# Multivariate Bayesian Analysis of Atmosphere–Ocean General Circulation Models

Reinhard Furrer,[1] Stephan R. Sain,[2] Douglas Nychka,[1]
Claudia Tebaldi,[1] and Gerald A. Meehl[1]

Numerical experiments based on atmosphere–ocean general circulation models (AOGCMs) are one of the primary tools in deriving projections for future climate change. Although each AOGCM has the same underlying partial differential equations, modelling large scale effects, they have different small scale parameterisations and different discretisations to solve the equations, resulting in different climate projections. This motivates climate projections synthesized from results of several AOGCMs' output. We combine present day observations, present day and future climate projections in a single hierarchical Bayes model. The challenging aspect is the modeling of the spatial processes on the sphere and the amount of data involved. We pursue a Bayesian hierarchical model that separates the spatial response into a large scale climate change signal and an isotropic process representing small scale variability among AOGCMs. Samples from the posterior distributions are obtained with computer-intensive MCMC simulations. The novelty of our approach is that we use gridded, high resolution data within a spatial hierarchical framework. The primary data source is provided by the Coupled Model Intercomparison Project (CMIP) and consists of 9 AOGCMs on a 2.8 by 2.8 degree grid under several different emission scenarios. In this article we consider mean seasonal surface temperature and precipitation as climate variables. Extensions for our model are also discussed.

*Keywords: Climate change; Spatial process; Spherical covariance, Hierarchical model; Large datasets; MCMC.*

## 1   Introduction

The influence of human activities on the Earth's climate is by now largely undisputed and the potential for even greater changes in this century confront us. A grand challenge facing the geosciences is to provide accurate predictions of these changes along with quantifications of uncertainties. Because future climate may be very different from the observational record a primary tool for assessing changes are large computer models that simulate the Earth's climate system under different circumstances. The results of such models are complex spatial fields and the use of statistical analysis is particularly useful for synthesizing the information from several models and providing statistical measures of uncertainty. The application of conventional statistics here is interesting because the models themselves are deterministic computer codes but the variation and biases among different models can fit into a probabilistic framework. Accordingly we will refer to the model output as "data" even though it may not fit the conventional perception of a statistical sample.

### 1.1   Climate change assessment

The impact of climate change can be persuasive from disrupting ecosystems to effecting economies to influencing public health. Significantly, each of these areas requires the analysis of climate at a regional spatial scale or at even smaller areas and part of the goal of current climate models is to provide such information. Although small spatial scales are the most useful for determining specific impacts these are precisely the scales where climate simulation is difficult and model biases increase. Thus, any approach to interpret climate projections at regional scales should include a quantification of the uncertainty in the estimated climate. The number of cutting edge climate system models is limited, however, and it is ironic that despite the voluminous spatial output for a given model the sample size for comparison across different

---

[1] National Center for Atmospheric Research, Boulder, CO 80307–3000; Corresponding e-mail: `furrerucar.edu`.
[2] University of Colorado at Denver, Denver, CO 80217–3364.

models is small. This work addresses this problem by spatial models that borrow strength across adjacent spatial regions and provide a more statistically accurate assessment of model bias and variability. Based on a statistical framework for the response of a suite of climate models it is possible to produce a synthesized estimate of climate change at a regional scale along with a measure of the uncertainty.

Assessing probabilistic climate change is not a new topic. Recent work in determining probabilities of global temperature change includes Allen *et al.* (2000), Schneider (2001), Wigley and Raper (2001), Forest *et al.* (2002), Gregory *et al.* (2002) Knutti *et al.* (2002, 2003), Frame *et al.* (2005). Tebaldi *et al.* (2005) evaluates probabilistic climate change on a regional level.

## 1.2   A random effects statistical model

As part of this introduction we give a overview of the main points of our statistical model for combining climate model experiments. Given experiments from $N$ climate models, for the $i^{th}$ model let $\mathbf{X}_i$ denote a vector of average temperatures representing current climate at a grid of points on the surface of the Earth. Let $\mathbf{Y}_i$ be the corresponding averages simulated at a future period under a specific scenario of climate change. Setting $\mathbf{D}_i = \mathbf{Y}_i - \mathbf{X}_i$ we are lead to the statistical model

$$\mathbf{D}_i = \mathbf{M}\boldsymbol{\theta}_i + \boldsymbol{\varepsilon}_i,$$

where $\mathbf{M}$ is a matrix of spatial basis functions and $\boldsymbol{\theta}_i$ are a vector of coefficients. We identify this regression function with the difference in the model's climate between present and future conditions. The key assumption is to interpret $\boldsymbol{\theta}_i$ as a random effect that is different for each model but whose expected value is the true difference in climate. The second term $\boldsymbol{\varepsilon}_i$ is assumed to be a mean zero spatial process. Besides being a focused example on interpreting climate model output, the flexibility of this spatial random effects models can have more general application. We believe that the representation of a spatial response into a targeted set of basis function and a more generic stationary model for finer detail is an effective model and is amenable to simple Markov Chain Monte Carlo techniques for an approximate Bayesian analysis.

## 1.3   Outline

This article is structured as follows. In the next section we discuss the model output, what we refer to as "data". Section 3 introduces a simple spatial model for climate change. Section 4 presents the results of our analysis. Extensions to the model and a discussion are given in Sections 5 and 6.

# 2   Climate Model Data

The *climate* at a given location is the joint distribution of meteorological variables describing the atmosphere averaged over a given period of time. In statistical language, given a stationary time series the climate is simply the marginal or stationary distribution. A common working definition of climate is a twenty to thirty year average around a particular time. The temporal variability of meteorology about a climatological mean is termed weather and it is weather that is observed both in the real world and also what is simulated by models. Thus, any analysis for differences in climate must account for the intrinsic variability of weather and the fact that climate can not be determined exactly with a finite sample.

## 2.1   Climate System Models

Climate models attempt to simulate the Earth's climate system including the complex interactions among the ocean, atmosphere, sea ice, biogeochemistry and the land surface. Technically these coupled models are referred as atmosphere–ocean general circulation models (AOGCMs). Large AOGCMs, such as the NCAR Community Climate System Model, involve hundreds of man years of scientific research and software engineering and require months of supercomputer time for numerical experiments. In these models the

motion of the atmosphere and ocean is the result of discretizing the partial differential equations for fluid dynamics. However, there are many geophysical processes that can not be resolved by the model but have important feedbacks to larger scales. These processes are *parameterized* in the models and because they are not derived directly from classical physics are potential sources of model bias. The AOGCMs are advanced on a short time step, typically on the order of minutes and they simulate weather. Over time they take as inputs a suite of variables that are considered external to the climate system. Some examples include greenhouse gases produced by human activities, changes in the sun's energy or changes in land use. These variables that drive the climate are termed *forcings*. Analogous to observational data, the weather produced by the model over a given period of simulation is then averaged to estimate the model's climate. This realized climate will be a function of the different magnitudes and sequences of forcings used in the experiment and so provides a tool for determining the effect of different curbs on greenhouse gas emissions on future climate. AOGCMs are also tested by using a sequence of external forcings that match the previous century or present conditions. In this way the climate produced by an AOGCM can be compared to the climate estimated from observations. However, it is well accepted that the bias of an AOGCM for present climate may not have a strong relationship with the magnitude or sign of model bias for future projections. The exact relationship between present and future model bias is still an open statistical problem.

## 2.2 Numerical Experiments

The different components of the AOGCMs generate a vast array of output. However, the most common atmospheric variables used for assessing the impacts of a changed climate are surface temperatures, precipitation, and to a lesser extent surface winds. These variables are often averaged to monthly or seasonal fields. In this article we focus on average surface temperature and precipitation for the boreal winter (December, January and February) and boreal summer (June, July and August).

We work with a set of 9 AOGCMs based on Coupled Model Intercomparison Project (CMIP) Meehl *et al.* (2000); Covey *et al.* (2000), summarized in Table 1. At present there are approximately a dozen research centers and universities supporting independent AOGCM development and these CMIP experiments represent a subset of these different modeling groups. CMIP began in 1995 by collecting output from model "control runs" in which climate forcing is held constant at present values. These experiments are thus simulations of current climate. Later versions of CMIP have collected output from an idealized scenario of global warming. Here, the atmospheric $CO_2$ is increased from current levels at a steady rate of 1% per year until it doubles at about year 70. The years 60–79 are averaged to estimate the climate at

| Model | Supporting research center |
|---|---|
| CCCM | Canadian Climate Centre, Canada |
| CSIR | Commonwealth Scientific and Industrial Research Organization, Australia |
| CSM | National Center for Atmospheric Research, USA |
| ECHAM | Deutsches Klimarechenzentrum, Germany |
| ECHO | Max Planck Institut für Meteorologie, Germany |
| GFDL | Geophysical Fluid Dynamics Laboratory, USA |
| HADCM | United Kingdom Meteorological Office, England |
| MRI | National Institute for Environmental Studies, Japan |
| PCM | Los Alamos National Laboratory, the Naval Postgraduate School, the US Army Corps of Engineers' Cold Regions Research and Engineering Lab and the National Center for Atmospheric Research, all USA |

Table 1: Summary of CMIP AOGCM models used. A more detailed description of the models and key references can be found at `http://www-pcmdi.llnl.gov/projects/cmip/Table.php`. See also (Houghton *et al.*, 2001).
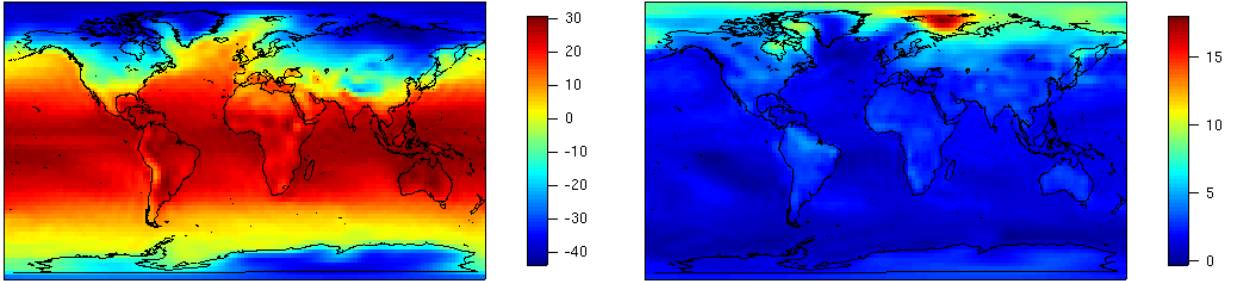
Figure 1: Winter temperature fields for control run (left panel) and transient minus control run (climate change, right panel) for model CSM (Unit: °C).
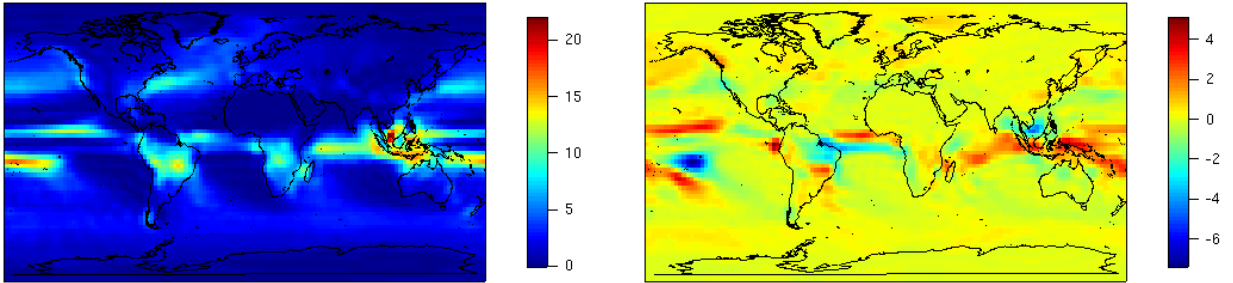


Figure 2: Winter precipitation fields for control run (left panel) and transient minus control run (climate change, right panel) for model CSM (Unit: mm/day).

the end of this period. This experiment is termed a transient run because the changing forcing (*i.e.* the amount of $CO_2$ is not constant). Hence, one would not expect the climate to be constant over time either. However, taking an average of 20 years is the standard format for summarizing the climate around a given time point. The different AOGCMs were run at a spatial resolution of $2.8 \times 2.8$ degrees at the equator. This resolution is converted into a standard gridded output format of $128 \times 64$ equally spaced points in longitude and latitude. Note that this griding convention for the output yields substantially smaller grid cells in area near the poles.

As an example of the model output, Figures 1 and 2 show winter temperature and precipitation fields for control run and transient minus control run (*i.e.* climate change) for model CSM. Compared to temperature fields, the precipitation fields exhibit a different pattern. The spatial structures have smaller scales and the fields are more heterogeneous. Exploratory data analysis suggests that a log transformation corrects for both of these shortcomings. Some models have a few grid cells with zero precipitation and we modified the data of these cells to 0.1mm/day precipitation, which can be justified by the climatological nature of the data and the size of the area of the grid cells.

As a useful summary, Figures 3 and 4 show the mean and standard deviation over all considered CMIP models of boreal winter climate change for the temperature and the log precipitation.

## 3   A Spatial Model for Climate Changes

In this section we present a spatial hierarchical model to synthesize the climate projections based on the data described in the last section. Similar models in slightly different contexts are presented in Wikle *et al.* (1998), Wikle and Cressie (1999) or Berliner *et al.* (2000). Banerjee *et al.* (2004) gives an excellent overview of hierarchical modeling for spatial data.
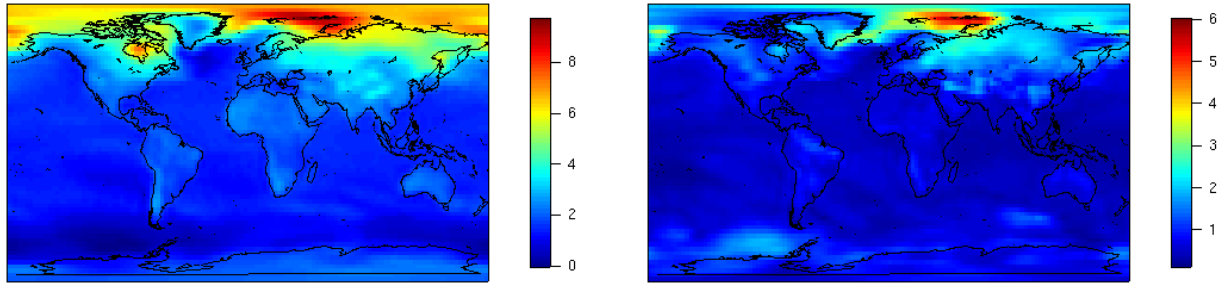
Figure 3: Mean and standard deviation of boreal winter temperature climate change (Unit: °C).
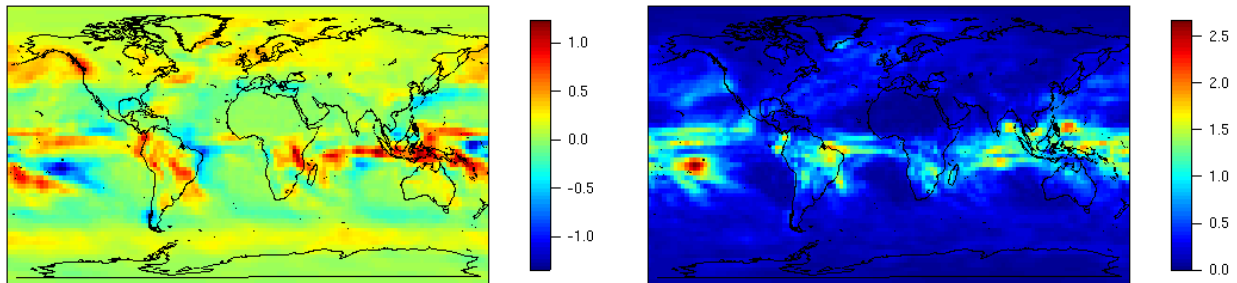


Figure 4: Mean and standard deviation of boreal winter log precipitation climate change (Unit: mm/day).

## 3.1    Simple Hierarchical Model

The temperature or transformed precipitation fields of the $N = 9$ models are stacked into vectors of length $n$ (total number of grid points), denoted as follows:

$$\mathbf{X}_i = \text{climatological field from control run of model } i,$$
$$\mathbf{Y}_i = \text{climatological field from transient run of model } i.$$

The difference is given by

$$\mathbf{D}_i = \mathbf{Y}_i - \mathbf{X}_i, \qquad i = 1, \dots, N,$$

which represents the climate change with respect to temperature or log precipitation. We use a standard hierarchical Bayes approach based on the following three levels: data, process, and priors.

*Data level*: The data level models the quantity of interest $\mathbf{D}_i$, $i = 1, \dots, N$, as a spatial process. We assume that the climate change is an additive decomposition of a large scale climate signal and small scale signals consisting of model bias and internal model variability. The single run CIMP data is not sufficient to model the small scale signals separately and we therefore represent both by a single spatial process. This decomposition is consistent with the more traditional decomposition in spatial modelling (Cressie, 1993; Banerjee *et al.*, 2004) where the mean corresponds to global (first-order) behavior and the error captures local (second-order) behavior. Based on these ideas:

$$\mathbf{D}_i = \boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_i, \qquad [\,\mathbf{D}_i \mid \boldsymbol{\mu}_i, \phi_i\,] \overset{\text{iid}}{\sim} \mathcal{N}_n(\boldsymbol{\mu}_i, \phi_i \boldsymbol{\Sigma}), \qquad \phi_i > 0, \qquad i = 1, \dots, N, \tag{1}$$

where | means "conditioned on" and $\mathcal{N}_n$ is an $n$-dimensional normal density. In our approach, the correlation matrix $\boldsymbol{\Sigma}$ is specified (see Section 3.3) and $\phi_i$ are scale parameters. Model (1) implies that the density of the entire dataset, conditional on $\boldsymbol{\mu}_i$ and $\phi_i$, $i = 1, \dots, N$, is the product of $N$ Gaussian density functions dictated by $\mathcal{N}_n(\boldsymbol{\mu}_i, \phi_i \boldsymbol{\Sigma})$.

*Process level*: The second hierarchical level models the large scale climate signals $\boldsymbol{\mu}_i$, $i = 1, \ldots, N$. We use a dimension reduction technique and assume that $\boldsymbol{\mu}_i = \mathbf{M}\boldsymbol{\theta}_i$, where the given "design" matrix $\mathbf{M}$ contains $p$ basis functions with $p \ll n$. The choice of $\mathbf{M}$ is further discussed in Section 3.2. We denote the "true" large scale climate change pattern as $\mathbf{M}\boldsymbol{\vartheta}$ and the $\boldsymbol{\theta}_i$ are modelled as:

$$[\boldsymbol{\theta}_i \mid \boldsymbol{\vartheta}, \psi_i] \stackrel{\text{iid}}{\sim} \mathcal{N}_p(\boldsymbol{\vartheta}, \psi_i\boldsymbol{\Omega}), \qquad \psi_i > 0, \qquad i = 1, \ldots, N.$$

Centering these coefficients vectors about the true climate one ($\boldsymbol{\vartheta}$) represents our assumption that the sample of climate models do not exhibit systematic large scale errors. However, we do expect departures between each model and the true climate and this variation is captured by the covariance term ($\psi_i\boldsymbol{\Omega}$). In particular the variation of $\psi_i$ across different models reflects different levels of bias and internal variability for a given AOGCM. Although the correlation matrix $\boldsymbol{\Omega}$ might have arbitrary structure, for orthogonal or close to orthogonal basis functions we believe $\boldsymbol{\Omega} = \mathbf{I}$ is a reasonable choice.

*Prior level*: The last level puts priors on the process parameters.

$$[\phi_i] \stackrel{\text{iid}}{\sim} I\Gamma(\xi_1, \xi_2), \qquad \xi_1, \xi_2 > 0, \qquad i = 1, \ldots, N;$$
$$[\psi_i] \stackrel{\text{iid}}{\sim} I\Gamma(\xi_3, \xi_4), \qquad \xi_3, \xi_4 > 0, \qquad i = 1, \ldots, N;$$
$$[\boldsymbol{\vartheta}] \sim \mathcal{N}_p(\mathbf{0}, \xi_5\mathbf{I}), \qquad \xi_5 > 0;$$

where $I\Gamma$ denotes the inverse Gamma distribution and where $\xi_1, \ldots, \xi_5$ are hyperparameters. If we include more basis functions in $\mathbf{M}$ or if the climatological fields are smooth then we expect $\phi_i$ to be small. The scaling parameters $\psi_i$ are related to the internal variability of the model and should also be small. When several model runs or ensembles are available, they can be used as prior information for the magnitude of $\psi_i$. Since we do not have any independent information about the truth $\boldsymbol{\vartheta}$ we choose a large value for $\xi_5$, reflecting an uninformative prior and justifying the identity as its correlation matrix.

## 3.2 Choice of Basis Functions

We now discuss the basis functions used to construct the design matrix $\mathbf{M}$. These functions need to be sufficiently flexible to represent the mean structure of the difference fields, and to achieve this goal we use three different types.

Any real valued random field on a sphere that has finite variance and has realizations which are square integrable over the surface of the sphere may be represented as an infinite series of spherical harmonics (*e.g.* Jones, 1963). The spherical harmonics are a generalisation of a sin-cosine decomposition of a real valued function to the sphere. It is therefore natural to assume that the large scale signal is a linear combination of $p_s$ spherical harmonics. The spherical harmonics can be obtained with an iterative procedure and each additional level has smaller scales. For a given level $\ell$, there are $2\ell + 1$ basis functions. For example, the first spherical harmonics corresponds to the global mean ($\ell = 0$), level $\ell = 1$ consists of three single sin/cosine structures on the sphere and so on. Figure 5 gives four examples from the first, second, third and seventh level.
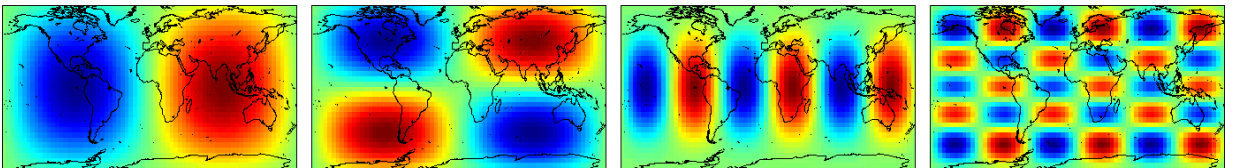


Figure 5: Examples of spherical harmonics. The fields correspond to spherical harmonics from levels $\ell = 1, 2, 3, 7$.

Temperature fields differ over ocean, land or sea ice and the AOGCM climate change fields echos these patterns (see Figure 3). These patterns based on land forms are not easily represented by spherical

harmonics. We therefore also introduce indicator basis functions linked to land forms and sea ice. Figure 6 gives four examples of these indicator functions. Incorporating these indicator functions results in the error processes $\varepsilon_i$ exhibiting more stationary and isotropic behavior, *i.e.* the covariance between two points depends only on the great angle distance of the points and not on their location.
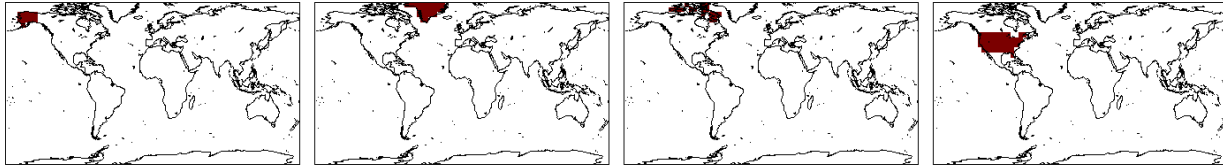


Figure 6: Examples of land indicator basis functions. From left to right Alaska, Greenland, Canadian arctic ice and lower USA.

Finally, temperature or precipitation change is related to the respective field itself (*e.g.* Figure 4). We therefore use a basis function consisting of present climate derived from observed data, provided from the reanalysis of the National Centers for Environmental Prediction (NCEP) (Kalnay *et al.*, 1996).

## 3.3   Models for Correlation Matrices on the Sphere

We will assume that the remaining small scale structure $\varepsilon_i$ in equation (1) is an isotropic and stationary processes on the sphere. Although this is a strong assumption, it is supported on a global scale by exploratory graphical analysis with $\varepsilon_i$ (*e.g.* Cressie, 1993). However, locally a few models exhibit slight non-stationary behaviour in different regions. In Figure 7 we show localized variograms (*e.g.* Cressie, 1993) of the error processes $\varepsilon_i$ of the model ECHAM for the temperature field and the log precipitation field. The differences in the temperature variograms can be traced to localized variability of this AOGCM around Greenland and in the Canadian arctic. For the log-precipitation the residuals in the Intertropical Convergence Zone (ITCZ), a zone of trade wind convergence and excessive precipitation, boosts the variogram for the equatorial zone (see also Figure 13 lower right panel)

Although the variogram is useful in identifying a functional form for the covariance function, the precise specification of a covariance must be derived from considering families of functions that are positive definite when restricted to a spherical domain. We discuss two approaches to the construction a such functions. Further information can be found in Yaglom (1987), Weber and Talkner (1993), Gaspari and Cohn (1999), and Gneiting (1999).

Schoenberg (1942) showed that a function belongs to the class of positive definite functions on the sphere if and only if it can be represented as an infinite series of Legendre polynomials. This representation is only
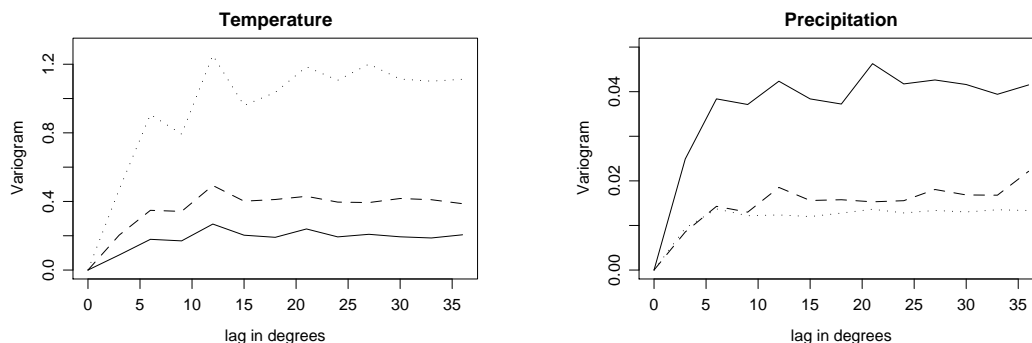


Figure 7: Localized empirical variograms of residual process $\varepsilon_i$ for model ECHAM for three different latitude bands (80 to 30 dotted, 30 to $-30$ solid, $-30$ to $-80$ dashed). We used the robust variogram estimator as given by equation (2.4.13) in Cressie (1993).

useful in practice if we can rewrite the expression in a low-order parameterized form. However, there are only a few known closed form expressions where the infinite series can be summed. One, refereed to as the Poisson kernel, is

$$c(\theta; \eta) = \frac{(1-\eta)^2(1-\eta^2)}{(1+\eta)(1 - 2\eta \cos(\theta) + \eta^2)^{-3/2}}, \qquad \eta \in (0,1), \qquad \theta \in [0, \pi], \tag{2}$$

where $\eta$ is related to the range parameter, *i.e.* how quickly the correlation decays, and $\theta$ is the great circle distance distance, *i.e.* the shortest angular distance between two points on the sphere.

A second more general approach is to restrict homogeneous and isotropic correlation functions on $\mathbb{R}^3$ to the sphere. In this case the Euclidean distance is replaced by $2R\sin(\theta/2)$ where $R$ is the globe's radius and $\theta$ is the great circle distance. A simple example is the exponential correlation function

$$c(\theta; \tau) = \exp\big(-2R\sin(\theta/2)/\tau\big), \qquad \tau > 0, \qquad \theta \in [0, \pi], \tag{3}$$

where $\tau$ is the range parameter. This approach can be generalized to Matérn family (Matérn, 1986; Handcock and Wallis, 1994), that allows parameterizing the smoothness of the process. We prefer this second approach and the choice of the covariance function is discussed in Section 4.

In Section 3.1 we proposed a model with given correlation matrices $\boldsymbol{\Sigma}$. This implies that the range parameters $\eta$ or $\tau$ are not modeled within the hierarchical model. However, the values of these parameters can be chosen according to an "empirical Bayes" approach or with restricted maximum likelihood techniques (REML). Both approaches consider the data model as a linear regression with correlated errors and estimate the error parameters with an iterative procedure as follows. To begin the algorithm assume an initial covariance matrix $\boldsymbol{\Sigma}$, then one estimates the mean structures, *i.e.* the vector $\boldsymbol{\theta}_i$, via weighted least squares (WLS). Given the estimate $\hat{\boldsymbol{\theta}}_i^*$ of the mean, $\mathbf{D}_i - \mathbf{M}\hat{\boldsymbol{\theta}}_i^*$ is used to estimate the covariance structure $\phi_i \boldsymbol{\Sigma}^*$ with a REML or a method-of-moments approach. Now $\hat{\boldsymbol{\theta}}_i^*$ is updated using WLS with the covariance matrix $\hat{\phi}_i^* \boldsymbol{\Sigma}^*$ and we obtain a second estimate $\hat{\boldsymbol{\theta}}_i^*$. These two steps are repeated until both $\hat{\boldsymbol{\theta}}_i^*$ and $\hat{\phi}_i^* \boldsymbol{\Sigma}^*$ converged according to some criterion. Given the similarity of both approaches it is no surprise that the estimates are also very similar.

## 3.4   Implementing a Gibbs Sampler

The goal of our hierarchical modelling approach is to obtain the posterior distribution of $\mathbf{M}\boldsymbol{\vartheta}$ given the model observations $\mathbf{D}_i$, *i.e.* $[\mathbf{M}\boldsymbol{\vartheta} \mid \mathbf{D}_1, \ldots, \mathbf{D}_N]$. The posterior density can be derived via Bayes' theorem (*e.g.* Bernardo and Smith, 1994), synthesized as

[ process | data, parameters ] $\propto$ [ data | process, parameters ] $\cdot$ [ process | parameters ] $\cdot$ [ parameters ].

The densities on the right-hand side are, of course, given by the three levels of the hierarchical model. The joint posterior is often a complicated distribution that has no closed form or from which it is impossible to draw directly. However, the posterior can be sampled using a Markov chain Monte Carlo (MCMC) procedure known as the Gibbs sampler (Geman and Geman, 1984; Gelfand and Smith, 1990). The essence of the MCMC approach is to simulate complex joint probability distributions by sampling from a Markov chain with a stationary, ergodic distribution that is identical to the posterior distribution (see also Gilks *et al.*, 1998; Robert and Casella, 1999).

Essentially the Gibbs sampler works as follows. For each parameter in the model its distribution conditional on all the other random quantities in the model is identified. Such distributions are called full conditionals because only the parameter of interest is allowed to be random and the entire remaining part of the model is fixed (or conditioned upon). The Monte Carlo algorithm cycles among the parameters by simulating a new value for each parameter based on the full conditional distribution and the current values of the other parameters. Under weak assumptions the sequence converges to the intended distribution.

Since the hierarchical model is based on multivariate normal and inverse gamma distributions, it is possible to derive the full conditionals in a closed from. To do so we use the property that if a random

variable $\mathbf{X}$ has a density proportional to $\exp\!\big(-1/2(\boldsymbol{x}^{\mathsf{T}}\mathbf{A}\boldsymbol{x} - 2\boldsymbol{x}^{\mathsf{T}}\boldsymbol{b})\big)$, for a positive definite symmetric matrix $\mathbf{A}$ and a vector $\boldsymbol{b}$, then $\mathbf{X}$ has a multivariate normal distribution with mean $\mathbf{A}^{-1}\boldsymbol{b}$ and variance $\mathbf{A}^{-1}$. Starting from the joint density $\big[\,\boldsymbol{\vartheta}, \{\boldsymbol{\theta}_i\}, \{\phi_i\}, \{\psi_i\}, \{\mathbf{D}_i\} \mid \xi_1, \ldots, \xi_5\,\big]$, the full conditionals are:

$$[\,\boldsymbol{\vartheta}\mid\ldots] \sim \mathcal{N}_p(\mathbf{A}^{-1}\boldsymbol{b}, \mathbf{A}^{-1}), \qquad \mathbf{A} = \frac{1}{\xi_5}\mathbf{I} + \sum_{i=1}^{N}\frac{1}{\psi_i}\mathbf{I}, \qquad \boldsymbol{b} = \sum_{i=1}^{N}\frac{1}{\psi_i}\boldsymbol{\theta}_i;$$

$$i = 1, \ldots, N: \quad [\,\boldsymbol{\theta}_i \mid \ldots] \sim \mathcal{N}_p(\mathbf{A}^{-1}\boldsymbol{b}, \mathbf{A}^{-1}), \qquad \mathbf{A} = \frac{1}{\psi_i}\mathbf{I} + \frac{1}{\phi_i}\mathbf{M}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{M}, \quad \boldsymbol{b} = \frac{1}{\psi_i}\boldsymbol{\vartheta} + \frac{1}{\phi_i}\mathbf{M}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{D}_i;$$

$$i = 1, \ldots, N: \quad [\,\phi_i \mid \ldots] \sim \mathrm{I\Gamma}\!\Big(\xi_1 + \frac{n}{2}, \xi_2 + \frac{1}{2}(\mathbf{D}_i - \mathbf{M}\boldsymbol{\theta}_i)^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{D}_i - \mathbf{M}\boldsymbol{\theta}_i)\Big);$$

$$i = 1, \ldots, N: \quad [\,\psi_i \mid \ldots] \sim \mathrm{I\Gamma}\!\Big(\xi_3 + \frac{p}{2}, \xi_4 + \frac{1}{2}(\boldsymbol{\theta}_i - \boldsymbol{\vartheta})^{\mathsf{T}}(\boldsymbol{\theta}_i - \boldsymbol{\vartheta})\Big),$$

where the '...' at the right of the conditioning sign refers to all the random quantities in the model, apart from the parameter to be drawn, and the data. Given the closed form of all the full conditionals, it is straightforward to implement a Gibbs sampler in any numerical software program.

We programmed the Gibbs sampler with the freely available computer software R (Ihaka and Gentleman, 1996; R Development Core Team, 2004). R ("GNU S") is similar to the S system, which was developed at Bell Laboratories by John Chambers and coauthors. It provides a wide variety of statistical and graphical techniques (linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering, *etc.*), as well as a fully functional programming environment. For the numerical experiments presented here, we run the sampler for a total of 20,000 iterations discarding the first half of the simulated values to obtain convergence and saving only every 20th draw. Thus, we base our conclusions on a total of 500 values for each parameter, representing a sample from its posterior distribution. On a reasonable desktop PC this task can be performed within a few hours.

The parameter space of the entire model is $(N+1) \times p + 2N$ roughly of the order of 1000. It is therefore not possible to perform formal tests of simultaneous convergence of the sampled chains. However, looking at trace plots, correlation plots, we strongly believe that the posterior samples are stationary and reached convergence.

## 4   Results

In this section we present the results concerning surface temperature and log precipitation output fields from the 9 AOGCM models.

The original CMIP $2.8 \times 2.8$ degree resolution is too large to be handled in a direct way and we also do not expect significant agreement in the models at this resolution. We use a simple bilinear interpolator to reduce the climate fields to a $5 \times 5$ degree resolution. Although we work on the sphere, our results are much more stable if we do not include the poles. Therefore we only consider the fields from $-80$ to $80$ degree latitude.

Table 2 summarizes the simulation parameters and hyperparameters for the simulations. The most critical parameter is the number of basis functions. We found that the key choices are the number of spherical harmonics functions, inclusion of indicator functions for land forms and sea ice, and the field of present climate derived from observed data. The correlation range $\tau$ for $\boldsymbol{\varepsilon}_i$ was found to be directly related to the number of basis functions. If we include more basis functions in $\mathbf{M}$ the resulting processes $\boldsymbol{\varepsilon}_i$ has a smaller spatial range. This result is not unexpected because a richer basis will remove more structure in the field leaving less for the small scale component. Based on empirical variograms we were lead to model the covariance using the exponential form in (3). In addition, simulations showed that the correlation matrix based on (3) yielded slightly more stable results compared to (2). This is probably due to the linear behavior of the exponential covariance function at the origin. However, we added a small nugget effect to the covariance with fixed signal-to-noise ratio. The choice of hyperparameters does not appear to be crucial to the results.

| | | Simulation parameters | | | Hyperparameters | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $p$ ($p_s$) | range $\tau$ | s/n ratio | $\xi_1$ | $\xi_2$ | $\xi_3$ | $\xi_4$ | $\xi_5$ |
| Temperature | DJF | 103(81) | 0.15 | 19 | 2.1 | 0.11 | 2.1 | 0.11 | 1 |
| | JJA | 103(81) | 0.15 | 19 | 2.1 | 0.11 | 2.1 | 0.11 | 1 |
| Precipitation | DJF | 191(169) | 0.1 | 9 | 2.1 | 0.11 | 2.1 | 0.11 | 1 |
| | JJA | 191(169) | 0.1 | 9 | 2.1 | 0.11 | 2.1 | 0.11 | 1 |

Table 2: Parameter and hyperparameter specifictions for the Gibbs sampler. An inverse gamma distribution with $\xi_1 = 2.1$, $\xi_2 = 0.11$ has mean 0.1 and variance 0.1.
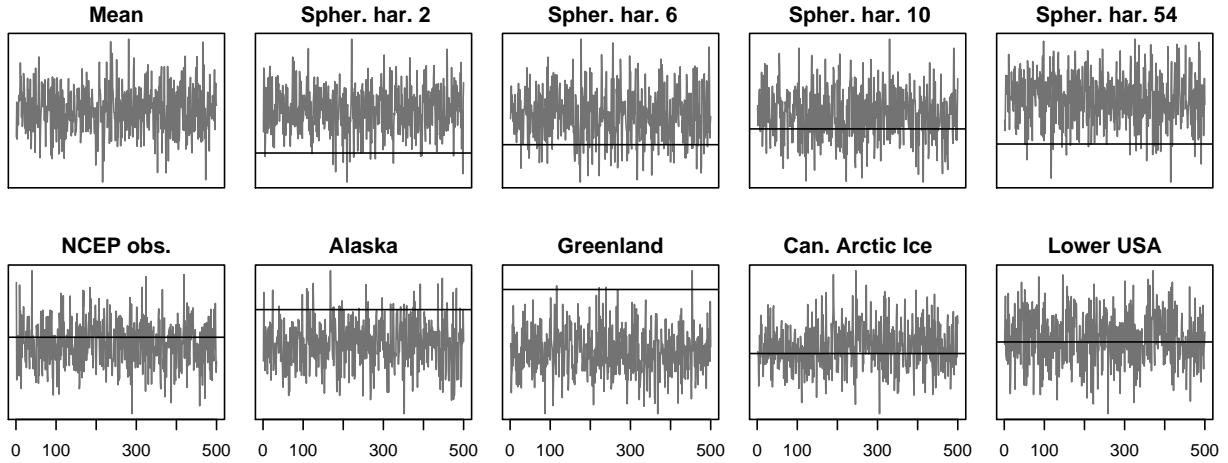


Figure 8: Trace plots for the mean, the spherical harmonics given in Figure 5 (top row), the NCEP climate observation and the land indicator fields given in Figure 6 (bottom row) of the parameter $\boldsymbol{\vartheta}$ for the DJF temperature change. The horizontal line represents zero.

Figure 8 and 9 depict trace plots and kernel estimates of the posterior densities for the mean, the spherical harmonics, the NCEP climate observations and the land indicator fields given in Figure 5 and 6 of the parameter $\boldsymbol{\vartheta}$.
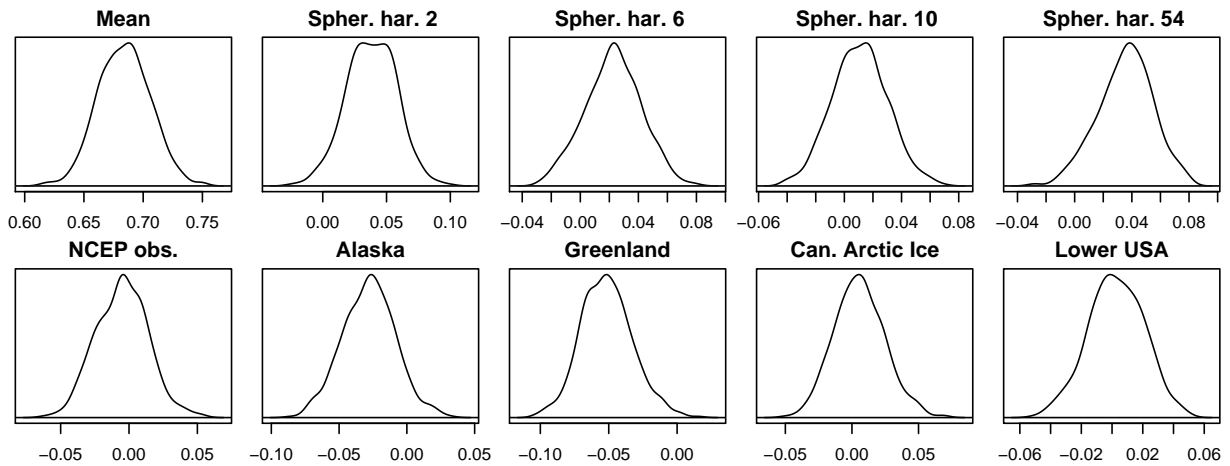


Figure 9: Kernel estimates of the posterior densities for the mean, the spherical harmonics given in Figure 5 (top row), the NCEP climate observation and the land indicator fields given in Figure 6 (bottom row) of the parameter $\boldsymbol{\vartheta}$ for the DJF temperature change.

## 4.1   Analysis of Surface Temperature and Precipitation Fields

Surface temperature is a rather smooth field and only about 100 basis functions ($\ell = 8$) are required to obtain a good fit of the model. Here, we judge adequacy of the fit by assessing whether the small scale residual field is consistent with our assumption on isotropy. Figure 10 gives the temperature change that occurs with at least 80% probability in 70 years with 1% $CO_2$ increase. Figure 11 shows the probability that the temperature change exceeds 2°C. The numbers 80% and 2°C are arbitrary but this particular format was found to be useful in presenting results to a general scientific audience.
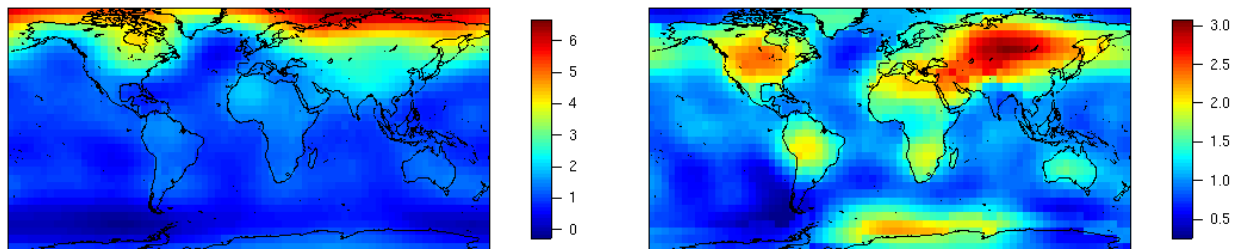


Figure 10: The DJF (left panel) and JJA (right panel) temperature change (degree Celsius) that occurs with at least 80% probability in 70 years with 1% $CO_2$ increase.
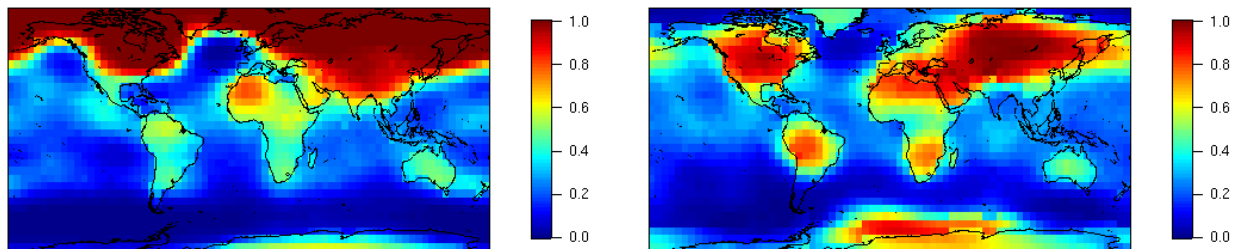


Figure 11: Probability that DJF (left panel) and JJA (right panel) temperature change exceeds 2°C.

The statistical model was also applied to smaller regions than the full globe with the goal of making inferences at a regional scale. The hierarchical model still works well at these small scales, but for very small regions (100 grid cells or less) the results are more sensitive to the number of basis functions.

Log precipitation fields have more small scale variability than temperature and are highly non-stationary, therefore we were lead to use more basis functions compared to temperature ($\ell = 12$). As example for the analysis of the precipitation fields, Figure 12 gives the median precipitation change that occurs in 70 years with 1% $CO_2$ increase.
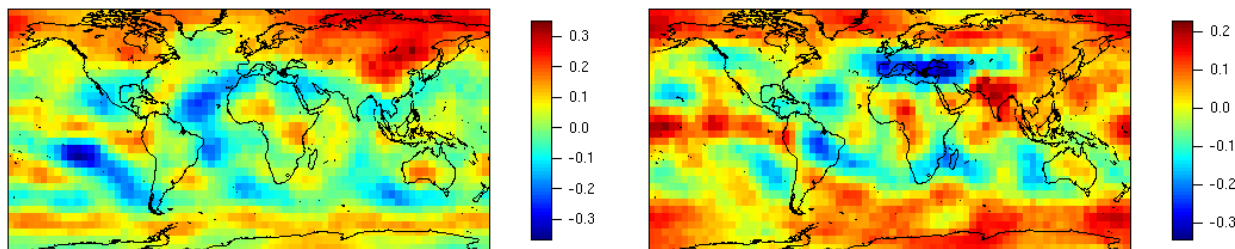


Figure 12: The median DJF (left panel) and median JJA (right panel) log precipitation change (mm/day) that occurs in 70 years with 1% $CO_2$ increase.

## 4.2   Model Checking

Although the posterior fields for each model, $[\mathbf{M}\boldsymbol{\theta}_i \mid \mathbf{D}_1,\ldots,\mathbf{D}_N]$, are not directly interesting we believe that they can nevertheless be useful for diagnostic purposes. Samples from these posterior fields should resemble the observed fields. Figure 13 shows posterior draws from the ECHAM model, where the lower right panel depicts the CMIP data. Overall the realizations from the posterior look qualitatively similar to the actual model output. This agreement confirms the adequacy of our modeling approach. Since the data panel exhibit a smoother behaviour in the arctic sea, slight improvements would probably be achieved if we generalize the errors $\boldsymbol{\varepsilon}_i$ to anisotropic, non-homogeneous spatial processes. The resulting covariance should have smaller scales and larger ranges in high latitudes compared to the equatorial zone.
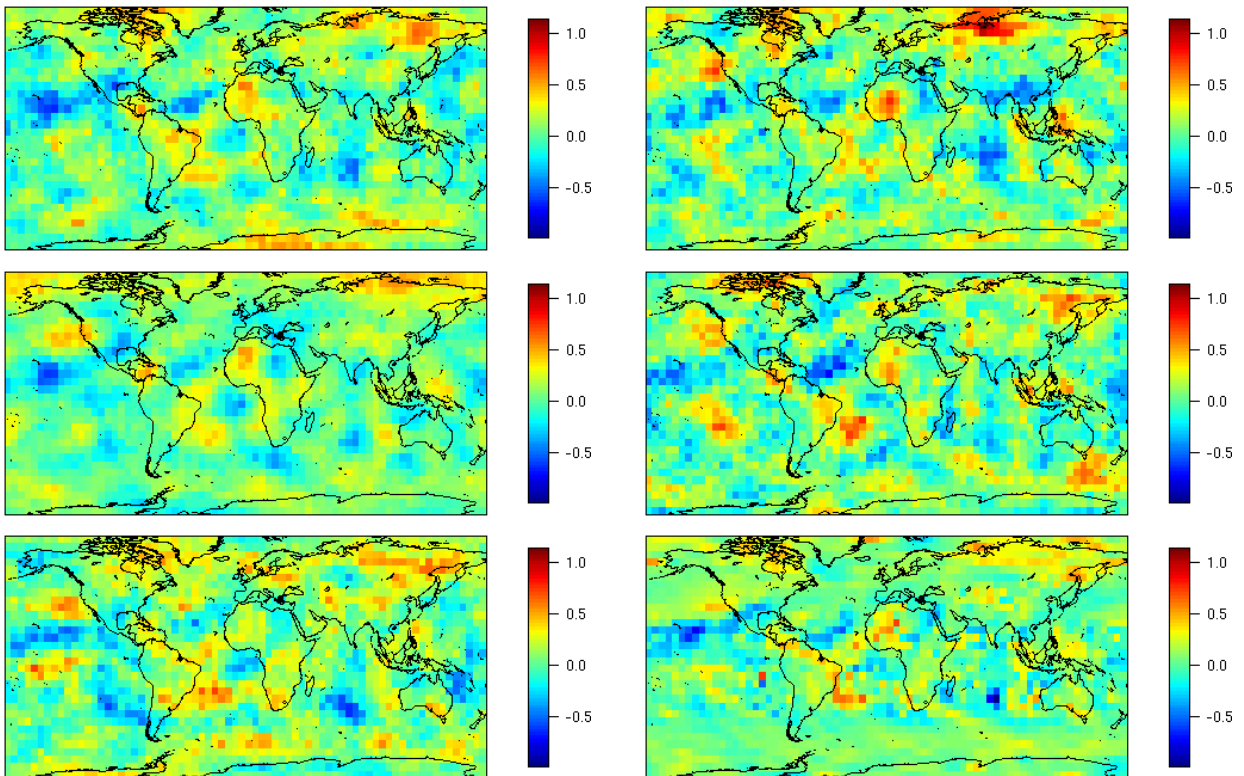


Figure 13: Posterior fields of precipitation climate change for the ECHAM model. The lower right panel gives the AOGCM result (Unit: mm/day).

## 5   Extensions

This paper has focused on the simplest model for comparing climate model output, assuming single realizations from each AOGCM and differenced (future minus present) climate fields. Although this presentation emphasizes the basic spatial elements of this problem, there are several important extensions to consider to make this analysis appropriate for application in current IPCC activities.

The presented model can be extended in three intuitive ways: we could 1) include the the covariance parameters of $\boldsymbol{\Sigma}$ as part of the hierarchical model, 2) include several different runs of the same AOGCM and 3) model the control and transient runs separately.

## 5.1 Parameterization of the Covariance Structure

A natural way to parametrize the covariance is to model the range parameter and the signal to noise ratio in the Gibbs sampler. With correlation model (3), the exponential covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\phi})$ would be based on

$$c(\theta; \boldsymbol{\phi}) = \phi_1 I_{\{\theta=0\}} + \phi_2 \exp\left(-2R \sin(\theta/2)/\phi_3\right), \qquad \phi_i > 0, \qquad \theta \in [0, \pi],$$

where $I$ is the indicator function. The resulting full conditionals for the parameters $\phi_i$ do not have a closed form and hence it would be necessary to include Metropolis–Hastings steps in the Gibbs sampler (*e.g.* Banerjee *et al.*, 2004). Simulations suggested that this additional computational burden does not offer significant improvements in the performance of the model.

## 5.2 Different Ensemble Runs from Each AOGCM

The new generation of model data for the forth assessment report (AR4) of the IPCC has several runs for different models. These are termed ensembles of runs and can be interpreted as multiple realizations of weather from the same climate. The hierarchical approach can be extended by

$$\mathbf{X}_{i,j} = \text{climatological field from control run } j \text{ of model } i,$$
$$\mathbf{Y}_{i,j} = \text{climatological field from transient run } j \text{ of model } i,$$

and modelling the difference $\mathbf{D}_{i,j} = \mathbf{Y}_{i,j} - \mathbf{X}_{i,j}$. Here, the variability between separate runs in an ensemble is due to the internal variability of the model and not due to differences in simulated climate. In this case, we would have the possibility to estimate the internal variability of each model (*e.g.* $\phi_i$).

## 5.3 Separate Control and Transient Run Modelling

The data level of the presented hierarchical model addresses the climate change directly. Another approach is modelling the control and transient fields individually and linking them via a correlation structure. Again, we use a large scale and small scale separation for the control and transient runs:

$$\mathbf{X}_i = \mathbf{M_X}\boldsymbol{\theta}_i + \boldsymbol{\sigma}_i, \qquad\qquad \mathbf{Y}_i = \mathbf{M_Y}\boldsymbol{\eta}_i + \boldsymbol{\nu}_i,$$

where the errors satisfy

$$[\,\boldsymbol{\sigma}_i \mid \boldsymbol{\Sigma}_i\,] \overset{\text{iid}}{\sim} \mathcal{N}_n(\mathbf{0}, \boldsymbol{\Sigma}_i),$$
$$\boldsymbol{\nu}_i = \boldsymbol{\omega}_i + \rho_i \boldsymbol{\sigma}_i \quad \text{with} \quad [\,\boldsymbol{\omega}_i \mid \boldsymbol{\Omega}_i\,] \overset{\text{iid}}{\sim} \mathcal{N}_n(\mathbf{0}, \boldsymbol{\Omega}_i) \text{ and } \boldsymbol{\omega}_i \perp \boldsymbol{\sigma}_i.$$

If $\rho_i = 0$, then $\boldsymbol{\nu}_i \perp \boldsymbol{\sigma}_i$ and if $\rho_i = 1$, then $\boldsymbol{\sigma}_i \perp (\boldsymbol{\nu}_i - \boldsymbol{\sigma}_i)$. The data level in this approach can be written as a $2n$-dimensional normal model. The advantage of using this conditional approach compared to a direct $2n$-dimensional normal model is the simple way of parameterizing the correlation structure without constraints on $\rho_i$. An important advantage of this separation is the ability to use the observed climate as an additional hierarchical component. For example, let $\mathbf{X}_0$ denote the NCEP observation field, then we also include

$$\mathbf{X}_0 = \mathbf{M}\boldsymbol{\vartheta} + \boldsymbol{\sigma}_i.$$

in the statistical modeling approach. This model is a spatial extension of Tebaldi *et al.* (2005). The process level and prior level are similar to what has been discussed in Section 3.1. Note that model (1) is a special case of this more general approach if $\boldsymbol{\Sigma}_i = \phi_{\mathbf{X}_i}\boldsymbol{\Sigma}$, $\boldsymbol{\Omega}_i = \phi_{\mathbf{Y}_i}\boldsymbol{\Sigma}$ and $\mathbf{M_X} = \mathbf{M_Y}$.

## 6   Discussion

A key feature of this work is not only an assessment of the difference in the spatial fields but also a measure of the uncertainty of this difference based on the posterior distributions. From the analysis of the climate model experiments, there is considerable warming over the landmasses and some warming over the oceans. These inferences are illustrated in Figures 10 and 11, where potential temperature shifts are related to differences that have high posterior probability. Precipitation change is harder to interpret. However, most of the sub-tropical areas have a decreased mean precipitation and in tropical areas and in the high latitudes the mean precipitation increases (Figure 12). Again, the posterior distribution is useful in characterizing the uncertainty of these results.

Although the presented analysis has used the CMIP experiment an important direction for future work is to apply the same statistical approach to a comprehensive set of AOGCM experiments prepared for the next IPCC report (AR4). Part of the computational challenge of this work will be to implement the statistical models for the full resolution of the climate models.

Even though our presentation has included many details specific to climate model output we have illustrated a general methodology that can be used in other contexts. The basic ingredient is to separate the spatial response into significant features represented by a small number of basis functions and a small scale process that has little structure. Although this separation may appear difficult to achieve, we should note that variogram analysis of the residuals and the posterior distribution provides detailed diagnostics and implied in our case the adequacy of the model. It is particularly efficient to use a hierarchical statistical model to implement these ideas. Moreover, for a Bayesian analysis many components of the model are easy to handle using a Gibbs sampler. Although we have chosen to estimate some parts of the model using maximum likelihood there is always the option to include a richer Bayesian model that accounts for these parameters. The extensions of our model outlined in Section 5 involve a more complicated MCMC algorithm but also include important features not accounted for by a simple approach.

We conclude by noting one outstanding problem that is an area of future statistical research. The choice of basis functions was done in a subjective manner based on the scientific background and diagnostic of the fit. However, it would be useful to have the model selection included as part of the Bayesian analysis. In this way an important component, the basis functions choice, is included in the overall uncertainty encapsulated by the posterior distribution.

## References

Allen, M. R., Stott, P. A., Mitchell, J. F. B., Schnur, R., and Delworth, T. L. (2000). Quantifying the uncertainty in forecasts of anthropogenic climate change. *Nature*, **407**, 617–620.   2

Banerjee, S., Gelfrand, A. E., and Carlin, B. P. (2004). *Hierarchical Modeling and Analysis for Spatial Data.* Chapman & Hall/CRC.   4, 5, 13

Berliner, L. M., Wikle, L., and Cressie, N. (2000). Long-lead prediction of Pacific SST via Bayesian dynamic modelling. *Journal of Climate*, **13**, 3953–3968.   4

Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory.* John Wiley & Sons Inc., Chichester.   8

Covey, C., AchutaRao, K. M., Lambert, S. J., and Taylor, K. E. (2000). Intercomparison of present and future climates simulated by coupled ocean-atmosphere GCMs. PCMDI Report No. 66, UCRL-ID-140325, Lawrence Livermore National Laboratory, Livermore, CA.   3

Cressie, N. A. C. (1993). *Statistics for Spatial Data.* John Wiley & Sons Inc., New York, revised reprint.   5, 7

Forest, C. E., Stone, P. H., Sokolov, A. P., Allen, M. R., and Webster, M. (2002). Quantifying uncertainties in climate system properties with the use of recent climate observations. *Science*, **295**, 113–117.   2

Frame, D. J., Booth, B. B. B., Kettleborough, U. K. J. A., Stainforth, D. A., Gregory, J. M., Collins, M., and Allen, M. R. (2005). Constraining climate forecasts. Submitted to *Journal of Geophysical Research*. 2

Gaspari, G. and Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, **125**, 723–757.  7

Gelfand, A. and Smith, A. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398–409.  8

Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741.  8

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1998). *Markov Chain Monte Carlo in Practice*. Chapman & Hall.  8

Gneiting, T. (1999). Correlation functions for atmospheric data analysis. *Quarterly Journal of the Royal Meteorological Society*, **125**, 2449–2464.  7

Gregory, J. M., Stouffer, R. J., Raper, S. C. B., Stott, P. A., and Rayner, N. A. (2002). An observationally based estimate of the climate sensitivity. *Journal of Climate*, **15**, 3117–3121.  2

Handcock, M. S. and Wallis, J. R. (1994). An approach to statistical spatial-temporal modeling of meteorological fields. *Journal of the American Statistical Association*, **89**, 368–390.  8

Houghton, J. T., Ding, Y., Griggs, D. J., Noguer, M., van der Linden, P. J., Dai, X., Maskell, K., and Johnson, C. A. (2001). *Climate Change 2001: The Scientific Basis*. Cambridge University Press, Cambridge. 3

Ihaka, R. and Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, **5**, 299–314.  9

Jones, R. H. (1963). Stochastic processes on a sphere. *The Annals of Mathematical Statistics*, **34**, 213–218. 6

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woolen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetma, A., Reynolds, R., Jenne, R., and Joseph, D. (1996). The NCEP/NCAR 40-year reanalysis project. *American Meteorological Society Bulletin*, **77**, 437–471.  7

Knutti, R., Stocker, T. F., Joos, F., and Plattner, G.-K. (2002). Constraints on radiative forcing and future climate change from observations and climate model ensembles. *Nature*, **416**, 719–723.  2

Knutti, R., Stocker, T. F., Joos, F., and Plattner, G.-K. (2003). Probabilistic climate change projections using neural networks. *Climate Dynamics*, **21**, 257–272.  2

Matérn, B. (1986). *Spatial Variation*. Springer-Verlag, Berlin, second edition.  8

Meehl, G. A., Boer, G. J., Covey, C., Latif, M., and Stouffer, R. J. (2000). The coupled model intercomparison project (CMIP). *American Meteorological Society Bulletin*, **81**, 313–318.  3

R Development Core Team (2004). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.  9

Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer.  8

Schneider, S. H. (2001). What is dangerous climate change? *Nature*, **411**, 17–19.  2

Schoenberg, I. J. (1942). Positive definite functions on spheres. *Duke Mathematical Journal*, **9**, 96–108.  7

Tebaldi, C., Smith, R. L., Nychka, D., and Mearns, L. O. (2005). Quantifying uncertainty in projections of regional climate change: a Bayesian approach to the analysis of multimodel ensembles. *Journal of Climate*, **18**, forthcoming.   2, 13

Weber, R. O. and Talkner, P. (1993). Some remarks on spatial correlation function models. *Monthly Weather Review*, **121**, 2611–2137; Corrigendum: (1999), **127**, 576.   7

Wigley, T. M. L. and Raper, S. C. B. (2001). Interpretation of high projections for global-mean warming. *Science*, **293**, 451–454.   2

Wikle, C., Berliner, L., and Cressie, N. (1998). Hierarchical Bayesian space-time models. *Environmental and Ecological Statistics*, **5**, 117–154.   4

Wikle, C. K. and Cressie, N. (1999). A dimension-reduced approach to space-time Kalman filtering. *Biometrika*, **86**, 815–829.   4

Yaglom, A. M. (1987). *Correlation theory of stationary and related random functions. Vol. I.* Springer Series in Statistics. Springer-Verlag, New York.   7