WHAT IS THE SKILL OF CLIMATE PARAMETER ESTIMATION METHODS? A CASE STUDY WITH GLOBAL AVERAGE OBSERVATIONAL CONSTRAINTS

A Thesis in
Geosciences
by
Roman Tonkonojenkov

© 2010 Roman Tonkonojenkov

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

August 2010
The thesis of Roman Tonkonovenkov was reviewed and approved* by the following:

Klaus Keller  
Associate Professor of Geosciences  
Thesis Advisor

Raymond G. Najjar, Jr.  
Associate Professor, Meteorology/Geosciences

James F. Kasting  
Professor of Geosciences

Chris J. Marone  
Professor of Geosciences  
Associate Head for Graduate Programs in Geosciences

*Signatures are on file in the Graduate School.
Abstract

Future climate model projections are deeply uncertain. One key driver of this uncertainty is the uncertainty about values of key model parameters, such as climate sensitivity. Recent studies have used ensembles of model runs together with observations to estimate these key parameters. In these studies, Markov Chain Monte Carlo (MCMC) methods are often employed to obtain posterior probability distributions for the parameters. Despite ubiquitous use of such methods, their skill at recovering known parameter values has not been thoroughly evaluated.

This study quantifies the skill of an MCMC parameter estimation method to recover true parameter values using pseudo-observations generated from an University of Victoria Earth System Climate Model (UVic ESCM). Specifically, the work addresses three key questions. First, what is the effect of reducing the combined model and observational error on the skill of the method? Second, what is the skill of different pseudo-observations to constrain model parameter estimates? Third, what is the effect of random realizations of the combined model and observational error on the results of the parameter estimation?

I first run an ensemble of UVic ESCM model runs spanning the last two centuries. I vary the parameters of climate sensitivity, background vertical ocean diffusivity and the strength of effects of anthropogenic aerosols between the ensemble members. I then implement a simple MCMC method to estimate the parameters using global observations of temperature and upper ocean heat content. The inversion accounts for uncertainty in the statistical properties of the potentially correlated model-data residuals, and reduces biases due to sparse sampling of the parameter space using emulation. I perform a set of perfect model experiments using pseudo-observations of globally average temperature and ocean heat content, derived from the UVic model using various assumptions about the observational and model error.

I show that at current estimates of combined model and observational error the results of parameter estimation hinge critically on random realizations of the
combined error process, but that the skill of the method increases rapidly if the combined error decreases. Using both temperature and ocean heat uptake observations improves the skill of the method compared to cases where only individual observations are used, except for the case of background vertical ocean diffusivity at current combined error estimates. Implications of the results for parameter estimation work are discussed and strategies for future research are outlined.

In addition, I compare the probabilistic UVic ESCM hindcasts of global average near-surface temperature with those from more complex General Circulation Models (GCMs). I show that a well calibrated intermediate complexity model such as UVic ESCM can perform comparably to GCMs in terms of skill at reproducing global mean historical temperature observations.
# Table of Contents

List of Figures vii

List of Tables x

Preface xi

Acknowledgments xii

Chapter 1
Introduction 1
1.1 Introduction ...................................................... 1
1.2 Feedbacks in the Earth System ................................. 2
1.3 Quantifying Climate Sensitivity ............................... 3
1.4 Recent Papers on Climate Parameter Estimation ............. 5
1.5 Current Challenges and Ways Forward ....................... 7

Chapter 2
What is the Skill of Climate Parameter Estimation Methods?
A Case Study with Global Average Observational Constraints 14
2.1 Abstract .................................................................... 16
2.2 Introduction ............................................................ 17
2.3 Methods ................................................................. 18
  2.3.1 Model Description .............................................. 18
List of Figures

2.1 Climate sensitivity as a function of $f$ and $K_{bg}$ as diagnosed from the CO$_2$ doubling runs. The locations of the CO$_2$ doubling runs in the parameter space are specified by the black circles. The red tick marks correspond to the $f$ and $K_{bg}$ values where the hindcast ensemble model runs (described in section 2.3.3.1) were run. These runs were run on a quazi-linear grid of the $f$ and $K_{bg}$ model parameters. Specifically, the $f$ values used in this ensemble have a slight dependence on $K_{bg}$, which results in the clustering of the red tick marks. .............................................. 20

2.2 The relationship between the index on the climate sensitivity grid (CS Index) where the ensemble runs were performed, and the climate sensitivity (CS) itself. The slight spread of climate sensitivity values at the same CS Index is due to the weak dependence of the CS on $K_{bg}$ at the same CS Index on the grid. ......................... 22
2.3 UVic ESCM model hindcasts (grey lines), and corresponding observations (red lines or markers): (a) global average atmospheric surface temperature [K], and observations of global average near surface/ocean surface temperature anomalies from the HadCRUT3 dataset (Brohan et al., 2006) with arbitrary offset added; (b) ocean heat content anomaly in the 0-700m layer [J], and observations from Domingues et al. (2008); black lines delineate observational 1σ error intervals. (c) Ocean carbon flux [GtC yr$^{-1}$] and decadal observational constraints (1.6 GtC yr$^{-1}$ for the 1980s and 2.0 GtC yr$^{-1}$ for the 1990s) from McNeil et al. (2003); (d) atmospheric CO$_2$ concentrations; the observations are from Mauna Loa observatory (Keeling et al., 2009) and Law Dome ice core (Etheridge et al., 1996); (e) Atlantic meridional overturning circulation (AMOC) intensity [m$^3$ s$^{-1}$], with observations from Bryden et al. (2005) and Cunningham et al. (2007), and their error estimates from Kanzow et al. (2007) and Lumpkin and Speer (2007).

2.4 Taylor diagram representing the skill of UVic ESCM ensemble runs (red dots) and GCMs from CMIP-3 multi-model dataset (blue crosses) at simulating global mean near surface temperature for period 1900-1999 (Brohan et al., 2006). The temperatures were used as anomalies with respect to mean of period 1900-1949. We used the following models from the CMIP3 multi-model dataset (Meehl et al., 2007): BCCR–BCM2.0, CGCM3.1(T47), CSIRO–Mk3.0, CSIRO–Mk3.5, GISS–ER, MIROC3.2(medres), ECHO–G, ECHAM5/MPI–OM and MRI–CGCM2.3.2. We neglect the difference between near-surface atmospheric temperatures and ocean surface temperatures in this analysis. Dotted lines are isolines of standard deviations, semi-circle dashed lines are isolines of correlation coefficient, and straight dashed – of centered root mean square difference.
2.5 Output of UVic ensemble and GCMs from the CMIP-3 dataset together with observations of temperature anomalies with respect to the 1900-1949 period. See caption of Figure 2.4 for the list of models analyzed. Observations have yearly resolution; they are plotted as lines for visual clarity. .......................................................... 30

2.6 Posterior pdfs for model parameter indices obtained from perfect model experiments using temperature pseudo-observations only. Top row: \((\sigma, \rho)/10\) experiment, middle row: \((\sigma, \rho)/3\) experiment, bottom row: \((\sigma, \rho)\) experiments. Dashed lines represent the true parameter values. .......................................................... 32

2.7 Posterior pdfs for model parameter indices obtained from perfect model experiments using ocean heat content pseudo-observations only. Top row: \((\sigma, \rho)/10\) experiment, middle row: \((\sigma, \rho)/3\) experiment, bottom row: \((\sigma, \rho)\) experiments. Dashed lines represent the true parameter values. .......................................................... 33

2.8 Posterior pdfs for model parameter indices obtained from perfect model experiments using both temperature and ocean heat content pseudo-observations together. Top row: \((\sigma, \rho)/10\) experiment, middle row: \((\sigma, \rho)/3\) experiment, bottom row: \((\sigma, \rho)\) experiments. Dashed lines represent the true parameter values. .......................................................... 35

2.9 Skill \(S\) of the MCMC method for the different parameters under different assumptions about the magnitude of combined model and observational error. Left: for \(K_{bg}\) index; middle: for \(CS\) Index; right: for \(A_{SCL}\) index. .......................................................... 36
List of Tables

2.1 Prior ranges for statistical parameters used in the assimilation. $b_T$ refers to the bias parameter for temperature data, and $b_{OHC}$ refers to the bias parameter for ocean heat content data. The ranges for $\sigma$ and $\rho$ are the same for both observations. 26

2.2 Mean posterior estimates for the statistical parameters $\sigma$ and $\rho$, for each observation type, as obtained from the assimilation experiment which uses both types of real observations together. 27

2.3 Method skill $S$ for various positive control experiments. Please refer to the text for definition of $S$. For the $(\sigma, \rho)$ experiments each skill represents an average for five realizations of the random error process. 31

B.1 Sample uncertainty of the method $U$ for different $(\sigma, \rho)$ experiments for each of the parameters. As described in text, $U$ is defined as sample variance of the mode of the posterior pdf for the five different realizations of the error process. 50
Preface

This thesis includes a manuscript projected to be submitted to a peer-reviewed journal. Roman Tonkonogojenkov, the candidate for Master of Science, is the first author, was leading the project and contributed most of the work. The thesis advisor, Klaus Keller, is the last author. Roman Tonkonogojenkov has implemented most model changes, performed the ensemble of transient model simulations, implemented the assimilation method and the perfect model experiments, and analyzed the results. Marlos Goes has performed equilibration and double CO$_2$ model simulations and helped to implement model changes. Nathan Urban provided the assimilation method, as well as essential insights. Damon Matthews helped implement some model changes and shared knowledge about the model. Klaus Keller has provided logistical support and general oversight of this study, as well as insightful comments and stimulating discussions. All authors had inputs into the design of the study and into the analysis.
Acknowledgments

I thank the UVic model developers and in particular Michael Eby for providing the model and for helpful discussions. I thank the scientists who gathered and compiled observations used in this study. I acknowledge the modeling groups, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP’s Working Group on Coupled Modelling (WGCM) for their roles in making available the WCRP CMIP3 multi-model dataset. Support of this dataset is provided by the Office of Science, U.S. Department of Energy.

I thank Murali Haran, Sham Bhat, David Pollard, and Andreas Schmittner for thought-provoking discussions and essential insights. I acknowledge financial support from NSF and USGS. Of course, all errors and opinions in the manuscript are that of the authors.

I am greatly indebted to my advisor, Klaus Keller, for his patience, encouragement and advice. I thank my committee members for the guidance of the study. I am very grateful to my family members as well as friends for their support and encouragement.
Chapter 1

Introduction

1.1 Introduction

Human emissions of carbon dioxide are influencing global atmospheric and ocean temperatures, causing global sea levels to rise, glaciers and snowcover to melt, sea ice to shrink, and aspects of extreme weather to change (IPCC, 2007). Atmospheric carbon dioxide has risen from about 285 ppm to about 380 ppm between years 1854 and 2004 (Keeling et al., 2009; Etheridge et al., 1996). Atmospheric surface temperatures have risen by approximately 0.8°C during the last century (Brohan et al., 2006), and are projected to rise further.

The way by which carbon dioxide affects the atmospheric radiative balance is well known. Carbon dioxide molecules allow the downward shortwave radiation from the Sun to penetrate through the Earth’s atmosphere, keeping the planet’s surface warm; however they impede transfer of outgoing longwave radiation. They do so by absorbing some of this radiation and then emitting a part of it back to the surface, thereby causing an increase in surface temperatures. Radiative effects of carbon dioxide are fairly well known, yet future predictions of atmospheric temperatures are quite uncertain. What causes this uncertainty are the feedbacks in the Earth system, whereby an initial perturbation in atmospheric temperature can be either enhanced or diminished through other processes that depend on the temperature. A variety of feedback processes operate within the Earth system. The next sections describes several of them in more detail.
1.2 Feedbacks in the Earth System

A feedback loop is a collection of couplings within the Earth System that results in an initial change (or perturbation) in one climate component bringing about further change in itself, either diminishing or enhancing the initial change. One popular example of a climate feedback is the ice-albedo feedback. Atmospheric surface temperature is negatively coupled to sea ice extent. Thus, a rise in atmospheric temperature causes sea ice cover to shrink. Sea ice cover is negatively coupled to the amount of radiation absorbed by the Earth surface through albedo changes. Therefore this decrease in ice causes increased absorption of solar incoming radiation by the surface. This closes the feedback loop and causes a further rise in the atmospheric surface temperature.

Once the anthropogenic greenhouse gases are introduced into the atmosphere, atmospheric temperature increases due the increased greenhouse effect. In the case with no feedbacks, this perturbation of the Earth radiative balance would lead to a net flux of energy into the Earth system, thereby eventually resulting in infinite warming. Fortunately, there are considerable negative feedbacks in the Earth system, which would keep the change in Earth temperature finite. In particular, the major negative feedback operating on temperature involves the Stefan-Boltzmann law. According to this law the amount of energy emitted by a black (or grey) body is proportional to the $4^{th}$ power of temperature. Thus, as the Earth surface temperature increases, emission of the outgoing longwave radiation also increases. This has a cooling effect on the Earth’s temperature and offsets the original warming.

Many of the other feedbacks, including positive feedbacks, operate within the Earth system. For example, as temperatures warm, the microbial soil respiration increases. Since this process converts organic matter to carbon dioxide, it results in increasing atmospheric CO$_2$ stocks and further warms the atmosphere. This is a positive feedback. 'Carbon dioxide fertilization' is a potentially important negative carbon cycle feedback (Li et al., 2009). It operates as follows. An increase in temperature can cause Net Primary Productivity (NPP) to increase. This process would convert more carbon dioxide and store it in plant matter, thereby diminishing the original change in temperature. Yet another feedback involves water vapor. As temperature increases, the water content in the atmosphere increases.
Because water is an important greenhouse gas, this leads to further increases in atmospheric temperature.

Many other feedbacks involving temperature have been identified. They can involve cloud properties (i.e. Zhu et al. (2007)), lapse rate, Earth’s albedo, vegetation, ocean biology, etc. The feedbacks differ in the timescales on which they operate. This thesis is concerned with only those feedbacks which operate on decadal to century time scales.

If the only feedback in the Earth system was the Stefan-Boltzmann Law feedback, it would be possible to calculate the changes in temperature resulting from doubling of CO\textsubscript{2} with relatively high precision. However, the magnitude of the combined feedbacks is poorly known. For example, at least partly due to differences in simulated carbon cycle feedbacks, model projections disagree on whether the land will be a sink or a source of additional carbon to the atmosphere (Friedlingstein et al., 2006). Such uncertainties make the task of estimating future rise in temperature due to changes in CO\textsubscript{2} more complicated.

The concept of feedbacks is closely linked to the concept of climate sensitivity. Climate sensitivity linearly relates equilibrium change in near surface air temperature $\Delta T$ to radiative forcing $\Delta F$ according to:

$$\Delta T = \lambda \Delta F,$$

where $\lambda$ is climate sensitivity (Andronova et al., 2007). Climate sensitivity is a measure of the combined feedbacks in the Earth system. In this thesis, however, climate sensitivity ($CS$) refers to the equilibrium change in global near-surface temperature in response to the doubling of atmospheric CO\textsubscript{2} concentrations.

### 1.3 Quantifying Climate Sensitivity

The concept of climate sensitivity to carbon dioxide was introduced by Arrhenius (1896) in his landmark paper “On the influence of carbonic acid in the air upon the temperature of the ground”. In 1979, the Charney report established a range from 1.5 to 4.5 K for climate sensitivity (NAS, 1979). The ‘likely range’ given in the Intergovernmental Panel of Climate Change (IPCC’s) Fourth Assessment Report (AR4) is from 2 to 4 K (Solomon et al., 2007). Despite rapid advances in
understanding and modeling climate, little progress has been made in determining the value of climate sensitivity. For example, the standard deviation of climate sensitivities in General Circulation Models (GCMs) with mixed layer upper ocean, as reported in the IPCC Third Assessment Report (TAR) (Cubasch et al., 2001), increased from 0.78 to 0.92 between the IPCC Second Assessment Report (SAR) and the TAR. The range of possible climate sensitivities obtained from individual studies remains considerable and spans beyond the range given in the IPCC reports (Edwards et al., 2007; Knutti and Hegerl, 2008).

Various approaches have been used to estimate climate sensitivity. Manabe and Wetherald (1975) calculated a response of a General Circulation Model to doubling of CO$_2$ by comparing a doubled CO$_2$ run with a control run. Such an approach is deterministic as it ignores uncertainty in values of other important model parameters. Biased parameter values or incomplete parameterizations of all relevant processes can result in biased estimates of climate sensitivity. An improvement on this approach is to make use of climate observations and an ensemble of model simulations, where key climate parameters are systematically varied, to provide constraints on these model parameters. This approach can provide probability distributions for the parameters, including climate sensitivity. One observation type that can provide constraint on climate sensitivity is historical near-surface temperature records (Tomassini et al., 2007). Everything else being equal, higher rates of near surface warming will imply higher climate sensitivities.

The transient temperature response a climate model depends on other parameters besides climate sensitivity. Therefore, for this approach to be able to estimate climate sensitivity, a knowledge of these other parameters (or climate system properties) has to be obtained. In particular, one needs to know (i) past forcings of climate change (Knutti and Hegerl, 2008; Forest et al., 2002) and (ii) anthropogenic heat uptake by the oceans and the parameter that typically controls it in the models – vertical ocean diffusivity (Stouffer et al., 2006; Forest et al., 2006).

Estimates of climate sensitivity hinge critically on our knowledge of past radiative forcings. For example, if the past forcings (assumed net positive) of the climate were low, then high climate sensitivity is required to simulate the observed temperature increase. However, if past radiative forcings were high, then a low climate sensitivity is needed (Knutti and Hegerl, 2008). The major past forcing uncertainty is the one associated with anthropogenic sulfate aerosols. Other forc-
ings, such as changes in solar irradiance, as well as anthropogenic forcings by CO₂, ozone, land use, black carbon on snow, etc. are also uncertain (Forster et al., 2007).

Climate sensitivity estimates are also strongly dependent on the ability of the oceans in the models to take up the heat associated with anthropogenic warming (Forest et al., 2006). In the models this property of the oceans is typically controlled by the vertical ocean diffusivity parameter. Vertical ocean diffusivity parameterizes sub-grid scale vertical mixing processes in the ocean. Simulations with higher vertical ocean diffusivity result in higher heat uptake by the ocean but slower surface warming (Urban and Keller, 2009). As the Charney report summarized it (NAS, 1979) ”the equilibrium warming [of the atmosphere] will eventually occur, it will merely have been postponed”. One can explain the observed trend in historical surface temperature records either by low climate sensitivity and a low ocean diffusivity; or by high climate sensitivity and high vertical ocean diffusivity (Urban and Keller, 2009). By determining the value of vertical ocean diffusivity, one can therefore better constrain climate sensitivity.

In addition, it is important to know the characteristics of natural (unforced) variability in the climate system in order to estimate climate sensitivity (Tomassini et al., 2007).

One promising approach is to simultaneously estimate climate sensitivity, vertical ocean diffusivity and anthropogenic sulfate aerosol effects using model runs constrained with historical data (Forest et al., 2002). We will now review some of the recent relevant studies taking this approach.

### 1.4 Recent Papers on Climate Parameter Estimation

Knutti et al. (2003) estimated numerous model parameter probability distribution functions (pdfs) using the Bern 2.5D climate model (1D atmosphere and 2.5D ocean). To construct the pdf, they selected only those model runs that were consistent with observed surface temperature, and oceanic heat uptake. They varied climate sensitivity, a multitude of atmospheric forcings, carbon feedback, and ocean mixing parameterizations; and presented posterior probabilistic information for climate sensitivity, radiative forcings and certain future model predictions.

Tomassini et al. (2007) used the same model to perform Bayesian parameter
estimation and robust Bayesian analysis. They obtained multivariate pdfs for twelve model parameters. They used a likelihood function, together with a realistic statistical model for the observations. Compared with Knutti et al. (2003), they were able to better constrain climate sensitivity. This study performed extensive sensitivity analysis of the results with respect to statistical assumptions (i.e. prior distributions) and commented on the potential of ocean heat uptake data to constrain climate sensitivity. Tomassini et al. (2007) also analyzed sensitivity of the results to the magnitude of the variability and observational error in surface temperatures and ocean heat uptake.

Another important paper on climate parameter estimation is Forest et al. (2002). They constrained the runs of the MIT 2D statistical-dynamical climate model with observations of the trend in oceanic heat uptake together with expanded surface and upper-air temperatures. They estimated the joint pdf for the rate of heat uptake by the deep ocean ($K_v$), climate sensitivity and anthropogenic aerosol forcing. The model and parameter likelihoods were calculated from an $r^2$ statistic. Forest et al. (2006) improved on their previous work by including some new forcings, which changed the posterior probability pdf for model parameters.

Drignei et al. (2008) expanded on work of Forest et al. (2006) by implementing a fast surrogate for the model (emulator). Emulators are tools that approximate models by efficiently mapping model input(s) into model outputs. Emulators can be used to interpolate model output between the parameter values at which the model was run. The advantage of the emulators is that they can be run much faster than the original model. The study also used a statistical model which accounts for different sources of uncertainty (e.g. climate model internal variability, etc.) and for correlation in various dimension of the problem (e.g. space-time correlation). Drignei et al. (2008) assumed independence of the three different observation datasets conditioned on the true climate signal. The study showed tighter confidence intervals compared to previous work, especially for climate sensitivity and ocean diffusivity parameters.

A more recent paper by Sanso and Forest (2009) also implemented a robust statistical framework where both physical parameters and statistical parameters pertaining to various sources of uncertainty were simultaneously estimated in an Markov Chain Monte Carlo (MCMC) procedure. This study used only surface temperature diagnostics. Sanso and Forest (2009) performed a sensitivity analysis
of the results with respect to priors for the true physical parameter values and for the spatial covariance matrix of the emulator. A variable amount of information from control runs of a GCM was used to define several priors on the spatial covariance matrix. The different priors on the true parameter values and on the covariance matrix had a large effect on the posterior pdf for climate model parameters. For example, one particular choice of priors yielded a median of 2.5 K for climate sensitivity whereas another choice lead to a median of 7.2 K. Other cases resulted in considerable probability mass above 10 K, although misconvergence of the MCMC chain was a probable cause in one of these cases (Forest, 2010, personal communication).

1.5 Current Challenges and Ways Forward

Despite the recent remarkable progress in our ability to model climate, estimates of climate sensitivity remain deeply uncertain. As summarized in (Edwards et al., 2007) this is

"due both to problems inherent in using the historic period as a reference, which include the small climate signal and uncertainties in the forcings and ocean heat uptake, and to problems inherent in the ensemble approach, which include the sensitivity to experimental choices, the uncertainty associated with interpolation between members of a small ensemble (if used), and the uncertainty inherent in the model itself."

MCMC methods also present a potential source of error, for example, due to (i) misconvergence of the chain or (ii) equifinality (Beven, 2006). Equifinality occurs when, given the same observational constraints, multiple models or parameter values are feasible. A simple case is when only a function (i.e. product) of two parameters is used within the model. In this example, only the product of the parameters can be estimated, but not the individual parameters independently.

Few attempts to quantify errors inherent in such Bayesian methods have been carried out to date. Tomassini et al. (2007) perform a sensitivity analysis of real data assimilation results to the magnitude of observational errors and natural variability in the surface temperature and ocean heat content change. Sanso and Forest (2009) present sensitivity analysis of the results to priors for some param-
eters. These studies provide important new insights but they do not address two key questions: (i) What is the skill of the method to recover known parameter values? (ii) Can random realizations of combined model and observational error introduce biases into the assimilation results? Urban and Keller (2009) use ’control’ or ’perfect model experiments’ to assess the skill of observations to constrain parameter estimates. Specifically, Urban and Keller (2009) generate observations using a model with known parameter values and subsequently re-estimate these parameters. However, they do not use the MCMC method to arrive at posterior pdfs; they evaluate the likelihood function on a grid of parameter values. Furthermore, their analysis does not account for uncertainty in aerosol forcing, employs a simple model, and uses only a very small subset of the relevant data.

In this Thesis I present a more comprehensive set of perfect model experiments for a Markov Chain Monte Carlo method that estimates the three climate parameters – climate sensitivity, background vertical ocean diffusivity, and relative strength of anthropogenic sulfate aerosol forcing. The method uses runs of an Earth System Model of Intermediate Complexity (EMIC) and globally averaged observational constraints of temperature and ocean heat content, to estimate posterior pdfs of the model parameters. I improve on previous studies in three main ways. First, I use a more realistic model and more observations in the perfect model experiments compared with Urban and Keller (2009). Second, I account for the uncertainty in the aerosol forcing. Third, I quantify the effect of observation type (surface temperature vs. ocean heat uptake vs. both) on the skill of the method to recover known parameter values given different assumptions about model and observational errors.

I find that for current estimates of combined model and observational error the results of the perfect model experiments (i.e. posterior pdfs for the model parameters) hinge critically on the random realizations of the combined error process. The skill of the method improves considerably if the combined error is reduced. In hypothetical cases where the combined observational and model error is assumed to be reduced (compared to its current estimate), using both observational constraints improves the skill of the method compared to when just the temperature or the ocean heat content data are used individually.

I also present hindcasts from the ensemble of EMIC runs and assess the skill of the model relative to more complex GCMs at simulating the historical global
average temperature record. I find that a well-calibrated EMIC can perform as well as GCMs at simulating the global average temperatures.


IPCC (2007). Summary for Policymakers. In Solomon, S., Qin, D., Manning, M.,
Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., editors,
Group I to the Fourth Assessment Report of the Intergovernmental Panel on
Climate Change. Cambridge University Press, Cambridge, United Kingdom and
New York, NY, USA.

pheric CO$_2$ records from sites in the SIO air sampling network. In Trends: A
Compendium of Data on Global Change. Carbon Dioxide Information Analysis
Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge,
Tenn., U.S.A.


Cycles, 23.

Manabe, S. and Wetherald, R. T. (1975). Effects of doubling of CO$_2$ concentra-

US National Academy of Sciences, Washington, DC.


Solomon, S., Qin, D., Manning, M., Alley, R. B., Berntsen, T., Bindoff, N. L.,
Chen, Z., Chidthaisong, A., Gregory, J. M., Hegerl, G. C., Heimann, M., Hewit-
son, B., Hoskins, B. J., Joos, F., Jouzel, J., Kattsov, V., Lohmann, U., Matsuno,
T., Molina, M., Nicholls, N., Overpeck, J., Raga, G., Ramaswamy, V., Ren, J.,
Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., editors,


Chapter 2

What is the Skill of Climate Parameter Estimation Methods? A Case Study with Global Average Observational Constraints
What is the skill of climate parameter estimation methods? A case study with global average observational constraints

R. Tonkononenkov *,1, M. Goes1, N. M. Urban1, H. D. Matthews2 and K. Keller1,3

1 Department of Geosciences, Penn State University, University Park, PA, USA
2 Department of Geography, Planning and Environment, Concordia University, Montreal, QC, Canada
3 Earth and Environmental Systems Institute, Penn State University, University Park, PA, USA

* Corresponding author email: rzt2-wrk [at] psu [dot] edu

Projected for submission to the Journal of Geophysical Research
2.1 Abstract

An important driver of uncertainty in future climate projections is that key climate model parameters (such as climate sensitivity, vertical ocean diffusivity and anthropogenic sulfate aerosol radiative forcings) are poorly constrained. Recent parameter estimation studies employed inverse Bayesian methods and runs of Earth System Models of Intermediate Complexity (EMICs) to constrain these model parameters. These studies provide new key insights, but they are typically silent on two key questions: (i) What is the skill of different observations to constrain model parameters and how can it change if models and observations improve? (ii) How robust are the inversion methods given current estimates of combined model and observational errors?

Here we address these key questions using the University of Victoria Earth System Climate Model (UVic ESCM), a Markov Chain Monte Carlo (MCMC) parameter estimation method, and globally averaged pseudo-observations. Specifically, we run an ensemble of transient simulations of the UVic model that spans the last two centuries. We systematically vary parameter values of climate sensitivity, background vertical ocean diffusivity, and the strength of anthropogenic aerosol forcing within the ensemble. We then present a Bayesian method for estimating these parameters. We test the skill of the method with perfect model experiments using pseudo-observations of near-surface temperature and ocean heat content.

We demonstrate that at current estimates of combined observational and model error, posterior parameter pdfs hinge critically on random realizations of the error process. However, reducing observational and model error can drastically improve the skill of the method. If the combined error is reduced, surface temperature pseudo-observations have more skill to estimate the model parameters than upper ocean heat content data. Method skill is generally greatest when both pseudo-observations are used.

In addition, by comparing model output from the UVic ensemble runs with several General Circulation Models (GCMs) we show that a well calibrated EMIC can perform as well as GCMs at simulating global mean surface temperature over the past century.
2.2 Introduction

Climate sensitivity, vertical ocean diffusivity and anthropogenic sulfate aerosol forcing are important uncertain climate model (and climate) parameters (Forster et al., 2007; Cubasch et al., 2001; Forest et al., 2002). A number of model-based studies have recently aimed at producing joint probability distribution functions for these parameters (Forest et al., 2002, 2006; Knutti et al., 2003; Tomassini et al., 2007; Drignei et al., 2008, and others). Knutti et al. (2003) and Tomassini et al. (2007) use the Bern2.5D climate model, whereas Forest et al. (2002, 2006) and Sanso and Forest (2009) use the MIT 2D statistical-dynamical climate model. These studies typically use an ensemble of model runs, and a likelihood function that quantifies the degree to which the model runs match certain climate observations, to present probability density functions for the key climate parameters.

Methodologically, Markov Chain Monte Carlo (MCMC) methods can be employed to evaluate the likelihood function in model and statistical parameter space. Despite the ubiquitous use of these methods in the literature (Sanso et al., 2008; Tomassini et al., 2007), the skill of these methods to specifically estimate climate parameters has not been thoroughly investigated.

One way to evaluate these methods is to use pseudo-observations (derived in some way from a model run with known climate parameter values), and observe whether the inverse MCMC method can recover these parameter values. To our knowledge, such evaluations of inverse methods for climate parameter estimation have not been reported in the literature. Urban and Keller (2009) present such a set of perfect model experiments. However, instead of the sampling parameters using MCMC, they evaluate likelihoods on a grid of parameter values. Their analysis, while providing key new insights, uses limited data, does not account for uncertainty in aerosol forcing, and uses a simple model.

The present research expands on previous analyses in several ways. First, we account for uncertainty in anthropogenic sulfate aerosol forcing. Second, we employ a more realistic EMIC compared to analysis of Urban and Keller (2009). Third, we use more information from observational constraints. Fourth, we compare the performance of the method for different observational constraints under various assumptions about combined model and observational error.

The rest of this paper is organized as follows. Section 2.3 details the Earth
System model, its simulations, and physical observations used as the constraints. Section 2.4 describes the assimilation method and the perfect model experiments. Section 2.5 presents and discusses the results of the perfect model experiments. In this section we also compare the UVic output with results from General Circulation Models (GCMs). The paper concludes with Caveats and Conclusions sections.

2.3 Methods

2.3.1 Model Description

University of Victoria Earth System Climate Model (UVic ESCM) is an intermediate complexity model developed at the University of Victoria (Weaver et al., 2001). We use the most recent version of the model available at the time of the study, version 2.8. The ocean component of the model is MOM2 (Pacanowski, 1995) with a 1.8° (lat) × 3.6° (lon) resolution in the horizontal and 19 depth levels. The atmospheric component is a one-layer atmospheric energy-moisture balance model. The model is forced by prescribed winds from the NCAR/NCEP climatology, and no flux corrections are applied. Also included are thermodynamic sea-ice component, the terrestrial dynamic vegetation model TRIFFID (Cox, 2001), and oceanic biogeochemistry based on the ecosystem model of Schmittner et al. (2005).

We modify the default 2.8 version of the model to incorporate additional greenhouse gas (AGGs), volcanic, and anthropogenic sulfate aerosol forcings as detailed in subsection 2.3.3.2. Some of the forcings that we use for transient historical climate simulations differ from those in the original UVic 2.8 distribution; section 2.3.3.2 describes the forcings in more detail.

2.3.2 Model Parameters

2.3.2.1 Climate Sensitivity

We introduce an additional perturbation term into the parameterization of planetary outgoing longwave radiation in order to vary climate sensitivity in the model according to:

\[
Q_{PLW}^* = Q_{PLW} + f(T - T_0).
\]  

(2.1)
In this equation $Q_{PLW}$ represents the planetary outgoing longwave radiation as calculated in the original 2.8 version of the model and $Q_{PLW}^*$ represents the modified outgoing longwave radiation. The original version of the model assumes a certain longwave radiation feedback (through, for example, the dependence of $Q_{PLW}$ on temperature). The processes contributing to this feedback, such as changing cloud altitude, cloud height, or lapse rate, are not explicitly incorporated into the model. We modify this feedback by the term $f(T - T_0)$. This feedback term is proportional to the deviation of local near-surface atmospheric temperature from its initial local value (in this case it is temperature at year 1800), and to the $f$ parameter. We vary the parameter $f$ to achieve different climate sensitivities. The parameterization loosely follows Knutti et al. (2003) and Matthews and Caldeira (2007). However, in our formulation the extra feedback is a function of local, rather than global temperatures.

Climate sensitivity ($CS$) in the model does not solely depend on $f$. At a given value of $f$, it is weakly dependent on the background vertical ocean diffusivity $K_{bg}$. This is because differences in oceanic circulation between model versions with different $K_{bg}$ can change additional feedbacks (e.g. ice-albedo feedback). To determine the relationship between $K_{bg}$, $f$ and climate sensitivity, we ran a small ensemble of $2 \times \text{CO}_2$ runs with varying $K_{bg}$ and $f$. Climate sensitivity was diagnosed for each run of the ensemble by fitting the function

$$\bar{T} = \bar{T}(\alpha_1, \ldots, \alpha_4, t) = \alpha_1 (1 - \exp (\alpha_2 t)) + \alpha_3 (1 - \exp (\alpha_4 t))$$

(2.2)

to global mean temperature as a function of time $t$. For each run of the small ensemble, the parameters $\alpha_1, \ldots, \alpha_4$ were optimized, and then climate sensitivity was estimated by evaluating $\bar{T}(\hat{\alpha}_1, \ldots, \hat{\alpha}_4)$ at $t = \infty$, where $\hat{\alpha}_1, \ldots, \hat{\alpha}_4$ were the optimized parameters. The fitting allowed us to evaluate climate sensitivity on a small grid of $K_{bg}$ and $f$ values. These climate sensitivities as a function of $K_{bg}$ and $f$ are plotted in Figure 2.1. To obtain climate sensitivity at the values of the parameters where the transient model ensemble was run, kriging was applied. Specifically, we fitted a Matern covariance function to log($CS$). We stretched the $K_{bg} - f$ coordinates during the fitting to correct for anisotropy.

The prior range of $f$ is such that the resulting climate sensitivities span the
Figure 2.1: Climate sensitivity as a function of $f$ and $K_{bg}$ as diagnosed from the CO$_2$ doubling runs. The locations of the CO$_2$ doubling runs in the parameter space are specified by the black circles. The red tick marks correspond to the $f$ and $K_{bg}$ values where the hindcast ensemble model runs (described in section 2.3.3.1) were run. These runs were run on a quazi-linear grid of the $f$ and $K_{bg}$ model parameters. Specifically, the $f$ values used in this ensemble have a slight dependence on $K_{bg}$, which results in the clustering of the red tick marks.

range from 0.564 to 14.808 K (per doubling of CO$_2$).

2.3.2.2 Vertical Ocean Diffusivity

Vertical diffusivity $K_v$ in the UVic model has background and tidal components (Schmittner et al., 2009):

$$K_v = K_{tidal} + K_{bg}. \quad (2.3)$$

The background diffusivity parameter $K_{bg}$ is assumed constant in space, and it is $K_{bg}$ that we vary and estimate in this study. We have modified the model to limit
$K_v$ to $\geq 1 \text{ cm}^2 \text{s}^{-1}$ in the Southern Ocean below 500 m as in Schmittner et al. (2009) so $K_{bg}$ does not affect mixing there. We use the prior range from 0.1 to 0.5 cm$^2$ s$^{-1}$ for $K_{bg}$.

### 2.3.2.3 Anthropogenic Aerosol Scaling Factor

Our modified 2.8 version of the model incorporates anthropogenic sulfate aerosols as perturbations to local albedo following Matthews et al. (2004), using the parameterization of Charlson et al. (1991). We refer to these perturbations as ‘sulfate albedoes’. Please see Appendix A for more information about this parameterization, including the correct version of Equation 7 of Matthews et al. (2004). Assuming proportionality between the direct and the indirect effects of the aerosols, we scale the original ‘sulfate albedo’ representing direct effects of aerosols to obtain total ‘sulfate albedo’ which incorporates the total effects:

$$\alpha_{total} = A_{scl} \times \alpha_{direct}. \quad (2.4)$$

We vary the strength of the total anthropogenic sulfate aerosol effects via the parameter $A_{scl}$. We consider a prior range for $A_{scl}$ between 0 and 3.5.

### 2.3.3 Description of Model Runs and Observations

#### 2.3.3.1 Run Design

The ensemble consists of 210 model runs. We vary the aerosol scaling factor $A_{scl}$ and $K_{bg}$ between the ensemble members such that the parameter values form a uniform grid. Specifically, for $K_{bg}$ we use values of (0.1, 0.2, 0.3, 0.4 and 0.5 cm$^2$ s$^{-1}$). For $A_{scl}$ we use values of (0, 0.7, 1.4, 2.1, 2.8 and 3.5). We vary $f$ on a non-uniform quazi-rectangular grid to result in a roughly uniform equidistant grid in climate sensitivity, but with slightly higher sampling in areas of high expected posterior probability (Figure 2.1). The relationship between the $f$ index (which is the same as the $CS$ Index) on this grid and the corresponding climate sensitivity itself ($CS$) is plotted in Figure 2.2. To the first approximation, the relationship between the $CS$ Index and $CS$ is roughly linear. In the rest of this paper we will work with the $CS$ Index.
2.3.3.2 Run Description

In order to generate initial conditions for our transient simulations, we first spin the model up for 3000 years for each value of $K_{bg}$ at conditions of year 1800 (for the first 2000 years of the spinup runs atmospheric CO$_2$ concentration is fixed, but later it is allowed to vary). We then perform transient model simulations for years 1800-2010. We force the model by a number of historical climate forcings. Carbon dioxide emissions from fossil fuels are prescribed following Zickfeld et al. (2008). Historical land use emissions are complemented by projections from IPCC A1FI Scenario after year 2002 (Nakicenovic and Swart, 2000). Compared with UVic ESCM 2.8, we use different solar irradiance forcing data. Specifically, the model is forced with total solar irradiance, reconstructed from Group sunspot number.

Figure 2.2: The relationship between the index on the climate sensitivity grid (CS Index) where the ensemble runs were performed, and the climate sensitivity (CS) itself. The slight spread of climate sensitivity values at the same CS Index is due to the weak dependence of the CS on $K_{bg}$ at the same CS Index on the grid.
(Krivova et al., 2007). Volcanic forcings (converted to anomaly) for years 1800-1850 are taken from Crowley (2000a,b), and for years 1850 to 2000 from Sato et al. (1993) and GISS (2007a). Additional greenhouse gas forcings include contributions from CH$_4$, N$_2$O and CFCs (Hansen et al., 1998; Hansen and Sato, 2004; Eby, 2007, and others); they are the same as used in the University of Victoria Earth System Model version 2.9 (UVic Climate Modelling Group, 2010). Anthropogenic sulfate aerosol forcings are input into the model as gridded optical depth data (Koch et al., 1999; Tegen et al., 2000; GISS, 2007b).

Figure 2.3 shows the UVic hindcasts (1800-2010) for several globally averaged variables. We find that changing the parameters has considerable effects on atmospheric CO$_2$, temperature and heat uptake by the upper ocean (Figure 2.3 (a), (b), (d)). Preliminary parameter scans (not shown) indicate that at typical parameter values higher climate sensitivity and/or smaller sulfate aerosol forcing leads to higher near-surface and upper ocean warming. These scans also showed that at typical parameter values temperature and upper ocean heat uptake are less sensitive to $K_{bg}$ as to the other model parameters.

Changing $K_{bg}$ leads to considerable changes in the simulated strength of Atlantic meridional overturning circulation (AMOC). Runs with highest $K_{bg}$ have highest AMOC at the beginning of the simulation at equilibrium (Figure 2.3 (e)). This is in agreement with other studies (Bryan, 1987, Goes et al., in review). In many runs the AMOC slows down toward the end of the simulation period. However, in some model runs the AMOC increases in strength over time. This happens when strong anthropogenic sulfate forcing is applied.

The carbon uptake by the ocean (Figure 2.3 (c)) is less sensitive to the parameters that we vary compared to the other model outputs.

2.3.3.3 Observational Constraints

For the Markov Chain Monte Carlo parameter estimation, we use the following two observational constraints. First, we use global average atmospheric above-surface/ocean surface temperature time series from the Hadley Center (Had-CRUT3) (Brohan et al., 2006). These data have a yearly resolution and span the time period from year 1850 to 2006. These observations are used as anomalies with respect to mean of 1850-1899 period. The second dataset is global average ocean heat content anomaly in the 0-700 m layer (1 yr average) (Domingues et al.,
Figure 2.3: UVic ESCM model hindcasts (grey lines), and corresponding observations (red lines or markers): (a) global average atmospheric surface temperature [K], and observations of global average near surface/ocean surface temperature anomalies from the HadCRUT3 dataset (Brohan et al., 2006) with arbitrary offset added; (b) ocean heat content anomaly in the 0-700m layer [J], and observations from Domingues et al. (2008); black lines delineate observational 1σ error intervals. (c) Ocean carbon flux [GtC yr\(^{-1}\)] and decadal observational constraints (1.6 GtC yr\(^{-1}\) for the 1980s and 2.0 GtC yr\(^{-1}\) for the 1990s) from McNeil et al. (2003); (d) atmospheric CO\(_2\) concentrations; the observations are from Mauna Loa observatory (Keeling et al., 2009) and Law Dome ice core (Etheridge et al., 1996); (e) Atlantic meridional overturning circulation (AMOC) intensity [m\(^3\) s\(^{-1}\)], with observations from Bryden et al. (2005) and Cunningham et al. (2007), and their error estimates from Kanzow et al. (2007) and Lumpkin and Speer (2007).
The ocean heat content observations extend from year 1950 to 2003, and they enter as anomalies with respect to the mean of the observational period (years 1950 to 2003).

2.4 Statistical Methods

2.4.1 Statistical Model and Markov Chain Monte Carlo

We denote observations by $y_{t,i}$, where $t = 1, ..., N_i$ is time index, $N_i$ is number of observational datapoints and $i$ is the tracer index (1 refers to temperature and 2 to ocean heat content). We also denote $y_i = (y_{1,i}, ..., y_{N_i,i})$ and $Y = (y_1, y_2)$. Likewise, we denote model output by $\mu_{t,i}(\theta)$ where $\theta$ is a vector of model parameters ($CS, K_{bg}, A_{sc}$). We assume that observations follow the statistical model

$$y_{t,i} = \mu_{t,i} + b_i + \epsilon_{t,i}, \quad (2.5)$$

where $b_i$ is a time-constant bias, and $\epsilon_{t,i}$ is time dependent error term. The time-constant bias may be, for example, due to time-independent model inadequacies. $\epsilon_{t,i}$ represents the combined effects of observational error, time dependent model structural error, and climate variability which is not represented in the model. We assume that $\epsilon_{t,i}$ is a first order autoregressive process with innovation variance $\sigma^2_i$ and AR(1) coefficient $\rho_i$.

Define whitened errors as $w_{t,i} = \epsilon_{t,i} - \rho_i \epsilon_{t-1,i}$, $t > 1$, and the stationary process variance as $\sigma^2_{p,i} = \sigma^2_i / (1 - \rho^2_i)$. The likelihoods for observations coming from this model can then be expressed (Bence, 1995) as:

$$p(y_i|\theta, \sigma_i, b_i, \rho_i) = (2\pi \sigma^2_{p,i})^{-1/2} \exp \left( -\frac{\epsilon^2_{t,i}}{2 \sigma^2_{p,i}} \right) \times \exp \left( -\frac{1}{2 \sigma^2_i} \sum_{j=2}^{N_i} w_{j,i}^2 \right). \quad (2.6)$$

Assuming independence of error terms $\epsilon$ between the different types of observations, the likelihood for all observations (conditional on the parameter values) can be expressed as $L(Y) = L(y_1) \times L(y_2)$.

We denote all model and statistical parameters by $\Theta$. According to the Bayes
Table 2.1: Prior ranges for statistical parameters used in the assimilation. $b_T$ refers to the bias parameter for temperature data, and $b_{OHC}$ refers to the bias parameter for ocean heat content data. The ranges for $\sigma$ and $\rho$ are the same for both observations.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma$</td>
<td>0.01</td>
<td>inf</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.01</td>
<td>0.99</td>
</tr>
<tr>
<td>$b_T [K]$</td>
<td>-0.51</td>
<td>0.50</td>
</tr>
<tr>
<td>$b_{OHC} [1 \times 10^{22} J]$</td>
<td>-7</td>
<td>7</td>
</tr>
</tbody>
</table>

Theorem, the posterior probability for all parameters can be expressed as:

$$p(\Theta|Y) \propto L(Y|\Theta) \times p(\Theta),$$

(2.7)

where $p(\Theta)$ is prior probability distribution for the model and statistical parameters. In our case all prior probabilities were uniform.

We use the Markov Chain Monte Carlo (MCMC) (Metropolis et al., 1953; Hastings, 1970) method to draw samples from joint probability distribution $p(\Theta|Y)$. We use information from previous chain covariance to construct proposal distributions following Roberts and Rosenthal (2009). We run three chains (two prechains and a final chain), and use the information from covariance in each previous chain to construct a proposal distribution for each subsequent chain. We use model parameter prior ranges specified in Sections 2.3.2.1, 2.3.2.2 and 2.3.2.3; and statistical parameter ranges, shown in Table 2.1. We obtain marginal probability distributions for each model and statistical parameter from the individual parameter chains.

2.4.2 Interpolation

Model parameters were sampled continuously over their prior ranges. Model output was interpolated between the ensemble parameter values to each sampled point in model parameter space using simple interpolation (linear in $K_{bg}$ and bilinear in the $A_{scl} - f$ index parameter space).

2.4.3 Perfect Model Experiments

The objective of the perfect model experiments is to analyze the performance of the MCMC when pseudo-observations are used. To construct the pseudo-observations,
Table 2.2: Mean posterior estimates for the statistical parameters $\sigma$ and $\rho$, for each observation type, as obtained from the assimilation experiment which uses both types of real observations together.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Temperature</th>
<th>Ocean Heat Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\sigma}_i$</td>
<td>0.10 [K]</td>
<td>$2.72 \times 10^{22}$ [J]</td>
</tr>
<tr>
<td>$\hat{\rho}_i$</td>
<td>0.61</td>
<td>0.21</td>
</tr>
</tbody>
</table>

a noise process needs to be superimposed on model output. The purpose of the noise process is to represent combined effects of model and observational error. For this, an estimate of the noise properties needs to be made.

To do so, we first run a real assimilation with both observations together, using the MCMC method described above. We assume the statistical model as in Eqn. 2.5 and estimate statistical parameters $\sigma_i$ and $\rho_i$ of the noise process $\epsilon_i$ for both temperature and ocean heat uptake. Table 2.2 lists the estimated means of these noise parameters.

For our perfect model experiments, we use model output at model parameter indices (3,3,3) on our grid of model parameters $K_{bg}$, $f$ and $A_{scl}$. These indices correspond to parameter values $K_{bg} = 0.3 \text{ cm}^2\text{s}^{-1}$, $CS = 2.74 \text{ K}$ and $A_{scl} = 1.4$. We construct pseudo-observations from this model run by contaminating model output from this run with noise processes with various properties. We use the same time period for the pseudo-observations as for the real observations described in Section 2.3.3.3. We perform the analysis in the space of model parameter indices. Note that the relationship between parameter indices and parameters is linear for $K_{bg}$ and $A_{scl}$, but slightly nonlinear for $CS$ (Figure 2.2).

Using the perfect model experiments, we analyze three effects. First, we analyze the effects of potentially reducing observational and model error on the skill of the method. To do so, we run three sets of experiments, called $(\sigma, \rho)$, $(\sigma, \rho)/3$ and $(\sigma, \rho)/10$ respectively. In the $(\sigma, \rho)$ experiment we use the AR(1) noise with current estimates of combined error parameters from from Table 2.2 to construct the pseudo-observations. This experiment is meant to represent the case of current knowledge about the system. In the $(\sigma, \rho)/3$ and $(\sigma, \rho)/10$ experiments we repeat the analysis with both innovation standard deviation of the error $\sigma_i$ and its autoregressive AR(1) coefficient $\rho_i$ divided by factors of 3 and 10. These experiments represent hypothetical cases of improving model and/or observation systems such that the combined model and observational error decreases in magnitude and
becomes less autocorrelated.

Second, we analyze the skill of the methods when different observations are used. To this end, we repeat the \((\sigma, \rho)\), \((\sigma, \rho)/3\) and \((\sigma, \rho)/10\) using just the temperature data, just ocean heat uptake data, and then both pseudo-observations together.

Finally, we analyze the effects of random realizations of the error process on the skill of the method. We do so by repeating the assimilation of pseudo-data with the current combined error estimates five times.

\section{2.5 Results}

\subsection{2.5.1 Comparison with GCM Output}

The adopted EMIC has a more simplified representation of many physical processes compared to more complex GCMs considered in the IPCC report (Meehl et al., 2007). However, EMICs are more computationally efficient to run. This allows for better exploration of parameter space compared to the GCMs for a fixed computational effort. While key climate parameters, diagnosed from GCMs, such as climate sensitivity, can be biased compared with results of Bayesian estimations (Forest et al., 2006), these parameters can be adjusted more easily in EMICs. We hypothesize that an EMIC with the optimized values of climate parameters can be more skillful than GCMs at simulating the global average response of climate to anthropogenic forcings.

To test this hypothesis, we analyze the skill of UVic model runs at simulating global mean annual temperatures over the past century. Specifically, we compare the temperatures from the UVic runs with the output from the World Climate Research Programme’s (WCRP’s) Coupled Model Intercomparison Project phase 3 (CMIP3) multi-model dataset (Meehl et al., 2007) using a Taylor diagram (Figure 2.4). We use observations of near-surface / ocean surface temperatures for period 1900-1999 (Brohan et al., 2006) as reference. Many UVic models runs that have extreme parameter values perform quite poorly in terms of standard deviation. However, some UVic ESCM runs have similar performance as the CMIP3 archive runs (based on the centered root mean square differences) in predicting global mean atmospheric temperature (Figure 2.4). These results are further supported
Figure 2.4: Taylor diagram representing the skill of UVic ESCM ensemble runs (red dots) and GCMs from CMIP-3 multi-model dataset (blue crosses) at simulating global mean near surface temperature for period 1900-1999 (Brohan et al., 2006). The temperatures were used as anomalies with respect to mean of period 1900-1949. We used the following models from the CMIP3 multi-model dataset (Meehl et al., 2007): BCCR–BCM2.0, CGCM3.1(T47), CSIRO–Mk3.0, CSIRO–Mk3.5, GISS–ER, MIROC3.2(medres), ECHO–G, ECHAM5/MPI–OM and MRI–CGCM2.3.2. We neglect the difference between near-surface atmospheric temperatures and ocean surface temperatures in this analysis. Dotted lines are isolines of standard deviations, semi-circle dashed lines are isolines of correlation coefficient, and straight dashed – of centered root mean square difference.

by the plot of model temperature hindcasts, and the observations (Figure 2.5).

The fact that an EMIC with optimized parameter values can perform as well as GCMs in terms of simulating globally averaged output both increases confidence in using EMICs as modeling tools and underlines the importance of properly estimating the key climate parameters for the purpose of making future climate projections.

Of course, the Taylor diagram is just one possible way to evaluate model performance. Hence, the results would need to be further confirmed using fully likelihood-based data comparison methods.
Figure 2.5: Output of UVic ensemble and GCMs from the CMIP-3 dataset together with observations of temperature anomalies with respect to the 1900-1949 period. See caption of Figure 2.4 for the list of models analyzed. Observations have yearly resolution; they are plotted as lines for visual clarity.

2.5.2 Positive Control Experiments

2.5.2.1 Effect of Combined Error

We define skill of the method as $S = 100/R$, where $R$ is the 95% credible interval for a parameter expressed as % of its prior range. At current estimates of combined
model and observational error, the method skill is fairly low (Table 2.3, Figure 2.9). Decreasing the combined model and observational error leads to a rapid improvement in the skill of the method (Table 2.3, Figure 2.9) and to convergence of the mode of the posterior pdf to true parameter values (Figures 2.6, 2.7, 2.8). In fact, when both $\rho$ and $\sigma$ are decreased by a factor of three from their current values, the method can recover the true parameters very well. The only exception is background ocean diffusivity when only ocean heat uptake observations are used. In this case, a marked improvement in method skill is only reached in the $(\sigma, \rho)/10$ experiment. The result of rapid improvement of method skill when decreasing the combined error underlines the need for improving both the observation systems and the climate models.

### 2.5.2.2 Method Skill for Different Parameters

The method has higher skill for estimating the aerosol scaling parameter, compared to vertical ocean diffusivity (Table 2.3). This result is observed for all three cases of pseudo-observations: temperature, ocean heat content, and both. We hypothesize that method skill for a model parameter is affected by variability of model output as a function of that parameter.

**Table 2.3**: Method skill $S$ for various positive control experiments. Please refer to the text for definition of $S$. For the $(\sigma, \rho)$ experiments each skill represents an average for five realizations of the random error process.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>$K_{bg}$</th>
<th>$CS$</th>
<th>$A_{scl}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(\sigma, \rho)/10$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>22.10</td>
<td>52.64</td>
<td>79.45</td>
</tr>
<tr>
<td>OHC</td>
<td>7.62</td>
<td>11.84</td>
<td>22.96</td>
</tr>
<tr>
<td>Both</td>
<td>34.75</td>
<td>58.07</td>
<td>89.15</td>
</tr>
<tr>
<td>$(\sigma, \rho)/3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>5.95</td>
<td>13.21</td>
<td>18.81</td>
</tr>
<tr>
<td>OHC</td>
<td>1.32</td>
<td>1.87</td>
<td>3.85</td>
</tr>
<tr>
<td>Both</td>
<td>8.37</td>
<td>21.62</td>
<td>32.17</td>
</tr>
<tr>
<td>$(\sigma, \rho)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>1.34</td>
<td>1.57</td>
<td>2.63</td>
</tr>
<tr>
<td>OHC</td>
<td>1.33</td>
<td>1.41</td>
<td>2.63</td>
</tr>
<tr>
<td>Both</td>
<td>1.35</td>
<td>2.22</td>
<td>3.47</td>
</tr>
</tbody>
</table>

In the case of current estimate of combined error $[(\sigma, \rho)$ experiments$]$, the interpretation of method skill as defined in this paper is confounded by potential biases in posterior pdfs (Figures 2.6, 2.7 and 2.8). We observe that these results are also dependent on prior ranges for the model parameters (i.e. choosing a dis-
Figure 2.6: Posterior pdfs for model parameter indices obtained from perfect model experiments using temperature pseudo-observations only. Top row: $(\sigma, \rho)/10$ experiment, middle row: $(\sigma, \rho)/3$ experiment, bottom row: $(\sigma, \rho)$ experiments. Dashed lines represent the true parameter values.

A proportionately high prior for a model parameter will result in perceived higher method skill. We also observe that, given the slightly nonlinear relationship between the CS Index and the climate sensitivity $(CS)$ itself (Figure 2.2), method skill at estimating CS Index is expected to differ from the skill at estimating the
Figure 2.7: Posterior pdfs for model parameter indices obtained from perfect model experiments using ocean heat content pseudo-observations only. Top row: $(\sigma, \rho)/10$ experiment, middle row: $(\sigma, \rho)/3$ experiment, bottom row: $(\sigma, \rho)$ experiments. Dashed lines represent the true parameter values.

climate sensitivity itself. Combined, these effects complicate comparing method skill across different model parameters.
2.5.2.3 Sensitivity of Results to Individual Error Realizations

At current levels of observational and model error, method results are highly sensitive to random realizations of the error noise process \( ((\sigma, \rho) \text{ plots in Figures 2.6, 2.7 and 2.8}) \). For the case of temperature pseudo-observations, one out of five realizations of the error process leads to the failure of the method to retrieve climate sensitivity index. Likewise, the method fails to recover \( CS \) index two out of five times when ocean heat content pseudo-observations are used, and one out of five times for both pseudo-observations together. Besides skill measure \( S \) another way to assess the method performance is using the uncertainty measure \( U \). This measure characterizes the sensitivity of the results to random realizations of the error process. We define the method uncertainty \( U \) as the sample variance of the mode of the posterior pdf for five different realizations of the error process. Method uncertainties for the current estimate of combined error are considerable (Table B.1 in Appendix B). We note that while this sample statistic is a useful demonstration of method limitations, it is not a robust estimate of the true method uncertainty. A robust estimate would require many more realizations of the error process.

The sensitivity of results to random error realizations underlines the need for caution when interpreting results of Bayesian parameter estimation for key climate model parameters, and highlights the need for performing perfect model experiments as part of scientific research.

2.5.2.4 Method Skill for Different Observations

For the reduced error experiments \( ([\sigma, \rho]/10 \text{ and } (\sigma, \rho)/3] \) the method works better in terms of skill \( S \) for temperature than for ocean heat uptake data (Figure 2.9). For these experiments, combining two pseudo-observations improves the skill of the method.

For the current estimated error experiments \( (\sigma, \rho) \), combining both pseudo-observations increases the skill of the method at estimating \( CS \) and aerosol scaling, but does not have an appreciable effect on estimating \( K_{bg} \) (Table 2.3). Ocean heat uptake and temperature have a similar skill to constrain model parameters. In these experiments, the skill concept becomes more limited due to biases in posterior pdfs (Figures 2.6, 2.7 and 2.8). Hence model uncertainty \( U \) needs to be also considered. We cannot conclude that using more observations decreases
Figure 2.8: Posterior pdfs for model parameter indices obtained from perfect model experiments using both temperature and ocean heat content pseudo-observations together. Top row: \((\sigma, \rho)/10\) experiment, middle row: \((\sigma, \rho)/3\) experiment, bottom row: \((\sigma, \rho)\) experiments. Dashed lines represent the true parameter values.

model uncertainty (Table B.1 in Appendix B). Considerably more realizations of the error process are required to robustly estimate the effects of combining pseudo-observations on model uncertainty.
Figure 2.9: Skill $S$ of the MCMC method for the different parameters under different assumptions about the magnitude of combined model and observational error. Left: for $K_{bg}$ index; middle: for $CS$ Index; right: for $A_{SCL}$ index.

### 2.6 Caveats

Our forthcoming conclusions are subject to several caveats. First, we only consider pseudo-observations at a specified model parameter setting. It is possible that the skill of the method depends on the values of the parameter that are used to generate the pseudo-observations. Second, we only consider a limited number of combined error process realizations. It is possible that the method might still fail to recover true parameter values for some realizations for the reduced noise cases.

When evaluating the performance of the UVic and GCM model runs, we neglect non-stationarity in the temperature observations. The non-stationarity makes it difficult to interpret standard deviations. The results obtained based on the Taylor diagram may change when a likelihood-based approach is used.

The eventual application of the model runs and the statistical method described here is to perform assimilation of real observations and jointly estimate background ocean diffusivity, climate sensitivity, and anthropogenic sulfate aerosol scaling. If these model runs are to be used for this purpose, an additional set of caveats emerges. First, it is necessary to keep in mind that the model outputs might be
influenced by other model parameters (besides the three physical parameters used in this work). If this is the case, the method would result in biased estimates of the included parameters. Goes et al. (in review), for example, show the sensitivity of estimates of $K_{bg}$ on Southern Ocean mixing parametrization for several types of globally averaged vertical ocean tracer profiles, including temperature, as observational constraints. Second, additional errors can be introduced because of imperfect knowledge of model forcings (Forster et al., 2007). Third, other potential model inadequacies need consideration. In particular, in this study we neglect potential spatial variability in $K_{bg}$ parameter. Also we assume a proportionality between direct and indirect sulfate aerosol effects. In reality the relationship between indirect and direct aerosol effects is likely non-linear. In addition, the model does not include all radiative forcings (Forster et al., 2007; Mickley et al., 2001). Some of the neglected forcings (tropospheric ozone, black carbon and organic carbon) might have spatial emission or concentration patterns similar to the anthropogenic sulfates. This can introduce biases when estimating the $A_{scl}$ parameter for the anthropogenic sulfate aerosols.

2.7 Conclusions

We perform a series of runs of an Earth System model of Intermediate Complexity (UVic model), where we systematically vary climate sensitivity, background vertical ocean diffusivity and the strength of anthropogenic sulfate aerosol forcing – the key climate parameters that control decadal- to century-scale response of the climate to radiative forcings. We compare the performance of the UVic model in terms of simulating global mean near-surface / ocean surface temperature with several GCMs that are part of the CMIP-3 multi-model dataset. Using a Taylor diagram and other model diagnostics we show that an optimized EMIC can perform as well as GCMs in terms of simulating globally averaged surface temperatures. This result needs to be confirmed using more statistically sound likelihood-based data inter-comparison methods.

We also present and test the skill of an assimilation method designed to estimate the model parameters using global surface temperature and upper ocean heat content as observational constraints. To test the method we assimilate pseudo-observations derived from a typical UVic model output, contaminated with noise
with various properties. We show that (i) decreasing combined model and observational error considerably improves the skill of the method; (ii) the method consistently performs better in terms of skill measure $S$ for anthropogenic aerosol scaling parameter compared to the background vertical ocean diffusivity; (iii) when realistic estimates of combined model and observational error are used in the analysis, the parameter estimation results hinge critically on random realizations of the error process; (iv) if the combined model and observational error was reduced, the temperature pseudo-observations would have more power to constrain the model parameters than the ocean heat uptake pseudo-observations; and (v) using both observations together increases the skill of the method in all cases except for $K_{bg}$ at current estimates of combined error.

The main implications of these results are twofold. First, improving model performance and reducing observational errors is crucial to reducing biases in model parameter estimates. Second, as part of parameter estimation work, positive control experiments need to be performed and skill of the assimilation methods needs to be, at least qualitatively, assessed.
Bibliography


Crowley, T. J. (2000b). Causes of Climate Change Over the Past 1000 Years, IGBP PAGES/World Data Center for Paleoclimatology Data Contribution Series # 2000-045. NOAA/NGDC Paleoclimatology Program, Boulder CO, USA.


Chapter 3

Conclusions and Future Work

3.1 Conclusions

We run an ensemble of runs of the UVic Earth System model where we vary parameters of climate sensitivity, background vertical ocean diffusivity and anthropogenic sulfate aerosol scaling. These are key uncertain model parameters that control its response to anthropogenic radiative forcings. The runs can be suitable for a range of analyzes such as model sensitivity studies, parameter estimation, inter-model comparisons, etc. We compare the global average near surface temperatures from both the ensemble runs and the models of the CMIP-3 multi-model dataset (Meehl et al., 2007) with observations of near surface / ocean surface temperatures. The comparison suggests that an Earth System model of intermediate complexity can perform as well as GCMs for the case of global output. The global average temperature rise in the UVic model is a strong function of the parameters that were varied in the ensemble. These results together both indicate the importance of estimating such parameters, and underline the validity of intermediate complexity models as useful tools. Since the Taylor diagram used to deduce this result was produced using non-stationary data, the findings need to be confirmed using more statistically sound likelihood-based methods.

We perform a set of perfect model experiments for a simple Markov Chain Monte Carlo method. This method was designed to estimate the probabilities for the physical model parameters based on how well the model fits near surface / ocean surface temperature and upper ocean heat content data. These perfect model experiments lead to several important results. First, decreasing combined
model and observational error improves the skill of the method as measured by the skill metric $S$. Second, the method performs consistently better (in terms of skill $S$) for anthropogenic aerosol scaling parameter compared to the background vertical ocean diffusivity. Third, when realistic estimates of combined model and observational error are used to construct the pseudo-observations, the posterior pdfs depend strongly on the random realizations of the error process. This implies that it is unlikely, but possible, to obtain a strongly biased pdf for climate sensitivity for similar MCMC methods simply due to the uncertainties in the observations and in the model (i.e. observational error, model structural error, interpolation error, etc.). Consider one conceptual example of how this can occur. Since the ocean heat uptake observations are not known precisely, it is possible, although unlikely, that, for example, the real ocean heat uptake was considerably lower than what is suggested by the observations (which are used for the assimilation). This would imply a considerably smaller true values for climate sensitivity or for vertical ocean diffusivity than what would be estimated using the observations. Our experiments explicitly demonstrate this effect. Fourth, in the reduced error experiments the method has higher skill when temperature observations are used compared to the case where ocean heat uptake observations are used. Last, when both observations are used, the method performs better compared to when only the individual observations are used. The only exception is the parameter $K_{bg}$ at the current estimates of combined error. It is unclear whether this result would hold for other datasets and/or methods. This is because when more observations are used, the number of parameters would increase, which can, everything else being equal, increase the complexity of estimation.

These results imply that an evaluation of the skill of the inverse Bayesian methods is needed in parameter estimation studies. They also demonstrate the importance of decreasing observational and model error, since this results in marked increase of method skill.

3.2 Future Work

Our results point to several potential avenues of future research. First, we only perform a relatively qualitative estimate of the effect of reducing the combined error on the method skill. It might be of interest to express more quantitatively
the gain associated with the reduction of model and observational error of ocean heat uptake and surface temperature data. Arguably, such a study would require separation of various sources of error as it is done in Tomassini et al. (2007) or Sanso and Forest (2009). It might also benefit from using a more realistic, heteroskedastic treatment of errors in the ocean heat uptake observations.

Second, the reasons behind the differential skill of MCMC methods at estimating different model parameters need to be investigated. One avenue of potential future research is to examine a relationship between the variability of model output as a function of a particular parameter, and the skill at estimating that parameter.

Third, a larger set of realizations of the error process might be desirable. If a large number of realizations are used, then the probability of method failure for current and potential future pseudo-observations can be obtained, and method uncertainty can be better assessed.

Finally, a question that is interesting from the perspective of method design is whether using more observations can improve the skill of the method. The current study shows that for the particular datasets we use, for cases of reduced combined error and arguably for one of the model parameters at realistic combined error estimates, this is the case. However, to make more general conclusions about the effects of more data, repetition of this experiment using more statistically advanced methods or different, potentially spatially resolved datasets (i.e. as in Goes et al., in revision) is desirable.
Bibliography


Aerosol Scaling Parameterization

The aerosol scaling parameterization implementation follows Matthews et al. (2004). The original parameterization was developed by Charlson et al. (1991). The sulfate aerosols effects are represented by spatially-dependent $\Delta a_s$. Specifically,

$$\Delta a_s = A_{scl} \frac{\beta \tau (1 - \alpha_s)^2}{\cos(Z_{eff})},$$

where $\beta = 0.29$ is an upward scattering parameter, $\tau$ is the spatially-varying aerosol optical depth and $\alpha_s$ is surface albedo. $Z_{eff}$ is an effective solar zenith angle such that $\cos(Z_{eff})$ is the diurnally averaged cosine of the zenith angle. I have modified the original parameterization by introducing the new $A_{scl}$ parameter, which scales the direct effects, represented by the original parameterization, to total effects of the aerosols. I vary $A_{scl}$ in the ensemble to achieve different sulfate aerosol effects.
### Method Uncertainty

**Table B.1**: Sample uncertainty of the method $U$ for different $(\sigma, \rho)$ experiments for each of the parameters. As described in text, $U$ is defined as sample variance of the mode of the posterior pdf for the five different realizations of the error process.

<table>
<thead>
<tr>
<th>Observations</th>
<th>$K_{bg}$ Index</th>
<th>$CS$ Index</th>
<th>$A_{SCL}$ Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>0.49</td>
<td>0.31</td>
<td>0.19</td>
</tr>
<tr>
<td>Ocean Heat Content</td>
<td>0.95</td>
<td>1.65</td>
<td>0.71</td>
</tr>
<tr>
<td>Both</td>
<td>1.24</td>
<td>0.94</td>
<td>0.28</td>
</tr>
</tbody>
</table>