Inference for Climate System Properties

Charles T. Curry*, Bruno Sansó* and Chris E. Forest†
[ctc,bruno]@ams.ucsc.edu, ceforest@mit.edu

Abstract:

Modern climate system models make use of numerous parameters to control their behavior. By matching model outputs with observed data we can perform inference over such parameters. Our work extends previous related analyses through the application of model interpolation, Markov-chain Monte Carlo simulation and model selection. The analysis is focused on three input parameters: climate sensitivity, rate of deep ocean heat uptake, and net aerosol forcing. Climate model output was sampled on a non-uniform grid over the 3D parameter space. We interpolate the climate model output over the empty regions of these grids, providing an approximation to the continuous climate model response for use in a likelihood function. The covariance matrix that is required by the likelihood is estimated from general circulation model output. Several model selection criteria are applied to estimate the number of non zero eigenvalues of such matrix. We find significant posterior uncertainty in all three climate model properties. Posterior parameter distributions differ according to the choice of climate model, likelihood and covariance estimate.

1. Introduction

Recently, the climate change research community summarized some of the key unanswered questions that complicate predictions of future global climate (Watson and the Core Writing Team, 2001). Four of the identified uncertainties include: (i) Climate sensitivity, as measured by the temperature change under a specified forcing; (ii) Climate forcings due to natural effects and anthropogenic aerosols; (iii) Magnitude and character of natural climate variability; (iv) Spatiotemporal patterns of change in climate variables. An honest assessment of such uncertain climate properties is key to the support of any scientific statement about the current state of Earth's climate, and to the construction of forward-looking projections that may be used in policy decisions.

Computer climate models contain parameterizations that allow for the exploration of climate system properties. Usually, for 3D models, climate system properties are adjusted by modifying multiple parameters or parameterizations. In the 2D models, like the one that is considered in this paper, climate system properties are controlled via single parameters. Climate sensitivity, defined as the equilibrium global mean temperature response to a doubling of CO_2 and denoted as $\mathcal{S}$, has been singled out as a critical parameter with extensive uncertainty. Another important uncertain parameter is the deep ocean temperature diffusion rate, controlled by varying a diffusion coefficient: $\mathcal{K}_v$. A third parameter to be considered is the net anthropogenic aerosol and unmodeled forcings, written here as $\mathcal{F}_{aer}$.

The MIT 2D climate model provides simulations of ocean, surface and upper atmosphere behavior. It operates on latitude zonal bands as small as 4° latitude. 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).

Forest et al. (2000, 2001, 2002, 2005) run the MIT 2D climate model for many choices of the uncertain parameters $\mathcal{S}, \mathcal{K}_v$ and $\mathcal{F}_{aer}$, selected systematically on a non-uniform grid. They compare the climate model response with the observed climate record using several "diagnostics". These are quadratic forms measuring the difference between the model response and the climate record. They depend on appropriate covariance matrices. Three diagnostics are considered. They are based on deep ocean temperature trend (Levitus et al., 2000), surface temperature time series (Jones et al., 1999) and upper air temperature anomalies (Parker et al., 1997). The present paper builds on these analyses.

We interpolate the climate model outputs over the unsampled regions of the input parameter grid. The interpolator provides a continuous approximation to the climate model response for use in a likelihood function, but also adds an additional source of error. Markov Chain Monte Carlo (MCMC) simulation is used with the interpolated model to explore the posterior distribution of the parameters given the data. The samples from the posterior distribution allow data-based inference about $\mathcal{S}, \mathcal{K}_v$ and $\mathcal{F}_{aer}$ as mod-

*University of California, Santa Cruz
†Massachusetts Institute of Technology

eled by the MIT 2D climate model. This approach is in line with the one proposed in O'Hagan et al. (1999); Kennedy and O'Hagan (2001).

We estimate the covariance among the elements of each of the multivariate diagnostics from "unforced" model control runs. These runs represent the behavior of a climate model in a dynamic equilibrium, spatiotemporal patterns of diagnostic change, variance due to observation error, natural climate model variability and interpolator error. The control run length is insufficient to properly estimate all of the covariance matrix entries, so, following an approach similar to that of optimal fingerprinting (Allen and Tett, 1999), we select a subset of eigenvalues and orthogonal components. We tackle the problem of selecting the significant eigenvalues using model selection criteria.

2. Methods

2.1 Model diagnostic distributions

Both the observations and model outputs were constructed using extensive weighted averaging of variables. Let $\theta = (\mathcal{S}, \mathcal{K}_v, \mathcal{F}_{aer})^T$, $T(\theta)$ be the p-variate climate model output for the chosen diagnostic, and T_{obs} be the observed value of the p-variate diagnostic. Averaging over system variability is expected to produce a normal error distribution around the actual underlying state of the climate system. For the multivariate diagnostics (surface and upper air), the normal errors are expected to be correlated according to some spatiotemporal dependency structure. We denote it as Σ.

Given θ and Σ, we expect that the observations and the model outputs vary normally:

$$T_{obs} - T(\theta)|\theta, \Sigma \sim \mathcal{N}_p(0, 2\Sigma) \quad (1)$$

where $\mathcal{N}_p$ denotes the multivariate normal distribution on $\mathbb{R}^p$. For the deep ocean temperature trend $p = 1$, for the surface temperature $p = 20$ (4×5), and for the upper air temperatures $p = 218$ ($36 \times 8 - 70$ missing observations). We then have that

$$\left(T_{obs} - T(\theta)\right)^T \Sigma^{-1} \left(T_{obs} - T(\theta)\right) |\Sigma, \theta \sim 2\chi^2_p \quad (2)$$

where χ^2_p is the chi-square distribution with p degrees of freedom.

For each of the observation-based diagnostics, we have only a single data point. So, we can not perform effective inference on the components of Σ, which contains $p(p+1)/2$ unknowns using only the observational data. Instead, we construct a point estimate, C_N from a long control run of a GCM which is assumed to reflect the unforced natural variability (Section 2.4). It is customary in climate analysis to retain only the eigenvalues that provide substantial information about the spatio-temporal signal in the data. We write $C_N(\kappa)$ to mean the point estimate of Σ with rank κ.

The analysis in Forest et al. (2002) is based on the assumption that:

$$(T_{obs} - T_\theta)^T C_N(\kappa)^{-1} (T_{obs} - T_\theta) \sim F_{\kappa,\nu-\kappa} \quad (3)$$

where ν is the number of independent samples from the control run used to estimate Σ. We considered this likelihood in our analysis, although to our knowledge, Equation (3) is only valid asymptotically in ν.

We present some results for both the likelihood in (1) and (3). Results for (2) are not shown here, as they are very similar to those of the normal likelihood (1). In the normal and χ^2 cases we assume that Σ is known and equal to $C_N(\kappa)$ and, for the likelihood in (2), adjust the degrees of freedom accordingly. For each value of κ we have a different model for the same data, so the determination of the most likely value of κ can be treated as a model comparison problem.

2.2 Interpolation of model outputs

Despite the acceleration provided by the MIT 2D climate model simplifications, a full model run given a specific set of values for $\mathcal{S}$, $\mathcal{K}_v$ and $\mathcal{F}_{aer}$ takes several hours of CPU on a standard workstation. Since using a MCMC method requires thousands of evaluations of $T(\theta)$, direct evaluation of the likelihood has not been feasible.

Forest et al. (2002) evaluated $T(\theta)$ using the MIT 2D climate model at a number of points on a non-uniform grid over the range of values for $\mathcal{S}$, $\mathcal{K}_v$ and $\mathcal{F}_{aer}$. We estimate $T(\theta)$ for values of θ between those grid values with an interpolator, say, $\hat{T}(\theta)$. We use a thin-plate spline (TPS) (Wahba, 1990) over the sampled θ, after scaling to [0,1] in each dimension.

A TPS uses all available input samples, weighting each datum according to the basis function (Wahba, 1990; Kent and Mardia, 1994). The weighting is calculated in a manner that balances the smoothness of the response function with the error between the interpolated response and the available samples. A smoothing parameter, say, $\lambda_i, i = 1, \ldots, p$, is required for each dimension of the response. We choose these parameters to minimize the generalized cross-validation errors. If $\lambda_i = 0, \forall i$, then the interpolated response function goes exactly through the sample responses, but may "overshoot" the true function between samples. As $\lambda \to \infty$, the interpolated response tends toward the least-squares mul-

tivariate regression of the response with respect to the parameters.

We assume that the interpolation is an unbiased estimator of the response $T(\theta)$. We then have that the likelihood in (1) can be modified to

$$T_{obs} - \hat{T}(\theta) \sim \mathcal{N}(0, 2\Sigma + \Sigma_{interp}(\theta))$$

$\Sigma_{interp}(\theta)$ is obtained by noticing that the interpolator the interpolation is a weighted linear combination of the model outputs. Thus the error is an expansion of the underlying covariance Σ coupled with an estimate of the uncertainty in the choice of smoothing.

2.3 Prior distributions for the parameters

All climate models have built-in range restrictions for their input parameters. Physical constraints and understanding of the Earth's atmospheric and oceanic composition impose additional limitations on what values the parameters might take. Prior attempts to fit climate model outputs to data or to match other climate models further hint at "good" values for the parameters. We compute posterior distributions for the parameters using two different priors: a uniform prior over the range of parameter values for which we have computed model outputs (to avoid extrapolation), and a set of more informative scaled beta distributions developed in Webster and Sokolov (2000). The expert prior can have a very strong influence on the results. For example, Forest et al. (2002) found that $\mathcal{K}_v$ was not well constrained by the data. High values of $\mathcal{K}_v$ were likely, but have physical implications that are not realistic. This can be accounted for by the prior. Nevertheless, in this paper we only report posteriors arising from the uniform priors.

2.4 Estimation of the covariance

In order to estimate the covariance matrix for the diagnostics we considered vectors extracted from a long global climate model control run (GFDL model). We were able to obtain 28 such consecutive vectors of size 36×8 for the upper air diagnostic and 95 overlapping vectors of size 4×5 for the surface temperatures. These limited data sets produce estimates C_N that are ill-conditioned. We use singular value decomposition of C_N to obtain its generalized inverse . We write the decomposition: $C_N = Q\Lambda Q^T$, where Q is an orthogonal matrix, Λ is diagonal and $\Lambda_{ii} \geq \Lambda_{jj}, i > j$. Each eigenvalue, Λ_{ii}, estimates the control run variance under projection onto the corresponding eigenvector $Q_{\cdot i}$.

Λ_{ii} is a biased estimate of the corresponding eigenvalue of Σ (Muirhead, 1982). The largest eigenvalues are inflated and the smallest are deflated. Our control runs do not contain enough independent samples to overcome such biases. Even if we did have access to sufficiently-long control runs to ignore the bias, climate modelers assume that the small eigenvalues represent low variance noise components. Unfortunately, the low variance in the covariance estimate implies that we know those components to high precision, an artifact of under-representing our uncertainty in the covariance structure. As a result, the smallest eigenvalues have the largest effect on the likelihood. Fitting these noise components blurs the climate change signal that we are trying to detect. To eliminate the noise, we zero out the smallest eigenvalues.

Let κ be the number of non-zero eigenvalues, then $C_N(\kappa)^{-1} = Q\tilde{\Lambda}^{-1}Q^T$ where $\tilde{\Lambda}_{ii}^{-1} = 1/\Lambda_{ii}, i \leq \kappa$ and $\tilde{\Lambda}_{ii}^{-1} = 0, i > \kappa$. This generalized inverse is used in the computation of the likelihood in (3). It is also used in Eqs. (1) and (2) as a "plug in" estimate of Σ^{-1}. The likelihood decreases monotonically with increasing κ, as high values of κ include components that show unrealistically low variance in the control samples (Allen and Tett, 1999).

Figure 1 shows some common methods for selecting the number of usable eigenvalues for the surface diagnostic control run. (Mardia et al., 1979). Note that Kaiser's rule selects smaller values of κ than a criterion that attempts to maintain a proportion of the total variance. Visual analysis of such a screeplot is typically more complex. One might attempt to find the value of κ after which the graph "flattens out." For the surface diagnostic, choosing either $\kappa < 13$ or $\kappa < 17$ might satisfy this criterion.

We consider several methods to find an appropriate rank. Traditional model selection criteria, like the Akaike Information Criterion (Akaike, 1973) and Schwartz' Bayesian Information Criterion (Schwarz, 1978), break down for this problem, as the maximum likelihood is a monotonically-decreasing function of the rank. The model size penalty further exacerbates this effect. According to these criteria, the model with one diagnostic dimension is the best model. The break-down is likely caused because these criteria have established meaning only in the limiting case of many independent data points, which are not available for this application.

3. Results

Model output came from two different configurations of the MIT 2D climate model. We denote the two sets of output GSO and GSOLSV. GSO refers to a model that included changes in concentrations of greenhouse gases, sulfate aerosols, and

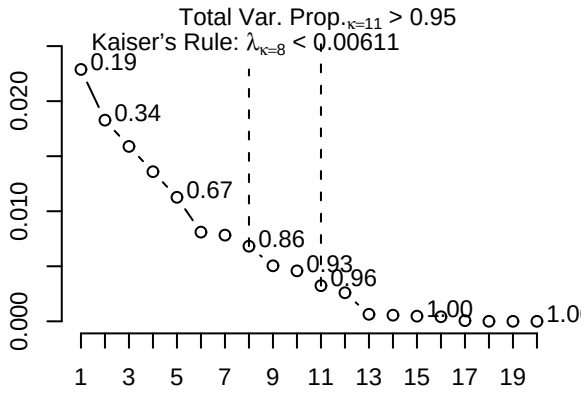

Figure 1: Screeplot of the surface diagnostic covariance eigenvalues with two selection criteria annotated. Fractions above each point i are the proportion of the total variance represented by considering all eigenvalues up to $\kappa = i$. Kaiser's rule consists of comparing the eigenvalues against the mean.

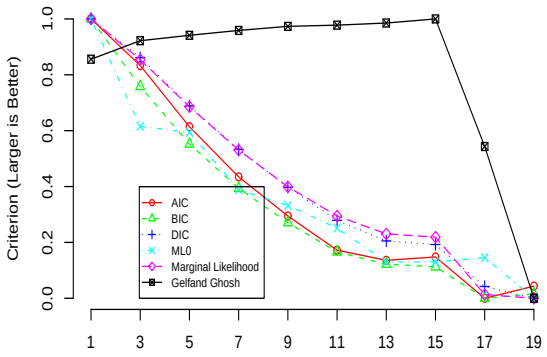

Figure 2: Model selection criteria calculated for κ given the surface diagnostic, the GSOLSV model, and the F likelihood. Note that most of the criteria are monotonically decreasing, as is the maximum likelihood.

ozone. GSOLSV added to the former, changes in land-use vegetation, solar irradiance, and stratospheric aerosols resulting from volcanic eruptions (Forest et al., 2005). The GSO model was interpolated from 306 runs on a bounded non-uniform parameter space grid. The GSOLSV model was interpolated from 499 points in a slightly-larger parameter space. We report primarily GSOLSV results, as the GSO results are similar to those published in Forest et al. (2002).

Reported results were produced from MCMC traces of 100,000 samples. In all cases, only the latter half of the samples were kept, to avoid startup biases. Proposed parameter samples were drawn from normal proposal distributions truncated by the uniform prior bounds. Mixing, as measured qualitatively by visualization of the traces, was good. We considered all three likelihoods using a covariance matrix that accounted for the interpolation error.

Reporting results is complicated by the lack of a clear choice for κ, the number of orthogonal components kept in the error covariance. As mentioned above, AIC and BIC suggest maintaining only the first principal component. The Deviance Information Criterion (Spiegelhalter et al., 2002), computed from the sampled posterior, produced similar results. Chib and Jeliazhov (2001) provide a way to estimate the probabilities of the data given the model that should not depend directly on the quantity of available data. These "marginal likelihoods" provide support only for low values of κ. Only the posterior-predictive loss-based criterion proposed in Gelfand and Ghosh (1998) supported a choice of $\kappa > 1$. The behavior of each criterion (after rescaling for comparability) can be seen in Figure 2 (using the GSOLSV model, surface diagnostic and F likelihood). Under the normal and chi-square likelihoods, all criteria (including the posterior-predictive loss) decreased monotonically.

In some cases, posteriors obtained using only the first error component ($\kappa = 1$) were poorly constrained, yielding posterior credible intervals that nearly covered the entire allowable parameter space (Figure 3, surface diagnostic with $\kappa = 1$). Such posteriors would not yield valuable insight into climate sensitivity, deep ocean heat uptake or aerosol forcing. Those constructed using more components (Figure 3, surface, upper air and joint diagnostics with $\kappa = 16$) were more constrained and more in line with previously-published results, but justification of such choices of κ is difficult. Uncertainty in κ is not represented in these posterior results.

Figure 4 shows marginal posteriors of the parameters conditional on a range of κ choices. For all of these marginals, the posterior narrows with increasing κ. Note that in some cases, including additional error components has little effect on the posterior (climate sensitivity and rate of deep ocean heat uptake with the surface diagnostic). In other cases, there is a noticeable shift once a certain number of components are included, suggesting one of the low-variance components has a significant impact on the posterior (climate sensitivity and aerosol forcing with the upper air diagnostic).

4. Discussion

We have performed a careful analysis of the problem of estimating $\mathcal{S}$, $\mathcal{K}_v$ and $\mathcal{F}_{aer}$ using the MIT 2D climate model. We obtained posterior distribu-

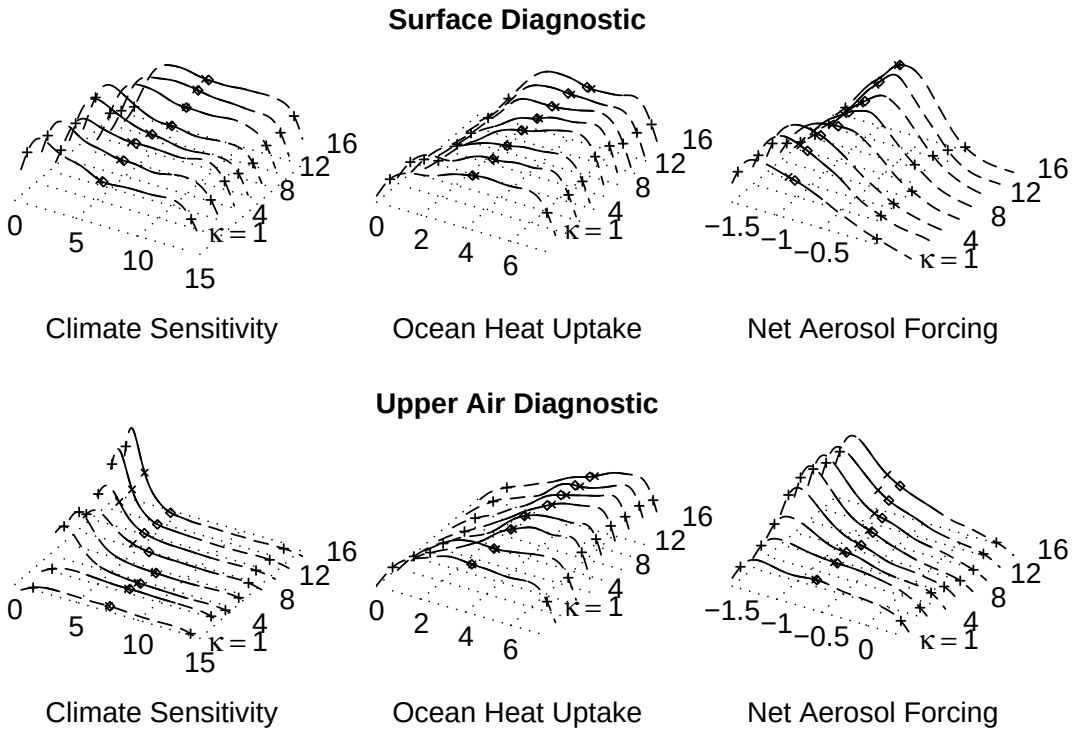

Figure 4: Marginal posterior densities of $\mathcal{S}$, $\mathcal{K}_v$ and $\mathcal{F}_{aer}$ as a function of κ. Note that the marginal posteriors tend to be stable over ranges of κ, and then change shape when the number of included components crosses some threshold. Samples were taken with from the MCMC using the specified diagnostic (surface, upper air), GSOLSV model and normal likelihood.

tions for all three climate parameters using three different likelihoods, three different diagnostics and considering different priors. Our methods account for the error introduced by a TPS interpolator. We consider the estimation of the number of significant eigenvalues using an array of model selection criteria.

Observation error, while available for some of the measurements, was not accounted for in the likelihood (except in the case of uncertainty in the deep ocean temperature trend coefficient). Serious issues with data quality make modeling this important. The interdependency between diagnostics was not modeled. Assuming the diagnostics are independent does not correctly weight their relative importance.

The choice of interpolator is not automatic. Other interpolation processes with more explicit error models may provide better approximations to the climate model outputs.

Many asymptotic and Monte Carlo integration-based criteria failed to automatically select a non-trivial subset of these eigenvalues (κ). This may imply that only the first few components should be used, as the higher components represent only low-variance noise. A better assessment of the significance of the eigenvalues requires the incorporation of the uncertainty in the covariance. A Bayesian approach with explicit structural assumptions will be needed. All these issues are currently being considered by the authors.

References

Akaike (1973) Information theory and an extension of the maximum likelihood principe. *2nd International Symposium on Information Theory,Akademia Kiad: Budapest*, 267–281.

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

Chib, S. and Jeliazhov, I. (2001) Marginal likelihood from the metropolis-hastings output. *Journal of the AMerican Statistical Association*, **96**, 270–281.

Forest, C., Stone, P. and Sokolov, A. (2005) Estimated PDFs of climate system properties includ-

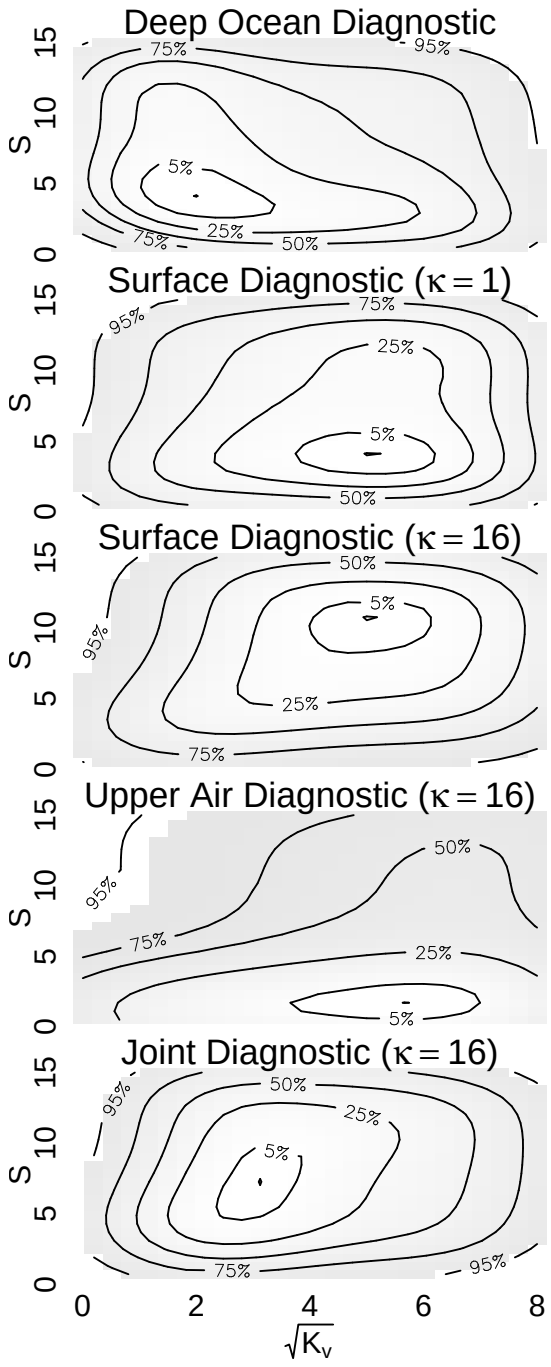

Figure 3: Joint posteriors of climate sensitivity versus rate of deep ocean heat uptake for selected diagnostics and values of κ. Contour regions correspond to posterior credible intervals for the parameters from the MCMC using the named diagnostic (deep ocean, surface, upper air, joint), the GSOLSV model and the normal likelihood.

ing natural and antrophogenic forcings. *Tech. rep.* Submitted to GRL.

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., 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.

Gelfand, A. E. and Ghosh, S. K. (1998) Model choice: A minimum posterior predictive loss approach. *Biometrika*, **85**, 1–11.

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.

Kent, J. T. and Mardia, K. V. (1994) The link between kriging and thin-plate splines. In *Probability, Statistics and Optimization* (ed. F. Kelley). New York: Wiley.

Levitus, S., Antonov, J., Boyer, T. P. and Stephens, C. (2000) Warming of the world ocean. *Science*, **287**, 2225–2229.

Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979) *Multivariate Analysis*. London: Academic Press.

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.

Schwarz, G. (1978) Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464.

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

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002) Bayesian measures of model complexity and fit (Pkg: p583-639). *Journal of the Royal Statistical Society, Series B, Methodological*, **64**, 583–616.

Wahba, G. (1990) *Spline Models for Observational Data*. CBMS-NSF.

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.