Statistical Calibration of Climate System Properties

Bruno Sansó, Chris Forest and Daniel Zantedeschi *

Abstract

The behavior of modern climate system simulators is controlled by numerous parameters. By matching model outputs with observed data we can perform inference on such parameters. This is a calibration problem that usually requires the ability to evaluate the computer code at any given configuration of the parameters. As the climate system simulator attempts to describe very complex physical phenomena, the task of running the model is very computationally demanding. Thus, a statistical model is required to approximate the model output. In this work, we use output from the MIT 2D climate model (MIT2DCM), historical records and output from a three-dimensional climate model, to obtain estimates of the effective thermal diffusivity in the deep-ocean, the climate sensitivity and the net aerosol forcing that control the MIT2DCM behavior. We use a Bayesian approach that allows for the use of scientifically based information on the climate parameters to be used in the calibration process. The model tackles the problem of dealing with multivariate computer model output and incorporates all estimation uncertainties into the posterior distributions of the climate parameters. Additionally we obtain estimates of the correlation structure of the unforced variability of temperature change patterns. These results 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. These results include uncertainties in the estimation of the multivariate covariance matrices for the first time.

Key words: model calibration, climate change, climate sensitivity, Bayesian methods

1 Introduction

An honest assessment of uncertain climate system properties is key to the support of any scientific statement about the current state of Earth's climate, and to the construction of

*Bruno Sansó, Associate Professor, Department of Applied Mathematics and Statistics, University of California, 1156 High Street MS: SOE2, Santa Cruz CA-95064, U.S.A. bruno@ams.ucsc.edu, www.ams.ucsc.edu/∼bruno. Chris Forest, Research Scientist, Joint Program on the Science and Policy of Global Change, Rm E40-427 MIT, 77 Mass. Ave., Cambridge, MA 02139, U.S.A. ceforest@mit.edu, web.mit.edu/globalchange/www/, Daniel Zantedeschi, Graduate Student, Department of Applied Mathematics and Statistics, University of California, 1156 High Street MS: SOE2, Santa Cruz CA-95064, U.S.A. danielz@ams.ucsc.edu, www.ams.ucsc.edu/∼danielz.

forward-looking projections that may be used in policy decisions. The Intergovernmental Panel on Climate Change's (IPCC) Third Assessment Report (Watson and the Core Writing Team, 2001) summarized some of the key unanswered questions that complicate predictions of future global climate. Four of the identified uncertainties include: (i) Climate sensitivity, as measured by the change in surface air temperature at equilibrium under a specified external forcing (A climate "forcing" represents a perturbation to the energy budget of the climate system that moves it away from its current equilibrium. This forcing is typically a change in the radiative transfer component of the energy budget and is measured in Watts per square meter for the net change in the radiative flux at the top of the atmosphere. For further discussion see Watson and the Core Writing Team (2001)); (ii) Variations in the climate due to either natural effects or anthropogenic aerosols; (iii) Magnitude and character of natural climate variability; (iv) Spatiotemporal patterns of change in climate variables. Each of these contributes to the uncertainty in climate system properties in various ways as will be discussed later. We consider climate system properties as quantities that control the surface-air temperature change at equilibrium and the rate of heat uptake by the deep-ocean. These are quantities that cannot be derived from first principles in models but must be diagnosed from climate simulations of the transient response to forcings. This means determining the time-dependent response as the system approaches the new equilibrium under the specified forcings.

Computer models or simulators are used to predict the evolution of the state of the global climate system. These models discretize the earth's atmosphere, oceans, and land into grid boxes that have a typical size of 250 km $\times$ 250 km. They must account for localized phenomena, referred to as sub-gridscale phenomena, like the influence of clouds, which happen at scales orders of magnitude smaller than the grid box size. These factors are dealt with using so called *parameterizations*, which simulate the large-scale effects of the sub-gridscale processes. Most components of the climate system can be adjusted within 3D models by modifying individual parameters or parameterizations. In 2D models, like the one that is considered in this paper, climate system properties can be controlled via single parameters. In this work we consider three such parameters. A critical parameter with extensive uncertainty is climate sensitivity, defined as the equilibrium global mean surface temperature response to a doubling of CO_2 and denoted as $\mathcal{S}$. Within climate models, this parameter is adjusted by altering the cloud feedback, which changes the effect of clouds on the radiative transfer in the energy budget. Another important uncertain parameter is the rate of diffusion for heat anomalies into the deep-ocean, controlled by varying a diffusion coefficient, called, $\mathcal{K}_v$. A third parameter to be considered is the net anthropogenic aerosol forcing, written here as $\mathcal{F}_{aer}$. Although this represents uncertainty in all unmodeled influences on the climate system, the uncertainty is dominated by that in the net aerosol forcing and so we call it such.

The MIT2DCM provides simulations of ocean, surface, and upper atmospheric composition, heat, moisture and momentum on a latitude-height coordinate system. To account for the missing effects of longitudinal variations, the transports of heat, moisture and momentum include an eddy-diffusion parameterization (see Sokolov and Stone, 1998). This parameterization specifies the contribution of turbulent fluxes to the longitudinally-averaged equations of motion for the atmosphere (see Section 13.10 in Gill, 1982). Thus the model is run on

two dimensions, which refer to altitude and latitude bands. Despite the averaging away of longitude, the model is sufficiently complex to match longitudinally-averaged observations of the climate and to make similar predictions to those of full 3D atmosphere-ocean general circulation models (GCM) (Sokolov and Stone, 1998). More specifically, the output from the MIT2DCM, considered in this paper, consists of temperatures over a grid of zonal bands corresponding to 46 latitudes, averaging over all longitudes in the band. It has 11 vertical layers for a grid of 506 cells for every time step. Typical output corresponds to periods of 140 years with monthly-average data (although the model time step is 30 minutes.)

Forest et al. (2000, 2001, 2002, 2006) run the MIT2DCM for many choices of the uncertain parameters $\mathcal{S}$, $\mathcal{K}_v$ and $\mathcal{F}_{aer}$, selected systematically on a non-uniform grid. The grid considered in this paper consists of 426 points and is illustrated in Figure 1. To summarize the 426 time series in a way that is useful to understand possible global climate changes, three different multivariate statistics are used and called "diagnostics". We will use such terminology throughout the paper.

- *Upper-air temperature changes.* Temperature differences between the means of [1986-1995] and those of [1961-1980] for 36 latitudes and 8 levels, producing a 288 component vector.

- *Surface temperature change.* Temperature differences between the means of [1906-1995] and the 5 decadal means in the period [1946-1995], for 4 zonal bands, producing a 20 component vector.

- *Deep-ocean temperature trend.* This is calculated for the period [1948-1995] from time series of pentadal means, and produces a one component vector.

In correspondence to these three diagnostics for the climate model output, it is possible to obtain statistics based on historical observations. In this paper Levitus et al. (2005) is followed for the deep-ocean temperature trend, Jones et al. (1999) for surface temperature and Parker et al. (1997) for the upper-air temperature anomalies. The observations are compared to the climate model diagnostics in order to perform the calibration of the climate model parameters. Figure 2 illustrates the pattern of temperature change in the observational data and some of the simulations for the surface temperature diagnostic.

In this paper we focus on the surface temperature change as it is considered one of the most informative diagnostics. A similar analysis to the one presented here can be applied to the other two diagnostics. In this context, the estimation of the climate system properties consists of a computer model calibration problem. Following O'Hagan et al. (1999) and Kennedy and O'Hagan (2001), we adopt a Bayesian approach to learn about the values of the climate parameters. Generally, we let $F(\theta)$ be the output from a computer model for parameter value θ. Suppose data corresponding to $F(\theta)$ are available, then we build a statistical model so that information on the most likely values of θ is obtained via $p(\theta|Data)$.

The first difficulty that we face is that $F(\cdot)$ is evaluated at a limited number of points. Evaluating $F(\cdot)$ is very time consuming, so it is impractical to consider statistical procedures that require additional evaluations. For all the other values of θ, $F(\cdot)$ is an unknown function. Since it is unknown, we regard it as a random function and use a Gaussian process to create a statistical equivalent model. A second difficulty is that the surface temperature change

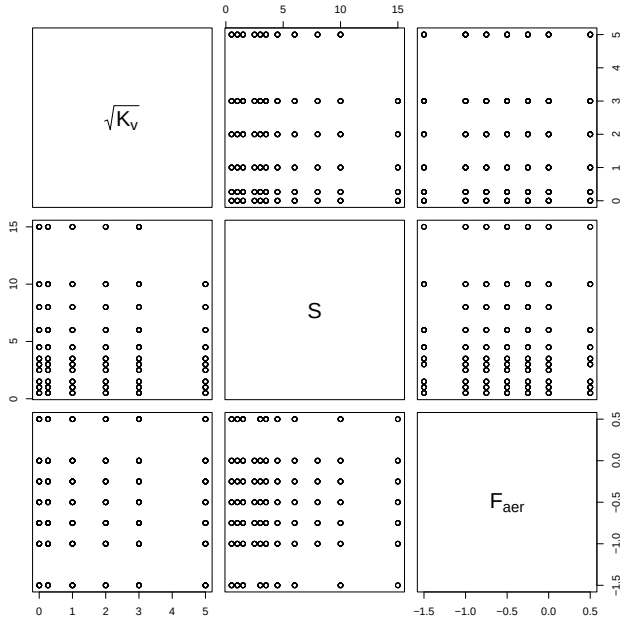

Figure 1: Configuration of the 426 different combinations of $\mathcal{S}$, $\mathcal{K}_v$ and $\mathcal{F}_{aer}$ used for simulations obtained with the MIT2DCM.

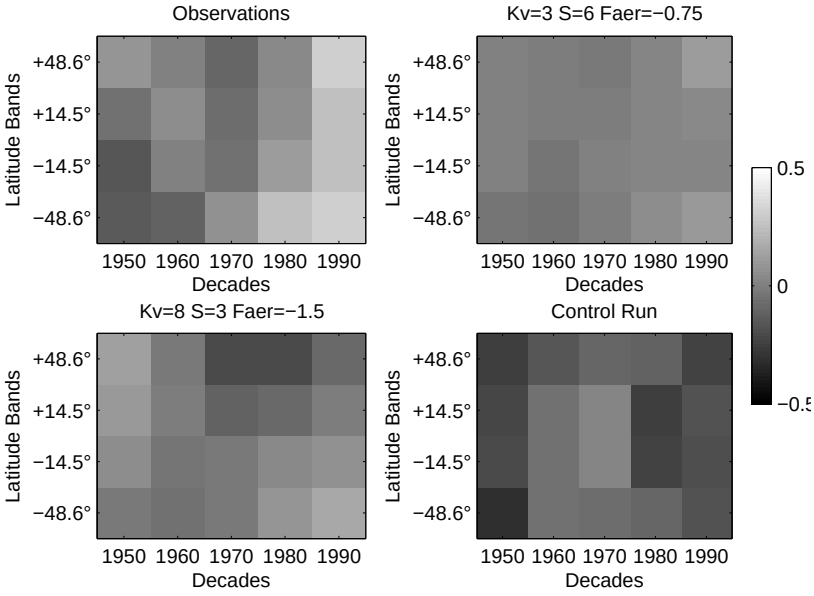

Figure 2: Comparison of the historically observed surface temperature change (top left panel) to MIT2DCM simulations obtained with two different parameter configurations (top right and bottom left panels) and a sample from the HadCM2 GCM (bottom right). All four figures are in the same scale in degrees Celsius.

is a multivariate diagnostic, so we have to account for the correlation between the different components. This is a problem that has an interest on its own. In fact, the estimation of the correlation structure of temperature change patterns is a fundamental problem for the detection and attribution of the causes of climate change (see, for example Hasselmann, 1979, 1993, 1997; Bell, 1982, 1986; Allen and Tett, 1999). Our approach is to assume a common correlation structure between the different sources of information. In particular we use control runs from models that the climatologists refer to as "unforced". These runs are obtained from 3D General Circulation Models (GCM) and represent the behavior of a climate model in a dynamic equilibrium or steady state. In summary, we have three sources of information: Historical records, MIT2DCM output and GCM model output.

In the next section we define a statistical model that provides the framework for the calibration of the climate model parameters and discuss the implementation details. In Section 3 we present the results obtained from the model and in the final section we present conclusions and discussion.

2 Statistical model

The available data, observed and simulated, correspond to space and time locations $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, $\boldsymbol{x}_i \in \mathbb{R}^2$. (i.e., latitude and decade, see Figure 2) We denote the historical records as $\boldsymbol{z} = (z_1, \ldots, z_n)$. MIT2DCM output at location $\boldsymbol{x}$ and parameter values $\boldsymbol{t} = (\mathcal{K}_v\ , \boldsymbol{\mathcal{S}}\ , \mathcal{F}_{aer}\ ) \in \mathbb{R}^3$ is denoted as $\eta(\boldsymbol{x}, \boldsymbol{t})$. The available model runs are denoted as $y_{ij} = \eta(\boldsymbol{x}_i, \boldsymbol{t}_j), i = 1, \ldots, n; j = 1, \ldots, p$, $\boldsymbol{y}_j = (y_{1j}, \ldots, y_{nj})'$ and $Y = [\boldsymbol{y}_1, \ldots \boldsymbol{y}_p] \in \mathbb{R}^{n \times p}$. The GCM model output is denoted as $\boldsymbol{w}_l = (w_{j1}, \ldots, w_{jn})', l = 1, \ldots, k$ and $W = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k] \in \mathbb{R}^{n \times k}$. For the surface temperature change $n = 20$, since there are five decades and four zonal bands. $p = 426$ is the number of different runs of the MIT2DCM. (NB: This is a reduction from the $p = 499$ simulations used in Forest et al. (2006) due to the lowering of the upper bound on $\mathcal{K}_v$ from 64 to 36 cm^2/s). $k = 162$ corresponds to the number of replications obtained from the GCM developed at the Hadley Centre, Bracknell, UK, code-named HadCM2. Table 1 shows a summary of the available information and the notation used.

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

Table 1: Summary of the available information for the Surface Temperature Diagnostics.

We denote as $\zeta(\boldsymbol{x})$ the true, unobserved, diagnostic at location $\boldsymbol{x}$. We note that the three diagnostics refer to linear functionals, or filters, of the temperature. For example, the difference between the average temperature for an area at a given time and the average for that area over a number of decades. This is referred to as an anomaly. We denote $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ as the 'true' value of the climate system properties $(\mathcal{K}_v\ , \boldsymbol{\mathcal{S}}\ , \mathcal{F}_{aer}\ )$. Then we

assume that

$$z_i = \zeta(\boldsymbol{x}_i) + \varepsilon_i + \xi_i = (\rho\eta(\boldsymbol{x}_i, \boldsymbol{\theta}) + \delta(\boldsymbol{x}_i)) + \varepsilon_i + \xi_i \quad . \tag{1}$$

This means that the observations deviate from the true value of the diagnostic by a measurement error ε_i. ξ_i represents the errors on spatial and temporal scales that are smaller than the grid cell size and that are introduced by the filter. In the atmospheric science literature this is known as representativeness error. Given the global scale of the data used in this application, we find it difficult to separate these two sources of error. So we will consider only one observational error, say, $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)' \sim N_n(0, \sigma^2\Sigma)$, where Σ is a covariance matrix and σ^2 is a scale parameter. Additionally, the code output is proportional to the true value, but there is a model inadequacy $\delta(\boldsymbol{x})$. We assume that $\boldsymbol{\delta} = (\delta(x_1), \ldots, \delta(x_n))' \sim N_n(0, \sigma^2\tau^2\Sigma)$. In a calibration problem like the one we consider here is very difficult to identify the multiplicative and additive biases ρ and δ. In the absence of strong prior information we fix $\rho = 1$ and $\tau^2 = 0$.

Since the value of $\eta(\boldsymbol{x}, \boldsymbol{\theta})$ is unknown. We assume that $\eta(\boldsymbol{x}, \cdot)$ 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. Gaussian processes provide the right balance between flexibility and tractability for this purpose. We assume that $E(\eta(\boldsymbol{x}, \boldsymbol{t})) = h(\boldsymbol{x}, \boldsymbol{t})'\boldsymbol{\beta}$, where h and $\boldsymbol{\beta}$ are q-dimensional vectors. They define a linear combination of effects due to locations and parameter values. β_0 is a baseline, β_1, β_2 and β_3 are the coefficients of each of the components of $\boldsymbol{t}$. The remaining elements of $\boldsymbol{\beta}$ correspond to three coefficients that measure the effect of the latitude, with respect to the band that includes the North Pole, and four coefficients that measure the effect of each decade with respect to the first one. This yields a total of $q = 11$ coefficients in $\boldsymbol{\beta}$. The covariance function is given by

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

Thus, we are assuming separability between $\boldsymbol{x}$ and $\boldsymbol{t}$.

Notice that the different sources of error share a common correlation structure through Σ. σ^2 and γ^2 are scale factors accounting for possible amplifications. Additionally we assume that $\boldsymbol{w}_j \sim N_n(0, \Sigma)$, giving the prior distribution for Σ,

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

where $S = 1/k \sum_{j=1}^{k} \boldsymbol{\omega}_j\boldsymbol{\omega}_j'$. Our assumptions are based on the idea that temperature has a correlation structure that is realized in the observations and it is also captured by the climate models. 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. So our model is using all three available types of data to extract information about the unforced variability.

To obtain the likelihood of the proposed model we need the joint density of $\boldsymbol{z}$ and Y. To proceed, we calculate $\operatorname{cov}(z_i, y_{jk}), i, j = 1, \ldots, n;\ k = 1, \ldots, p$, using equations (1) and (2). 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 $H(\cdot) = [h(\boldsymbol{x}_1, \cdot), \ldots, h(\boldsymbol{x}_n, \cdot)] \in \mathbb{R}^{q \times n}$, and let $H = [H(\boldsymbol{t}_1), \ldots, H(\boldsymbol{t}_p)]' \in \mathbb{R}^{np \times q}$, then

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

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

The space of climate parameters is not naturally geometrical, so we assume that the random process is separable, and

$$r(\boldsymbol{t}, \boldsymbol{t}') = r_1(t_1 - t_1'; \phi_1, \nu_1) r_2(t_2 - t_2'; \phi_2, \nu_2) r_3(t_3 - t_3'; \phi_3, \nu_3),$$

where

$$r_i(t_i, t_i') = \frac{1}{2^{\nu_i - 1}\Gamma(\nu_i)} \left(\frac{|t_i - t_i'|}{\phi_i} \right)^{\nu_i} K_{\nu_i} \left(\frac{|t_i - t_i'|}{\phi_i} \right).$$

This is the Matèrn correlation function (See, for example, Stein, 1999). 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. ν_i measures the correlation smoothness, which determines the smoothness of the random field. $\nu_i = 0.5$ corresponds to the popular exponential correlation function which produces processes that are not mean square differentiable. $\nu \to \infty$ corresponds to the Gaussian correlation, which produces infinitely smooth processes. 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 in r in the Matèrn class.

2.1 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 support on $(0, 6)$ and parameters $(3.5, 6)$. The prior for $\theta_2 = \mathcal{S}$ is specified as a beta distribution with support on $(0, 15)$ and parameters $(2.85, 14)$. The prior for $\theta_3 = \mathcal{F}_{aer}$ is a beta distribution on $(-1.5, .5)$ with parameters $(4, 4)$. 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 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 20^{th} 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 all terms in the covariance matrix depend only on the product $\sigma^2\gamma^2$, except the first 20×20 block. This implies near lack of identifiability of the two parameters. A possible solution is to use prior distributions that carry very precise information about the parameters. In this application we fix γ^2 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 the problem is not so severe when separability of the correlation function is assumed. For the smoothness parameters, ν_i, we used three independent normal priors centered around 0.5 with standard deviation 0.1.

2.2 Implementation

We fit the model using a Markov chain Monte Carlo (MCMC) method to explore the joint posterior distribution of all parameters (see, for example, Gamerman and Lopes, 2006).

Notice that it is important to avoid the explicit computation of the covariance matrix of the joint distribution of $(\boldsymbol{z}, \mathrm{vec}(Y))$. The dimension of such matrix is $n(1+p) \times n(1+p)$. For the surface temperature diagnostics, $n = 20$, so $n(1+p) = 8,540$, making such computations unfeasible within an iterative method. For the upper-air diagnostic the situation would be much worse. Lengthy computations and storage of large matrices is avoided using the properties of Kronecker products. Let

$$V = \begin{pmatrix} (1+\gamma^2)\Sigma & \gamma^2 \boldsymbol{r}(\boldsymbol{\theta})' \\ \gamma^2 \boldsymbol{r}(\boldsymbol{\theta}) & \gamma^2 R \end{pmatrix}, \quad \boldsymbol{v} = \begin{pmatrix} \boldsymbol{z} \\ \mathrm{vec}(Y) \end{pmatrix}, \quad K = \begin{pmatrix} H(\boldsymbol{\theta})' \\ H \end{pmatrix}$$

and $\boldsymbol{\mu} = K\boldsymbol{\beta}$. Then $\boldsymbol{v} \sim N_{n(1+p)}(\boldsymbol{\mu}, \sigma^2 V \otimes \Sigma)$. Let D be the $n \times (p+1)$ matrix such that $\mathrm{vec}(D) = \boldsymbol{v}$. Similarly, let M be such that $\mathrm{vec}(M) = \boldsymbol{\mu}$. The density of $\boldsymbol{v}$ is proportional to

$$\exp\left\{-\frac{1}{2\sigma^2}\mathrm{tr}\left(V^{-1}(D-M)'\Sigma^{-1}(D-M)\right)\right\} \frac{|\Sigma|^{-(p+1)/2}|V|^{-n/2}}{(\sigma^2)^{n(p+1)/2}}.$$

The full conditional densities of $\gamma^2, \theta, \phi_1, \phi_2, \phi_3$ and ν can be evaluated using the above expression within Metropolis-Hastings steps. Notice that it involves the solution of an $n \times n$ and a $(p+1) \times (p+1)$ linear systems as opposed to that of a $n(p+1) \times n(p+1)$ one, needed when the full matrix $V \otimes \Sigma$ is considered.

The full conditional of Σ, say, $p(\Sigma|\ldots)$ is proportional to

$$\exp\left\{-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}(kS + (D-M)V^{-1}(D-M)'/\sigma^2)\right)\right\} |\Sigma|^{-(p+1+k)/2} \quad .$$

This corresponds to an inverse Wishart, $IW(\Sigma|(p+k-n), kS + (D-M)V^{-1}(D-M)'/\sigma^2)$. So, Σ can be sampled directly. We use Bartlett's decomposition (Muirhead, 1982, Section

3.2.4) to obtain a sample of the Cholesky factor of Σ^{-1}. This saves the computational cost of decomposing and inverting Σ at each step of the MCMC.

The full conditional of $\boldsymbol{\beta}$ is a $N_q(\boldsymbol{\beta}|\hat{\boldsymbol{\beta}}, V_{\hat{\boldsymbol{\beta}}})$. To obtain $\hat{\boldsymbol{\beta}}$ and $V_{\hat{\boldsymbol{\beta}}}$ we observe that, conditional on all the other parameters, $\boldsymbol{v} = K\boldsymbol{\beta} + \boldsymbol{e}$, $\boldsymbol{e} \sim N_{n(1+p)}(0, \sigma^2 V \otimes \Sigma)$. Let the Cholesky decomposition of V and Σ be $L_1 L_1'$ and $L_2 L_2'$, respectively, then $V \otimes \Sigma = (L_1 \otimes L_2)(L_1' \otimes L_2')$. Let $\boldsymbol{u} = (L_1^{-1} \otimes L_2^{-1})\boldsymbol{v}$, $J = (L_1^{-1} \otimes L_2^{-1})K$ and $\boldsymbol{f} = (L_1^{-1} \otimes L_2^{-1})\boldsymbol{e}$, then

$$\boldsymbol{u} = J\boldsymbol{\beta} + \boldsymbol{f}, \quad \boldsymbol{f} \sim N_{n(p+1)}(0, \sigma^2 I).$$

$\hat{\boldsymbol{\beta}}$ is the solution of $(J'J)\hat{\boldsymbol{\beta}} = J'\boldsymbol{u}$ and $V_{\hat{\boldsymbol{\beta}}} = \sigma^2(J'J)^{-1}$. These can be obtained using ordinary least squares. To avoid computations with large matrices, notice that $\boldsymbol{u} = (L_1^{-1} \otimes L_2^{-1})\boldsymbol{v} = \mathrm{vec}(L_2^{-1} D L_1^{-t})$. Analogous calculations can be done for each column of K to obtain the columns of J.

Finally, σ^2 can be sampled from its full conditional, consisting of an inverse gamma $IG(\sigma^2 | n(p+1)/2, \mathrm{tr}(V^{-1}(D-M)'\Sigma^{-1}(D-M)/2))$.

3 Results

We run the MCMC for 100,000 iterations with a burn in of 10,000 iterations. For output analysis we use the package Bayesian Output Analysis Program (`BOA`) (Smith, 2005) within `R` (R Development Core Team, 2005). As for all problems where the posterior distribution corresponds to a large dimensional space, it is difficult to have a global assessment of convergence. We explore the traces of several parameters as well as the trace of the log-likelihood.

A summary of the posterior distributions of $\sigma^2, \phi_1, \phi_2, \phi_3$ and ν is presented in Table 2. These results are obtained by setting $\gamma^2 = 10$. We run the model for different values of γ^2 and observed that the resulting samples for σ^2 were essentially rescaled. Most importantly, the posterior distribution for $\boldsymbol{\theta}$ changed very little. We observe that the ranges and the smoothness parameters differ substantially for the three different climate parameters. This is not surprising because they correspond to different physical quantities. We notice in passing that the posterior distributions of the smoothness parameters are concentrated well below their prior mean, indicating that the data provide substantive information about the roughness of the Gaussian field. The posterior distributions for the components of $\boldsymbol{\beta}$ are summarized in Table 3. Recall that the surface temperature diagnostic consists of a 20 dimensional vector corresponding to four zonal bands, whose effects are modeled by $\beta_4, \beta_5, \beta_6$, and five decades, whose effects are modeled by $\beta_7, \ldots, \beta_{10}$. We observe that the latitude effect is essentially negligible, whilst the effect of the decades is clearly positive. As for the climate parameters, the increases of $\mathcal{S}$ and $\mathcal{F}_{aer}$ have a positive effect on the temperature (β_2 and β_3), while that of $\mathcal{K}_v$ has a negative one (β_1). The baseline is not significantly different from zero (β_0). This is sensible given that the response variables are temperature anomalies.

To assess the validity of the Gaussian model fit to the MIT2DCM output we sampled at random 43 points from the set of available parameter configurations used to run the MIT2DCM. We denote those points as $\tilde{Y}$. These correspond to about 10% of the total points. We fitted the model excluding $\tilde{Y}$. We obtained samples of the joint predictive posterior distribution, say $p(\tilde{Y}|Y, \boldsymbol{z})$ by sampling from $p(\tilde{Y}|Y, \boldsymbol{z}, \boldsymbol{\theta}^{(i)}, \boldsymbol{\phi}^{(i)}, \boldsymbol{\nu}^{(i)}, \sigma^{2(i)}, \boldsymbol{\beta}^{(i)}, \Sigma^{(i)}) =$

quantile	σ^2	ϕ_1	ϕ_2	ϕ_3	ν_1	ν_2	ν_3
2.5%	1.56	156.9	30.60	2.011	0.1093	0.3725	0.2036
50%	1.63	190.6	32.53	2.176	0.1145	0.3816	0.2154
97.5%	2.05	342.8	29.20	2.732	0.1335	0.4128	0.2592

Table 2: Summary of the posterior distributions of the covariance parameters.

	Baseline	Climate parameters		
quantile	β_0	β_1	β_2	β_3
2.5%	-0.1257	-0.0239	0.0054	0.0765
50%	0.0172	-0.0146	0.0093	0.1252
97.5%	0.1599	-0.0057	0.0132	0.1745

	Latitude			Decades			
quantile	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}
2.5%	-0.4267	-0.2457	-0.2994	-0.0954	-0.0846	0.0449	0.1909
50%	-0.0505	0.0849	-0.0591	-0.0111	-0.0095	0.1219	0.2654
97.5%	0.3220	0.4129	0.1813	0.0746	0.0656	0.2001	0.3400

Table 3: Summary of the posterior distributions of the components of $\boldsymbol{\beta}$

$N_{20\times 43}(\boldsymbol{m}^{(i)}, \sigma^{2(i)} W^{(i)} \otimes \Sigma^{(i)})$, where the superscript (i) denotes samples from the i-th iteration of the MCMC. To calculate $\boldsymbol{m}^{(i)}$ and $W^{(i)}$ we build the matrices V and K and the vector $\boldsymbol{v}$ that correspond to stacking Y and $\tilde{Y}$. We use subindex 1 to denote the blocks within those arrays that correspond to Y and subindex 2 for those that correspond to $\tilde{Y}$. Then $\boldsymbol{m}^{(i)} = K_2^{(i)}\boldsymbol{\beta}^{(i)} + (V_{21}^{(i)}(V_{11}^{(i)})^{-1} \otimes \boldsymbol{I})(\boldsymbol{v}_1 - K_1^{(i)}\boldsymbol{\beta}^{(i)})$ and $\boldsymbol{W}^{(i)} = V_{22}^{(i)} - V_{21}^{(i)}(V_{11}^{(i)})^{-1}V_{12}^{(i)}$. Figure 3 shows the posterior predictive means for two randomly selected configurations. The images are based on 1,000 samples from the joint posterior predictive distribution of the 43 configurations. From the left hand panels we observe that the predictive mean matches the MIT2D model output very closely. Predictions are not equally accurate for the parameter values in the right hand panels, but the temperature patterns follow a very similar layout. For the remaining 41 configurations we observed that predictions have a structure that is in correspondence with the output from the MIT2DCM and their values are fairly close. In fact the 25%, 50% and 75% quantiles of the differences between the predicted means and the model output are, respectively, $-0.0486, 0.0046$ and 0.0542. This is to be compared to model output values ranging from -0.346 to 0.753. A graphical exploration of the predictive marginals for each of the 860 components of $\tilde{Y}$ produced results similar to the ones shown in figure 4. We conclude that the predictive means are pretty close to the MIT2DCM output. So the model is not introducing any predictive bias. The predictive uncertainty is somewhat large and is introducing some variability in the analysis that is accounted for due to the Bayesian nature of the method.

Figure 5 presents the posterior distribution for each of the three climate parameters. These are compared to the histograms of the values used to run the MIT2DCM and the corresponding priors. From the figure it is clear that the posterior distribution of $\mathcal{K}_v$ follows the prior very closely and that large values of $\mathcal{K}_v$ are completely discarded. The posterior median is 1.97 cm/$\sqrt{\text{s}}$. The posterior density of $\mathcal{S}$ is shifted to the right in relation to the

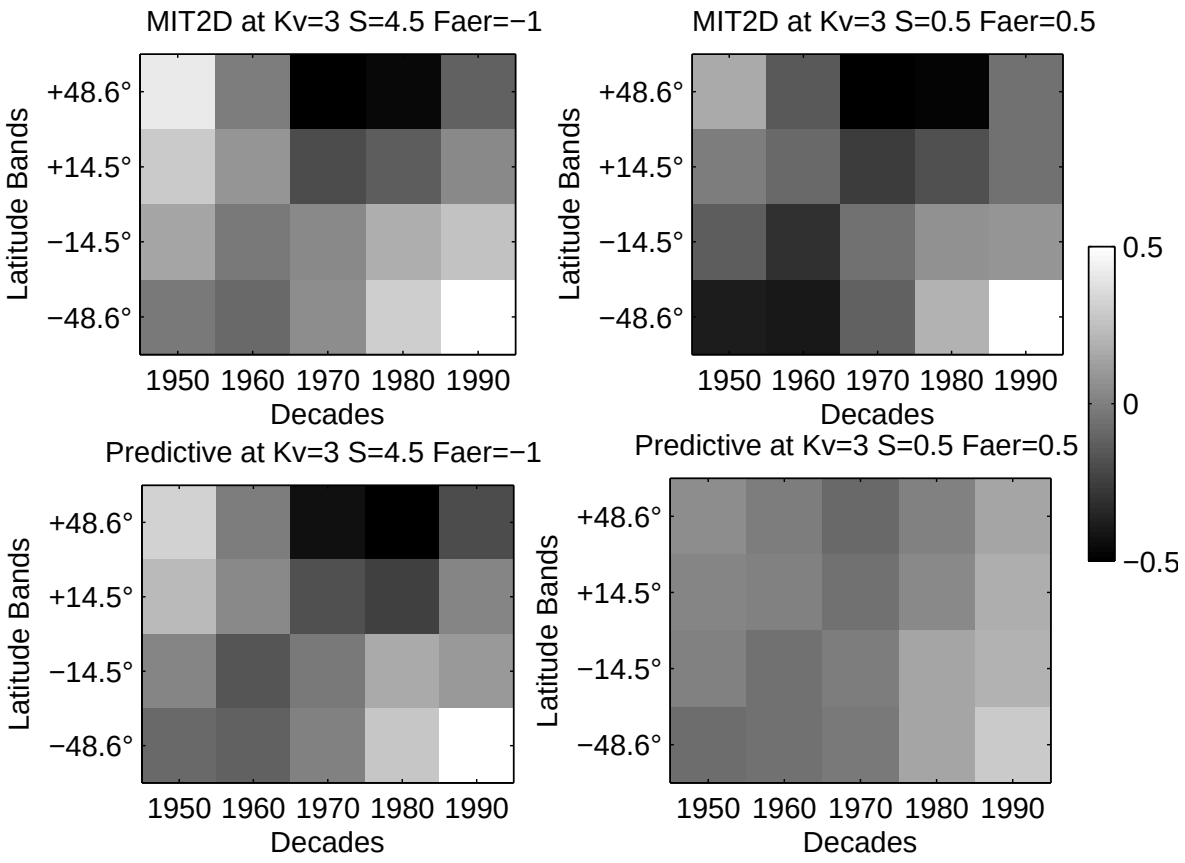

Figure 3: Posterior predictive means of the surface temperature produced by the MIT2DCM for two different configurations of the climate parameters. The top panels correspond to actual model output. The bottom panel correspond to the averages of 1,000 samples from the joint posterior predictive distribution. The gray scale for this figure and Figure 2 are the same.

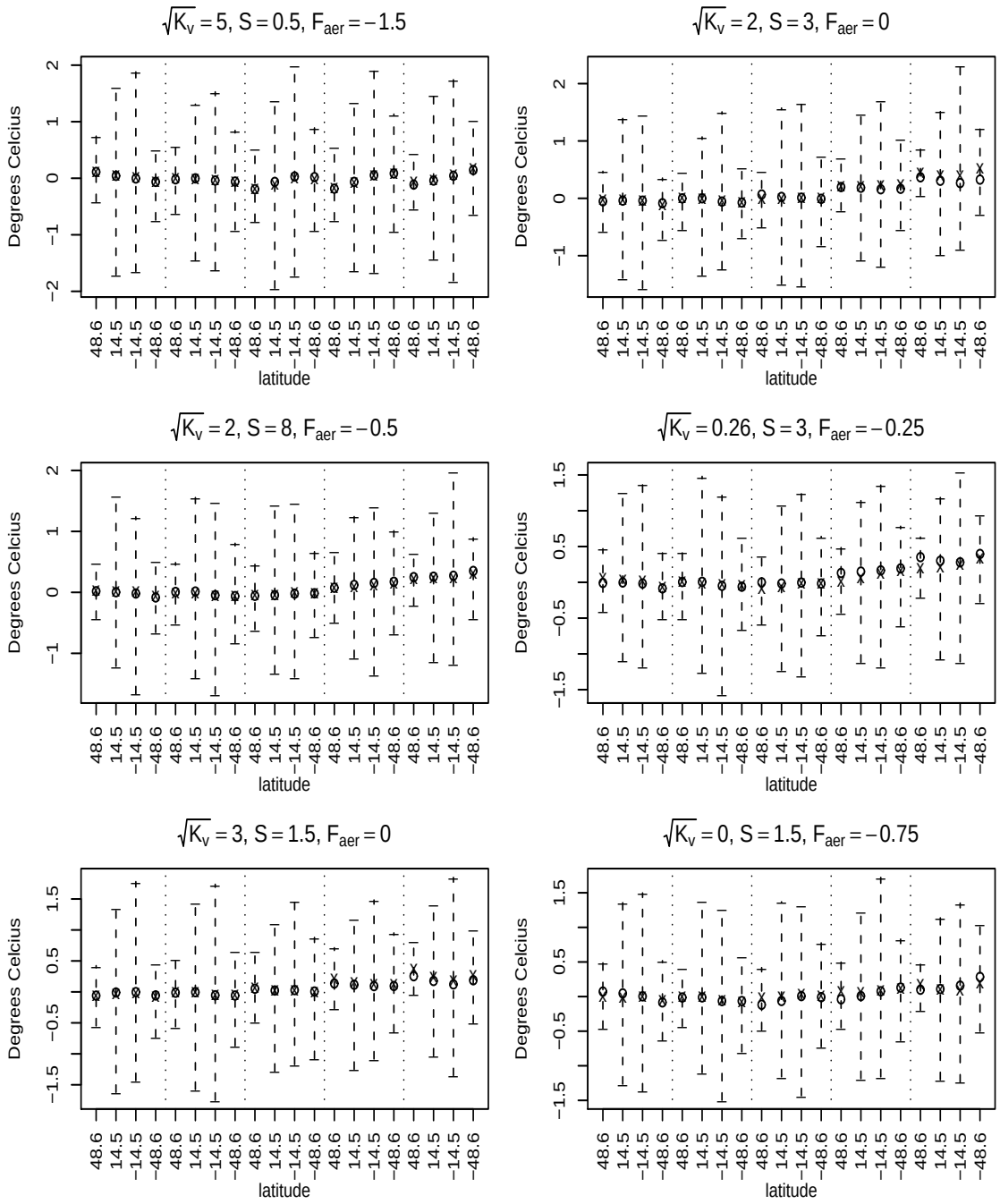

Figure 4: Marginal posterior predictive distributions for six combinations of the climate parameters randomly selected among the 43 used for validation. The dotted lines are used to group the five different decades, with latitude decreasing from left to right within each decade. The extremes of the intervals correspond to the 2.5% and 97.5% predictive quantiles. The crosses to the predictive means and the circles to the MIT2DCM output (not used to obtain the predictive distributions).

prior, the prior median being 2.36 K and the posterior median being 2.61 K. We observe the probability of small values is much smaller under the posterior than under the prior. The posterior density of $\mathcal{F}_{aer}$ has a median of -0.49 W/m^2, which is almost identical to the prior median. Nevertheless the posterior is much more concentrated and gives smaller probability to low values of $\mathcal{F}_{aer}$ than the prior. Two main inferences can be drawn from these results. The modal values of these distributions do not change from the priors except for $\mathcal{S}$. The surface temperature diagnostic has valuable information for constraining the lower bound on $\mathcal{S}$ and reducing the width of $\mathcal{F}_{aer}$ distribution. Further comparisons with previous results will be discussed in the conclusions. Given the use of only one climate change diagnostic, we cannot elaborate fully on the implication of these results on the projection of future climate but the consistency of these results for central values suggests that previous results appear robust to this more complete accounting for potential errors.

Figure 6 presents the three bivariate posterior marginals. We observe that modal values in the distributions correspond with the modes in the marginal distributions in Figure 5. The bivariate densities show similar correlations as expected in previous work. Namely, the surface temperature patterns are inconsistent with a combination of large $\sqrt{K_v}$ and low S or low $\sqrt{K_v}$ and high S. This implies that both high and low temperature responses are rejected as inconsistent with the observational record. Likewise, the accepted combinations of F_{aer} and S indicate a strong aerosol cooling requires a high climate sensitivity (and vice versa) to remain consistent with 20th century observations.

As mentioned in Section 2, the estimation of the natural variability plays a central role in the literature on detection and attribution of climate change. Of particular interest is the estimation of the largest eigenvalues and eigenvectors. We obtain the spectral decomposition of the samples of the posterior distribution of Σ, available as output from the MCMC, to estimate the eigenvalues and eigenvectors of Σ. For identifiability purposes, we impose the restriction that all samples of a given eigenvector are such that their first component is of the same sign. Figure 7 we present a summary of the eigenvalue estimation. We observe that the first eigenvalue is very clearly separated from the rest and that it explains about 50% of the variability. Most of the remaining variance is explained by the eigenvalues second through sixth or seventh. Figure 8 provides a graphical representation of the eigenvectors corresponding to the six largest eigenvalues. We observe that the components of the first eigenvector vary mostly from decade to decade. This is more clearly seen in Figure 9. We notice that the components that correspond to the south pole are shrunk towards zero in all decades. Back to Figure 8, we observe that for most of the eigenvectors, the effect of latitude is not very strong. The sixth eigenvector is a clear exception to this fact. When the vector of observations $\boldsymbol{z}$ is projected onto the space generated by the first six eigenvectors, we observe that the coordinate that corresponds to the first eigenvector has 2.5%, 50% and 97.5% posterior quantiles equal to $(0.26, 0.43, 0.49)$, respectively. For all the other eigenvectors the .25% and 97.5% quantiles have opposite signs. This is an indication that the variability in the historical records is captured mainly by the first eigenvalue of Σ.

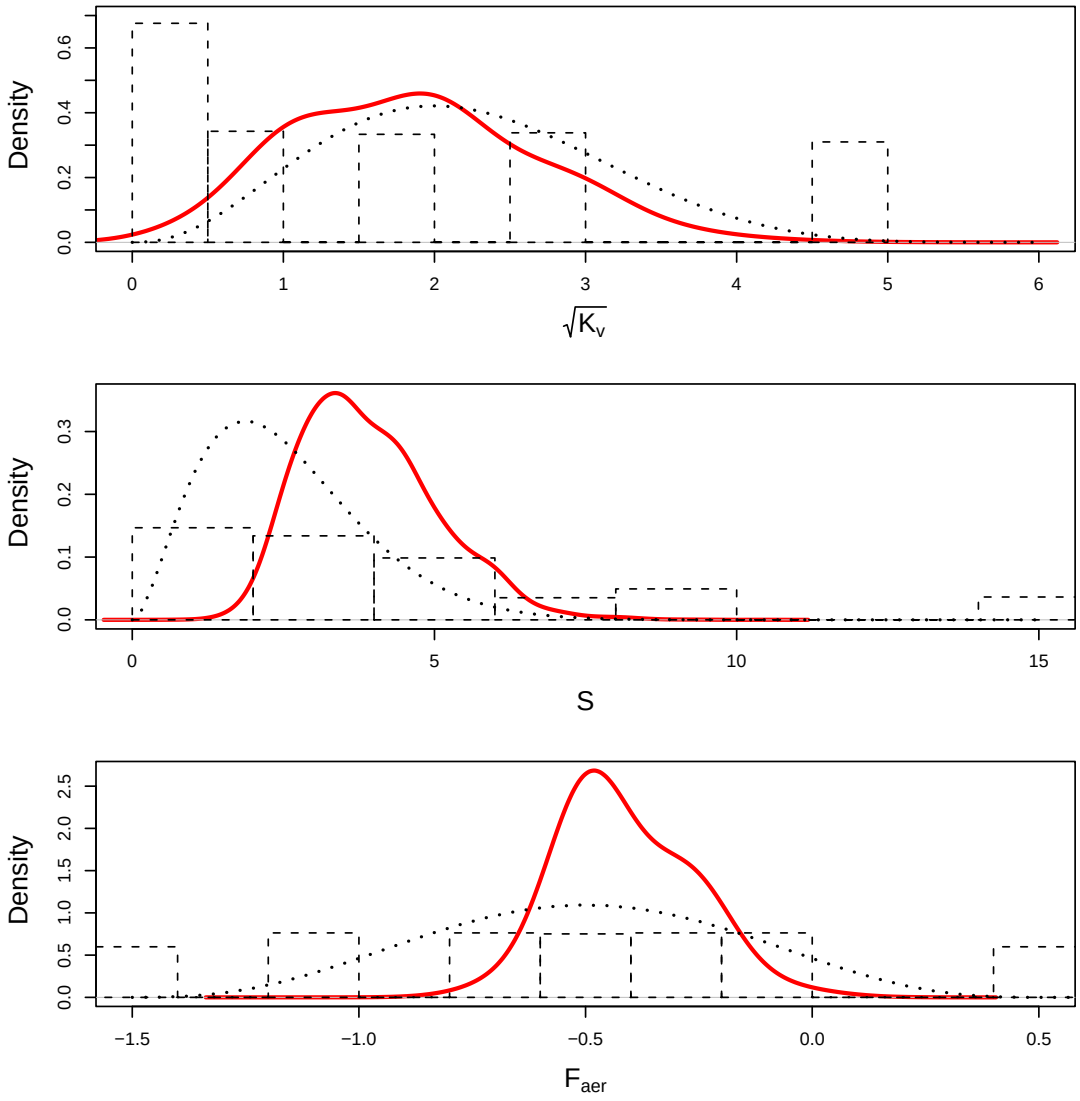

Figure 5: Distributions of the three climate parameters. The solid lines correspond to marginal posterior densities. The dotted lines correspond to the prior distributions. The histograms correspond to the values of each of the components of $\boldsymbol{t}_i, i = 1, \ldots, p$.

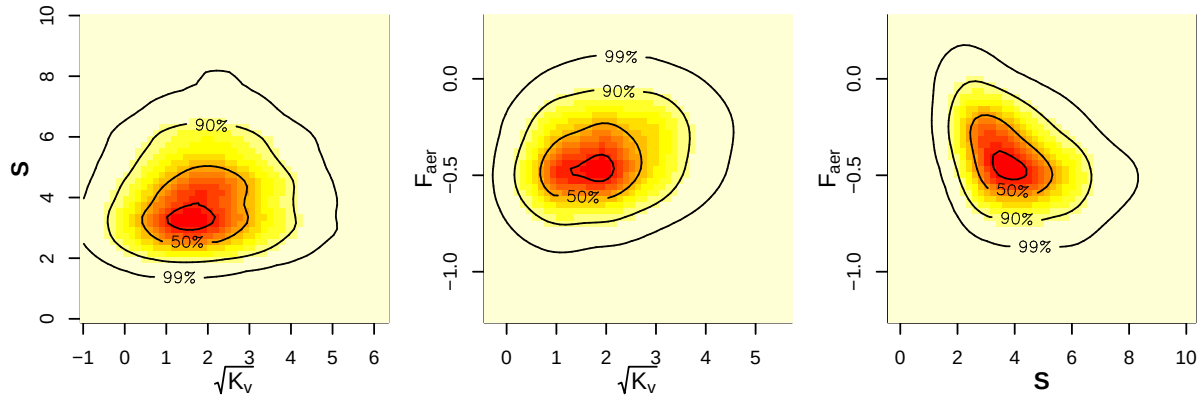

Figure 6: Bivariate posterior distributions of the three climate parameters. The contours correspond to regions of 95%, 90%, 50% and 10% posterior probability.

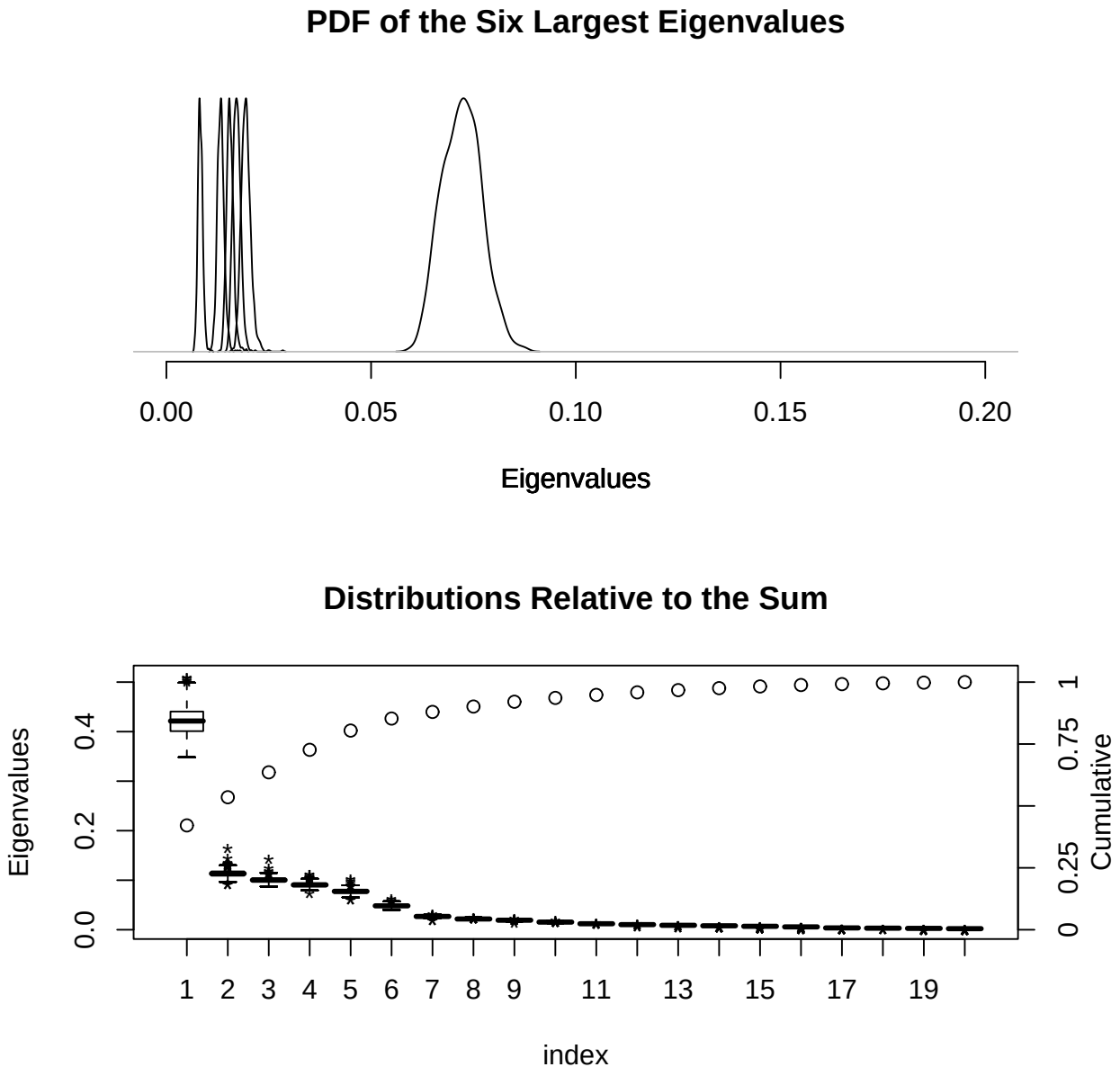

Figure 7: Posterior distribution of the eigenvalues of Σ. The top panel presents the posterior densities of the six largest eigenvalues. The curves have been rescaled so that all have the same height. The bottom panel shows the box-plots of the values of the 20 eigenvalues divided by the sum of the means of all 20 eigenvalues. The circles represent the cumulative sum of the means relative to the total.

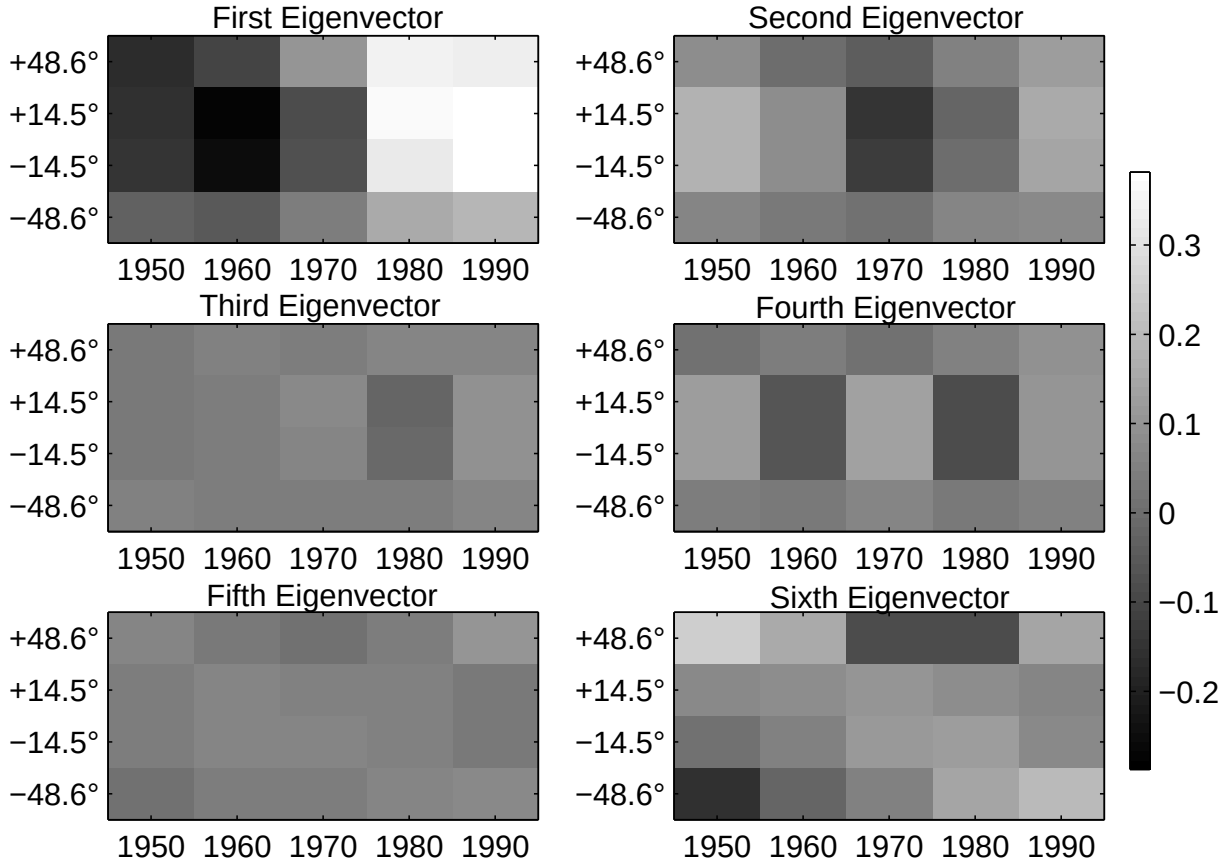

Figure 8: Posterior means of the eigenvectors of Σ corresponding the six largest eigenvalues.

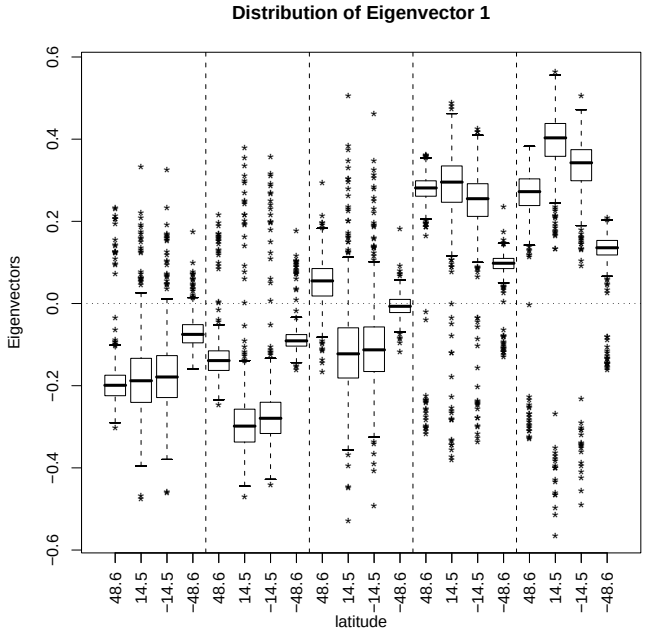

Figure 9: Posterior distribution of the components of the first eigenvector. The dashed lines are used to group the five different decades, with latitude decreasing from left to right within each decade.

4 Conclusions and future work

In this paper we have considered the problem of calibrating the properties of the climate system using historical records and output from the MIT2DCM. Our approach uses information from different sources, including key expert knowledge. It considers a multivariate output on a large number of points. By using a Bayesian method the uncertainty about the calibration parameters is expressed probabilistically and estimation uncertainties are accounted for.

As with most computer model calibration problems we were faced with the issue of building a statistical equivalent model to obtain fast approximations to the model output. We used more than 400 points to estimate the parameters in a Gaussian process, indexed on three dimensions, with values on a twenty dimensional space. Such a problem offers a formidable number of degrees of freedom and we were forced to strike a balance between accuracy and complexity. Given the dimensionality of the problem, it is unlikely that high predictive accuracy will be attainable with a simple model. Potentially more structured models incorporating more physical principles could be used. But adding too much complexity to a multidimensional model deceives the purpose of using a fast approximation that can be evaluated very many times as a proxy for the actual computer model. So, it is important to account for the variability introduced by the statistical equivalent model, as we do within a Bayesian framework.

The estimation of Σ provides evidence that only a handful of eigenvalues can describe the variability in the data not explained by the mean. This fact is commonly used in

	Current results			Forest et al. (2006)		
	$\mathcal{S}$	$\mathcal{K}_v$	$\mathcal{F}_{aer}$	$\mathcal{S}$	$\mathcal{K}_v$	$\mathcal{F}_{aer}$
5%	2.4	0.72	-0.59	1.9	0.2	-0.69
50%	3.8	1.9	-0.45	2.8	0.73	-0.43
95%	5.9	3.2	-0.19	4.7	1.90	-0.14
mode	3.3	1.9	-0.48	2.8	0.60	-0.45

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

climate studies where the dimensionality of the data is large compared to the number of data points, making the estimation of the full covariance matrix problematic. We are currently investigating methods of using only a limited number of eigenvalues and how this affects the calibration.

The method presented in this paper can be easily adapted to the diagnostics from deep-ocean and upper atmospheric temperatures. Ideally we could use all of them together, since they all provide information about the climate parameters. Qualitatively, the deep-ocean temperatures offer information on the rate of heat uptake and thus, will place limits on combinations of high $\mathcal{S}$ and high $\mathcal{K}_v$. The upper-air temperatures provide information similar to the surface temperature diagnostic although from independent data sources. Results in Forest et al. (2006) were obtained assuming independence of the three diagnostics. From a statistical point of view, considering all three diagnostics jointly presents the challenge of creating a likelihood that encompasses all temporal and spatial dependencies. A difficult task since they are summaries of information covering large areas and time periods.

For the climate sciences it is useful to consider what the surface temperature alone indicates for the posterior distribution of the climate parameters. Table 4 reports the 5%, 50%, 95% quantiles and modes of the marginal posterior distributions from the current model and the model in Forest et al. (2006). In addition to using all three diagnostics, the latter model differs from the one developed in this paper in the priors on $\mathcal{K}_v$ and $\mathcal{F}_{aer}$, which were assumed uniform with specified ranges, the estimation of Σ, based solely on the control runs, and the fact that it does not include a statistical equivalent model for the MIT2DCM. From the comparison, two features quickly appear. First, the current $p(\mathcal{K}_v)$ has a much higher mode than in previous results although the shape has been changed little from the data. This indicates the deep-ocean temperatures are critical for setting the most likely values of $\mathcal{K}_v$. Second, while the modes are similar for $\mathcal{S}$ and $\mathcal{F}_{aer}$, the ranges are slightly wider in the current results with the exception of the lower bound on $\mathcal{S}$. This could result from a combination of factors including the effect of the additional diagnostics. The higher lower bound on $\mathcal{S}$ is consistent with the surface temperature diagnostic implying a higher response given the $\mathcal{K}_v$ distribution. With deep-ocean temperatures included, the $\mathcal{S}$ and $\mathcal{K}_v$ distributions shift to lower values. So this is consistent with Forest et al. (2006) and suggests that the surface temperatures have more influence on $\mathcal{S}$ and $\mathcal{F}_{aer}$ distributions. These two features show both the value of the additional diagnostics and the potential effect of including the additional uncertainties in the calibration algorithm.

In this paper we have assumed that dependence on θ is present only in the first moment of the distributions of the variables under study. This implies that the covariance matrix

Σ does not depend on θ. Since the information from the GCM runs is summarized using anomalies, the mean of the control run is assumed to be zero and no information about θ is obtained from those simulations. Control runs from the GCM data are obtained from a model with specific values of the climate parameters and it would be desirable to incorporate this fact in the statistical analysis. The control runs from alternate GCMs with different θ values would include such information.

We recall that the MIT2DCM represents a class of climate models for which the climate system properties, θ, can be specified by single variables. This differs from more complex GCMs, such as that used for the prior on Σ, which contain multiple parameterizations each with a set of adjustable parameters. In both the MIT2DCM and the GCMs, the climate system properties are diagnosed from the response of each model to specified forcings. At some level, the equivalence between the MIT model and the GCMs must be established. The key difference is that a single parameter can be varied in the MIT model whereas a subset of adjustable variables in GCMs must be first identified (as in Allen, 1999) and then sampled to estimate the correspondence with a given θ_i (e.g., Stainforth et al., 2005). As a climate modeling project to establish these equivalences, this requires significant computing resources. As a computational statistical calibration project, the required resources would be far larger. The approach taken in this paper (to build a consistent emulator and parameter calibration) could be applied to a more complex climate model provided the resources and data were available. Such an exercise would be very useful to the climate science community.

Acknowledgments

The authors were partially supported by the National Science Foundation grant NSF-Geomath 0417753.

References

Allen, M. R. (1999) Do-it-yourself climate prediction. *Nature*, **401**, 627.

Allen, M. R. and Tett, S. F. B. (1999) Checking for model consistency in optimal fingerprinting. *Climate Dynamics*, **15**, 419–434.

Bell, T. L. (1982) Optimal weighting of data to detect climatic change: Application to the carbon dioxide problem. *J. Geophys. Res.*, **87**, 11161–11170.

— (1986) Theory of optimal weighting of data to detect climatic change. *J. Atmos. Sci.*, **43**, 1694–1710.

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.

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

Forest, C. E., Allen, M. R., Stone, P. H. and Sokolov, A. P. (2000) Constraining uncertainties in climate models using climate change detection methods. *Geophysical Research Letters*, **27**, 569–572.

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**, doi:10.1029/2005GL023977.

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.

Gamerman, D. and Lopes, H. F. (2006) *Markov Chain Monte Carlo - Stochastic Simulation for Bayesian Inference*. London, UK: Chapman and Hall, second edn.

Gill, A. E. (1982) *Atmosphere-Ocean Dynamics*. Academic Press, San Diego, CA.

Hasselmann, K. (1979) On the signal-to-noise problem in atmospheric response studies. In *Meteorology of Tropical Oceans* (ed. Shawn), 251–259. Royal Meteorological Society.

— (1993) Optimal fingerprints for the detection of time dependent climate change. *J. Climate*, **6**, 1957–1971.

— (1997) On multifingerprint detection and attribution of anthropogenic climate change. *Clim. Dyn.*, **13**, 601–611.

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.

Kennedy, M. C. and O'Hagan, A. (2001) Bayesian calibration of computer models. *Journal of the Royal Statistical Society, Series B*, **63**, 425–464.

Levitus, S., Antonov, J. and Boyer, T. P. (2005) Warming of the world ocean, 1955–2003. *Geophys. Res. Let.*, **32**, doi:10.1029/2004GL021592.

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

Muirhead, R. J. (1982) *Aspects of Multivariate Statistica Theory*. New York, USA: John Wiley and Sons.

O'Hagan, A., Kennedy, M. C. and Oakley, J. E. (1999) Uncertainty analysis and other inference tools for complex computer codes. In *Bayesian Statistics 6* (eds. J. M. Bernardo, J. O. Berger, A. P. Dawid and A. . F. M. Smith), 503–524. Oxford University Press.

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.

Paulo, R. (2005) Default priors for Gaussian processes. *The Annals of Statistics*, **33**, 556–582.

R Development Core Team (2005) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: `http://www.R-project.org`. ISBN 3-900051-07-0.

Smith, B. J. (2005) *BOA: Bayesian Output Analysis Program (BOA) for MCMC*. URL: `http://www.public-health.uiowa.edu/boa`. R package version 1.1.5-2.

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.

Sokolov, A. P. and Stone, P. H. (1998) A flexible climate model for use in integrated assessments. *Climate Dynamics*, **14**, 291–303.

Stainforth, D. A., Aina, T., Christensen, C., Collins, M., Faull, N., Frame, D. J., Kettleborough, J. A., Knight, S., Martin, A., Murphy, J. M., Piani, C., Sexton, D., Smith, L. A., Spicer, R. A., Thorpe, A. J. and Allen, M. R. (2005) Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, **433**, 403–406.

Stein, M. (1999) *Interpolation of Spatial Data*. New York, USA: Springer-Verlag.

Watson, R. T. and the Core Writing Team (eds.) (2001) *Third Assessment Report on Climate Change 2001: Synthesis Report of the Intergovernamental Panel on Climate Change*. Cambridge, UK: Cambridge University Press.

Webster, M. D. and Sokolov, A. P. (2000) A methodology for quantifying uncertainty in climate projections. *Climatic Change*, **46**, 417–446.