface Temperature (ERSST) dataset (Huang et al., 2015). Furthermore, the base period used to calculate temperature anomalies differs among the datasets. A 1951-1980 base period is used for BEST and GISTEMP250, a 1961-1990 base period is used for HadCRUT4, and a 1971-2000 base period is used for MLOST. Lastly, the research centers differ in how they fill in sparse data regions.

We derive the surface temperature diagnostic by temporally and spatially averaging the gridded data. In the following calculation, we assume uncertainty in the observations is zero, relying on using multiple datasets to account for uncertainty in the observed record. Due to data scarcity and missing values in some regions, we set threshold criteria for each spatial and temporal average in the derivation. First, the annual mean for each 5x5 degree grid box is calculated, provided that at least eight months of the year have non-missing data. From these annual averages, decadal mean time series are calculated for both the period being used in the diagnostic and the chosen climatology base period. For these calculations, we require at least eight years of defined data for a decadal mean to be defined. We also extract from the annual mean time series a data mask indicating where observations are present or missing. We use this mask on the model output to match the coverage of the observations.

Once the data mask and decadal mean time series are calculated, each time series is zonally averaged on the five-degree grid. The zonal mean is marked as undefined if there is less than 20 percent longitudinal coverage in a given latitude band. We calculate temperature anomalies for each zonal band by subtracting the mean of the climatological time series for the given band from each decade of the comparison period time series. The resulting time series of decadal mean, five-degree resolution temperature anomalies are then averaged into the four equal-area zones. When aggregating to larger areas, the mean is calculated as the area-weighted average of the zonal bands contained within the larger zone.

For ocean heat content observations, we use the estimated global mean ocean heat content in the 0-2000 m layer from Levitus et al. (2012). This dataset replaces the Levitus et al. (2005) 0-3000 m global mean dataset because the latter ends in 1998 and we aim to extend the diagnostic into the 21st century. Data are presented as heat content anomalies in five-year running means, starting with the 1955-1959 pentad and ending in the 2011-2015 pentad. Also included is a time series of the standard error of the pentadal mean estimate for the global mean heat content. The procedure for deriving the standard error estimates is described in the supplemental material of Levitus et al. (2012) and is based on observational error estimates in 1-degree gridded data.

For a given diagnostic period, we calculate the linear trend in the global mean ocean heat content as the slope of the best-fit linear regression line. In the calculation of the regression line, all deviations from the mean are assigned a weight inversely proportional to the square of the standard error at that point in the time series. For example, the standard deviation of y from the mean,

$$\sigma_y = \sqrt{\frac{\sum_i (y_i - \bar{y})^2}{n-1}} \tag{3.1}$$

is modified by multiplying each term in the summation by its weight, giving the weighted standard deviation of y from the mean of

$$\sigma_{y,w} = \sqrt{\frac{\sum_i w_i * (y_i - \bar{y})^2}{n-1}}, \tag{3.2}$$

where w_i is the weight assigned to each point y_i based off of the observational error estimate. All summation terms in the regression are replaced by the corresponding weighted version. By doing such, the regression is weighed more towards portions of the time series for which the standard error of the observations is small. Because observational errors decrease in latter years, more recent observations have a stronger influence on the trend estimate.

Following Forest et al. (2006), we calculate a goodness-of-fit statistic, r^2, for each model run to determine how well it matches the diagnostics. The statistic is the weighted sum of squared residuals between the observed diagnostic and the corresponding diagnostic calculated from the model output. The weights applied to the model-to-observation residuals are estimated from the unforced climate variability in a fully coupled, three-dimensional climate model and represent how we expect the diagnostics to behave in the absence of external forcings. Prior to this study, separate models were used for the surface and ocean diagnostics, potentially yielding inconsistent variability estimates. We eliminate that problem by using the Community Climate System Model, version 4 (CCSM4, Gent et al., 2011) to estimate the natural variability for both the surface and ocean diagnostics.

We convert goodness-of-fit statistics to probability distribution functions (PDFs) for the model parameters using the likelihood function described in Libardoni and Forest (2011) and modified by Lewis (2013). We calculate goodness-of-fit statistics for each diagnostic, with the surface diagnostic based off of Δr^2, the difference between r^2 and the minimum r^2 in the domain, and the ocean diagnostic based off of r^2 itself. Through an application of Bayes' Theorem (Bayes, 1763), we combine the likelihoods to estimate a joint PDF for the model parameters. We apply the expert prior derived in Webster and Sokolov (2000) to ECS and uniform priors to K_v and F_{aer}. Probability distributions for individual parameters are calculating by integrating the joint PDF over the other two parameter dimensions.

Prior to calculating the likelihood function, we interpolate the goodness-of-fit statistics onto a finer grid in the parameter space. This interpolation fills in the gaps between θ's where the model was run and increases the density of points within the domain. Forest et al. (2006) presented an interpolation method that was implemented in Libardoni and Forest (2011). The interpolation is first carried out on ECS-$\sqrt{K_v}$ planes via a spline interpolation on all F_{aer} levels to a finer mesh of points. A second set of spline interpolations at every ECS-$\sqrt{K_v}$ point on the fine mesh then fills in the fine grid in the F_{aer} dimension.

In this study, we implement an alternate interpolation method based off of radial basis functions (RBFs, Powell, 1977). The RBF method approximates the value of a function based off a set of node points where the functional value is known and is a variation of kriging that does not allow the data to inform the internal parameters of the algorithm. The function value at any point in the domain is calculated as the weighted sum of the value at all nearby node points. The weight assigned to each node is related to the radial distance between the location that is being interpolated to and the node. We view this method as an improvement because it is a three-dimensional method and does not require multiple steps. We will also show in Section 3.3.1 that this leads to a smoother interpolation surface.

For our implementation, we use the 1800 r^2 values at the points θ where the model has been run as nodes. We interpolate the r^2 values onto the same fine grid from the spline method with resolution 0.1 °C in ECS, 0.1 cm s$^{-1/2}$ in $\sqrt{K_v}$ and 0.05 Wm^{-2} in F_{aer}. For weights, we choose Gaussian basis functions, with the weight assigned to each node given by:

$$\phi(d) = e^{-(\epsilon d)^2}, \tag{3.3}$$

where ϕ is the weight, r is the radial distance between the two points, and ϵ is a scaling parameter that determines how quickly the weight decreases with distance. As the value of ϵ increases, the radius of influence of a given node point decreases. We calculate the r^2 value at any point in the domain as:

$$r^2(\theta) = \sum_i^N \phi_i r_i^2, \tag{3.4}$$

where the sum is over all $N = 1800$ node values. Typically, RBFs are calculated in physical space, where the distance between points, d, is well defined. However, in this application, we need to apply the concept of distance in model parameter space. Because the spacing between nodes in each dimension of the parameter space is different, we normalize all distances by the range in a given parameter dimension. We recognize that this choice of normalization constant is arbitrary and in the future should be determined by a physical metric. After accounting for the normalization, we treat each parameter dimension as isometric, so that the distance between two points is represented by:

$$|d|\left(\theta_i, \theta_n\right) = \sqrt{\left(\frac{ECS_i - ECS_n}{NORM_{ECS}}\right)^2 + \left(\frac{\sqrt{KV_i} - \sqrt{KV_n}}{NORM_{KV}}\right)^2 + \left(\frac{FA_i - FA_n}{NORM_{FA}}\right)^2}, \tag{3.5}$$

where subscript i refers to the interpolated point, subscript n refers to the node points, and the normalization constants are 9.5 °C in ECS, 8 cm s$^{-1/2}$ in $\sqrt{K_v}$, and 2.25 Wm^{-2} in F_{aer}.

In summary, we have made a number of changes and updates to the methodology. (i) To account for a change in observational dataset, we have modified the ocean diagnostic to be estimated from the 0-2000 m layer, as opposed to the 0-3000 m layer. (ii) We now estimate the natural variability from a common model, as opposed to using different models for the surface and ocean diagnostics. (iii) We implement a new interpolation scheme where radial basis functions are used to interpolate goodness-of-fit statistics from the coarse grid of model runs to the fine grid used to derive the joint probability distribution functions.

Using the updated methodology, we show how temporal and spatial information impact the PDFs of the model parameters. We address the temporal component by adding more recent temperature data to the model diagnostics in one of two ways. First, we extend the diagnostics by fixing the starting date while shifting the end date forward in time. To maximize the amount of data that we use in the surface diagnostic while also assuring good observational data coverage, we take decadal mean temperature anomalies with respect to the 1906-1995 base period starting in 1941. We then shift the end date from 1990 to 2000 to 2010 to change the diagnostics from five to six to seven decades, respectively. For the ocean diagnostic, we choose 1955 as the starting date of the first pentad to correspond to the beginning of the observational dataset. Similar to the surface diagnostic, we increase the length of the ocean diagnostic by changing the end date of the last pentad from 1990 to 2000 to 2010.

In a second test, we fix the length of the diagnostics while shifting the end date forward in time. This maintains a five decade diagnostic for the surface diagnostic by shifting the 50-year window from 1941-1990 to 1951-2000 to 1961-2010 and a 35 year ocean diagnostic by shifting the period we use to estimate the linear trend from 1955-1990 to 1965-2000 to 1975-2010. By deriving PDFs with each pair of diagnostics corresponding to a given end date, we determine the impact of recent temperature trends on the parameter distributions in both the extension and sliding window cases.

In a third test, we derive PDFs with different structures for the surface diagnostic. In these new diagnostics, we maintain the decadal mean temporal structure, but reduce the dimensionality of the spatial structure by replacing the four zonal bands with global mean or hemispheric mean temperatures. In the former case, we have a one-dimensional spatial structure and the latter a two-dimensional structure.

3.3 Results

We present our findings as follows. In Section 3.3.1 we show (i) the difference in the ocean diagnostic due to changing to the 0-2000 m data, (ii) provide justification for using the RBF interpolation method, and (iii) present the impact of the methodological changes described in Section 3.2 on the parameter distributions. In Section 3.3.2, we (i) analyze how the model diagnostics change due to the inclusion of more recent data and (ii) assess how those changes impact the distributions. In Section 3.3.3, we show how including spatial patterns of surface temperature change impact the distributions.

3.3.1 Methodological Changes

We first identify the difference in the ocean diagnostic derived from the 0-3000 m and 0-2000 m layers for the common period of 1955-1996 (Figure 3.1). This period is chosen to coincide with the ocean diagnostic in Libardoni and Forest (2013) and allows for a direct comparison of distributions presented later in this section. We observe a stronger warming trend of 3.6 ± 0.50 ZJ/yr in the 0-2000 m layer compared to the estimate of 2.7 ± 0.39 ZJ/yr in the 0-3000 m layer, suggesting that the rate of heat penetration into the deep ocean decreases with depth.

Second, we present the switch to the radial basis function interpolation algorithm. We note from Equation 3.3 that the weight assigned to each node point depends on the radial distance between the points, d, and a scaling factor, ϵ. Because the distance between any two points in the parameter space is always the same, the choice of ϵ plays a critical role in determining the behavior of the algorithm. We demonstrate this impact by running the RBF algorithm using six different choices of ϵ on a random surface temperature diagnostic (see Appendix A for choices of ϵ). We show the resulting r^2 patterns and compare them against the surface derived using the Forest et al. (2006) spline interpolation method and the original pattern (Figure 3.2). We observe that the old method is very successful at matching the r^2 values at points where they were run (Figure 3.2b). However, the surfaces are not always smooth and in some instances the location of the minimum value of r^2 shifts to a new, nearby location in the interpolated space.

We aim to improve upon the shortcomings of the old interpolation method by identifying ϵ so that not only is the spatial pattern of r^2 maintained, but the resulting response surface is smooth. We observe smoother interpolated surfaces for lower values of ϵ because of the relationship between ϵ and the radius of influence of each node point (Figure 3.2c-h). Because we do not require the interpolated values to pass exactly through the node points, the smoothness comes at the expense of increasing the interpolation

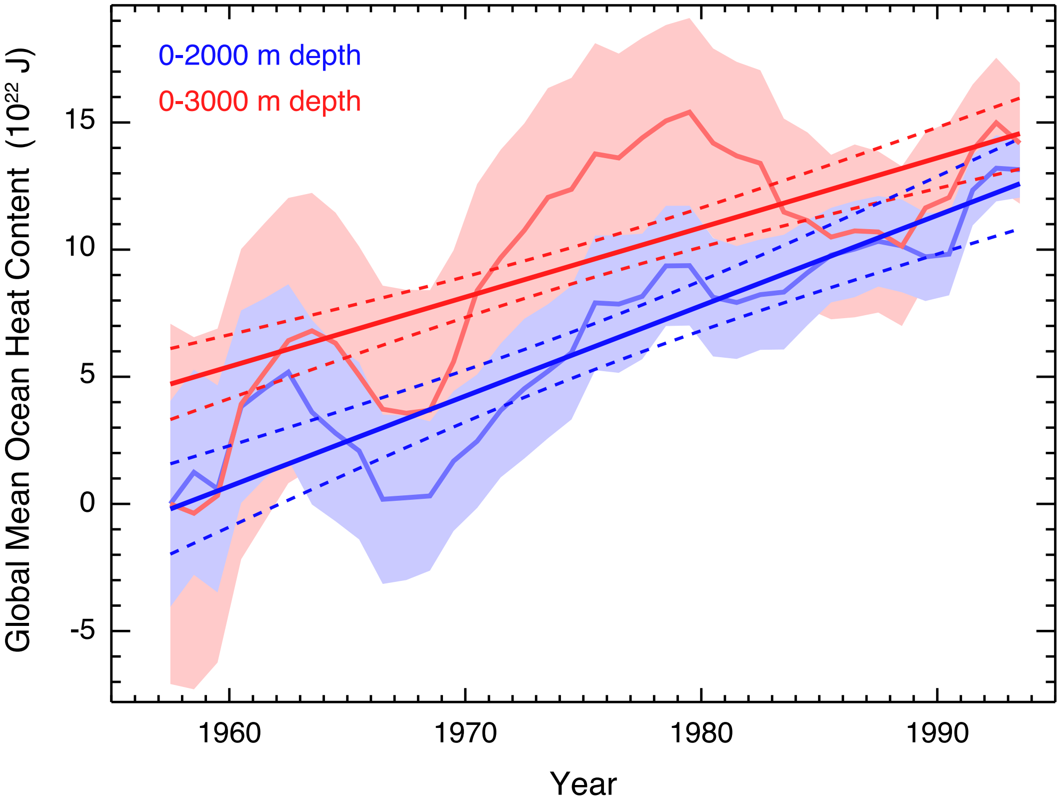

Figure 3.1: Global mean ocean heat content for the 0-3000 m layer (Levitus et al., 2005) and 0-2000 m layer (Levitus et al., 2012). Shading indicates twice the standard error on either side of the estimate. Dashed lines indicate the 95-percent confidence interval for the point estimate for a given year based on the best fit line and its uncertainty.

error at the node points. Unlike the old interpolation method, the errors at node points do not lead to a change in the rank order of r^2 values at the node points, however. The location of the minimum remains the same, as well as all subsequent comparisons.

We also observe a reduction in the range of r^2 values within the domain. The reduction occurs because regions where r^2 is originally low are now influenced by areas further away in the parameter space where r^2 is high, and vice versa. This is true of the algorithm in general, with the errors at each node point and the reduction of the range diminishing as ϵ increases and the radii of influence of each node point decreases. However, as ϵ increases and the radius of influence for a given node decreases, the response surface becomes less smooth. Thus, there is a trade off, in that decreasing the interpolation error at node points leads to a decrease in the smoothness of the surface. Small ϵ's provide the desired smoothness, while large ϵ's provide the the truest fit to the actual values at the node

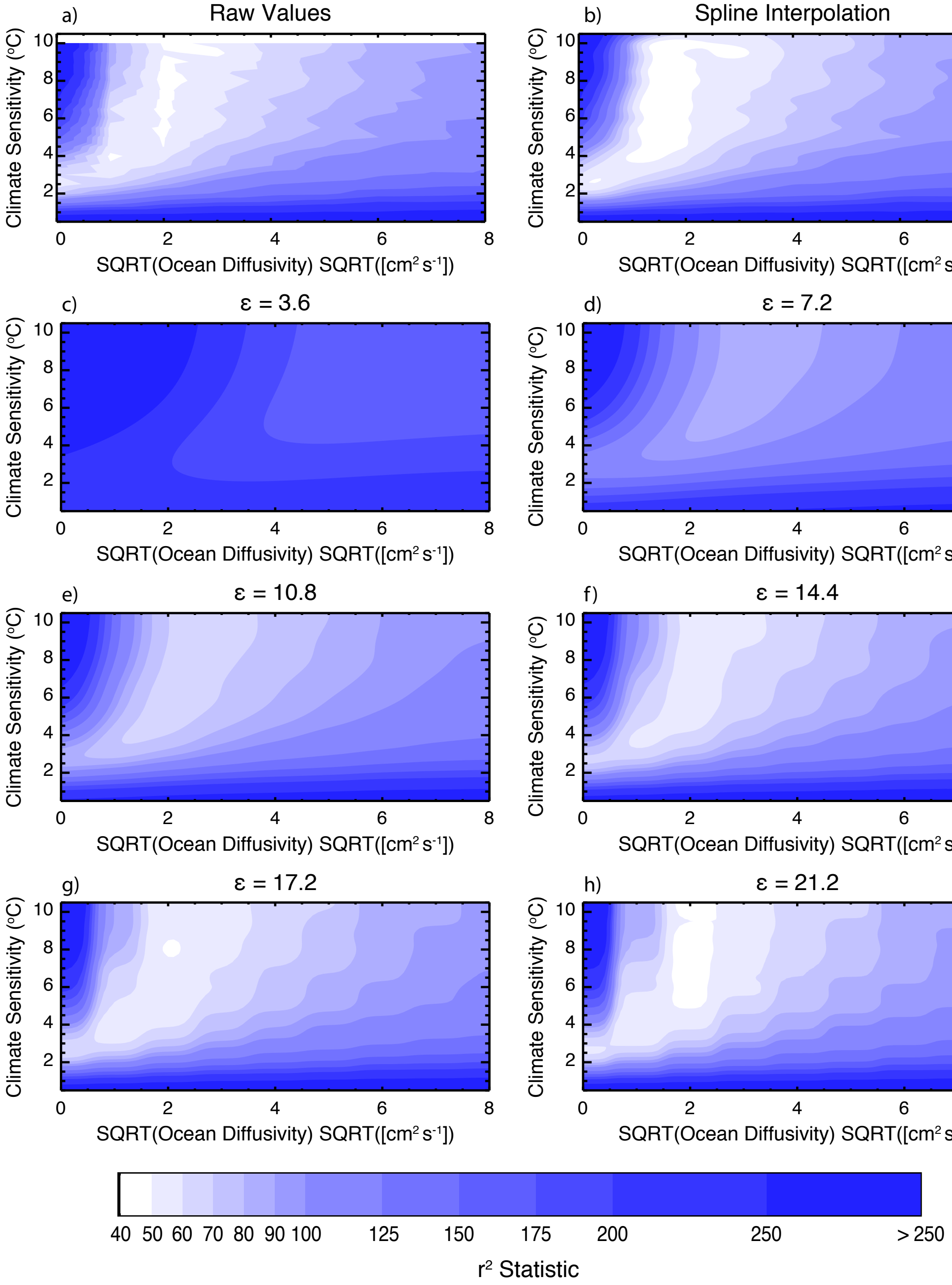

Figure 3.2: Example of the differences between the algorithms to interpolate goodness-of-fit statistics from the coarse grid of model runs to the finer grid used for the derivation of parameter distributions. Calculated r^2 values (a) are shown along with the interpolated values using the algorithm from Libardoni and Forest (2011) (b) and the radial basis function interpolation with six different values of ϵ (c-h).

points. This indicates that intermediate values of ϵ (e.g., 10.8 or 14.4) are appropriate.

Thus far, we have only looked at the impact of ϵ on the fit of the interpolated r^2 values to the raw values. In practice, our likelihood function for the surface diagnostic is a function of the residual between r^2 at a given point and the minimum value of r^2 in the domain, a statistic we refer to as Δr^2. Plotting the Δr^2 field as a function of ϵ confirms our assessment that intermediate values of ϵ lead to the best fit to the raw values (Figure 3.3). Both $\epsilon = 10.8$ and $\epsilon = 14.4$ fit the raw Δr^2 values quite well as the inflation of low r^2 values is normalized out by the subtraction of the minimum value (which is also interpolated to a greater value). However, for $\epsilon = 14.4$, the region of best fit (Δr^2 less than 10) is larger than the raw values and there are regions where the interpolated surface is not as smooth as when $\epsilon = 10.8$. In some situations, this lack of smoothness leads to PDFs that are also not smooth and display bumps at values for the parameter settings of the node points (not shown). For these reasons, we choose $\epsilon = 10.8$ for our analysis.

To further test our choice of ϵ, we perform an out-of-sample test on 300 runs of MESM that were not included in the 1800 member ensemble used in this study. The parameter settings for the out-of-sample runs were the result of two separate 150-member Latin Hypercube Samples (McKay et al., 1979) and did not correspond to the settings of any of the node points. For each run, we calculate Δr^2 for the surface diagnostic matching the one used in Figures 3.2 and 3.3 and compare those against the values calculated using the RBF interpolation method with $\epsilon = 10.8$ and the 1800 runs as nodes (Figure 3.4).

With a few exceptions, we see good agreement between Δr^2 calculated from model output and Δr^2 estimated from the RBF algorithm. The biggest discrepancies are typically found for Δr^2 values greater than 50, where the likelihood function for the diagnostic approaches 0. We also note that the differences are small in regions of the parameter space where the likelihood function approaches its maximum, namely for small Δr^2. Lastly, we find an almost equal number of runs where the difference between the value calculated from the model output and the value estimated from the RBF method is greater than zero and where the difference is less than zero, indicating no substantial bias in the RBF algorithm. Because we see good agreement of the RBF interpolated surface with the out-of-sample test runs and observe a smooth response surface with a good fit to the data (Figures 3.2 and 3.3), we argue that choosing $\epsilon = 10.8$ is appropriate.

To test the impact of the methodological changes just described, we start from a previously published probability distribution and apply the changes one at a time. For a

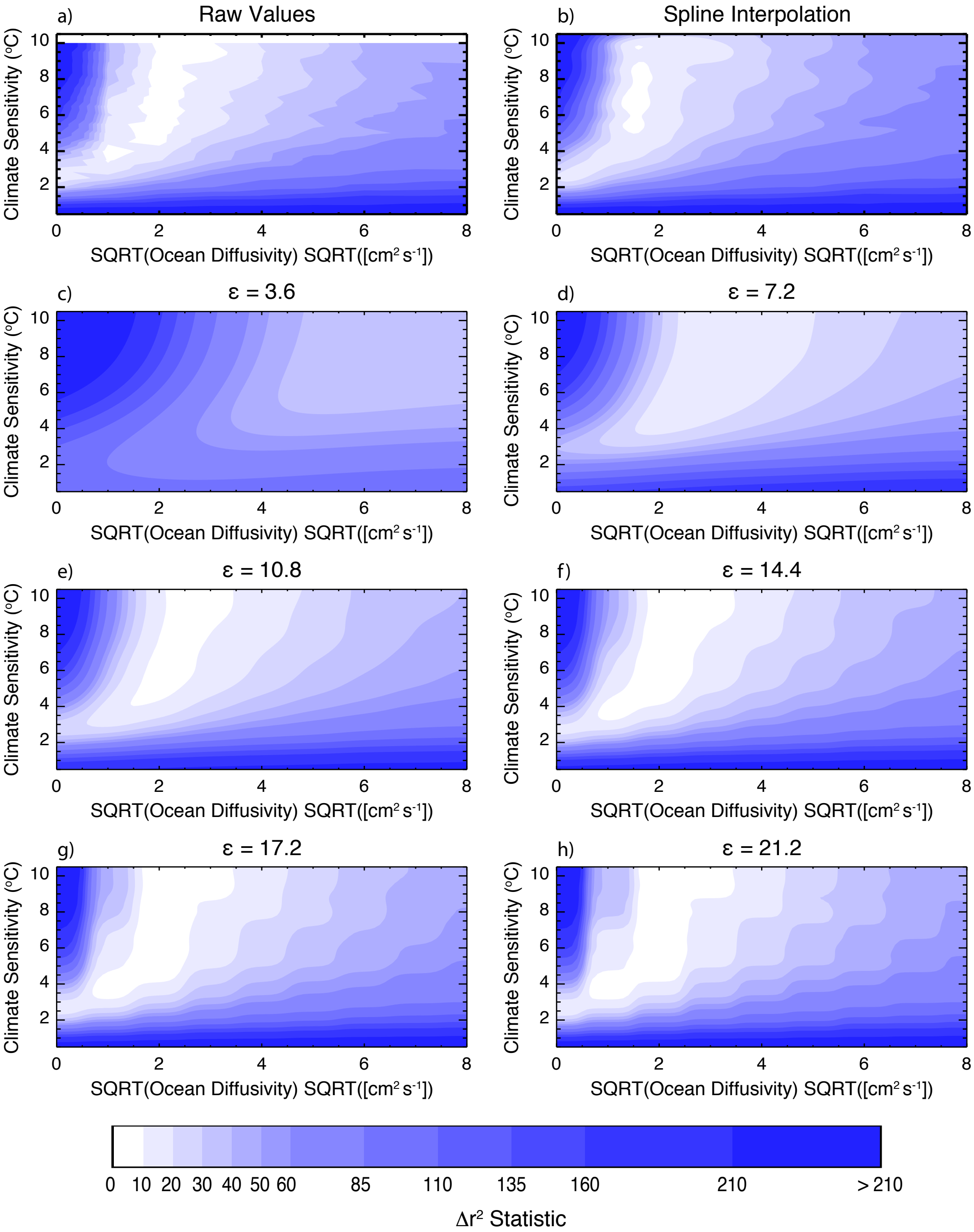

Figure 3.3: As in Figure 3.2, except for Δr^2, the difference between r^2 and the minimum r^2 value in the domain.

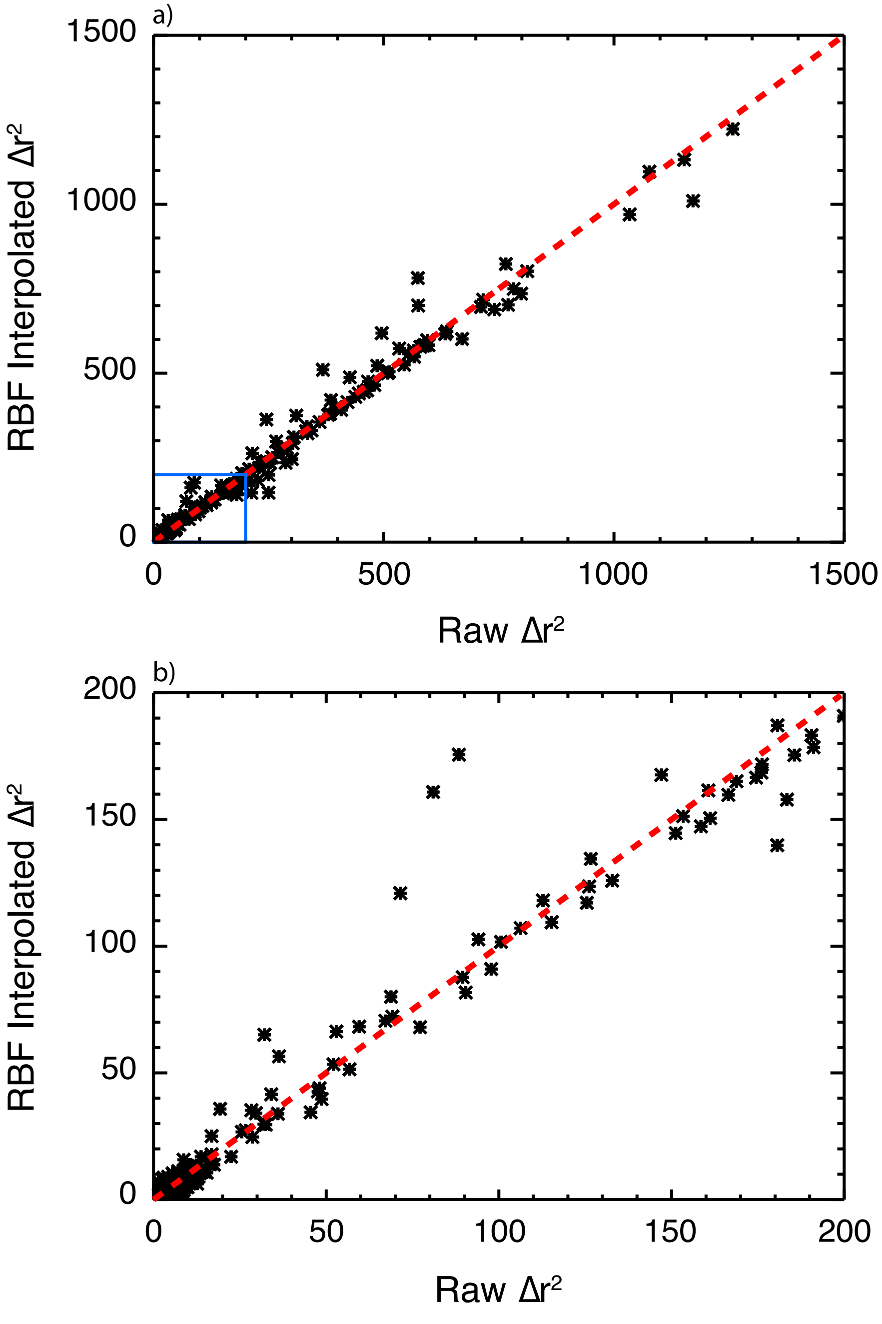

Figure 3.4: Comparison of Δr^2 values calculated from out-of-sample model runs and those calculated using the RBF interpolation method. (a) All 300 runs. (b) All runs with Δr^2 less than 200. The one-to-one line is plotted for reference (red, dashed line).

reference point, we start with the PDF from Chapter 2 derived using the HadCRUT3 surface dataset (Brohan et al., 2006) and the likelihood function presented earlier in Section 3.2. The changes we implement are: (i) change the ocean diagnostic from the 0-3000 m layer to the 0-2000 m layer, (ii) replace the interpolation method of Forest et al. (2006) with the RBF interpolation method, and (iii) change from using natural variability estimates from different control run models for the surface and ocean diagnostics to a common model for both estimates. To better illuminate the changes, we derive an additional PDF changing both the control run model and the interpolation method simultaneously. We summarize the resulting distributions in Figure 3.5.

When changing the ocean diagnostic from the 0-3000 m layer to the 0-2000 m layer, we observe the largest change as a shift towards higher K_v. As measured by the 90-percent confidence interval for the marginal distribution of $\sqrt{K_v}$, our estimate increases from 0.29-1.90 to 0.81-3.22 cm $s^{-1/2}$. We also note that the wider interval indicates a weaker constraint on the estimate of K_v. In MESM, K_v controls how fast heat is mixed into the the deep ocean. Thus, we trace the shift towards higher K_v to the stronger heating rate in the ocean diagnostic due to estimating the trend from the 0-2000 m data (Figure 3.1). We observe a small shift towards higher ECS and almost no change in estimates of F_{aer}.

For the second change, we explore the implementation of the RBF interpolation algorithm. In Figure 3.5, we observe that the parameter distributions are indeed smoother when the RBF method is used. This is particularly evident in the climate sensitivity distributions. We also note changes to the constraints on model parameters. In general, we see a flattening of the of center of the distributions, as measured by the interquartile range (IQR). In particular, the IQR for $\sqrt{K_v}$ increases from 0.59 to 0.71 cm $s^{-1/2}$ (ranges of 0.71-1.3 to 0.86-1.57 cm $s^{-1/2}$) and for F_{aer} from 0.07 to 0.11 Wm^{-2} (-0.25 - -0.18 to -0.32 - -0.21 Wm^{-2}) when comparing the reference PDF using the old interpolation method to the PDF estimated using the RBF method. This increase is consistent with our previous discussion that the RBF method tends to adjust low r^2 values upwards and high r^2 values downwards. In this situation, the maximum likelihood region of the joint PDF, where r^2 is a minimum, impacts all points within its radius of influence.

In general, we observe tighter constraints on all of the distributions when a common control run model is used for the surface and ocean diagnostics. For all three parameters, the width of the 90-percent confidence interval decreases. However, in the climate sensitivity distribution, we see multiple minima and maxima in the distribution. All of the local extrema occur at values of ECS where the model has been run. We attribute

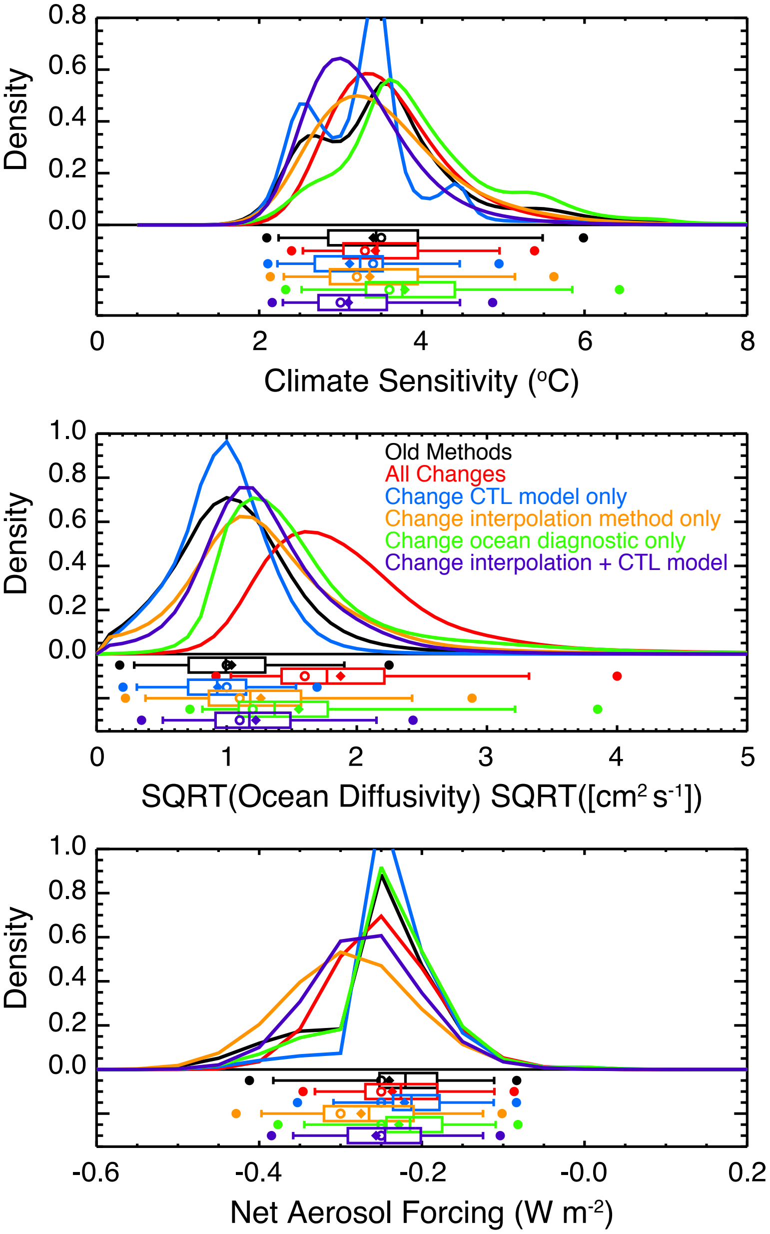

Figure 3.5: Marginal probability distribution functions for ECS (a), $\sqrt{K_v}$ (b), and F_{aer} (c) derived with changes in methodology. A comparison between the HadCRUT3 distribution derived in Libardoni and Forest (2013) (black) with those derived using all changes outlined in the text (red) and individual changes to the control run used to estimate natural variability (blue), the ocean diagnostic (green), and the interpolation method (orange). Also shown is the case where the natural variability estimate and interpolation method are changed together (purple). Whisker plots indicate boundaries for the 2.5-97.5 (dots), 5-95 (vertical bar ends), 25-75 (box ends), and 50 (vertical bar in box) percentiles. Distribution means are represented by diamonds and modes are represented by open circles.

these oscillations to the spline interpolation method attempting to pass through r^2 exactly at all of the points and observe them in plots similar to Figure 3.2 for different aerosol levels (not shown). Because the assumed impact of the old interpolation method leads to the spurious ECS marginal distribution, we also show the case where both the control run and interpolation method are changed together (purple curve in Figure 3.5). This test also separates the impacts of changing datasets and diagnostics (ocean dataset) from the technical details of the derivation (interpolation method and variability estimate).

We summarize the net impact of the changes by implementing all three simultaneously (red curve in Figure 3.5). When comparing the ECS and F_{aer} distributions, we observe very little change in the estimates of central tendency and stronger constraints on the parameters. Here, we measure central tendency by the median of the distribution and the constraint by the width of the 90-percent confidence interval. Before implementing the changes, we estimate the median ECS to be 3.44 °C with a 90-percent confidence interval of 2.24-5.48 °C. After the changes, we estimate a median of 3.45 °C and a 90-percent confidence interval of 2.54-4.96 °C. Similarly, for F_{aer} we estimate a median of -0.22 Wm^{-2} and 90-percent confidence interval of -0.38 - -0.11 Wm^{-2} before and a median of -0.23 Wm^{-2} and 90-percent confidence interval of -0.38 - -0.11 Wm^{-2} after the changes. This pattern does not hold for the K_v distribution. In $\sqrt{K_v}$, we estimate the median to increase from 1.00 to 1.77 cm $\mathrm{s}^{-1/2}$ and the 90-percent confidence interval to change from 0.29-1.90 to 1.03-3.32 cm $\mathrm{s}^{-1/2}$ when implementing the new methodology. We previously showed that the change in ocean dataset led to higher K_v estimates without changing the central estimates of the other two parameters. Combining this with the findings from the ECS and F_{aer} distributions leads us to conclude that the central estimates of the distributions change with the diagnostics, and that the technical changes, namely the unforced variability estimate and the interpolation method, impact the uncertainty estimates.

3.3.2 Temporal Changes to Model Diagnostics

Before presenting new PDFs using the methods analyzed in the previous section, we present the model diagnostics used to derive them. We show the time series of decadal mean temperature anomalies with respect to the 1906-1995 climatology in the four equal-area zonal bands of the surface temperature diagnostic (Figure 3.6). We plot the time series from 1941-2010 with the decadal mean plotted at the midpoint of the decade

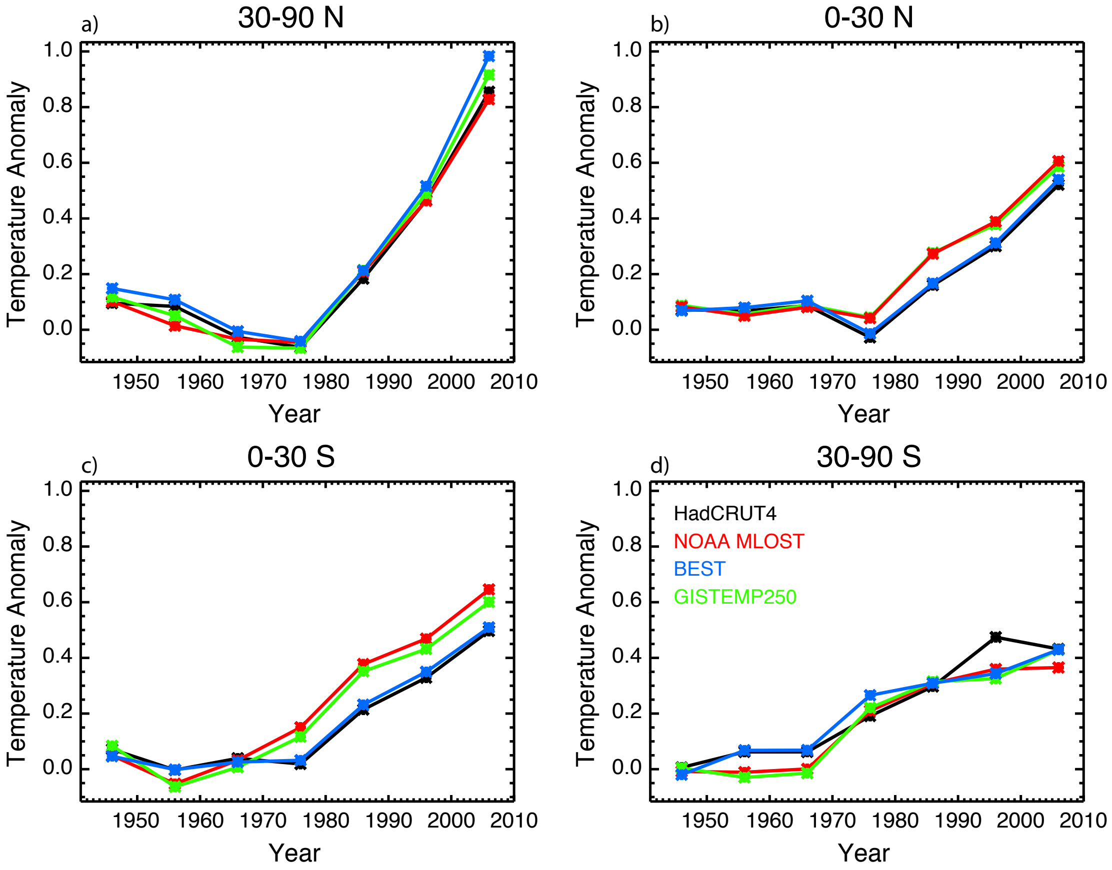

Figure 3.6: Decadal mean temperature anomaly time series derived from the HadCRUT4, NOAA MLOST, BEST, and GISTEMP 250 datasets. Time series are for the four equal-area zonal bands spanning 30-90 °N (a), 0-30 °N (b), 0-30 °S (c), and 30-90 °S (d). Temperatures are plotted as anomalies with respect to the 1906-1995 base period at the midpoint of each decade.

it represents. In tests where we extend the model diagnostics by holding the start date fixed and add additional data, we add an additional data point to the end of each time series. In tests where we hold the length of the diagnostics fixed while adding recent data, we change which five data points are used.

From the time series, we see that while general similarities exist, the model diagnostic depends on which surface observations are used. Across all datasets, we observe the largest signal in the 30-90 °N zonal band, consistent with the polar amplification of warming. We also note that the highest agreement across the datasets is observed in this band. We find that there is a separation between the time series in the 0-30 °N and 0-30 °S zonal bands based on which SST dataset a group used for the temperatures over the ocean. When considering this split, we see similar patterns in the tropical bands between datasets using HadSST (HadCRUT4 and BEST) and datasets using ERSST (MLOST and GISTEMP250). Although not shown, we observe similar patterns in the hemispheric and global mean time series, with a stronger warming signal in the northern hemisphere and the time series showing sensitivity to the dataset.

We illustrate how additional data impact the estimate of the linear increase in ocean heat content (Figures 3.7 and 3.8). In both figures, we plot the time series from Levitus et al. (2012) with the pentadal mean plotted at the midpoint of the five-year period defining the pentad. In Figure 3.7, we fix the starting date in 1955 and shift the end date further ahead. In Figure 3.8, we fix the length of time over which the linear trend is calculated and shift the end of the period forward.

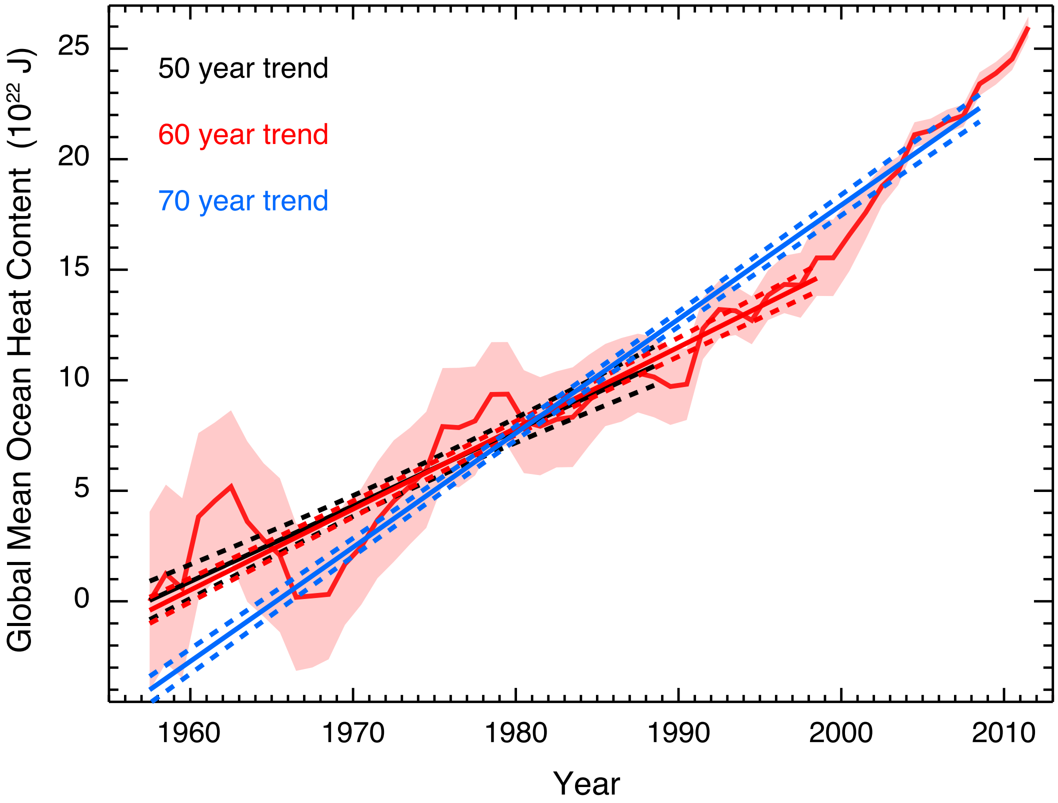

Figure 3.7: Global mean ocean heat content for the 0-2000 m layer. Shading indicates twice the standard error on either side of the estimate. Also shown are the best fit linear trend lines for the trend beginning in 1955 and ending in 1990 (black), 2000 (red), and 2010 (blue). Dashed lines indicate the 95-percent confidence interval for the point estimate for a given year based on the best fit line and its uncertainty.

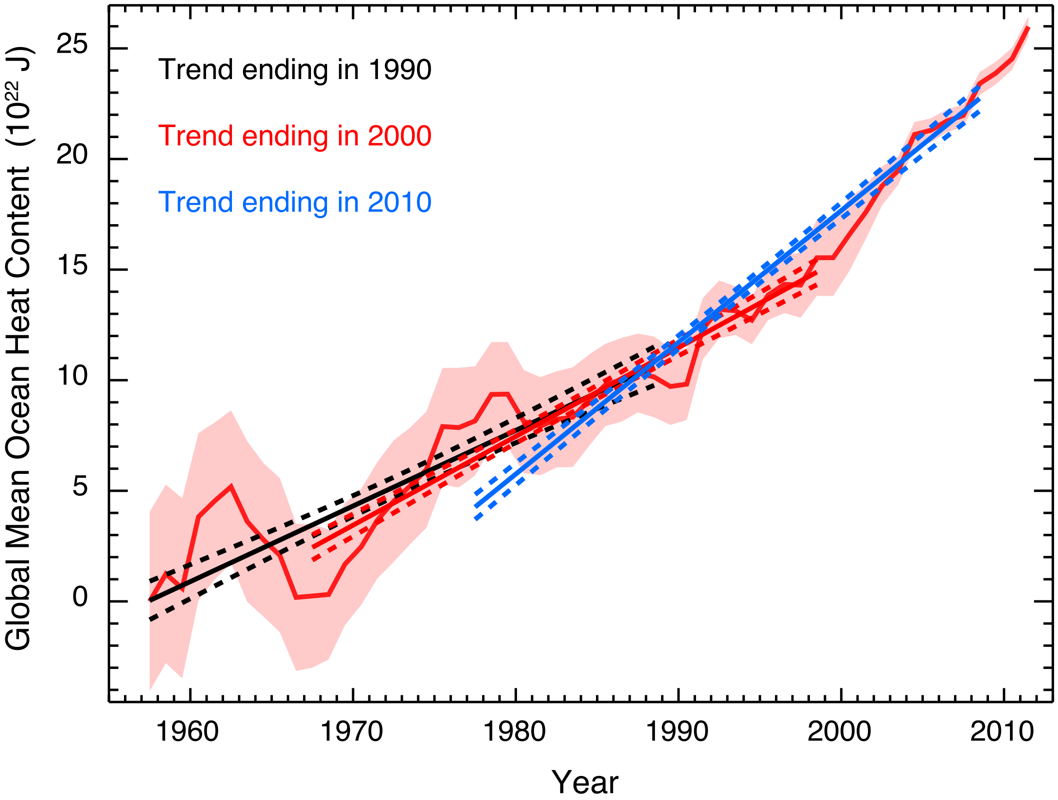

Figure 3.8: As in Figure 3.8, except the diagnostic length is held fixed. Linear trend estimates are for the 1955-1990 (black), 1965-2000 (red), and 1975-2010 periods.

The recent acceleration of heat stored in the deep ocean is well documented (Levitus et al., 2012; Gleckler et al., 2016), and as expected, we find that the trend estimate depends on both the end points of the period used for estimation and the length of the period used for estimation. As previously stated, more recent observations have a stronger influence on the trend estimate because the standard error of the observations decreases with time. We calculate higher trend estimates when holding the period length fixed while including more recent data compared to when the period is extended to include more recent data. We estimate a trend of 3.4 ± 0.28 ZJ/yr when considering the period from 1955 to 1990. For diagnostics ending in 2000, we estimate a trend of 4.0 ± 0.19 ZJ/yr if the starting date is shifted to 1965 and a trend of 3.7 ± 0.15 ZJ/yr if the starting date is held at 1955. Trends of 6.0 ± 0.18 ZJ/yr and 5.2 ± 0.12 ZJ/yr are estimated when using data up to 2010 and holding the diagnostic length fixed and extending the diagnostic length, respectively. By shifting the diagnostic rather than extending it, the accelerated warming signal is stronger because periods of slower warming earlier in the time series are replaced by periods of faster warming later in the time series.

For each surface and ocean diagnostic set, we derive joint probability distributions according the experiments laid out in Section 3.2. When calculating the surface diagnostic for a given period, we derive a PDF using each of the four datasets as observations. We combine the four PDFs into a single estimate by taking the average likelihood at each point in the joint PDF. In offline calculations, we confirmed that the marginal PDFs for each parameter using the average joint PDF were nearly identical to the marginal PDFs resulting from the merging method used to submit the distributions from Libardoni and Forest (2013) for inclusion in the Intergovernmental Panel on Climate Change Fifth Assessment Report (IPCC AR5, Collins et al., 2013). For the IPCC AR5 estimates, we drew a 1000-member Latin Hypercube Sample from each distribution and calculated marginal distributions for each parameter from the histogram of the drawn values. By including an equal number of samples from each distribution, we assign equal weight to each surface temperature dataset. Taking the average of the four PDFs is the limit of this method as the number of draws approaches infinity. We justify using the average of the four PDFs by noting that the same general conclusions are drawn from the combined PDF as would be drawn from the PDFs derived from individual datasets.

We show that incorporating more recent data in the temperature diagnostics has a significant impact on the parameter estimates (Figure 3.9). Unless otherwise noted, we again approximate the central estimate of the distributions as the median and use the 90-percent confidence intervals to estimate the uncertainty. Across all three parameters, we generally observe sharper PDFs as more recent data are added. Furthermore, the constraints are stronger when the data are used to extend the diagnostics as opposed to when the diagnostic lengths are fixed. We attribute the general tightening of the distributions with recent data to the strong climate signals that have emerged in the observations. Further, we argue that the uncertainty bounds tend to be tighter when the diagnostic lengths are increased because the model output is being compared against more detailed temperature patterns with additional data points to match. Runs that do not match the added points are rejected for being inconsistent with the observations.

For climate sensitivity, we find that extending the data beyond 1990 leads to higher climate sensitivity estimates when compared to the estimate shown in Figure 3.5 that incorporates all of the methodological changes. However, we find that the inclusion of more recent data does not always lead to an increase in the estimate of ECS. Our estimate of ECS for diagnostics ending in 2000 is greater than the estimate for the diagnostics ending in 2010, regardless of whether the diagnostic length is extended or fixed. For the case where the diagnostics are extended, we estimate a median climate sensitivity of 4.04 °C with data ending in 2000 and 3.73 °C with data ending in 2010. When the diagnostic length is fixed, we estimate median climate sensitivities of 4.08 °C and 3.72 °C for diagnostics ending in 2000 and 2010, respectively. We hypothesize that the lowering of the estimate for ECS with diagnostics ending in 2010 can be attributed to the global warming hiatus, which led to slower warming trends in surface air temperature in the 2000s as more heat was stored in the deep ocean. We also note the uncertainty in the estimate of ECS decreases as more recent data are added and the tighter uncertainty

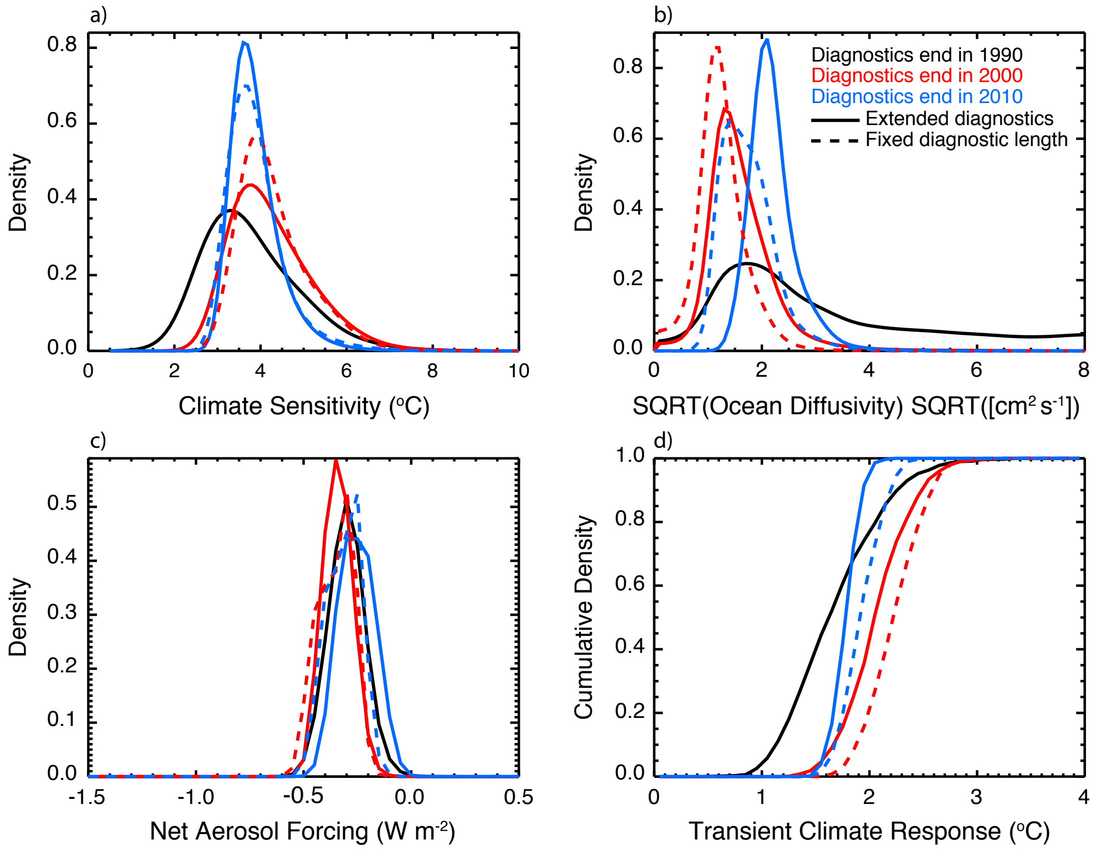

Figure 3.9: Marginal probability distribution functions for ECS (a), $\sqrt{K_v}$ (b), and F_{aer} (c) and cumulative distribution function for TCR (d) when changing the end date of model diagnostics. Distributions with diagnostics ending in 1990 (black), 2000 (red), and 2010 (blue) are shown. Solid lines indicate an extension of the diagnostic and dashed lines indicate that the length of the diagnostics remain fixed when incorporating more recent data.

bounds come predominantly from a reduction in the upper tail of the distribution. There is also a slight increase in the estimate of the lower bound, however.

Our estimates of K_v show large shifts in response to changes in the diagnostics. When the diagnostics end in 1990, we find very weak constraint on K_v, with a non-zero tail throughout the domain. As more recent data are included, we see a large reduction in the upper tail of the distributions. We also see shifts towards higher K_v with the inclusion of data from 2001-2010. When including these data, we estimate $\sqrt{K_v}$ to increase from 1.45 to 2.08 cm $s^{-1/2}$ when the diagnostic lengths increase and from 1.16 to 1.62 cm $s^{-1/2}$ when the diagnostic lengths are fixed. Because K_v sets how fast heat is mixed into the deep ocean in the model, we attribute the higher estimates to the recent acceleration of heat storage in 0-2000 m layer (see Figures 3.7 and 3.8).

We also see shifts in the F_{aer} distribution in response to the changes in model diagnostics. We reiterate that in MESM, F_{aer} sets the amplitude of the net anthropogenic aerosol forcing and represents the sum of all unmodeled forcings. We observe shifts towards stronger cooling (more negative values of F_{aer}) when diagnostics end in 2000 and shifts back towards weaker values (less negative) when the diagnostics end in 2010. When the diagnostics are extended, F_{aer} estimates shift from -0.28 Wm^{-2} when data up to 1990 are used to -0.32 Wm^{-2} and -0.23 Wm^{-2} when data up to 2000 and 2010 are used, respectively. Similarly, we observe shifts to -0.32 Wm^{-2} and -0.28 Wm^{-2} when the diagnostic lengths are held fixed and include data up to 2000 and 2010, respectively. Thus, the observed change from the 2000 to 2010 estimate is larger in the case where the diagnostics are extended rather than of fixed length.

Although not shown, we observe these shifts in the F_{aer} distributions for each of the PDFs derived using the different datasets individually, but note that we see smaller changes with the merged PDF. Also, from the individual PDFs, we see a grouping of the F_{aer} distributions based on the SST dataset used by the research center. We find the HadCRUT4 and MLOST distributions (HadSST) and the BEST and GISTEMP250 distributions (ERSST) to be similar.

We attribute the shift towards stronger cooling for the 1991-2000 decade to the cut-off of the high K_v tail. When K_v decreases, excess heat in the Earth system is stored in the ocean less efficiently. In response to this excess heating, surface and atmospheric temperatures would rise unless an external factor is active and opposes the heating. In MESM, negative values of F_{aer} reduce the net forcing and contribute to balancing the global energy budget. The spatial pattern of the net aerosol forcing in MESM leads to the forcing being stronger in the northern hemisphere than the southern hemisphere. With this pattern, we observe stronger temperature responses in the northern hemisphere when we adjust F_{aer} than we do in the southern hemisphere. We attribute the shift back towards weaker aerosol cooling when adding the 2001-2010 trends to the northern hemispheric polar amplification signal noted earlier in this section.

For each pair of model parameters, we calculate the two-dimensional marginal distribution by integrating over the third parameter (Figure 3.10). From these distributions, we observe correlations between the the pairs of parameters, independent of the diagnostic length and end date. We show ECS and K_v to be positively correlated, ECS

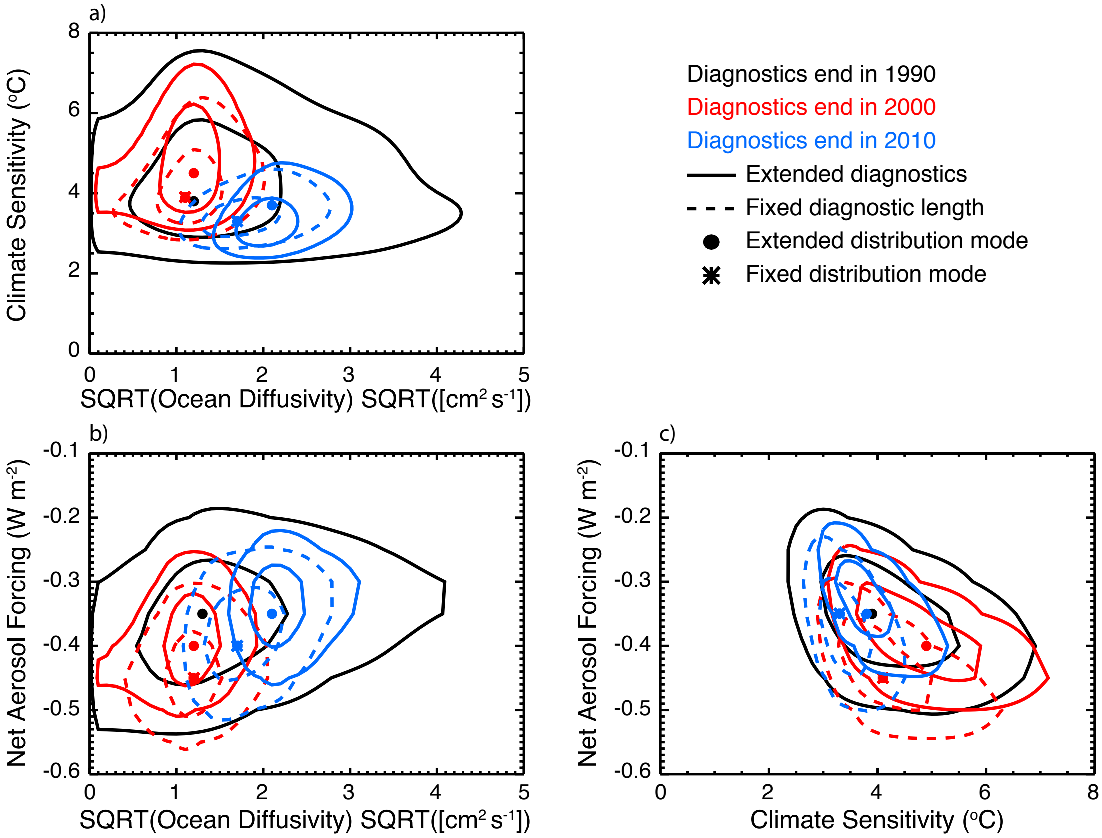

Figure 3.10: 2D joint probability distribution functions for each pair of parameters: (a) ECS-$\sqrt{K_v}$, (b) F_{aer}-$\sqrt{K_v}$, (c) F_{aer}-CS. Distributions with diagnostics ending in 1990 (black), 2000 (red), and 2010 (blue) are shown. Solid contours indicate an extension of the diagnostic and dashed contours indicate that the length of the diagnostics remain fixed when incorporating more recent data. Contours show the 90- and 50-percent confidence intervals and symbols indicate the distribution modes.

and F_{aer} to be negatively correlated, and K_v and and F_{aer} to be positively correlated. These correlations are similar to the shifts we observed when considering the parameters individually in the marginal distributions.

We also derive estimates of transient climate response from the PDFs discussed above (Figure 3.9d). From each PDF, we draw a 1000-member Latin Hypercube Sample and calculate TCR for each of the ECS-$\sqrt{K_v}$ pairs using the model response surface derived in Chapter 2. The PDFs of TCR are estimated from the histogram of TCR values with bin size = 0.1 °C. We show that the TCR estimates reflect changes in the parameter distributions. In particular, TCR and climate sensitivity are positively correlated and TCR and K_v are negatively correlated. Furthermore, the uncertainty in the TCR distribution is correlated with the uncertainty in ECS and K_v. Thus, we find that TCR estimates are greater when more recent data are added due the higher climate sensitivity estimates, but are smaller in 2010 than in 2000 due to the shift towards higher K_v. Furthermore, TCR estimates are higher when the diagnostic lengths are fixed compared to when they are extended.

3.3.3 Spatial Changes to Model Diagnostics

Until now, we have only considered how the temporal component of the diagnostics impacts the parameter estimates. As the last case study, we reduce the spatial dimension of the surface temperature diagnostic by replacing the four zonal band diagnostic with either global mean surface temperature or hemispheric mean temperatures using the 1941-2010 diagnostic period (Figure 3.11). Similar to the PDFs shown when changing the temporal structure of the diagnostic, we present the distributions calculated from the average of the four individual PDFs derived using the different surface temperature datasets.

We find little sensitivity in the central estimates of the ECS and K_v distributions to the spatial structure of the surface diagnostic using data up to 2010. For ECS, the median estimate for when global mean temperatures, hemispheric means, and four zonal bands are used are 3.81 °C, 3.75 °C, and 3.72 °C, respectively. Similarly, median estimates for $\sqrt{K_v}$ are 2.06 cm s$^{-1/2}$, 1.94 cm s$^{-1/2}$, and 2.08 cm s$^{-1/2}$ when global mean, hemispheric mean, and four zonal mean temperatures are used. However, we observe a tightening of the distributions as the spatial resolution of the surface diagnostic increases. The narrowest distributions are derived using the four zonal band diagnostic and the

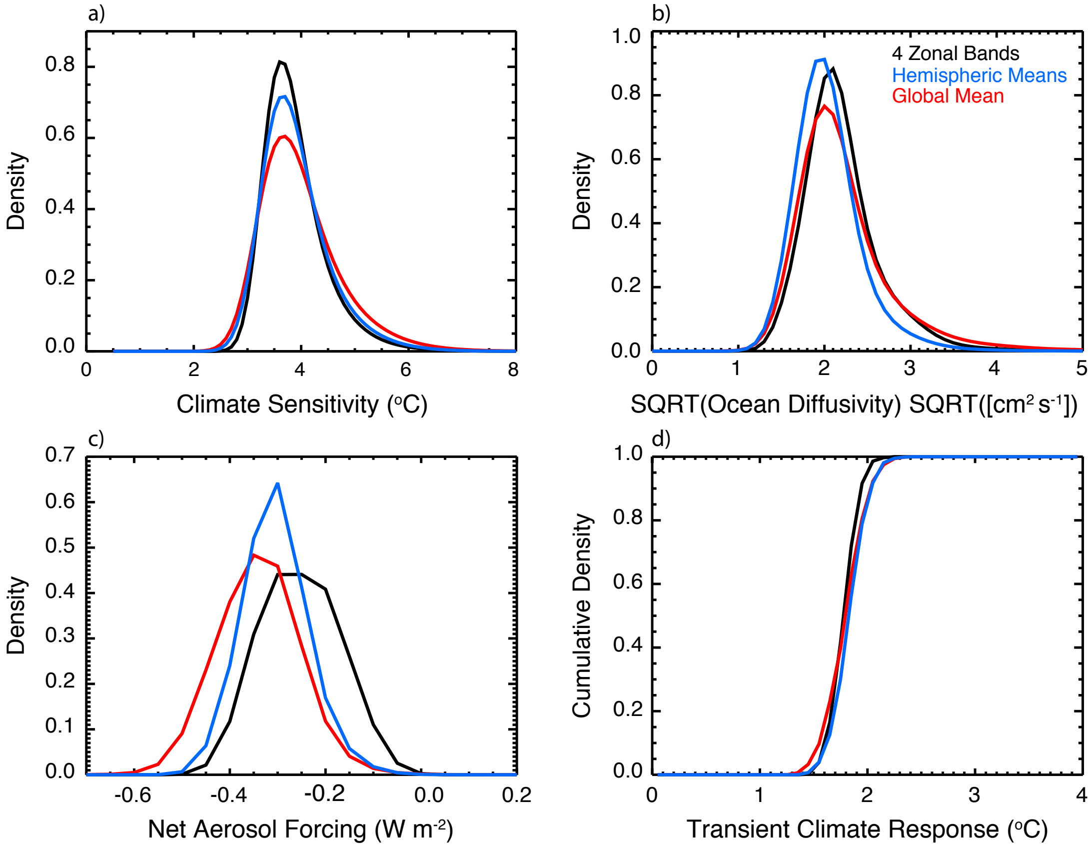

Figure 3.11: Marginal probability distribution functions for ECS (a), $\sqrt{K_v}$ (b), and F_{aer} (c) and cumulative distribution function for TCR (d) derived from different spatial diagnostics. Diagnostics end in 2010 and data is added by extending the diagnostics.

widest distributions are derived using global mean temperatures. We note that the TCR distributions follow the shifts in ECS and K_v. Thus, the central estimates do not change significantly, but the width of the distribution shrinks as spatial information is added to the surface diagnostic.

Unlike with the ECS and K_v distributions, we observe a sensitivity to the surface diagnostic structure in the F_{aer} distributions. In particular, we observe that the estimate derived using global mean temperature leads to the strongest (most negative) aerosol forcing and the estimate derived using the four zonal bands leads to the weakest aerosol forcing. When considering only global mean temperature, we remove the polar amplification signal from the temperature diagnostic. Removing this signal means that we ignore the spatial dependence of the aerosol distribution and only consider the net effect on the global energy budget. However, as we include variations of temperature with latitude, the spatial pattern of the aerosol forcing pattern matters. As a result, the median estimate of F_{aer} shifts from -0.31 Wm^{-2} to -0.28 Wm^{-2} to -0.23 Wm^{-2} when global mean, hemispheric mean, and four zonal bands are used. Thus, while the spatial structure has only a small influence on ECS and K_v, it has a strong influence on F_{aer}.

3.4 Conclusions

We implement a number of methodological changes to improve probability estimates of climate system properties. Changes include switching to an interpolation based on radial basis functions, estimating natural variability from a common model across diagnostics, using new observational datasets, and incorporating recent temperature changes in model diagnostics. We show that the parameter estimates follow signals in the data and depend on the model diagnostics. Furthermore, we show that the technical changes, namely the interpolation method and the natural variability estimate, do not considerably change the central estimate of the parameters but do impact the uncertainty estimates of the distributions.

We have shown that the RBF interpolation method is successful in smoothing the distributions while not changing the central estimate. The success of the RBF method is an encouraging sign for future research. Due to the two-dimensional interpolation method previously used, our work until now has been restricted to running ensembles on a uniform grid of points in the parameter space. The RBF method is three-dimensional and can be applied to any collection of node points. We can thus run the full model at any set of non-gridded nodes and interpolate the goodness-of-fit statistics to estimate the values at intermediate points. Other studies (Sansó and Forest, 2009; Olson et al., 2012) have built statistical emulators to approximate model output at non-node parameter settings for each point in the diagnostic time series and then calculate the likelihood function by comparing the emulator output to observations. We argue that by interpolating the metrics, rather than model output at individual points in the time series, we approximate the impact of all feedbacks on the diagnostic together, rather than individually at different spatial and temporal scales.

Our results suggest that the spatial structure of model diagnostics plays a key role in the estimation of parameters with spatial variation. When adding spatial structure to the diagnostics, we observed little change in parameters representing global mean quantities (ECS and K_v), but the distributions of F_{aer} differed depending on whether global mean temperature, hemispheric mean temperatures, or temperatures in four equal-area zonal bands were used. When global diagnostics are used, we ignore the spatial variation of forcing patterns and fail to account for regional influences on climate change. Our estimates provide an assessment of the importance of these spatial patterns when estimating probability distributions for model parameters.

Overall, our work highlights that recent temperature trends have a strong influence on the parameter distributions. In particular, we observe a shift in the distributions towards higher climate sensitivity due to the addition of recent surface temperature warming trends relative to 1990, but with a reduction in the estimate when using data up to 2010 as opposed to 2000. We also observe that the distributions of K_v shift towards higher values. The uncertainty in our estimates decreases as more recent data are used in the temperature diagnostics. Our estimates of transient climate response reflect the changes in ECS and K_v and are correlated with ECS and anticorrelated with K_v. By incorporating more recent data, which are of higher quality, and using improved methodology, we are more confident in our estimates of the model parameters and transient climate response.

Chapter 4

Underestimating Internal Variability Leads to Narrow Estimates of Climate Sensitivity

4.1 Introduction

Despite improvements to climate models, observational systems, and methodologies, the estimated likely range of equilibrium climate sensitivity has remained relatively unchanged from the Charney Report (Charney, 1979) to the most recent Intergovernmental Panel on Climate Change Fifth Assessment Report (IPCC AR5, Collins et al., 2013). In both reports, the central estimate for climate sensitivity is 3 °C with a likely range of 1.5 to 4.5 °C. The lack of convergence towards a single value does not imply a lack of knowledge, however, but rather the inherent difficulty of the problem. Estimates of climate sensitivity remain uncertain for a number of reasons, some of which are outlined in Hegerl et al. (2000) and remain relevant today. The first reason is uncertainty in observations. Many studies estimate climate sensitivity using historical observations of climate change. Due to instrumental errors, coverage gaps, and inhomogeneity in measurement practices, there is irreducible error in the time series of climate variables. The second reason is systematic/structural errors in models. Many studies derive estimates of climate sensitivity by comparing model output to observations (e.g., Knutti et al., 2003; Forest et al., 2008; Libardoni and Forest, 2013; Olson et al., 2013). All models are approximations of reality, and thus do not exactly match the climate system. The last reason is chaotic or internal variability in the climate system. In the absence of external forcing, fluctuations in the atmosphere and ocean still occur due to processes and feedbacks active in the climate system. This unforced internal variability is embedded in any climate signal, and due to its chaotic nature, represents an irreducible uncertainty.

In this study, we show how estimates of climate sensitivity are affected by different estimates of natural variability. In previous work (Chapters 2 and 3), the goodness-of-fit statistic used to evaluate model performance weighs model-to-observation residuals in temperature diagnostics by the internal variability of the climate system (Eq. 4.1). In many methods (Sansó and Forest, 2009; Aldrin et al., 2012; Olson et al., 2013), internal variability is estimated from the pre-industrial control runs of atmosphere-ocean general circulation models (AOGCMs). In these runs, forcing patterns are fixed over long simulations and the coupled atmosphere and ocean system interacts in the absence of changes in external forcings. Thus far, we have only used one model to estimate the internal variability. Over 20 AOGCMs submitted pre-industrial controls runs to the Coupled Model Intercomparison Project Phase 5 (CMIP5, Taylor et al., 2012). Due to structural differences, the natural variability is not the same across all models and a single model does not span the full range of variability. Here we show that this has a significant impact on the calculation of probability distributions for climate sensitivity and leads to estimates of climate sensitivity and other climate system properties that have uncertainty bounds that are too narrow.

We present the study as follows. In Section 4.2, we identify the models that are used and how estimates of the internal variability are derived from long control runs. Section 4.3 presents probability distributions derived from each of the variability estimates and proposes a method for combining estimates across multiple models. We conclude the study in Section 4.4.

4.2 Methods

Using the 1800-member ensemble of the Massachusetts Institute of Technology Earth System Model (MESM) described in Chapter 2, we derive estimates of the joint probability distribution (PDF) of effective climate sensitivity (ECS), effective ocean diffusivity (K_v), and net anthropogenic aerosol forcing (F_{aer}). Each model simulation is run for a set of the three parameters, $\theta = (ECS, K_v, F_{aer})$, and simulates past climate when driven by historical forcings. We evaluate the likelihood of a given model run by comparing the model output to time series of observed climate change. In particular, we use the decadal mean time series of surface temperature anomalies in four equal-area zonal bands from 1941 to 2010 with respect to a 1906-1995 climatology and the linear trend in the 0-2000 m global mean ocean heat content change between 1955 and 2010 as diagnostics.

For each diagnostic, we calculate a goodness-of-fit statistic, r^2, as the weighted sum-of-squares residuals between model output and observations. Mathematically, r^2 is expressed as

$$r^2 = (\mathbf{x}(\theta) - \mathbf{y})^T \mathbf{C}_{\mathbf{N}}^{-\mathbf{1}} (\mathbf{x}(\theta) - \mathbf{y}), \tag{4.1}$$

where $\mathbf{x}(\theta)$ is the vector of model output for a set of model parameters, $\mathbf{y}$ is the vector of observed data, and $\mathbf{C}_N^{-1}$ is the inverse of the noise-covariance matrix. In our method, the noise-covariance matrix represents the temperature patterns we would expect in the model diagnostics in the absence of external forcing and observational noise. The observations and model output are projected onto the natural variability patterns, with the weight assigned to each element of the residual vector inversely proportional to the size of the deviations expected in the unforced climate. A more thorough discussion of the goodness-of-fit statistic calculation and its properties is provided in Forest et al. (2001). From this statistic, the likelihood of a given model run representing past climate change is calculated and used to derive a joint PDF for the model parameters (Libardoni and Forest, 2011; Lewis, 2013).

We estimate the unforced variability patterns from the pre-industrial control runs of AOGCMs submitted to the CMIP5 archive. We choose to use pre-industrial control runs for two reasons. First, all forcings are fixed at pre-industrial levels, allowing the climate system to evolve without influence from external forcings. Second, control runs tend to be long, allowing for many samples of the model diagnostics to be extracted. From these samples, we are able to calculate a large statistical sample of the internal variability.

In recent work (Chapters 2 and 3), our estimates have used the variability from only one model, the Community Climate System Model, version 4 (CCSM4, Gent et al., 2011), in the derivation of parameter distributions. All models are approximations of reality, and thus have internal variability inconsistent with that of the true climate system. We test the sensitivity of the parameter distributions to the internal variability of the climate system by deriving PDFs using natural variability estimates from many different models. To reflect the current state of models across multiple research groups, we choose to use the pre-industrial control runs from 25 models from the CMIP5 archive. In Table 4.1, we present the 25 models used in this study, in addition to the length of the pre-industrial

Table 4.1: CMIP5 models used for internal variability estimates

CMIP5 Institute ID*	Model Name**	Length (years)	Color (Fig 4.1)
BCC	bcc-csm1-1	500	Blue
BCC	bcc-csm1-1-m	400	Green
CCCMA	CanESM2	996	Blue
NCAR	CCSM4	1051	Black
NSF-DOE-NCAR	CESM1-BGC	500	Black
NSF-DOE-NCAR	CESM1-FASTCHEM	222	Green
NSF-DOE-NCAR	CESM1-WACCM	200	Green
CNRM-CERFACS	CNRM-CM5	850	Purple
LASG-CESS	FGOALS-s2	501	Black
NOAA GFDL	GFDL-CM3	500	Blue
NOAA GFDL	GFDL-ESM2G	500	Blue
NOAA GFDL	GFDL-ESM2M	500	Black
NASA GISS	GISS-E2-H	780	Black
NASA GISS	GISS-E2-H-CC	251	Red
NASA GISS	GISS-E2-R	2113	Blue
NASA GISS	GISS-E2-R-CC	251	Green
IPSL	IPSL-CM5A-LR	1000	Blue
IPSL	IPSL-CM5A-MR	300	Red
IPSL	IPSL-CM5B-LR	300	Red
MPI-M	MPI-ESM-LR	1000	Blue
MPI-M	MPI-ESM-MR	1000	Black
MPI-M	MPI-ESM-P	1156	Blue
MRI	MRI-CGCM3	500	Blue
NCC	NorESM1-M	501	Blue
NCC	NorESM1-ME	252	Orange

*See Table B.1 for participating institutions

**See Table B.2 for full model names

control run.

From each model, we extract multiple surrogate estimates of the temperature diagnostics to determine the patterns of unforced variability. To treat the monthly mean surface temperature data from the models as if they were observations, we first regrid the model output from its native grid to the 5x5 degree grid of the observations (Hansen et al., 2010; Vose et al., 2012; Morice et al., 2012; Rohde et al., 2013). In the regridding process, we take the area-weighted average of all grid boxes in the model grid that fall partially or completely within the 5x5 degree grid box of the observations. From the 5-degree resolution data, we extract 105-year segments to correspond to the 1906-2010 period used in the surface temperature diagnostic. To maximize the number of segments extracted from each control run, we shifted the start year of each segment forward four years in the simlation. We choose four years as a compromise between wanting to maximize the number of samples and maintaining a degree of independence between the samples.

After the output is regridded and the segments extracted, we mask the model output using the missing data mask from each of the observational datasets. For each control run segment, the output is masked so that the coverage matches that of the observational dataset for the length of the time series. We then treat individual segments as if they are observations and calculate the decadal mean temperature anomaly time series as in Section 3.2. The noise-covariance matrix used in the goodness-of-fit statistic calculation is estimated by calculating the spatial and temporal correlations across all segments from an individual model. Thus, each model has its own noise-covariance estimated from its internal variability patterns.

We extract noise estimates of the ocean diagnostic following methods similar to those for the surface data. From the gridded data and depth field provided in the model documentation, we calculate the average global mean potential temperature in the 0-2000 m layer. We ignore the small differences between potential temperature and temperature observed at these depths. From the global mean average time series, we extract 55-year segments corresponding to the 1955-2010 ocean diagnostic. We again separate the starting date of the samples by four years. Before calculating the trend, we convert the mean temperature in the layer to ocean heat content using the conversion factor 900/0.09 EJ/°C from Levitus et al. (2012). Because the ocean diagnostic is a single point, namely the linear trend in global mean ocean heat content, the noise-covariance matrix simplifies to the standard deviation in the estimate of the slope.

Once the PDFs are calculated using the variability estimates derived from each CMIP5 model, we extract the median value of each parameter from its marginal distribution. We run simulations of MESM using these parameter sets to further test the distributions *a postiori*. Namely, we distinguish models by whether the MESM runs with the median estimates are consistent with observed changes in global mean temperature change and/or ocean heat content by comparing long-term temperature changes simulated by the model with the changes observed in the historical records.

4.3 Results

We present probability distributions for the model parameters using the variability estimates from each of the CMIP5 models in Table 4.1 (Figure 4.1). As in Chapter 3, we estimate the joint PDF using each of the four surface temperature datasets individually and merge the estimates by taking the average of the four PDFs. Although presented without medians or 90-percent confidence intervals, it is clear that the PDFs are sensitive to which CMIP5 model is used to estimate the variability. The sensitivities arise because the patterns of internal variability estimated from each CMIP5 model are different, and the elements of the $\mathbf{x}(\theta) - \mathbf{y}$ vector project differently onto these patterns. Thus, a difference between the observations and model output that is assigned little weight with the pattern from one model may be assigned a large weight with the pattern from a second model.

We assume that the internal variability of each CMIP5 model is an equally-likely representation the climate system. We thus assign equal weight to each estimate and do not judge one model as better than any other. For this reason, we have chosen to not label the individual distributions. However, we have color-coded the distributions (see Table 4.1) based on three selection criteria as described next and summarized in Table 4.2. For criterion one, we identify models based on their control run length being longer or shorter than 500 years. As the length of the control run increases, the number of estimates of the diagnostics to determine the noise-covariance matrix increases. For the second and third criteria, we use the MESM simulations with parameters set to the median values from each of the marginal distributions shown in Figure 4.1. We calculate the difference between the global mean surface temperature in the last decade of the simulation (2001-2010) and the first 20 years of the simulation (1861-1880) and compare the difference to the observed climate record (Morice et al., 2012). We have identified models where the absolute value of the difference between the temperature change estimated from the model and calculated from the observations is less than 0.05 °C. The overall range of 0.1 °C is slightly wider than the spread in the estimates (0.08 °C) when the temperature change is calculated using different surface temperature datasets for a common period (Hansen et al., 2010; Morice et al., 2012; Vose et al., 2012; Rohde et al., 2013). Similarly, we have identified models by whether the simulated change in global mean ocean heat content matches the observed change.

For most of the models, we find climate sensitivity to be centered between 3.3 and

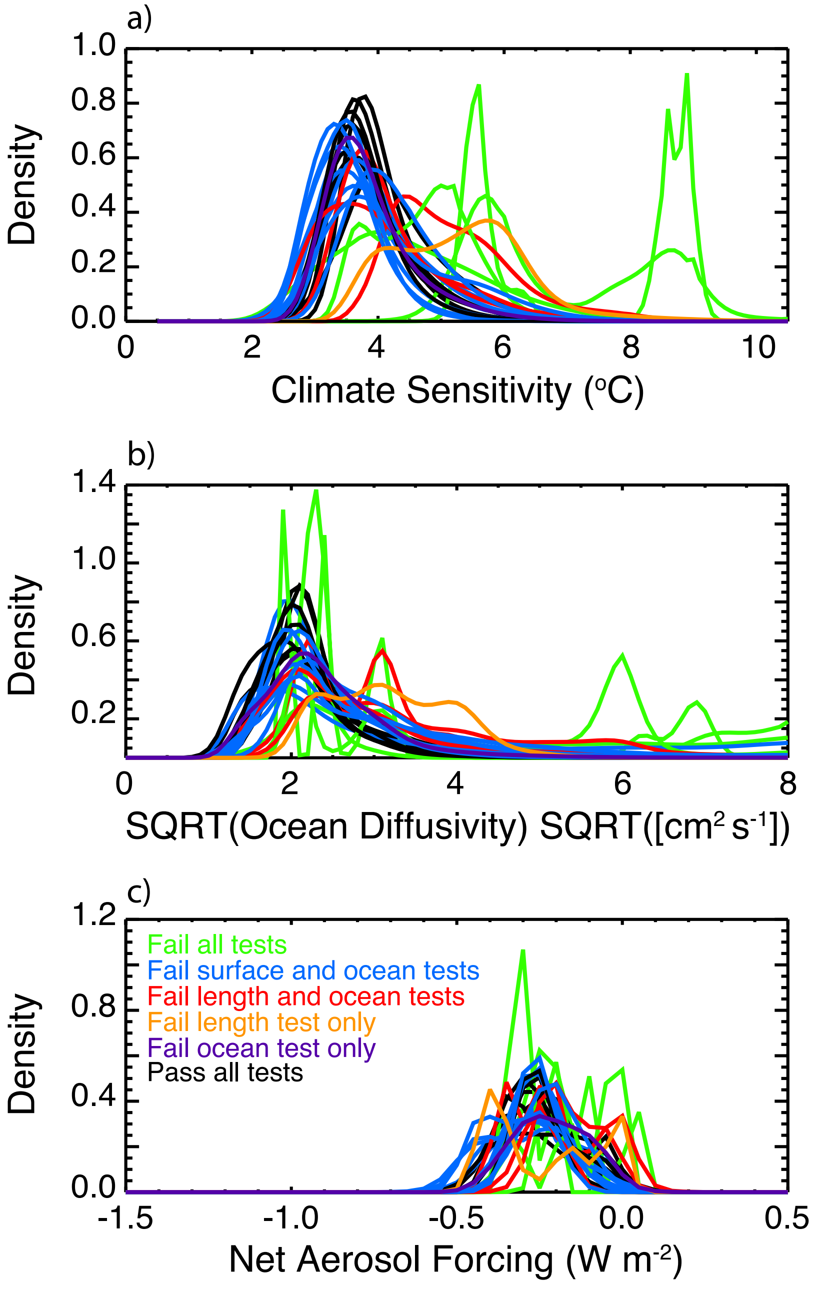

Figure 4.1: Marginal probability distribution functions for (a) ECS, (b) K_v, and (c) F_{aer} derived using variability estimates from each of the 25 CMIP5 models. Distributions are colored based on the selection criteria in Table 4.2.

Table 4.2: Selection criteria for CMIP5 models. A 1 indicates that a model does and a 0 indicates that a model does not pass a given test.

Model Name	Color	Length	Surface	Ocean
bcc-csm1-1	Blue	1	0	0
bcc-csm1-1-m	Green	0	0	0
CanESM2	Blue	1	0	0
CCSM4	Black	1	1	1
CESM1-BGC	Black	1	1	1
CESM1-FASTCHEM	Green	0	0	0
CESM1-WACCM	Green	0	0	0
CNRM-CM5	Purple	1	1	0
FGOALS-s2	Black	1	1	1
GFDL-CM3	Blue	1	0	0
GFDL-ESM2G	Blue	1	0	0
GFDL-ESM2M	Black	1	1	1
GISS-E2-H	Black	1	1	1
GISS-E2-H-CC	Red	0	1	0
GISS-E2-R	Blue	1	0	0
GISS-E2-R-CC	Green	0	0	0
IPSL-CM5A-LR	Blue	1	0	0
IPSL-CM5A-MR	Red	0	1	0
IPSL-CM5B-LR	Red	0	1	0
MPI-ESM-LR	Blue	1	0	0
MPI-ESM-MR	Black	1	1	1
MPI-ESM-P	Blue	1	0	0
MRI-CGCM3	Blue	1	0	0
NorESM1-M	Blue	1	0	0
NorESM1-ME	Orange	0	1	1

4.3 °C (Figure 4.2). However, we do observe variation between the individual estimates and outliers above 5 °C and as high as 8 °C. We see a similar spread in the marginal distributions of $\sqrt{K_v}$. While the distributions tend to cluster towards lower values between 1.8 and 2.8 cm $s^{-1/2}$, there are a few estimates with much higher $\sqrt{K_v}$. We note that all of the outlier distributions are derived from models with less than 500 years of control run data. For these models, there are not enough control run segments to accurately estimate the noise-covariance matrix. We see the most agreement in the F_{aer} distributions. There are no extreme outliers and the distributions all tend to center between -0.3 and -0.1 Wm^{-2}. However, the range in the estimates is greater than those found when considering other factors such as the end date of the diagnostic and the surface temperature dataset used (see Chapters 2 and 3).

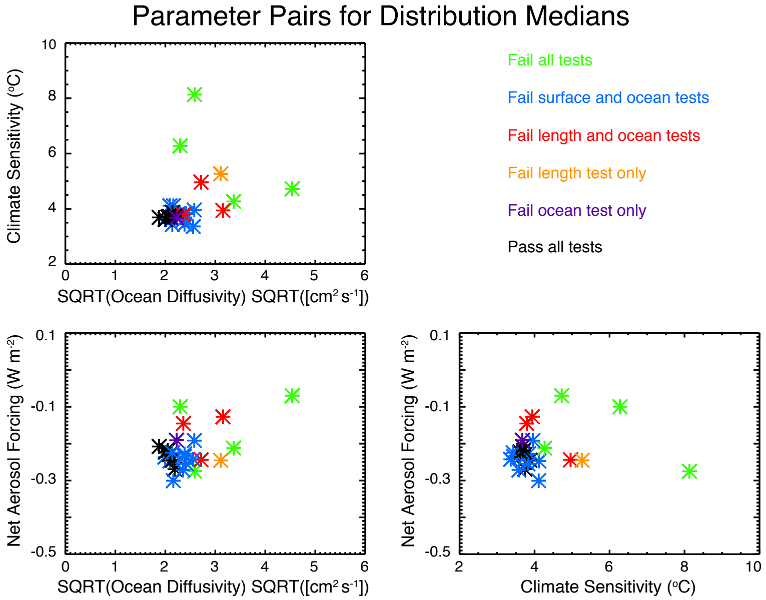

Figure 4.2: Median parameter values extracted from each distribution in Figure 4.1 for each combination of parameters. Colors are as described in Tables 4.1 and 4.2.

For each of the separation criteria— length of the control simulation, MESM simulation with median parameters matching global mean temperature change, and MESM simulation matching global mean ocean heat content change —we group the diagnostic segments from all models that pass the test into a single, merged pool of estimates. In doing so, we combine variability estimates for models with similar length and variability characteristics, thus increasing the number of segments available to calculate the covariance matrix. From the increased collection of segments, we estimate a noise-covariance matrix and derive an additional PDF using the estimate. In total, four groupings are evaluated: (1) all models, (2) all models with a 500 year or greater control run, (3) all models where the median distribution values lead to a MESM simulation consistent with the observed global mean temperature change, and (4) all models where the median parameter simulation is consistent with both the global mean surface temperature and ocean heat content changes. The models that are in each grouping are given in Table 4.3 and the resulting distributions are shown in Figure 4.3.

We observe a much smaller spread in the parameter estimates when the merged variability estimates are used to calculate the noise-covariance matrix compared to when individual CMIP5 models are used. Climate sensitivity estimates are nearly independent of the criteria used to pool the variability estimates. Across the four options, the 5th

Table 4.3: Model groupings by selection criteria

Parsing Method	Models Used	# of Models	Colors (Table 4.1)
Simulation Length	bcc-csm1-1, CanESM2, CCSM4, CESM1-BGC, CNRM-CM5, FGOALS-s2, GFDL-CM3, GFDL-ESM2G, GFDL-ESM2M, GISS-E2-H, GISS-E2-R, IPSL-CM5A-LR, MPI-ESM-LR, MPI-ESM-MR, MPI-ESM-P, MRI-CGCM3, NorESM1-M	17	Black, Blue, Purple
Surface Test	CCSM4, CESM1-BGC, CNRM-CM5, FGOALS-s2, GFDL-ESM2M, GISS-E2-H, GISS-E2-H-CC, IPSL-CM5A-MR, IPSL-CM5B-LR, MPI-ESM-MR, NorESM1-ME	11	Black, Purple, Red, Orange
Surface and Ocean Tests	CCSM4, CESM1-BGC, FGOALS-s2, GFDL-ESM2M, GISS-E2-H, MPI-ESM-MR, NorESM1-ME	7	Black, Orange

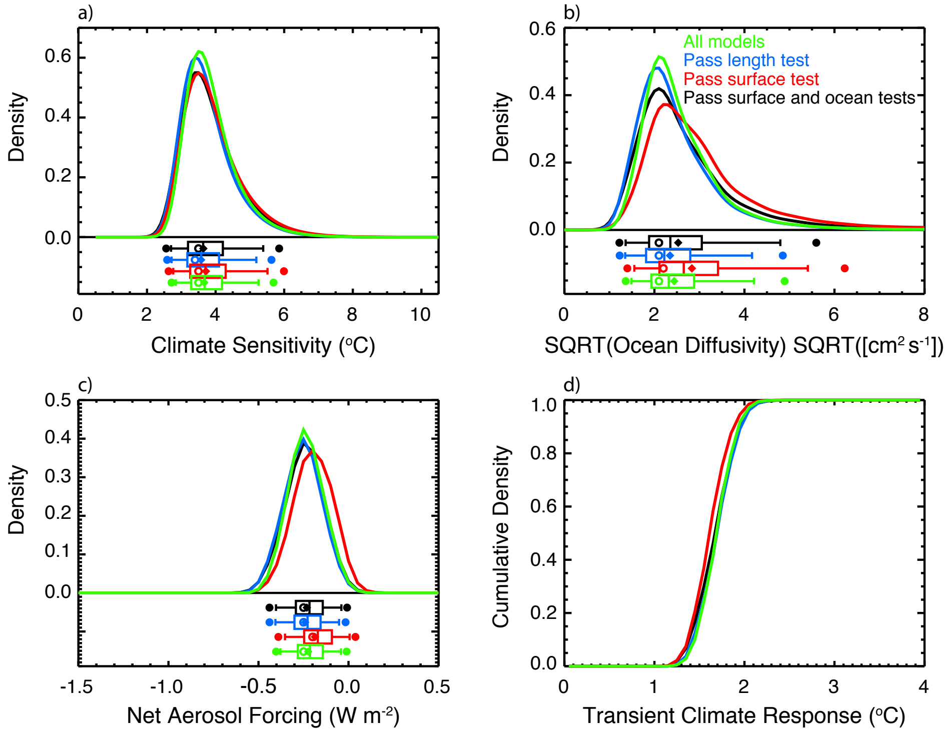

Figure 4.3: Marginal probability distribution functions for ECS (a), K_v (b), and F_{aer} (c) and cumulative distribution function for TCR (d) resulting from merging the variability estimates from the model groupings in Table 4.3. Groupings are for all models, all models with length greater than 500 years, all models with median parameters consistent with global mean temperature change, and all models with median parameters consistent with global mean temperature change and ocean heat content change. Colors for the merged groups are chosen so that the color of the last test passed before failure matches that from Figure 4.1 (e.g., models that fail all tests are plotted green in Figure 4.1 and therefore only fit in the "all models" group).

and 95th percentiles of the distributions vary between 2.7 and 2.8 and 5.2 and 5.6 °C, respectively. When compared to the 90-percent confidence interval of 3.1 to 4.9 °C estimated from the distribution using only the CCSM4 internal variability derived in Chapter 3, this represents a widening of the distribution. We note that similar to the climate sensitivity distributions, the F_{aer} distributions show strong agreement regardless of which tests are used to group variability estimates. Unlike with the climate sensitivity and aerosol distributions, we do not observe the K_v distributions collapsing towards a single distribution with the different merged variability estimates. We estimate 90-percent confidence intervals of 1.4-4.8, 1.4-4.2, 1.6-5.4, and 1.5-4.2 cm $s^{-1/2}$ when variability from all models, models longer than 500 years, models that pass the surface temperature test, and models that pass both the surface temperature and ocean heat content tests are used to estimate the noise-covariance matrix, respectively. We also note a much longer upper tail in distributions when merged variability estimates are used compared to the distribution in Chapter 3 which used variability from CCSM4 alone (1.5-3.0 cm $s^{-1/2}$ 90-percent confidence interval).

Following the methods of Chapter 2, we draw 1000-member Latin Hypercube Samples

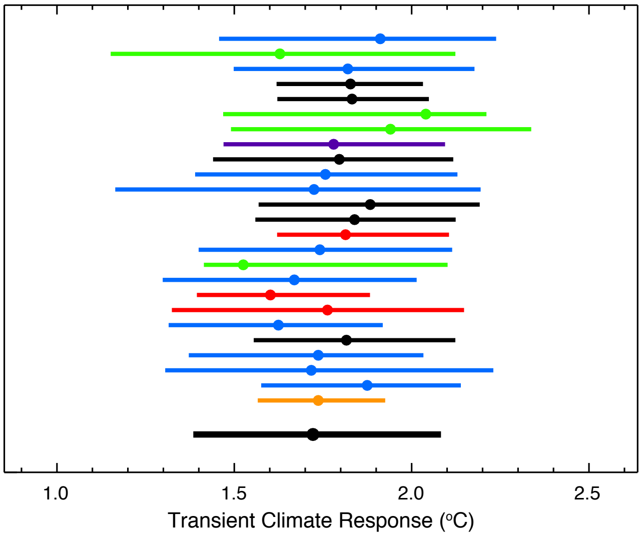

Figure 4.4: 90-percent confidence interval (horizontal line) and median (dots) from the probability distribution functions of transient climate response derived from 1000-member Latin Hypercube Samples drawn from the joint PDFs derived from the internal variability estimates of individual models. Colors are as in Tables 4.1 and 4.3. Also shown is the estimate from the PDF derived using the merged variability segments from all models (bottom bold black line).

(McKay et al., 1979) from the joint PDF derived using the variability estimate from each model individually. We convert each ECS-$\sqrt{K_v}$ pair to transient climate repsonse using the functional fit derived in Chapter 2. We show the median and 90-percent confidence intervals for the TCR distributions derived from the variability estimates from each individual model along with the TCR distribution derived using the merged variability estimate including all models in Figure 4.4. Similar to the ECS and K_v distributions, the estimates of TCR are sensitive to the model used to estimate the internal variability. When compared to the transient climate response PDF using only CCSM4 (Chapter 3), using the merged variability across all models broadens the 90-percent confidence interval from 1.6-2.0 °C to 1.4-2.1 °C. This wider distribution suggests that using a single model can under sample the range of natural variability and lead to TCR estimates that are too narrow.

4.4 Conclusions

In this study, we show that the natural variability estimate used in the likelihood function to weigh the residuals between model output and observations has a strong impact on parameter estimation. By estimating parameters using natural variability from a single model, we observe that the posterior distributions are highly sensitive to the control run data. We have explored several criteria to account for these differences. In particular, some models have control runs that are too short to extract enough samples of variability to provide a good estimate of the noise-covariance matrix. Combining the variability from multiple models provides enough samples for a stable covariance estimate and leads to a convergence of the distributions. Furthermore, if a single model is used, the range of natural variability is under-sampled, leading to narrower estimates of model parameters and projections. In particular, the constraints on all model parameters and transient climate response were tighter when using the CCSM4 model alone, as opposed to when the variability estimates from 25 CMIP5 models are merged to estimate the noise-covariance matrix.

Chapter 5

Summary and Conclusions

In this dissertation, we derived joint probability distributions for three model parameters—effective climate sensitivity (ECS), effective ocean diffusivity (K_v) and the net anthropogenic aerosol forcing (F_{aer}) —representing key climate system properties by comparing model output to historical observations. In Equation 1.5, we presented the goodness-of-fit statistic, r^2, that is used to evaluate model performance for a given set of the model parameters. This statistic depends on three components:

1. Climate model output
2. Observations of past climate change
3. An estimate of the internal chaotic variability of the climate system

We have improved our estimates of the distributions by systematically changing and improving each component in the evaluation procedure. The dissertation is structured as a set of three stand-alone papers, each focusing on an individual component of the goodness-of-fit statistic: the climate model (Chapter 2), the observations used to compare the model against (Chapter 3), and the estimate of internal variability in the climate system (Chapter 4). By taking this systematic approach, we are able to identify the causes of the shifts we observe in the distributions of the climate model parameters.

Chapter 2 is the foundation of this dissertation. In Chapter 2, we tested how the probability distributions for climate system properties changed due to updating the climate component of the MIT Integrated Systems Model (IGSM, Sokolov et al., 2005) to the MIT Earth System Model (MESM, Sokolov et al, in prep). Updating the model involved changing the land surface component from CLM 2.1 to CLM 3.5 and extending the historical forcing records so that model runs can simulate the climate up to 2010.

Due to the discontinuation of the solar and ozone datasets used to force IGSM, new time series of solar and ozone forcing were used in MESM. Two separate ensembles were run with the fully updated model. The first was an 1800-member ensemble of historical simulations from 1860 to 2010. Model runs were chosen to sample the parameter space densely and uniformly over a wide range of the parameter values. The second was a 372-member ensemble of transient simulations where all forcings were held fixed and CO_2 concentrations were increased at the rate of 1 %/yr. From the transient simulations, we derived a response surface for transient climate response as a function of ECS and $\sqrt{K_v}$. Chapters 3 and 4 use the historical runs and the TCR response surface.

Using the 1800-member ensemble, we tested how the parameter distributions change as a result of the switch from IGSM to MESM. By using identical model diagnostics to evaluate model performance and subsampling the model runs to a common set of parameter values, we made a direct comparison between the models. Because of the weaker forcing trends over the historical period from the new solar and ozone datasets, estimates of the 90-percent confidence interval for F_{aer} shifted from -0.83 to -0.19 Wm^{-2} in IGSM (Libardoni and Forest, 2011) to -0.53 to -0.03 Wm^{-2} in MESM. F_{aer} represents the negative radiative forcing due to aerosols, and because the solar and ozone forcing components are reduced with the new datasets, the aerosol cooling does not need to be as strong to achieve the same total net forcing. The decrease in historical forcing also led to a shift towards higher ECS because the same temperature patterns needed to be matched with a weakened forcing signal. This requires an increase in the temperature response to a unit change in forcing. Estimates increased from 1.2 to 5.3 °C in IGSM to 1.5 to 5.7 °C in MESM. Estimates of K_v went from unconstrained with an unbounded upper tail using IGSM (Libardoni and Forest, 2011) to being bounded within the ensemble domain for three of the surface temperature datasets in MESM (Figure 2.4). These findings not only give us insight into the impact that the change in model has on the parameter distributions, but also gives a reference point from which to assess the impact of adding more recent data to the model diagnostics (Chapter 3) or changing the unforced variability estimate (Chapter 4).

In Chapter 3, we applied a number of methodological changes to improve the parameter estimates. First, we implemented a radial basis function interpolation algorithm to produce smooth distributions and eliminate the need for ensembles to be run on a uniform grid. Through a case study with a sample distribution of goodness-of-fit r^2 values, we tuned the algorithm to find a balance between smoothing the values throughout the parameter space and making sure the values at the node points do not significantly change. We showed that the RBF algorithm does indeed produce smooth surfaces and does not change the central estimate of the distribution (Section 3.3.1). A second change we implemented was an averaging of the distributions derived using individual surface temperature datasets. In offline calculations, we found the average distribution to be very similar to distributions drawn by sub-sampling individual distributions and merging the sample. We take the average to account for uncertainty in the estimates due to the observations.

Using the new methods, we showed that the incorporation of recent temperature changes in the surface temperature and ocean heat content diagnostics significantly impact the parameter distributions. When considering the diagnostics separately, the rise in surface temperatures observed over the last two decades led to higher climate sensitivity estimates than when the warming signals were not included. The steady rise in global mean ocean heat content led to higher estimates of ocean diffusivity and lower estimates of climate sensitivity. Estimates of F_{aer} tended to follow shifts in the other two parameters. When climate sensitivity increased or ocean diffusivity decreased, our estimate of the magnitude of F_{aer} increased, corresponding to a stronger aerosol cooling. Overall, we found the estimated distributions are less uncertain as recent temperature changes are included.

In Chapter 4, we tested the sensitivity of the distributions to changes in the estimates of unforced variability in the climate system. Because the observational record is too short and confounded by forced climate signals, we estimated the variability from pre-industrial control runs of CMIP5 models. We showed that the distributions can change considerably when different estimates are used. By applying a series of tests related to the length of the control run and the agreement between model runs at the median parameter settings from each distribution and long-term temperature changes, we pooled variability estimates across models with similar behavior. We showed that the parameter distributions were much less sensitive to the pooled variability estimates than to the variability estimate from individual models. Furthermore, the uncertainty in our parameter estimates increased when the pooled variability estimates were used compared to when the variability from a single model was used. This indicates that estimating the natural variability from a single model under-samples the natural variability and leads to parameter distributions that are too narrow.

Throughout the dissertation, we have shown that distributions of transient climate response follow changes in climate sensitivity and ocean diffusivity. All estimates of TCR are derived from the functional fit derived in Chapter 2. We observe higher TCR when climate sensitivity increases or ocean diffusivity decreases and lower TCR when climate sensitivity decreases or ocean diffusivity increases. Furthermore, as the uncertainty in either parameter increases, the uncertainty in the TCR distribution also increases.

Given what we have learned in this dissertation, our best estimate of the climate system properties derived from the MIT Earth System Model is with model diagnostics that are extended to include data up to 2010 (Figure 5.1). For the estimate of the noise-covariance matrix, we take the merged unforced variability estimates across all 25 models considered in Chapter 4. Lastly, we merge the distributions derived from each of the four surface temperature datasets. In choosing to extend the diagnostics, we use more detailed

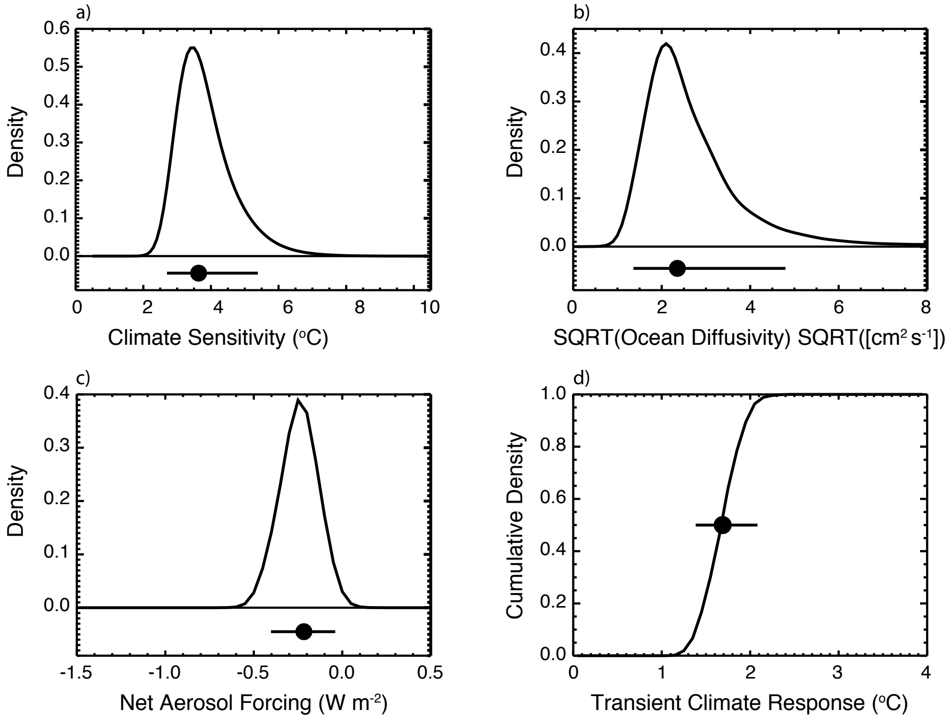

Figure 5.1: Marginal probability distribution functions for ECS (a), K_v (b), and F_{aer} (c) and cumulative distribution function for TCR (d) for our best estimate of the model parameters. Distributions are derived by extending the model diagnostics to include data up to 2010 and are merged across all surface datasets. The pooled variability estimate of natural variability across all CMIP5 models is used to estimate the noise-covariance matrix. Horizontal lines show the 90-percent confidence interval and circles are plotted at the distribution medians.

temperature patterns to evaluate model performance, leading to better constraints on the parameter distributions. By using all of the natural variability estimates and information across all surface datasets, we account for structural uncertainty in the variability estimate and observational uncertainty in the surface temperature record.

Our final estimates of the model parameters and TCR have changed substantially from the distributions in Libardoni and Forest (2011, 2013). Our estimate of climate sensitivity has shifted from 1.2 to 5.3 °C (90-percent confidence interval) with median values ranging from 1.6 to 2.6 °C depending on the surface temperature dataset in Libardoni and Forest (2011) to 2.7 to 5.4 °C with a median value of 3.6 °C and mode of 3.5 °C. In Libardoni and Forest (2011, 2013), ocean diffusivity was poorly constrained with no upper bound to the distribution. Due to this lack of constraint, no numerical estimate was provided. Our analyses have improved the constraint on ocean diffusivity considerably. We estimate the square root of ocean diffusivity to lie between 1.4 and 4.8 cm $s^{-1/2}$ with a median value of 2.4 cm $s^{-1/2}$ and mode of 2.1 cm $s^{-1/2}$. Libardoni and Forest (2011, 2013) estimated the amplitude of the net aerosol forcing in the 1980s to be -0.83 to -0.19 Wm^{-2} with medians ranging between -0.69 and -0.47 Wm^{-2}. We estimate a weaker amplitude of -0.40 to -0.04 Wm^{-2} with a median of -0.22 Wm^{-2} and mode of -0.25 Wm^{-2}. This reduction is partially attributable to the decrease in the strength of the solar and ozone forcings. Lastly, TCR estimates have changed from 0.87 to 2.31 °C with medians ranging between 1.2 and 1.8 °C to 1.4 to 2.1 °C with a median of 1.7 °C and mode of 1.8 °C.

5.1 Commentary and Future Work

Although comprehensive in many aspects, this work has also left open several questions that merit further research. Thus far, we have considered only one ocean heat content dataset in the ocean diagnostic (Levitus et al., 2012). While Levitus et al. (2012) is the only dataset with an estimate of global mean ocean heat content that both starts in 1955 and extends below 700 m, there are a number of global mean ocean heat content datasets for the 0-700 m layer (Gleckler et al., 2016). Analogous to the change in ocean heat content diagnostic presented in Chapter 3, future research could consider replacing the global mean ocean heat content in the 0-2000 m layer with the heat content in the 0-700 m layer. Deriving distributions using each of the datasets and merging them as was done with the surface temperature datasets in Chapters 3 and 4 would allow for an accounting of observational uncertainty in the ocean diagnostic. This could replace or work in conjunction with the methods presented here that use the observational errors as weights in calculating residuals in the regression slope.

The development of the RBF interpolation algorithm opens the door to a vast array of future research. For our work, we are no longer confined to a gridded sampling of the parameter space. This allows for additional flexibility in future ensemble design choices. In our 1800-member ensemble, many runs are in regions of the parameter space where the likelihood function evaluates to zero. While this is helpful in evaluating the tails of the distributions, it also inefficiently allocates computational resources by over-sampling low probability regions. If fewer runs are dedicated to these regions, more runs can be run in regions where the likelihood function changes rapidly. With the ability to sample from anywhere in the domain without being restricted by the gridded design, different strategies can be tested to develop a more efficient means of sampling the parameter space. One such strategy would be to use a Latin Hypercube Sample of the parameter space. Urban and Fricker (2010) showed that such a sampling performs better for an ensemble of size n when compared to a gridded design of the same size. The Latin Hypercube Sampling strategy is just one of many non-gridded strategies and may not be the most efficient. Developing, testing, and implementing new strategies to reduce the number of runs to achieve the same result is an exciting problem and one that would be of great benefit to the scientific community. By using fewer runs, researchers can move down a contour of constant computational resources in Figure 1.1 into regions of greater model complexity. Thus, with efficient sampling strategies, models of higher complexity with more components can be used in probabilistic studies, much like the MESM is used in this work.

A surprising result was the lack of sensitivity in the climate sensitivity and ocean diffusivity distributions to the spatial structure of the surface diagnostic for the period ending in 2010. Before starting this dissertation, we expected there to be more spread among the results derived using the global mean, hemispheric mean, and four zonal mean surface temperature patterns. Initial tests using the 1990 and 2000 end dates showed sensitivity to the diagnostic structure, making the convergence of the distributions in 2010 quite surprising. Whether this convergence is just a coincidence or that global mean parameters in fact show no sensitivity with the longer and more recent temperature patterns is an interesting question that merits a deeper analysis. If it is found that zonal structure does matter, the scientific community may have to rethink the merit of using energy balance models that predict only global mean surface temperature, or else risk averaging over signals crucial to improving estimates of climate system properties.

Appendix A

Radial Basis Functions in Model Parameter Space

A.1 Grid Spacing in Normalized Model Parameter Space

As discussed in Section 3.2, when estimating the r^2 value at intermediate points, the weight assigned to values at node points in the radial basis function interpolation is a function of the distance between the two points. We have chosen to normalize the parameter space for each parameter by the range sampled by the 1800-member ensemble of MESM runs so that each dimension is isometric in the distance calculation. In this normalized space, the grid spacing for each model parameter is:

$$\Delta ECS = \frac{0.5\,^{\circ}\mathrm{C}}{10\,^{\circ}\mathrm{C}} = 0.05 \tag{A.1}$$

$$\Delta K_v = \frac{1\ \mathrm{cm\ s^{-1/2}}}{8\ \mathrm{cm\ s^{-1/2}}} = 0.125 \tag{A.2}$$

$$\Delta F_{aer} = \frac{0.25\ \mathrm{Wm^{-2}}}{2.25\ \mathrm{Wm^{-2}}} = 0.111 \tag{A.3}$$

A.2 Interpretation of ϵ Values in Model Parameter Space

The weight of any node point in the calculation of r^2 at an interpolated point was given in Eq. 3.3 and is a function of the distance between the points and the scaling parameter ϵ. When first developing the algorithm, we hypothesized that having each node point influence the r^2 value at an interpolated point within three grid points in model parameter space would achieve the fit and smoothness we sought from the interpolation. Because the grid spacing in normalized space is not equal for the three parameters, we chose an average of the three individual spacings and used 0.1 as the approximate distance of one grid space. Setting $d = 0.3$ and $\phi = 0.01$ to account for the distance between three nodes and the weight approaching zero at that distance, respectively, we solve for $\epsilon = 7.2$.

To test other ϵ values, we scaled the original choice by factors of 0.5, 1.5, and 2. For $\epsilon = 3.6$, we calculate an e-folding distance of 0.27. This implies a large sphere of influence, as the weight decays to 0.37 at a distance of approximately three grid points away in normalized parameter space. Thus, rather than decay to zero as for the original estimate, there is still significant influence from the node point at $d = 0.3$. This leads to the over smoothing of the r^2 pattern observed in Figure 3.2. In similar calculations, we determine e-folding distances in normalized parameter space of 0.09 and 0.07 for $\epsilon = 10.8$ and $\epsilon = 14.4$, respectively. For $\epsilon = 10.8$, this implies an e-folding distance of approximately one grid space in the $\sqrt{K_v}$ and F_{aer} dimensions, while for $\epsilon = 14.4$, the weight has decayed to 0.13 at a distance of one grid space in those dimensions. Using larger values of ϵ leads to further decay of the weighting function one normalized grid point away from the nodes.

Appendix B
Selected Coupled Model Intercomparison Project Phase 5 Models

Table B.1: CMIP5 Participant Identification Codes

CMIP5 Institute ID	Modeling Center or Group
BCC	Beijing Climate Center, China Meteorological Administration
CCCMA	Canadian Centre for Climate Modelling and Analysis
NCAR	National Center for Atmospheric Research
NSF-DOE-NCAR	Community Earth System Model Contributors
CNRM-CERFACS	Centre National de Recherches Météorologiques / Centre Européen de Recherche et Formation Avancée en Calcul Scientifique
LASG-CESS	State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics (LASG), Institute of Atmospheric Physics, Chinese Academy of Sciences, and Center for Earth System Science, Tsinghua University
NOAA GFDL	National Oceanic and Atmospheric Administration Geophysical Fluid Dynamics Laboratory
NASA GISS	National Aeronautics and Space Administration Goddard Institute for Space Studies
IPSL	Institut Pierre-Simon Laplace
MPI-M	Max-Planck-Institut für Meteorologie (Max Planck Institute for Meteorology)
MRI	Meteorological Research Institute
NCC	Norwegian Climate Centre

Table B.2: CMIP5 Models

Model Abbreviation	Full Model Name
bcc-csm1-1	Beijing Climate Center, Climate System Model, version 1.1
bcc-csm1-1-m	Beijing Climate Center, Climate System Model, version 1.1 (moderate resolution)
CanESM2	Second Generation Canadian Earth System Model
CCSM4	Community Climate System Model, version 4
CESM1-BGC	Community Earth System Model, version 1 (Biogeochemistry)
CESM1-FASTCHEM	Community Earth System Model, version 1 [Community Atmosphere Model with Chemistry (CAM-chem)]
CESM1-WACCM	Community Earth System Model, version 1 (Whole Atmosphere Community Climate Model)
CNRM-CM5	Centre National de Recherches Météorologiques Coupled Global Climate Model, version 5
FGOALS-s2	Flexible Global OceanÐAtmosphere-Land System Model, second spectral version
GFDL-CM3	Geophysical Fluid Dynamics Laboratory Climate Model, version 3
GFDL-ESM2G	Geophysical Fluid Dynamics Laboratory Earth System Model with GOLD component
GFDL-ESM2M	Geophysical Fluid Dynamics Laboratory Earth System Model with MOM, version 4 component
GISS-E2-H	Goddard Institute for Space Studies Model E2, coupled with Hybrid Coordinate Ocean Model
GISS-E2-H-CC	Goddard Institute for Space Studies Model E2, coupled with Hybrid Coordinate Ocean Model and interactive terrestrial carbon cycle (and oceanic biogeochemistry)
GISS-E2-R	Goddard Institute for Space Studies Model E2, coupled with the Russell ocean model
GISS-E2-R-CC	Goddard Institute for Space Studies Model E2, coupled with the Russell ocean model and interactive terrestrial carbon cycle (and oceanic biogeochemistry)
IPSL-CM5A-LR	L'Institut Pierre-Simon Laplace Coupled Model, version 5A, low resolution
IPSL-CM5A-MR	L'Institut Pierre-Simon Laplace Coupled Model, version 5A, mid resolution
IPSL-CM5B-LR	L'Institut Pierre-Simon Laplace Coupled Model, version 5B, low resolution
MPI-ESM-LR	Max Planck Institute Earth System Model, low resolution
MPI-ESM-MR	Max Planck Institute Earth System Model, medium resolution
MPI-ESM-P	Max Planck Institute Earth System Model, paleoclimate
MRI-CGCM3	Meteorological Research Institute Coupled Atmosphere-Ocean General Circulation Model, version 3
NorESM1-M	Norwegian Earth System Model, version 1 (intermediate resolution)
NorESM1-ME	NorESM1-M with carbon cycling and biogeochemistry

Bibliography

Aldrin, M., M. Holden, P. Guttorp, R. B. Skeie, G. Myhre, and T. K. Bernstein, 2012: Bayesian estimation of climate sensitivity based on a simple climate model fitted to observations of hemispheric temperatures and global ocean heat content. *Environmetrics*, **23**, 253–271.

Allen, M. R. and D. J. Frame, 2007: Call off the quest. *Science*, **318**, 582–583.

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., M. Schlesinger, S. Dessai, M. Hulme, and B. Li, 2007: The concept of climate sensitivity: History and development. *Human–induced climate change: An interdisciplinary assessment*, M. E. Schlesinger, H. S. Kheshqi, J. Smith, F. C. de la Caesnaye, J. M. Reilly, T. Wilson, and C. Kolstad, Eds., Cambridge University Press, Cambridge, UK, 5–17.

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.

Bellouin, N., O. Boucher, J. Haywood, and M. Reddy, 2005: Global estimate of aerosol direct radiative forcing from satellite measurements. *Nature*, **438 (22)**, 1138–1141.

Bellouin, N., J. Quaas, J.-J. Morcrette, and O. Boucher, 2013: Estimates of aerosol radiative forcing from the MACC re-analysis. *Atmos. Chem. Phys.*, **13**, 2045–2062.

Bindoff, N. L., et al., 2013: Detection and attribution of climate change: From global to regional. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, T. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels,

Y. Xia, V. Bex, and P. Midgley, Eds., Cambridge University Press, Cambridge, UK and New York, NY, USA, 867–952.

Bony, S., et al., 2006: How well do we understand and evaluate climate change feedback processes? *J. Clim.*, **19**, 3445–3482.

Boucher, O., et al., 2013: Clouds and aerosols. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, T. Stocker, D. Quin, G.-K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Zia, V. Bext, and P. Midgley, Eds., Cambridge University Press, Cambridge, UK and New York, NY, USA, 571–658.

Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones, 2006: Unceratainty estimates in regional and global observed temperature changes: A new data set from 1850. *J. Geophys. Res.*, **111 (D12106)**, doi:10.1029/2005JD006 548.

Charney, J., 1979: Carbon dioxide: A scientific assessment. National Academy of Sciences Press, Washington, D.C., 33 pp.

Cionni, I., et al., 2011: Ozone database in support of CMIP5 simulations: Results and corresponding radiative forcing. *Atmos. Chem. Phys.*, **11**, 11 267–11 292.

Collins, M., et al., 2013: Long-term climate change: Projections, commitments and irreversibility. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, T. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. Midgley, Eds., Cambridge University Press, Cambridge, UK and New York, NY, USA, 1029–1136.

Conley, A. J., J.-F. Lamarque, F. Vitt, W. D. Collins, and J. Kiehl, 2013: PORT, a CESM tool for the diagnosis of radiative forcing. *Geosci. Model Dev.*, **6**, 469–476.

Cowtan, K. and R. G. Way, 2014: Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Q.J.R. Meterol. Soc.*, **140**, 1935–1944.

Cubasch, U., et al., 2001: Projections of future climate change. *Climate Change 2001: The Physical Science Basis. Contribution of Working Group I to the Second Assessment Report of the Intergovernmental Panel on Climate Change*, J. Houghton, Y. Ding, D. Griggs, M. Noguer, P. van der Linden, X. Dai, K. Maskell, and C. Johnson, Eds., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 526–582.

Drignei, D., C. E. Forest, and D. Nychka, 2008: Parameter estimation for computationally intensive nonlinear regression with an application to climate modeling. *Ann. Appl. Stat.*, **2 (4)**, 1217–1230.

Flato, G., et al., 2013: Evaluation of climate models. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, T. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. Allen, J. Doschung, A. Nauels, Y. Xia, V. Bex, and P. Midgley, Eds., Cambridge University Press, Cambridge, UK and New York, NY, USA, 741–866.

Forest, C. E., M. R. Allen, A. P. Sokolov, and P. H. Stone, 2001: Constraining climate model properties using optimal fingerprint detection methods. *Clim. Dyn.*, **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. Let.*, **33 (L01705)**, doi:10.1029/2005GL023 977.

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.

Forster, P. M., 2016: Inference of climate sensitivity from analysis of earth's energy budget. *Ann. Rev. of Earth and Planetary Sciences*, **44 (1)**, 85–106.

Forster, P. M., T. Andrews, P. Good, J. M. Gregory, L. S. Jackson, and M. Zelinka, 2013: Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models. *J. Geophys. Res.: Atmos.*, **118**, 1139–1150.

Forster, P. M. F. and J. Gregory, 2006: The climate sensitivity and its components diagnosed from Earth Radiation Budget data. *J. Clim.*, **19**, 39–52.

Gent, P. R., et al., 2011: The Community Climate System Model version 4. *J. Climate*, **24**, doi:10.1175/2011JCLI4083.1.

Gleckler, P. J., P. J. Durack, R. J. Stouffer, G. C. Johnson, and C. E. Forest, 2016: Industrial-era global ocean heat uptake doubles in recent decades. *Nature Clim. Change*, **6**, 394–398.

Gregg, M., 1987: Diapycnal mixing in the thermocline. *J. Geophys. Res.*, **92 (C5)**, 5249–5286.

Gregory, J. M., 2000: Vertical heat transports in the ocean and their effect on time-dependent climate change. *Clim. Dyn.*, **16**, 501–515.

Gregory, J. M., T. Andrews, and P. Good, 2015: The inconstancy of the transient climate response parameter under increasing CO_2. *Phil. Trans. of the R. Soc. of London*, **373 (2054)**, doi:10.1098/rsta.2014.0417.

Gregory, J. M., R. J. Stouffer, S. C. B. Raper, P. A. Stott, and N. A. Rayner, 2002: An observationally based estimate of the climate sensitivity. *J. Clim.*, **15 (22)**, 3117–3121.

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., R. Ruedy, M. Sato, and K. Lo, 2010: Global surface temperature change. *Rev. Geophys.*, **48 (RG4004)**, doi:10.1029/2010RG000 345.

Hansen, J., et al., 2007: Climate simulations for 1880–2003 with GISS modelE. *Clim. Dyn.*, **29**, doi:10.1007/s00 382–007–0255–8.

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.

Huang, B., et al., 2015: Extended Reconstructed Sea Surface Temperature version 4 (ERSST.v4). Part I: Upgrades and intercomparisons. *J. Clim.*, **28**, 911–930.

Huber, M. and R. Knutti, 2014: Natural variability, radiative forcing and climate response in the recent hiatus reconciled. *Nature Geosci.*, **7**, 651–656.

Johansson, D. J. A., B. C. O'Neill, C. Tebaldi, and O. Haggstrom, 2015: Equilibrium climate sensitivity in light of observations over the warming hiatus. *Nature Clim. Change*, **5**, 449–453.

Jones, P. and A. Moberg, 2003: Hemispheric and large-scale surface air temperature variations: An extensive revision and an update to 2001. *J. Clim.*, **16**, 206–223.

Karl, T., et al., 2015: Possible artifacts of data biases in the recent global surface warming hiatus. *Science*, **348**, 1469–1472.

Kattenberg, A., et al., 1996: Climate models – projections of future climate. *Climate Change 1995: The Science of Climate Change. Contribution of Working Group I to the Second Assessment Report of the Intergovernmental Panel on Climate Change*, J. Houghton, L. M. Filho, B. Callander, N. Harris, A. Kattenberg, and K. Maskell, Eds., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 285–357.

Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby, 2011a: Reassessing biases and other uncertainties in sea surface temperature observations measured *in situ* since 1850: 1. Measurement and sampling uncertainties. *J. Geophys. Res.: Atmospheres*, **116 (D14103)**, doi:10.1029/2010JD015 218.

Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby, 2011b: Reassessing biases and other uncertainties in sea surface temperature observations measured *in situ* since 1850: 2. Biases and homogenization. *J. Geophys. Res.: Atmospheres*, **116 (D14104)**, doi:10.1029/2010JD015 220.

Klimont, Z., S. J. Smith, and J. Cofala, 2013: The last decade of global anthropogenic sulfur dioxide: 2000–2011 emissions. *Environ. Res. Lett.*, **8**, doi:10.1088/1748–9326/8/1/014 003.

Klocke, D., J. Quaas, and B. Stevens, 2013: Assessment of different metrics for physical climate feedbacks. *Clim. Dyn.*, doi:10.1007/s00 382–013–1757–1.

Knutti, R. and G. C. Hegerl, 2008: The equilibrium sensitivity of the Earth's temperature to radiation changes. *Nature Geosci.*, **1**, 735–743.

Knutti, R., T. F. Stocker, F. Joos, and G. Plattner, 2003: Probabilistic climate change projections using neural networks. *Clim. Dyn.*, **21**, 257–272.

Knutti, R., T. F. Stocker, F. Joos, and G.-K. Plattner, 2002: Constraints on radiative forcing and future climate change from observations and climate model ensembles. *Nature*, **416**, 719–723.

Knutti, R. and L. Tomassini, 2008: Constraints on the transient climate response from observed global temperature and ocean heat uptake. *Geophys. Res. Let.*, **35 (L09701)**, doi:10.1029/2007GL032 904.

Kopp, G. and J. L. Lean, 2011: A new, lower value of total solar irradiance: Evidence and climate significance. *Geophys. Res. Let.*, **38 (L01706)**, doi:10.1029/2010GL045 777.

Kriegler, E., 2005: Imprecise probability analysis for integrated assessment of climate change. Ph.D. thesis, University of Potsdam, Potsdam, Germany.

Kunze, E., E. Firing, J. M. Hummon, T. K. Chereskin, and A. M. Thurnherr, 2006: Global abyssal mixing inferred from lowered ADCP shear and CTD strain profiles. *J. Phys. Oceanogr.*, **36 (8)**, 1553–1576.

Lean, J. L., 2000: Evolution of the sun's spectral irradiance since the Maunder Minimum. *Geophys. Res. Let.*, **27 (16)**, 2425–2428.

Ledwell, J., A. Watson, and C. Law, 1993: Evidence for slow mixing across the pycnocline from an open-ocean tracer-release experiment. *Nature*, **364**, 701–703.

Ledwell, J., A. J. Watson, and C. S. Law, 1998: Mixing of a tracer in the pycnocline. *J. Geophys. Res.*, **103 (C10)**, 21,499–21,529.

Levitus, S., J. Antonov, and T. Boyer, 2005: Warming of the world ocean, 1955–2003. *Geophys. Res. Let.*, **32 (L02604)**, doi:10.1029/2004GL021 592.

Levitus, S., et al., 2012: World ocean heat content and thermosteric sea level change (0–2000 m), 1955–2010. *Geophys. Res. Let.*, **39 (L10603)**, doi:10.1029/2012GL051 106.

Lewis, N., 2013: An objective Bayesian improved approach for applying optimal fingerprint techniques to climate sensitivity. *J. Clim.*, **26**, doi:10.1175/JCLI–D–12–00 473.1.

Lewis, N. and J. A. Curry, 2014: The implications for climate sensitivity of AR5 forcing and heat uptake estimates. *Clim. Dyn.*, **45 (1009)**, doi:10.1007/s00 382–014–2342–y.

Libardoni, A. G. and C. E. Forest, 2011: Sensitivity of distributions of climate system properties to the surface temperature dataset. *Geophys. Res. Let.*, **38 (L22705)**, doi:10.1029/2011GL049 431.

Libardoni, A. G. and C. E. Forest, 2013: Correction to "Sensitivity of distributions of climate system properties to the surface temperature data set". *Geophys. Res. Let.*, **40**, doi:10.1002/grl.50 480.

Lorenz, A., H. Held, E. Bauer, and T. S. von Deimling, 2010: Constraining ocean diffusivity from the 8.2 ka event. *Clim. Dyn.*, **34**, 719–734.

MacKinnon, J., L. St. Laurent, and A. Naveria Garabato, 2013: Diapycnal mixing processes in the ocean interior. *Ocean Circulation and Climate.*

Masters, T., 2014: Observational estimate of climate sensitivity from changes in the rate of ocean heat uptake and comparison to CMIP5 models. *Clim. Dyn.*, doi:10.1007/s00 382–013–1770–4.

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.

Meehl, G., et al., 2007: Global climate projections. *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. Averyt, M. Tignor, and H. Miller, Eds., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 747–846.

Meehl, G. A., J. M. Arblaster, J. T. Fasullo, A. Hu, and K. E. Trenberth, 2011: Model-based evidence of deep-ocean heat uptake during surface-temperature hiatus periods. *Nature Clim. Change*, **1**, 360–364.

Meinshausen, M., N. Meinshausen, W. Hare, S. Raper, K. Frieler, R. Knutti, D. Frame, and M. Allen, 2009: Greenhouse-gas emission targets for limiting global warming to 2 degrees C. *Nature*, **458 (7242)**, 1158–1162.

Millar, R. J., A. Otto, P. M. Forster, J. A. Lowe, W. J. Ingram, and M. R. Allen, 2015: Model structure in observational constraints on transient climate response. *Clim. Change*, **131**, 199–211.

Miller, R. L., et al., 2014: CMIP5 historical simulations (1850–2012) with GISS ModelE2. *J. Adv. Model. Earth Syst.*, **6**, 441–477.

Mitchell, J., S. Manabe, V. Meleshko, and T. Tokioka, 1990: Equilibrium climate change – and its implications for the future. *Climate Change. The IPCC Scientific Assessment. Contribution of Working Group 1 to the First Assessment Report of the Intergovernmental Panel on Climate Change*, J. T. Houghton, G. J. Jenkins, and J. J. Ephraums, Eds., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 137–164.

Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones, 2012: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *J. Geohpys. Res.*, **117**, doi:10.1029/2011JD017 187.

Munk, W., 1966: Abyssal recipes. *Deep-Sea Res.*, **13**, 707–730.

Munk, W. and C. Wunsch, 1998: Abyssal recipes II: Energetics of tidal and wind mixing. *Deep-Sea Res.*, **45**, 1977–2010.

Murphy, D. M., S. Solomon, R. W. Portmann, K. H. Rosenlof, P. M. Forster, and T. Wong, 2009: An observationally based energy balance for the Earth since 1950. *J. Geophys. Res.*, **114 (D17107)**, doi:10.1029/2009JD012 105.

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. Clim.*, **8**, 496–514.

Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth, 2004: Quantification of modeling uncertainties in a large ensemble of climate change simulations. *Nature*, **430**, 768–722.

Oleson, K. W., et al., 2008: Improvements to the Community Land Model and their impact on the hydrological cycle. *J. Geophys. Res.*, **113 (G01021)**, doi:10.1029/2007JG000 563.

Olson, R., R. Sriver, W. Chang, M. Haran, N. M. Urban, and K. Keller, 2013: What is the effect of unresolved internal climate variability on climate sensitivity estimates? *J. Geophys. Res.: Atmos.*, **118**, 1–11.

Olson, R., R. Sriver, M. Goes, N. M. Urban, H. D. Matthews, M. Haran, and K. Keller, 2012: A climate sensitivity estimate using Bayesian fusion of instrumental observations and an Earth System model. *J. Geophys. Res.*, **117 (D04103)**, doi:10.1029/2001JD016 620.

Otto, A., et al., 2013: Energy budget constraints on climate response. *Nature Geosci.*, **6**, 415–416.

Piani, C., D. J. Frame, D. A. Stainforth, and M. R. Allen, 2005: Constraints on climate change from a multi-thousand member ensemble of simulations. *Geophys. Res. Let.*, **32 (L23825)**, doi:10.1029/2005GL024 452.

Polzin, K. L., J. M. Toole, J. R. Ledwell, and R. W. Schmitt, 1997: Spatial variability of turbulent mixing in the abyssal ocean. *Science*, **276**, 93–96.

Powell, M. J. D., 1977: Restart procedures for the conjugate gradient method. *Math. Prog.*, **12 (1)**, 241–254.

Rohde, R., et al., 2013: A new estimate of the average Earth surface and land temperature spanning 1753 to 2011. *Geoinfor. Geostat: An Overview*, **1:1**, doi:10.4172/gigs.1000 101.

Sanderson, B. M., 2011: A multimodel study of parametric uncertainty in predictions of climate response to rising greenhouse gas concentrations. *J. Clim.*, **24**, 1362–1377.

Sansó, B. and C. Forest, 2009: Statistical calibration of climate system properties. *Appl. Statist.*, **58**, 485–503.

Sato, M., J. E. Hansen, M. P. McCormick, and J. B. Pollack, 1993: Stratospheric aerosol optical depths. *J. Geophys. Res.*, **98 (D12)**, 22 987–22 944.

Schmidt, G., D. T. Shindell, and K. Tsigaridis, 2014: Reconciling warming trends. *Nature Geosci.*, **7**, 158–160.

Shindell, D. T., et al., 2013: Radiative forcing in the ACCMIP historical and future climate simulations. *Atmos. Chem. Phys.*, **13**, 2939–2974.

Skeie, R., T. Bernsten, M. Aldrin, M. Holden, and G. Myhre, 2014: A lower and more constrained estimate of climate sensitivity using updated observations and detailed radiative forcing time series. *Earth System Dynamics*, **5 (1)**, 139–175.

Smith, S. J., J. van Aardenne, Z. Klimont, R. J. Andres, A. Volke, and S. D. Arias, 2011: Anthropogenic sulfur dioxide emissions: 1850–2005. *Atmos. Chem. Phys.*, **11**, 1101–1116.

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. Clim.*, **21**, 2283–2296.

Sokolov, A., S. Dutkiewicz, P. Stone, and J. Scott, 2007: Evaluating the use of ocean models of different complexity in climate change studies, MIT JP Report 128. *Joint Program Report Series*, **Report 128**, 23 pages.

Sokolov, A., et al., 2005: The MIT Integrated Global System Model (IGSM) version 2: Model description and baseline evaluation. *Joint Program Report Series*, **Report 124**, 40 pages.

Sokolov, A. P., 2006: Does model sensitivity to changes in CO_2 provide a measure of sensitivity to other forcings? *J. Clim.*, **19**, 3294–3306.

Sokolov, A. P., C. E. Forest, and P. H. Stone, 2003: Comparing oceanic heat uptake in AOGCM transient climate change experiments. *J. Clim.*, **16**, 1573–1582.

Sokolov, A. P. and E. Monier, 2012: Changing the climate sensitivity of an atmospheric general circulation model through cloud radiative adjustment. *J. Clim.*, **25**, 6567–6584.

Sokolov, A. P. and P. H. Stone, 1998: A flexible climate model for use in integrated assessments. *Clim. Dyn.*, **14**, 291–303.

Sokolov, A. P., et al., 2009: Probabilistic forecast for twenty-first century climate based on uncertainties in emissions (without policy) and climate parameters. *J. Clim.*, **22**, 5175–5204.

Sriver, R. L., N. M. Urban, R. Olson, and K. Keller, 2012: Toward a physically plausible upper bound of sea-level rise projections. *Clim. Change*, **115**, 893–902.

St. Laurent, L. and H. Simmons, 2006: Estimates of power consumed by mixing in the ocean interior. *J. Climate*, **19**, 4877–4890.

Stainforth, D. A., et al., 2005: Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, **433**, 403–406.

Stevenson, D. S., et al., 2013: Tropospheric ozone changes, radiative forcing and attribution to emissions in the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP). *Atmos. Chem. Phys.*, **13**, 3063–3085.

Stone, P. H. and M.-S. Yao, 1987: Development of a two–dimensional zonally averaged statistical–dynamical model. Part II: The role of eddy momentum fluxes in the general circulation and their parameterization. *J. Atmos. Sci*, **44 (24)**, 3769–3786.

Stone, P. H. and M.-S. Yao, 1990: Development of a two–dimensional zonally averaged statistical–dynamical model. Part III: The parameterization of the eddy fluxes of heat and moisture. *J. Clim.*, **3**, 726–740.

Su, W., N. Loeb, G. Schuster, M. Chin, and F. Rose, 2013: Global all-sky shortwave direct radiative forcing of anthropogenic aerosols from combined satellite observations and GOCART simulations. *J. Geophys. Res.: Atmos.*, **118**, 655–66 905.

Tanaka, K., T. Raddatz, B. O'Niell, and C. Reick, 2009: Insufficient forcing uncertainty underestimates the risk of high climate sensitivity. *Geophys. Res. Let.*, **36 (L16709)**, doi:10.1029/2009GL039 642.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experimental design. *Bull. Am. Meteorol. Soc.*, **93**, 485–498.

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. Clim.*, **20**, 1239–1254.

Trenberth, K. E. and J. T. Fasullo, 2013: An apparent hiatus in global warming? *Earth's Future*, **1**, 19–32.

Urban, N. M. and T. E. Fricker, 2010: A comparison of Latin Hypercube and grid ensemble designs for the multivariate emulation of an Earth system model. *Comp. and Geosci.*, **36 (6)**, 746–755.

Urban, N. M., P. B. Holden, N. R. Edwards, R. L. Sriver, and K. Keller, 2014: Historical and future learning about climate sensitivity. *Geophys. Res. Let.*, **41**, 2543–2552.

Urban, N. M. and K. Keller, 2009: Complementary observational constraints on climate sensitivity. *Geophys. Res. Let.*, **36 (L04708)**, doi:10.1029/2008GL036 457.

Vose, R. S., et al., 2012: NOAA's merged land–ocean surface temperature analysis. *Bull. Am. Meteorol. Soc.*, **93**, 1677–1685.

Waterhouse, A., et al., 2014: Global patterns of diapycnal mixing from measurements of the turbulent dissipation rate. *J. Phys. Oceanogr.*, **44**, 1854–1872.

Webb, M. J., et al., 2006: On the contribution of local feedback mechanisms to the range of climate sensitivity in two GCM ensembles. *Clim. Dyn.*, **27 (1)**, 17–38.

Webster, M., L. Jakobovits, and J. Norton, 2008: Learning about climate change and implications for near-term policy. *Clim. Change*, **89**, 67–85.

Webster, M. and A. Sokolov, 2000: A methodology for quantifying uncertainty in climate projections. *Clim. Change*, **46**, 417–446.

Yao, M.-S. and P. H. Stone, 1987: Development of a two–dimensional zonally averaged statistical–dynamical model. Part I: The parameterization of moist convection and its role in the general circulation. *J. Atmos. Sci.*, **44 (1)**, 65–82.

Yokohata, T., M. J. Webb, M. Collins, K. D. Williams, M. Yoshimori, J. C. Hargreaves, and J. Annan, 2010: Structural similarities and differences between climate responses to CO_2 increase between two perturbed physics ensembles. *J. Clim.*, **23**, 1392–1410.

Yu, H., P. K. Quinn, G. Feingold, L. A. Remer, R. A. Kahn, M. Chin, and S. E. Schwartz, 2009: Remote sensing and *in situ* measurements of aerosol properties, burdens, and radiative forcing. *Atmospheric Aerosol Properties and Climate Impacts: A Report by the U.S. Climate Change Science Program and the Subcommittee on Global Change Research.*, M. Chin, R. Kahn, and S. Schwartz, Eds., National Aeronautics and Space Administration, Washington, D.C., USA.

Vita

Alex Gordon Libardoni

Education

Ph.D. in Meteorology, 2017 (Expected), The Pennsylvania State University, PA
Dissertation title: Improving estimates of climate system properties with additional data and new methods

M.S. in Meteorology, 2011, The Pennsylvania State University, PA
Thesis title: Sensitivity of distributions of climate system properties to surface temperature datasets

B.S. in Physics, 2009, The University of Vermont, VT
Minors: Applied Mathematics and Geography

Assistantships

2009-2010, 2015: Teaching Assistant, The Pennsylvania State University

- METEO 436, Atmospheric Physics I (Radiation and Climate)
- METEO 440W, Principles of Atmospheric Measurements
- METEO 421, Atmospheric Dynamics

2010-present: Research Assistant, The Pennsylvania State University

- Research under the guidance of and in collaboration with Prof. Chris E. Forest

Publications and Presentations

1. Libardoni, A. G., and C. E. Forest (2011), Sensitivity of distributions of climate system properties to the surface temperature dataset, Geophys. Res. Lett., 38, L22705, doi:10.1029/2011GL049431.

2. Libardoni, A. G. and C. E. Forest (2011), Sensitivity of distributions of climate system properties to the surface temperature dataset, Abstract GC11B-0915 presented at 2011 Fall Meeting, AGU, San Francisco, Calif., 5-9 Dec.

3. Libardoni, A. G. and C. E. Forest (2012), Assessment of the sensitivity of distributions of climate system properties to methodology and model diagnostics, Abstract GC43E-1083 presented at 2012 Fall Meeting, AGU, San Francisco, Calif., 3-7 Dec.

4. Libardoni, A. G., and C. E. Forest (2013), Correction to "Sensitivity of distributions of climate system properties to the surface temperature data set", Geophys. Res. Lett., 40, 2309–2311, doi:10.1002/grl.50480.

5. Libardoni, A. G., C. E. Forest, and A. P. Sokolov (2014), Estimates of climate system properties based on recent climate records up to 2010, Abstract GC41F-0655 presented at 2014 Fall Meeting, AGU, San Francisco, Calif, 15-19 Dec.