

7-9-2009

# Monte Carlo strategies for calibration in climate models

Alejandro Villagran-Hernandez

Follow this and additional works at: [https://digitalrepository.unm.edu/math\\_etds](https://digitalrepository.unm.edu/math_etds)

---

## Recommended Citation

Villagran-Hernandez, Alejandro. "Monte Carlo strategies for calibration in climate models." (2009).  
[https://digitalrepository.unm.edu/math\\_etds/69](https://digitalrepository.unm.edu/math_etds/69)

This Dissertation is brought to you for free and open access by the Electronic Theses and Dissertations at UNM Digital Repository. It has been accepted for inclusion in Mathematics & Statistics ETDs by an authorized administrator of UNM Digital Repository. For more information, please contact [disc@unm.edu](mailto:disc@unm.edu).

Alejandro Villagran-Hernandez

Signature

Mathematics and Statistics

Department

This dissertation is approved, and it is acceptable in quality and form for publication;

*Approved by the Dissertation Committee:*

Dr. Gabriel Huerra



Chairperson

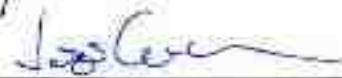
Dr. Edward John Bedrick



Dr. Michele Guindani



Dr. Joseph Galewsky



# Monte Carlo Strategies for Calibration in Climate Models

by

**Alejandro Villagran-Hernandez**

B.Sc., Actuarial Science, Universidad Nacional  
Autónoma de México, 2000  
M.Sc., Statistics, Universidad de Guanajuato, 2003  
M.Sc., Statistics, University of New Mexico, 2006

DISSERTATION

Submitted in Partial Fulfillment of the  
Requirements for the Degree of

Doctor of Philosophy  
Statistics

The University of New Mexico

Albuquerque, New Mexico

May, 2009

©2009, Alejandro Villagran-Hernandez

# Dedication

*To my wife Zazil, for all her love and patience through this journey.*

*To my family, for being the most precious treasure I would ever asked.*

*“To call in the statistician after the experiment is done may be no more than asking him to perform a post-mortem examination: he may be able to say what the experiment died of.”*

*– Ronald Fisher.*

# Acknowledgments

First of all, I would like to thank my advisor, Professor Gabriel Huerta, for his teachings and his priceless friendship over the years. I really appreciate all the time and effort that he has spent to form me not only as a statistician but also as a professional with ethical values. I will always be in debt with him for being so supportive even in my darkest hour.

I would also like to thank professors Edward Bedrick, Michele Guindani, and Joseph Galewsky as members of my committee for all their valuable comments to improve the quality of this manuscript. A special thanks to professor Bedrick for being a role model through my years as graduate student at UNM.

I would like to thank all my family, my parents (Socorro and Carlos), my siblings (Juan Carlos, Claudia, and Gabriela), my nephews (Carlos Emiliano, Sebastian, and Joaquin), my nieces (Alejandra and Miranda), and my family in law (Enrique, Joaquin, and Graciela), for all the love and support over these long years. I will never have the right words to let them know how proud I am to be a part of them.

Thanks to my friends (Gris, Kathya, Juan Miguel, Erika, and Zibonele) for being there for me despite the distance in some way or another. Special thanks to Alvaro Nosedal, my compilla over the last 9 years, I feel really fortunate to have you as my best friend. I consider you all part of my family. I am grateful for all the moments I have shared with you.

This work has been funded by the National Science Foundation grant OCE-0415251. I was also partially supported by the CONACyT-Mexico grant 159764.

Additionally, I would like to thank professors Alejandro Aceves, Michele Guindani, and Marina Vannucci for their help at crucial moments of my career. Thanks to The University of New Mexico and all the people related that help me to fulfill this goal. A special thanks to Dr. Charles S. Jackson for all his immense help through the learning process of climate models, and for being part of the enlightening experience of collaborative research.

# Monte Carlo Strategies for Calibration in Climate Models

by

**Alejandro Villagran-Hernandez**

ABSTRACT OF DISSERTATION

Submitted in Partial Fulfillment of the  
Requirements for the Degree of

Doctor of Philosophy  
Statistics

The University of New Mexico

Albuquerque, New Mexico

May, 2009

# Monte Carlo Strategies for Calibration in Climate Models

by

**Alejandro Villagran-Hernandez**

B.Sc., Actuarial Science, Universidad Nacional

Autónoma de México, 2000

M.Sc., Statistics, Universidad de Guanajuato, 2003

M.Sc., Statistics, University of New Mexico, 2006

Ph.D., Statistics, University of New Mexico, 2009

## Abstract

Intensive computational methods have been used by Earth scientists in a wide range of problems in data inversion and uncertainty quantification such as earthquake epicenter location and climate projections. To quantify the uncertainties resulting from a range of plausible model configurations it is necessary to estimate a multidimensional probability distribution. The computational cost of estimating these distributions for geoscience applications is impractical using traditional methods such as Metropolis/Gibbs algorithms as simulation costs limit the number of experiments that can be obtained reasonably. Several alternate sampling strategies have been proposed that could improve on the sampling efficiency including Multiple



Very Fast Simulated Annealing (MVFSA) and Adaptive Metropolis algorithms. As a goal of this research, the performance of these proposed sampling strategies are evaluated with a surrogate climate model that is able to approximate the noise and response behavior of a realistic atmospheric general circulation model (AGCM). The surrogate model is fast enough that its evaluation can be embedded in these Monte Carlo algorithms. The goal of this thesis is to show that adaptive methods can be superior to MVFSA to approximate the known posterior distribution with fewer forward evaluations. However, the adaptive methods can also be limited by inadequate sample mixing. The Single Component and Delayed Rejection Adaptive Metropolis algorithms were found to resolve these limitations, although challenges remain to approximating multi-modal distributions. The results show that these advanced methods of statistical inference can provide practical solutions to the climate model calibration problem and challenges in quantifying climate projection uncertainties. The computational methods would also be useful to problems outside climate prediction, particularly those where sampling is limited by availability of computational resources.

# Contents

<b>List of Figures</b>	<b>xi</b>
<b>List of Tables</b>	<b>xvii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 A Physical Surrogate Climate Model</b>	<b>6</b>
2.1 Introduction . . . . .	6
2.2 Milankovitch Cycles . . . . .	7
2.3 Climate Model . . . . .	9
2.4 Surrogate Climate Model . . . . .	11
2.5 The Cost Function . . . . .	13
<b>3 Monte Carlo Methods</b>	<b>17</b>
3.1 Introduction . . . . .	17
3.2 Gibbs and Metropolis samplers . . . . .	18

*Contents*

3.3	Multiple Very Fast Simulated Annealing (MVFSA) . . . . .	22
3.4	Adaptive Methods . . . . .	25
3.4.1	SCAM . . . . .	26
3.4.2	FAM . . . . .	27
3.4.3	DRAM . . . . .	30
3.5	Multiple-Try Metropolis . . . . .	33
3.6	Simulated Tempering . . . . .	34
3.7	Swapping . . . . .	36
<b>4</b>	<b>Computational Approaches Applied to a Climate Model</b>	<b>39</b>
4.1	Introduction . . . . .	39
4.2	Root Mean Square (RMS) Probability Error . . . . .	42
4.3	Appraisal of a few forward evaluations on the climate model . . . . .	46
4.4	Relevance of the parameter $S$ . . . . .	48
<b>5</b>	<b>Conclusions and future work</b>	<b>61</b>
	<b>References</b>	<b>65</b>

# List of Figures

1.1	A map of predicted global warming at the end of the 21st century according to the HadCM3 climate model with a business-as-usual emissions scenario (IS92a). (Source: Global Warming Art (2007)). . . . .	3
1.2	A comparison of predictions of temperature anomalies from 8 different climate models assuming the SRES A2 emissions scenario (Source: Global Warming Art (2007)). . . . .	4
2.1	Earth's orbital geometry parameters: Obliquity (Earth's axial tilt), Eccentricity (how elliptical the Earth's orbit around the Sun is), and longitude of Perihelion (point of closest approach to the Sun). (Source: Windows for the Universe (UCAR) (2007)). . . . .	7
2.2	Variations in Earth's orbit, the resulting changes in solar energy flux at high latitude, and the observed glacial cycles (Source: Global Warming Art (2007)). . . . .	8
2.3	Eccentricity ( $e(t)$ ), Obliquity ( $\Phi(t)$ ), and longitude of Perihelion ( $\lambda(t)$ ) time series: 500,000 yrs before present. . . . .	10

*List of Figures*

2.4 AGCM response of annual and zonal mean lowest level temperature (Kelvin degrees) from 70° S to 90° S (over Antarctica) to known continuous changes in Earth’s orbital geometry for the past 165 thousand years ( $e(t)$ ,  $\lambda(t)$ , and  $\Phi'(t)$  are known time series). Panel (a) shows the model response (in gray) with least squares fitted solution (black curve) given by equation (1). The least squares fitted solution includes an obliquity component (b), and a precessional component (c). The residual between the least squares fitted solution and the AGCM is shown in panel (d). . . . . 12

2.5 Likelihood for orbital forcing parameters. Left column: S=1. Right column: S=47. . . . . 16

3.1 Target distribution with 2 local modes and 1 global mode:  $\pi(m) = \frac{3}{9}N(-45, 25) + \frac{1}{9}N(2, 9) + \frac{5}{9}N(30, 4)$  . . . . . 18

3.2 Gibbs sampler applied to the target distribution  $\pi(m) = \frac{3}{9}N(-45, 25) + \frac{1}{9}N(2, 9) + \frac{5}{9}N(30, 4)$ . Top: The black solid line represents the actual pdf, and the white bars, the plot of the histogram. Bottom: Trace plot of iterations. . . . . 20

3.3 Metropolis-Hastings algorithm applied to the target distribution  $\pi(m) = \frac{3}{9}N(-45, 25) + \frac{1}{9}N(2, 9) + \frac{5}{9}N(30, 4)$ , using as proposal distribution a  $N(m^{(k-1)}, V)$ . Top: The black solid line represents the actual pdf, and the white bars, the plot of the histogram. Bottom: Trace plot of iterations. . . . . 21

List of Figures

3.4 Multiple Very Fast Simulated Annealing algorithm applied to the target distribution  $\pi(m) = \frac{3}{9}N(-45, 25) + \frac{1}{9}N(2, 9) + \frac{5}{9}N(30, 4)$ , using 400 independent initial points. Top: The black solid line represents the actual pdf, and the white bars, the plot of the histogram. Bottom: Trace plot of iterations. . . . . 23

3.5 Metropolis-Hastings algorithm applied to the target distribution  $\pi(m) = \frac{3}{9}N(-45, 25) + \frac{1}{9}N(2, 9) + \frac{5}{9}N(30, 4)$ , using as proposal distribution a  $N(m^{(k-1)}, 1)$ . Top: The black solid line represents the actual pdf, and the white bars, the plot of the histogram. Bottom: Trace plot of iterations. . . . . 26

3.6 Adaptive Metropolis algorithm applied to the target distribution  $\pi(m) = \frac{3}{9}N(-45, 25) + \frac{1}{9}N(2, 9) + \frac{5}{9}N(30, 4)$ , using as initial proposal distribution a  $N(m^{(t-1)}, 1)$ . Top: The black solid line represents the actual pdf, and the white bars, the plot of the histogram. Bottom: Trace plot of iterations. . . . . 29

3.7 Adaptive Metropolis algorithm applied to the target distribution  $\pi(m) = \frac{3}{9}N(-45, 25) + \frac{1}{9}N(2, 9) + \frac{5}{9}N(30, 4)$ , using as initial proposal distribution a  $N(m^{(t-1)}, 2)$ . Top: The black solid line represents the actual pdf, and the white bars the plot of the histogram. Bottom: Trace plot of iterations. . . . . 30

3.8 DRAM algorithm applied to the target distribution  $\pi(m) = \frac{3}{9}N(-45, 25) + \frac{1}{9}N(2, 9) + \frac{5}{9}N(30, 4)$ . Strategy used: First and second proposal distributions use same variance  $V = 1$ . Top: The black solid line represents the actual pdf, and the white bars, the plot of the histogram. Bottom: Trace plot of iterations. . . . . 31

*List of Figures*

3.9	DRAM algorithm applied to the target distribution $\pi(m) = \frac{3}{9}N(-45, 25) + \frac{1}{9}N(2, 9) + \frac{5}{9}N(30, 4)$ . Strategy used: the variance on the second stage proposal is a factor of the variance of the first stage proposal, $V_2 = 0.1 * V_1$ , where $V_1 = 10$ . Top: The black solid line represents the actual pdf, and the white bars, the plot of the histogram. Bottom: Trace plot of iterations. . . . .	32
3.10	Multiple-try Metropolis algorithm applied to the target distribution $\pi(m) = \frac{3}{9}N(-45, 25) + \frac{1}{9}N(2, 9) + \frac{5}{9}N(30, 4)$ , using $k = 10$ different proposed new values at every time. Top: The black solid line represents the actual pdf, and the white bars, the plot of the histogram. Bottom: Trace plot of iterations. . . . .	34
3.11	Simulated Tempering algorithm applied to the target distribution $\pi(m) = \frac{3}{9}N(-45, 25) + \frac{1}{9}N(2, 9) + \frac{5}{9}N(30, 4)$ , using 10 heated distributions with a cooling schedule similar to MVFSA. Top: The black solid line represents the actual pdf, and the white bars, the plot of the histogram. Bottom: Trace plot of iterations. . . . .	35
3.12	Swapping algorithm applied to the target distribution $\pi(m) = \frac{3}{9}N(-45, 25) + \frac{1}{9}N(2, 9) + \frac{5}{9}N(30, 4)$ , using 3 parallel chains starting from different points over the parameter space. Top: The black solid line represents the actual pdf, and the white bars, the plot of the histogram. Bottom: Trace plot of iterations. . . . .	38
4.1	Bivariate scatter plots of orbital forcing parameters. First column: FAM. Second column: SCAM. Third column: DRAM. Fourth column: MVFSA. Fifth column: METSA. . . . .	42
4.2	Comparison for Obliquity parameter. Top left: PPD estimation. Bottom: Ergodic quantile estimation (97.5%). . . . .	50

*List of Figures*

4.3	Comparison for Obliquity parameter. Top: RMS as function of iterations. Bottom: RMS as function of time (seconds). . . . .	51
4.4	Comparison for Longitude of Perihelion parameter. Top left: PPD estimation. Bottom: Ergodic quantile estimation (97.5%). . . . .	52
4.5	Comparison for Longitude of Perihelion parameter. Top: RMS as function of iterations. Bottom: RMS as function of time (seconds). . . . .	53
4.6	Comparison for Eccentricity parameter. Top left: PPD estimation. Bottom: Ergodic quantile estimation (97.5%). . . . .	54
4.7	Comparison for Eccentricity parameter. Top: RMS as function of iterations. Bottom: RMS as function of time (seconds). . . . .	55
4.8	Box plots of the samples from different methods. Top: Obliquity. Middle: Longitude of Perihelion. Bottom: Eccentricity. . . . .	56
4.9	Autocorrelation function of orbital forcing parameters. First column: FAM. Second column: SCAM. Third column: DRAM. Fourth column: MVFSA. Fifth column: METSA. . . . .	57
4.10	Bivariate scatter plots of orbital forcing parameters with just 500 iterations. First column: FAM. Second column: SCAM. Third column: DRAM. Fourth column: MVFSA. Fifth column: METSA. . . . .	58
4.11	Histograms with just 500 iterations (black bars). PPD Estimated using 100,000 iterations from the FAM algorithm (red line). First column: FAM. Second column: SCAM. Third column: DRAM. Fourth column: MVFSA. Fifth column: METSA. . . . .	59



*List of Figures*

4.12	First row: Estimated PPD using an adaptive method. Using the $S$ parameter (solid line) and using a non-informative prior (dotted line). Second row: Box plots of posterior samples. . . . .	60
5.1	Surrogate Climate Model with two solutions (modes). Comparative PPD estimation using Adaptive Metropolis methods and Simulated Annealing based techniques. Top: Comparison for Obliquity parameter. Bottom: Comparison for Longitude of Perihelion parameter. . .	64

# List of Tables

4.1	Time in seconds spent per computational method to compute one iteration with the climate surrogate model. . . . .	41
4.2	Comparative estimation after 500 forward evaluations. $E(m^*)$ is the minimum of the cost function, $\Phi$ is the Obliquity, $\lambda$ is the Longitude of Perihelion and $e$ is the Eccentricity. . . . .	47

# Chapter 1

## Introduction

Monte Carlo inversion techniques were first used by Earth scientists more than 30 years ago as a method to estimate the parameters of computer models that simulate real, physical systems. Given randomly selected parameter values, the best computer model results were tested for their fit to the observed data and then the model was accepted or rejected to finally make predictions about the physical system of interest. As more computational power became available, Monte Carlo methods have shown to be important in the analysis of nonlinear inverse problems where simple gradient descent algorithms fail and multi-modality of the cost function results in multiple possible solutions.

Monte Carlo techniques can be divided as sampling methods and as optimization methods. Monte Carlo sampling is useful where calculus-based methods fail to search an optimal solution and characterize uncertainty. The Metropolis algorithm and the Gibbs sampler are the most widely used Monte Carlo samplers for this purpose. Monte Carlo optimization methods are powerful tools when searching for global optimal solutions amongst numerous local optima. Simulated annealing and genetic algorithms have shown their strengths in this respect. Some important areas where

## Chapter 1. Introduction

Monte Carlo inversion became feasible are problems in seismology in the latter part of the 1960s. Since that time, they have been applied to a wide range of problems such as earthquake epicenter location, Kozlovskaya (2000), atmospheric remote sensing, Haario et al. (2004), sea-borne radar applications, Yardim et al. (2006) and so forth.

One area of paramount importance for having more quantitative approaches to evaluating parametric uncertainties in Earth sciences is prediction of global warming. Models referenced by the Intergovernmental Panel on Climate Change Third Assessment Report (IPCC-TAR (2001)) predict that global temperatures are likely to increase (Figure 1.1) by 1.1 to 6.4°C (2.0 to 11.5°F) between 1990 and 2100. The uncertainty in this range comes from both the difficulty in predicting the amount of future greenhouse gas emissions and uncertainties regarding climate sensitivity. There has been limited progress in understanding and quantifying sources of this uncertainty. What has been done stems mainly from the analysis of multiple model responses to similarly applied forcings (e.g. Gates et al. (1999); Joussaume and Taylor (2000); Meehl et al. (2000)). The 2001 IPCC report, in its assessment of current research needs, calls for “a much more comprehensive and systematic system of model analysis and diagnosis, and a Monte Carlo approach to model uncertainties associated with parameterizations” (Section 8.10, McAvaney et al. (2001)). There has been some recent progress along these lines including work with models of reduced complexity (Forest et al., 2000, 2001, 2002) and perturbed physics ensembles with a general circulation model (Allen (1999); Murphy et al. (2004); Stainforth et al. (2005); Collins et al. (2006); Jackson et al. (2008)).

A large disparity exists among various climate models (Figure 1.2) in their prediction of global mean surface air temperature when atmospheric  $CO_2$  is doubled compared to present concentrations. There is an overwhelming number of reasons why these differences could exist. Although each climate model has been optimized to reproduce observational means, each model contains slightly different choices of model

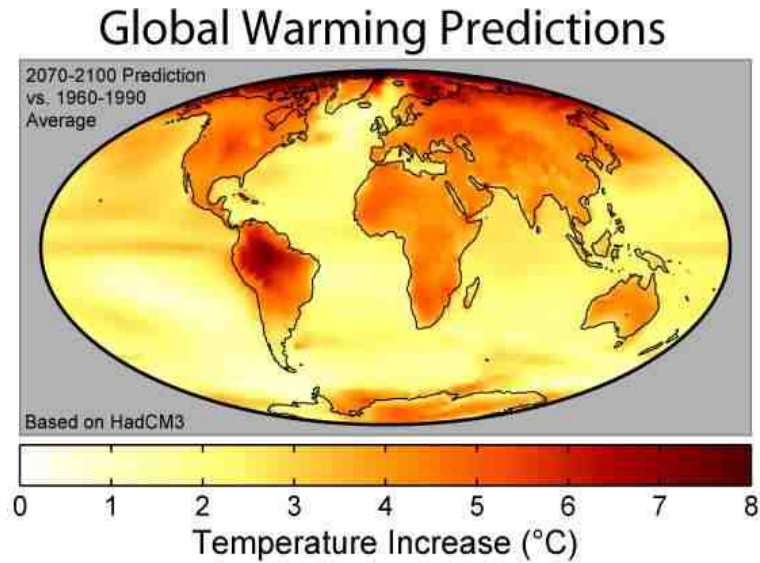


Figure 1.1: A map of predicted global warming at the end of the 21st century according to the HadCM3 climate model with a business-as-usual emissions scenario (IS92a). (Source: Global Warming Art (2007)).

parameter values as well as different parameterizations of under-resolved physics. Multi-model systems could be more reliable than single-model systems. In this matter, Tebaldi et al. (2005) propose a Bayesian statistical model that combines information from a multi-model ensemble of atmospheric ocean general circulation models (AOGCM) and observations to determine the probability distribution of future climate change. Barnett et al. (2006) use multiple versions of the HadAM3 GCM to quantify the uncertainty in changes in extreme event frequency in response to doubled  $CO_2$ . Collins et al. (2006) also compare multi-model ensembles of models from the IPCC-4AR (2007) with the predictions using the HadCM3 to quantify uncertainties in transient climate change using a perturbed physics approach in which modeling uncertainties are sampled systematically by perturbing uncertain parameters. Lopez et al. (2006) develop a Bayesian statistical model to produce probabilistic projections of regional climate change using observations and ensembles of GCMs.

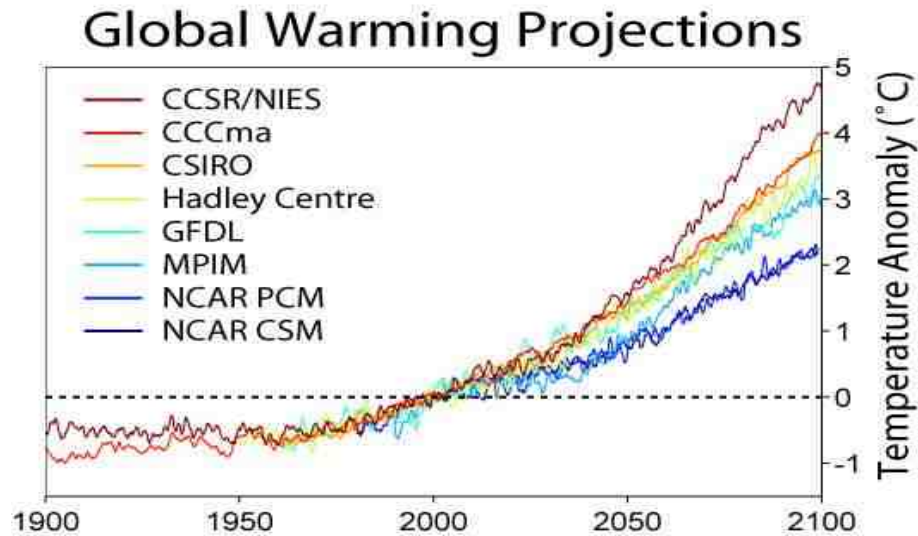


Figure 1.2: A comparison of predictions of temperature anomalies from 8 different climate models assuming the SRES A2 emissions scenario (Source: Global Warming Art (2007)).

Kettleborough et al. (2007) discuss a method for estimating uncertainty in future climate change using Monte Carlo Sampling.

A range of model hierarchies have been used to quantify the sources and impacts of climate modeling uncertainties: general circulation models, models of reduced complexity, and surrogate or emulator models. General circulation models are the most demanding computationally and simulate the detailed interactions among the atmospheric, oceanic, land surface, and sea ice components of the climate system and are usually developed by national model development centers such as the Hadley Center and their version 3 coupled Atmosphere-Ocean system (HadCM3) and the National Center for Atmospheric Research and their version 3 Community Climate System Model (CCSM3). As an example of the typical computational expense of these models, it takes 16 processors of a computational cluster 24 hours to simulate 10 years of climate. This expense has motivated some researchers to consider models of reduced

## *Chapter 1. Introduction*

complexity where one or more spatial dimensions of a climate model are eliminated (e.g. Forest et al., 2000, 2001, 2002). The present research uses a surrogate climate model that mimics the equilibrium space-time response of an Atmospheric GCM to changes in multiple model parameters from a set of previously run experiments to test different sampling strategies for quantifying parametric uncertainties.

The main goal of this thesis is to study how we may estimate probability distributions for parameters in climate models and assess which methods are more adequate for this purpose. However, the computational algorithms used in this work can be applied more generally to sample target distributions in any statistical inference problem. The methods used are in no way standard for the current state of the art within the climate literature. By applying adaptive methods we can approximate the posterior probability distribution (PPD) of the climate model parameters with few forward evaluations. The results obtained not only could be used to improve the calibration of a climate model, but also to test the strength of scientific inferences from observational data. Moreover, the strategically chosen samples could also serve as the basis for creating a statistical climate emulator model on which other, more standard MCMC sampling strategies could be used for generating accurate measures of the posterior distribution.

# Chapter 2

## A Physical Surrogate Climate Model

### 2.1 Introduction

Global warming by definition is the increase in the average temperature of the Earth's near-surface air and the oceans. The purpose of this chapter is to unfold the main elements to understand the relationship between the Earth's physical system and the change of the surface air temperature in the planet. Section 2.2 explains the theory that relates the orbital mechanics of the Earth with the variation in the surface air temperature. Section 2.3 describes a procedure to use the output from an Atmospheric General Circulation Model to build a surrogate climate model. Section 2.4 explains the details of the surrogate climate model, and finally Section 2.5 provides a description about the concept of cost function, one of the key elements to make inference about the parameters on the climate system.



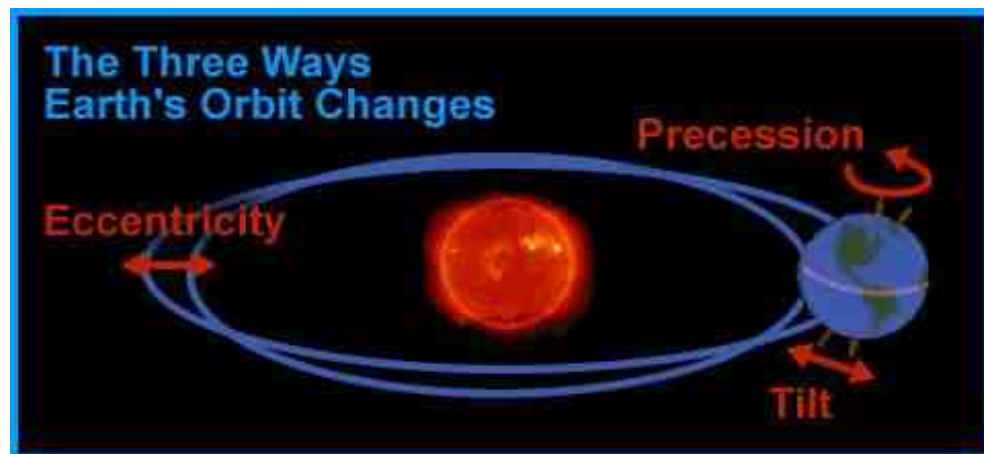


Figure 2.1: Earth's orbital geometry parameters: Obliquity (Earth's axial tilt), Eccentricity (how elliptical the Earth's orbit around the Sun is), and longitude of Perihelion (point of closest approach to the Sun). (Source: Windows for the Universe (UCAR) (2007)).

## 2.2 Milankovitch Cycles

Milankovitch (1941) proposed that variations in the Earth's orbit cause climate variability through a local thermodynamic response to changes in insolation. The Earth's orbital geometry parameters (obliquity, longitude of perihelion and eccentricity) are astronomical factors that influence the timing and intensity of the seasons. The properties of the solar forcing result from variations in the obliquity of the Earth's spin axis relative to the plane of the Earth's orbit about the Sun, precession of the Earth's spin axis, and the eccentricity (non-circularity) of the Earth's orbit (Figure 2.1). The obliquity varies on a time cycle of about 40,000 yrs. This changes the geographical distribution of insolation on both a seasonal and annual mean basis. The Earth's spin axis completes one precessional cycle in about 20,000 yrs. The precession effect acts to increase insolation during the season the Earth is at its closest approach to the Sun (the perihelion). Because insolation is greater for all latitudes

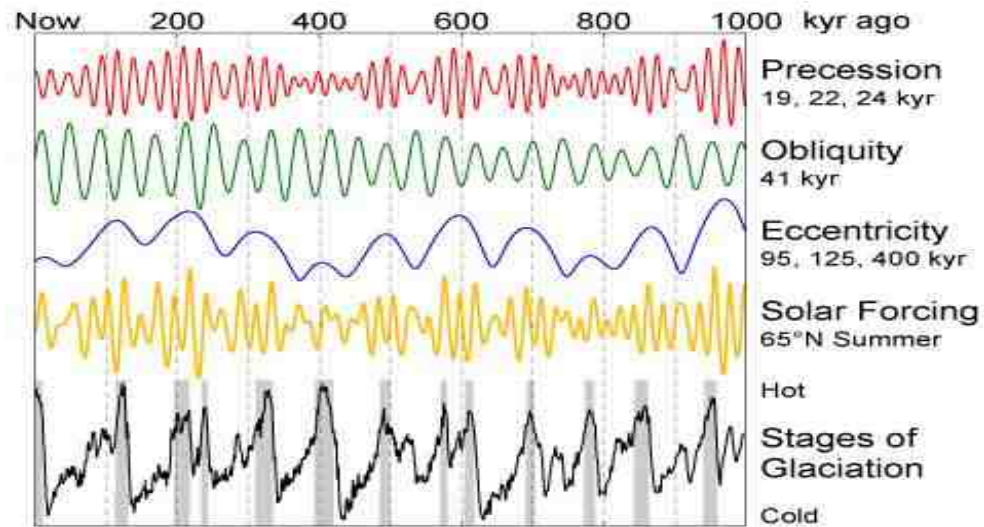


Figure 2.2: Variations in Earth's orbit, the resulting changes in solar energy flux at high latitude, and the observed glacial cycles (Source: Global Warming Art (2007)).

at perihelion, the precessional forcing is in phase globally for any given time of year. Unlike obliquity, precession does little to alter the geographical distribution of annual mean insolation. Variations in eccentricity, occurring on approximately 100,000 year time scales, has a small influence on annual mean insolation. Their main effect is to modulate the strength of the precessional forcing. According to Milankovitch theory, the precession of the equinoxes, variations in the tilt of the Earth's axis (obliquity) and changes in the eccentricity of the Earth's orbit are responsible for causing the observed 100,000 yrs cycle in ice ages by varying the amount of sunlight received by the Earth at different times and locations, particularly high northern latitude summer, (Figure 2.2).

## 2.3 Climate Model

Jackson and Broccoli (2003) taking advantage of the short equilibration time (10 yrs) of an atmospheric general circulation model (AGCM), land surface model and a static mixed-layer ocean model (i.e. including a thermodynamic model of sea ice), derive the equilibrium climate response to accelerated variations in Earth's orbital configuration over the past 165,000 yrs. More precisely, they estimate the sensitivity of each orbital component by fitting a time series of the evolution of each orbital component with the model output. They can estimate an amplitude of that component within the time series. This amplitude represents the sensitivity of the region and season to changes in that orbital component.

The sensitivity of surface air temperature to obliquity forcing  $A_{o,ijk}$  and precessional forcing  $A_{p,ijk}$  can be defined for particular latitudes  $i$ , longitudes  $j$ , and seasons  $k$ . They represent the climate model's response to the seasonally and latitude varying changes in insolation for a given unit change in orbital parameter values. They are derived from an ordinary multiple least squares fitting procedure between modeled variations in climate found within a climate model integration of the past 165,000 yrs forced only by changes in Earth's orbital geometry and two basis functions representing the known temporal variations in obliquity and precession. In particular, the obliquity basis function  $A_{o,ijk}\Phi'(t)$  consists of an unknown sensitivity  $A_{o,ijk}$  and the time series of obliquity variations  $\Phi'(t)$  over the past 165,000 yrs, where  $\Phi'(t) = \Phi(t) - \Phi_o$ , is the deviation of obliquity from its 165,000 yrs mean ( $\Phi_o = 23.3515^\circ$ ). The precessional basis function  $A_{p,ijk}e(t)\cos(\phi_{p,ijk} - \lambda(t))$  consists of an unknown sensitivity  $A_{p,ijk}$ , an unknown phase angle of response  $\phi_{p,ijk}$ , the time series of eccentricity  $e(t)$ , and the time series of the longitude of the perihelion  $\lambda(t)$ . The time series  $e(t)$ ,  $\lambda(t)$ , and  $\Phi'(t)$  are known from orbital mechanics and were used as input values in the AGCM which calculates the changes in insolation as a function of latitude and season for each year of the experiment (Figure 2.3).

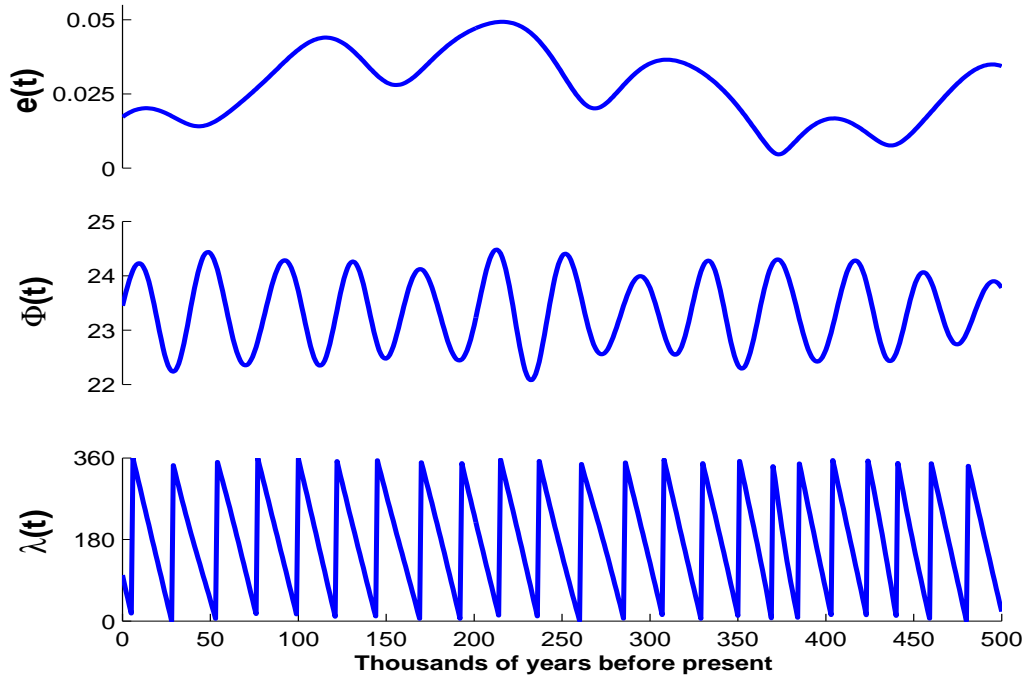


Figure 2.3: Eccentricity ( $e(t)$ ), Obliquity ( $\Phi(t)$ ), and longitude of Perihelion ( $\lambda(t)$ ) time series: 500,000 yrs before present.

The multiple least squares fitting procedure provides estimates of  $A_{o,ijk}$ ,  $A_{p,ijk}$ , and  $\phi_{p,ijk}$  that best represent the climate model's response to the time evolving changes in orbital forcing. For instance, the variations in surface air temperature with respect to the 165,000 yrs annual mean for a given region and season  $T_{ijk}(t)$  may be represented by,

$$T_{ijk}(t) = A_{o,ijk}\Phi'(t) + A_{p,ijk}e(t)\cos(\phi_{p,ijk} - \lambda(t)) + R_{ijk}(t), \quad (2.1)$$

where  $R_{ijk}(t)$  is a residual. The fitting procedure described above also allows one to construct a surrogate climate model using the estimated latitude, longitude, and seasonal obliquity and precessional forcing sensitivities. Figure 2.4 gives a comparison of the ability of the least squares fitting procedure with imposed time variations in

Earth's orbital geometry to reproduce the AGCM's response to the annual mean air temperature in Antarctica averaged from  $70^\circ$  S to  $90^\circ$  S and separated into its obliquity, precessional, and residual components. This is done by averaging together the sensitivities of all latitude, longitude, and seasons for this region and estimating the response by imposing the changes in the obliquity and precessional components.

## 2.4 Surrogate Climate Model

The surrogate model is based on surface air temperature fields generated by an AGCM in its response to changes in three parameters specifying Earth's orbital geometry over the past 165,000 yrs. The response can be approximated in terms of obliquity and precession components by using the multiple least squares fitting procedure described in the previous section. We will use this surrogate climate model to test sampling strategies as a follow-up study to Jackson et al. (2004), which consider the same surrogate model as in this research but mostly compares Multiple Very Fast Simulated Annealing (MVFSa) with the Metropolis/Gibbs and Grid search algorithms. One of the main goals of my thesis was to evaluate factors affecting the efficiency and accuracy of alternative sampling strategies to the MVFSa and Gibbs/Metropolis algorithms.

We denote the Earth's orbital geometry parameters and their physical range as obliquity,  $\Phi \in (22^\circ, 25^\circ)$ , eccentricity,  $e \in (0, 0.05)$  and longitude of perihelion,  $\lambda \in (0^\circ, 360^\circ)$ , therefore the dependency on  $t$  is omitted for the surrogate climate model. The observed data is a 3D array  $d_{obs,ijk}$  which represents the observed surface temperature anomalies with respect to the long term 165,000 yrs mean at latitude  $i$ , longitude  $j$ , and season  $k$ . The grid spacing is approximately  $4.5^\circ$  latitude by  $7.5^\circ$  longitude, then the latitude can take  $I = 40$  different values, and the longitude  $J = 48$ . The season takes  $K = 12$  values, which are selected days throughout the

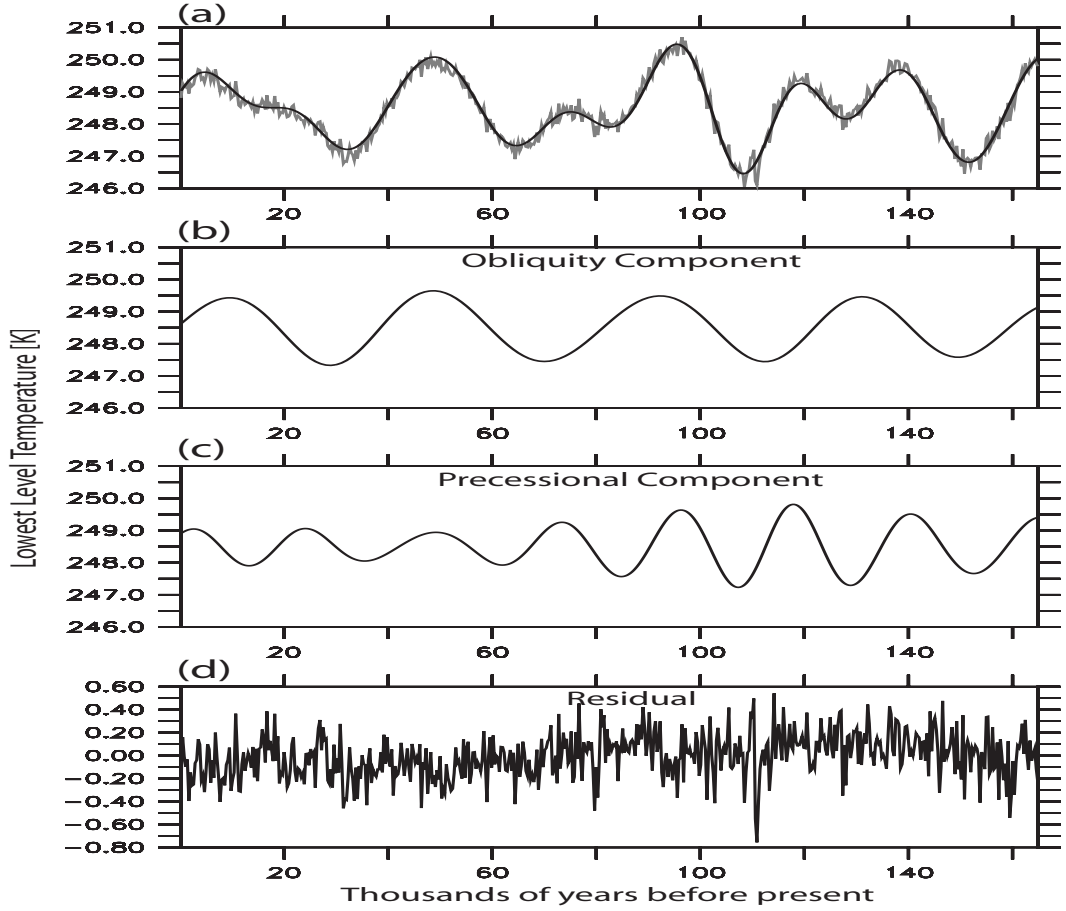


Figure 2.4: AGCM response of annual and zonal mean lowest level temperature (Kelvin degrees) from  $70^\circ$  S to  $90^\circ$  S (over Antarctica) to known continuous changes in Earth's orbital geometry for the past 165 thousand years ( $e(t)$ ,  $\lambda(t)$ , and  $\Phi'(t)$  are known time series). Panel (a) shows the model response (in gray) with least squares fitted solution (black curve) given by equation (1). The least squares fitted solution includes an obliquity component (b), and a precessional component (c). The residual between the least squares fitted solution and the AGCM is shown in panel (d).

year. Each value of  $k$  would apply for that season and for all time  $t$  over the past 165,000 years. The observed data are simulated using  $\Phi = 22.625$ ,  $e = 0.043954$ , and  $\lambda = 75.93$ , as ideal values for the climate model. We approximate the data using the relationship,  $d_{obs,ijk} = g_{ijk}(m) + \eta_{ijk}$ , where  $m = (\Phi, e, \lambda)$  is the vector of parameters,

$g$  represents the forward operator and it has the same dimensionality as  $d_{obs,ijk}$ . The definition of the function  $g$  is crucial since it is completely defined by the physical system. The term  $\eta_{ijk}$  is a Gaussian error with estimated variance given by  $B_{ijk}$ ; this array represents the variance of the observations at each grid point. This variability comes from the 1,500 year integration of the model itself, but with the appropriate seasonal and climatological averages (i.e. 10 year means of particular seasons). Typically, in Earth science models the observational uncertainties are assumed as Gaussian, see Jackson et al. (2004), Tebaldi et al. (2005), and Lopez et al. (2006).

In this thesis, the surface air temperature anomaly to a given change in the three parameters that define the Earth's orbital geometry is  $g_{ijk}(m)$ . The surrogate climate model is defined as follows,

$$g_{ijk}(m) = \hat{A}_{o,ijk}\Phi' + e\hat{A}_{p,ijk}\cos(\hat{\phi}_{p,ijk} - \lambda) + \hat{R}_{ijk}, \quad (2.2)$$

where  $\hat{\phi}_{p,ijk}$  is the phase of the response to precessional forcing and  $\hat{R}_{ijk}$  are the residuals averaged over time obtained from the AGCM in (2.1). This term is added to represent the effects of internal variability on 10 year seasonal means. Repeated experiments of the climate model will cycle through 1 of 150 possible values of  $\hat{R}_{ijk}$  that come from a 1,500 year long control integration of the AGCM.  $\hat{A}_{o,ijk}$  and  $\hat{A}_{p,ijk}$  are the sensitivity of temperature to changes in obliquity and precession obtained using the time series fitting procedure in (2.1).

## 2.5 The Cost Function

The cost function or misfit function is a measure of the deviation generated from the observed data ( $d_{obs}$ ) and the data generated from the model ( $g(m)$ ). In general, the cost function can be represented as  $E(m) = ||d_{obs} - g(m)||$ , where  $m$  is any given vector of parameters of interest from the physical system, and the difference between

the data and the model is given by a specific metric. The cost function can be defined in many ways. For instance, in Mu et al. (2003),

$$E(m) \approx \frac{1}{2N} \sum_{i=1}^N \left( \sum_{j=1}^K \frac{a_j^2}{\lambda_j^2} \right)_i,$$

where  $a_j$  are the coefficients of a series of empirical orthogonal functions that reconstruct modeled-observational differences and  $\lambda_j$  is the variance accounted for by the  $j$ th component, in other fields this is known as principal component analysis (PCA).

Wang (2007) uses the following cost function to investigate El Nino Southern Oscillation (ENSO) sensitivity,

$$E(m) = \frac{1}{5} \left\{ \frac{(var_{mod} - var_{obs})^2}{\sigma_{var}} + \frac{(skw_{mod} - skw_{obs})^2}{\sigma_{skw}} + \frac{(kur_{mod} - kur_{obs})^2}{\sigma_{kur}} + \frac{1}{6} \sum_{\tau} \frac{(cor_{mod}(\tau) - cor_{obs}(\tau))^2}{\sigma_{cor}(\tau)} \right\},$$

where  $m = (var_{mod}, skw_{mod}, kur_{mod}, cor_{mod})$ , here  $var$ ,  $skw$ ,  $kur$ , and  $cor$  are variance, skewness, kurtosis, and autocorrelation of the time series;  $\sigma$  is the data uncertainty and  $\tau$  is the time lag. This cost function is defined as a weighted average of the mean square error of multiple characteristics of ENSO statistics.

In the surrogate climate model considered here, the cost function is defined as,

$$E(m) = \frac{1}{2} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K B_{ijk}^{-1} (d_{obs,ijk} - g_{ijk}(m))^2, \quad (2.3)$$

where  $m = (\Phi, e, \lambda)$  is the vector of Earth's orbital geometry parameters (obliquity, eccentricity, and longitude of perihelion). On the climate model studied here there is just one field, surface air temperature anomalies, however we can have  $N$  different sets of observations, such as seasonal and annual mean surface air temperature, precipitation, winds, and clouds at different latitudes. It is proposed that, the likelihood function takes the form,

$$L(d_{obs}|m, S) \propto \exp\{-SE(m)\}. \quad (2.4)$$



## Chapter 2. A Physical Surrogate Climate Model

The parameter  $S$  is connected to  $B_{ijk}$  according to Jackson et al. (2004) as a scaling factor.  $S$  performs the function of weighing the significance of model-data differences. Large values of  $S$  would imply small errors between the data and the model and would result in highly peaked probability distributions. To illustrate this, we fixed  $S = 47$  based on the expertise and initial knowledge given in Jackson et al. (2004) as an appropriate value for this parameter. To gain some initial insight of the likelihood function for the surrogate climate model, we can plot the profile likelihood for each parameter. Since we already know the optima values for the simulation study about this model ( $\Phi = 22.625$ ,  $e = 0.043954$ ,  $\lambda = 75.93$ ), we fix two parameters on their optimum value and we evaluate the third one using a 20,000 point grid evaluation. Figure 2.5 shows the profile likelihoods for each parameter,  $\Phi$ ,  $e$  and  $\lambda$  for two different values of  $S$ . Due to internal variability, the climate model can take on a range of likelihood values for any given combination of orbital parameter values. This scatter from internal variability (noise term in (2.2)) is seen within Figure 2.5 as the thin vertical lines that follow the broader scale variations in likelihood values. These broad scale variations reflect the smoothly evolving changes in climate that accompany changes in orbital geometry.

Chapter 2. A Physical Surrogate Climate Model

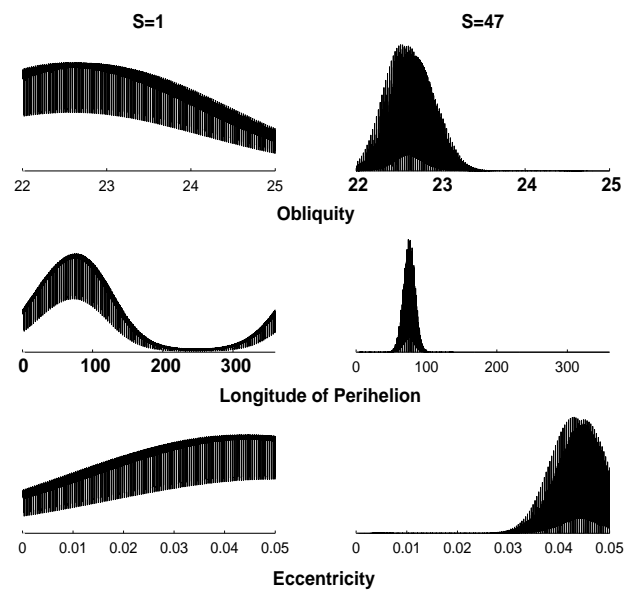


Figure 2.5: Likelihood for orbital forcing parameters. Left column:  $S=1$ . Right column:  $S=47$ .

# Chapter 3

## Monte Carlo Methods

### 3.1 Introduction

This chapter is a review of some Monte Carlo (MC) and Markov Chain Monte Carlo (MCMC) algorithms used to get samples from target distributions. To illustrate the different methods, we consider an example of a multi-modal univariate probability density which is challenging to sample using standard approaches. The target distribution is a mixture of univariate normal distributions with different means, variances, and component weights. The target density function is defined by,

$$\begin{aligned}\pi(m) = & \left(\frac{3}{9}\right) \frac{1}{\sqrt{(25)(2\pi)}} \exp\left\{-\frac{1}{2(25)}(m+45)^2\right\} + \\ & \left(\frac{1}{9}\right) \frac{1}{\sqrt{(9)(2\pi)}} \exp\left\{-\frac{1}{2(9)}(m-2)^2\right\} + \\ & \left(\frac{5}{9}\right) \frac{1}{\sqrt{(4)(2\pi)}} \exp\left\{-\frac{1}{2(4)}(m-30)^2\right\}.\end{aligned}\tag{3.1}$$

The mixture weights  $(\frac{3}{9} + \frac{1}{9} + \frac{5}{9} = 1)$  must add up to 1 so  $\pi(m)$  is a probability density function (pdf). Figure 3.1 shows a graph of  $\pi(m)$ . We can notice that  $\pi(m)$  has two

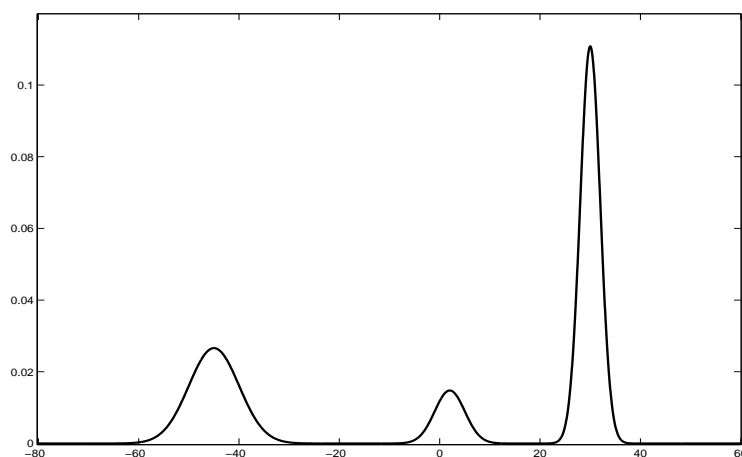


Figure 3.1: Target distribution with 2 local modes and 1 global mode:  $\pi(m) = \frac{3}{9}N(-45, 25) + \frac{1}{9}N(2, 9) + \frac{5}{9}N(30, 4)$

local modes and one global mode. Numerical methods used to find the global mode like the Newton-Raphson algorithm can get easily stuck at a local mode, in this kind of situation.

The cost or misfit function is defined through the relationship  $\pi(m|\cdot) \propto \exp\{-E(m)\}$ , where  $E(m)$  is the cost function. In the case of a univariate Normal distribution with mean zero and variance equal to 4, the cost function is  $E(m) = -\frac{1}{2(4)}m^2$ . In the case of a mixture of Normals,  $E(m) = -\log(\pi(m))$  since we cannot factor any of the terms due to the different weights and variances of the mixture components.

## 3.2 Gibbs and Metropolis samplers

Geman and Geman (1984) consider that a Gibbs distribution  $\pi$  can be uniquely

determined by

$$\pi(x_s|x_r, r \neq s) = \frac{\pi(\omega)}{\sum_{x_s \in \Delta} \pi(\omega)} \quad s \in \mathbf{S}, \quad \omega \in \Omega.$$

Where  $\mathbf{S} = \{s_1, \dots, s_N\}$  is a set of sites and  $\Omega = \{\omega = (x_{s_1}, \dots, x_{s_N}) \mid x_{s_i} \in \Delta, \quad 1 \leq i \leq N\}$  is the set of all possible configurations. In geosciences, Sambridge and Mosegaard (2002), use the right side of this formula to define the Gibbs sampler algorithm while Gelfand and Smith (1990) proposed to use the left side to iteratively sample the full conditional distributions of each parameter. The Gibbs sampler is a version of an importance sampling technique that improves the efficiency of the calculation by sampling model parameters sets from the Gibbs distribution which is, in effect, equivalent to the desired posterior probability distribution (PPD). This approach requires the parameter space to be subdivided into a number of equally spaced intervals.

According to Sen and Stoffa (1996), the probability distribution function of a parameter  $m$  can be approximated via the Gibbs Sampler algorithm as follows,

- First we define,

$$P(m = m_j) = \frac{\exp\{\frac{1}{T}E(m_j)\}}{\sum_{j=1}^N \exp\{\frac{1}{T}E(m_j)\}}, \quad j = 1, \dots, N, \quad (3.2)$$

where  $N$  is the number of values that the parameter  $m$  can take, so  $m$  is being discretized.  $T$  is a temperature parameter. A canonical value is  $T = 1$ .  $E(m)$  is the misfit function or cost function.

- A value is drawn from (3.2) and is always retained.
- This procedure is repeated for the  $N$  values that  $m$  can take. ( $N$  forward evaluations).

Figure 3.2 shows that this version of the Gibbs algorithm can work very well in a univariate case, but is also extremely inefficient since the grid defined on the param-

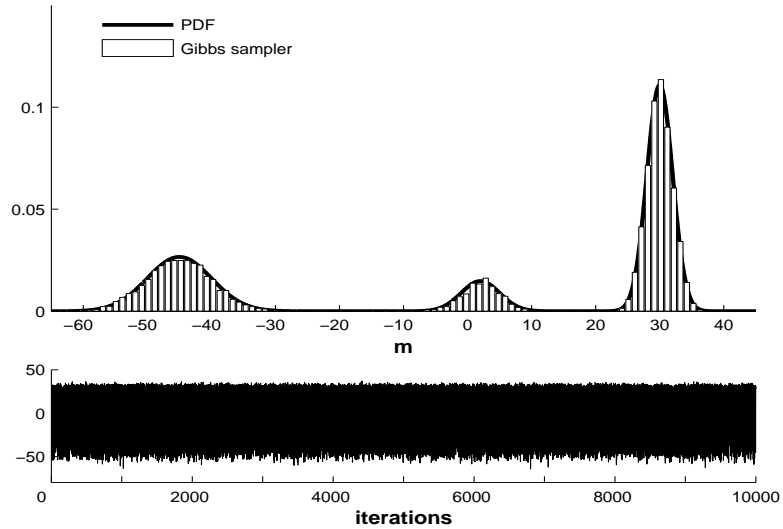


Figure 3.2: Gibbs sampler applied to the target distribution  $\pi(m) = \frac{3}{9}N(-45, 25) + \frac{1}{9}N(2, 9) + \frac{5}{9}N(30, 4)$ . Top: The black solid line represents the actual pdf, and the white bars, the plot of the histogram. Bottom: Trace plot of iterations.

eter space required 10,000 equally spaced points. The selected  $N$  values of  $m_j$  are completely arbitrary and chosen before the algorithm is implemented. This procedure is different from the statistical version proposed by Gelfand and Smith (1990), which can be used only in problems of dimensionality greater or equal than 2.

The Metropolis-Hastings (M-H) algorithm (Hastings (1970)) is a variation of the Metropolis scheme (Metropolis et al. (1953)), that requires a probability function  $q$  as a proposal. This proposal or jump distribution affects the way in which new models are accepted. The rule is to accept a new model  $m^{(k+1)}$  with probability,

$$\alpha(m^{(k)}, m^{(k+1)}) = \min\left(1, \frac{\pi(m^{(k+1)})q(m^{(k+1)}, m^{(k)})}{\pi(m^{(k)})q(m^{(k)}, m^{(k+1)})}\right).$$

We applied the M-H algorithm with a random walk proposal distribution to the target distribution  $\pi(m)$  of (3.1). The algorithm steps are,

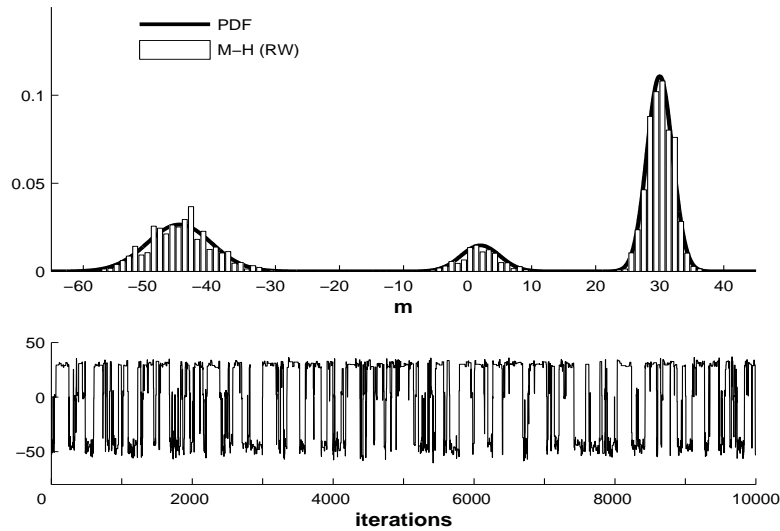


Figure 3.3: Metropolis-Hastings algorithm applied to the target distribution  $\pi(m) = \frac{3}{9}N(-45, 25) + \frac{1}{9}N(2, 9) + \frac{5}{9}N(30, 4)$ , using as proposal distribution a  $N(m^{(k-1)}, V)$ . Top: The black solid line represents the actual pdf, and the white bars, the plot of the histogram. Bottom: Trace plot of iterations.

- Set an initial value  $m^{(0)}$ .
- At iteration  $k + 1$  sample a possible new value  $y$  for  $m^{(k+1)}$  from a Normal distribution with mean  $m^{(k-1)}$  and variance  $V$ . We will use the notation  $q(\cdot)$  to denote the proposal distribution for the new possible values of  $m$ .
- We accept  $y$  with probability  $\alpha(m^{(k)}, m^{(k+1)})$ .

Figure 3.3 shows that the M-H algorithm has a good performance, but one of its main drawbacks is setting the value of  $V$ , the variance of the proposal distribution must take. In order to get good results this value must be tuned carefully. For the mixture of Normals example,  $V$  was set equal to 1000 to allow efficient sampling from the parameter space. The acceptance rate resulted in a value around 40%.

Another version of the Metropolis algorithm proposed by Sen and Stoffa (1996) is the Multiple Metropolis Simulated Annealing (METSAs). This method is started from several independent initial points to improve the sampling of  $\pi$ . At every point, candidates are drawn at random. The acceptance/rejection rule is to accept a new value with probability  $\min(1, \exp(-(E(m^{(k+1)}) - E(m^{(k)}))/T))$ . Adding  $T$  to the algorithm allows to sample regions of the parameter space with high density.

### 3.3 Multiple Very Fast Simulated Annealing (MVFSAs)

One may use the temperature parameter within the Metropolis algorithm to take advantage of the well known features of stochastic optimizers from Simulated Annealing (Kirkpatrick et al. (1983)) and the Very Fast Simulated Annealing (Ingber (1989)) to locate the global minimum of the cost function  $E(m)$  by very slowly lowering the temperature parameter. Ingber (1989) propose the selection of model parameters given a current selection  $m_i^{(k)}$  within VFSA so that  $m_i^{(k+1)} = m_i^{(k)} + y_i(m_i^{max} - m_i^{min})$ ,  $y_i$  is generated according to a Cauchy distribution which is dependent on the cooling schedule at iteration  $k$ ,  $T_k = T_o \exp(-\alpha(k - 1)^{1/d})$ . The scalars  $m_i^{max}$  and  $m_i^{min}$  are the maximum and minimum values that the parameter  $i$  can take. The acceptance criterion for successive model selections is the same as for the Metropolis algorithm. Also Sen and Stoffa (1996) and Jackson et al. (2008) argue that one may allow for numerous repetitions of the minimization procedure to strike a balance between estimating a multidimensional posterior probability density (PPD) and finding the global minimum. This is known as the MVFSAs algorithm and it has the following features:

- **Optimization:** Take advantage of the well known characteristics of stochastic



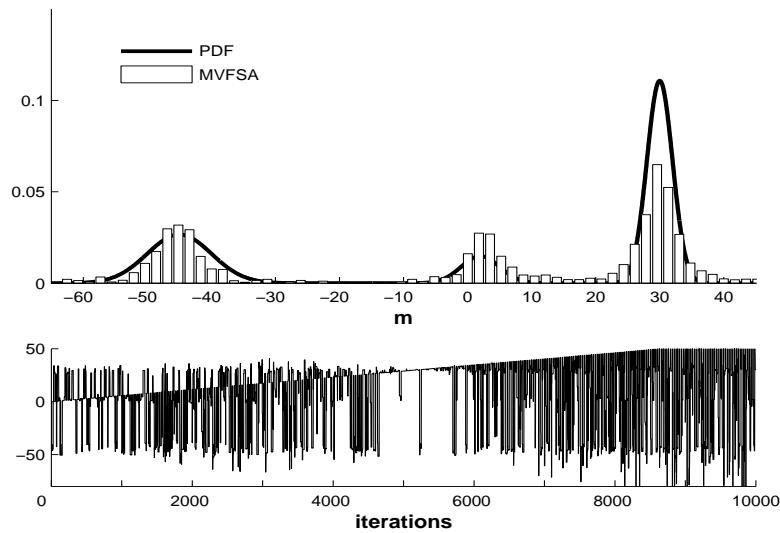


Figure 3.4: Multiple Very Fast Simulated Annealing algorithm applied to the target distribution  $\pi(m) = \frac{3}{9}N(-45, 25) + \frac{1}{9}N(2, 9) + \frac{5}{9}N(30, 4)$ , using 400 independent initial points. Top: The black solid line represents the actual pdf, and the white bars, the plot of the histogram. Bottom: Trace plot of iterations.

optimizers from Simulated Annealing (SA) and VFSA.

- **Multiple:** Instead of using a single initial value as SA or VFSA, we can use “multiple” initial values over the parameter space.
- **Sampling:** For each single initial value, propose a new candidate using a random walk combined with a random number sampled from a Cauchy distribution that is dependent on the cooling schedule. Acceptance of the proposed value is similar to a Metropolis step.
- **Flexibility:** Two additional parameters such as *moves/temperature* and *ntarget* add flexibility to the VFSA algorithm to control sampling efficiency for a specified number of dimensions.

In our example, MVFSA follows the next steps:

- Set an initial value  $m^{(0)}$ .
- Set  $m^{min}$  as the minimum value that  $m$  can take and  $m^{max}$  as the maximum value that  $m$  can take. Therefore  $m^{min} \leq m^{(k)} \leq m^{max}$ .
- Compute  $m^{(k+1)} = m^{(k)} + y(m^{max} - m^{min})$ , where  $y$  is drawn from a Cauchy distribution. The realization of  $y$  is obtained using the next formula

$$y = \text{sign}(u - \frac{1}{2})T_k \left( \left(1 + \frac{1}{T_k}\right)^{|2u-1|} - 1 \right)$$

where  $u$  is a random number from a Uniform distribution over  $(0, 1)$  and  $T_k$  is the temperature parameter and is defined by  $T_k = T_0 e^{-\alpha(k-1)^{1/d}}$ , where  $T_0 = 1$ ,  $\alpha = 0.9$ , and  $d = 2$ .

- Define *moves/temperature* as the number of times that  $m$  can change before lowering the temperature  $T$ .
- Evaluate  $E(m^{(k+1)})$
- Evaluate  $\Delta E = E(m^{(k+1)}) - E(m^{(k)})$  and if  $\Delta E \leq 0$  then accept  $m^{(k+1)}$  as a new value for  $m$ . If  $\Delta E > 0$ , then accept  $m^{(k+1)}$  as a new value with probability  $\min(1, \exp\{-\frac{\Delta E}{T}\})$ .
- Define *ntarget* as the maximum number of failed attempts at finding an acceptable set of values for  $m$  before stopping.
- Repeat the previous steps for every single point in the grid defined over the parameter space. This step is what turns the VFSA into a MVFSA.

As we can see in Figure 3.4, although MVFSA sampling is based on stochastic optimization, the PPD derived through the MVFSA algorithm is broader than what

may be obtained through the Metropolis/Gibbs sampler as a result of the multiple independent initial values over the parameter space. In order to get these results, we set *moves/temperature* equal to 5, *ntarget* equal to 25 and the number of points on the grid equal to 400. The unusual behavior in the trace plot is explained by the fact that MVFSA is not an MCMC method since its design uses multiple independent initial points over the parameter space.

### 3.4 Adaptive Methods

One of the issues with the Metropolis or the Metropolis-Hastings algorithm is the choice or tuning of an effective proposal distribution to keep acceptance rates at a 30 – 40% level, Figure 3.5 shows samples from a M-H using a random walk proposal with variance  $V = 1$ . Alternatively Haario et al. (1999) suggested a method called Adaptive Proposal (AP) that basically updates the proposal distribution with the knowledge we have so far learned about the target distribution. Furthermore Haario et al. (2001) proposed a variation of the AP algorithm called Adaptive Metropolis (AM) which is a non-Markovian algorithm that has the correct ergodic properties.

The AM has two versions, one called Single Component Adaptive Metropolis (SCAM) and the full component version (FAM). Haario et al. (2004) applied the SCAM algorithm to a 90 dimension inversion problem about gas profiles (GOMOS) in a successful way. Another powerful variant is called Delayed Rejection Adaptive Metropolis (DRAM) that combines the Delayed Rejection scheme (Tierney and Mira (1999)) with the AM.

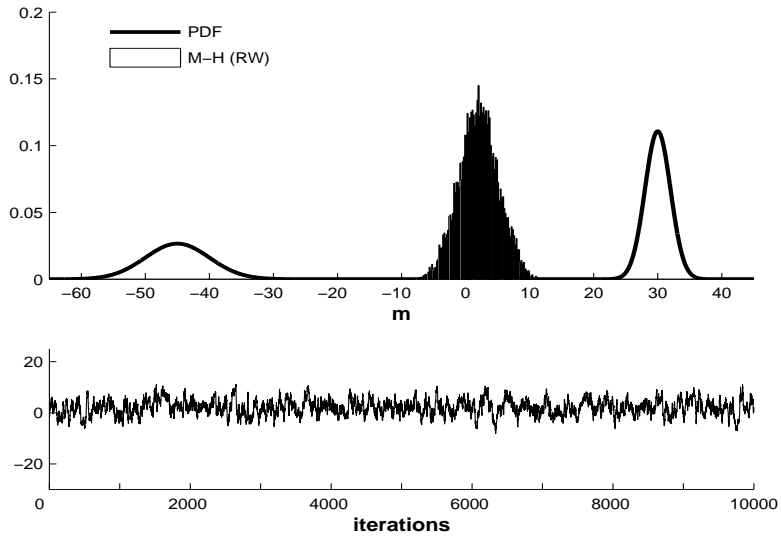


Figure 3.5: Metropolis-Hastings algorithm applied to the target distribution  $\pi(m) = \frac{3}{9}N(-45, 25) + \frac{1}{9}N(2, 9) + \frac{5}{9}N(30, 4)$ , using as proposal distribution a  $N(m^{(k-1)}, 1)$ . Top: The black solid line represents the actual pdf, and the white bars, the plot of the histogram. Bottom: Trace plot of iterations.

### 3.4.1 SCAM

Let  $\pi$  denote the density of our target distribution, typically a PPD which we can evaluate up to a normalizing constant. The sequence  $m^{(0)}, m^{(1)}, \dots$  denotes the full states of the process, that is, we consider a new state updated as soon as all the  $d$  components of the state have been separately updated. We denote  $m_{i,k-1} = (m_i^{(0)}, \dots, m_i^{(k-1)})$  as the sampled vector for the  $i$ -th parameter up to state  $k-1$ . The adaptive scheme is done through the variance equation  $V_i^{(k)} = s_d V(m_{i,k-1}) + s_d \epsilon$ . Where  $V(m_{i,k-1}) = \frac{1}{k-1} \sum_{r=0}^{k-1} (m_i^{(r)} - \bar{m}_i)^2$ ,  $\bar{m}_i = \frac{1}{k} \sum_{r=0}^{k-1} m_i^{(r)}$ ,  $s_d$  is a positive constant and  $0 < \epsilon < 1$ . When updating on the  $i$ -th parameter  $m_i^{(k)}$  at state  $k$ , we apply a one dimensional Metropolis step:

- Sample  $z_i \sim N(m_i^{(k-1)}, V_i^{(k)})$

- Accept the candidate point  $z_i$  with probability

$$\min\left(1, \frac{\pi(m_1^{(k)}, \dots, m_{i-1}^{(k)}, z_i, m_{i+1}^{(k-1)}, \dots, m_d^{(k-1)})}{\pi(m_1^{(k)}, \dots, m_{i-1}^{(k)}, m_i^{(k-1)}, \dots, m_d^{(k-1)})}\right)$$

in which case we set  $m_i^{(k)} = z_i$ , and otherwise  $m_i^{(k)} = m_i^{(k-1)}$ .

When parameters of the target function are highly correlated the SCAM can suffer from poor mixing as well as the Metropolis algorithm. A remedy to this problem is to rotate the proposal distribution during the early stage of the sampling (i.e. during the burn-in period). The rotation can be done by computing the covariance matrix of the chain so far detected and then computing the eigenvalues of the covariance matrix. We sort the eigenvalues from the largest to the smallest one and we use this order as sampling directions for the parameters in the SCAM algorithm.

### 3.4.2 FAM

Suppose that at time  $t - 1$  we have sampled the states  $m^{(0)}, \dots, m^{(t-1)}$  where  $m^{(0)}$  is the initial state and is a vector of dimension  $d$ . Then a candidate point  $z$  is sampled from the proposal distribution  $q_t(\cdot | m^{(0)}, \dots, m^{(t-1)})$ , which now may depend on the whole history. The candidate  $z$  is accepted with probability,

$$\alpha(m^{(t-1)}, z) = \min\left(1, \frac{\pi(z)}{\pi(m^{(t-1)})}\right),$$

in which case we set  $m^{(t)} = z$ , and otherwise  $m^{(t)} = m^{(t-1)}$ .

The proposal distribution  $q_t(\cdot | m^{(0)}, \dots, m^{(t-1)})$  employed in the FAM algorithm is a multivariate Gaussian distribution with mean at the current point  $m^{(t-1)}$  and covariance matrix  $C_t$ . The matrix  $C_t$  is computed using the sampled covariance matrix of the parameters up to time  $t$ . The crucial aspect regarding the adaptation is how the covariance of the proposal distribution depends on the history of the chain.

In the FAM algorithm this is solved by setting  $C_t = s_d C_{t-1} + s_d \epsilon I_d$  after an initial period, where  $s_d$  is a positive constant that depends only on the dimension  $d$  and  $\epsilon > 0$  is a constant that we may choose very small. The role of  $\epsilon$  is to ensure that  $C_t$  does not become singular. Let us define the matrix  $\mathbf{M}_t$  with dimensions  $d$  by  $t$  as the matrix of sampled values up to  $t$ .

$$\mathbf{M}_t = \begin{pmatrix} m_1^{(1)} & m_1^{(2)} & \dots & m_1^{(t)} \\ \vdots & \vdots & \vdots & \vdots \\ m_d^{(1)} & m_d^{(2)} & \dots & m_d^{(t)} \end{pmatrix}$$

We select an initial time  $t_0$  for the length of an initial period and define

$$C_t = \begin{cases} C_0, & \text{if } t \leq t_0 \\ s_d C_{t-1} + s_d \epsilon I_d, & \text{if } t > t_0 \end{cases}$$

We can compute the covariance matrix for time  $t \leq t_0$  as

$$C_t = \frac{1}{t-1} \mathbf{M}_t \left( I_t - \frac{1}{t} \mathbf{1}_t \mathbf{1}_t' \right) \mathbf{M}_t',$$

where  $I_t$  is an identity matrix of size  $t$ , and  $\mathbf{1}_t$  is a row vector of ones with length  $t$ . To avoid too much computational cost we can use recursive formulas for the mean and the covariance. Then we can easily define the recursion for the vector of means as

$$\bar{m}_t = \frac{t-1}{t} \bar{m}_{t-1} + \frac{1}{t} m^{(t)}$$

and for  $t > t_0$ . The recursion for the covariance matrix is,

$$C_t = \frac{t-2}{t-1} C_{t-1} + \bar{m}_{t-1} \bar{m}_{t-1}' + \frac{1}{t-1} \left[ m^{(t)} m^{(t)'} - t \bar{m}_t \bar{m}_t' \right]$$

As a basic choice for the scaling factor  $s_d$  we can adopt the value  $s_d = (2.4)^2/d$  as in Gelman et al. (1996). In contrast to the SCAM algorithm all parameters are sampled at once in the FAM scheme. The AM chain defined above simulates properly

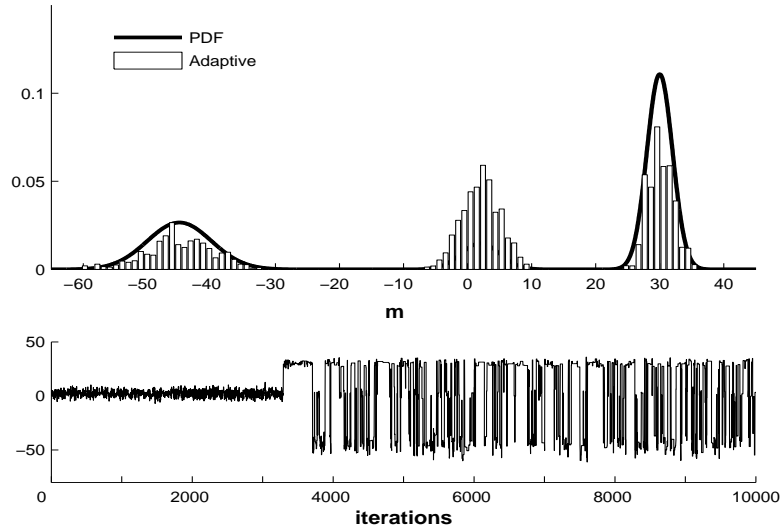


Figure 3.6: Adaptive Metropolis algorithm applied to the target distribution  $\pi(m) = \frac{3}{9}N(-45, 25) + \frac{1}{9}N(2, 9) + \frac{5}{9}N(30, 4)$ , using as initial proposal distribution a  $N(m^{(t-1)}, 1)$ . Top: The black solid line represents the actual pdf, and the white bars, the plot of the histogram. Bottom: Trace plot of iterations.

the target distribution  $\pi$ : for any bounded and measurable function  $f : S \rightarrow \mathfrak{R}$ , it holds almost surely that

$$\lim_{n \rightarrow \infty} \frac{1}{n+1} (f(m_0) + f(m_1) + \dots + f(m_n)) = \int_S f(m) \pi(dm).$$

For a detailed proof of this result, see Haario et al. (2001). To illustrate how the adaptive Metropolis works, we show a couple of examples: (1) Figure 3.6 uses a random walk proposal with initial variance  $V = 1$ , and while the results are not completely accurate, we can say that it does a better job estimating the target distribution than a traditional M-H. (2) Figure 3.7 shows how a little increase in the variance of the proposal distribution ( $V = 2$ ) improves a lot the results. In both Figures we did not use a burn-in period to emphasize that the adaptation starts from the very beginning.

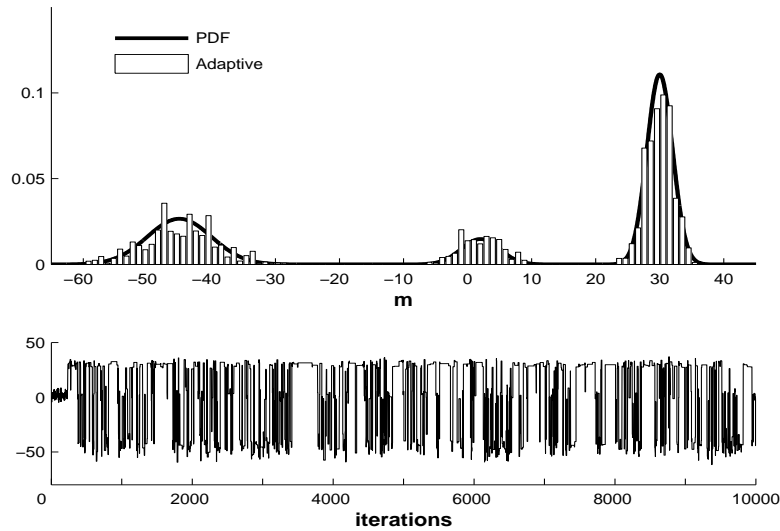


Figure 3.7: Adaptive Metropolis algorithm applied to the target distribution  $\pi(m) = \frac{3}{9}N(-45, 25) + \frac{1}{9}N(2, 9) + \frac{5}{9}N(30, 4)$ , using as initial proposal distribution a  $N(m^{(t-1)}, 2)$ . Top: The black solid line represents the actual pdf, and the white bars the plot of the histogram. Bottom: Trace plot of iterations.

### 3.4.3 DRAM

Haario et al. (2006) combine the ideas of Delayed Rejection (DR) and Adaptive Metropolis to improve the efficiency of MCMC algorithms. The basic idea of DR is that upon rejection in a M-H, instead of advancing time and retaining the current position, a second state move is proposed. The acceptance probability of the second state candidate is computed so that reversibility of the Markov chain relative to the distribution of interest is preserved. The process of delaying rejection can be iterated for a fixed or random number of stages. The DR can be also considered as a way of combining different proposals to allow the sampler to explore the parameter space more efficiently.

Suppose the current position of the Markov chain is  $m^{(t-1)}$ . A new candidate  $z_1$



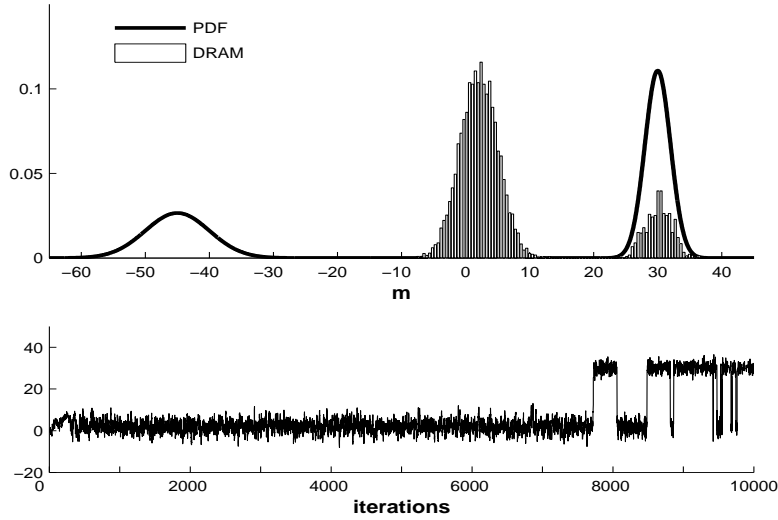


Figure 3.8: DRAM algorithm applied to the target distribution  $\pi(m) = \frac{3}{9}N(-45, 25) + \frac{1}{9}N(2, 9) + \frac{5}{9}N(30, 4)$ . Strategy used: First and second proposal distributions use same variance  $V = 1$ . Top: The black solid line represents the actual pdf, and the white bars, the plot of the histogram. Bottom: Trace plot of iterations.

is generated from a proposal  $q_1(m^{(t-1)}, \cdot)$  and accepted with probability

$$\alpha_1(m^{(t-1)}, z_1) = \min\left(1, \frac{\pi(z_1)q_1(z_1, m^{(t-1)})}{\pi(m^{(t-1)})q_1(m^{(t-1)}, z_1)}\right).$$

Upon rejection, instead of retaining  $m^{(t)} = m^{(t-1)}$ , as we would do in a standard M-H, a second state move  $z_2$  is proposed. The second state proposal is allowed to depend not only on the current position of the chain but also on what we have just proposed and rejected  $q_2(m^{(t-1)}, z_1, \cdot)$ . The second state proposal value  $z_2$  is accepted with probability

$$\alpha_2(m^{(t-1)}, z_1, z_2) = \min\left(1, \frac{\pi(z_2)q_1(z_2, z_1)q_2(z_2, z_1, m^{(t-1)})(1 - \alpha_1(z_2, z_1))}{\pi(m^{(t-1)})q_1(m^{(t-1)}, z_1)q_2(m^{(t-1)}, z_1, z_2)(1 - \alpha_1(m^{(t-1)}, z_1))}\right).$$

There is a number of different strategies to combine DR and AM. In order to be efficient, the proposal distribution must somehow be tuned to the shape and size of

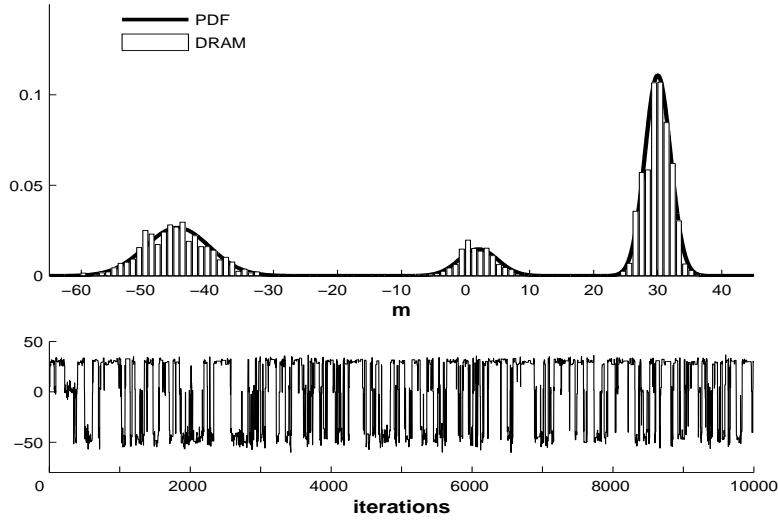


Figure 3.9: DRAM algorithm applied to the target distribution  $\pi(m) = \frac{3}{9}N(-45, 25) + \frac{1}{9}N(2, 9) + \frac{5}{9}N(30, 4)$ . Strategy used: the variance on the second stage proposal is a factor of the variance of the first stage proposal,  $V_2 = 0.1 * V_1$ , where  $V_1 = 10$ . Top: The black solid line represents the actual pdf, and the white bars, the plot of the histogram. Bottom: Trace plot of iterations.

the target distribution. One solution is to use adaptation only for the burn-in period and discard the part of the chain where adaptation has been used. In that respect, the adaptation can be thought as an automatic burn-in. The idea of diminishing adaptation is that when adaptation works well, its effect gets smaller and we might be able to prove the ergodicity properties of the chain even when adaptation is used throughout the whole simulation. This is the idea behind AM adaptation. On the other hand, the DR method allows us the use of the current rejected values without losing the Markovian property and thus allows to adapt locally to the current location of the target distribution. As an example of two different strategies to use DRAM we will use DR on the first 200 iterations and we will update the variance of the proposals after this period and every 100th iteration using the AM. Figure 3.8 is using the first stage and second stage proposals with initial variance  $V = 1$ . In

Figure 3.9, we used a first stage proposal with initial variance  $V_1 = 10$  while the second stage proposal has an initial variance  $V_2 = 0.1 * V_1$ , this strategy allows us to make both global and local steps through the parameter space.

### 3.5 Multiple-Try Metropolis

Liu et al. (2000) introduced the multiple-try Metropolis (MTM) algorithm. They claim that the algorithm improves exploration of neighboring regions defined by the transition kernel  $q(\theta, \phi)$ . They define weights  $w(\theta, \phi) = \pi(\phi)q(\theta, \phi)\lambda(\theta, \phi)$ , where  $\lambda(\theta, \phi)$  is a nonnegative symmetric function in  $\theta$  and  $\phi$  with  $\lambda(\theta, \phi) > 0$  whenever  $q(\theta, \phi) > 0$ . Suppose that  $\theta$  is the current state of the Markov chain, then one iteration of the algorithm proceeds according to the following steps:

- Draw a random sample  $\phi_1^*, \dots, \phi_k^*$  from  $q(\theta, \cdot)$ ;
- Draw  $\phi_k$  from  $\{\phi_1^*, \dots, \phi_k^*\}$  with probability proportional to  $w(\phi_j^*, \theta)$ ;
- Draw  $\phi_1, \dots, \phi_{k-1}$  from  $q(\phi_k, \cdot)$ ;
- Calculate the acceptance probability

$$\alpha = \min\left(1, \frac{w(\phi_1^*, \theta) + \dots + w(\phi_k^*, \theta)}{w(\phi_1, \phi_k) + \dots + w(\phi_k, \phi_k)}\right).$$

If the move is accepted, set  $\phi = \phi_k$ . If the move is not accepted, set  $\phi = \theta$ .

Two of the  $\lambda$  functions introduced by Liu et al. (2000) are  $\lambda_1(\theta, \phi) = 2q(\theta, \phi) + q(\phi, \theta)^{-1}$  and  $\lambda_2(\theta, \phi) = q(\theta, \phi)q(\phi, \theta)^{-\alpha}$ . When  $\alpha = 1$ ,  $w(\theta, \phi)$  can be thought of as weights obtained in a weighted resampling algorithm with proposal density  $q(\theta, \phi)$  and target  $\pi$ . Figure 3.10 shows that MTM samples pretty well the target distribution with  $k = 10$ , and the proposal is a random walk with variance  $V = 100$ . While the number of iterations is the same as for the other examples, the number of evaluations of

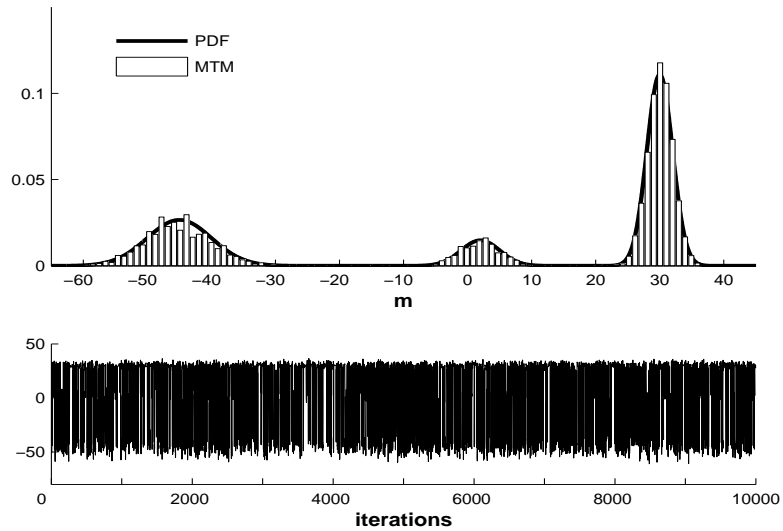


Figure 3.10: Multiple-try Metropolis algorithm applied to the target distribution  $\pi(m) = \frac{3}{9}N(-45, 25) + \frac{1}{9}N(2, 9) + \frac{5}{9}N(30, 4)$ , using  $k = 10$  different proposed new values at every time. Top: The black solid line represents the actual pdf, and the white bars, the plot of the histogram. Bottom: Trace plot of iterations.

the target distribution is approximately 200,000 since we need to evaluate the target distribution at each of the  $\phi_1^*, \dots, \phi_k^*$ , and at each of the  $\phi_1, \dots, \phi_{k-1}$  to compute the acceptance probability  $\alpha$ . This could be a major concern with this technique if the cost to evaluate  $\pi$  is too high.

### 3.6 Simulated Tempering

This algorithm proposed by Marinari and Parisi (1992) and later by Geyer and Thompson (1995) is motivated by the idea of using a temperature parameter to control the simulation of the target distribution  $\pi$  in the same spirit as the Simulated Annealing method. In the Simulated Tempering (ST) method the target distribution

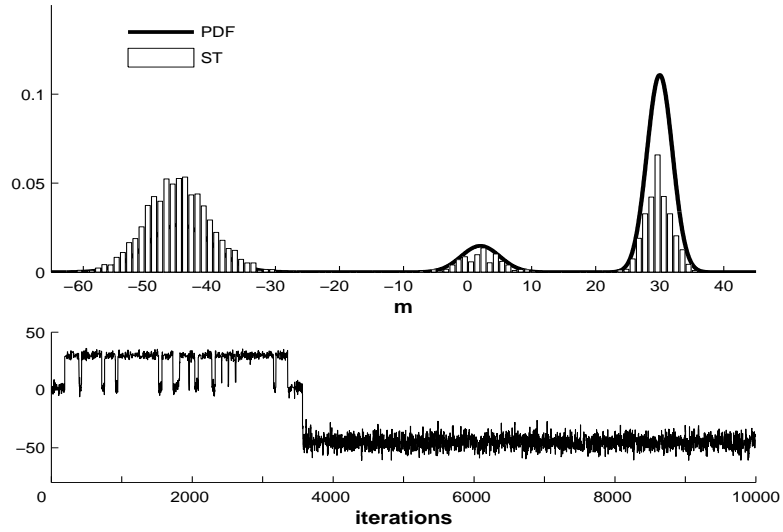


Figure 3.11: Simulated Tempering algorithm applied to the target distribution  $\pi(m) = \frac{3}{9}N(-45, 25) + \frac{1}{9}N(2, 9) + \frac{5}{9}N(30, 4)$ , using 10 heated distributions with a cooling schedule similar to MVFSA. Top: The black solid line represents the actual pdf, and the white bars, the plot of the histogram. Bottom: Trace plot of iterations.

is assumed to be  $\pi(m) \propto h_i(m)p(i)$ , where  $h_i(m)$  is the  $i$ th heated distribution. For instance,  $h_i(m) = \pi(m)^{1/T_i}$ , with  $i = 1, \dots, k$  possible heated distributions. The values  $p(i)$  are called pseudo-priors and they must be chosen in advance. The ST algorithm follows the next steps:

- Set an initial value  $m^{(0)}$ .
- Set the pseudo-priors as  $p(i) = \frac{1}{k}$ ,  $i = 1, \dots, k$ .
- Set a finite cooling schedule for the temperature of size  $k$ . Let say  $(T_1, \dots, T_k)$ , where  $T_1 > T_2 > \dots > T_k$ . The index  $i$  will represent any of these  $k$  values.
- At iteration  $t$  update  $m$  using a M-H or GS step for  $\pi_i(m)$ . Note that the index  $i$  is fixed.

- Set  $j = i \pm 1$  according to probabilities  $q_{i,j}$ , where  $q_{1,2} = q_{k,k-1}$  and  $q_{i,i+1} = q_{i,i-1} = \frac{1}{2}$  if  $1 < i < k$ . The interpretation of  $q_{i,j}$  is the probability to move from  $T_i$  to  $T_j$ .
- Accept the transition from  $i$  to  $j$  with probability

$$r = \min\left(1, \frac{h_j(m)p(j)q_{j,i}}{h_i(m)p(i)q_{i,j}}\right)$$

- Adjust pseudo-priors  $p(i)$ . This can be done by looking at the time spent by the sampler at each heated distribution.

Figure 3.11 shows that ST can obtain acceptable results, but we can appreciate that one of the local modes is underestimated and the other one is overestimated. Disadvantages on this algorithm are the calculation of the auxiliary probabilities  $p(i)$ , selecting properly the proposal distribution to sample from that we have to use on the M-H step and the definition of the cooling schedule. In this example, the  $p(i)$ 's were adjusted after an exploratory run by looking at the time spent per each heated distribution. The variance  $V$  on the M-H step was equal to 100 and the cooling schedule was chosen from the MVFSA algorithm.

## 3.7 Swapping

As mentioned in Holloman (2002), in the swapping algorithm the idea is to run parallel chains using M-H steps and then propose a way to exchange the values on each chain. We will discuss the case when there are only three chains, by following the next steps:

- Set an initial value  $m_i^{(0)}$  for each chain,  $i = 1, 2, 3$ .

- Run parallel M-H steps and accept the new values  $y_i$  with its respective probability

$$\min\left(1, \frac{\pi(m_i^{(k)})q(m_i^{(k)}, m_i^{(k-1)})}{\pi(m_i^{(k-1)})q(m_i^{(k)}, m_i^{(k-1)})}\right), \quad i = 1, 2, 3.$$

- At iteration  $k$  perform a swapping step, randomly choose a neighboring pair, say  $i$  and  $j$ , propose swapping  $m_i^{(k)}$  and  $m_j^{(k)}$  and accept it with probability

$$\min\left(1, \frac{\pi_i(m_j^{(k)})\pi_j(m_i^{(k)})}{\pi_i(m_i^{(k)})\pi_j(m_j^{(k)})}\right).$$

Each parallel M-H uses as proposal distribution a random walk with adaptive variance ( $N(m_i^{(k-1)}, V_k)$ ). The initial variance  $V$ , is set to be equal to 1. Allowing exchange of information among chains makes more flexible the selection of initial points to implement this method. Figure 3.12 shows very good results for this methodology. For this example we chose completely different initial values for each chain and far away from each other on the parameter space. The variances on each proposal distribution were calibrated via adaptive schemes.

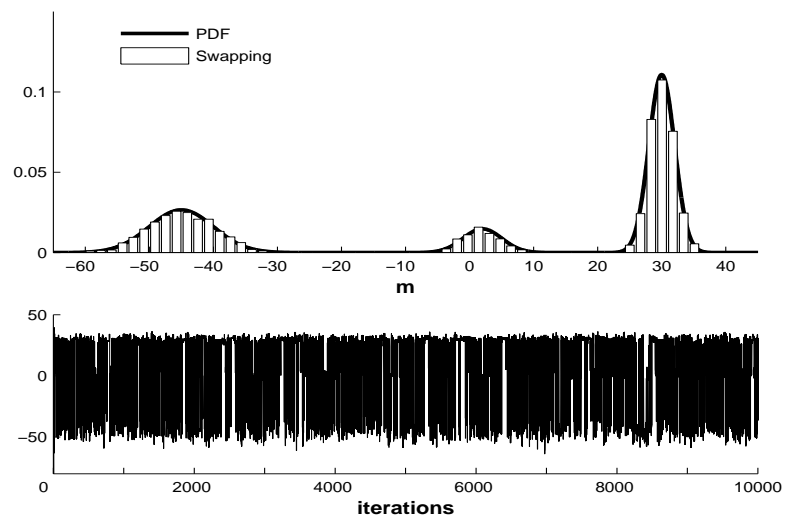


Figure 3.12: Swapping algorithm applied to the target distribution  $\pi(m) = \frac{3}{9}N(-45, 25) + \frac{1}{9}N(2, 9) + \frac{5}{9}N(30, 4)$ , using 3 parallel chains starting from different points over the parameter space. Top: The black solid line represents the actual pdf, and the white bars, the plot of the histogram. Bottom: Trace plot of iterations.



# Chapter 4

## Computational Approaches Applied to a Climate Model

### 4.1 Introduction

The computational effort of evaluating the cost function in climate models has made the Metropolis/Gibbs traditional schemes useless in this context. Other methods of the previous chapter were discarded as well because they need either too much tuning or too much iterations to be made in the context of climate model calibration. In the former case, Simulated Tempering requires too much tuning of the pseudo-priors and the proposals to be considered. Among the latter, the Swapping method requires multiple chains to be run in parallel and an extra step to exchange information among chains to be implemented properly. Multiple-try Metropolis requires a lot of forward evaluations due to the use of multiple proposals which makes it not suitable for the climate model. On the other hand, the MVFSA algorithm provides fast, approximate but biased answers to solve the problem of mapping the multidimensional PPD. In order to reach a balance between efficiency and precision we consider the use of

adaptive methods.

In the case of the climate model described in Chapter 2, we simulate samples from the posterior distribution of interest  $\pi(m, S|d_{obs})$  using a two step scheme. We considered flat priors on all the orbital forcing parameters and since  $S$  can be seen as a precision parameter it is reasonable to choose a Gamma distribution as a prior. The hyperparameters  $\alpha_0 = 552.25$  and  $\beta_0 = 11.75$  were fixed by expertise to provide a prior distribution with  $E(S) = 47$  and  $Var(S) = 4$ . We will discuss this issue further on Section 4.3.  $\pi(m|S, d_{obs})$  is sampled accordingly through the SCAM and FAM algorithms and then we simulate from  $\pi(S|m, d_{obs})$  directly from a Gamma distribution with parameters  $\alpha^* = \alpha_0$  and  $\beta^* = \beta_0 + E(m^{(k)})$ .

We compare five different computational techniques, three adaptive methods (FAM, DRAM and SCAM) and two algorithms based on the simulated annealing scheme (MVFSA and METSA) in the context of the surrogate climate model described in Chapter 2. Our main goal is to provide alternatives to traditional algorithms such as the Gibbs sampler and the Metropolis-Hastings (Hastings (1970)) that have little chances to succeed in the climate model problem either because they need a lot of forward evaluations to solve the inverse problem or because we may have difficulties finding a good proposal distribution to sample the parameter space. To make a fair comparison among methods we use a Sun Ultra 2 (Solaris 8) server with 2400 Mhz in CPU and 2048 MB in memory to run the FORTRAN code of each algorithm.

One of the main concerns on climate models is the time spent in performing forward computations. We did several runs of 100,000 iterations to estimate the time in seconds that every method uses to make one iteration for the climate model. We found there is no significant difference among FAM, METSA and MVFSA. DRAM is slower than these three algorithms because of the delayed rejection. SCAM is three times slower than FAM to compute one evaluation of the climate model as

one would expect since it is sampling each parameter conditional on the other parameters. Should the dimension of the parameter space increase to  $N$ , then the SCAM algorithm would spend  $N$  times what FAM would have used to make one iteration. Table 1 shows the time in seconds that each algorithm takes to compute one iteration.

<b>Method</b>	<b>Time (sec)</b>
FAM	0.06054
DRAM	0.10146
SCAM	0.18148
METSA	0.06340
MVFSA	0.06291

Table 4.1: Time in seconds spent per computational method to compute one iteration with the climate surrogate model.

Additionally, we looked at the bivariate scatter plots (Figure 4.1) of the samples from the different methods corresponding to each orbital parameter. For the adaptive methods, we use a burn in period of 20,000 iterations while in the case of MVFSA and METSA there is no burn in period and their sampling design is based on multiple independent starting points over the parameter space. These scatter plots reveal that all methods sample the regions of the parameter space where the optimum is located. However, the dispersion of the samples based on the Simulated Annealing methods is higher than the one from adaptive methods. While the adaptive methods use one single point and adapt themselves to reach the target distribution, the MVFSA and METSA need to start at different independent initial points to collect all the information that they require to estimate the PPD.

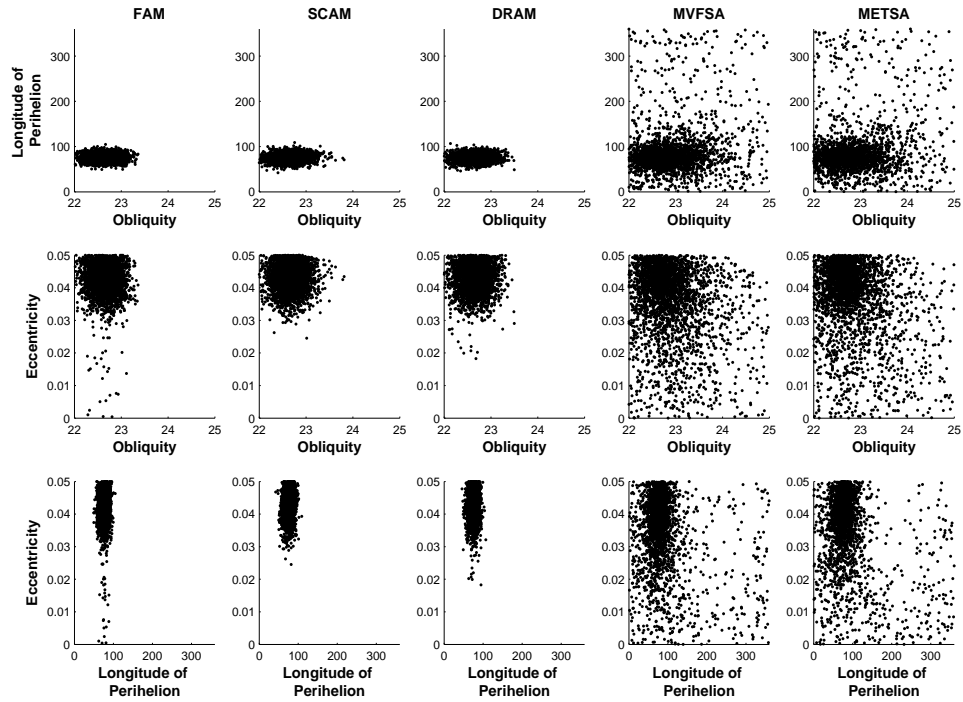


Figure 4.1: Bivariate scatter plots of orbital forcing parameters. First column: FAM. Second column: SCAM. Third column: DRAM. Fourth column: MVFSA. Fifth column: METSA.

## 4.2 Root Mean Square (RMS) Probability Error

Besides assessing the time that each method uses to make forward evaluations of the climate model, we consider an empirical measure of convergence for any proposed algorithm. For every parameter  $\theta$ , the RMS probability error is defined as follows:

$$RMS_i(\theta) = \|Prob_i^{(\theta)} - Prob_{\pi}^{(\theta)}\|$$

where  $i$  goes from 0 to the maximum number of iterations and  $\|\cdot\|$  denotes the Euclidean norm.  $Prob_i^{(\theta)}$  is a vector that contains the frequencies to plot a histogram using the samples generated from a specific method (FAM, SCAM, DRAM, MVFSA

or METSA) at iteration  $i$ .  $Prob_{\pi}^{(\theta)}$  is a vector of frequencies from the target distribution  $\pi$  based on the same bins used for  $Prob_i^{(\theta)}$ . Since we do not have available the actual frequencies based on  $\pi$ , we replace them by those obtained with the whole sample simulated from the FAM since this algorithm has ergodic properties. RMS has the desired property that it goes to zero as  $i$  goes to infinity when the method being used converges to the target distribution. The RMS also provides a measure of how fast an estimated PPD from the samples of any particular method stops changing over iterations.

As shown in Villagran et al. (2008) every method provides a PPD centered on the target value (see, Figure 4.2) of obliquity,  $\Phi = 22.625$ . MVFSA and METSA provide a distribution with long tails due to their design based on multiple initial values and Simulated Annealing compared to the distribution of obliquity obtained with FAM, SCAM and DRAM. A goal for climate models is to estimate properly the uncertainty about parameters of interest. For instance, we consider the estimation of the 97.5% quantile for every parameter. We estimated the 97.5% quantile per iteration and for all the methods. The quantile estimation via MVFSA and METSA is at a value of 24.2, while with the other methods the estimated quantile is at 23.

One comment in Sen and Stoffa (1996) and Jackson et al. (2004) is that MVFSA is preferred over Metropolis/Gibbs algorithms because of the reduction in required number of forward evaluations and the time needed to estimate the PPD. In Figure 4.3, we can see that DRAM and SCAM have the lowest RMS up to iteration 30,000, after which FAM has the minimum RMS value. The RMS values of MVFSA converge as quick or quicker than those of DRAM and SCAM but they do not reach zero. One concern with the RMS as a measure to compare how fast the algorithms converge to the target distribution, is that it is presented as a function of the iterations. This could be misleading since it does not consider that every method requires a different number of model evaluations to make one iteration. To overcome this issue

we propose to look at the RMS as a function of time. In this case DRAM fairs better than SCAM within the first hundreds of seconds. Perhaps the most surprising aspect of these results is the slow convergence rate of FAM. One would have expected FAM to be fast, similar to the DRAM and SCAM results, since it integrates information about previous samples to generate an improved proposal distribution. We hypothesize that the calibration of the covariance matrix for the proposal distribution in this problem is particularly difficult because of the need to restrict orbital forcing parameter values to a particular range. We suspect that the DRAM algorithm is able to overcome this problem by adding a second stage proposal with more precision which allows us to improve the mixing at the beginning of the algorithm. It is not entirely clear to us why SCAM was able to perform so well relative to FAM. However, SCAM handles one parameter at each stage and therefore it is simpler to deal both with the parameter physical restrictions and with the correlation among parameters.

For the Longitude of Perihelion parameter, Figures 4.4 and 4.5 show similar results as described for the previous Figures. In this case DRAM and SCAM converge as fast to an answer compared to MVFSA and METSA. The estimated 97.5% quantile via MVFSA and METSA is now 3 times greater than the same estimated quantile for the other methods. In the case of the Eccentricity parameter (Figures 4.6 and 4.7), FAM and SCAM have some differences on the estimation of the PPD, this is because the FAM uses as a proposal distribution a multivariate normal density which could lead into many rejections of new candidates points due to the physical restrictions on the orbital forcing parameters. On the other hand, the estimated PPD under DRAM is similar to the one obtained via SCAM. For the Eccentricity parameter, there are no differences on the estimation of the 97.5% quantile. However, this is not the case for the estimation of the 2.5% quantile as we can see from the box plots in Figure 4.8.

We looked at the sample autocorrelation function (ACF) of each parameter (Fig-

ure 4.9) for all the algorithms that are being considered. This is a well known way to assess how every algorithm is mixing along iterations, the smaller the autocorrelations the better the method is moving across the parameter space. We noticed that FAM has the worst autocorrelations, this is no surprise since the algorithm is sampling all the parameters at once from a multivariate Gaussian distribution and the physical restrictions on each parameter make difficult to achieve acceptances that satisfy the restrictions. By using DRAM, we reduce the autocorrelations of the FAM but they are not as good as the autocorrelations of SCAM. In fact, SCAM is very efficient in terms of chain mixing even though it is the slowest one in computer time spent per iteration. The adaptive methods have acceptance rates at around 45%. METSA and MVFSA are similar in terms of autocorrelation, their autocorrelation function values are not so large because the sampling design on these algorithms is based on multiple initial points that are not connected to each other along the iterations.

### 4.3 Appraisal of a few forward evaluations on the climate model

Even though the surrogate climate model used here is cheap enough to make forward evaluations without consideration of the execution time, climate modelers have an interest in making inferences about models that may take hours to days to execute a single iteration of a stochastic sampler. In these cases one may only have a limited number of samples to work with. Looking for an algorithm that is efficient to sample from becomes a mandatory task. From a statistical standpoint, a statistical climate model or emulator can avoid computational limitations as shown in Sansó et al. (2008). Depending on the problem characteristics, this may not always be an easy task especially if the region of acceptability is a small fraction of the parameter space volume. Therefore, inefficient samplers fail to capture the most important regions. We considered a Gaussian spatial model to approximate the surface of the surrogate climate model parameters but we did not obtain good results. We believe this was a result of not having enough information or understanding about the climate model characteristics to formulate a statistical version. Also, stationary models are limited for the output arising from our surrogate model. Therefore, there will likely continue to be a need for efficiently sampling schemes to achieve good inferences with few evaluations. We chose only 500 forward evaluations for testing the sampling strategies.

In Figure 4.10, we see that based on 500 iterations MVFSA and METSA sample various points over the parameter space, but a few hundreds iterations are not enough to establish a pattern on the bivariate plots. On the other hand, SCAM and DRAM show a very good concentration of the samples. These bivariate plots look much alike the scatter plots we get using the entire run from the adaptive methods. The behavior of the FAM algorithm although disappointing, is expected since the proposal



cannot be easily tuned with a small number of iterations. There is no burn-in period considered in the samples of FAM, DRAM and SCAM used to produce this Figure.

In Table 4.2, we compare the uncertainty estimation of the parameters and the minimum values of the cost function. The performance of SCAM and DRAM are remarkable since they not only find a minimum cost with few forward evaluations, but they also provide acceptable estimates of the 95% credible intervals of the orbital forcing parameters. The 95% credible intervals for the methods based on Simulated Annealing are not informative at all because they practically covered the entire parameter space.

<b>Method</b>	<b><math>E(\mathbf{m}^*)</math></b>	$\Phi_{2.5\%}$	$\Phi_{97.5\%}$	$\lambda_{2.5\%}$	$\lambda_{97.5\%}$	$e_{2.5\%}$	$e_{97.5\%}$
FAM	0.8411	22.6817	22.7813	0.2248	3.4413	0.0101	0.0331
SCAM	0.1912	22.1644	23.1309	60.6625	91.5434	0.0312	0.0495
DRAM	0.1943	22.1665	22.9816	62.7528	141.1550	0.0113	0.0491
MVFSA	0.2019	22.1595	24.5072	18.3329	311.9383	0.0089	0.0482
METSA	0.2029	22.0631	24.2744	38.6806	353.6220	0.0029	0.0481

Table 4.2: Comparative estimation after 500 forward evaluations.  $E(\mathbf{m}^*)$  is the minimum of the cost function,  $\Phi$  is the Obliquity,  $\lambda$  is the Longitude of Perihelion and  $e$  is the Eccentricity.

In Figure 4.11, we compare the estimation of the marginals using 500 iterations to the estimation made with 100,000 samples from the FAM. With only 500 iterations, FAM does a poor job estimating the marginals while DRAM and SCAM seem to do a very good job for Longitude of Perihelion and Eccentricity. For Obliquity the estimation is acceptable. The factor  $h$  on the second stage proposal for DRAM is producing a large impact compared to AM for the results based with only a few iterations. MVFSA and METSA present a lot of mass on the tails in comparison to the results of the FAM with 100,000 iterations. From a practical standpoint, the main problem with FAM is its difficulty to tune up the covariance matrix from

the proposal distribution in a parameter space where there are physical restrictions. There are different strategies to combine both the delayed rejection and the adaptive Metropolis algorithms. The strategy that achieves best results for the climate model is to implement the AM for a short period of time to compute the covariance matrix  $C_t$  of the first state proposal distribution and then construct the matrix  $V_t = hC_t$ , which has more precision given that we selected  $h \leq 0.01$ .  $V_t$  is used in the second stage proposal. By doing this, we allow DRAM the chance to take two different steps: (first stage) one that allows global moves to sample along the parameter space and (second stage) another step that allows to sample with more precision from points in the parameter space where the posterior distributions have more density.

Using a small number of forward evaluations we have gained appreciation of how well the computational methods presented in this research work in this context. The main strengths and weaknesses of every sampling scheme has been emphasized. While Simulated Annealing based sampling schemes present a very good tool to optimize the problem, SCAM and DRAM reach the balance between sampling and optimization with few iterations.

## 4.4 Relevance of the parameter $S$

As mentioned in Section 2.4, the parameter  $S$  performs the function of weighing the significance of model-data differences and was included in our inferences of parametric uncertainty.  $S$  has its own prior probability distribution that provides information about observational and other uncertainties that are hard to quantify within the metric of model skill  $E(m)$  such as correlations among a suite of observations. In Figure 4.12, we can see what happens if we use a non-informative prior on  $S$ , our results only reflect the form of the likelihood function of each parameter when we set  $S = 1$ . In our case, the prior mean value of  $S$  is elicited to be 47, a number

much larger than 1 which reflects the fact that many surface temperature points are highly correlated. This correlation increases the ability to detect the effects of Earth's orbital geometry on observations of surface air temperature and sharpens the PPD around the correct orbital configuration.

The results in this thesis are consistent with the findings of Sansó et al. (2008) where they use an emulator of their climate model based on a Gaussian process. They use the MIT 2D climate model that controls the large-scale response of the climate system to external forcings. In that paper, the authors find that when they use non-informative priors on the climate model parameters, they obtain vague posteriors. Since  $S$  is used to assist on the optimization process of the cost function of the climate model, we may think that it has a positive support. A convenient form for a prior on  $S$  is the Gamma probability distribution. The simplicity of this assumption makes possible to sample directly from the full conditional  $P(S|m, d)$ . More on the interpretation and the motivation of this  $S$  parameter can be found in Jackson et al. (2008).

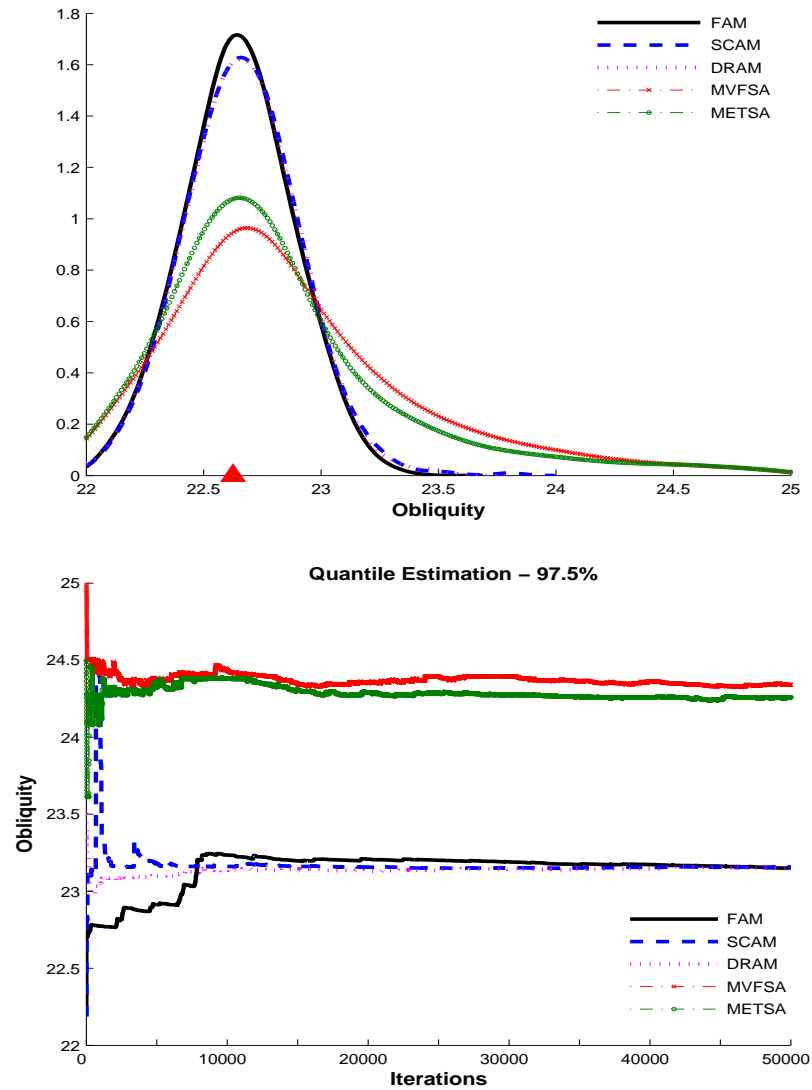


Figure 4.2: Comparison for Obliquity parameter. Top left: PPD estimation. Bottom: Ergodic quantile estimation (97.5%).

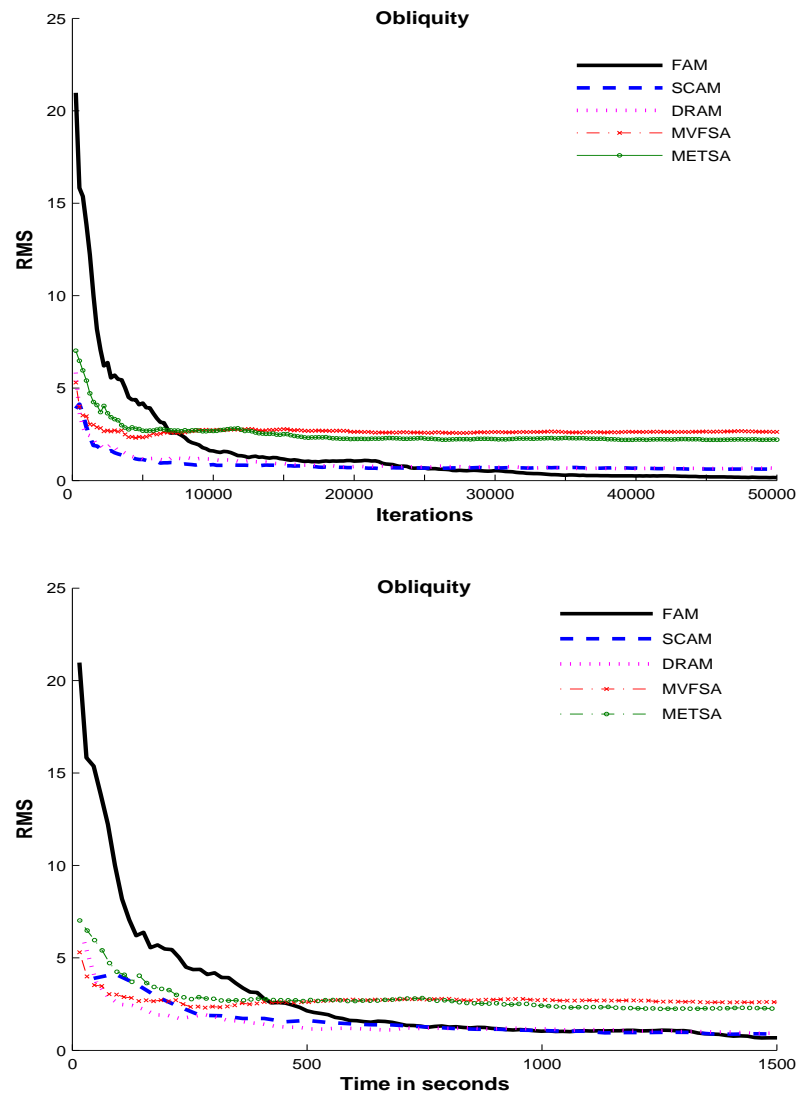


Figure 4.3: Comparison for Obliquity parameter. Top: RMS as function of iterations. Bottom: RMS as function of time (seconds).

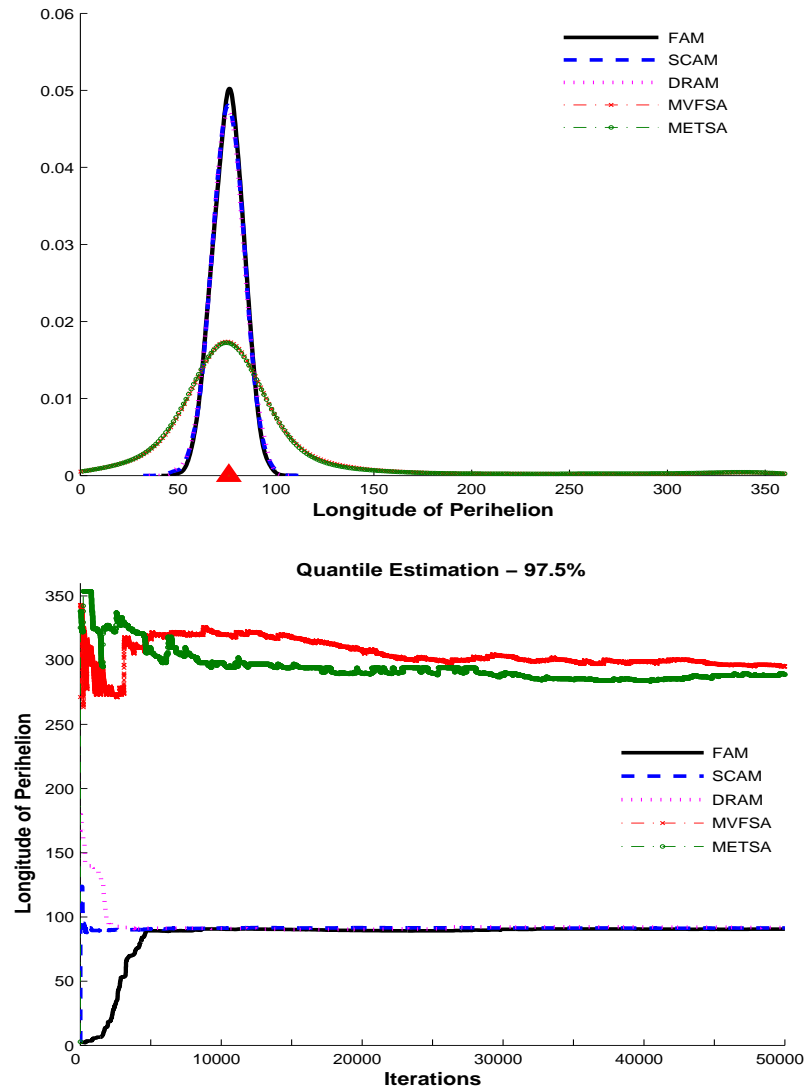


Figure 4.4: Comparison for Longitude of Perihelion parameter. Top left: PPD estimation. Bottom: Ergodic quantile estimation (97.5%).

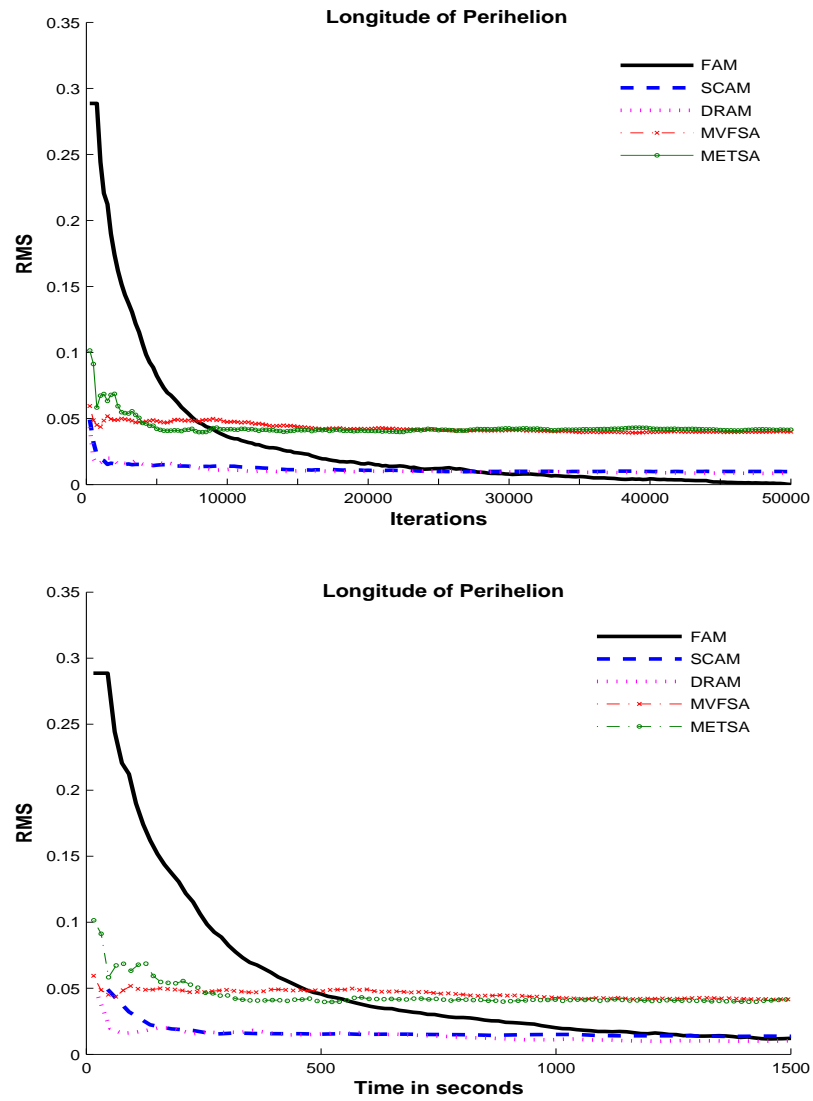


Figure 4.5: Comparison for Longitude of Perihelion parameter. Top: RMS as function of iterations. Bottom: RMS as function of time (seconds).

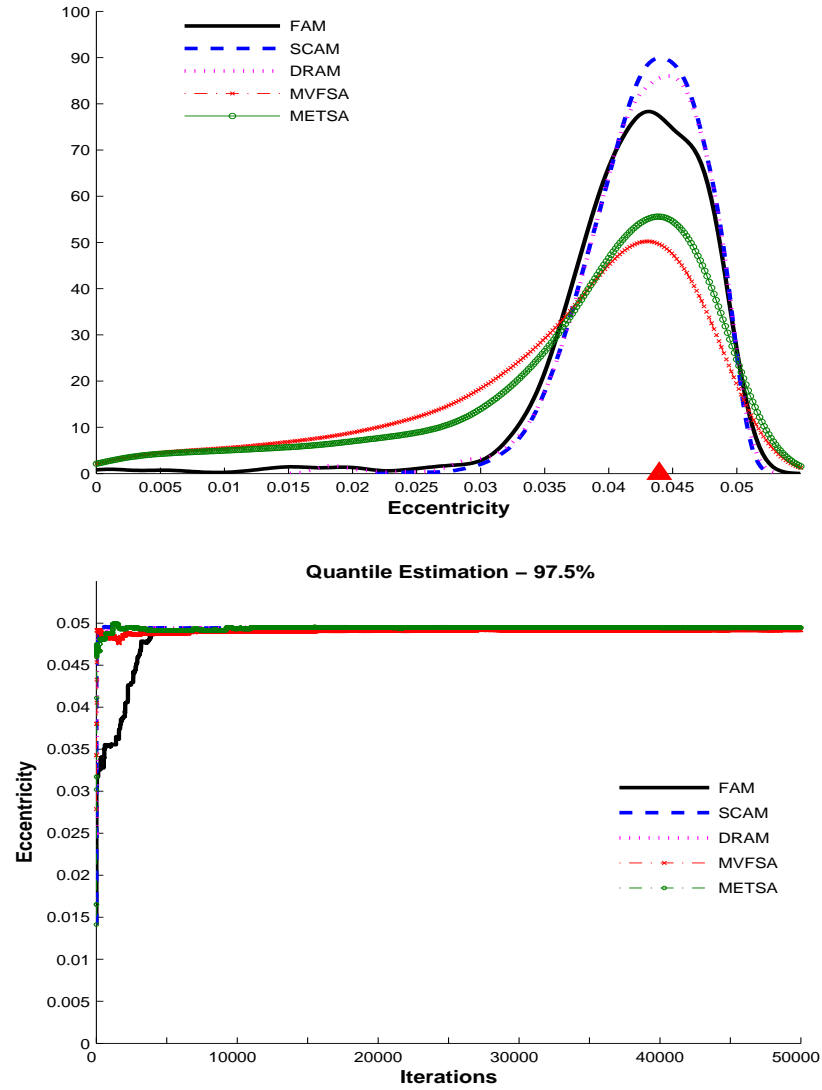


Figure 4.6: Comparison for Eccentricity parameter. Top left: PPD estimation. Bottom: Ergodic quantile estimation (97.5%).



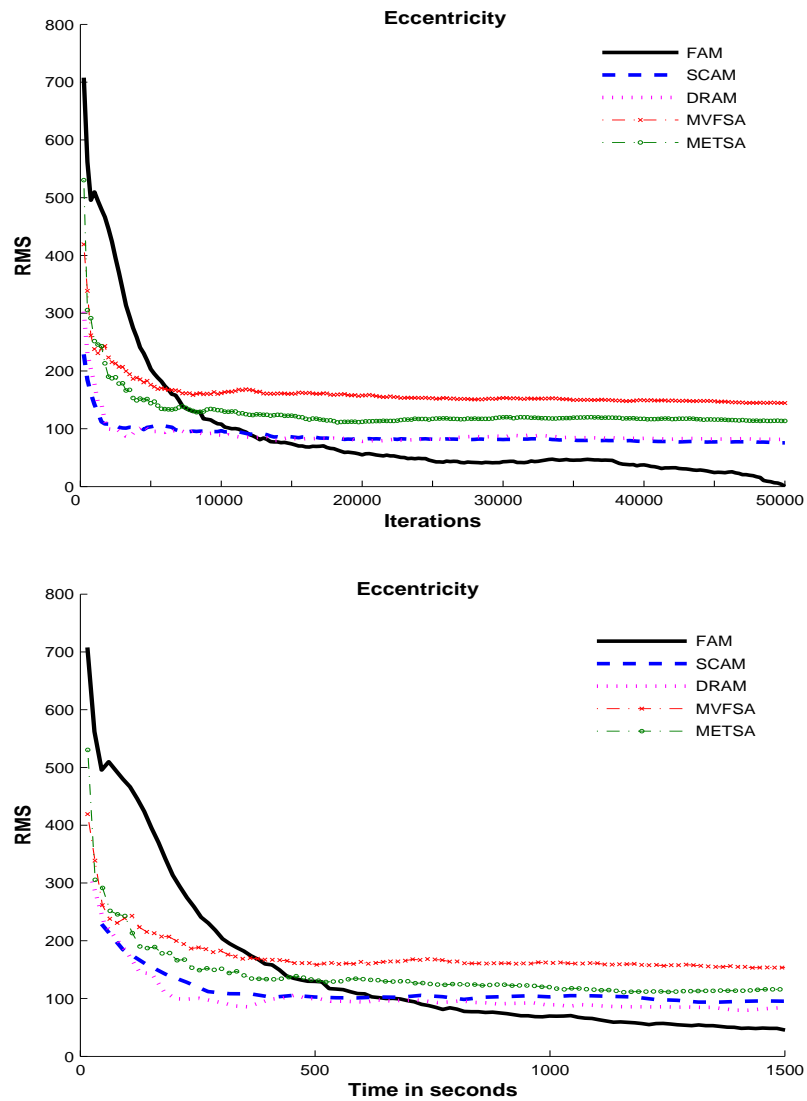


Figure 4.7: Comparison for Eccentricity parameter. Top: RMS as function of iterations. Bottom: RMS as function of time (seconds).

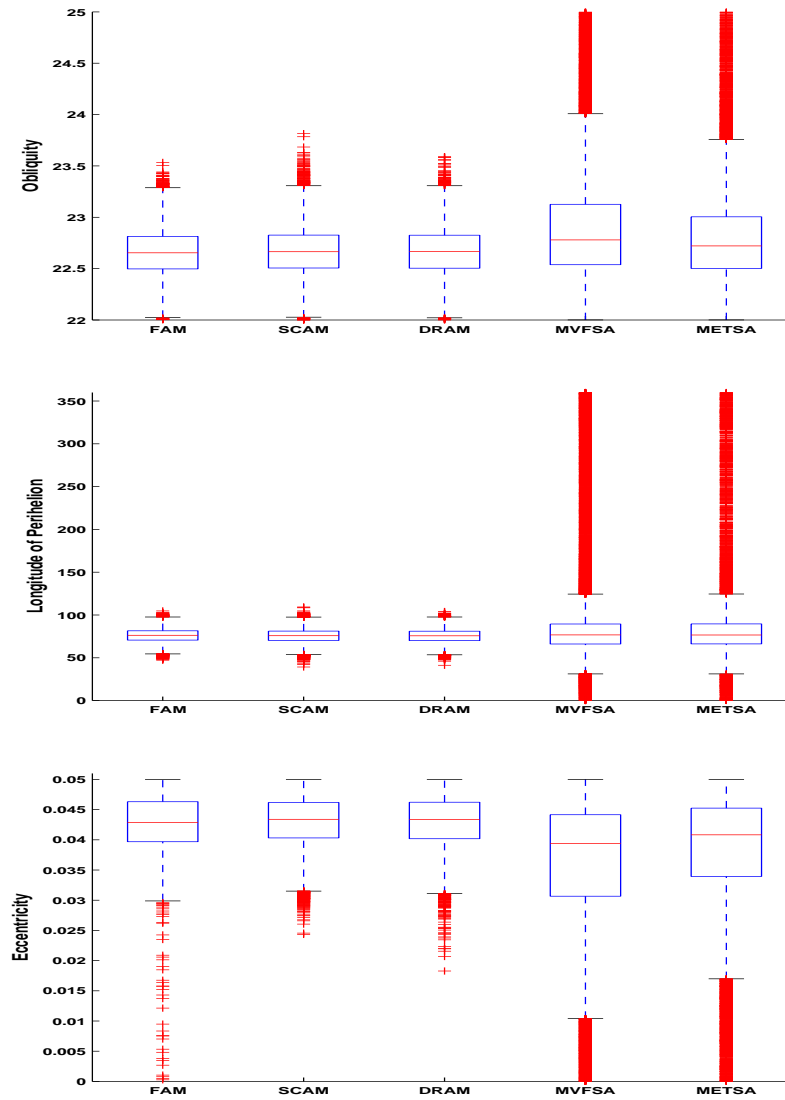


Figure 4.8: Box plots of the samples from different methods. Top: Obliquity. Middle: Longitude of Perihelion. Bottom: Eccentricity.

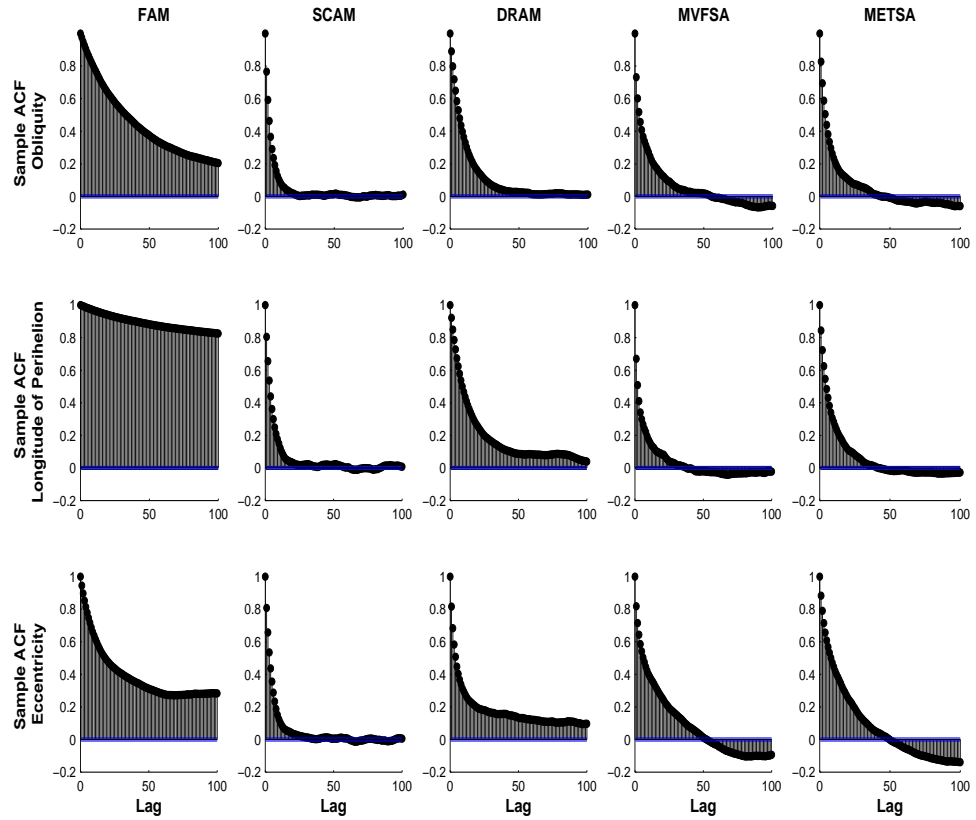


Figure 4.9: Autocorrelation function of orbital forcing parameters. First column: FAM. Second column: SCAM. Third column: DRAM. Fourth column: MVFSA. Fifth column: METSA.

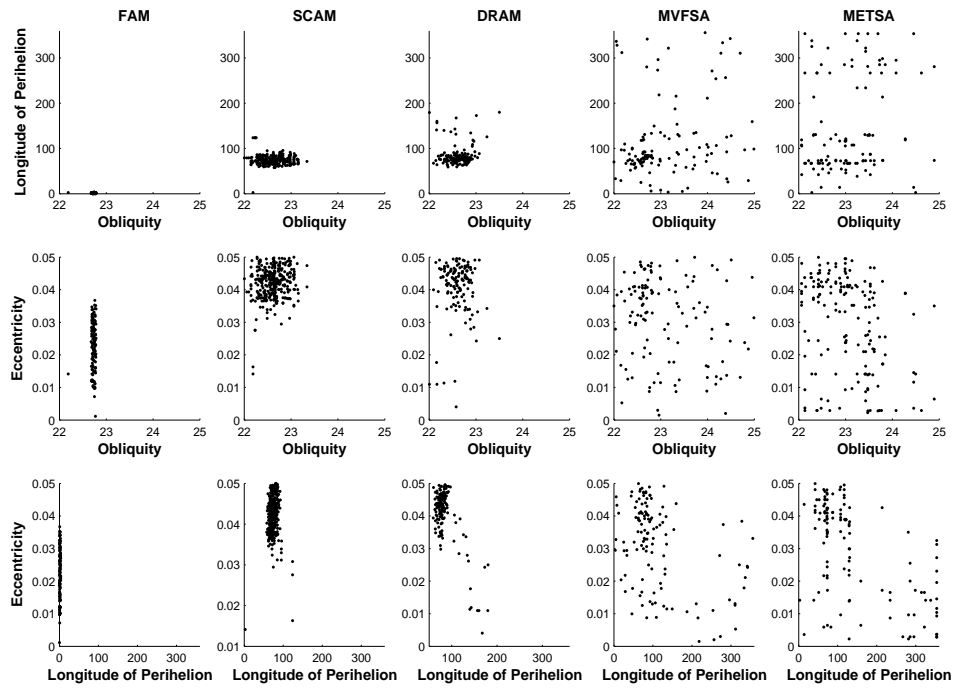


Figure 4.10: Bivariate scatter plots of orbital forcing parameters with just 500 iterations. First column: FAM. Second column: SCAM. Third column: DRAM. Fourth column: MVFSA. Fifth column: METSA.

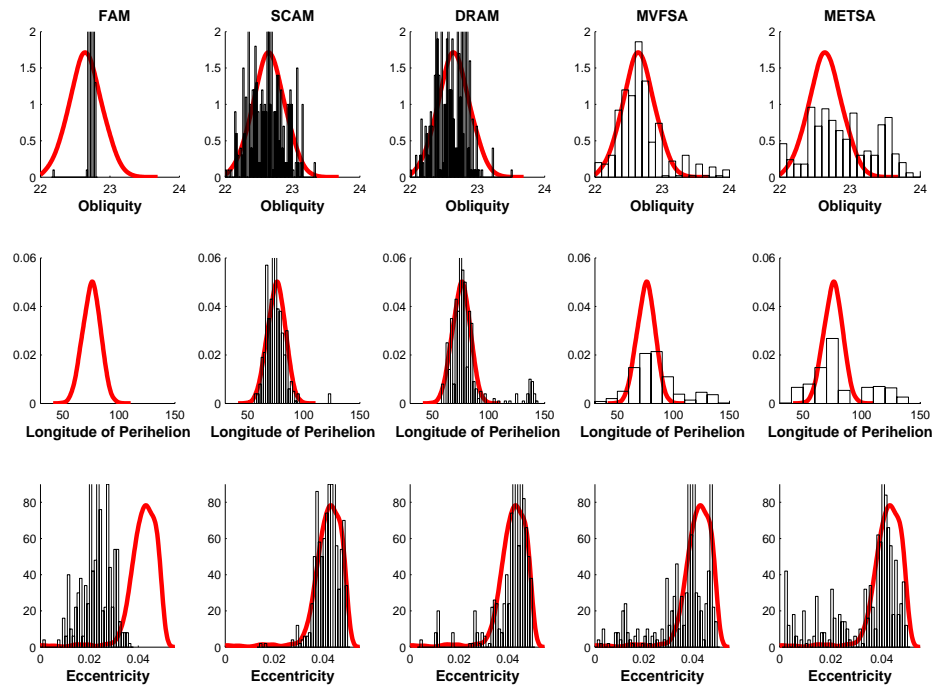


Figure 4.11: Histograms with just 500 iterations (black bars). PPD Estimated using 100,000 iterations from the FAM algorithm (red line). First column: FAM. Second column: SCAM. Third column: DRAM. Fourth column: MVFSA. Fifth column: METSA.

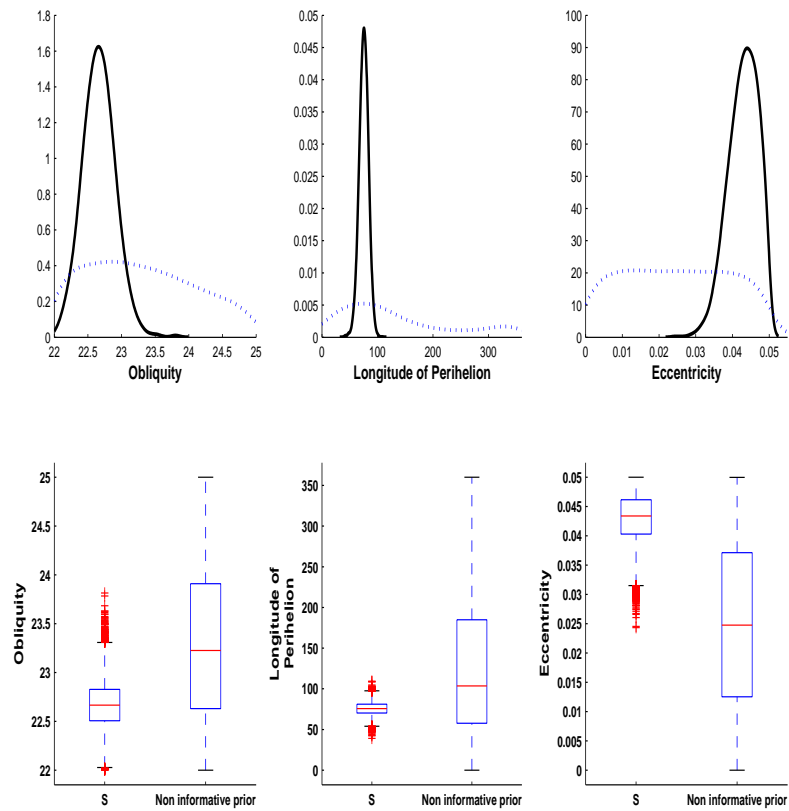


Figure 4.12: First row: Estimated PPD using an adaptive method. Using the  $S$  parameter (solid line) and using a non-informative prior (dotted line). Second row: Box plots of posterior samples.

# Chapter 5

## Conclusions and future work

Sen and Stoffa (1996) stated that traditional methods such as Metropolis or Gibbs sampler were not sufficiently practical for finding inversion solutions due to their high cost in terms of the time required making forward evaluations and the amount of tuning these algorithms need. MVFSA and METSA were proposed as algorithms to overcome these problems. Indeed they provide approximate and fast answers to estimating the PPD. These methods share the same design based on multiple independent initial starting values, cooling schedule and a Metropolis acceptance rule for new candidates. Adding multiple initial points to these algorithms helps to provide more information about the parameter space similar to a grid point search. They differ in the proposal distribution since MVFSA uses a Cauchy distribution dependent on the cooling schedule to sample candidates from high density regions while METSA uses a random walk proposal and the temperature cooling schedule is present only in the acceptance/rejection step. However, by its design MVFSA has biases estimating the tails of the PPD of the climate model parameters. Annan and Hargreaves (2007) noted that we can consider either MVFSA or METSA as sophisticated heuristic methods to estimate the PPD.

## Chapter 5. Conclusions and future work

The series of adaptive methods proposed by Haario et al. (1999, 2001, 2004, and 2005) are setting a breakthrough in Monte Carlo methods by introducing adaptive schemes that are successful in finding the target distribution not only in theory but also in practice despite the fact that these are non Markovian algorithms.

The results in Chapter 4 show that compared to FAM, DRAM, or SCAM, which all provided nearly identical estimates of the marginal PPD, marginals derived from MVFSA and METSA sampling had similar modes, but with broader 95% credible intervals. For instance, the Longitude of Perihelion ( $\lambda = 75.93$ ) parameter, the 95% posterior credible interval using the FAM scheme is (60.291, 90.676) compared to (23.274, 280.675) that was obtained using MVFSA. The inclusion of the parameter  $S$  in the estimation process provides prior information about the significance of model differences with the target observations.

The main goal of this dissertation was to compare sampling efficiencies and accuracies among different proposed methods for estimating parametric uncertainties of a climate model. The DRAM and SCAM sampling algorithms are particularly effective on both of these accounts. In terms of the RMS criterion, the adaptive methods were as efficient as MVFSA to reach convergence, but without its sampling biases.

The decision to choose either the full component version (FAM) or the single component version (SCAM) of the adaptive methods should be based on the problem at hand. FAM has great speed but also needs to calibrate the covariance matrix of the parameters of interest. As we observed, this could be a drawback in parameter spaces with physical restrictions. We propose to employ DRAM when such problem exists and obtained substantial improvements on the results. On the other hand, SCAM does not have to deal with inversion of covariance matrices but it pays a price regarding computational time as soon as the dimensionality of the problem increases. As a caveat, the results obtained from the methods used in this research



only apply to non-linear inverse problems with unimodal posteriors. On the other hand, the results of Chapter 3 provide good evidence of how these methods work with a one-dimensional multi-modal distribution. In general, finding a good proposal distribution to sample from is difficult even for adaptive methods when the posterior distribution of interest has multiple modes, Jackson et al. (2004) proposed a modification to (2.2),

$$g_{ijk}(m) = 2\text{sgn}(\Phi' \hat{A}_{o,ijk} e^{\hat{A}_{p,ijk} \cos(\lambda - \hat{\phi}_{p,ijk})}) \sqrt{|\Phi' \hat{A}_{o,ijk} e^{\hat{A}_{p,ijk} \cos(\lambda - \hat{\phi}_{p,ijk})}| + \hat{R}_{ijk}},$$

to permit the existence of two solutions. Figure 5.1 shows that adaptive methods fail to find the second solution because of the combination of non-linearity and multimodality in this problem. Even though MVFSA found both solutions, it is unclear how well the PPD is estimated due to the inherent biases of this method.

Ongoing work is being done to develop computational methods that have a balance between sampling a multi-modal PPD and reducing the cost in making forward evaluations in climate models. As part of future research is the study of an MCMC based over a nonparametric surface that approximates the PPD, such idea has been developed by Sambridge (1999) using Voronoi cells.

In terms of efficiency of estimation and ensemble generation in climate modeling, a statistical emulator is a computationally efficient approximation to a complex computer model as mentioned in Annan and Hargreaves (2007) and depicted in Sansó et al. (2008). Although, we share the same goals as Sansó et al. (2008) our problem is different since we are making evaluations directly on the climate model and not based on a statistical emulator.

One can use the results obtained in this research to suggest how with relatively few model integrations, one may draw inferences of how observational data may constrain uncertain model parameters or physical hypotheses about how nature works. The potential applications are broad and may prove invaluable for problems that are

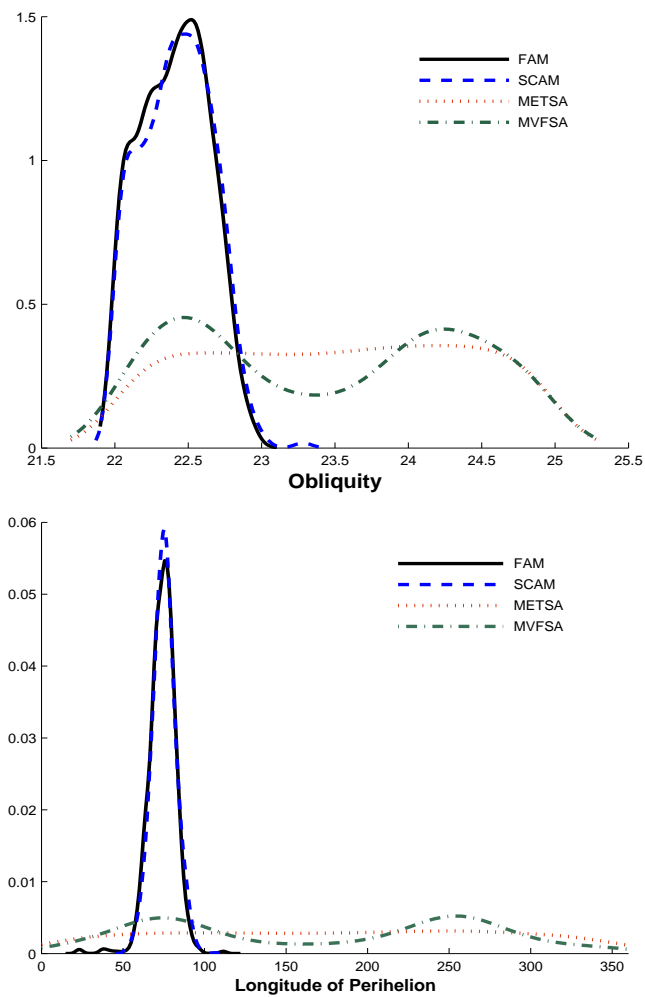


Figure 5.1: Surrogate Climate Model with two solutions (modes). Comparative PPD estimation using Adaptive Metropolis methods and Simulated Annealing based techniques. Top: Comparison for Obliquity parameter. Bottom: Comparison for Longitude of Perihelion parameter.

currently limited by computational requirements of the forward model.

# References

- ALLEN, M. (1999). *Do-it-yourself climate prediction*, Nature, 401, 642.
- ALLEN, M., STOTT, P., MITCHELL, J., SCHNUR, R. AND DELWORTH, T. (2000). *Quantifying the uncertainty in forecasts of anthropogenic climate change*, Letters to Nature, 407, 617-620.
- ANNAN, J.D. AND HARGREAVES, J.C. (2007). *Efficient estimation and ensemble generation in climate modelling*, Philosophical Transactions of the Royal Society A, 365, 2077-2088.
- BARNETT, D., BROWN, S., MURPHY, J., SEXTON, D. AND WEBB, M. (2006). *Quantifying uncertainty in changes in extreme event frequency in response to doubled CO<sub>2</sub> using a large ensemble of GCM simulations*, Climate Dynamics, 26, 489-511.
- BERLINER, M. (2003). *Uncertainty and Climate Change*, Statistical Science, 18, 430-435.
- BERLINER, M., LEVINE, R. AND SHEA, D. (2000). *Bayesian Climate Change Assessment*, Journal of Climate, 13, 3805-3820.
- CLAUSSEN, M., MYSAK, L., WEAVER, A., CRUCIFIX, M., FICHEFET, T., LOUTRE, M., WEBER, S., ALCAMO, J., ALEXEEV, V., BERGER, A., CALOV, R., GANOPOLSKI, A., GOOSE, H., LOHMANN, G., LUNKEIT, F., MOKHOV,

- I., PETOUKHOV, V., STONE, P. AND WANG, Z. (2002). *Earth system models of intermediate complexity: closing the gap in the spectrum of climate system models*, Climate Dynamics, 18, 579-586.
- COLLINS, M., BOOTH, B., HARRIS, G., MURPHY, J., SEXTON, D. AND WEBB, M.. (2006) *Towards quantifying uncertainty in transient climate change*, Climate Dynamics, 127-147.
- FOREST, C., M. R. ALLEN, P. H. STONE, AND A. P. SOKOLOV (2000). *Constraining uncertainties in climate models using climate change detection techniques*, Geophys. Res. Lett., 27(4), 569-572.
- FOREST, C., M. R. ALLEN, A. P. SOKOLOV, AND P. H. STONE (2001). *Constraining climate model properties using optimal fingerprint detection methods*, Climate Dynamics, 18, 277-295.
- FOREST, C., P. H. STONE, A. P. SOKOLOV, M. R. ALLEN, AND M. D. WEBSTER (2002). *Quantifying uncertainties in climate system properties with the use of recent climate observations*, Science, 295, 113-117.
- GATES, W. L., J. S. BOYLE, C. COVEY, C. G. DEASE, C. M. DOUTRIAUX, R. S. DRACH, M. FIORINO, P. J GLECKLER, J. J. HNILO, S. M. MARLAIS, T. J. PHILLIPS, G. L. POTTER, B. D. SANTER, K. R. SPERBER, K. E. TAYLOR, D. N. WILLIAMS (1999). *An Overview of the Results of the Atmospheric Model Intercomparison Project (AMIP I)* Bulletin of the American Meteorological Society 80(1), 29-56.
- GELFAND, A.E. AND SMITH, A.F.M. (1990). *Sampling-Based Approaches to Calculating Marginal Densities*, Journal of the American Statistical Association, 85, 398-409.

- GELMAN, A.G., ROBERTS, G.O. AND GILKS, W.R. (1996). *Efficient Metropolis jumping rules*, Bayesian Statistics V, 599-608.
- GEMAN, S. AND GEMAN, D. (1984). *Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images*, IEEE Transactions PAMI-6, 721-741.
- GEYER, CH.J. AND THOMPSON, E.A. (1995). *Annealing Markov Chain Monte Carlo with Applications to Ancestral Inference*, Journal of the American Statistical Association, 90, 909-920.
- ROHDE, R.A. (2007). *Global Warming Art*, <http://www.globalwarmingart.com>
- HAARIO, H., LAINE, M., MIRA, A. AND SAKSMAN, E. (2006). *DRAM: Efficient adaptive MCMC*, Statistics and Computing, 16, 339-354.
- HAARIO, H., LAINE, M., LEHTINEN, M., SAKSMAN, E. AND TAMMINEN, J. (2004). *Markov Chain Monte Carlo methods for high dimensional inversion in remote sensing*, Journal of Royal Statistical Society, 66, 591-607.
- HAARIO, H., SAKSMAN, E. AND TAMMINEN, J. (1999). *Adaptive proposal distribution for random walk Metropolis algorithm*, Computational Statistics, 14, 375-395.
- HAARIO, H., SAKSMAN, E., TAMMINEN, J. (2001). *An Adaptive Metropolis algorithm*, Bernoulli, 7, 223-242.
- HAARIO, H., SAKSMAN, E., TAMMINEN, J. (2005). *Componentwise adaptation for high dimensional MCMC*, Computational Statistics.
- HALL, A., CLEMENT, A., THOMPSON, D., BROCCOLI, A., AND JACKSON, C. (2005). *The Importance of Atmospheric Dynamics in the Northern Hemisphere Wintertime Climate Response to Changes in the Earth's Orbit.*, Journal of Climate, 18, 1315-1325.

- HASTINGS, W. (1970). *Monte Carlo sampling methods using Markov chains and their applications*, Biometrika, 57, 97-109.
- HOLLOMAN, C.H. (2002). *Parameter Estimation Algorithms for Computationally Intense Spatial Problems*, PhD Thesis Duke University.
- INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE (2001). *Third Assessment Report*, <http://www.ipcc.ch/pub/reports.htm>
- INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE (2007). *Fourth Assessment Report*, <http://www.ipcc.ch/pub/reports.htm>
- INGBER, L. (1989). *Very fast simulated re-annealing*, Mathematical Computational Modelling, 12, 967-973.
- JACKSON, C. AND BROCCOLI, A. (2003). *Orbital forcing of Arctic climate: mechanisms of climate response and implications for continental glaciation*, Climate Dynamics, 21, 539-557.
- JACKSON, C.S., SEN, M.K., HUERTA, G., DENG, Y., AND BOWMAN, K.P. (2008). *Error Reduction and Convergence in Climate Prediction*, (Submitted to Journal of Climate).
- JACKSON, C., SEN, M. AND STOFFA, P. (2004). *An Efficient Stochastic Bayesian Approach to Optimal Parameter and Uncertainty Estimation for Climate Model Predictions*, Journal of Climate, 17, 2828-2840.
- JOUSSAUME, S., AND K.E. TAYLOR (2000). *The Paleoclimate Modeling Intercomparison Project (PMIP)*, proceedings of the third PMIP workshop, Canada, 4-8 October 1999, P. Braconnot, Ed., WCRP-111, WMO/TD-1007, pp. 271.
- KETTLEBOROUGH, J., BOOTH, B., STOTT, P. AND ALLEN, M. (2007). *Estimates of Uncertainty in Predictions of Global Mean Surface Temperature*, Journal of Climate, 20, 843-855.

- KIRKPATRICK, S., GELATT JR., AND VECCHI, M. (1983). *Optimization by simulated annealing*, Science, 220, 671-680.
- KOZLOVSKAYA, E. (2000). *An algorithm of geophysical data inversion based on non-probabilistic presentation of a priori information and definition of Pareto-optimality*, Inverse Problems, 16, 839-861.
- LIU, J.S., LIANG, F., AND WONG, W.H. (2000). *The use of multiple-try method and local optimization in Metropolis sampling*, Journal of the American Statistical Association, 95, 121-134.
- LIU, J. AND SABATTI, C. (1999). *Simulated sintering: MCMC with spaces of varying dimension*, Bayesian Statistics VI, 398-413.
- LOPEZ, A., TEBALDI, C., NEW, M., STAINFORTH, D., ALLEN, M. AND KETTLEBOROUGH, J. (2006). *Two Approaches to Quantifying Uncertainty in Global Temperature Changes*, Journal of Climate, 19, 4785-4796.
- MARINARI, E. AND PARISI, G. (1992). *Simulated Tempering: A new Monte Carlo Scheme*, Europhysics letters, 19, 451-458.
- MCAVANEY, B.J., COVEY, S. JOUSSAUME, V. KATTSOV, A. KITOH, W. OGANA, A.J. PITMAN, A.J. WEAVER, R.A. WOOD, AND Z.-C. ZHAO (2001). *Model Evaluation. Climate Change 2001: The Scientific Basis Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, Houghton, J.T., Y. Ding, D.J. Griggs, M. Noguer, P.J. van der Linden, X. Dai, K. Maskell, and C.A. Johnson, Eds., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 881pp.
- MEEHL, G.A., BOER, G.J., COVEY, C., LATIF, M., STOUFFER, R.J. (2000). *The Coupled Model Intercomparison Project (CMIP)*, Bulletin of the American Meteorological Society 81(2), 313-318.

- METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M.N., TELLER, A.H., AND TELLER, E. (1953). *Equations of State Calculations by Fast Computing Machines*, Journal of Chemical Physics, 21, 1087-1091.
- MILANKOVITCH, M. (1941). *Canon of insolation and the ice age problem*, Israel Program for Scientific Translations, Jerusalem.
- MU, Q., C. S. JACKSON, P. L. STOFFA (2003). *A Multivariate EOF-based measure of climate model performance*, (Submitted to Journal of Geophysical Research).
- MURPHY, JAMES M., D. M. H. SEXTON, D. N. BARNETT, G. S. JONES, M. J. WEBB, M. COLLINS, AND D. A. STAINFORTH (2004). *Quantification of modelling uncertainties in a large ensemble of climate change simulations*, Nature 430, 768-772.
- SAMBRIDGE, M. (1999). *Geophysical Inversion with a Neighbourhood Algorithm -I. Searching a parameter space*, Geophys. J. Int., 138, 479-494.
- SAMBRIDGE, M. AND MOSEGAARD, K. (2002). *Monte Carlo Methods in Geophysical Inverse Problems*, Review of Geophysics, 40, 1-29.
- SANSÓ, B., FOREST, C.E. AND ZANTEDESCHI, D. (2008). *Inferring Climate System Properties Using a Computer Model (with discussion)*, Bayesian Analysis, 3, 1, 1-62 .
- SEN, M. AND STOFFA, P. (1996). *Bayesian Inference, Gibbs sampler and uncertainty estimation in geophysical inversion*, Geophysical Prospecting, 44, 313-350.
- STAINFORTH, D.A., AND COAUTHORS (2005). *Uncertainty in predictions of the climate response to rising levels of greenhouse gases*, Nature 433, 403-406.
- TEBALDI, C., SMITH, R., NYCHKA, D., AND MEARN, L. (2005). *Quantifying Uncertainty in Projections of Regional Climate Change: A Bayesian Approach to the Analysis of Multimodel Ensembles*, Journal of Climate, 18, 1524-1540.



- TIERNEY, L. AND MIRA, A. (1999). *Some Adaptive Monte Carlo Methods for Bayesian Inference*, *Statistics in Medicine*, 18, 2507-2515.
- UNIVERSITY CORPORATION FOR ATMOSPHERIC RESEARCH (2007). *Windows for the Universe*, <http://www.windows.ucar.edu>
- VILLAGRAN, A., HUERTA, G., JACKSON, C.S., AND SEN, M. (2008). *Computational Methods for Parameter Estimation in Climate Models*, *Bayesian Analysis*, 3, 823-850.
- WANG, F. (2007). *Investigating ENSO sensitivity to mean climate in an intermediate model using a novel statistical technique*, *Geophysical Research Letters*, 34, L07705.
- YARDIM, C., GERSTOFT, P. AND HODGKISS, W. (2006). *Statistical Sea-Borne Duct Estimation Using a Hybrid Genetic Algorithms-Markov Chain Monte Carlo Method*, Technical Report, University of California.