Statistical Calibration of Climate System Properties - II

Bruno Sansó and Chris Forest *

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 climate sensitivity, the effective thermal diffusivity in the deep-ocean and the net aerosol forcing that control the MIT2DCM. 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, 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/

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 state of the art complex climate models by modifying individual parameters or parameterizations. In 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 adjusts the cloud feedback, which controls the effect of clouds on the radiative transfer in the energy budget. Another important uncertain process 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 forcing factors in 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, different statistics are used to form "diagnostics". This will be explained in the next paragraph. In this paper we focus on *Surface temperature change* (STC), as it is considered one of the most informative summaries. This is obtained as 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. Although other diagnostics have been used successfully, the STC provides the strongest restriction on the model response (i.e., the model parameter) related to surface temperature changes. The fact that the STC is 20 dimensional vector, requires us to focus on the assessment of the uncertainty in the covariance structure. Additional diagnostics are needed to understand the limits on a model's deep-ocean temperature response which is where more than 90% of the heat energy is stored (Levitus et al., 2005), but this will not be the focus of this paper.

In correspondence to the summaries of the climate model output, it is possible to obtain statistics based on historical observations. In this paper the surface temperature data are obtained from Jones et al. (1999). To obtain a "diagnostic", the observations are compared to the climate model summaries. We can then perform the calibration of the climate model parameters. Figure 2 illustrates the pattern of temperature change in the observational data and in the simulations of the computer model. Intuitively, we want to find the combination of parameters that produces an output that is as close as possible to the top left panel.

Let $\boldsymbol{z}$ be the vector of observations corresponding to the historical STC and $\boldsymbol{F}(\boldsymbol{\theta})$ the vector corresponding to STC obtained from the MIT2DCM evaluated at a given combination of parameter values $\boldsymbol{\theta} = (\mathcal{K}_v\ , \mathcal{S}\ , \mathcal{F}_{aer}\ )$. Throughout the paper light face with lower case is used for scalars, bold face is used for vectors and capital letters are used for matrices. The approach followed in Forest et al. (2000) and subsequent work, is to obtain a probability distribution for $\boldsymbol{\theta}$ based on the quadratic form $(\boldsymbol{z} - \boldsymbol{F}(\boldsymbol{\theta}))^t \Sigma^{-1} (\boldsymbol{z} - \boldsymbol{F}(\boldsymbol{\theta}))$. The covariance matrix Σ has been typically estimated from a number of control runs, corresponding to the output of a GCM. These are simulations of the climate system without changes in the external forcings. They provide a sample of the unforced climate variability. In Figure 3 three different assumptions for Σ are considered and we observe that inference for $\boldsymbol{\theta}$ is strongly affected by these. Sansó et al. (2008) develop a model where all the available sources of information, i.e. historical records, MIT2DCM output and GCM model output, have covariance matrices that are multiples of Σ. In this paper we consider more flexible and realistic error structures. This is a problem that has an interest on its own. In fact,

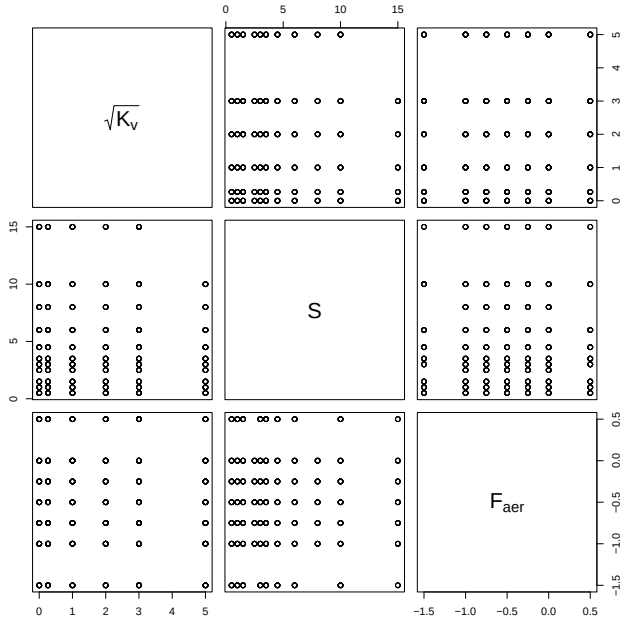

Figure 1: Configuration of the 426 different combinations of $\mathcal{S}$, $\mathcal{K}_v$ and $\mathcal{F}_{aer}$ used for simulations obtained with the MIT2DCM.

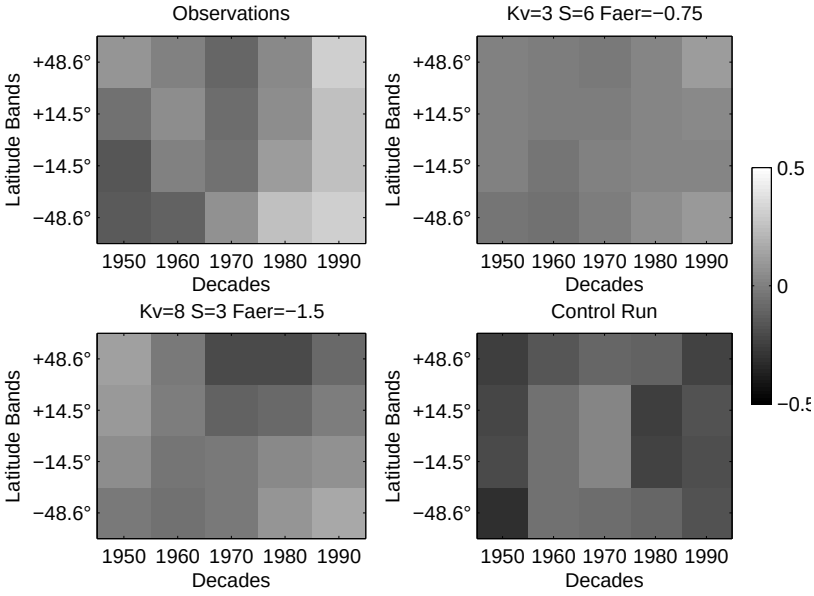

Figure 2: Comparison of the historically observed surface temperature change (top left panel) to MIT2DCM simulations obtained with two different parameter configurations (top right and bottom left panels) and a sample from the HadCM2 GCM (bottom right). All four figures are in the same scale in degrees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

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).

In the context of this paper, 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. 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 $\boldsymbol{F}(\cdot)$ is evaluated at a limited number of points. Evaluating $\boldsymbol{F}(\cdot)$ is very time consuming, so it is impractical to consider statistical procedures that require additional evaluations. For all the other values of $\boldsymbol{\theta}$, $\boldsymbol{F}(\cdot)$ 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. In most computer model applications such a surrogate model is fitted to the model output. In a second stage, the statistical model is used to perform the calibration of the computer model parameters. There is a large body of literature on the analysis of computer models. The books by Santner et al. (2003) and Fang et al. (2006) and the paper by Bayarri et al. (2007) contain a number of useful references. In this paper we take a comprehensive approach, where all parameters are estimated within the estimation procedure and so all the uncertainty produced by the estimation of the statistical model is considered in the calibration phase.

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 $\boldsymbol{t} = (\mathcal{K}_v\ , \mathcal{S}\ , \mathcal{F}_{aer}\ ) \in \mathbb{R}^3$ is denoted as $f(\boldsymbol{x}, \boldsymbol{t})$. The available model runs are denoted as $y_{ij} = f(\boldsymbol{x}_i, \boldsymbol{t}_j), i = 1, \ldots, n; j = 1, \ldots, p$, $\boldsymbol{y}_j = (y_{1j}, \ldots, y_{nj})'$ 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 STC $n = 20$, 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 HadCM2 model, a GCM developed at the Hadley Centre, UK. Table 1 shows a summary of the available information and the notation used.

We denote as $\zeta(\boldsymbol{x})$ the true, unobserved, STC at location $\boldsymbol{x}$, and $\boldsymbol{\zeta} = (\zeta(\boldsymbol{x}_1), \ldots, \zeta(\boldsymbol{x}_n))$. We denote $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ as the true value of the climate system properties $(\mathcal{K}_v\ , \mathcal{S}\ , \mathcal{F}_{aer}\ )$. Then we assume that $z_i = \zeta(\boldsymbol{x}_i) + \varepsilon_i$. So that observations deviate from the true value of STC by an observational error ε_i. In this formulation ε_i encompasses both, measurement errors and so called representativeness errors, or errors on spatial and temporal scales that

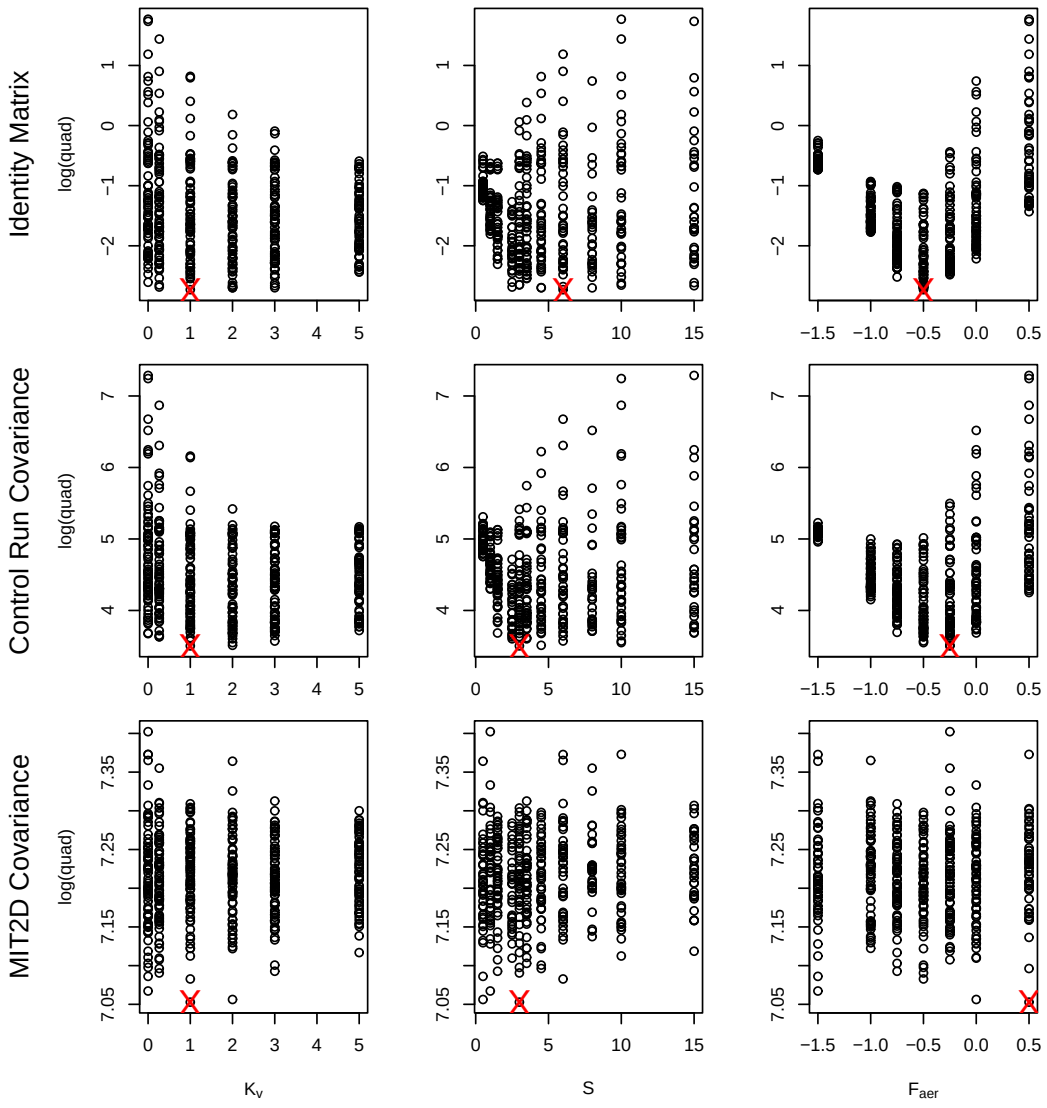

Figure 3: Values of the quadratic form given by the difference between observed and simulated STCs for different choices of covariance matrices. The first row corresponds to $\Sigma = I$. The second row corresponds to Σ estimated from 162 HadCM2 control runs. The third row corresponds to Σ estimated from the 426 available MIT2DCM runs. The parameter values that correspond to the minima are marked with a cross.

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.

are smaller than the grid cell size. We assume that $(\varepsilon_1, \ldots, \varepsilon_n)' \sim N_n(0, \tau^2 I)$, where $N_n(\cdot, \cdot)$ denotes an n-dimensional normal distribution.

The code output is proportional to the true value, but there is a model inadequacy. In a calibration problem like the one we consider here it is very difficult to identify both multiplicative and additive biases, in the absence of substantive prior information. So we limit the model inadequacy to an additive term. Let $\boldsymbol{F}(\boldsymbol{\theta}) = (f(\boldsymbol{x}_1, \boldsymbol{\theta}), \ldots, f(\boldsymbol{x}_n, \boldsymbol{\theta}))$, and $\delta(\boldsymbol{x})$ be the discrepancy between the truth and the computer model value, then

$$\zeta(\boldsymbol{x}_i) = f(\boldsymbol{x}_i, \boldsymbol{\theta}) + \delta(\boldsymbol{x}_i) \quad . \tag{1}$$

We assume that $\boldsymbol{\delta} = (\delta(\boldsymbol{x}_1), \ldots, \delta(\boldsymbol{x}_n))$ has mean zero. This is justified by the fact that climate models can predict temperature changes with reasonable accuracy (see, for example Tebaldi and Sansó, 2008). $\boldsymbol{\delta}$ represents variability not modeled by the MIT2DCM that we assume has a covariance structure proportional to that in the control runs. Thus, we have that $\boldsymbol{w}_l \sim N_n(0, \Sigma)$ for $l = 1, \ldots, k$ and $\boldsymbol{\delta} \sim N_n(0, \sigma^2\Sigma)$. This effectively implies that we will use the control runs as prior information for the estimation of the covariance of the MIT2DCM inadequacy.

To model $f(\boldsymbol{x}, \boldsymbol{t})$, for any value of $\boldsymbol{t}$, we assume that $f(\boldsymbol{x}, \cdot)$ 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 $E(f(\boldsymbol{x}, \boldsymbol{t})) = \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{t})'\boldsymbol{\beta}$, where $\boldsymbol{h}$ and $\boldsymbol{\beta}$ are q-dimensional vectors. They define a linear combination of effects due to locations and parameter values. Thus $f(\boldsymbol{x}, \boldsymbol{t}) = \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{t})'\boldsymbol{\beta} + \eta(\boldsymbol{x}, \boldsymbol{t})$, with $E(\eta(\boldsymbol{x}, \boldsymbol{t})) = 0$ and

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

Here Ψ_{ij} are the components of a covariance matrix Ψ and $r(\cdot, \cdot)$ is a correlation function. Thus, we are assuming separability between $\boldsymbol{x}$ and $\boldsymbol{t}$. Furthermore, as the space of climate parameters is not naturally geometrical, we assume separability for r as well. Thus

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

where

$$r_i(t_i, t_i') = \frac{1}{2^{\nu_i - 1}\Gamma(\nu_i)} \left(\frac{|t_i - t_i'|}{\phi_i} \right)^{\nu_i} K_{\nu_i} \left(\frac{|t_i - t_i'|}{\phi_i} \right) \quad i = 1, 2, 3.$$

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.

To obtain the likelihood of the proposed model we need the joint density of $\boldsymbol{z}$ and Y. To proceed, we condition on $\boldsymbol{\zeta}$ and $\boldsymbol{\delta}$, and calculate $\mathrm{cov}(\zeta_i, y_{jk}), i, j = 1, \ldots, n;\ k = 1, \ldots, p$, 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 $H(\cdot) = [\boldsymbol{h}(\boldsymbol{x}_1, \cdot), \ldots, \boldsymbol{h}(\boldsymbol{x}_n, \cdot)] \in \mathbb{R}^{q \times n}$, and let $H = [H(\boldsymbol{t}_1), \ldots, H(\boldsymbol{t}_p)]' \in \mathbb{R}^{np \times q}$, then the model is expressed by the following hierarchy

$$\begin{array}{ccl} \boldsymbol{z} & \sim & N_n(\boldsymbol{\zeta}, \tau^2 I) \\ \begin{pmatrix} \boldsymbol{\zeta} - \boldsymbol{\delta} \\ \mathrm{vec}(Y) \end{pmatrix} & \sim & N_{n(p+1)}\left(\begin{pmatrix} H(\boldsymbol{\theta})^t \\ H \end{pmatrix} \boldsymbol{\beta},\ \begin{pmatrix} 1 & \boldsymbol{r}(\boldsymbol{\theta})' \\ \boldsymbol{r}(\boldsymbol{\theta}) & R \end{pmatrix} \otimes \Psi \right) \\ \boldsymbol{\delta} & \sim & N_n(0, \sigma^2 \Sigma)\ , \end{array} \tag{3}$$

where $\mathrm{vec}(\cdot)$ denotes the operation of stacking the columns of matrix into a vector and $\otimes$ denotes the Kronecker product. We recall that Y is the output from all MIT2D model simulations and z is the set of historical records. The control simulations from the GCM model is used in the prior for Σ and does not appear directly in these equations. Table 2 reports a summary of the different errors considered in the model.

Error	Variable	Covariance
Observational	$\boldsymbol{\varepsilon}$	$\tau^2 I$
Model discrepancy	$\boldsymbol{\delta}$	$\sigma^2 \Sigma$
Surrogate model	$\boldsymbol{\eta}$	$R \otimes \Psi$

Table 2: Summary of the different errors considered in the model and their different covariances.

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 with parameters $(3.5, 6)$ distribution stretched to have support on $(0, 6)$. The prior for $\theta_2 = \mathcal{S}$ is specified as a beta distribution with parameters $(2.85, 14)$ on the support $(0, 15)$. The prior for $\theta_3 = \mathcal{F}_{aer}$ is a beta distribution with parameters $(4, 4)$ stretched and shifted to have 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 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 paleoclimate records for about the past 500,000 years.

We considered a second prior corresponding to the product of three beta distributions, shifted and stretched as above, with parameters $(2.5, 2.5); (1.5, 2.25)$ and $(2.5, 2.5)$. This prior has mass that is substantially more spread around the support than the prior above. As illustrated in Sansó et al. (2008) a uniform prior over the support provides very weak information for the model to learn about the posterior distribution of $\boldsymbol{\theta}$. We shall refer to the two priors for $\boldsymbol{\theta}$ as TP1 and TP2, respectively.

As a prior for Σ we used a density proportional to

$$\exp\left\{-\frac{k-n-1}{2}\mathrm{tr}\left(\Sigma^{-1}S\right)\right\}|\Sigma|^{-(k+n+1)/2} \tag{4}$$

where $S = 1/k\sum_{l=1}^{k}\boldsymbol{w}_l\boldsymbol{w}_j^t$. This corresponds to an inverse Wishart with k degrees of freedom and scale matrix $S^{-1}/(k-n-1)$. The parameterization in Equation (4) implies that the prior mean of Σ is equal to the covariance estimated empirically from the control runs.

We considered three different priors for Ψ. The first one is a non informative prior $p(\Psi) \propto |\Psi|^{-(n+1)/2}$, indicating that we expect the inference for Ψ to be dominated by the MIT2DCM runs. The second prior is the same as the one used for Σ, which implies that some information from the control runs will be used directly in the estimation of Ψ. Finally we considered a prior as in (4) with $20,000$ degrees of freedom (an arbitrary number that is very large compared to the others). This choice of degrees of freedom implies that the distribution is strongly concentrated around S. So this third prior mimics the traditional approach of estimating the covariance matrix from a set of control runs. We shall refer to these three priors for Ψ as PP1, PP2 and PP3 respectively. A summary of the priors for $\boldsymbol{\theta}, \Sigma$ and Ψ is reported in Table 3.

$\boldsymbol{\theta}$	Σ	Ψ
TP1=B(3.5,6)B(2.85,14)B(4,4)	$IW(k, \frac{S^{-1}}{(k-n-1)})$	PP1$\propto |\Psi|^{-(n+1)/2}$
TP2=B(2.5,2.5)B(1.5,2.25)B(2.5,2.5)		PP2= $IW(k, \frac{S^{-1}}{(k-n-1)})$
		PP3= $IW(20,000, \frac{S^{-1}}{(k-n-1)})$

Table 3: Summary of the most relevant prior distributions. $\mathrm{B}(\cdot,\cdot)$ denotes a beta density appropriately shifted and scaled.

For the regression parameters β we assume a flat prior $p(\boldsymbol{\beta}) \propto 1$. After an exploration of the MIT2DCM runs we found a transformation that linearizes the relationship between $\boldsymbol{\theta}$ and Y. Our choice of $\boldsymbol{h}$ is then given by

$$\boldsymbol{h}(\boldsymbol{x},\boldsymbol{t})' = \left(1, \frac{1}{t_1+1}, \log t_2, t_3, x_1, x_2\right)$$

where x_1 takes the values $(-2,-1,1,2)$ and x_2 takes the values $(-2,-1,0,1,2)$.

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 present when separability of the correlation function is assumed. Additionally, for comparison purposes, we used three independent inverse gamma priors with means equal, respectively, to 3, 7 and 1, which correspond to about 50% of the length of the supports of $\mathcal{K}_v$, $\mathcal{S}$ and $\mathcal{F}_{aer}$. For the smoothness parameters, ν_i, we used three independent normal priors centered around 3 with standard deviation 1. The rationale is that, as is commonly assumed in computer modeling applications, we expect the correlation to be fairly smooth, but we wanted to avoid the numerical problems associated with very large values of ν.

We considered a prior for σ^2 given by an inverse gamma with parameters 100 and 0.01. This implies that based on the prior means of Σ and σ^2 we expect that the values of $\boldsymbol{\delta}$ would be within $(-10^{-3}, 10^{-3})$°C, which seems reasonable given that $|z_i| \in (0.0012, 0.3165)$°C. We fixed the value of τ at 0.005, which implies that the observational error is within $(-0.015, 0.015)$°C.

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 $(\boldsymbol{\zeta}, \mathrm{vec}(Y))$. The dimension of such matrix is $n(1+p)\times n(1+p)$. Given that $n(1+p) = 8,540$, such computations are unfeasible within an iterative method. Lengthy computations and storage of large matrices is avoided using the properties of Kronecker products. Let

$$V = \begin{pmatrix} 1 & \boldsymbol{r}(\boldsymbol{\theta})' \\ \boldsymbol{r}(\boldsymbol{\theta}) & R \end{pmatrix}, \quad \boldsymbol{v} = \begin{pmatrix} \boldsymbol{\zeta} - \boldsymbol{\delta} \\ \mathrm{vec}(Y) \end{pmatrix}, \quad K = \begin{pmatrix} H(\boldsymbol{\theta})^t \\ H \end{pmatrix}$$

and $\boldsymbol{\mu} = K\boldsymbol{\beta}$. Then $\boldsymbol{v} \sim N_{n(1+p)}(\boldsymbol{\mu}, V \otimes \Psi)$. 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

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

Multiplying by the appropriate priors, the full conditional densities of $\phi_1, \phi_2, \phi_3, \nu_1, \nu_2$ and ν_3 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 $V \otimes \Psi$ is considered.

$\boldsymbol{\theta}$ and Ψ can be sampled jointly by noticing that

$$p(\boldsymbol{\theta}, \Psi|\ldots) \propto \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\Psi^{-1}A(\boldsymbol{\theta})\right)\right\}p(\boldsymbol{\theta})|\Psi|^{-\frac{p+d+n+2}{2}} \quad ,$$

where the dots denote the data and all the other parameters, d denote the degrees of freedom in the inverse Wishart prior for Ψ and $A(\boldsymbol{\theta}) = (d-n-1)S + (D-M)V^{-1}(D-M)'$.

Integrating Ψ we obtain that $p(\boldsymbol{\theta}|\ldots) \propto |A(\boldsymbol{\theta})|^{-(p+d+1)/2} p(\boldsymbol{\theta})$. This expression can be used within a Metropolis step to sample $\boldsymbol{\theta}$. If the candidate, say $\boldsymbol{\theta}^*$, is accepted, we can then sample Ψ from an inverse Wishart with $(p+d+1)$ degrees of freedom and scale matrix $A(\boldsymbol{\theta}^*)^{-1}$.

In a separate step we sample Σ from its full conditional that is proportional to

$$\exp\left\{-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}(\boldsymbol{\delta}\boldsymbol{\delta}'/\sigma^2 + (k-n-1)S)\right)\right\} |\Sigma|^{-(k+n+2)/2} \quad .$$

This corresponds to an inverse Wishart, $IW(\Sigma|k+1, (\boldsymbol{\delta}\boldsymbol{\delta}'/\sigma^2+(k-n-1)S)^{-1})$. To sample the full conditionals of Ψ and Σ we use Bartlett's decomposition (Muirhead, 1982, Section 3.2.4) to obtain a sample of the Cholesky factor of Ψ^{-1} and Σ^{-1}. This saves the computational cost of decomposing and inverting the matrices 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, $\boldsymbol{v} = K\boldsymbol{\beta} + \boldsymbol{e}$, $\boldsymbol{e} \sim N_{n(1+p)}(0, V \otimes \Psi)$. Consider Cholesky decompositions of V and Ψ^{-1}. So $V = L_V L_V^t$ and $\Psi^{-1} = L_\Psi L_\Psi^t$, respectively, then $V \otimes \Psi = (L_V \otimes L_\psi^{-t})(L_V^t \otimes L_\Psi^{-1})$. Let $\boldsymbol{u} = (L_V^{-1} \otimes L_\Psi^t)\boldsymbol{v}$, $J = (L_V^{-1} \otimes L_\Psi^t)K$ and $\boldsymbol{f} = (L_V^{-1} \otimes L_\Psi^t)\boldsymbol{e}$, then

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

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

Samples of $\boldsymbol{\zeta}$ are obtained from an n-variate normal with mean $\hat{\boldsymbol{\zeta}}$, that is the solution of the equation

$$\left(\frac{I}{\tau^2} + \Psi^{-1}\right)\hat{\boldsymbol{\zeta}} = \left(\frac{\boldsymbol{z}}{\tau^2} + \Psi^{-1}(H(\boldsymbol{\theta})^t\boldsymbol{\beta} + \boldsymbol{\delta})\right) \quad ,$$

and covariance matrix $(I/\tau^2 + \Psi^{-1})^{-1}$. Samples of $\boldsymbol{\delta}$ are also obtained from an n-variate normal. The mean $\hat{\boldsymbol{\delta}}$ is the solution of the equation

$$\left(\frac{\Sigma^{-1}}{\sigma^2} + \Psi^{-1}\right)\hat{\boldsymbol{\delta}} = \Psi^{-1}(\boldsymbol{\zeta} - H(\boldsymbol{\theta})^t\boldsymbol{\beta})$$

and the covariance matrix is $(\Sigma^{-1}/\sigma^2 + \Psi^{-1})^{-1}$. Finally, σ^2 is sampled from its full conditional, consisting of an inverse gamma, $IG(\sigma^2|n/2 + a_\sigma, \mathrm{tr}(\Sigma^{-1}\boldsymbol{\delta}\boldsymbol{\delta}') + b_\sigma)$.

3 Results

We run the MCMC for 100,000 iterations with a burn in of 10,000 iterations as coded in FORTRAN95. 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.

quantile	σ^2	ϕ_1	ϕ_2	ϕ_3	ν_1	ν_2	ν_3
2.5%	8.29×10^{-5}	10.089	12.317	0.908	0.102	0.131	0.158
50%	1.00×10^{-4}	16.145	16.219	1.142	0.126	0.156	0.199
97.5%	1.23×10^{-4}	29.821	22.456	1.482	0.150	0.178	0.246

Table 4: Summary of the posterior distributions of the covariance parameters for TP1 and PP2.

	Baseline	Climate parameters			Lat.	Decade
quantile	β_0	β_1 ($\mathcal{K}_v$)	β_2 ($\mathcal{S}$)	β_3 ($\mathcal{F}_{aer}$)	β_4	β_5
2.5%	0.021	0.012	0.007	0.003	0.002	0.048
50%	0.034	0.021	0.009	0.012	0.005	0.063
97.5%	0.046	0.029	0.012	0.021	0.008	0.077

Table 5: Summary of the posterior distributions of the components of $\boldsymbol{\beta}$ for TP1 and PP2.

A summary of the posterior distributions of $\sigma^2, \phi_1, \phi_2, \phi_3, \nu_1, \nu_2$ and ν_3 is presented in Table 4 for the run based on TP1 and PP2. The posterior distribution of σ^2 is very close to its prior distribution. A similar behavior was observed when different prior distributions were used, revealing that the data provide little information about σ^2. The posterior distributions for ϕ_1 and ϕ_2 concentrate on values that are substantially larger than the range of the prior distribution. The distributions of ν_1, ν_2 and ν_3 concentrate around values that are much smaller than those indicated by the prior, implying that the data favor high levels of roughness of the Gaussian field. The posterior distributions for the components of $\boldsymbol{\beta}$ are summarized in Table 5.

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 $p(\tilde{Y}|\ldots) = N_{20\times 43}(\boldsymbol{m}^{(i)}, \sigma^{2(i)} W^{(i)} \otimes \Sigma^{(i)})$, 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 $\boldsymbol{W}^{(i)} = V_{22}^{(i)} - V_{21}^{(i)}(V_{11}^{(i)})^{-1}V_{12}^{(i)}$. 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 distributions are pretty close to the MIT2DCM output. So the model is not introducing any predictive bias. The predictive uncertainty is reasonably small and the variability that it introduces in the analysis is accounted for due to the Bayesian nature of the method.

Figure 5 presents the predictive intervals for $\boldsymbol{z}$, given the full matrix $\boldsymbol{Y}$, under the models that use the TP1 and all three different priors for Ψ. This represents the ability for the model to reproduce the observations. We observe that in the first three decades the predictive intervals under PP3 are wider than those under the other priors while this reverses for the last two decades where PP1 is wider than PP2 or PP3 results. The predictive means under

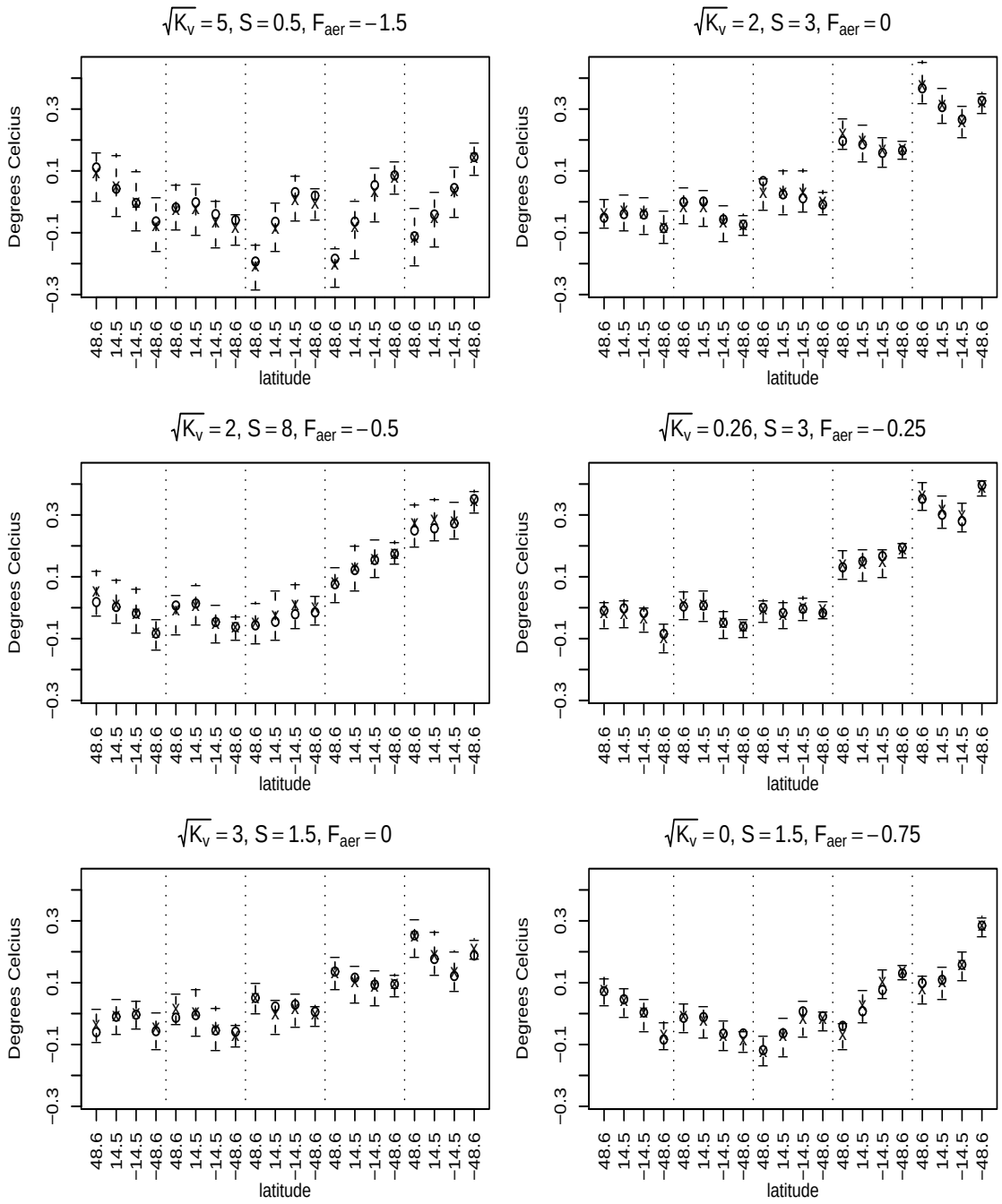

Figure 4: Marginal posterior predictive distributions for six combinations of the climate parameters randomly selected among the 43 used for validation. The dotted lines are used to group the five different decades, with latitude decreasing from left to right within each decade. The extremes of the intervals 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).

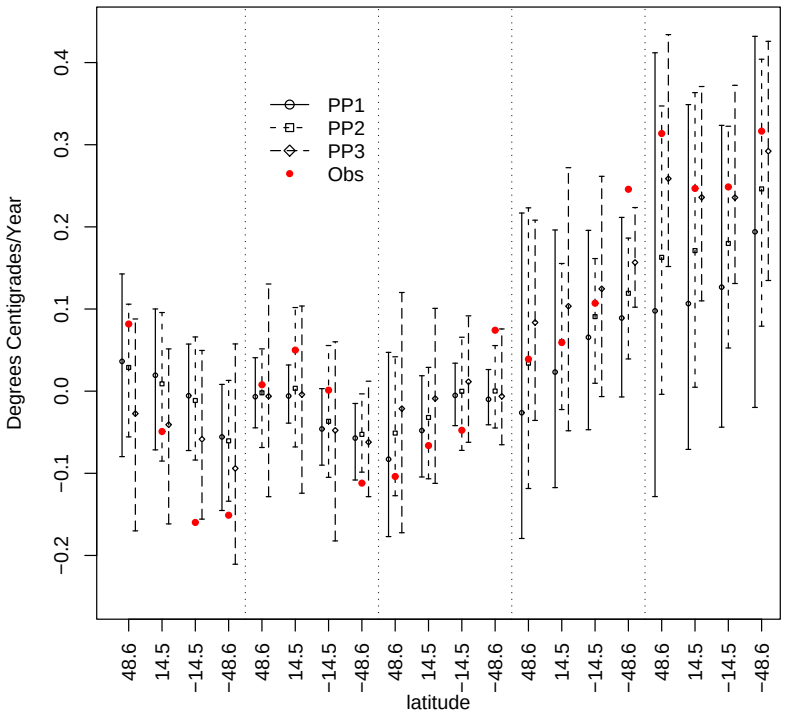

Figure 5: 95% posterior predictive intervals for $\boldsymbol{z}$ using TP1 and the three different priors for Ψ compared to the actual observations. The symbols at the center of the intervals correspond to the predictive means.

PP3 for the first decade have lower values than those under PP1 and PP2, for all latitudes. On Decades 2 and 3 all predictive means are very close. The predictive uncertainty for the last two decades is in most cases larger than in the first three decades. Additionally, the predictive means under each of the three priors are different and they have a tendency to underestimate, that is more pronounced for PP1 than for the other two. A numerical summary of the comparison in Figure 5 is provided by $1/n \sum_{i=1}^{n} (\tilde{z}_i - z_i)^2$, where $\tilde{z}_i$ denotes a sample from the predictive distribution. We calculate this summary for each sample. The results in Table 6 indicate that PP2 and PP3 produce almost equivalent errors. These are on average smaller than those produced under PP1.

Figure 6 shows the posterior marginal distributions of the three climate parameters under

	PP1	PP2	PP3
mean	0.0149	0.0086	0.0079
median	0.0133	0.0077	0.0075

Table 6: Summary of the posterior predictive errors.

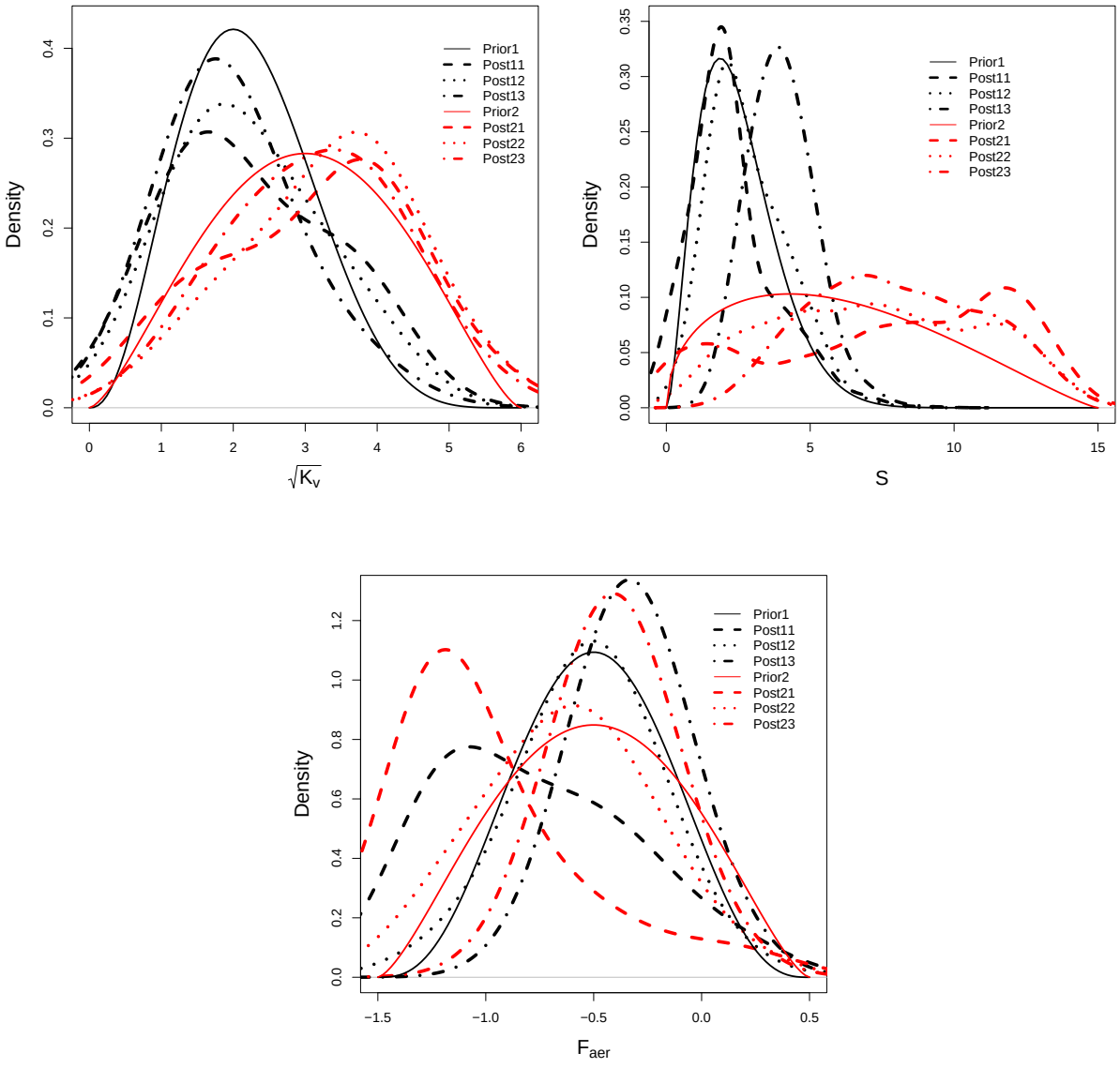

Figure 6: Marginal posterior distributions of the three climate parameters for each of the six combinations of priors for $\boldsymbol{\theta}$ and Ψ. The first index of the labels for the posterior distribution correspond to the prior for $\boldsymbol{\theta}$, the second index to the prior for Ψ.

the two different priors for $\boldsymbol{\theta}$ and the three priors for Ψ. In addition to the six cases illustrated in Figure 6 we considered the case where both Σ and Ψ are given the prior PP3. (Recall that in all cases Σ has used a prior the same as PP3.) The results are almost identical to the ones obtained under PP3 for both TP1 and TP2. We observe that the posterior corresponding to TP1 + PP3 has a similar shape to the ones obtained in Forest et al. (2006). This is not surprising, as PP3 corresponds to the second row in Figure 3, which is the basis to the method used in Forest et al. (2006). The posterior distributions of $\mathcal{F}_{aer}$ corresponding to TP1 + PP1 and TP2 + PP1 assign large masses to unrealistically low values of $\mathcal{F}_{aer}$ and, in the second case a poorly defined distribution for $\mathcal{S}$ (see densities marked with dashed lines in Figure 6) . Together, these indicate that a non-informative prior for Ψ is not appropriate.

Figure 7 illustrates the effect of the three priors for Ψ on the posterior distribution of the eigenvalues of Σ and Ψ. We observe that the eigenvalues of Σ show small changes when the different priors for Ψ are considered. Furthermore, the posterior distribution of its eigenvalues is driven by the eigenvalues of the prior mean, S. Under a non-informative or mildly informative prior (i.e. PP1 or PP2), the posterior for Ψ has only two dominant eigenvalues. The second row corresponds to the case where Σ and Ψ have the same prior, nevertheless, their posterior distributions are remarkably different. Only for an extremely informative prior on Ψ we obtain similar posteriors for both matrices. This indicates that there is little support in the data for the assumption that the different errors have covariance structures equal to the control runs. Another view is that information in ζ and $\boldsymbol{Y}$ has the most influence on Ψ, not Σ. This is not surprising given that $\boldsymbol{Y}$ is the only data depending on $\boldsymbol{\theta}$ which is ultimately driving the sampling.

4 Discussion

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. We build a statistical model based on a Gaussian process that is used as a surrogate for the computer model. The estimation of the surrogate model parameters is performed jointly with the estimation of the computer model parameters, as opposed to the approach commonly used. As we use a Bayesian method, the uncertainty about the calibration parameters is expressed probabilistically and estimation uncertainties are accounted for.

As indicated in the introduction, the estimation of Σ is an interesting problem of its own. This is usually referred to in the climate literature as the problem of estimating the natural climate variability. The posterior distribution of Σ indicates that the four largest eigenvalues of Σ are more separated than the values obtained from the empirical covariance matrix based on the control runs. In particular, the largest eigenvalue is likely to be 40% larger than the largest eigenvalue of S. An interesting question is how to use samples from different GCMs to obtain a pooled estimate of the natural variability.

The differences between the posteriors for Σ and Ψ indicate that our proposed model is able to separate the variability due to the Gaussian process to that due to model discrepancy. This is very reassuring when we consider that the covariance matrices have significantly

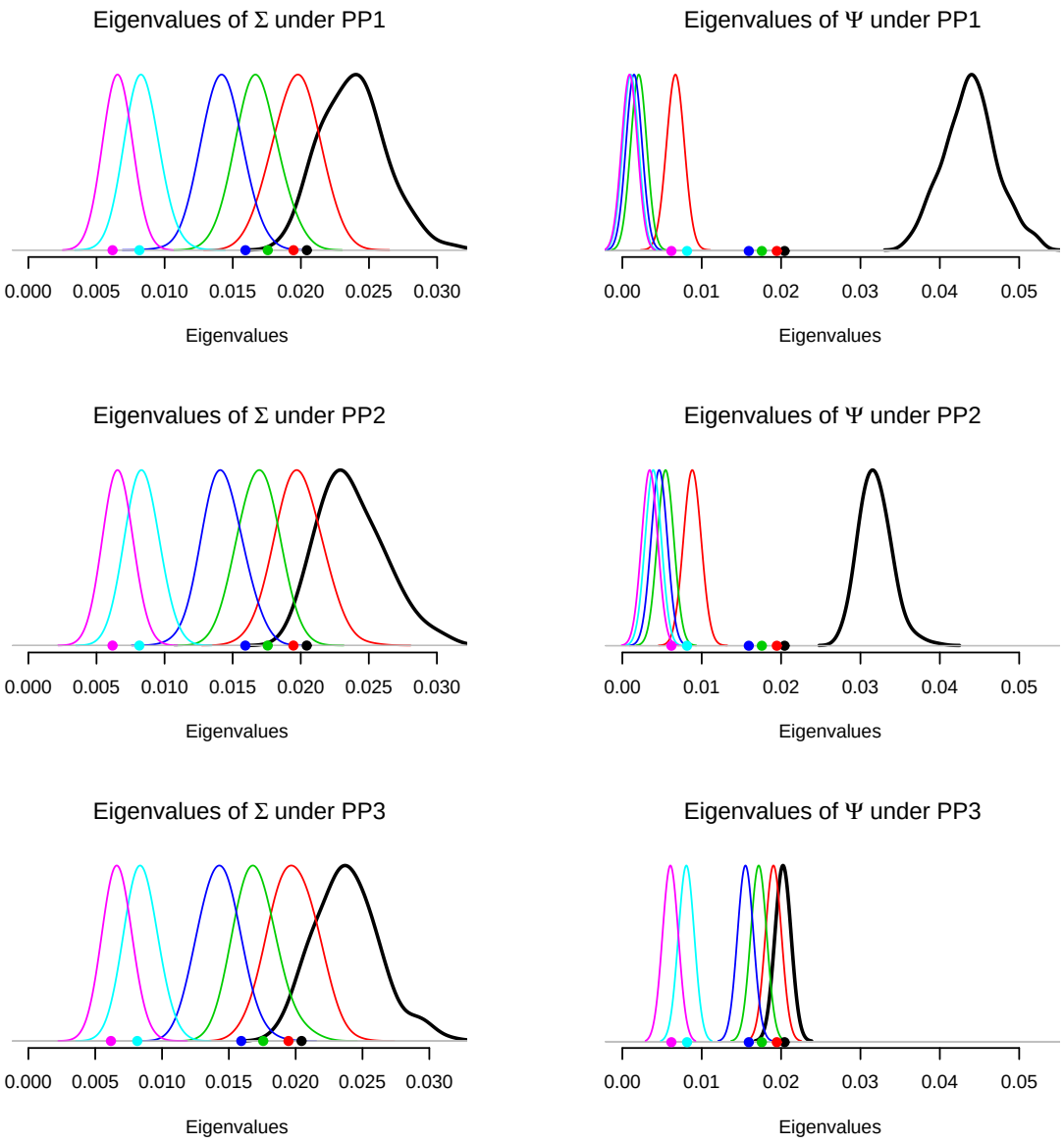

Figure 7: Posterior distribution of the six largest eigenvalues of Σ and Ψ. Left column corresponds to Σ and right column to Ψ. The rows correspond to the three different priors for Ψ. The curves have been rescaled so that all have the same height. The dots correspond to the eigenvalues of S.

	TP1 + PP2			TP2 + PP2			TP1 + PP3		
	$\mathcal{S}$	$\sqrt{\mathcal{K}_v}$	$\mathcal{F}_{aer}$	$\mathcal{S}$	$\sqrt{\mathcal{K}_v}$	$\mathcal{F}_{aer}$	$\mathcal{S}$	$\sqrt{\mathcal{K}_v}$	$\mathcal{F}_{aer}$
5%	1.1	0.9	-1.08	1.9	1.3	-1.22	2.29	0.79	-0.59
50%	2.5	2.2	-0.52	7.2	3.5	-0.61	3.95	1.74	-0.34
95%	5.2	3.9	-0.11	13.9	5.2	-0.05	6.10	3.66	-0.06

	Forest et al. (2006)			Sansó et al. (2008)		
	$\mathcal{S}$	$\sqrt{\mathcal{K}_v}$	$\mathcal{F}_{aer}$	$\mathcal{S}$	$\sqrt{\mathcal{K}_v}$	$\mathcal{F}_{aer}$
5%	1.9	0.2	-0.69	1.75	0.82	-0.71
50%	2.8	0.73	-0.43	3.21	1.99	-0.39
95%	4.7	1.90	-0.14	5.68	3.08	-0.05

Table 7: Summary of the posterior distributions of the climate system properties and comparison to previous results.

different roles in the calibration algorithm. From Figure 7 it is clear that the the prior PP2 has the effect of shrinking the eigenvalues, particularly the largest one. Our results show that such shrinkage has a strong effect on the posterior distribution of the climate parameters. A method commonly used for the analysis of multivariate computer output is to reduce the dimension of the problem by projecting the output on some of the components of an orthogonal basis (see, for example, Higdon et al., 2008). A principal component analysis based on a covariance matrix estimated empirically from the sample is a popular way to obtain such an orthogonal basis. As illustrated by the posteriors obtained using PP1, this must be done with caution, as it is likely that inflated estimates of the eigenvalues will have an effect in the distribution of the calibration parameters.

For the climate sciences it is useful to consider what the surface temperature alone indicates for the posterior distribution of the climate parameters. Table 7 reports the 5%, 50%, 95% quantiles of the marginal posterior distributions from the current model, the model in Forest et al. (2006) and that in Sansó et al. (2008). In addition to already mentioned differences, Forest et al. (2006) used two more diagnostics, one for the deep ocean temperature trend and another for upper air temperatures. We observed substantial differences between the four models. Some are due to the priors, as already discussed, some to the fact that additional summaries of the MIT2DCM were used. Most importantly for this paper, it is clear that assessing the variability in the estimation of the covariance matrix of the errors has a large impact and will require further analysis.

In this paper we have assumed that dependence on $\boldsymbol{\theta}$ 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 $\boldsymbol{\theta}$. 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 $\boldsymbol{\theta}$ 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 $\boldsymbol{\theta}$ values would include some such information but there are a limited number available.

We recall that the MIT2DCM represents a class of climate models for which the climate system properties, $\boldsymbol{\theta}$, can be specified by single variables. This differs from more complex

GCMs 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.

Bayarri, M., Berger, J., Paulo, R., Sacks, J., Cafeo, J., Cavendish, J., Lin, C. and Tu, J. (2007) A framework for validation of computer models. *Technomnetrics*, **49**, 138–154.

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.

Fang, K., Li, R. and Sudjianto, A. (2006) *Design and modeling for computer experiments.* Chapman and Hall.

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.

Higdon, D., Gattiker, J., Williams, B. and Rightley, M. (2008) Computer model calibration using high dimensional output. *Journal of the American Statistical Association*, **103**, 570–583.

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.

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.

Sansó, B., Forest, C. and Zantedeschi, D. (2008) Inferring climate system properties using a computer model (with discussion). *Bayesian Analysis*, **03**, 1–62.

Santner, T., Williams, B. and Notz, W. (2003) *The Design and Analysis of Computer Experiments*. Springer-Verlag.

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.

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, A*. To appear.

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.