The Pennsylvania State University

The Graduate School

STATISTICAL SKILL IN THE EMULATION OF CLIMATE MODELS

A Thesis in

Meteorology

by

Ashley Warner

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science

August 2014

The thesis of Ashley Warner was reviewed and approved* by the following:

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

Klaus Keller
Associate Professor of Geosciences

Eugene E. Clothiaux
Professor of Meteorology

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

*Signatures are on file in the Graduate School.

Abstract

A climate model of low complexity can be used to emulate the performance of one with higher complexity by identifying the parameters for each model that yield similar model responses. In this thesis, an energy balance model, the Diffusive Ocean Energy balance CLIMate model (DOECLIM) was calibrated to match the output from 639 simulations of the MIT Integrated Global System Model (IGSM), where each IGSM simulation has a different set of values for three key climate parameters: climate sensitivity, vertical ocean diffusivity and aerosol forcing. The energy balance model estimates the globally averaged climate state based on simplified model physics. The IGSM estimates the zonal mean state of the atmosphere and ocean based on physics in higher complexity climate models, and estimates climate changes that include significant internal variability. The DOECLIM parameters were estimated for each IGSM run to find the parameter settings for the simpler model that would best match results from the more complex model. This allows for the simpler model to be used as an emulator over a range of parameter settings. Two model calibration techniques were used and compared. These techniques are Differential Evolution, a genetic algorithm that produces a single set of parameters providing best fit values, and Markov Chain Monte Carlo, which produces a joint probability distribution for the parameters. The study analyzed the statistical skill, including potential biases, that exist when calibrating the energy

balance model to IGSM output. In particular, the estimated DOECLIM climate sensitivity values tended to be lower than their corresponding IGSM values, particularly for runs with low ocean diffusivity. The parameter estimates also vary depending on the choice of noise model, AR(1) or AR(0) for the atmosphere and ocean temperatures.

Table of Contents

List of Figures

List of Tables

Acknowledgments

This work was partially supported by the National Science Foundation and the United States Department of Energy through the following funding sources: DOE award DE-SC0005171, DOE award DE-SC0004956, NSF award SES-0825915, and the NSF cooperative agreement GEO-1240507. I would also like to thank my thesis advisor, Chris Forest, and my thesis committee, Klaus Keller and Eugene Clothiaux for guidance and discussion through the course of this project.

CHAPTER 1

Introduction

Climate models are used to make climate projections for a variety of applications. A range of climate model complexity exists, and each application requires a choice of climate model with the appropriate balance between skill at representing real-world physics and simplifications to ensure computational efficiency. Climate model complexity ranges from simple energy balance climate models (EBMs) to more complex Atmosphere-Ocean General Circulation Models (AOGCMs) or Earth System Models (ESMs). Energy balance models enable quick computing times compared to other climate models by reducing dimensions (occasionally down to zero dimensional, globally averaged models) and simplifying the model physics [McGuffie and Henderson-Sellers, 2013]. Consider two examples. For projections of regional climate change where one requires information for a particular city or region, one would use model projections from AOGCMs or ESMs with high model resolutions. This provides estimates at an appropriate scale for local and regional

decision makers to make informed responses to climate change. For projections of global mean climate change, where high spatial resolution is not required, an EBM may be used. These estimates may be useful for understanding the climate implications for global policy decisions, and devising strategies to limit global greenhouse gas emissions. In this thesis, EBMs are of particular interest, as they are used in integrated assessment models: models which reach across disciplines to consider climate science, economics and human activity to find optimal policy strategies. Integrated assessment models such as the Dynamic Integrated model of Climate and the Economy (DICE) model [Nordhaus, 2008] require relatively simple climate components that can be efficiently paired with economic models to determine the potential impact of various environmental and economic decisions. Involving climate model projections in future policy decisions requires the uncertainty in those projections to be quantified, such that potential climate risks can be assessed across a range of possible climate futures.

Uncertainties in climate model projections arise partly due to uncertainty in key model parameters. Three of these uncertain model parameters are climate sensitivity, vertical diffusivity in the ocean and aerosol forcing. Multiple studies have used historical observational data to constrain these three model parameters, thereby obtaining estimated ranges for their actual values [Andronova and Schlesinger, 2001; Knutti et al., 2002; Forest et al., 2002; Forest et al., 2006; Forest et al., 2008; Urban and Keller, 2010; Libardoni and Forest, 2011; Olson et al., 2012; Urban et al., 2014]. These studies make use of EBMs or Earth System Models of Intermediate Complexity (EMICs), which can be efficiently run for many different parameter settings. While the exact calibration methods in these studies differ, the methodologies are similar in that the models are run for a large variety of parameter settings and the model responses are then compared to observational data. All three of these key parameters influence surface temperature and ocean heat responses, and different combinations of parameters may produce similar responses.

The goal of the calibration process is to produce a joint probability distribution for the uncertain parameters.

Climate sensitivity, defined here as the equilibrium surface temperature response to a doubling of atmospheric CO_2, is an uncertain parameter of particular interest. Generally, climate sensitivity measures the surface warming due to radiative forcing, taking into account feedback effects that will either increase or decrease the warming that would result from a change in the radiative forcing by CO_2 alone. The IPCC Fifth Assessment Report (AR5) estimates a climate sensitivity that is likely to be in the range of $1.5 - 4.5°C$ [Bindoff et al., 2013].

In addition to studies where climate models are calibrated to observational data, several model emulation exercises have been performed. These studies seek to calibrate simpler models to models of higher complexity, such as AOGCMs, to determine the skill of the simple models in emulating more complex model responses. In addition, the "true" values of model parameters, such as climate sensitivity, are already known for the complex model. Model emulation can therefore be a helpful precursor to emulating the actual climate system and estimating its parameters. Furthermore, the computational efficiency of simple and intermediate complexity climate models allows for these models to be run in large ensembles over a variety of scenarios or parameter ranges more quickly than their complex counterparts. A simple, well-calibrated emulator may allow for a more thorough sampling of parametric uncertainty. It therefore becomes important to determine how well simpler models can emulate complex model responses. Because AOGCMs include the most complete representation of physical understanding, these emulations would not include the complete set of climate system feedbacks that emerge and interact in a changing climate. Olivie and Stuber [2010] calibrated three simple climate models to AOGCM simulations and found differences in the simple climate models' ability to emulate AOGCM results depending on the radiative forcing scenario used. Meinshausen et al. [2011] calibrated the simple Model for

the Assessment of Greenhouse Gas Induced Climate Change (MAGICC) to 19 AOGCMs for use as an emulation tool. Furthermore, Aldrin et al. [2012] used a simple EBM to fit both observational data and data generated by AOGCMs where the "true" climate sensitivity is known, to validate the simple model's ability to estimate real-world climate sensitivity. Results from Aldrin et al. [2012] suggest that when both surface temperature and ocean heat content data were used to constrain the model parameters, the climate sensitivity estimate was satisfactorily close to the expected AOGCM value. One important goal of climate model emulation exercises is to test a simple model's parameter estimation ability by calibrating it to a complex model response where the parameters are known.

Motivation

In this thesis, the primary goal is to calibrate a simple energy balance model to emulate the response of a climate model of higher complexity. I chose to emulate the MIT Integrated Global System Model (IGSM), as it has significant internal variability compared with other EMICs (model details are contained in Chapter 2). While the initial intent was to understand the mapping of the parameters of the MIT IGSM onto the parameters of the EBM, the project developed into a study of the calibration processes and potential biases that arise when calibrating the EBM. In Chapter 3 I present the calibration methods used, in Chapter 4 I present the results from these calibration exercises and I summarize the results in Chapter 5.

CHAPTER 2

Data and Models

In this chapter, I present the two models used in this study, as well as details of the data sets used in the model calibration.

2.1 Simple Non-Linear Earth System (SNEASY) Model

The Simple Non-Linear Earth System (SNEASY) Model is a globally-averaged energy balance model, described in detail in Urban and Keller [2010]. The simplicity of the model allows for fast running times, making it suitable for studies demanding many iterations of the model, whether those iterations involve alterations to the model code or multiple settings of the model parameters. SNEASY consists of a coupled atmosphere-ocean component, a carbon cycle component and a meridional

overturning circulation (MOC) component. The atmosphere-oceean component is the Diffusive Ocean Energy balance CLIMate model (DOECLIM); DOECLIM is described in Kriegler [2005] and Tanaka et al. [2007]. This climate component consists of an energy balance atmospheric model coupled to a one-dimensional, purely diffusive ocean model, and calculates the annual average change in global surface air temperature and global ocean heat content. The model requires an annual time series of the net radiative forcing to drive the climate change scenario. The parameters passed into the DOECLIM component are climate sensitivity (S), defined here as the global average surface temperature response to a doubling in atmospheric CO_2, the ocean's vertical thermal diffusivity (κ) and a scale factor for aerosol forcing (α), which multiplies the aerosol forcing time series. For this study, I used only the DOECLIM climate component of SNEASY, so that the simple model would be analogous to the climate-only component of the MIT IGSM which provided the pseudo-observational data.

2.2 MIT Integrated Global System Model (IGSM)

In this study, I used data obtained from the climate component of the IGSM Version 2 [Sokolov et al., 2005]. Much like DOECLIM, this model consists of an atmospheric component coupled to an ocean component. However, DOECLIM is a highly simplified model compared to the IGSM. The atmospheric model is a zonally-averaged GCM based on the Goddard Institute for Space Studies (GISS) Model II [Hansen et al., 1983]. This model version has 4° resolution in latitude and 11 layers in the vertical. The model includes representations of land, ocean, land-ice and sea-ice. The ocean component consists of a two-dimensional mixed layer model, with vertical diffusion of ocean heat anomalies into a three-dimensional ocean below. Unlike the purely diffusive one-dimensional ocean in DOECLIM, the ocean component accounts for both upwelling and diffusion. Furthermore,

the mixed layer maintains the pole-to-equator temperature gradient by prescribing horizontal heat flux terms such that sea surface temperatures match observed climatology. The model also includes an adjustment to allow for alteration of the model cloud feedback, which in turn changes the climate sensitivity. This feedback is included by adjusting the cloud fraction used by the radiative transfer scheme, with the adjustment depending on the global mean temperature. Thus, as temperatures increase, the cloud adjustment has a greater impact on the radiation budget, thereby providing the altered feedback. This adjustment to the cloud feedback permits one to change the climate sensitivity in the IGSM while other feedbacks are left alone. Because the atmospheric model is derived from GISS Model II [Hansen et al., 1983], it includes the primary atmospheric feedbacks contributing to the non-linear response to anthropogenic forcings. Climate sensitivity is therefore an emergent property in the IGSM, unlike the DOECLIM model, where climate sensitivity is a parameter that is specified for each model run. Nonetheless, climate sensitivity in the IGSM will be treated like a parameter in this study, as the climate sensitivity value for each model run has been determined and can be compared against its corresponding DOECLIM estimate.

2.3 Data

Throughout this study, I used model outputs from the MIT IGSM as pseudo-observational data sets, in that they serve the same role that observational data would take if one were to calibrate DOECLIM to simulate real world climate changes. In a similar way, Sokolov et al. [2003] used outputs from the Coupled Model Intercomparison Project (CMIP) [Meehl et al., 2000] to identify the parameters of the IGSM that matched the response of the CMIP models using transient climate change simulations. Here, I perform a similar study to identify parameters in DOECLIM that yield changes in climate similar to those in the MIT IGSM.

This approach provides two important insights. The first is to test the statistical methodology in a perfect model setup where model parameters are known, prior to using observational data to constrain unknown model parameters (e.g., Forest et al. [2008], Urban and Keller [2010]). The second is to identify how response behaviors differ between models.

While the IGSM produces a large variety of output data, only surface temperature and ocean temperature were of particular interest to this study. These are the two output data sets that can be compared with output data from the DOECLIM model. Because the IGSM model parameters used to produce each simulation are known, I can test whether the DOECLIM calibration returns similar values to those of the MIT IGSM runs. The pseudo-observational data used to calibrate DOECLIM comes from work presented in Forest et al. [2008]. This data set spans the years 1861-2000. I used the annual global surface temperature anomalies and the annual global ocean temperature anomalies (averaged vertically, from the surface down to a depth of 3 km from the surface). This data set contains 639 runs of the IGSM, each run with a different combination of three parameters: climate sensitivity (Cs), ocean vertical diffusivity (Kv) and the aerosol forcing averaged over the 1980's (F_{aer}). The parameters were chosen from a grid sample.

CHAPTER 3

Methods

Here I present the mathematics and calibration techniques used to fit the DOECLIM model to IGSM output data. In total, I calibrated the DOECLIM climate component to emulate output from each of the 639 simulations of the IGSM in Forest et al. [2008]. I used two different calibration techniques, Differential Evolution (DE) and Markov Chain Monte Carlo (MCMC), which will be further outlined in this section.

3.1 Calibrating DOECLIM Parameters

Each IGSM simulation was executed with a different combination of three input parameters: climate sensitivity (Cs), vertical diffusion of ocean heat anomalies (Kv) and average aerosol forcing over the 1980's (F_{aer}). The goal of the calibration process was, for each IGSM simulation, to estimate the DOECLIM parameters

that produced output best matched to the IGSM results. Thus, each set of IGSM parameters corresponds to an analogous set of DOECLIM parameters. The three DOECLIM parameters allowed to vary during the calibration process are climate sensitivity (hereby denoted by S), vertical ocean heat diffusivity (κ), and aerosol forcing scale factor (α). While the climate sensitivities in both models have the same meaning (defined as the equilibrium change in global surface temperature that would result from a doubling of atmospheric CO_2), the parameters that control ocean heat uptake and aerosol forcing are not directly comparable. The ocean heat uptake parameter for the IGSM represents the vertical diffusion of ocean heat anomalies, while the ocean heat uptake parameter for the DOECLIM model represents the vertical diffusion of heat in the globally-averaged ocean. Additionally, the IGSM parameter F_{aer} is defined as the average value of aerosol forcing in W/m^2 over the 1980's, while the DOECLIM parameter α is a scale factor by which the input aerosol forcing time series is multiplied. Each run from Forest et al. [2008] used the same aerosol forcing time series scaled by some value which would result in a final time series where the average over the 1980's is equal to F_{aer}. The same original aerosol forcing time series was put into DOECLIM, scaled by the parameter α. F_{aer} and α can therefore be compared by calculating each value of the scaling α that produces a scaled aerosol forcing time series with an average over the 1980's equal to each value of F_{aer} used in the IGSM runs.

Two variables from the IGSM output data set were used to define the model response: global mean surface temperature anomaly and global mean ocean temperature anomaly (averaged over a depth of 0-3 km). DOECLIM calculates the change in global mean surface temperature and ocean heat content, so the DOECLIM output was converted from a change in heat content, given in 10^{22}J, to a temperature change in K.

3.2 Objective Function

The next step in calibrating DOECLIM to a set of IGSM output data is to define an objective function to be optimized. The objective function is a function of the set of parameters to be estimated, where the parameters that maximize the objective function are also the parameters that provide the best fit of DOECLIM to the IGSM output data. The optimization method (DE) involves systematically running DOECLIM with different combinations of parameters (denoted by the parameter vector θ) in order to find which parameter vector maximizes the objective function, and thus provides the best fit to the IGSM data. It is noted that the parameters S, κ and α are a subset of the parameter vector, θ, which is the vector of parameters being estimated and includes both model parameters and statistical parameters.

I followed the approach to defining the objective function presented in Urban and Keller [2010]. Defining the objective function requires an application of Bayes' Theorem [Bayes, 1763], given by Equation 3.1:

$$p(\theta|D) \propto p(D|\theta) \times p(\theta). \tag{3.1}$$

Bayes' Theorem states that the probability of a parameter vector θ given data D, $p(\theta|D)$, is proportional to the likelihood of the data given the parameter vector, $p(D|\theta)$, multiplied by the prior distribution of the parameters, $p(\theta)$. $p(\theta|D)$ is the posterior distribution, the function that is maximized by the optimal set of parameters.

To define the posterior distribution, it is necessary to first define the likelihood function and the prior distribution. To define the likelihood function, a number of assumptions are made. First, it is assumed that the residuals r_i (that is, the difference between the DOECLIM model results and the IGSM data at each time step) are drawn from an AR(1) first-order autoregressive process. This means that

the residuals include a lag-1 autocorrelation, so that each residual in the time series contains a scaled dependence on the residual one time step before it. The AR(1) noise model helps to account for discrepancies between the two models other than white noise. Residuals were corrected for autocorrelation before being incorporated into the likelihood function. This is known as residual whitening, because the effect of autocorrelation is removed and only white noise is left behind. Equation 3.2 for whitened residuals, w_i, is shown below, where r_i are the residual values and ρ is the lag-1 autocorrelation coefficient. It is then assumed that the whitened residuals w_i are all independent and normally distributed.

$$w_i = r_i - \rho r_{i-1} \tag{3.2}$$

Stationary process variance is defined as $\sigma_p^2 = \sigma^2/(1-\rho^2)$. It is also assumed that the surface temperature residuals are independent from the ocean temperature residuals, so the total likelihood function may be calculated as the product of the likelihood functions for each. Both functions take the same form:

$$p(D|\theta) = (2\pi\sigma_p^2)^{-1/2}\exp\left(-\frac{1}{2\sigma_p^2}r_1^2\right) \times (2\pi\sigma^2)^{-(N-1)/2}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=2}^{N} w_i^2\right). \tag{3.3}$$

These equations introduce several new parameters, which are estimated in the calibration process along with climate parameters S, κ and α. These parameters are the standard deviations of the normally distributed residuals for the surface and ocean temperature time series (σ_{ST} and σ_{OT}, respectively) and the autoregressive coefficients (ρ_{ST} and ρ_{OT}). This gives the parameter vector $\theta = (S, \kappa, \alpha, \sigma_{ST}, \sigma_{OT}, \rho_{ST}, \rho_{OT})$.

Uniform prior distributions with upper and lower boundaries are also defined for each of the parameters, such that the parameter vectors chosen must remain within these boundaries. The boundaries are defined in Table 3.1. As the prior

Table 3.1. Table of DOECLIM parameters and their upper/lower bounds

Parameter Name	Symbol	Units	Lower Bound	Upper Bound
Climate sensitivity	S	K	0	100
Vertical ocean diffusivity	κ	cm^2/s	0	100
Aerosol forcing scale factor	α	none	-10	10
Standard deviation (ST)	σ_{ST}	K	0	10
Standard deviation (OT)	σ_{OT}	K	0	10
Autocorrelation (ST)	ρ_{ST}	none	0	1
Autocorrelation (OT)	ρ_{OT}	none	0	1

distributions are uniform, the maximum of the likelihood function occurs at the same parameter vector as the maximum of the posterior distribution function. The negative of the likelihood function is defined as the function to be minimized.

3.3 Differential Evolution Method

The first method employed to attempt to calibrate DOECLIM to the IGSM results was Differential Evolution (DE). DE was developed by Price et al. [2006] and is a genetic algorithm used to search for a function's global minimum. Genetic algorithms are analogous to the process of evolution, consisting of a population that transforms over a number of iterations. An initial set of parameter vectors are chosen and then proceed through many "generations", with each subsequent generation made up of parameter vectors that better optimize the objective function. In the simplest sense, the population of points eventually converges on a global optimum after a number of iterations. A more detailed description of the algorithm follows.

The DE algorithm first selects an initial population of parameter vectors from within a prescribed set of upper and lower boundaries. Individual parameter vectors are denoted by $\underline{x}_{i,g}$, where i indexes population member number and g indexes generation number. The size of the population (the number of parameter vectors

in each generation) is denoted by the variable NP. A mutant population made up of NP vectors $v_{i,g}$ is then created at each generation. The creation of a single mutant vector $\underline{v}_{i,g}$ occurs as follows. A randomly chosen member of the population, $\underline{x}_{r1,g}$, is added to the scaled difference of two other randomly chosen members of the population, $\underline{x}_{r2,g}$ and $\underline{x}_{r3,g}$, to create a trial mutant parameter vector, $\underline{v}_{i,g}$.

$$\underline{v}_{i,g} = \underline{x}_{r1,g} + F \times \left(\underline{x}_{r2,g} - \underline{x}_{r3,g}\right) \tag{3.4}$$

This occurs NP times, to produce a mutant population with NP members of the form $\underline{v}_{i,g}$. Each mutant parameter vector $\underline{v}_{i,g}$ is crossed over with another vector $\underline{x}_{i,g}$, creating a trial vector $\underline{u}_{i,g}$ that contains some parameter values from $\underline{x}_{i,g}$ and some from $\underline{v}_{i,g}$. The fraction of parameters copied from the mutant vector $\underline{v}_{i,g}$ is determined by a crossover probability, CR. The objective function is evaluated at both $\underline{u}_{i,g}$ and $\underline{x}_{i,g}$. If $\underline{u}_{i,g}$ returns a lower value of the objective function than $\underline{x}_{i,g}$, it replaces $\underline{x}_{i,g}$ in the population. Thus, a new generation is created containing some vectors from the old generation and some new vectors. Overall, the new generation will contain members that better minimize the objective function. This new generation undergoes the same mutation and crossover process, thus creating another new generation, and this continues until either a maximum number of iterations has been completed or the value of the objective function has dipped below a certain threshold value. The DE routine then stops and returns an optimum value for the parameter vector, on which the population converges.

I used the *DEoptim* package written in R [Mullen et al., 2011]. Arguments in the call to DE include the function to be minimized, lower and upper boundaries for each parameter and a set of control parameters. The function to be minimized is defined here as the negative of the likelihood function, which accepts as arguments the parameters to be optimized. Control parameters include the maximum number of iterations to be completed before an optimum value is returned (this parameter was set to 500), the number of population members NP (which defaults to ten

times the length of the parameter vector), crossover probability CR (which defaults to 0.5 and must be in the interval [0,1]) and scaling factor F (which defaults to 0.8 and must be in the interval [0,2]). The DE routine was repeated for each of the 639 IGSM runs, producing one single best fit parameter vector corresponding to each individual IGSM run.

3.4 Markov Chain Monte Carlo Method

While the DE method was initially employed due to its fast running time and the simplicity of producing a one-to-one map between the IGSM and DOECLIM parameter spaces, it is possible to glean additional information about the posterior distribution by calibrating the IGSM using the Markov Chain Monte Carlo (MCMC) method. Rather than returning one set of parameters that provides a best fit to the IGSM output, the MCMC method produces a sample of the joint probability distribution for the parameters being estimated. In other words, DE returns the parameters that maximize the posterior distribution $\mathrm{p}(\theta|D)$, while the MCMC method returns a chain of points in parameter space that represent a sample of the posterior distribution. From this sample, probability distributions were constructed and the maxima of these probability distributions can be estimated. This provides uncertainty information for the parameters and allows one to use DOECLIM as a probabilistic emulator for the IGSM.

A Markov chain is constructed by the *metrop* function given by the *mcmc* package in R [Geyer and Johnson, 2014]. An initial parameter vector is specified, then a random walk algorithm is carried out so that a random path around parameter space is created in small steps, either adding a point to the chain or discarding it at each step. The add/reject criterion is such that the final chain of accepted points represents a sample from the posterior distribution, with many points located around the optimum point, but also some points located in the tails of the

distribution. The MCMC algorithm is further described in Hastings [1970].

A number of control parameters must be selected to be used in the *metrop* algorithm. These include step size, initial starting point, the number of points in the final Markov chain and the fraction of points accepted. Once the final Markov chain is complete, one can begin to analyze the information it contains. I calculated the mean, median and mode from the marginal distribution of each parameter being estimated, for the purpose of comparing these values to the best fits obtained by the DE method. I also plotted and analyzed both the marginal distribution for each single parameter and joint probability distributions of multiple parameters.

CHAPTER 4

Results

The DE method was initially used to calibrate the DOECLIM model. The following results demonstrate the best fits obtained using this method. I then utilized an MCMC technique, which provided a wider breadth of information regarding parameter estimation uncertainty.

4.1 Differential Evolution Results

4.1.1 Time series fit for a typical model run

A "typical" IGSM model run with parameters $Cs = 3$ K, $Kv = 4$ cm^2/s and $F_{aer} = -0.5$ W/m^2 was chosen out of the set of 639 runs, and this run will be presented from here on as an example to visualize the best fit results. This run was chosen as a standard example because its parameter values are closest to the

most likely values estimated in Forest et al. [2008].

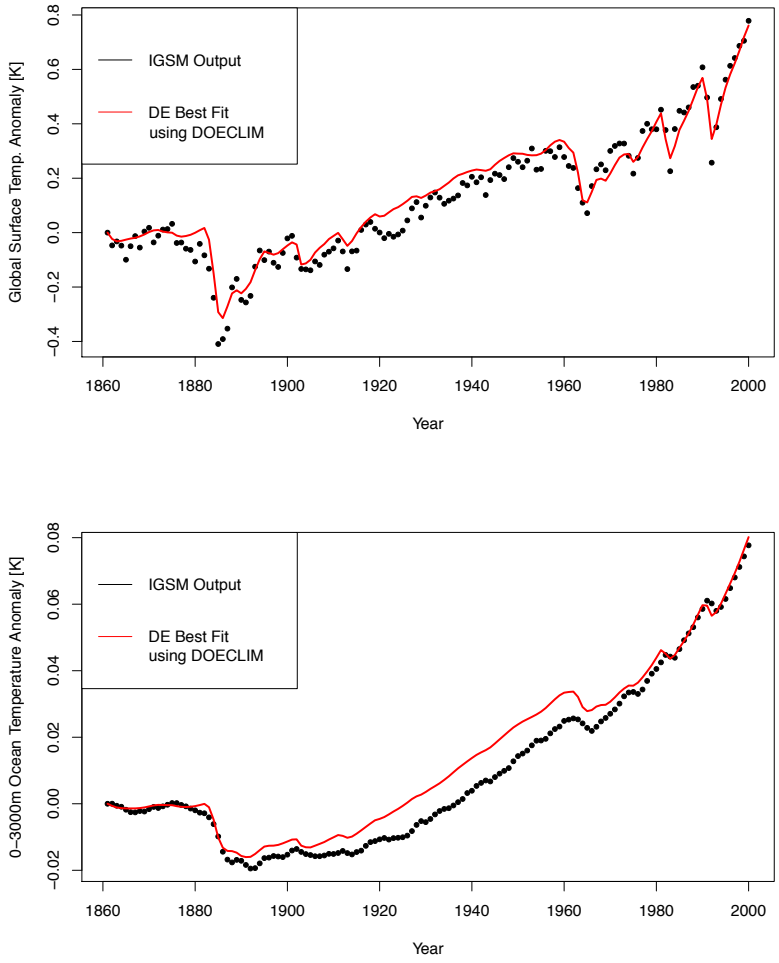

Figure 4.1. IGSM output data (black dots) plotted with DE best fit using DOECLIM (red lines). IGSM parameter values: $Cs = 3$ K, $Kv = 4$ cm^2/s, $F_{aer} = -0.5$ W/m^2. DE Best Fit: $S = 2.61$ K, $\kappa = 1.99$ cm^2/s and $\alpha = 1.65$.

Figure 4.1 shows the best fit of DOECLIM to the IGSM output for the typical model run, where the best fit values returned were $S = 2.61$ K, $\kappa = 1.99$ cm^2/s and $\alpha = 1.65$. While DOECLIM is able to provide a reasonably good match to the global surface temperature anomaly data, it overestimates ocean temperature from about 1890 to 1970. The AR(1) noise model allows for some flexibility in the

fits, as it allows for autocorrelation in the residuals.

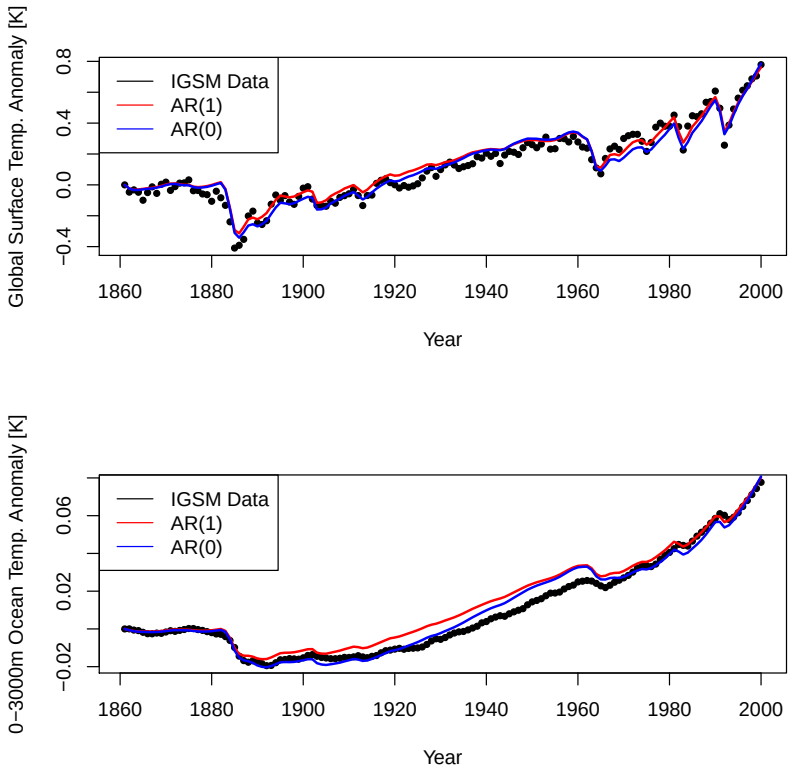

Figure 4.2. IGSM output data (black dots) plotted with DOECLIM best fit using AR(1) noise model (red lines) and AR(0) (blue lines). The DE method was used. IGSM parameter values: $Cs = 3$ K, $Kv = 4$ cm^2/s, $F_{aer} = -0.5$ W/m^2. DE best fit with AR(1) noise: $S = 2.61$ K, $\kappa = 1.99$ cm^2/s and $\alpha = 1.65$. DE best fit with AR(0) noise: $S = 4.23$ K, $\kappa = 2.29$ cm^2/s and $\alpha = 2.10$.

In Figure 4.2, I show the difference between fits when AR(0) noise is used rather than AR(1). The fit with AR(0) adheres more closely to the ocean temperature for the first half of the 20th century, while the AR(1) fit is a looser match. The best fit values found in the AR(0) case were $S = 4.23$ K, $\kappa = 2.29$ cm^2/s and $\alpha = 2.10$. Plots of the autocorrelation function are provided in Appendix B.

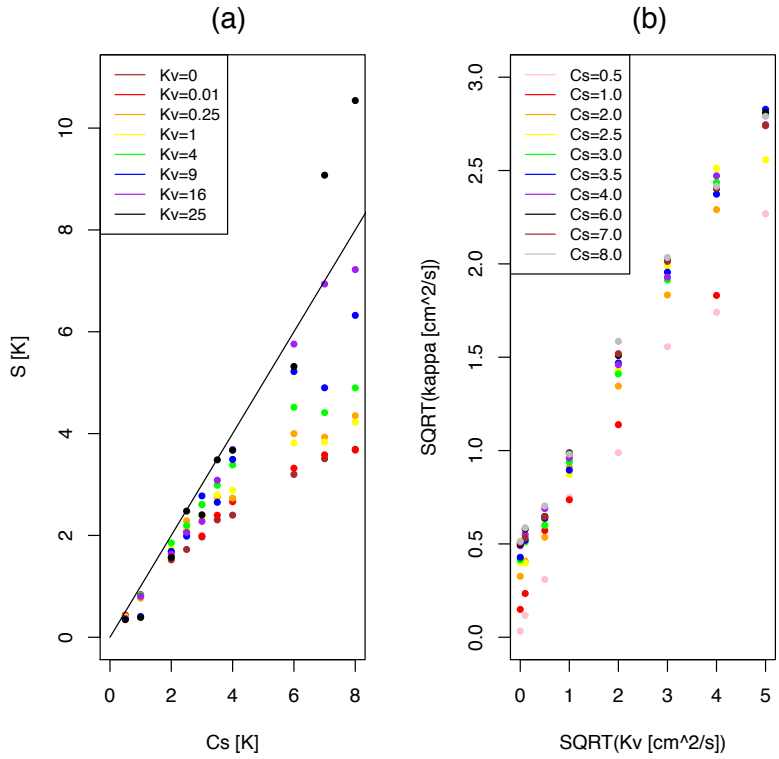

Figure 4.3. Comparison between IGSM parameter values (horizontal axis) and their best fit DOECLIM counterparts (vertical axis) using the DE method, for all runs where $F_{aer} = -0.5$ W/m^2. Comparison of climate sensitivity parameters is shown in (a), while comparison of vertical ocean diffusivity parameters is shown in (b).

4.1.2 Biases between IGSM and DOECLIM parameters

Figure 4.3 shows the relationship between IGSM run parameters Cs and Kv and their best fit DOECLIM counterparts for each run with $F_{aer} = -0.5$ W/m^2. Climate sensitivity parameters can be directly compared, while ocean diffusivity parameters are defined differently for each model and therefore differences are expected. Figure 4.4 shows each of the α estimates for this same set of runs. Almost all of the estimates are higher than expected for the aerosol forcing time series used in these IGSM runs. The expected value of α is 0.943, corresponding to $F_{aer} = -0.5$ W/m^2 averaged over the 1980's. Figures 4.5 and 4.6 show the cli-

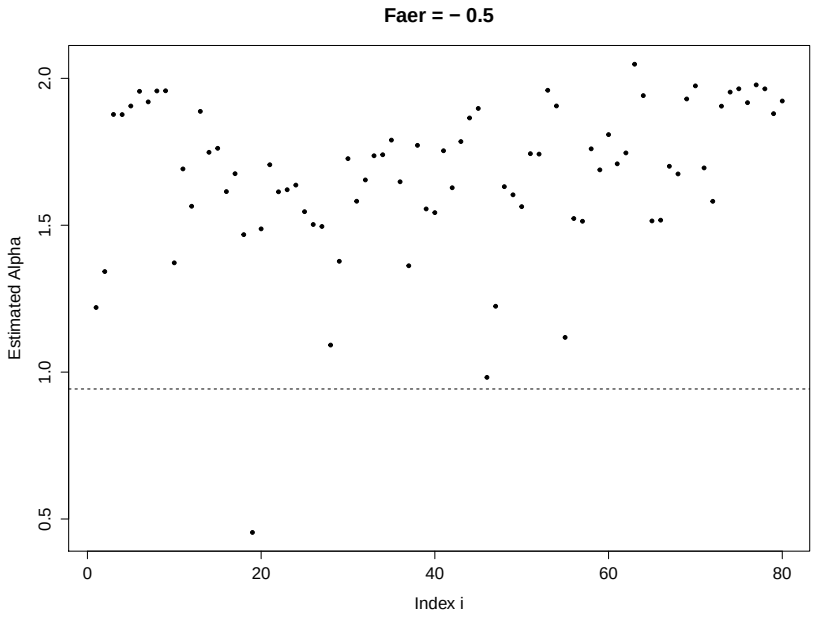

Figure 4.4. DOECLIM estimated alpha values using the DE method, for each run where $F_{aer} = -0.5$ W/m^2. The alpha value that would produce a match to the IGSM aerosol forcing time series (0.943) is indicated with a dotted line.

mate sensitivity and aerosol forcing parameter biases in $Kv - Cs$ space. Climate sensitivity is generally underestimated by the DOECLIM calibration. Biases in climate sensitivity estimates increase as the IGSM run value of climate sensitivity increases. Climate sensitivity estimates also vary depending on the Kv value. For low Kv, these biases tend to be more negative, while positive biases are observed for the few runs where Kv is very high. The greatest biases are observed at high Cs and low Kv values, where the estimated DOECLIM climate sensitivity is as much as four K lower than the IGSM value.

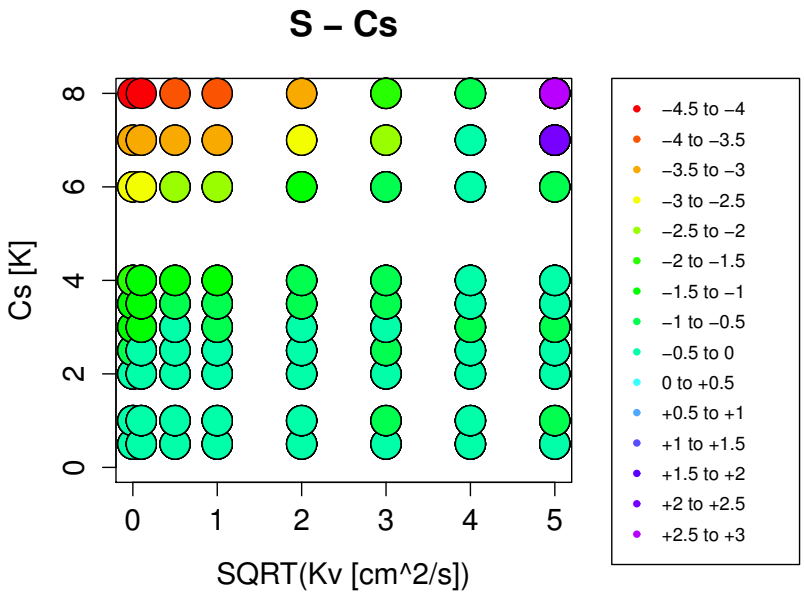

Figure 4.5. Climate sensitivity biases using DE for all runs where $F_{aer} = -0.5$ W/m^2.

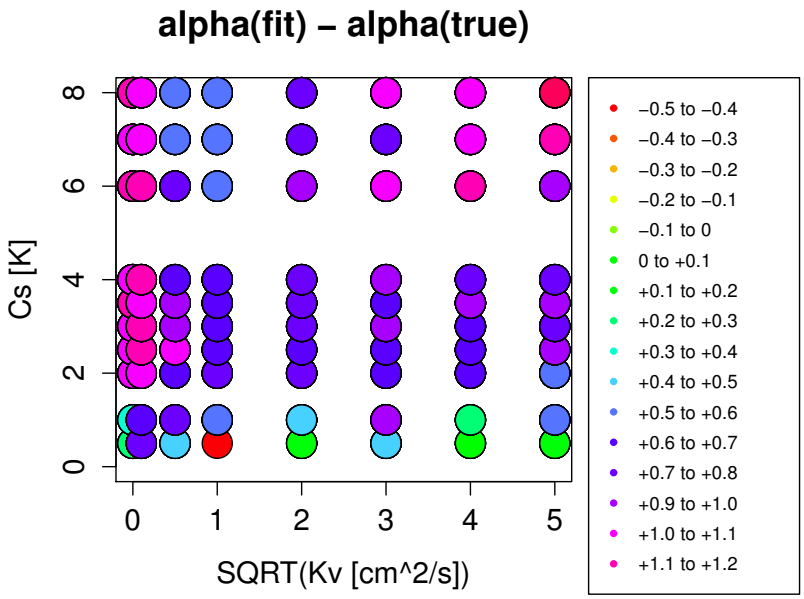

Figure 4.6. Aerosol forcing biases using DE for all runs where $F_{aer} = -0.5$ W/m^2.

4.2 MCMC Results

While the DE method produces a straightforward one-to-one translation of IGSM parameters to their DOECLIM counterparts, it does not provide information about parameter uncertainty. The MCMC calibration was used to obtain this uncertainty information, thereby expanding on the results obtained about the relationship between the parameters of the two models. Furthermore, it can be determined whether the best fit parameter values obtained through the DE method are consistent with the most likely values found by the MCMC method.

4.2.1 Time series fit for a typical model run

The MCMC calibration method was employed for the same "typical" model run from above, with $Cs = 3$ K, $Kv = 4$ cm^2/s and $F_{aer} = -0.5$ W/m^2. The MCMC method returns a sample from the joint distribution of parameter values. From this sample, the means, medians and modes of the marginal distributions of each individual parameter can be estimated. The DOECLIM simulations using the modes from the marginal distributions of S, κ and α are shown for surface and ocean temperatures in Figure 4.7, along with the results from the DE calibration for comparison. The DE best fit results and MCMC mode results are similar, indicating that the DE method is a suitable faster alternative to MCMC optimization for this problem, though it does not provide the uncertainty information obtained from the MCMC results.

4.2.2 Biases between IGSM and DOECLIM parameters

The MCMC method was then performed for each run with $F_{aer} = -0.5$ W/m^2. Figure 4.8 can then be compared with Figure 4.3 to see that similar bias patterns occur with both calibration methods. These biases are a result of translating between the two models, in that they occur regardless of which calibration method

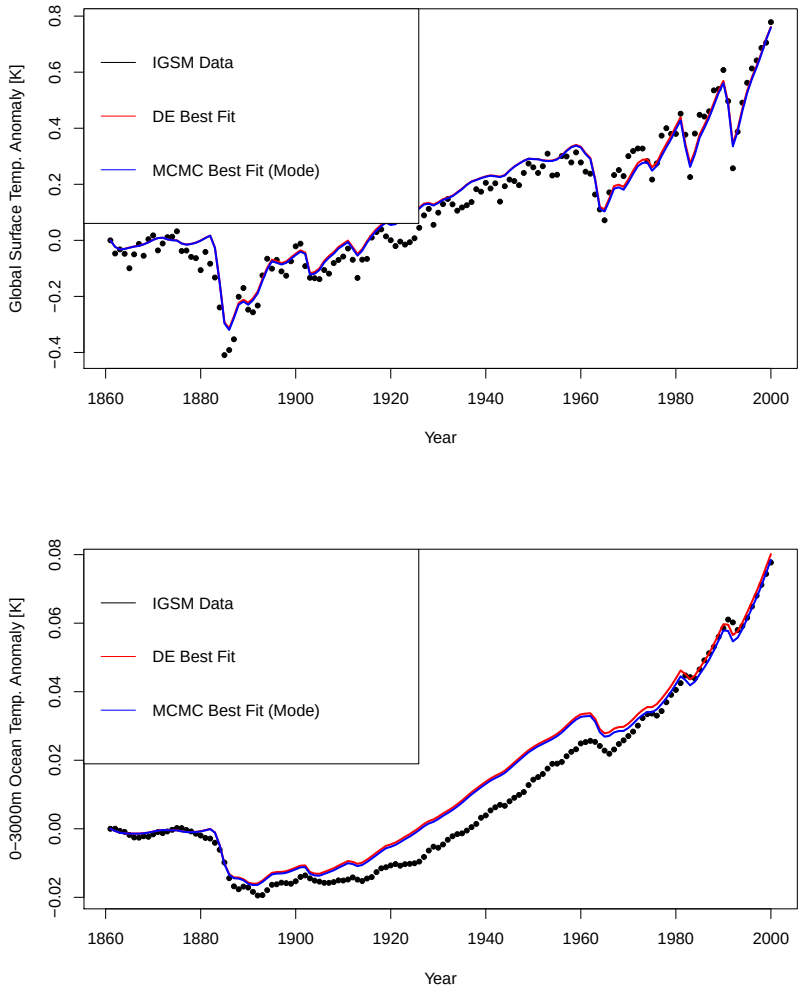

Figure 4.7. IGSM output data (black dots) plotted with output from DOECLIM with DE best fit parameters (red lines) and MCMC modes from the marginal distribution of each parameter (blue lines). IGSM parameter values: $Cs = 3$ K, $Kv = 4$ cm^2/s, $F_{aer} = -0.5$ W/m^2. DE best fit: $S = 2.61$ K, $\kappa = 1.99$ cm^2/s and $\alpha = 1.65$. MCMC modes from marginal distributions: $S = 2.70$ K, $\kappa = 1.93$ cm^2/s and $\alpha = 1.71$.

was used. Figure 4.9 shows just the runs with $F_{aer} = -0.5$ W/m^2 and $Kv = 4$ cm^2/s, with median estimates and 95% confidence intervals indicated. The confidence intervals become wider with increased climate sensitivity.

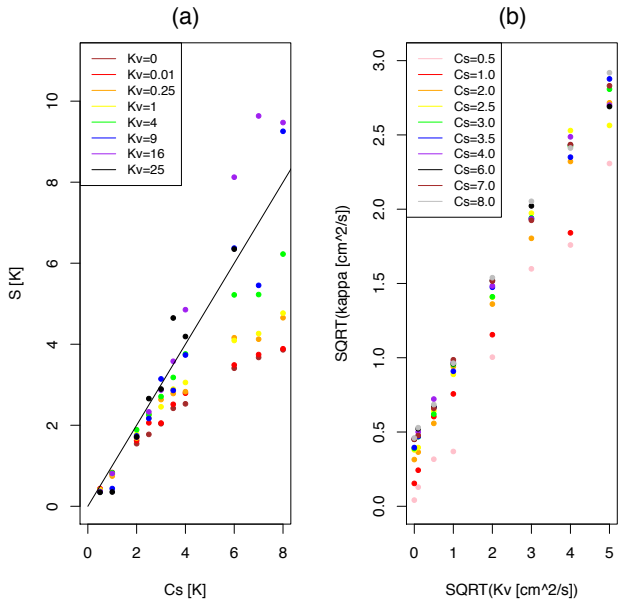

Figure 4.8. Comparison between IGSM parameters (horizontal axis) and MCMC median estimates using DOECLIM (vertical axis) for all runs where $F_{aer} = -0.5$ W/m^2. Climate sensitivity comparison is show in (a), while vertical ocean diffusivity comparison is shown in (b).

4.2.3 Marginal distributions

The marginal probability distributions for each individual parameter value are shown in Figure 4.10. The modes of the marginal distributions are close to the best fit values estimated by the DE method.

4.2.4 Bivariate histograms

I also constructed bivariate histograms of several parameter pairs. These plots, displayed in Figure 4.11, show how the frequency of parameter combinations are distributed in two-dimensional parameter space. This visualization is helpful due to the nature of calibrating multiple model parameters. One combination of parameters might be equally as good at matching a set of observations as another, slightly different combination of parameters. For instance, increased surface warm-

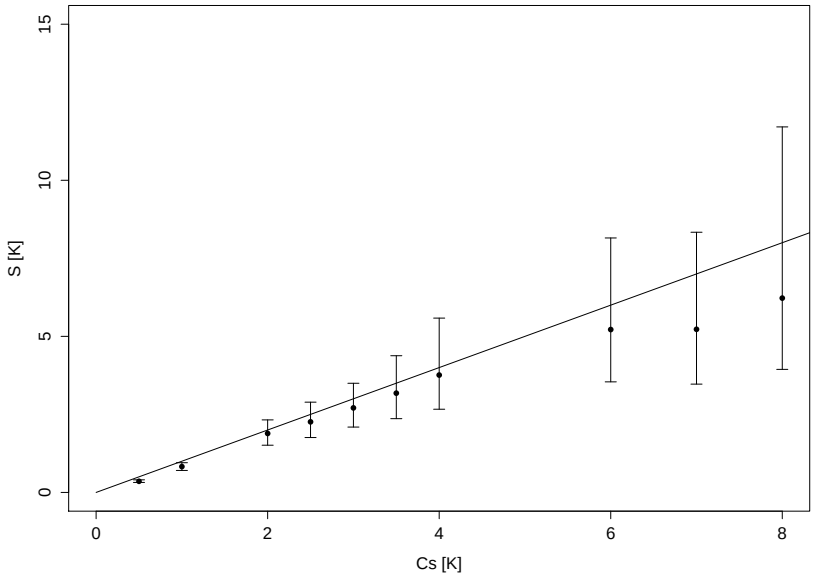

Figure 4.9. Comparison between Cs (IGSM) and S (DOECLIM) estimated using MCMC for all IGSM runs with $Kv = 4\ \mathrm{cm}^2/\mathrm{s}$ and $F_{aer} = -0.5\ \mathrm{W/m}^2$. Median estimates are plotted along with 95% confidence intervals.

ing due to a higher climate sensitivity value can compensate for decreased warming due to larger ocean heat uptake, and the model produces similar surface temperature output as it would have if climate sensitivity and ocean heat uptake were both lower.

4.2.5 Fixing κ and α

I performed another test with the standard model run data, where α in the DOECLIM model was set to a constant value of 0.943 (to match the aerosol forcing of the IGSM run) and κ was systematically set to a series of constant values, while climate sensitivity was allowed to be estimated by the calibration algorithm at each of these different κ values. This test was repeated, comparing the response of the two models using only surface temperature data, only ocean temperature data and both sets of data. This procedure allows one to test whether the MCMC esti-

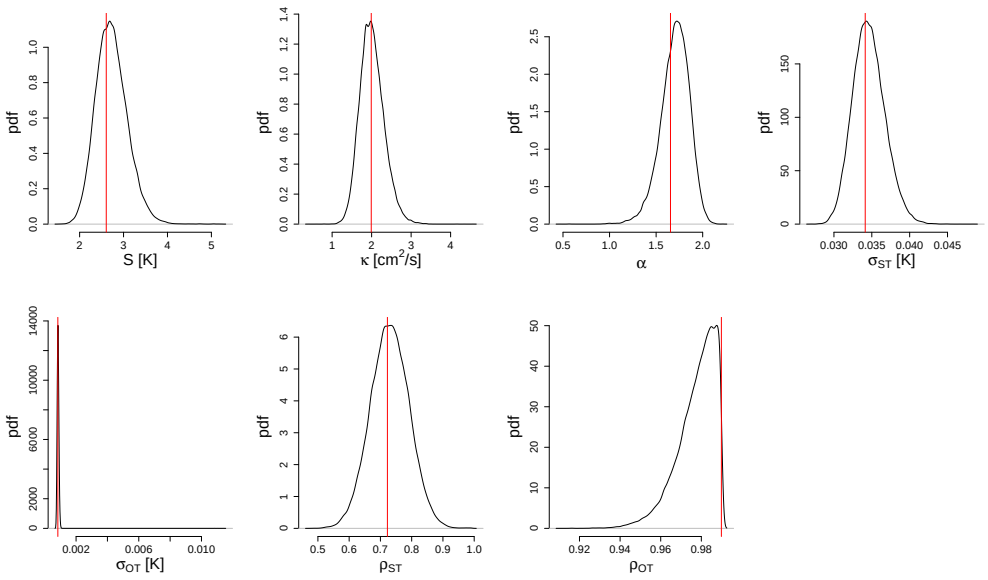

Figure 4.10. Marginal distributions for all parameters estimated by MCMC. DE best fit results shown in red.

mation algorithm as applied to the DOECLIM model is performing in an expected manner. When increasingly higher κ values are used for the DOECLIM model, the estimated climate sensitivity must compensate to match the same data, as seen in Figure 4.12.

As in Figure 4.9, the uncertainty in climate sensitivity estimates increases when the climate sensitivity estimate is larger. This increase is consistent with what one would expect, as an increased climate sensitivity value leads to increased internal variability in the model, and this results in a larger uncertainty range. Results also differ depending on whether I used just surface temperature, just ocean temperature or both types of temperature data in the calibration. For the test using just surface temperature, estimated climate sensitivity must increase as κ increases in order to match the same output data. When I used just ocean temperature, the estimated climate sensitivity must decrease. When both data sets are used in the calibration, the tendency for estimated climate sensitivity to decrease dominates.

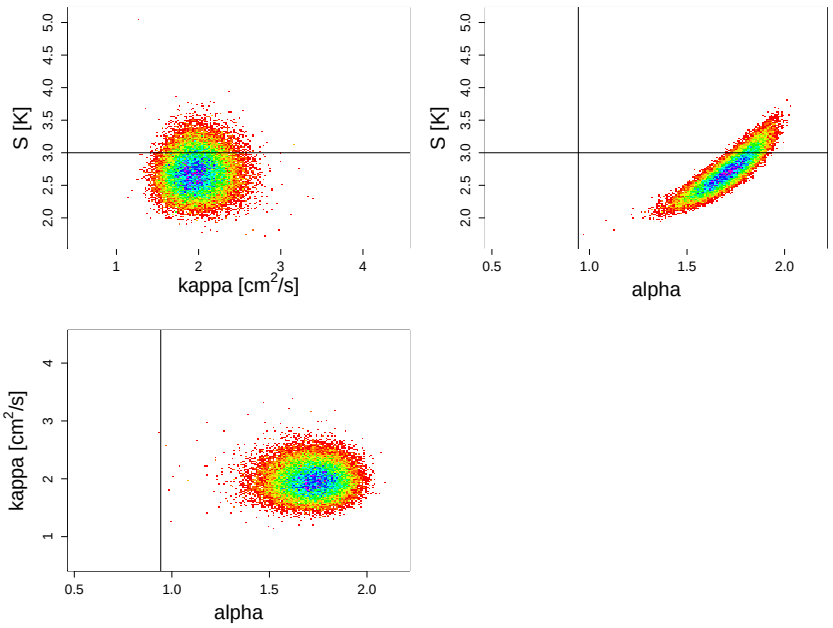

Figure 4.11. Two-dimensional histograms showing the frequency of parameter pairs in the final Markov chain produced using the MCMC method. IGSM values are shown with black lines.

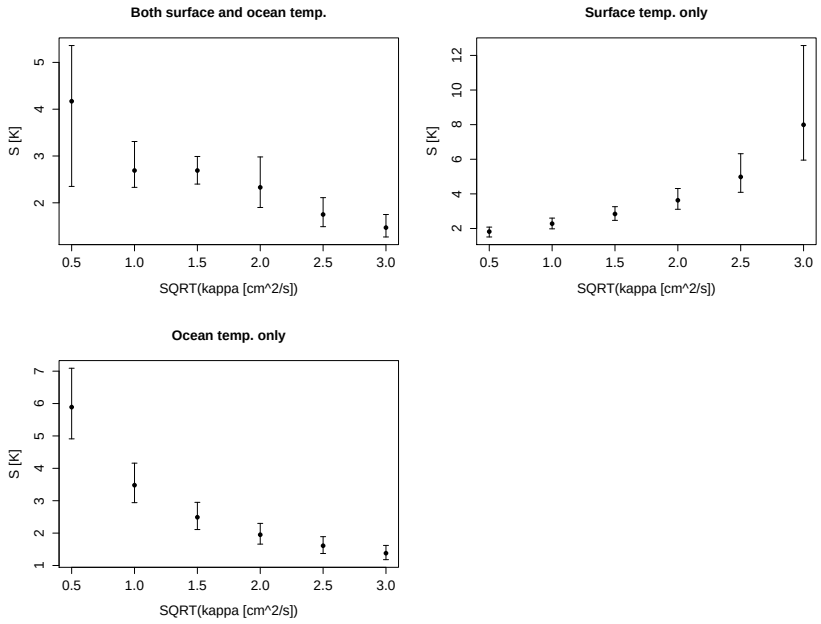

Figure 4.12. Estimated climate sensitivities using MCMC for systematically increased fixed kappa, fit to data from standard run where $Cs = 3$ K, $Kv = 4$ cm^2/s, $F_{aer} = -0.5$ W/m^2. Median estimates are plotted along with 95% confidence intervals.

4.3 Transient Climate Response Comparison

DOECLIM can also be compared to the IGSM by comparing their responses to the same external forcing across a selection of input parameter combinations. By plotting the model response in parameter space for the surface and the ocean, it is possible to see if the two models perform similarly when parameters move further away from the "standard" run. These response surface plots can also be used as an alternate method to determine corresponding parameters between the two models; one can note the IGSM's surface temperature and ocean heat content response to a certain set of parameters, find that same response on the matching DOECLIM figure and identify the parameters that produce the equivalent response.

Transient climate response (TCR) is the global surface temperature change in response to a particular transient scenario. Suppose CO_2 concentration is increased by 1% per year. TCR is then defined as the global surface temperature change at the time of CO_2 doubling. To estimate TCR for the DOECLIM model, I forced DOECLIM with a 1% increase in CO_2 concentration per year, then calculated the change in surface temperature and ocean heat content at the time of CO_2 doubling for pairs of ocean thermal diffusivity and climate sensitivity. This provides estimates for TCR on a grid sample for the model parameter space S-κ. S values range from 0.5 to 5 K, while $\sqrt{\kappa}$ values range from 0.1 to 4 W/m^2 (DOECLIM does not return output values when κ=0). The DOECLIM model response is shown in Figure 4.13. The same forcing scenario, applied to the version of the IGSM used in Forest et al. [2008], produces the transient climate response and sea level rise due to thermal expansion shown in Figure 4.14. On the $Cs - Kv$ parameter space shown, Cs ranges from 0.5 to 5 K and $\sqrt{Kv}$ ranges from 0 to 4 cm^2/s. Ocean heat content and sea level rise due to thermal expansion, though two different quantities, are proportionally related and the shapes of the contours may be compared directly. A distortion in the response surface plots is observed when comparing the results for the two models. One strong example of this distortion is seen in

the $TCR = 1.5K$ contour, which curves sharply upward in the DOECLIM figure and remains a relatively straight line in the IGSM figure. Along this contour the model parameters needed to produce the same climate response differ noticeably between the two models. This also highlights the structural differences between the two models.

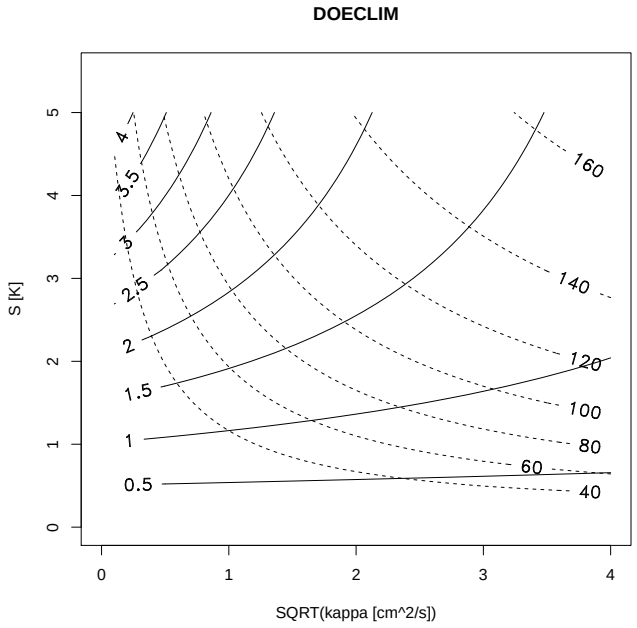

Figure 4.13. Transient Climate Response [K] (solid lines) and change in ocean heat content [10^{22}J] at time of CO_2 doubling (dotted lines) for DOECLIM model.

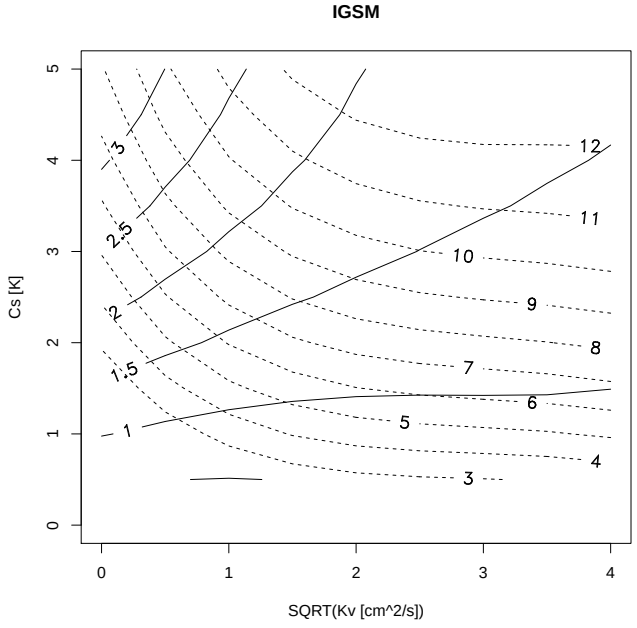

Figure 4.14. Transient Climate Response [K] (solid lines) and change in sea level rise due to thermal expansion [cm] at time of CO_2 doubling (dotted lines) for IGSM model.

CHAPTER 5

Conclusion

5.1 Conclusions

In this study, I set out to calibrate the DOECLIM energy balance model for use as an emulator for the MIT IGSM by finding a mapping relationship that would allow one to translate between the parameter spaces of the two models. I found that this exercise does not provide a simple mapping from one set of parameters to the other. Rather, there are noticeable biases in the estimated DOECLIM parameters, and the biases in a single parameter vary in magnitude depending on the other two parameters of the particular IGSM run. Of particular note are the biases in climate sensitivity. Almost all of the DOECLIM calibrations returned estimated climate sensitivity values lower than the "real" values (those of the corresponding IGSM run) when using both the DE and MCMC estimation methods. The 95% confidence intervals for estimated climate sensitivity from the MCMC results grow wider as

the IGSM run climate sensitivity increases. The value of Kv also affects climate sensitivity estimates, as IGSM runs with lower Kv show a stronger negative bias in DOECLIM climate sensitivity estimates. These results highlight the potential difficulties in using this kind of simple model for emulation purposes over a variety of parameter settings, particularly for parameters with values far from the most likely values for real world applications. Additionally, one may note the importance that the choice of noise model has on the parameter estimations. An AR(1) noise model was chosen to capture differences between the two models, but I found that when I used an AR(0) noise model, the estimated climate sensitivity was over a degree K larger (see Figure 4.2). The choice of an AR(0) noise model causes the ocean temperature data between the two models to fit more closely, not allowing for potential discrepancy between the ocean models. The importance of the noise model would also be relevant when climate models are calibrated with observational data to estimate the real world climate sensitivity, where the AR(1) process is often used to approximate observational error, model error and natural variability (see, for example, Urban and Keller [2010]). A different noise model can produce very different results.

There are also biases in the aerosol forcing estimation results. The estimated DOECLIM α values are generally higher than the expected scale factors, that is, the scale factors that would produce aerosol forcing time series matching those used in the corresponding IGSM runs. Further investigation is required in order to determine whether a temporal or spatially dependent pattern for aerosol forcing scaling would be more appropriate than the global scale factor used in this study.

While climate sensitivity and aerosol forcing parameters can be directly compared between the two models, the vertical ocean diffusivity parameters are defined differently and therefore biases cannot be estimated directly for these parameters. One could potentially compare the timescale responses in ocean temperature between the two models to investigate this further. The ocean temperature data

fits show discrepancies that the surface temperature data fits do not. The DOE-CLIM best fits for ocean temperature show ocean temperature values larger than the IGSM values for the first half of the 20th century. The AR(0) noise model allows for a slightly closer fit, though the best fit is still larger than the IGSM time series over the decades from 1930-1970. As there is a time lag inherent in ocean heat changes, investigating the timescale responses of both models would help to understand these differences.

Finally, this study involved a comparison between the Differential Evolution and Markov Chain Monte Carlo methods for parameter optimization. I found that the DE method is able to pick out the most likely values for each parameter, and is therefore a fast alternative to the MCMC method, though the additional information obtained from the MCMC method gave a greater understanding of the probability distributions for each parameter and the relationships among the model parameters based on their joint distributions.

5.2 Future Work

It would be pertinent to investigate whether the patterns observed with respect to estimated biases in climate sensitivity also occur when using other models (other EBM-to-EMIC calibrations, AOGCM-to-EMIC, AOGCM-to-EBM and so on). This would demonstrate whether these parameter estimation effects apply broadly to the problem of comparing parameters between models, or whether they occur due to the particular differences between the two models used in this study. It would also be useful to determine what effect the ocean model of the simpler model (purely vertically diffusive in this case) has on the parameter estimations, by performing the same study with an ocean model that contains meridional overturning circulation, spatial resolution or higher complexity physics. While this thesis studied the 20th century response to historical forcing values, it remains an open

question how results would vary for other forcing scenarios. Such scenarios could include the response to a 1% increase in CO_2 per year, as well as future projections based on greenhouse gas emissions scenarios. All of these tests, if they respond similarly in terms of parameter biases, would strengthen the need to investigate this issue for future climate model emulation studies.

APPENDIX A

Positive Control Test

An additional test was done to confirm that the model calibration process was performing properly. This test was done to show that the biases seen in the parameter estimation results are not due to the calibration itself. First, I ran the DOECLIM model with an S value of 5 K, a κ value of 2 cm^2/s and an α value of 1. I added Gaussian noise to the surface temperature and ocean heat content output data, with a standard deviation of 0.01 for surface temperature and a standard deviation of 1 for ocean heat content. I then carried out the MCMC calibration process with this output data, in order to ensure that the calibration is able to return the same parameters that were originally used to run the model. The resulting distribution shows results close to what was expected, demonstrating that we can recover DOECLIM's own results through model calibration.

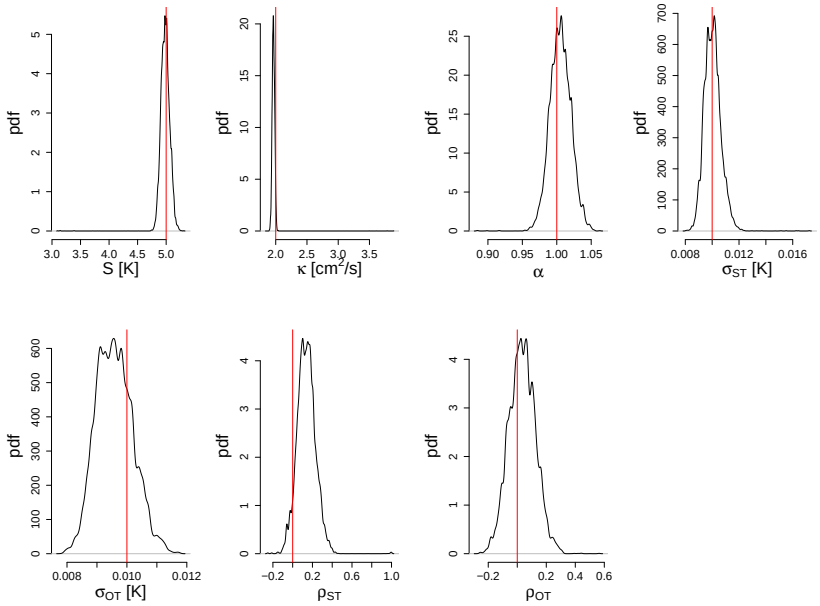

Figure A.1. Marginal distributions for parameters calibrated for the positive control test. Red lines show expected results, that is, parameters originally used to run the DOECLIM model.

APPENDIX B

Additional Tests on Statistical Estimates

B.1 Distribution of Residuals

I plotted histograms of the whitened residuals, w_i, found by comparing the standard run IGSM data with the DOECLIM output when run with the best fit parameters found by DE (Figures B.1 and B.2), and removing the autocorrelation using the autocorrelation parameters also found by DE. The residuals for each time series (surface temperature and ocean temperature) are expected to follow a normal distribution with mean zero and with a standard deviation equivalent to the best fit standard deviation found by the calibration process.

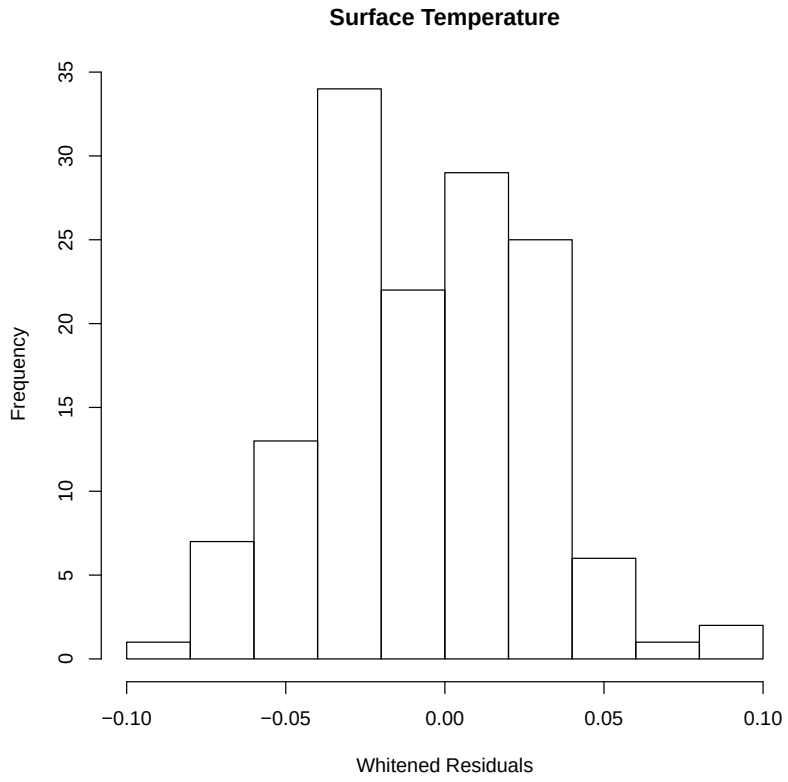

Figure B.1. Histogram showing frequency of whitened residual values for surface temperature time series, when DE is used to fit DOECLIM to the standard IGSM run.

B.2 Autocorrelation Function

Finally, I plotted the autocorrelation function for the whitened residuals, w_i (Figures B.3 and B.4). This autocorrelation function shows that the strong autocorrelation was removed by whitening the residuals, and that the AR(1) process is an appropriate choice of noise model for this study.

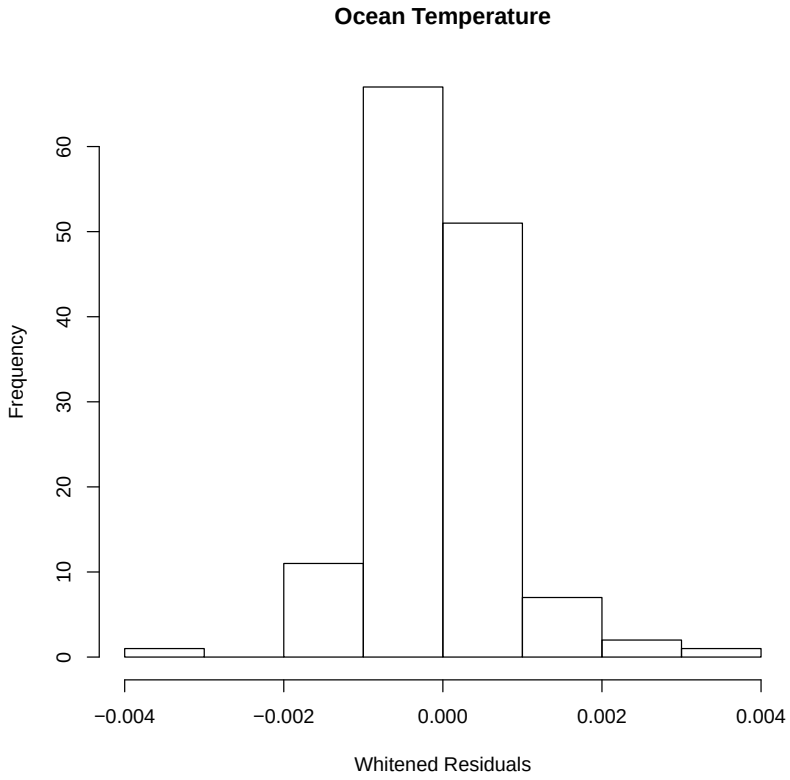

Figure B.2. Histogram showing frequency of whitened residual values for ocean temperature time series, when DE is used to fit DOECLIM to the standard IGSM run.

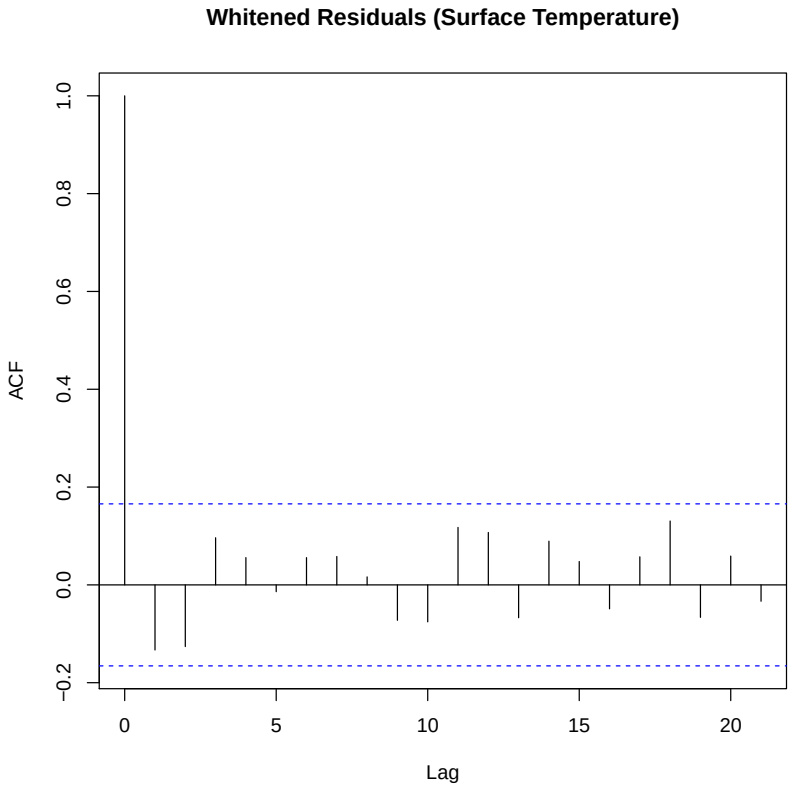

Figure B.3. Autocorrelation function of whitened residuals for surface temperature time series.

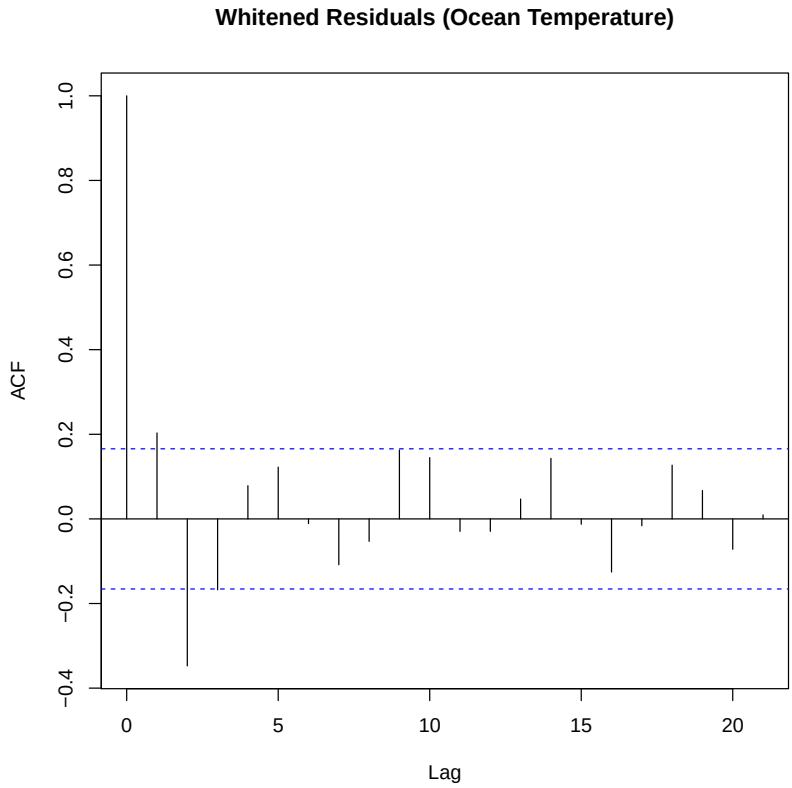

Figure B.4. Autocorrelation function of whitened residuals for ocean temperature time series.

Bibliography

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

N. G. Andronova and M. E. Schlesinger. Objective estimation of the probability density function for climate sensitivity. *J. Geophys. Res.*, 106(D19):22,605–22,612, 2001.

T. Bayes. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society*, 53:370–418, 1763.

N.L. Bindoff, P.A. Stott, K.M. AchutaRao, M.R. Allen, N. Gillett, D. Gutzler, K. Hansingo, G. Hegerl, Y. Hu, S. Jain, I.I. Mokhov, J. Overland, J. Perlwitz, R. Sebbari, and X. Zhang. Detection and Attribution of Climate Change: from Global to Regional. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley, editors, *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK, 2013.

C. E. Forest, P. H. Stone, A. P. Sokolov, M. R. Allen, and M. D. Webster. Quantifying uncertainties in climate system properties with the use of recent climate observations. *Science*, 295:113–117, 2002.

C. E. Forest, P. H. Stone, and A. P. Sokolov. Estimated PDFs of climate system properties including natural and anthropogenic forcings. *Geophys. Res. Let.*, 33 (L01705):doi:10.1029/2005GL023977, 2006.

C. E. Forest, P. H. Stone, and A. P. Sokolov. Constraining climate model parameters from observed 20th century changes. *Tellus*, 60A:911–920, 2008.

C. J. Geyer and L. T. Johnson. *Package 'mcmc'*, 2014. URL `http://CRAN.R-project.org/web/packages/mcmc`.

J. Hansen, G. Russell, D. Rind, P. Stone, A. Lacis, and co authors. Efficient three-dimensional global models for climate studies: models i and ii. *Mon. Weath. Rev.*, 111:609–662, 1983.

W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

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

E. Kriegler. *Imprecise probability analysis for integrated assessment of climate change*. PhD thesis, University of Potsdam, 2005.

A. G. Libardoni and C. E. Forest. Sensitivity of distributions of climate system properties to the surface temperature dataset. *Geophys. Res. Lett.*, 38(12): L22705, 2011.

K. McGuffie and A. Henderson-Sellers. *A Climate Modeling Primer, 3rd Edition*. Wiley, 2013.

G. A. Meehl, G. J. Boer, C. Covey, M. Latif, and R. J. Stouffer. The coupled model intercomparison project (cmip). *Bull. American Meteorological Society*, 81(2):313–321, 2000.

M. Meinshausen, S. C. B. Raper, and T. M. L. Wigley. Emulating coupled atmosphere-ocean and carbon cycle models with a simpler model, MAGICC6 – Part 1: Model description and calibration. *Atmos. Chem. Phys.*, 11:1417–1456, 2011.

K. M. Mullen, D. Ardia, D. L. Gil, D. Windover, and J. Cline. Deoptim: An R package for global optimization by differential evolution. *Journal of Statistical Software*, 40(6):1–26, 2011.

W. Nordhaus. *A Question of Balance: Weighing the Options on Global Warming Policies*. Yale University Press, 2008.

D. Olivie and N. Stuber. Emulating AOGCM results using simple climate models. *Clim. Dyn.*, pages 1257–1287, 2010.

R. Olson, R. Sriver, M. Goes, N. M. Urban, H. D. Matthews, M. Haran, and K. Keller. A climate sensitivity estimate using bayesian fusion of instrumental observations and an Earth System model. *J. Geophys. Res.*, 117, 2012.

K. V. Price, R. M. Storn, and Jouni A. Lampinen. *Differential Evolution - A Practical Approach to Global Optimization*. Natural Computing. Springer-Verlag, January 2006. ISBN 540209506.

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

A. P. Sokolov, C. A. Schlosser, S. Dutkiewicz, S. Paltsev, D. W. Kicklighter, and co authors. The MIT Integrated Global System Model (IGSM) Version 2: Model Description and Baseline Evaluation, MIT JP Report 124. Technical report, MIT, Joint Program on the Science and Policy of Global Change, Room E40-428, 77 Massachusetts Ave., Cambridge, MA 02139, 2005.

K. Tanaka, E. Kriegler, T. Bruckner, G. Hooss, W. Knorr, and co authors. Aggregated carbon cycle, atmospheric chemistry, and climate model (ACC2). Description of the forward and inverse models (Reports on Earth System Science No. 40). Technical report, Hamburg, Germany, 2007.

N. M. Urban and K. Keller. Probabilistic hindcasts and projections of the coupled climate, carbon cycle and Atlantic meridional overturning circulation system: a Bayesian fusion of century-scale observations with a simple model. *Tellus*, 62A: 737–750, 2010.

N. M. Urban, P. B. Holden, N. R. Edwards, R. L. Sriver, and K. Keller. Historical and future learning about climate sensitivity. *Geophys. Res. Lett.*, 2014.