Inferring Climate System Properties Using a Computer Model

Bruno Sansó[*], Chris E. Forest[†] and Daniel Zantedeschi[‡]

Abstract.   A method is presented to estimate the probability distributions of climate system properties based on a hierarchical Bayesian model. At the base of the model, we use simulations of a climate model in which the outputs depend on the climate system properties and can also be compared with observations. The degree to which the model outputs are "consistent" with the observations is used to obtain the likelihood for the climate system properties. We define the climate system properties as those properties of the climate model that control the large-scale response of the climate system to external forcings. In this paper, we use the MIT 2D climate model (MIT2DCM) to provide simulations of ocean, surface and upper atmospheric temperature behavior over zones defined by latitude bands. In the MIT2DCM, the climate system properties can be set via three parameters: Climate sensitivity (the equilibrium surface temperature change in response to a doubling of CO2 concentrations), the rate of deep-ocean heat uptake (as set by the diffusion of temperature anomalies into the deep-ocean below the climatological mixed layer), and net strength of the anthropogenic aerosol forcings. In this work, we use output from MIT2DCM coupled with historical temperature records to make inference about these climate system properties. Even though the MIT2DCM is far less computationally demanding than a full 3D climate model, the task of running the model for each combination of the climate parameters and processing its output is computationally demanding. Thus, a statistical model is required to approximate the model output. We obtain results that are critical for understanding uncertainty in future climate change and provide an independent check that the information contained in recent climate change is robust to statistical treatment.

Keywords: model calibration, climate change, climate sensitivity, Bayesian methods

1   Introduction

Two major points about our knowledge of the climate system were summarized in the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC 2007). First, the global average annual temperature of the Earth's surface has increased by as much as 0.7 degrees Celsius since the late 19th century. Second, most of this observed warming is attributed to human influence on the climate system since

[*]Department of Applied Mathematics and Statistics, University of California Santa Cruz, Santa Cruz, CA, http://www.ams.ucsc.edu/~bruno/

[†]Joint Program on the Science and Policy of Global Change, Massachussets Institute of Technology, Boston, MA, mailto:ceforest@mit.edu

[‡]Department of Applied Mathematics and Statistics, University of California Santa Cruz, Santa Cruz, CA, http://www.ams.ucsc.edu/~danielz/

the beginning of the industrial revolution with anthropogenic factors (primarily increasing greenhouse gas concentrations) being partly responsible for the warming. The first point is based on evidence in the observational records of surface temperatures with similar evidence from temperatures in both the upper-atmosphere and ocean temperature records beginning in the 1950s. The second point is supported by the literature on climate model experiments in which the simulations of the historical period ($\sim$1850 to present) are compared against the observational record. In short, the historical record cannot be explained by natural climate forcings or unforced climate variability. Thus, only simulations with both anthropogenic and natural forcings can best replicate the past changes in temperature. Furthermore, experiments with these same models predict even larger warming in the future if the concentrations of greenhouse gases are not stabilized by reducing emissions.

The subject of this paper is to use climate change observations to explore the distributions for properties of the climate system that control the simulations of future climate change in response to expected climate forcings. These simulations provide information on long-term climate predictions and are made by simulating the future state of the global climate system using a climate model. A hierarchy of climate models exists (Claussen et al. 2002) that classifies the range of model complexities according to the simulated spatial scales and climate-relevant processes. As discussed in Section 10.5.1 of Chapter 10 in the IPCC AR4 (Meehl et al. 2007), these classes of models are listed as Simple Climate Models (SCMs), Earth-system Models of Intermediate Complexity (EMICs), and Atmosphere-Ocean General Circulation Models (AOGCMs). Depending on the scale required or the key processes to be included, one can choose a model from one of these classes for a given project. For example, to simulate the full spatial distribution of climate variables (e.g., temperature, precipitation, or winds), three-dimensional AOGCMs are required. These models discretize the earth's atmosphere, oceans, and land into grid boxes that have a typical size of 250 km $\times$ 250 km. We note that many climate relevant phenomena having spatial scales smaller than 250 km (in particular clouds) occur within those grid boxes and are accounted for by *parameterizations*. Changing the parameterizations will change the sub-gridscale effects and can affect the large-scale model output substantially. We will return to this issue later.

On the other extreme, if only the global or hemispheric mean temperatures are required, a SCM can be an appropriate model for simulations provided details on smaller scales can be disregarded. In this case, processes smaller than hemispheric scales are parameterized (e.g., all mixing processes within and between the atmosphere or oceans) such as is done for processes within an AOGCM grid-box. These models are extremely computationally efficient and are well suited for testing climate policy scenarios when a full AOGCM is not required. The disadvantage is that many processes are highly parameterized and so key feedbacks between climate sub-systems may not be adequately represented in simulations. For example, the solubility of carbon dioxide in sea-water is highly temperature dependent and so hemispheric mean temperatures would not distinguish between polar and tropical regions where temperatures determine the flux of CO2 into the ocean. When the polar warming is 2-3 times stronger than the tropical warming, this polar amplification would need to be parameterized in an SCM.

This previous example highlights the need for the third class of climate models, EMICs. EMICs have been developed to explore research questions where SCMs are too simple and AOGCMs are too computationally inefficient. Typically, EMICs have high latitude resolution to resolve the tropical to polar temperature changes but only simplified longitude representation of the land and ocean distributions. This intermediate spatial resolution between SCMs and AOGCMs provides a more computationally efficient numerical structure for simulating the climate system. In this paper, we use an EMIC that includes parameterizations identical to those in AOGCMs while only simulating the zonal-mean state of the climate on a latitude-height grid. More details will be described in Section 2.

Common to all these climate models are properties that determine the large-scale response to climate forcings (e.g., increasing concentrations of greenhouse gases). These are typically characterized by two properties: Climate sensitivity and the rate of heat uptake by the deep-ocean. Climate sensitivity is defined as the *equilibrium* global-mean annual-mean surface temperature change in response to a doubling of CO2 concentration and is denoted by $\mathcal{S}$. This is the temperature response after all transient processes come into balance, such as ocean and land surface. The rate of heat uptake by the deep-ocean is the controlling process that limits the rate of surface warming as it approaches equilibrium. This is determined by all processes that mix excess heat from the surface ocean into the deeper layers and can be characterized by an effective diffusion coefficient denoted by $\mathcal{K}_v$. Together, these two properties determine the transient response of the climate system to changes in large-scale radiative forcings.

Depending on the climate model being used, these two properties can either be directly set via parameters in the model or be estimated from simulations using the full system. In the latter case, the combination of different parameterizations (and their internal parameter settings) ultimately determines these climate system properties and so the dynamical response of the model cannot be estimated *a priori*. In this case, there may be many combinations of parameters that lead to the same climate system properties and so a thorough search of the parameter space would be required for calibration purposes. In this paper we consider an EMIC climate model for which the climate system properties can be adjusted via single parameters. This will be discussed in Section 2.

An additional requirement is to include the uncertainty in the external radiative forcing. For the calibration of a climate model to the observed temperature record, we must include the uncertainty in the forcing over the same time period. This forcing uncertainty arises mainly from the poor knowledge of the historical record for aerosols. Aerosols are the suspended particulate matter in the atmosphere that have a short lifetime (approx. days to months) and can reflect or absorb solar and infrared radiation (see Forster et al. (2007).) A record of the concentrations for aerosols does not exist (unlike greenhouse gases) and so we rely on the emissions record for the various species to estimate the radiative forcing. As such we refer to the uncertainty in the net anthropogenic aerosol forcing, written here as $\mathcal{F}_{aer}$. This primarily represents the uncertainty in all aerosols but also can represent any unmodeled external forcings. Together with $\mathcal{S}$ and $\mathcal{K}_v$, these are three parameters that can be calibrated against the historical record of

observed temperatures. As alluded to before, our goal is to provide a full quantification of the probabilistic uncertainties for these three parameters using output from the MIT 2D climate model (Sokolov and Stone (1998a); Forest et al. (2006)) (MIT2DCM), historical records, GCM simulations and expert judgment.

The MIT2DCM provides simulations of ocean, surface and upper atmospheric temperature behavior. The model uses a system of latitude and height coordinates to simulate the average state of the climate over zones defined by latitude bands. Details of the structure of the model are presented in Section 2. The MIT2DCM averages over longitude, but its output matches climate observations and produces similar predictions as those of full 3D atmosphere-ocean general circulation models (GCM). The computer effort involved in running the MIT2DCM is a fraction of that of a full 3D model. So, it is relatively easy to explore different values for the climate system properties, that are represented by low dimensional parameters.

Typical output from a run of the MIT2DCM consists of temperatures at 46 different latitudes, with 11 vertical layers for the 1860-1995 period, every 30 minutes. In this paper we consider three summaries of such output that we refer to as "diagnostics": A vector of 288 components, consisting of the upper air temperature changes between the 1986-1995 and 1961-1980 periods at 36 latitudes and 8 levels; Surface temperature change, consisting of the difference between the decadal average temperatures for 1946-1995 periods and the average temperature of 1906-1995 at 4 different equal-area zonal bands, resulting in a 20 dimensional vector; and deep ocean temperature trend, calculated for the 1952-1995 period. Historical observations and GCM output are obtained in correspondence to the three diagnostics. We notice that in all three cases we avoid using absolute temperature values and consider some measure of temperature change. This is a common practice among climatologists to account for systematic biases in the model output. See Tebaldi and Sansó (2008) for an illustration of these facts.

Figure 1 shows summaries of the three diagnostics obtained from the MIT2DCM simulations performed on a three dimensional grid of 426 points for $\boldsymbol{\theta} = (\sqrt{\mathcal{K}_v}, \mathcal{S}, \mathcal{F}_{aer})$. For comparison purposes we superimpose each diagnostic, calculated from historical records, to the model results. In essence our problem is to find the values of $\boldsymbol{\theta}$ that make the model output as similar to the historical records as possible. More precisely, for a given diagnostic, denote by $\boldsymbol{z}$ the observation and by $\boldsymbol{\eta}(\boldsymbol{\theta})$ the output of the model for a given $\boldsymbol{\theta}$. A likelihood for $\boldsymbol{\theta}$ is obtained by assuming that the difference is Gaussian. Using prior information on the climate properties from the literature, we obtain a posterior for $\boldsymbol{\theta}$. Unfortunately, output from the MIT2DCM is available only at a few hundreds pre-specified points on an irregular grid. Also, running and post-processing times needed to calculate $\boldsymbol{\eta}(\boldsymbol{\theta})$ prevents us from embedding its evaluation within an iterative method. So, to fully explore the posterior distribution of $\boldsymbol{\theta}$ we create an auxiliary statistical model that provides an approximation to $\boldsymbol{\eta}$. For this purpose we use a Gaussian process. This is justified, from a Bayesian viewpoint, by the fact that for a given $\boldsymbol{\theta}$ the value of $\boldsymbol{\eta}$ is unknown, so we may consider it as a random process. The setting of our problem is that of calibration of computer model parameters as described in Kennedy and O'Hagan (2001). The focus is not on prediction or data assimilation, but on inference for the parameters that control the computer model and that have a
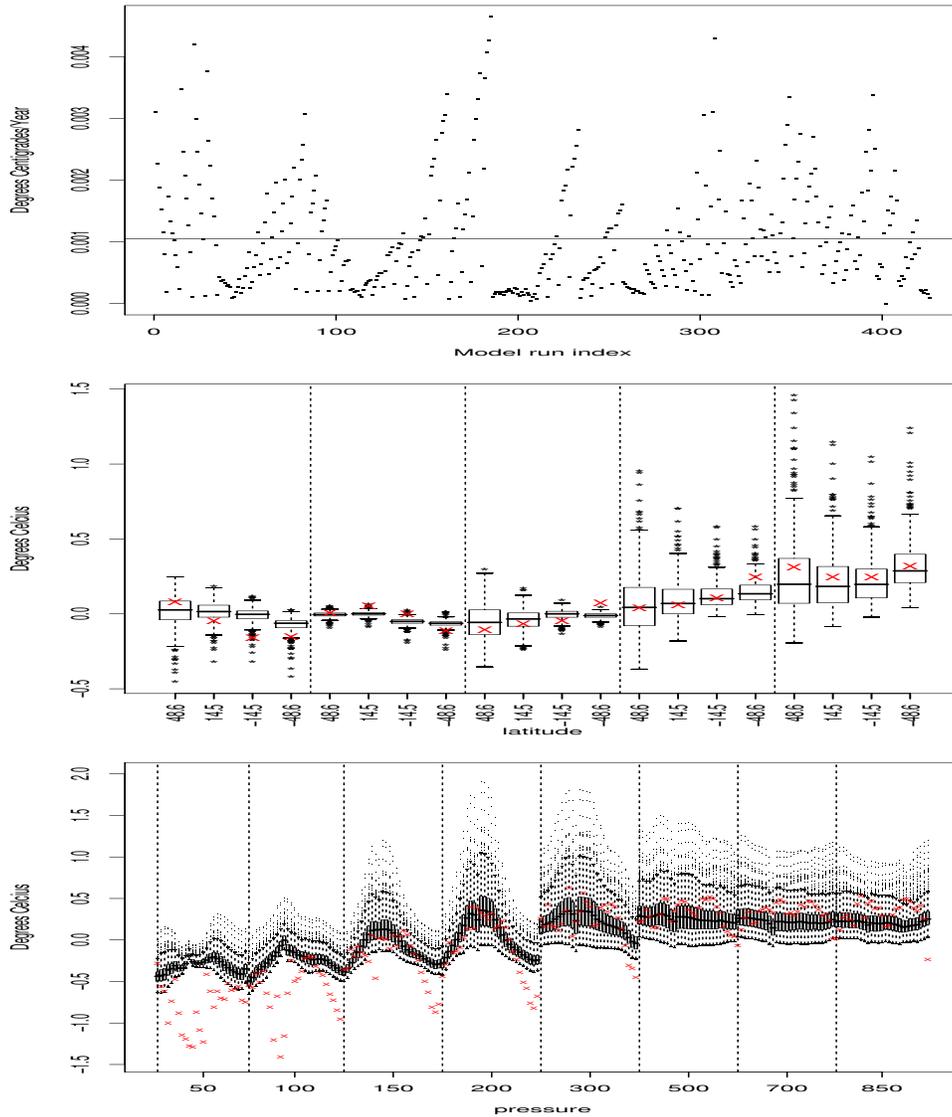
Figure 1: Top: Deep ocean temperature trends diagnostic. The index corresponds to the output for the 426 different combinations of the climate parameters. The horizontal line corresponds to the observation. Center: Surface temperature diagnostic. The boxplots correspond to the model output for the five decades and four latitude bands. The crosses correspond to the observations. Bottom: Upper air temperature diagnostics. The boxplots correspond to the model output for the different latitudes and pressure bands. The observations are marked with crosses.

precise physical meaning. For two of the diagnostics the output is multivariate. So the calibration procedure has to incorporate information about a covariance matrix. We use GCM output to elicit a prior distribution for such matrices.

In the following section we provide the scientific background for the MIT2DCM output that we use in this paper. Our statistical analysis builds on previous work aimed at estimating the posterior distributions of climate system properties. This is described in Section 2.4. The statistical model we consider in this work applies the methods proposed in Sansó et al. (2007), which are described in Section 3. As for most calibration problems, we expect the prior information on $\boldsymbol{\theta}$ to be crucial for the estimation of the climate system properties. In this paper we focus on estimating the amount of information that is provided by the data. We do so by quantifying the overlap between prior and posterior distributions using Bhattacharyya distances, which are defined in Section 3.2. Our results for each of the three diagnostics are presented in Section 4. Our conclusions are presented in the last section.

2   Scientific background

The description of the generation of the MIT2D data follows a series of papers. The climate model was initially described in Sokolov and Stone (1998a) with an update to the ocean model described in Forest et al. (2006). The description of the forcings applied for the 1860-1995 period is given in the supplemental material from Forest et al. (2006). Here we provide a brief description of the model, the historical forcings, the experimental design, and a discussion of other methods used for estimating the probability distributions for climate sensitivity.

2.1   Description of MIT 2D Climate Model

The MIT 2D climate model consists of a zonally-averaged atmospheric GCM coupled to a mixed-layer Q-flux ocean model, with heat anomalies diffused below the mixed-layer. The model details can be found in Sokolov and Stone (1998b). The atmospheric model is a zonally averaged version of the Goddard Institute for Space Studies (GISS) Model II general circulation model (Hansen et al. 1983) with parameterizations of the eddy transports of momentum, heat, and moisture by baroclinic eddies (Stone and Yao 1987, 1990). The model version we use has 46 latitude bands, for a resolution of 4°, and 11 vertical layers with 4 layers above the tropopause (above 200 hPa).

The model also employs the 2.5D Q-flux ocean mixed layer model with 4°x 5°latitude-longitude grid cells and diffusion of heat anomalies into the deep-ocean below the climatological mixed layer. Allowing for changing sea-ice in multiple grid cells, this provides smooth melting transitions as compared to a single zonal-mean ocean grid-cell. This ocean component model is a Q-flux mixed layer model and is further described by Hansen et al. (1984). Conceptually, this is similar to a "slab ocean" model used by many atmospheric GCMs when a full 3D ocean model is not required. The ocean's surface mixed layer temperature profile is represented in the MIT model which distinguishes

this from a typical slab-ocean model. Briefly, a Q-flux is defined as the additional heat flux required to maintain the seasonal cycle of sea-surface temperatures (SST) at a given latitude for the present-day climate. Physically, this primarily represents the horizontal heat transport not explicitly included in the model. The atmospheric model uses the GISS radiative transfer code which contains all radiatively important trace gases as well as aerosols and their effect on radiative transfer. The surface area of each latitude band is divided into a percentage of land, ocean, land-ice, and sea-ice with the surface fluxes computed separately for each surface type. This allows for appropriate treatment of radiative forcings dependent on underlying surface type such as anthropogenic aerosols. The atmospheric component of the model, therefore, provides most important nonlinear interactions between components of the atmospheric system.

As discussed previously, the MIT 2D model has two parameters that determine the timescale and magnitude of the decadal to century timescale response to an external forcing. These are the equilibrium climate sensitivity ($\mathcal{S}$) to a doubling of CO_2 concentrations and the global-mean vertical thermal diffusivity ($\mathcal{K}_v$) for the mixing of thermal anomalies into the deep ocean. We note that $\mathcal{S}$ is set by adjusting a cloud feedback parameter, k, that adjusts the cloud fractions used in the radiative transfer calculations. Specifically, we set the adjusted cloudiness, C', by multiplying the model calculated cloudiness, C_o, by a term proportional to the global mean surface temperature, ΔT to obtain: $C' = C_o(1 + k\Delta T)$. In this manner, we vary k to adjust $\mathcal{S}$ and have a one-to-one mapping for specifying $\mathcal{S}$. We also note that the vertical thermal diffusivity for heat anomalies has a latitude-longitude dependence that is scaled by the global-mean $\mathcal{K}_v$ value. This implies that adjustments for different relative rates of heat uptake in different regions are not permitted with this design. In sensitivity tests to changes in this pattern of deep-ocean heat uptake, we found the global mean changes in ocean temperatures were insensitive to such changes.

2.2 Temperature Change Diagnostics

As mentioned in the introduction, we have elected to use three climate change diagnostics, which are the same as used in Forest et al. (2006). This allows us to isolate the effect of the change in the calibration algorithm on the posterior distributions. The climate change diagnostics used in Forest et al. (2006) were:

- **Deep-ocean temperatures:** trend in global-mean 0–3km deep layer of pentadal averages from 1952–1995. Source of observational records: Levitus et al. (2005). We will refer to this diagnostic as DO.

- **Surface temperatures:** 4 equal-area latitude averages for each of five decades from 1946–1995 referenced to 1905-1995 climatology. Source of observational records: Jones et al. (1999). We will refer to this diagnostic as ST.

- **Upper-air temperatures:** Difference between 1986–1995 and 1961–1980 averages at eight standard pressure levels from 850-50 hPa on 5 degree grid. Source of observational records Parker et al. (1997). We will refer to this diagnostic as UA.

Observations for the upper air diagnostic are missing for a number of latitude-height coordinates, mostly in the Southern Hemisphere. This reduces the dimension of the diagnostic from 288 to 220. Additionally, by looking at the bottom panel of Figure 1 we observe that the distributions of modeled temperatures differ between the stratospheric (50-200 hPa) and the tropospheric layers (300-850 hPa). In the stratospheric data, the modeled temperatures are less sensitive to $\boldsymbol{\theta}$ (primarily, $\mathcal{S}$) as shown by the smaller range of temperature changes at these levels. The weak dependence on $\mathcal{S}$ is consistent with the local cooling being driven by the decrease in stratospheric ozone concentrations (not the changes in CO2 or other greenhouse gas concentrations). Recalling that different values of $\mathcal{S}$ are mainly due to different cloud feedbacks in the troposphere, stratospheric temperatures will not vary significantly as $\mathcal{S}$ is varied. But, we also see that the stratospheric temperatures depend more on latitude as compared to the tropospheric regions and so the information content of this data differs between regions. As such, the modeled tropospheric temperatures provide less information for the likelihood estimates. To address these differences, we restrict ourselves to using only the four layers that correspond to 50–200 hPa pressure levels (or altitudes from about 11 to 20 km). This amounts to requiring that we get the stratospheric cooling correct (and its latitude dependence) and neglect the weak tropospheric warming.

2.3   Summary of Applied Climate Forcings

The current set of simulations uses a set of historical climate forcings during the period 1860-1995. The external forcings are changes in: greenhouse gas concentrations et al. (2002), sulfate aerosol loadings scaled by SO2 emissions Smith et al. (2003), tropospheric and stratospheric ozone concentrations (et al. 2002), land-use vegetation changes (Ramankutty and Foley 1999), solar irradiance changes (Lean 2000), and stratospheric aerosols from volcanic eruptions (Sato et al. 1993, updated to 2001). GSOLSV is the shorthand notation for this set of forcings.

2.4   Previous Estimation Work

Section 9.6 in Chapter 9 of the IPCC AR4 (Hegerl et al. 2007) provides a summary of estimates of probability distributions for climate sensitivity taken from the recent literature. Section 10.5 in Chapter 10 of the IPCC AR4 (Meehl et al. 2007) further discusses the uncertainties in quantifying climate change projections by accounting for additional model uncertainties. In this paper, we are restricting the discussion to those methods that use the 20th century instrumental record to estimate the distribution of climate sensitivity. Five groups have published results in this category (Andronova and Schlesinger 2001; Gregory et al. 2002; Knutti et al. 2003; Forest et al. 2006; Frame et al. 2005). (Note we exclude Forster and Gregory (2006) because it does not use the record back to beginning of 20th century.) The underlying methods in all of these are very similar to the basic methodology used in this paper. Namely, a set of climate simulations were made with a climate model with various settings for $\mathcal{S}$. In each case, the simulations were compared against the observed temperature record and

assigned a likelihood that it is consistent with the observations. From the set of simulations, a distribution, $p(\mathcal{S})$, is estimated. Several choices are made in each method: What climate model is used to perform the simulations? What observational diagnostic is used? and What noise model is required to determine consistency between the simulation and observations? In all cases except Forest et al. (2006) and Frame et al. (2005), global mean temperature is used as the diagnostic variable and so only univariate statistical methods are required. We summarize these methods in Table 1. [1]

Reference	(1)	(2)	(3)
Model	EBM 24 zones	EMIC Bern-2D	EBM global
Diagnostic	global ΔT_{sfc} global ΔT_{ocean}	$\Delta T_{NH} + \Delta T_{SH}$, $\Delta T_{NH} - \Delta T_{SH}$	global ΔT_{sfc}, ΔT_{sfc} (for Faer), ΔT_{ocean}
Noise Model	Bootstrap using deviations from EBM	observational uncertainty	uncertainty in: ΔT_{sfc}, ΔF, ΔQ_{sfc}
Properties	$\mathcal{S}$, $\mathcal{F}_{aer}$	$\mathcal{S}$, $\mathcal{F}_{aer}$	$\mathcal{S}$

Reference	(4)	(5)	
Model	EBM global	EMIC MIT 2DLO	
Diagnostic	Attrib. Warming ΔT_{sfc}, ΔT_{sfc} (for Faer), ΔT_{ocean}	$\Delta T_{sfc}(lat., time)$, $\Delta T_{upper-air}$ (lat., height), ΔT_{ocean}	
Noise Model	Bootstrap using control data	observational lontrol data	
Properties	$\mathcal{S}$, $\mathcal{K}_v$	$\mathcal{S}$, $\mathcal{K}_v$, $\mathcal{F}_{aer}$	

Table 1: Summary of Previous Approaches for Estimates of $p()$. The references correspond to: (1) = Andronova and Schlesinger (2001); (2) = Knutti et al. (2002) ; (3) = Gregory et al. (2002); (4) = Frame et al. (2005) ; (5) = Forest et al. (2002)

Forest et al. (2002) presented an estimate of the joint probability density function (PDF) for uncertain climate system properties. Other groups (Andronova and Schlesinger 2001; Gregory et al. 2002; Knutti et al. 2003; Forest et al. 2006; Frame et al. 2005; Forster and Gregory 2006) have estimated similar PDFs although each uses different methods and data. However all are based on estimating the degree to which a climate model can reproduce the historical climate record. Parameters within each model are perturbed to alter the response to climate forcings and a statistical comparison is used to reject combinations of model parameters.

The method is based on the optimal fingerprint detection technique for comparing model and observational data. This technique consists of running a climate model under a set of prescribed forcings and using climate change detection diagnostics (i.e., spatio-temporal patterns of temperature changes) to determine whether the simulated climate

---

[1]Gregory et al. (2002) do make use of the optimal fingerprint detection results to estimate the net aerosol forcing from temperature change patterns but then only consider global mean temperature response to estimate $\mathcal{S}$. It appears that this amounts to double use of the surface temperature record.

change is observed in the climate record and is distinguishable from unforced variability of the climate system (see Mitchell et al. (2001) or International *ad hoc* Detection and Attribution Group (2005) and references therein). In statistical terms, letting T_{obs} be the historical records, $T(\theta)$ the model output that depends on parameters θ and C_N the covariance matrix, a likelihood function is obtained from the statistics $(T(\theta) - T_{obs})' C_N^{-1} (T(\theta) - T_{obs})$. It is not possible to estimate the true climate system variability, C_N, on century time-scales from observations and therefore, climate models are run with fixed boundary conditions for thousands of years to obtain estimates of the climate variability.

The work presented here is based on obtaining a likelihood for the climate sensitivity ($\mathcal{S}$), the rate of heat uptake by the deep ocean ($\mathcal{K}_v$), and the net aerosol forcing ($\mathcal{F}_{aer}$) and analyze the uncertainty in such parameters using a posterior distribution, after the incorporation of expert prior knowledge. This is the same framework of Forest et al. (2001, 2002). The uncertain parameters and the climate model data from the MIT2DCM are identical to those in Forest et al. (2006). The method used in Forest et al. (2006) is based on evaluating the likelihood only at a set of pre-specified combinations of the climate system properties. Also, the estimation of C_N^{-1} is done using an eigenvector decomposition of GCM runs, prior to forming the likelihood. As part of ongoing investigations into this method, Curry et al. (2005) considered the impact of using different truncation numbers in the estimate of the noise covariance inverse. This truncation represents the number of eigenvectors retained in the inverse estimate and various information criteria were used to determine an appropriate truncation value for the multivariate diagnostics. The methods developed in Sansó et al. (2007) and used in this paper interpolate over the space of climate parameters and account for most estimation uncertainties, in particular those due to the covariance matrix. Additional details are presented in Section 3.

3   Statistical model

We consider a statistical model along the lines of that developed in Sansó et al. (2007). In what follows vectors will be denoted in bold face and matrices will correspond to capital letters. Recalling the notation in the introduction, let $\boldsymbol{z} \in \mathbb{R}^n$ correspond to a summary of the historical data collected around the world over several decades. Thus $n = 1$ for DO, $n = 20$ for ST and $n = 109$ for UA (although our initial UA analysis was performed with $n = 220$). The model runs are denoted as $\boldsymbol{y}_j \in \mathbb{R}^n$, $j = 1, \ldots, p$. These are a collection of p n-dimensional vectors, one for each combination of the climate parameters. We denote this as $Y \in \mathbb{R}^{n \times p}$. For ST and UA, we consider additional information provided by CMG control runs, consisting of k n-dimensional vectors $\boldsymbol{w}_j$, and let $W \in \mathbb{R}^{n \times k}$ be the matrix of such control runs. Also, for ST and UA, each component of the n dimensional data vectors is indexed by a two dimensional location $\boldsymbol{x}_i, i = 1, \ldots, n$. $\boldsymbol{x}_i$ indicates latitude and decade for ST and latitude and altitude for UA. The values of the climate parameters where the MIT2DCM is evaluated will be denoted as $\boldsymbol{t} = (\sqrt{\mathcal{K}_v}, \mathcal{S}, \mathcal{F}_{aer}) \in \mathbb{R}^3$ so that the model output at location $\boldsymbol{x}$ and parameter values $\boldsymbol{t}$ is denoted as $\boldsymbol{\eta}(\boldsymbol{x}, \boldsymbol{t})$. A summary of the notation is presented in

Table 2.

Variable	Notation	Size	Dimension
Locations	$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$	$n = 1; 20; 109$	$\boldsymbol{x}_i \in \mathbb{R}^2$
Historical records	$z_1, \ldots, z_n$	$n = 1; 20; 109$	$z_i, \in \mathbb{R}$
Climate parameters	$\boldsymbol{t}_1, \ldots, \boldsymbol{t}_p$	$p = 426$	$\boldsymbol{t}_j \in \mathbb{R}^3$
Surface diagnostics	$\boldsymbol{y}_1, \ldots, \boldsymbol{y}_p$	$p = 426$	$\boldsymbol{y}_j \in \mathbb{R}^n$
Control runs	$\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k$	$k = 91; 162; 40$	$\boldsymbol{w}_l \in \mathbb{R}^n$

Table 2: Summary of the available information.

We denote as $\boldsymbol{\zeta}(\boldsymbol{x})$ the true, unobserved, diagnostic at location $\boldsymbol{x}$. We note that the three diagnostics refer to linear functionals, or filters, of the temperature. We denote $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ as the 'true' value of the climate system properties $(\sqrt{\mathcal{K}_v}, \mathcal{S}, \mathcal{F}_{aer})$. Then we assume that

$$z_i = \boldsymbol{\zeta}(\boldsymbol{x}_i) + \xi_i = \boldsymbol{\eta}(\boldsymbol{x}_i, \boldsymbol{\theta}) + \delta_i + \xi_i \quad i = 1, \ldots, n . \tag{1}$$

ξ_i represents observation errors, including errors on spatial and temporal scales that are smaller than the grid cell size and that are introduced by the filter. δ_i is the inadecuacy of the statistical emulator. We assume that $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)' \sim N_n(0, \tau^2 \sigma^2 \Sigma)$, where Σ is a covariance matrix and σ^2 is a scale parameter. We also assume that $\delta_i \sim N_n(0, \sigma^2 \Sigma)$. For the univariate diagnostic $n = 1$ and Σ is a scalar equal to one. Note that δ_i has mean zero. Thus we do not consider possible systematic additive biases of the computer code, since, as mentioned earlier, these should be accounted for by the fact that the diagnostics measure temperature change. In practice we found it difficult to separate observational and inadequacy errors without assuming τ^2 known. Thus we consider only one error term.

The value of $\boldsymbol{\eta}(\boldsymbol{x}, \boldsymbol{\theta})$ is unknown for general $(\boldsymbol{\eta}, \boldsymbol{\theta})$, so we treat it as a random process. More specifically, we assume that $\boldsymbol{\eta}(\boldsymbol{x}, \cdot)$ corresponds to a Gaussian process. We model such process by specifying a mean and a stationary covariance function. This is the statistical equivalent model that allows for fast approximations to the numerical simulator. We assume that $E(\boldsymbol{\eta}(\boldsymbol{x}, \boldsymbol{t})) = \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{t})' \boldsymbol{\beta}$, where $\boldsymbol{h}$ and $\boldsymbol{\beta}$ are q-dimensional vectors. They define a linear combination of effects due to locations and parameter values. The covariance function is given by

$$\text{cov}(\boldsymbol{\eta}(\boldsymbol{x}_i, \boldsymbol{t}), \boldsymbol{\eta}(\boldsymbol{x}_j, \boldsymbol{t}')) = r(\boldsymbol{t}, \boldsymbol{t}') \sigma^2 \gamma^2 \Sigma_{ij}, \tag{2}$$

for some correlation function $r(\cdot, \cdot)$. Thus, we are assuming separability between $\boldsymbol{x}$ and $\boldsymbol{t}$. Since there are 426 different points $\boldsymbol{t}$ the full covariance matrix for ST is of dimension $8,520 \times 8,520$ and of dimension $46,434 \times 46,434$ for UA, so, separability is key to building a model that can be fitted with an iterative procedure. For ST and UA, we assume that $\boldsymbol{w}_j \sim N_n(\boldsymbol{0}, \Sigma)$. This yields the prior distribution for Σ,

$$p(\Sigma) \propto \exp \left\{ -\frac{k}{2} \text{tr}(\Sigma^{-1} S) \right\} |\Sigma|^{-k/2} ,$$

where $S = 1/k \sum_{j=1}^{k} \boldsymbol{\omega}_j \boldsymbol{\omega}_j'$. Note that $p(\Sigma)$ is proper for $k \geq 2n + 1$. For ST we will use two control runs, one corresponds to $k = 162$ and the other to $k = 91$. So in both cases the prior is proper. For UA we use a control run with $k = 40$, implying that $p(\Sigma)$ is not proper. Note that we are assuming that the covariances of the observations and the climate models are proportional. Thus, the dependence structure in the temperature diagnostics is captured by all three sources of information. Forest et al. (2006) implicitly make the same assumption. This is usually referred to in the climate literature as unforced or natural variability and cannot be estimated from the observational data alone because the records are too short for the 50-100-year timescales of interest. An important output of our model is an estimate of the unforced variability, based on all three available types of data.

Define a matrix $R \in \mathbb{R}^{p \times p}$ such that $R_{ij} = r(||\boldsymbol{t}_i - \boldsymbol{t}_j||)$. Let $\boldsymbol{r}(\boldsymbol{\theta}) = (r(||\boldsymbol{t}_1 - \boldsymbol{\theta}||), \ldots, r(||\boldsymbol{t}_p - \boldsymbol{\theta}||))'$. Let $H(\cdot) = [\boldsymbol{h}(\boldsymbol{x}_1, \cdot), \ldots, \boldsymbol{h}(\boldsymbol{x}_n, \cdot)] \in \mathbb{R}^{q \times n}$, and let $H = [H(\boldsymbol{t}_1), \ldots, H(\boldsymbol{t}_p)]' \in \mathbb{R}^{np \times q}$, then using equations (1) and (2),

$$\begin{pmatrix} \boldsymbol{z} \\ \mathrm{vec}(Y) \end{pmatrix} \sim N_{n(1+p)} \left(\begin{pmatrix} H(\boldsymbol{\theta})' \\ H \end{pmatrix} \boldsymbol{\beta}, \; \sigma^2 \begin{pmatrix} (1+\gamma^2)\Sigma & \gamma^2 \boldsymbol{r}(\boldsymbol{\theta})' \otimes \Sigma \\ \gamma^2 \boldsymbol{r}(\boldsymbol{\theta}) \otimes \Sigma & \gamma^2 R \otimes \Sigma \end{pmatrix} \right), \quad (3)$$

where $\mathrm{vec}(\cdot)$ denotes the operation of stacking the columns of a matrix into a vector and $\otimes$ denotes the Kronecker product.

There is no natural distance in the space of climate parameters that will help us define a correlation function, so we consider each component separately and let

$$r(\boldsymbol{t}, \boldsymbol{t}') = r_1(t_1 - t_1'; \phi_1) r_2(t_2 - t_2'; \phi_2) r_3(t_3 - t_3'; \phi_3),$$

where $r_i(t_i, t_i') = \exp\{-1/\phi_i |t_i - t_i'|\}$. Here ϕ_i measures the correlation range in the same units as t_i. So, large values of ϕ_i imply that the correlation will be small only for points that are very far apart. The separability assumption is very common in the literature of statistical modeling of computer output, see, for example, Paulo (2005), who discusses the choice of default priors for the parameters of a separable correlation function in the Matèrn class.

3.1   Prior distributions

Calibration problems are known to be ill posed in the sense that often times different configurations of parameter values produce similar results. Fortunately in this application knowledge about likely values of the climate parameters is available. So we can specify scientifically sound priors for such parameters.

The prior for $\theta_1 = \sqrt{\mathcal{K}_v}$ corresponds to a beta distribution with parameters $(3.5, 6)$ that is transformed to have a support on $(0, 6)$ and . The prior for $\theta_2 = \mathcal{S}$ is specified as a beta distribution with parameters $(2.85, 14)$, transformed to have a support on $(0, 15)$. The prior for $\theta_3 = \mathcal{F}_{aer}$ is a beta distribution with parameters $(4, 4)$ transformed to have a support on $(-1.5, .5)$. With the exception of the prior on $\theta_2 = \mathcal{S}$, we based the distributions for $\sqrt{\mathcal{K}_v}$ and $\mathcal{F}_{aer}$ partly on the previous work in Forest et al. (2002)

and Forest et al. (2006). The widths were chosen to extend well outside the range suggested by likelihoods from Forest et al. (2006) while the shapes were designed to be rather diffuse in the interior (i.e., the cumulative density function approximately linear.) These ranges are also supported by the locations of the state-of-the-art 3D GCMs well within the parameter space (Sokolov et al. 2003). The likelihoods of a model outside these regions are near zero. For the prior on $\mathcal{S}$, we use the Webster and Sokolov (2000) estimate as based on the expert elicitation study of Morgan and Keith (1995). These results were based on the understanding of climate science experts in the early 1990's who would have considered model results as well as changes during the 20th century and the glacial-interglacial records for about the past 500,000 years.

Regarding the other parameters in the model, we have already discussed the prior we use for Σ, which is based on GCM simulations. For the regression parameters β we assume a flat prior $p(\boldsymbol{\beta}) \propto 1$. For the scale parameters σ^2 and γ^2 we notice from Equation (3) that the product $\gamma^2\sigma^2$ can be factorized from three of the four blocks in the covariance matrix. This implies that the two parameters are nearly un-identifiable. In practice, separate estimation of σ^2 and γ^2 is possible only by using very informative priors. In this application we fix $\gamma^2 = 1$ and let $p(\sigma^2) \propto 1/\sigma^2$. The priors for the range parameters ϕ_i are $p(\phi_i) \propto 1/\phi_i, i = 1, \ldots, 3$. Berger et al. (2001) showed that posterior impropriety could result from the choice of an improper prior for the range parameter of an isotropic Gaussian field. The results in Paulo (2005) show that impropriety of the posterior is not a problem when separability of the correlation function is assumed.

3.2 Bhattacharyya distances

A typical feature of calibration problems for computer models is that prior information on the parameters to be calibrated is usually fairly influential on the posterior results. Quantifying the overlap between prior and posterior distributions for the calibration parameters provides a measure of the information in the data and the model about such parameters.

A measure of the overlap between two probability densities is given by the Bhattacharyya distance. This is defined, for two densities $f_1(x)$ and $f_2(x)$ as

$$B(f_1, f_2) = \int \sqrt{f_1(x)f_2(x)}dx = \int \sqrt{\frac{f_1(x)}{f_2(x)}}f_2(x)dx = \int \sqrt{\frac{f_2(x)}{f_1(x)}}f_1(x)dx.$$

Clearly $B(f_1, f_2) = B(f_2, f_1)$, $B(f_1, f_2) \geq 0$ and, by Jensen's inequality, $B(f_1, f_2) \leq 1$. Moreover $B(f_1, f_2) = 1$ if and only if $f_1 = f_2$ and $B(f_1, f_2) = 0$ if and only if f_1 and f_2 have no overlap. So $B(f_1, f_2)$ provides a bounded symmetric measure of the similarity between two distributions. Zhou and Sansó (2008) use it two compare the outputs of an atmospheric transport model run under two different boundary conditions. They found that $B(f_1, f_2)$ provides similar information to the percentage of overlap between the highest density interval corresponding to f_1 and that corresponding to f_2 for a given probability. For multivariate settings, $B(f_1, f_2)$ is easier to calculate than interval overlaps.

In our application f_i correspond to either prior or posterior distributions of the climate parameters. The latter are known only up to a proportionality constant. Let $f_i(x) = q_i(x)/c_i, i = 1, 2$, then

$$B(f_1, f_2) = \left(\int \sqrt{f_1(x)f_2(x)}dx \int \sqrt{f_1(x)f_2(x)}dx \right)^{1/2} =$$

$$\left(E_{f_2}\sqrt{\frac{q_1(X)}{q_2(X)}\frac{c_2}{c_1}}E_{f_1}\sqrt{\frac{q_2(X)}{q_1(X)}\frac{c_1}{c_2}} \right)^{1/2} = \left(E_{f_2}\sqrt{\frac{q_1(X)}{q_2(X)}}E_{f_1}\sqrt{\frac{q_2(X)}{q_1(X)}} \right)^{1/2},$$

where E_{f_i} denotes the expectation with respect to the density f_i. So, it is possible to obtain an approximation to $B(f_1, f_2)$ from samples of f_1 and f_2, using the law of large numbers. When $X = (Z, W)$, so that X consists of two block of random variables, the marginals $f_i(z)$ can be can be calculated as

$$f_i(z) = \int f_i(z|w)f_i(w)dw \approx \frac{1}{J}\sum_{j=1}^{J} f(z|w^{(j)})$$

where $w^{(j)} \sim f_i$. Note that $f_i(z|w)$ will usually be the full conditional of z.

3.3   Implementation

The details of the implementation of the statistical model follow along the lines of Sansó et al. (2007). We run a similar MCMC to sample from the posterior distribution of all parameters in the model, for each one of the diagnostics separately. In order to obtain the posterior distribution of θ based on the information on more than one diagnostic, we assume conditional independence. Thus, using a super-index one or two to denote the quantities that define the statistical model for two different diagnostics, we have that the joint likelihood $f(z^1, Y^1, z^2, Y^2|\theta, \Xi_1, \Xi_2)$ is factorized as the product of $f_1(z^1, Y^1|\theta, \Xi_1)$ and $f_2(z^2, Y^2|\theta, \Xi_2)$, where Ξ_i denotes all the parameters other than θ used in the model for diagnostic i. Given a prior $p(\theta, \Xi_1, \Xi_2) = p(\theta)p(\Xi_1)p(\Xi_2)$ it is possible to run a MCMC to obtain samples from $p(\theta, \Xi_1, \Xi_2|z^1, Y^1, z^2, Y^2)$. This approach is computationally demanding and likely slow to converge. Additionally, we are interested in observing how the posterior distribution conditional on z^1, Y^1 is changed by the use of the diagnostic z^2, Y^2.

We start by running a MCMC to get samples of $p(\theta|z^1, Y^1)$. We then run a second MCMC for $p(\theta|z^1, Y^1, z^2, Y^2)$. In this case the full conditional for θ is proportional to $p(\theta|\Xi_1, z^1, Y^1)f_2(z^2, Y^2|\theta, \Xi_2)$. So, to obtain a sample of θ we can perform a Metropolis-Hasting step and draw a proposed sample, say θ^*, from $p(\theta|\Xi_1, z^1, Y^1)$. The acceptance ratio will be given by the likelihood ratio $f_2(z^2, Y^2|\theta^*, \Xi_2)/f_2(z^2, Y^2|\theta, \Xi_2)$. We use the approximation $p(\theta|\Xi_1, z^1, Y^1) \approx p(\theta|z^1, Y^1)$ and use the samples from the first MCMC as proposals for the second one.

		$\sqrt{\mathcal{K}_v}$	$\mathcal{S}$	$\mathcal{F}_{aer}$
	5%	0.84	0.73	-1.05
Informative Prior	50%	2.15	2.33	-0.5
	95%	3.78	5.03	0.05
	5%	0.3	0.75	-1.4
Uniform Prior	50%	3.0	7.59	-0.5
	95%	5.7	14.25	0.4

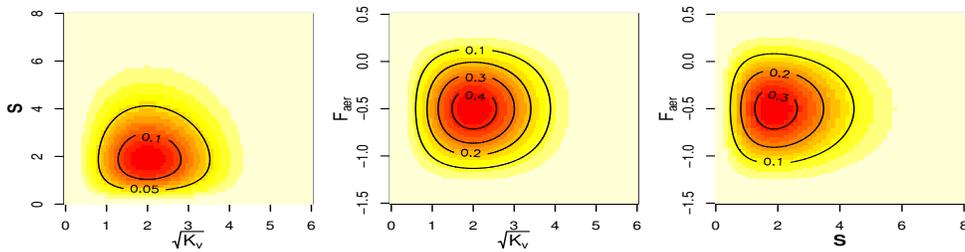Table 3: Summary of the prior distributions of the climate system properties.

Figure 2: Bivariate prior distributions for the three combinations of climate system properties.

4 Results

We considered two distributions for the climate system properties. One, that will be referred to as the uniform prior, consist of the product of three densities uniform over the ranges $(0, 6), (0, 15)$ and $(-1.5, .5)$. The second is the one described in Section 3.1, and will be referred to as the informative prior. Notice that, in both cases, we are a priori assuming that the three parameters are independent. Some quantiles that describe the characteristics of both distributions are reported in Table 3. The contours of the bivariate prior densities are presented in Figure 2.

4.1 Deep ocean temperature trend (DO)

We started by considering the DO diagnostic, the trend in deep-ocean temperatures. Here $\boldsymbol{x}$ is irrelevant (the diagnostic is a scalar), so we let $\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{t}) = \boldsymbol{h}(\boldsymbol{t}) = (1, t_1, t_2, t_3)'$, with $q = 4$. A posteriori we observe that the 2.5% and the 97.5% quantiles of the distributions of $\beta_i, i = 1, \ldots, 4$ are all positive for the informative prior. This is also the case for the 2.5% and the 97.5% posterior quantiles in the uniform prior case, with exception of the 2.5% quantile of β_1, which is equal to $-4 * 10^{-5}$. We note that this implies that β_1 is almost significantly positive, but not quite and that this is the only diagnostic where $\mathcal{K}_v$ has a positive coefficient. So, for both priors, the posterior carries substantial evidence that the baseline for DO is positive and increasing any of the three climate system properties increases DO.
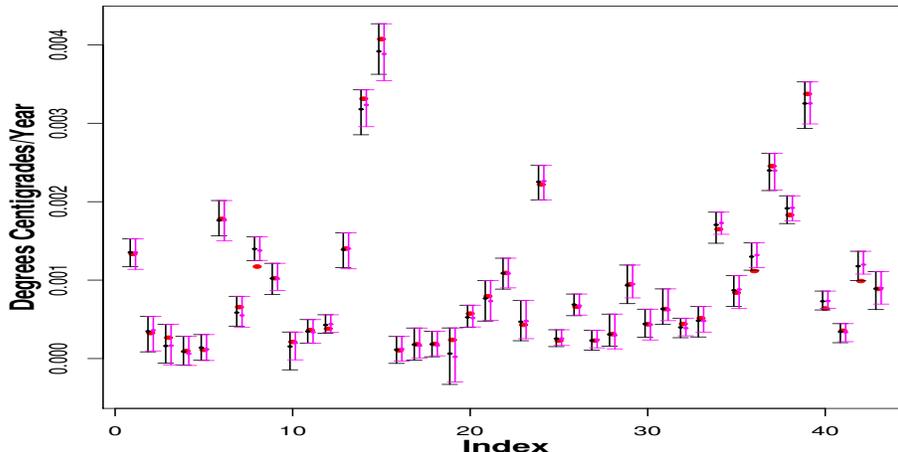
Figure 3: Predictive distribution for DO at 43 randomly selected combinations of the climate system parameters. The intervals are given by the 2.5% and 97.5% quantiles. Right hand (black) and left hand (red) intervals for informative and uniform priors respectively. The small dots correspond to the predictive mean and the larger dots to the MIT2DCM output.

To check the ability of the model to predict the values of DO we chose 43 combinations of climate parameters at random. We excluded those data from the estimation procedure and then obtained samples from their joint posterior prediction distributions. The results are shown in Figure 3. They are based on 100 samples from the predictive distribution taken from 10,000 iterations of the MCMC with a lag of 100, using as starting value the last iterates of a chain of 100,000 iterations. We observe that in all but two cases (nos. 8 and 42) the predictive interval contains the observation and that the widths of most intervals are small compared to the variability in the data.

A summary of the posterior distribution of the climate system properties for DO is presented in the first three columns of Table 4. We see substantial differences between the posteriors obtained under the informative and the uniform priors, especially for $\mathcal{S}$ and $\mathcal{F}_{aer}$. In these two cases the marginal posterior densities corresponding to the informative prior are concentrated around smaller values than the ones obtained for the uniform prior. When the values in Table 4 are compared to those in Table 3 we observe that the posterior densities follow the behavior of the priors. A better picture of this situation is provided by the plots in Figure 4. Here the posterior densities corresponding to DO are plotted with dotted cyan lines. For the informative prior we observe that the posterior densities of $\sqrt{\mathcal{K}_v}$ and $\mathcal{S}$ are very close to the priors. This is not the case for $\mathcal{F}_{aer}$, where the posterior shows substantially larger tails. For the uniform prior, all three posterior densities are pretty flat. Overall we find that the likelihood obtained from DO carries little information about the climate system properties. We observe that, for $\mathcal{S}$ the posterior density shows some irregularities, this is a consequence of the

		Deep ocean			Surface temperature			Upper air		
		$\sqrt{\mathcal{K}_v}$	$\mathcal{S}$	$\mathcal{F}_{aer}$	$\sqrt{\mathcal{K}_v}$	$\mathcal{S}$	$\mathcal{F}_{aer}$	$\sqrt{\mathcal{K}_v}$	$\mathcal{S}$	$\mathcal{F}_{aer}$
	5%	0.88	0.92	-0.99	0.82	1.75	-0.71	1.33	2.27	-0.79
Inf.	50%	2.13	2.31	-0.53	1.99	3.21	-0.39	2.59	3.79	-0.27
	95%	3.71	4.83	-0.005	3.08	5.68	-0.051	4.15	6.08	0.17
	5%	0.24	0.73	-1.37	0.19	2.26	-0.82	1.58	2.91	-1.33
Unfm.	50%	2.65	6.71	-0.80	2.03	7.23	-0.49	4.18	6.52	-0.32
	95%	5.61	14.11	0.21	5.07	14.05	-0.05	5.88	13.06	0.37

Table 4: Summary of the posterior distributions of the climate system properties.

	DO	DO Uni.	ST	ST Uni.	DO + ST	UA	UA Uni.
Prior	**0.97**	**0.89**	**0.70**	**0.59**	**0.72**	**0.49**	**0.43**
DO	1	0.48	0.67	0.34	0.76	0.30	0.10
DO Uni.		1	0.30	0.54	0.44	0.26	0.32
ST			1	0.55	0.99	0.63	0.11
ST Uni				1	0.54	0.38	0.20
DO + ST					1	0.78	0.10
UA						1	0.50

Table 5: First row in bold face: Bhattacharyya distances between the prior and the posterior distributions corresponding to the different diagnostics. All other rows: Bhattacharyya distances between the posterior distributions for different diagnostics and priors.

fact that the samples of θ_2 tended to fluctuate around the values of $\mathcal{S}$ used in the design for the runs of the MIT2DCM.

The bivariate posterior distributions for the informative prior are presented in the first row of Figure 5. These can be compared with the plots in Figure 2. We observe some elongation of the contours of the posterior density of θ_2 and θ_3, reflecting the fact that the tails of the posterior of θ_3 are larger than those of the prior. Figure 6 shows the contours of the bivariate posterior densities under the uniform prior. A quantitative measure of the similarity between the joint priors and posteriors is given by the Bhattacharyya distances reported in the first two entries of the first row of Table 5. We observe that the two numbers are close to 1, so that, especially in the case of the informative prior, the data are not changing the prior information substantially.

4.2   Surface temperature change

In order to define $\boldsymbol{h}$ for the analysis of ST, we considered $\boldsymbol{h}(\boldsymbol{t}, \boldsymbol{x}) = (1, t_1, t_2, t_3, x_1, x_2)$, where x_1 is latitude and x_2 is the decade. We used $k = 162$ control runs obtained from the GCM developed at the Hadley Centre, Bracknell, UK, code-named HadCM2. The posterior distributions of $\boldsymbol{\beta}$ show that the baseline for the surface temperature change
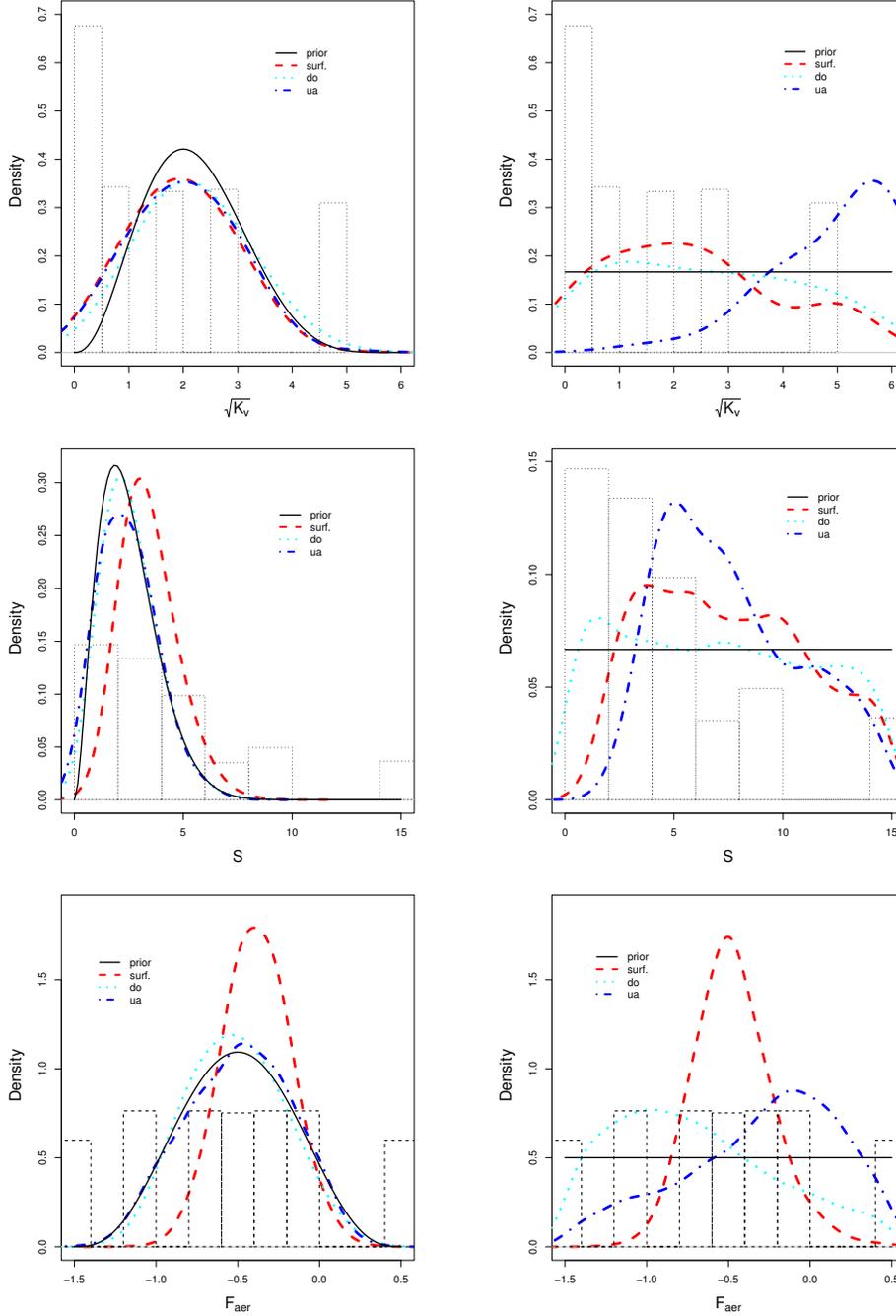
Figure 4: Marginal posterior distributions of the climate parameters for each of the three diagnostics. The left column corresponds to informative priors, the right column corresponds to uniform prior. The histograms in dashed lines correspond to the points in the experimental design.
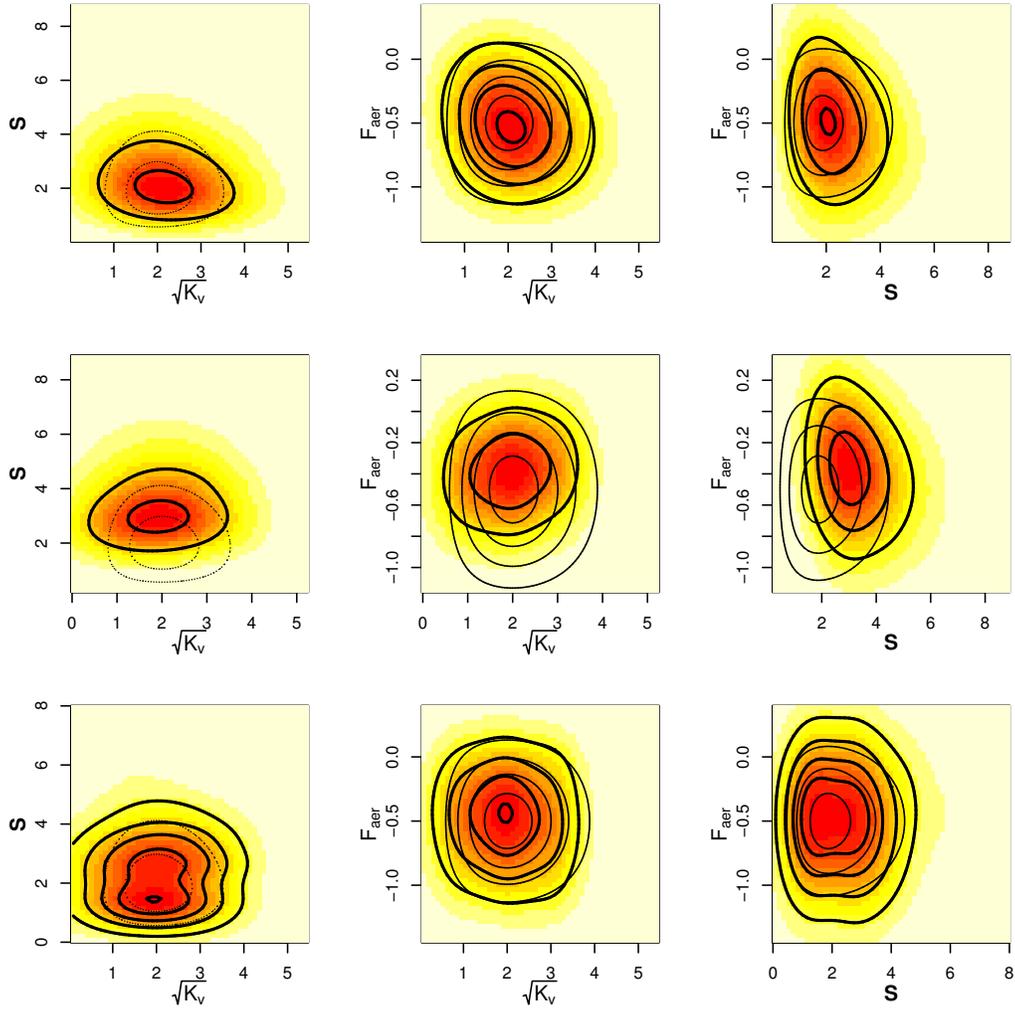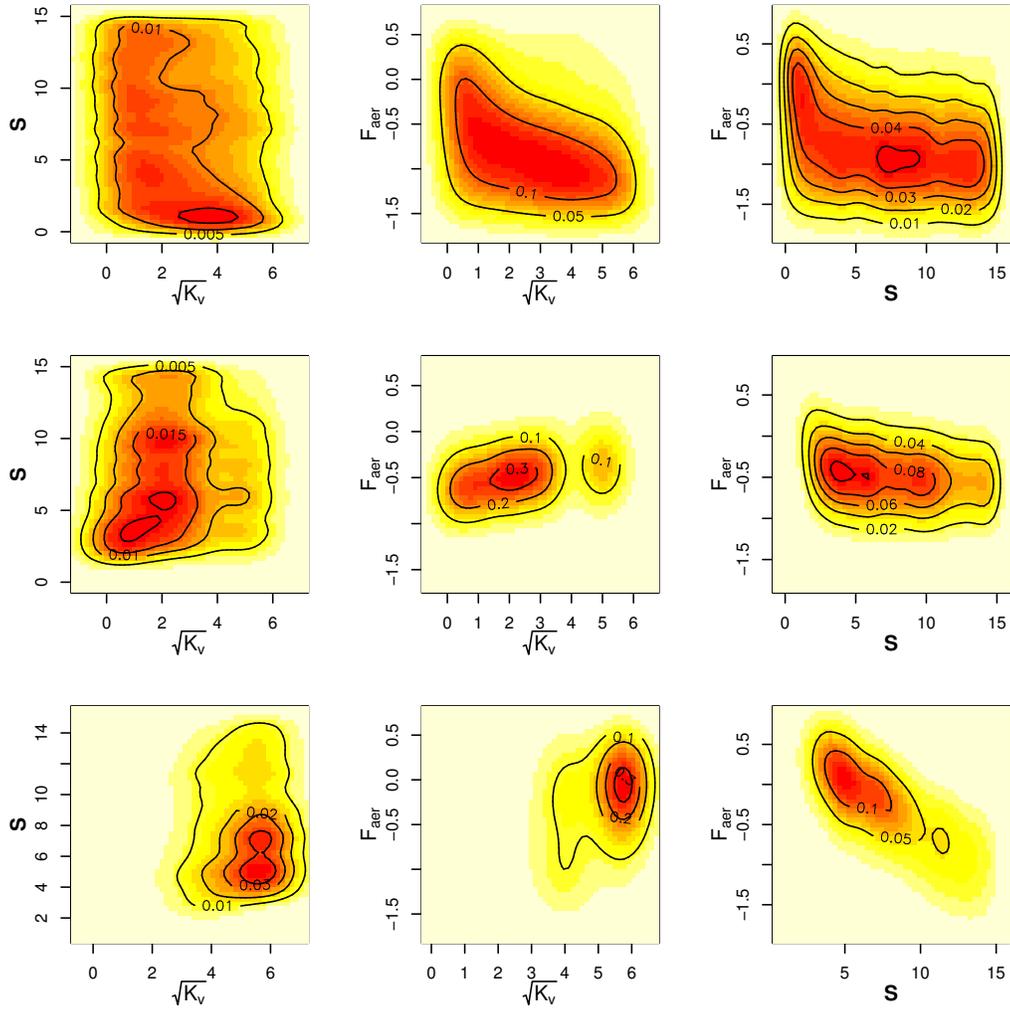
Figure 5: Marginal bivariate posterior distributions of the climate parameters using each of the three diagnostics with an informative prior. The priors and posteriors are shown by the thin and thick lines, respectively. First row: Deep ocean; Second row: Surface temperature; Third row: Upper air.

Figure 6: Marginal bivariate posterior distributions of the climate parameters using each of the three diagnostics with a uniform prior. First row: Deep ocean; Second row: Surface temperature; Third row: Upper air.

is not significantly different from zero. Increasing $\sqrt{\mathcal{K}_v}$ has a decreasing effect on ST, while for $\mathcal{S}$ and $\mathcal{F}_{aer}$ the effect is positive. Latitude does not seem to have a significant effect on ST. In contrast, the decades have a significant positive effect.

We performed a predictive assessment similar to the one presented for DO. In Figure 7 we present the predictive intervals for six combinations of the climate parameters of the 43 left out of the inference. We observe that most of the MIT2CM output are within the predictive intervals. These have a typical width of .2 degrees. Also the latitudinal and decadal patterns of the model simulations are well captured by the predictions.

Some quantiles of posterior distributions of the climate parameters are reported in the fourth, fifth and sixth columns of Table 4. When compared to the prior quantiles in Table 3 we see important differences both for the informative and the uniform prior. The marginal posterior densities in the first columns of Figure 4 show that the posterior for $\sqrt{\mathcal{K}_v}$ is slightly shifted to the right with respect to the prior. The posterior for $\mathcal{S}$ is shifted to the left, while the posterior for $\mathcal{F}_{aer}$ is much more concentrated around the median than the prior. This is in sharp contrast to the behavior of the posterior based on DO, which has much larger tails than the prior. In the second column of Figure 4 we present the posterior densities based on a uniform prior. As for the DO diagnostic we observe that the distributions tend to be diffuse and strongly influenced by the location of the design points. The posterior density of $\mathcal{F}_{aer}$ is a notorious exception. In fact this looks similar to the posterior obtained using the informative prior. The values of the posterior quantiles indicate that the uniform prior based posterior is shifted to the left with respect to the informative prior based posterior. These results indicate that the ST diagnostic contains useful information regarding $\mathcal{F}_{aer}$.

From the plots of the bivariate densities in Figures 5 and 6, we observe some differences between the densities from DO and those from ST. In the informative case we observe that the joint of $\sqrt{\mathcal{K}_v}$ and $\mathcal{S}$ for DO has a lower center of mass than that for ST. The joint of $\mathcal{S}$ and $\mathcal{F}_{aer}$ for ST is tilted downward and is less concentrated when compared to the one for DO. Far more dramatic differences are observed in the uniform prior case. Here it is clear that the densities obtained using ST have a range of $\mathcal{F}_{aer}$ far more concentrated around the median and a range of $\mathcal{S}$ higher than those of the densities obtained using DO. The values of the Bhattacharyya distance between the priors and the posteriors for the ST diagnostic, reported in Table 5, are smaller than those computed for the ST diagnostic. This confirms the analysis of the quantiles and the graphics of the univariate and bivariate posterior densities.
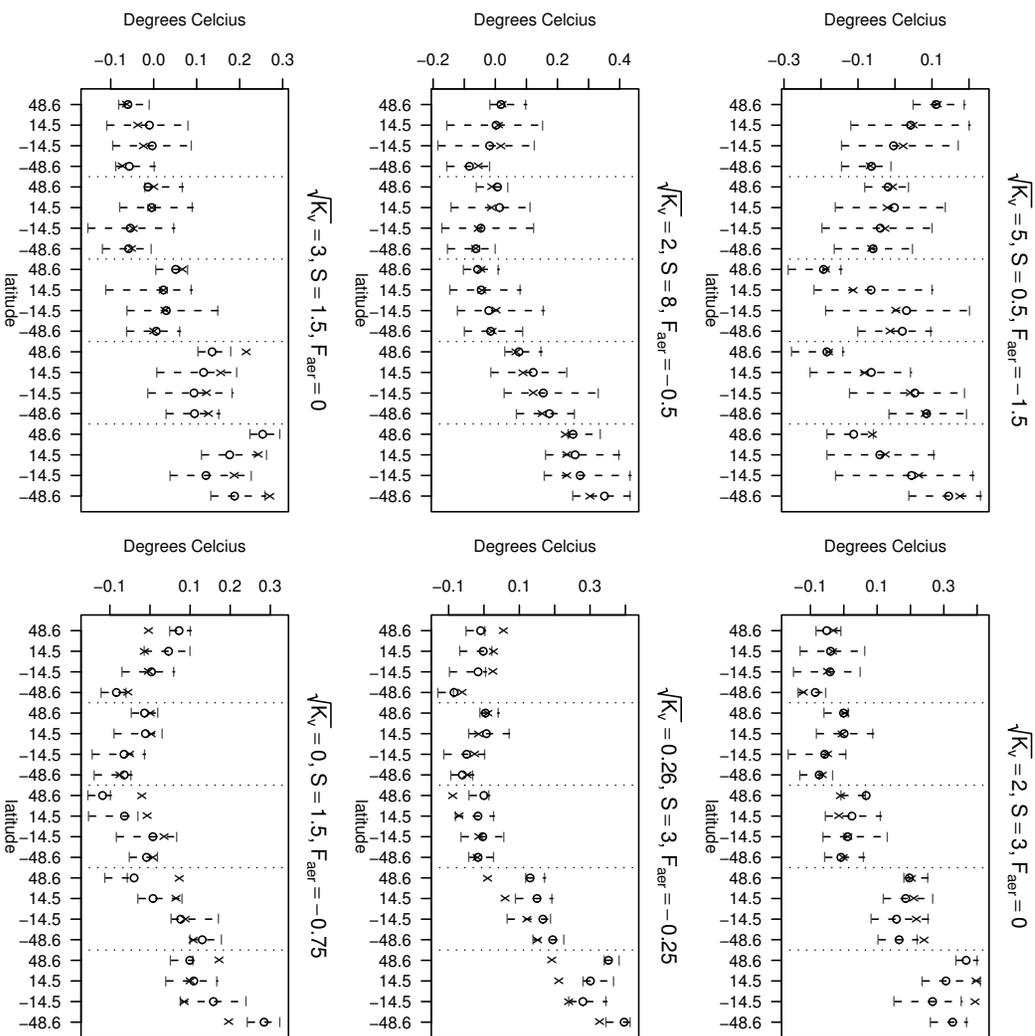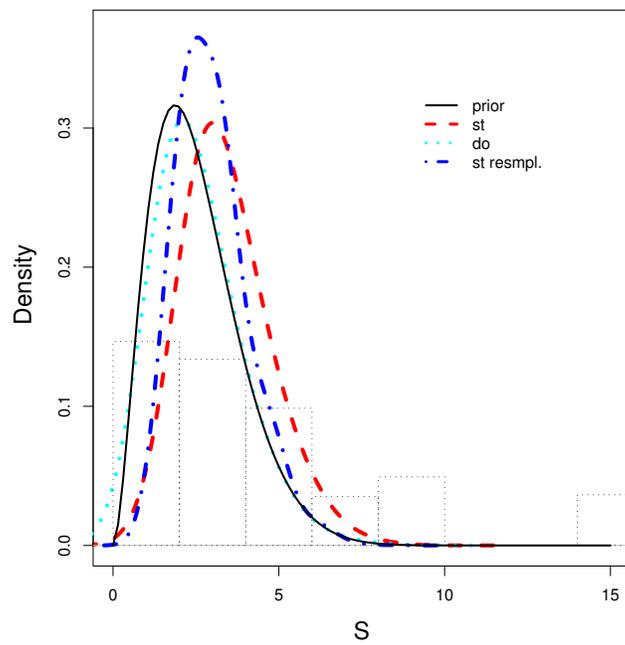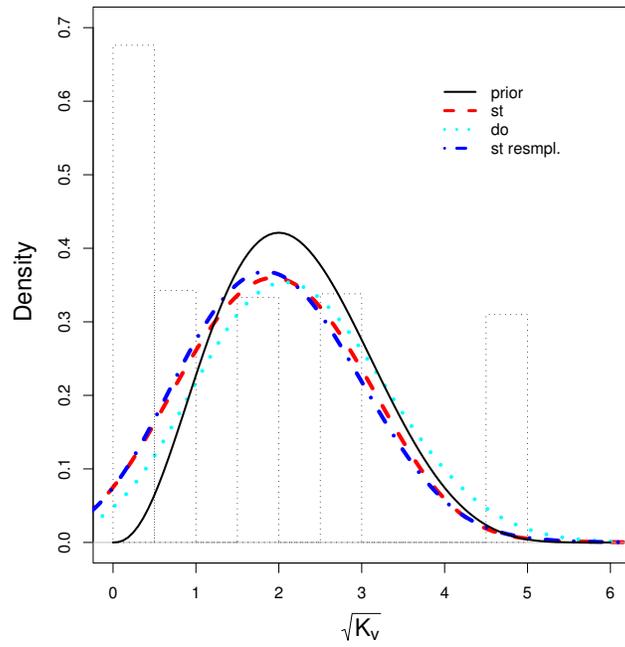
Inferring Climate Properties

Figure 7: Predictive distribution for ST at six combinations of the climate system properties. The crosses correspond to the MIT2DCM simulations. The extremes of intervals are given by the 2.5% and 97.5% posterior quantiles. Open circles correspond to predictive means.
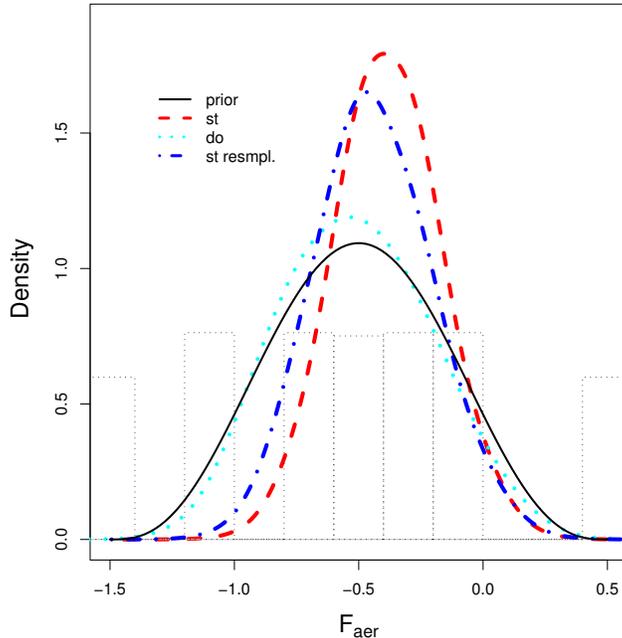
$\sqrt{K_v} = 5$, $S = 0.5$, $F_{aer} = -1.5$

$\sqrt{K_v} = 2$, $S = 8$, $F_{aer} = -0.5$

$\sqrt{K_v} = 3$, $S = 1.5$, $F_{aer} = 0$

$\sqrt{K_v} = 2$, $S = 3$, $F_{aer} = 0$

$\sqrt{K_v} = 0.26$, $S = 3$, $F_{aer} = -0.25$

$\sqrt{K_v} = 0$, $S = 1.5$, $F_{aer} = -0.75$

Degrees Celcius

latitude

Figure 8: Marginal posterior distributions of the climate parameters using the deep ocean and the surface temperature diagnostics jointly and comparison with the results using the diagnostics separately.

We considered two more posterior distributions obtained using ST. One is the result of resampling the posterior distribution from DO, following the method described in Section 3.3. The resulting univariate posterior densities are presented in Figure 8. For $\sqrt{\mathcal{K}_v}$ and $\mathcal{S}$ the posterior follows very closely the one obtained using only ST. The posterior density for $\mathcal{F}_{aer}$ is somewhat influenced by the DO results. For a numerical comparison we report some of the quantiles of these densities in Table 6. Additionally we notice that the Bhattacharyya distance between the joint posterior distribution and the informative prior is almost the same as the distance between the ST posterior and the informative prior. Furthermore, the distances between the those two posteriors is very close to one. The above discussion provides evidence that the posterior distribution is dominated by the ST diagnostic.

Finally, we consider the issue of using a different prior for Σ based on a different GCM. We use $k = 91$ replicates of $\boldsymbol{w}_j$ based on GFDL with the informative prior for $\boldsymbol{\theta}$. The plots of the univariate posterior distributions of $\sqrt{\mathcal{K}_v}$, $\mathcal{S}$ and $\mathcal{F}_{aer}$ show minimal differences to the posteriors obtained using the control runs from HadCM2. On the other hand we expect the posterior of Σ to be affected by the use of different control runs. In Figure 10 we present the posterior densities of the log-eigenvalues of Σ as well as boxplots for the posterior samples of the eigenvector corresponding to the largest eigenvalues. We

	$\sqrt{\mathcal{K}_v}$	$\mathcal{S}$	$\mathcal{F}_{aer}$
5%	0.84	1.45	-0.79
50%	1.97	2.86	-0.46
95%	3.15	5.04	-0.05

Table 6: Posterior quantiles of the climate system properties using DO and ST.

observe that in both cases the first eigenvalue has large variability, slightly larger in the HDCM case. For the eigenvector we observe a more regular pattern in the GFDL case than in the HDCM one. In particular it is noticeable that, for the GFDL case, the coefficients are increasing as a function of latitude for the first two decades, and decreasing for the last two ones. Although the overall patterns show warming for each latitude, the relative warming rates differ for the leading eigenmodes.

4.3   Upper air temperature change

As mentioned in Section 2.2, we restrict our analysis to the first four layers of the atmosphere. In Figure 1 we observe that, for the first four layers, the MIT2DCM simulations can have significant departures from the observations. In contrast to the ST, for which most simulated values differ one or two tenths of a degree Celsius from the observations, the UA simulations can be more than one degree away from the observations, especially at latitudes away from the poles. This implies that UA is likely to provide only weak information about the most likely values of $\boldsymbol{\theta}$. As in the previous two cases we considered $\boldsymbol{h}$ as a linear function of θ_1, θ_2 and θ_3. For UA the position variable $\boldsymbol{x}$ is two dimensional and corresponds to latitude, x_1, and altitude, x_2. So we fitted our model with $\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{t})$ equal to $(1, t_1, t_2, t_3, \cos(2x_2\pi/180), \log_{10} x_2)$. The posterior distribution of $\boldsymbol{\beta}$ shows that the baseline for the upper air temperature change is negative. As for the ST, increasing $\sqrt{\mathcal{K}_v}$ has a decreasing effect on ST, while for $\mathcal{S}$ and $\mathcal{F}_{aer}$ the effect is positive. The coefficient for the term $\cos(2x_2\pi/180)$ is significantly positive, implying a positive effect at the equator and negative effects at the poles. The coefficient for altitude is also positive, implying an increasing temperature change with increasing altitude.

To assess the predictive capability of the model we did an analysis similar to the one presented for DO and ST. In this case, we observed that the predictive intervals calculated from samples of the predictive distributions were extremely wide. In Figure 11 we present the predictive means for six combinations of the climate parameters. We observe that the predictive means track the simulations very well. We also observe that there are large differences between the simulations and the observations. Clearly these differences have no effect on the predictive mean but they produce very large predictive variances.

The last three columns of Table 4 report some of the quantiles of the posterior distributions for UA obtained with the informative and the uniform prior. For the informative prior, the quantiles as well as the plots of the univariate posterior density
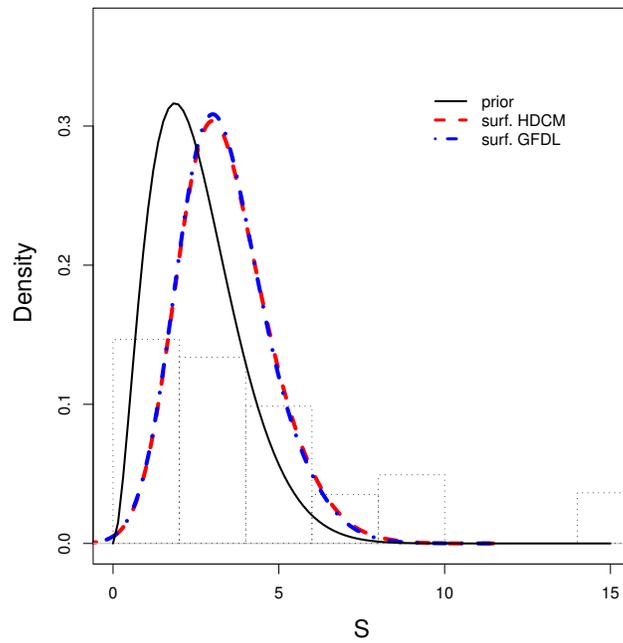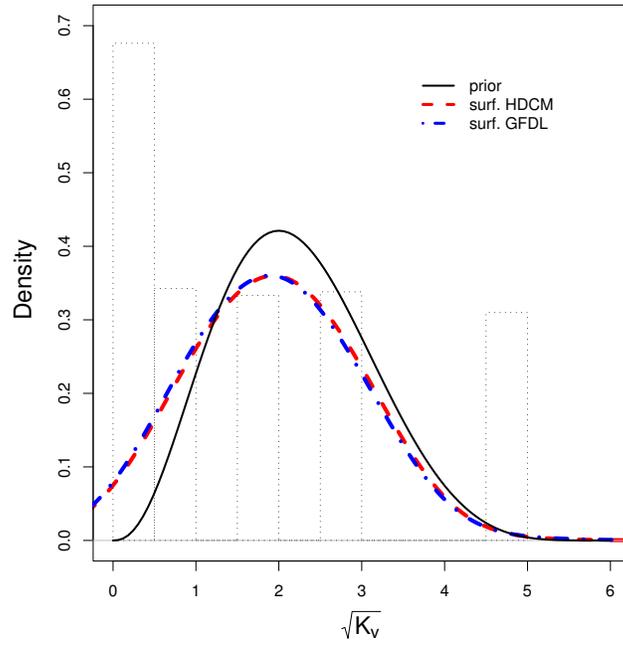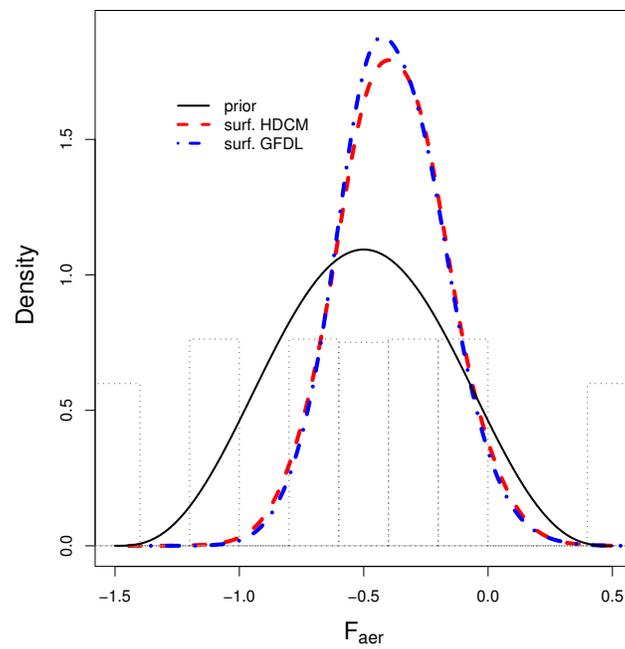
Figure 9: Marginal posterior distributions of the climate parameters using the deep ocean diagnostic with two different prior specifications for the covariance matrices based on different GCM model outputs.
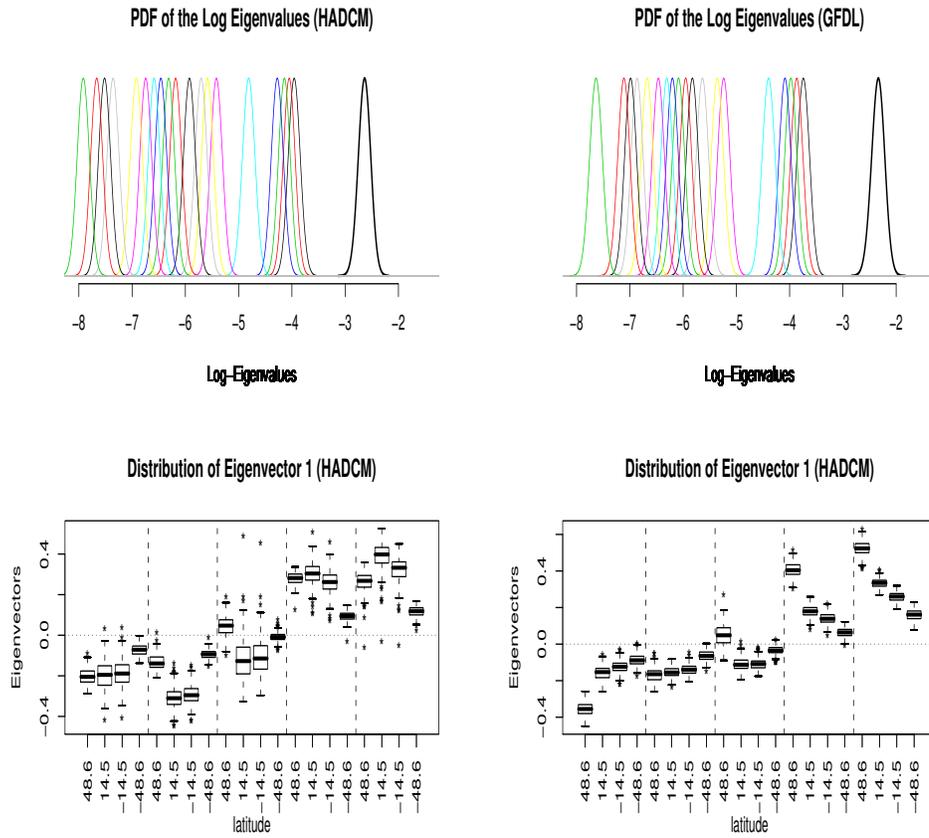
PDF of the Log Eigenvalues (HADCM)   **PDF of the Log Eigenvalues (GFDL)**

Log–Eigenvalues   Log–Eigenvalues

Distribution of Eigenvector 1 (HADCM)   **Distribution of Eigenvector 1 (HADCM)**

latitude   latitude

Figure 10: Posterior density of the log-eigenvalues for the Σ under the Surface Temperature diagnostic using two different control runs to obtain the prior (top two panels). The densities have been rescaled so that all have the same height. Distributions for the components of the first eigenvectors (bottom panels).
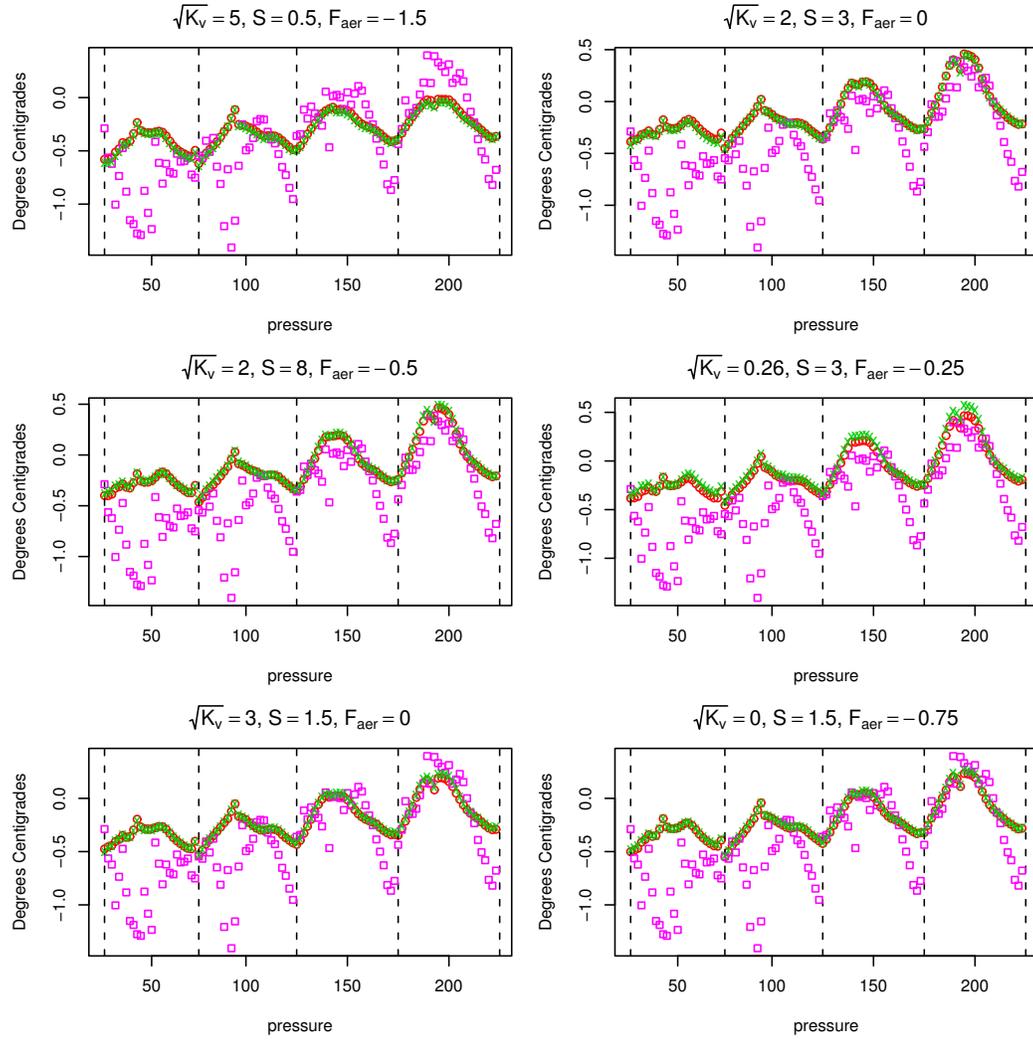
Figure 11: Predictive means for UA at six combinations of the climate system properties. ∘ correspond to the MIT2DCM simulations. '+' correspond to the predictive means. □ correspond to the observations.

in Figure 4 show a tendency to concentrate more values around large values of $\sqrt{\mathcal{K}_v}$ than the prior. A similar, but much less pronounced phenomenon is observed for $\mathcal{S}$ and $\mathcal{F}_{aer}$. In spite of this, the bivariate distributions in Figure 5 show that the joint density of $\mathcal{S}$ and $\sqrt{\mathcal{K}_v}$, obtained from UA, has the smallest model values for $\mathcal{S}$, when compared to the corresponding densities obtained from DO and ST.

The posterior distributions from UA using a uniform prior are shown in the second column of Figure 4 and Figure 6. These, together with the posterior quantiles indicate that UA produces a very strong shift of probability towards large values of the parameters, especially for $\sqrt{\mathcal{K}_v}$ and $\mathcal{F}_{aer}$. The values of the Bhattacharyya distances corresponding to the last two columns of Table 5 are the smallest of all the distances reported. This indicates that the posterior distributions obtained from UA have the least amount of overlap with either the priors or any of the posteriors produced with the other two diagnostics. Interestingly, there seems to be substantial overlap between the posterior based on UA with the informative prior and the posterior obtained from DO plus ST.

5   Discussion and Conclusions

This paper provides a successful implementation of a Bayesian model for the calibration of the parameters in the MIT2DCM. We have explored the posterior distribution of the climate system properties based on different summaries of climate data and climate simulation. Our model is able to blend information from different sources, including historical records, GCM output, MIT2DCM output and expert knowledge based on the scientific literature. All the variability from parameter estimation is accounted for in the results. So the present analysis provides an assessment of how posterior distributions differ from those in Forest et al. (2006) when such variability is considered. Results for the point estimates of the climate system properties are similar to the analysis in Forest et al. (2006), in particular for $\mathcal{S}$ and $\mathcal{F}_{aer}$. Nevertheless, we observe substantial differences in the shapes of the posterior distributions. Consistent with Forest et al. (2006), we find that ST is the most informative of the three diagnostics, which is where tightening of the distributions occurs most. Similarly, the least information is in the UA diagnostic where we see a widening of the distributions. The most robust feature in the results are the strong constraints on $\mathcal{F}_{aer}$ where the posterior is independent of the the priors. We observed very small differences in the posterior distributions obtained using different GCM runs.

Our graphical results as well as the numerical values of the Bhattachryya distances for the posterior distributions obtained with informative priors show a substantial overlap with the prior. To asses the significance of the differences between prior and posterior for $\mathcal{K}_v$ and $\mathcal{S}$ we consider the distributions of two quantities, related to these climate properties, used to characterize climate models predictions. These are the Transient Climate Response (TCR) and Sea Level Rise (SLR) due to thermal expansion. They are estimated by calculating the climate system response to a standardized forcing scenario ($F(t) = 1\%$/year increase in CO2 concentrations) and taking the change in tempera-

	Prior	ST	ST + DO
$Pr(\text{TCR} > 3\,C^\circ)$	0.67	0.94	0.88
$Pr(\text{SRL} > 10\,cm)$	0.40	0.63	0.51

Table 7: Probabilities of TCR and SLR being above specific thresholds using different pdfs for $\mathcal{S}$ and $\mathcal{K}_v$.

ture (TCR) and sea-level (SLR) at the time of CO2 doubling (year 70). We use samples from different distributions of $\mathcal{K}_v$ and $\mathcal{S}$ to estimate the distribution of TCR and SLR. The results for the informative prior, the ST and the ST+DO based posteriors are shown in Figure 12. As an illustration of how the survival functions change when the prior of $\boldsymbol{\theta}$ is updated wih either ST or ST + DO we calculated $Pr(\text{TCR} > 3\,C^\circ)$ and $Pr(\text{SRL} > 10\,cm)$. Table 7 shows a substantial increase in those probabilities when the information in ST and DO is included in the estimation of $\mathcal{S}$ and $\mathcal{K}_v$.

Regarding the UA diagnostic, we found that it is the most complex and difficult one to use. In fact we performed three separate analyses involving upper air temperature change. Our first attempt consisted of using the full 220 dimensional diagnostic, corresponding to the eight pressure layers. We then analyzed the first four and the last four layers separately. Of all three cases considered, the one that corresponds to the 50 to 200 hPa layers, is the one providing the most information. As mentioned in Section 4.3, the predictive variance in this case is very large. For the full UA diagnostic it is even larger. For the UA diagnostic based on the 300 to 850 hPa the mixing of the MCMC was very poor, with most of the samples corresponding to design points and a very low acceptance ratio.

For the UA, we have already commented on the discrepancies between the MIT2DCM simulations and the historical values, which tend to produce large predictive variances. As a consequence the statistical emulator performs poorly for points not in the original design grid. An additional problem of the UA is that it carries little information about the climate system properties in most of the layers. We have excluded any observational errors in our model. We justify this on the grounds that data sources are very heterogeneous and it is impractical to keep track of all possible sources of observational error. On the other hand the observations we use are the results of averaging large numbers of data, so the observational variability should be very small. In the atmosphere observations are both scarce and noisy, especially in comparison to the data used for ST. (n.b., The ocean data are similarly scarce and noisy although we have used a trend in the global-mean quantity to reduce these concerns.) So including an additional source of error for the analysis of the UA is an extension to be considered. To further complicate this picture, we know that the UA data have a known cold bias in the lower stratospheric region (50-200 hPa) in the Parker et al. (1997) data set (see Mears et al. (2006)). Taking this into account by using more recent data sets such as HadAT2 (Thorne et al. 2005) may eliminate this bias but perhaps not simplify the issues.

The current form of the emulator is purely statistical, and no physical considerations have been incorporated into it. One approach to incorporating a more realistic
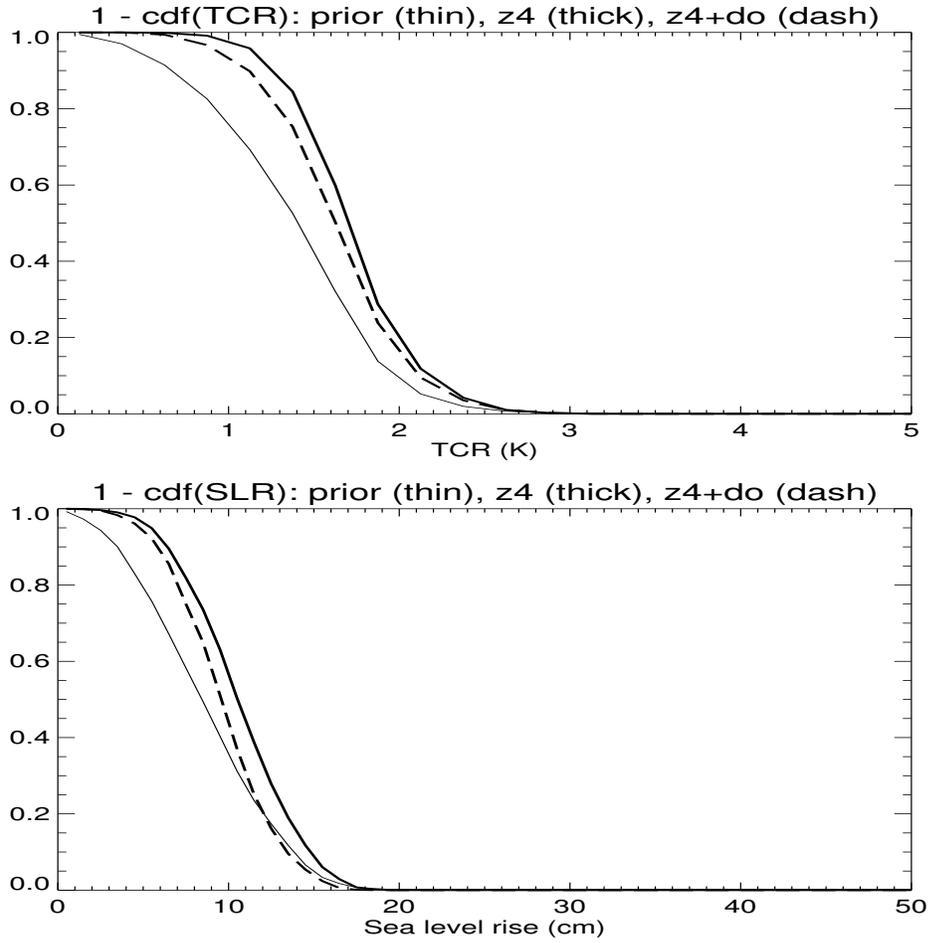
Figure 12: Survival functions for TCR (top) and SLR (bottom) using 5,000 from the prior distribution of $\mathcal{K}_v$ and $\mathcal{S}$ (thin line), the posterior distribution obtained using ST (thick) and the posterior distribution using ST + DO (dashed).

structure would be to use a simple energy balance model (EBM). For the surface and deep ocean temperatures, an EBM can be constructed with several latitude zones and an ocean model similar to the Q-flux mixed layer model used in the MIT 2D climate model. We can consider the energy balance equation for a given latitude zone as: $c_p \frac{dT}{dt} = F - \lambda T - \Phi_v - \Phi_h$ where c_p is the heat capacity per unit mass, $\frac{dT}{dt}$ is the time rate of change of the temperature anomaly, F is the local radiative forcing, related to $\mathcal{F}_{aer}$, λ is related to $\mathcal{S}$, and $\Phi_{v,h}$ are the vertical and horizontal heat fluxes out of the latitude zone, related to $\mathcal{K}_v$. The details of such a model are beyond the scope of this paper although Schneider (1992) provides a good introduction to the design of such models. For the upper-air temperature diagnostics, as already discussed, we need to consider simplifying the diagnostic to capture the large-scale behavior. Typically, an EBM does not include the vertical temperature structure in the atmosphere and so the relation between surface and upper-air temperature changes needs additional consideration. Fixed vertical temperature dependence is not desired in a standard EBM because the interactions between clouds, water vapor, temperature and winds lead to different feedbacks. It may be possible to address these vertical temperature dependence issues by allowing coefficients in the EBM to vary accordingly. MIT 2D climate model can provide sufficient quantities of data to explore alternative EBM structures such as these.

In this paper we have tackled a calibration problem with a multidimensional response variable. This involves the estimation of Σ, which is an important by-product of the analysis, since it provides information about the second order properties of the climate variability. For example, we have estimates of the eigen-decomposition which show a clear distinction between eigenspectrum from GFDL and HadCM2 models. Looking at the first two eigenvectors in each case we find that they differ in their patterns of latitude dependence, but not in those of decadal dependence. The task of estimating Σ is complicated when the diagnostic is of large dimensionality and there is weak prior information available, like in the case of UA. An alternative would be to consider a parametric structure for Σ, based on a valid space and time covariance function. In addition to providing less flexibility, this approach could run into the known problem that covariance parameters can be difficult to estimate. A different dimension reduction approach is that of truncating the number of eigenvalues of Σ. In fact, the posterior distribution of Σ indicates that only the first few eigenvalues are significantly large, so eliminating those modes of variability with very small amplitudes in the data is likely to be effective. We are currently working in a fully Bayesian implementation of this approach.

References

Andronova, N. G. and Schlesinger, M. E. (2001). "Objective Estimation of the Probability Density Function for Climate Sensitivity." *J. Geophys. Res.*, 106(D19): 22,605– 22,612. 8, 9

Berger, J., De Oliveira, V., and Sansó, B. (2001). "Objective Bayesian Analysis of Spatially Correlated Data." *Journal of the American Statistical Association*, 96: 1361–1374. 13

Claussen, M., Mysak, L. A., Weaver, A. J., Crucifix, M., Fichefet, T., Loutre, M., Weber, S. L., Alcamo, J., Alexeev, V. A., Berger, A., Calov, R., Ganapolski, A., Goosse, H., Lohman, G., Lunkeit, F., Mokhov, I. 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." *Clim. Dyn.*, 18(7): 579–586. DOI:10.1007/s00382-001-0200-1. 2

Curry, C., Sansó, B., and Forest, C. (2005). "Inference for climate system properties." Technical Report ams2005-13, Applied Mathematics and Statistics, University of California Santa Cruz. 10

et al., J. H. (2002). "Climate Forcings in GISS SI2000 Simulations." *J. Geophys. Res.*, 107: DOI:10.1029/2001JD001143. 8

Forest, C. E., Allen, M. R., Sokolov, A. P., and Stone, P. H. (2001). "Constraining climate model properties using optimal fingerprint detection methods." *Climate Dynamics*, 18: 277–295. 10

Forest, C. E., Stone, P. H., and Sokolov, A. P. (2006). "Estimated PDFs of climate system properties including natural and anthropogenic forcings." *Geophys. Res. Let.*, 33(L01705): DOI:10.1029/2005GL023977. 4, 6, 7, 8, 9, 10, 12, 13, 30

Forest, C. E., Stone, P. H., Sokolov, A. P., and Allen, M. R. (2002). "Quantifying uncertainties in climate system properties with the use of recent climate observations." *Science*, 295: 113–117. 9, 10, 12

Forster, P., Ramaswamy, V., Artaxo, P., Berntsen, T., Betts, R., Fahey, D., Haywood, J., Lean, J., Lowe, D., Myhre, G., Nganga, J., Prinn, R., Raga, G., Schulz, M., and Dorland, R. V. (2007). "Changes in Atmospheric Constituents and in Radiative Forcing." In Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., M.Tignor, and Miller, H. (eds.), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, 129–234. Cambridge University Press, Cambridge, UK and New York, NY, USA. 3

Forster, P. M. F. and Gregory, J. M. (2006). "The Climate Sensitivity and Its Components Diagnosed from Earth Radiation Budget Data." *J. Climate*, 19: 39–52. 8, 9

Frame, D. J., Booth, B. B. B., Kettleborough, J. A., Stainforth, D. A., Gregory, J. M., Collins, M., , and Allen, M. R. (2005). "Constraining climate forecasts: The role of prior assumptions." *Geophys. Res. Let.*, 32(L09702): DOI:10.1029/2004GL022241. 8, 9

Gregory, J., Stouffer, R., Raper, S., Stott, P., and Rayner, N. (2002). "An Observationally Based Estimate of the Climate Sensitivity." *J. Climate*, 15(22): 3117–3121. 8, 9

Hansen, J., Lacis, A., Rind, D., Russell, G., Stone, P., Fung, I., Ruedy, R., and Lerner, J. (1984). "Climate Sensitivity: Analysis of Feedback Mechanisms." In Hansen, J. E. and Takahashi, T. (eds.), *Climate Processes and Climate Sensitivity, Geophysical Monograph*, volume 29, 130–163. American Geophysical Union, Washington, D.C. 6

Hansen, J., Russell, G., Rind, D., Stone, P., Lacis, A., Lebedeff, S., Ruedy, R., and Travis, L. (1983). "Efficient Three-Dimensional Global Models for Climate Studies: Models I and II." *Mon. Weath. Rev.*, 111: 609–662. 6

Hegerl, G., Zwiers, F. W., Braconnot, P., Gillett, N., Luo, Y., Orsini, J. M., Nicholls, N., Penner, J., and Stott, P. (2007). "Understanding and Attributing Climate Change." In Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., M.Tignor, and Miller, H. (eds.), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, 663–746. Cambridge University Press, Cambridge, UK and New York, NY, USA. 8

International *ad hoc* Detection and Attribution Group, C. (2005). "Detecting and Attributing External Influences on the Climate System: A Review of Recent Advances." *J. Climate*, 18(9): 1291–1314. 10

IPCC (2007). *Climate Change 2007 – The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the IPCC*. Solomon, S. et al. (eds.), Cambridge University Press. 1

Jones, P., New, M., Parker, D., Martin, S., and Rigor, I. (1999). "Surface air temperature and its changes over the past 150 years." *Reviews of Geophysics*, 37: 173–199. 7

Kennedy, M. C. and O'Hagan, A. (2001). "Bayesian Calibration of Computer Models." *Journal of the Royal Statistical Society, Series B*, 63: 425–464. 4

Knutti, R., Stocker, T. F., Joos, F., and Plattner, G. (2002). "Constraints on radiative forcing and future climate change from observations and climate model ensembles." *Nature*, 416: 719–723. 9

— (2003). "Probabilistic climate change projections using neural networks." *Clim. Dyn.*, 21: 257–272. 8, 9

Lean, J. (2000). "Evolution of the Sun's Spectral Irradiance Since the Maunder Minimum." *Geophys. Res. Lett.*, 27: 2421–2424. 8

Levitus, S., Antonov, J., and Boyer, T. P. (2005). "Warming of the World Ocean, 1955–2003." *Geophys. Res. Let.*, 32(L02604): DOI:10.1029/2004GL021592. 7

Mears, C., Forest, C., Spencer, R., Vose, R., and Reynolds, R. (2006). "What is our understanding of the contribution made by observational or methodological uncertainties to the previously reported vertical differences in temperature trends?" In Karl, T. R., Hassol, S. J., Miller, C. D., and Murray, W. L. (eds.), *Temperature*

Trends in the Lower Atmosphere: Steps for Understanding and Reconciling Differences. A Report by the Climate Change Science Program and the Subcommittee on Global Change Research, Washington, DC.   31

Meehl, G., Stocker, T., Collins, W., Friedlingstein, P., Gaye, A., Gregory, J., Kitoh, A., Knutti, R., Murphy, J., Noda, A., Raper, S., Watterson, I., Weaver, A., and Zhao, Z.-C. (2007). "Global Climate Projections." In Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., M.Tignor, and Miller, H. (eds.), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, 747–846. Cambridge University Press, Cambridge, UK and New York, NY, USA.   2, 8

Mitchell, J. F. B., Karoly, D. J., Hegerl, G. C., Zwiers, F. W., Allen, M. R., and Marengo, J. (2001). "Detection of Climate Change and Attribution of Causes." In Houghton, J. T., Ding, Y., Griggs, D. J., Noguer, M., van der Linden, P. J., Dai, X., Maskell, K., and Johnson, C. A. (eds.), *Climate Change 2001: The Scientific Basis*, 695–738. Cambridge University Press, Cambridge, UK and New York, NY, USA.   10

Morgan, M. G. and Keith, D. W. (1995). "Subjective judgements by climate experts." *Environ. Sci. Technol.*, 29: 468A–476A.   13

Parker, D. E., Gordon, M., Cullum, D. P. N., Sexton, D. M. H., Folland, C. K., and Rayner, N. (1997). "A new global gridded radiosonde temperature data base and recent temperature trends." *Geophysical Research Letters*, 24: 1499–1502.   7, 31

Paulo, R. (2005). "Default Priors for Gaussian Processes." *The Annals of Statistics*, 33(2): 556–582.   12, 13

Ramankutty, N. and Foley, J. A. (1999). "Estimating historical changes in global land cover: croplands from 1700 to 1992." *Global Biogeochemical Cycles*, 13(4): 997–1027.   8

Sansó, B., Forest, C., and Zantedeschi, D. (2007). "Statistical Calibration of Climate System Properties." Technical Report ams2007-06, Applied Mathematics and Statistics, University of California Santa Cruz.   6, 10, 14

Sato, M., Hansen, J. E., McCormick, M. P., and Pollack, J. B. (1993). "Stratospheric aerosol optical depths." *J. Geophys. Res.*, 98: 22987–22994.   8

Schneider, S. H. (1992). "Introduction to climate modeling." In Trenberth, K. E. (ed.), *Climate System Modeling*, 3–26. Cambridge University Press, New York, NY.   33

Smith, S. J., Andres, R., Conception, E., and Lurz, J. (2003). "Historical Sulfur Dioxide Emissions 1850-2000." Technical Report ftp://jgcri.umd.edu/ssmith/Hist_SO2_Emissions/, Pacific Northwest National Laboratory, Joint Global Change Research Institute, 8400 Baltimore Avenue, College Park, Maryland 20740.   8

Sokolov, A. P., Forest, C. E., and Stone, P. H. (2003). "Comparing Oceanic Heat Uptake in AOGCM Transient Climate Change Experiments." *J. Climate*, 16: 1573–1582.   13

Sokolov, A. P. and Stone, P. H. (1998a). "A flexible climate model for use in integrated assessments." *Clim. Dyn.*, 14: 291–303. 4, 6

— (1998b). "A flexible climate model for use in integrated assessments." *Climate Dynamics*, 14: 291–303. 6

Stone, P. H. and Yao, M.-S. (1987). "Development of a two-dimensional zonally averaged statistical-dynamical model. Part II: the role of eddy momentum fluxes in the general circulation and their parametrization." *J. Atmos. Sci.*, 44(24): 3769–3786. 6

— (1990). "Development of a two-dimensional zonally averaged statistical-dynamical model. Part III: the parametrization of the eddy fluxes of heat and moisture." *J. Clim.*, 3(7): 726–740. 6

Tebaldi, C. and Sansó, B. (2008). "Joint Projections of Temperature and Precipitation Change from Multiple Climate Models: A Hierarchical Bayes Approach." *Journal of the Royal Statistical Society: Series A*. To appear. 4

Thorne, P. W., Parker, D. E., Tett, S. F. B., Jones, P. D., McCarthy, M., Coleman, H., and Brohan, P. (2005). "Revisiting radiosonde upper air temperatures from 1958 to 2002." *J. Geophys. Res.*, 110(D18105): DOI:10.1029/2004JD005753. 31

Webster, M. D. and Sokolov, A. P. (2000). "A Methodology for Quantifying Uncertainty in Climate Projections." *Climatic Change*, 46(4): 417–446. 13

Zhou, W. and Sansó, B. (2008). "Statistical Inference for Atmospheric Transport Models Using Process Convolutions." *Environmetrics*, 19: 87 – 101, DOI 10.1002/env.858. 13

Acknowledgments