Characterizing the Uncertainty of Climate Change Projections Using Hierarchical Models

Claudia Tebaldi[*]
&
Richard L. Smith[†]

January 1, 2009

Chapter Y of

The Handbook of Applied Bayesian Analysis

Eds: Tony O'Hagan & Mike West
Oxford University Press

Abstract

We present a suite of Bayesian hierarchical models that synthesize ensembles of climate model simulations, with the aim of reconciling different future projections of climate change, while characterizing their uncertainty in a rigorous fashion. Posterior distributions of future temperature and/or precipitation changes at regional scales are obtained, accounting for many peculiar data characteristics, like systematic biases, model-specific precisions, region-specific effects, changes in trend with increasing rates of greenhouse gas emissions. We expand on many important issues characterizing model experiments and their collection into multi-model ensembles. Also, we address the need of impact research, by proposing posterior predictive distributions as a representation of probabilistic projections. In addition, the calculation of the posterior predictive distribution for a new set of model data allows a rigorous cross-validation approach to confirm the reasonableness of our Bayesian model assumptions.

[*]Climate Central, Princeton NJ 08542-3718
[†]Dept. of Statistics and Operations Research, University of North Carolina, Chapel Hill NC 27599-3620

1 Climate change and human influences, the current state and future scenarios

There is substantial consensus on many aspects of climate change. It is already with us; we have a large responsibility in many facets of it; and future changes in the absence of significant curbing of greenhouse gas emissions are going to be much more dramatic than what is already experienced, with consequences that will be predominantly detrimental to social and natural systems. Adaptation and mitigation decisions need however detailed information about the reality and the future of climate change, and more often than not at regional, rather than global mean scales. To produce this information, deterministic numerical models are run under different hypothetical scenarios corresponding to alternative greenhouse gas emission pathways into the future. These extremely complex computer models discretize the surface of the Earth, the depths of the oceans and the layers of the atmosphere into regularly spaced grid boxes. Computations through differential equations provide the evolution of a high dimensional state vector representing climate variables, by applying our understanding of climate dynamics and climate interacting processes over land, water, air and sea ice (Washington and Parkinson, 2005). With the evolution of science and technology, more and more processes at increasingly finer scales can be represented explicitly in these simulations, but there still remains the need for approximations, for those processes that act at scales not explicitly represented. It is in these approximations that the source of large uncertainties resides.

1.1 Many climate models, all of them right, all of them wrong

Because of the complexities and high dimensional nature of global climate simulations, many alternative solutions to the representation and parameterization of processes are consistent with the state of our scientific understanding. Thus, different models approach the simulation of climate over the Earth with different strategies. It may be as basic as the choice of the resolution used to discretize the globe in its three dimensions (which also affects the time step of the finite difference equations). This choice has important consequences on the range of processes that can be directly simulated and those that have to be represented by parameterization of the sub-grid scales[1]. Of course there are less basic choices, like the processes that are included, independently of the computing limitation: is an interactive carbon cycle present? Is vegetation changing with climate? Are urban areas and their effect on climate represented? Even for a given resolution and a given set of processes, the actual computation of the time-evolving series of quantities that make up the climate system may differ because of different values in exogenous parameters of the equations (usually estimated by studying the individual processes in the field) or of different numerical solvers adopted.

Within this large population of model configurations we are hard pressed to find first and second class citizens. There are more than a dozen global climate models (many of which run at different resolutions) which are comparable in terms of overall performance, when validated against the observed record and paleo-climate proxy records. Of course some models are better than others for a given set of metrics (average temperature over North America, precipitation in the monsoon region, the frequency and intensity of the El Niño Southern Oscillation phenomenon) but for any set of "better than average models" a different set of metrics can be found for which those models will underperform compared to a different set (Gleckler et al., 2007).

Thanks to international efforts like the Intergovernmental Panel on Climate Change periodic assessment reports, the last of which was published in 2007 (IPCC, 2007), modeling centers participate in concerted sets of experiments, running their models under standardized scenarios of external forcings. These are standard greenhouse gas concentration pathways, derived

[1] Parameterization is performed by estimating the relation between the large scale, resolved processes and the small scale, unrepresented processes that happen within a model grid box. Observational or experimental studies provide the basis for the estimation of the parameters that govern the interaction between them. Uncertainties in observations and in the relations themselves are at the source of the uncertainties in parameterizations, causing a range of values for these parameters to be consistent with their function, but at the same time causing sugnificant differences in the evolution of quantities in the model integrations.

after scientists devised a set of alternative, but equally plausible, future story-lines about the social, political, technological and economic future of our world. From these story lines consequences in terms of greenhouse gases and other pollutants' emissions were derived using economic integrated modeling. For different families of scenarios (varying from a future of fast technological progress and international collaboration to one of slower integration and fossil-fuel-intensive economies) very different trajectories of greenhouse gas rates of emissions over this century and beyond are specified. These scenarios of greenhouse gas emissions, known as SRES scenarios (Nakicenovic, 2000), have been used by climate modelers as external inputs to run simulations of alternative climate scenarios in the future. By the second half of this century, different emission scenarios cause significantly different climate outcomes, starting from global average temperature change but riverberating in all aspects of Earth's climate. By the end of this century, the uncertainty across scenarios of greenhouse gas increases is larger than the inter-model differences under a specific scenario. However, the climate community feels it inappropriate to attach probabilities to different emission pathways, and we are left with performing our uncertainty analyses conditionally on a given SRES scenario.

After running model experiments under alternative SRES scenarios, the modeling centers are contributing the resulting simulations' output into open access archives. These collections of model simulations have been labelled ensembles of opportunity, i.e., multiple models' collections that are not the result of a statistically designed experiment or random sampling from a population of models, but a post facto collection of what is available, thanks to the voluntary and self-selecting participation of the world's largest and most advanced research centers. The most recent and largest archive of such data sets is maintained by the Program for Climate Model Diagnosis and Intercomparison (PCMDI) at Lawrence Livermore National Laboratory (LLNL), and can be found at `http://www-pcmdi.llnl.gov/ipcc/about_ipcc.php`

1.2 Goals and challenges of analyzing ensembles of opportunity

The most direct way to obtain regionally detailed future projections is to process the output of global climate models and determine statistics of the regional climate variables of interest. Determining which model to trust above all others is a daunting task, and one defensible strategy is to utilize all that are available, synthesizing the projections and their uncertainty through a rigorous statistical analysis. This will provide optimal estimates of the changes in store, and will quantify their uncertainty, conditionally on the available information. This kind of representation is of great value to decision makers and stakeholders, notwithstanding the need of communicating the underlying limitations of our current understanding and working hypotheses as embodied by these models, and the assumptions that are at the basis of any statistical representation of the data. Bayesian statistics has a natural advantage in this particular setting, not only because we are dealing with uncertain events that are not easily framed in a frequentist perspective, but more fundamentally because of its natural framework for incorporating expert judgement, and updating current assessment with the in-flow of additional pieces of information. There are great challenges underlying any statistical analysis of such multi-model ensembles (Tebaldi and Knutti, 2007). They stem from the non-systematic and conversely non-random nature of the sampling of models, which hampers a full representation of the uncertainties at stake; from the dependence among the models in the ensemble, some sharing components, some sharing even their full name, when the same modeling center contributes output from runs at different resolutions; from the lack of a theory linking model performance for current climate, that we can measure and validate, to model reliability for future projections. A particular difficulty is the impossibility of verifying models' future projections, which sets this problem well apart from weather forecasting, where feedback about the performance of the model can be as immediate as six hours after the forecast is issued.

All these challenges stand in the way of a robust representation of climate change projections, especially at regional scales and especially for variables other than temperature, which is relatively easier to model because of its smoothness. We are using "robust" here in a more generic sense than is usually understood by the phrase Bayesian robustness, though the latter may also be worth exploring in our setting. Climate scientists, hydrologists and agricultural modelers – among others – are often interested in studying the impacts of climate change and of adaptation measures. The traditional scientists' approach is through scenarios, where alter-

native futures used to span a range of outcomes are fed through impact models. In contrast with this approach, we argue for a rigorous uncertainty analysis via Bayesian statistics. In fact much of the impact research efforts are veering towards full probabilistic analysis, and probabilistic projection of climate change constitute their stepping stone(Tebaldi and Lobell, 2008). We think it is fair to say that the area is still in a phase of methodological development. We hope that this chapter will offer enough motivation, and open up enough interesting research directions to invite fresh perspectives to this important application of statistical analysis.

2 A world of data. Actually, make that 'many worlds'

The most up to date archive of a multimodel ensemble, hosted by PCMDI at LLNL, contains over 35 terabytes of data. It collects output from 23 models, run under a range of SRES emission scenarios. Scores of variables constitute the output of a climate model simulation, as familiar as temperature and precipitation or as esoteric as the mass of water that evaporates over an ice sheet, or metrics of ocean overturning. Many of the variables are archived as daily or monthly means, some are 6-hourly, some are yearly averages. The potential for statistical analysis of model output is practically infinite, when we consider that these quantities are produced at each point of grids that discretize the entire Earth surface, and at many layers in the atmosphere and oceans. The median resolution of the climate models in the PCMDI archive is about 2.75 degrees in latitude/longitude, making the standard grid output for each variable and each time step, when vectorized, 8,192 components in length. A typical climate change experiment consists of a simulation that starts from conditions describing the state of the system at a pre-industrial time, chosen often by convention as 1870, and run with only external forcing imposed, otherwise in a self consistent and isolated manner, until year 2100. External forcings are meant to represent changing greenhouse gas concentrations over time, aerosols, volcano eruptions and solar cycles. The system evolves (at a few minutes time step) according to the laws of physics known to govern climate dynamics. To the extent permitted by their resolution, climate models represent coastlines and topography, and they impose vegetation types, often evolving along the simulation timeline, in order to represent urbanization, deforestation, switches from natural vegetation to crop growing areas and vice versa. Even for a given experiment (i.e., a given SRES scenario of external forcings) a model is run for a handful of slightly different initial conditions. The trajectories of the members of these single model ensembles give a handle on the characterization of natural internal variability of the system, generated by the intrinsically chaotic nature of weather and, in the long run, climate processes. For an analysis of climate change, its signal has to be isolated and extracted from the background noise of natural variability by averaging the members of the initial-conditions ensemble, and by considering differences between two at least 20-year averages, usually. The idea is that there is initial condition uncertainty to take care of, and there are low frequency natural modes of variability (decadal oscillations and other phenomena like the alternating phases of El Niño/La Niña conditions in the Pacific) that need to be averaged out before starting to talk about anthropogenic (i.e., externally forced by human actions) climate change.

Averaging in time is one way of facilitating the extraction of the signal of change. Averaging over space is its natural counterpart. Local climate is noisy, and climate models, because of their coarse resolution, are not expected to reproduce faithfully the statistics of local climate. Even in the face of a constant push to provide regionally detailed information from these models, it is fair to say that our confidence in their ability to simulate climate is highest when large regions of sub-continental size are concerned, and their climate considered in terms of large area-averages.

3 Our simplified datasets

In the following sections of the chapter we present a suite of statistical analyses combining output from ensemble of climate models, from now on referred to as GCMs (General Circulation Models). We will significantly simplify the problem by using summary statistics of their output, regional means of temperature and precipitation, seasonally averaged and aggregated as 10 or 20 year means. We always condition our analysis to a given experiment, defined in terms

of the greenhouse gas emission scenario (SRES) used as part of the external forcings by the simulation. For a given scenario a number of GCMs have run simulations covering the period 1870 through 2100 and have archived temperature and precipitation output. We also have observed records (in some region of the world extending as far back) that can be used to gauge the GCM ability to reproduce historic conditions. The model runs take account of changes in forcings, including observed changes in emissions or solar output, but they are not directly calibrated to observational climate data. One consequence of this is that the model outputs include dynamical features such as El Niños, but the El Niños of the model do not correspond in time to the El Niños of the observational climate record. One reason for considering 20-year averages is that over a 20-year time span, such short term fluctuations are likely to average out.

Suppose there are M GCMs, X_j is a projection of some current climate variable generated by GCM j, and Y_j a projection of some future climate variable generated by GCM j. We also have an observation X_0 of the true current climate, with its associated standard error $\lambda_0^{-1/2}$ that we can estimate from the observations' series and fix in our model. In our typical application, X_j is the mean temperature or precipitation in a particular region for the period 1981–2000, X_0 is the corresponding value calculated from the observational climate record, and Y_j is the corresponding temperature average calculated from the 2081–2100 segment of the GCM simulation.

A modification of this simple setup will involve R regional averages at a time. Accordingly, we add a subscript $i = 1, \ldots, R$ to the variables, and consider $X_{ij},\ Y_{ij},\ X_{i0},\ \lambda_{0i}$.

Finally we will model the joint distribution of two variables, say temperature and precipitation, for a given region and season, and over the entire length of the simulation, as a bivariate time series of decadal averages. Accordingly we will consider X_{jt}, $t = 1, \ldots, 15$ a bivariate vector of temperature and precipitation averages derived from the j-th GCM output. Here the time index corresponds to the decades centered at $1955, 1965, \ldots, 2005, 2015, \ldots, 2095$, so that both historical and future periods will be modelled jointly. Similarly, O_t, $t = 1, \ldots 6$ will indicate a two-component vector of observed temperature and precipitation averages. In this case the time index, t, corresponds to the decades centered at $1955, 1965, \ldots, 2005$ (the last one at present is approximated by an estimate based on 8 years of data only).

4 A hierarchy of statistical models

Our strategy in presenting our approach to multi-model synthesis and uncertainty characterization is to start from a basic representation, highlight its shortcomings and increase the complexity of the statistical treatment gradually to account for as many additional details as possible. This way, we hope to highlight issues, limitations and solutions step by step. Hopefully this will help the reader achieve familiarity with the data and the ultimate goals of the analysis that will engender his or her constructive criticisms and original thinking in the face of this challenging application.

4.1 One region at a time

Let us then start from the simplest series of assumptions. We treat each region separately, and we assume the following likelihood model: For $j = 1, \ldots, M$,

$$\begin{aligned} X_j &\sim N(\mu, \lambda_j^{-1}) \\ Y_j &\sim N(\nu, (\theta\lambda_j)^{-1}) \end{aligned}$$

and for the observed mean climate variable (from now on we referred to it as temperature, for simplicity) ,

$$X_0 \sim N(\mu, \lambda_0^{-1}). \tag{1}$$

This likelihood model simply states that each GCM approximates the true mean temperature of the region (μ for current, ν for future climate) with a Gaussian error, whose variance is model specific. The parameter θ allows for the possibility that future precision will be different

than current precision, by a factor common to all GCMs (i.e., the likely degradation of the accuracy of future projections affects all GCMs equally). Notice that this is an assumption dictated by our data, which would not permit the identification of a parameter modeling GCM-specific change in precision. In fact the precision parameter λ_j will be estimated on the basis of the minimum necessary number of datapoints, two. Also, the likelihood model assumes no correlation between the error of GCM j in simulating current climate and its error in simulating future climate.

The use of uninformative improper priors for μ and ν, $U(-\infty, +\infty)$, and proper but very diffuse Gamma priors for the precision parameters λ_j and θ, $Ga(a, b)$, with $a = b = 0.01$ completes this basic model. A simple Gibbs sampler can be used to explore the joint posterior distribution of the parameters (see Appendix B1). This approach was first presented in Tebaldi et al. (2004, 2005).

Here we highlight a few features of the approximate posterior estimates for the parameters μ, ν and λ_j. The form of the means of μ and ν, given λ_0, λ_j, $j = 1, \ldots, M$ as well as given the Xs and Ys is respectively,

$$\tilde{\mu} = \frac{\lambda_0 X_0 + \sum \lambda_j X_j}{\lambda_0 + \sum \lambda_j}$$

and

$$\tilde{\nu} = \frac{\sum \lambda_j Y_j}{\sum \lambda_j}.$$

The posterior distribution of λ_j can be approximated by

$$\lambda_j \mid \text{rest} \sim Ga\left[a + 1, b + \frac{1}{2}(X_j - \mu)^2 + \frac{\theta}{2}\{Y_j - \nu\}^2\right].$$

An easy addition to this model may accommodate the obvious critique that we expect correlation between errors within a single GCM's simulations of current and future climate. We can substitute to the likelihood of Y_j the following:

$$Y_j | X_j \sim N(\nu + \beta(X_j - \mu), (\theta\lambda_j)^{-1}).$$

By so doing we estimate a possible correlation between X_j and Y_j . Here as before with θ we are forced to assume a common correlation across all GCMs. Perhaps this is too simplistic an assumption, but it is needed for identifiability of the parameter β. The posterior distribution of β will tell us if such assumption is indeed substantiated by the data. The interpretation of the parameter is particularly interesting if we note that $\beta = 0$ corresponds to X_j and Y_j being independent, while $\beta = 1$ corresponds to X_j and $Y_j - X_j$ being independent, conditionally on the other model parameters. We choose an uninformative prior for the correlation parameter by assuming $\beta \sim U(-\infty, +\infty)$, and the Gibbs sampler with an added step still converges to a stable set of posterior estimates for all random variables.

We are being extremely liberal, here, letting each of the parameters have their own diffuse prior. We will see in the application section that the result is indeed a Markov Chain converging to a stationary distribution, but the posterior distribution of temperature change appears extremely sensitive to the actual location of the GCMs' projections in a way that is not easily supported by the scientific understanding behind this application. The difficulty is illustrated by Figure 1 where we show some posterior distributions of $\nu - \mu$ based on actual climate models data, in comparison to two alternative approaches introduced later. The current approach (dashed line in Figure 1) often shows multimodal posterior distributions, with modes sometimes corresponding to single climate models' projections (see also the original analysis in Tebaldi et al. (2004, 2005), which is based on an older set of experiments, prepared for the third assessment report of the IPCC, published in 2001). However, we would expect change in temperature over a given time period to have a smooth distribution. This is the result of dealing with a limited set of climate models, that happen to populate the interval between the extremes of their range very sparsely, and of a statistical model that attributes an uneven set of weights to the participating members of the ensemble, as the λ_j parameters may actually be interpreted according to the form of $\tilde{\mu}$ and $\tilde{\nu}$.

Let us then formalize the fact that we do not expect such an uneven distribution of weights among the different models' precision parameters in this ensemble. The natural way to do that is to use a hierarchical approach, and hypothesize that all λ_js are samples from the same prior distribution with parameters that we estimate in turn, as in

$$\lambda_j \sim Ga(a_\lambda, b_\lambda)$$

with a_λ, b_λ sampled from a Gamma prior, $Ga(a^*, b^*)$.

This simple extra layer has the effect of smoothing out the estimates of λ_js and, accordingly, the shape of the posterior for $\nu - \mu$. Thus, sensitivity of the estimates to the precise location of X_j and Y_j values is significantly diminished as well, as Figure 1 demonstrates by comparing posterior estimates by the first model (dashed lines) to its hierarchical variation (dotted lines), for the same values of the GCMs' $Y_j - X_j$. The only operational consequence is that the joint posterior for this model is no longer estimated by a simple Gibbs sampler, but a Metropolis-Hastings step needs to handle the iterative simulation of a_λ, b_λ (see Appendix B1). This modification of the univariate approach was introduced in Smith et al. (2009), but there too the application uses the older data set from IPCC 2001.

4.2 Borrowing strength by combining projections over multiple regions

It is sensible to assume that the characteristics of each GCM, which in our development so far were represented by the parameters λ_j, θ and β, could be estimated by gathering information on the model's simulations over a set of regions, rather than just a single one. Consider then X_{ij} and Y_{ij}, which in addition to representing different models $j = 1, ..., M$, also represent different regions $i = 1, ..., R$. We also have a set of X_{i0}, the current observed mean temperature in region i which is an estimate of the true current mean temperature with standard deviation $\lambda_{0i}^{-1/2}$.

The likelihood model is a natural extension of the univariate case, where

$$\begin{aligned}
X_{i0} &\sim N[\mu_0 + \zeta_i, \lambda_{0i}^{-1}], \quad (\lambda_{0i} \text{ known}), && (2)\\
X_{ij} &\sim N[\mu_0 + \zeta_i + \alpha_j, (\eta_{ij}\phi_i\lambda_j)^{-1}], && (3)\\
Y_{ij} \mid X_{ij} &\sim N[\nu_0 + \zeta_i' + \alpha_j' + \beta_i(X_{ij} - \mu_0 - \zeta_i - \alpha_j), (\eta_{ij}\theta_i\lambda_j)^{-1}]. && (4)
\end{aligned}$$

We choose joint prior densities as in:

$$\begin{aligned}
\mu_0, \nu_0, \beta_i, \beta_0, \zeta_i, \zeta_i' &\sim U(-\infty, \infty), && (5)\\
\theta_i, \phi_i, \psi_0, \theta_0, c, a_\lambda, b_\lambda &\sim Ga(a, b), && (6)\\
\lambda_j \| a_\lambda, b_\lambda &\sim Ga(a_\lambda, b_\lambda) && (7)\\
\eta_{ij} \| c &\sim Ga(c, c), && (8)\\
\alpha_j | \psi_0 &\sim N[0, \psi_0^{-1}], && (9)\\
\alpha_j' \mid \alpha_j, \beta_0, \theta_0, \psi_0 &\sim N[\beta_0\alpha_j, (\theta_0\psi_0)^{-1}], && (10)
\end{aligned}$$

all mutually independent unless explicitly indicated otherwise.

This approach is presented in Smith et al. (2009). As can be noticed, there are a few substantial changes from the model of Section4.1: new parameters α_j and α_j', ζ_i and ζ_i' are introduced in the mean components of the likelihood, and the variances in (3) and (4) have a more complex structure.

The parameters α_j and α_j' represent model biases. They are model-specific quantities, but they are constant across regions, thus introducing correlation between projections from the same model in different regions. By (10) we introduce the possibility of a correlation between α_j and α_j', i.e, between the bias in the current period of the simulation and the future period, using the regression parameter β_0. Similarly, ζ_i and ζ_i' represent region-specific mean components. The idea is that different regions will tend to warm differently, and this region effect will be common to all GCMs. We treat these two sets of mean components in a fundamentally different manner by imposing on the α_j, α_j' Gaussian priors with mean zero, while letting the priors for ζ_i, ζ_i' be improper priors, $U(-\infty, +\infty)$. It is the difference between allowing for significantly different region effects, unconstrained by one another, and imposing a shrinkage effect across model biases according to the expectation that they should cancel each other out.

The variances in (3) and (4) are modeled as three multiplicative factors, one model-specific, one region-specific, one allowing for an interaction. In the case where $\eta_{ij} \equiv 1$ (a limiting case of (8) in which $c \to \infty$) the variance factorizes, with λ_j representing a model reliability and either ϕ_i or θ_i a region reliability. (We note here that two different parameters for current and future simulations of a given region have the same effect of what the model in Section 4.1 accomplished by using the parameter θ.) Compared with fitting a separate univariate model to each region, there are many fewer parameters to estimate, so we should get much improved precision. However there is a disadvantage to this approach: if model A has higher reliability than model B for one region, then it will for all regions (and with the same ratio of reliabilities). This is contrary to our experience with climate models, where it is often found that a model's good performance in one region is no guarantee of good performance in another. The parameter η_{ij}, then, may be thought of as an interaction parameter that allows for the relative reliabilities of different models to be different in different regions. As with the other reliability parameters, we assume a prior gamma distribution, and there is no loss of generality in forcing that gamma distribution to have mean 1, so we set the shape and scale parameters both equal to some number c that has to be specified. However, if c is too close to 0, then the model is in effect no different from the univariate model, allowing the variances for different models in different regions to be completely unconstrained. Our recommendation is that c should be chosen not too close to either 0 or ∞, to represent some reasonable judgment about the strength of this interaction term. A different solution will be to let c be a random variable, and let the data give us an estimate of it. We can give c a diffuse prior distribution, as with $c \sim Ga(0.01, 0.01)$.

The Gibbs sampler, with a Metropolis Hastings steps to sample values for c, is described in detail in Appendix B2. Together with the rest of the discussion of the multivariate approach, it was first detailed in Smith et al. (2009).

4.3 A bivariate, region-specific model

The statistical treatment described in the preceding section can be applied separately to temperature and precipitation means. The latter does not need to be logarithmically transformed. In our experience, since the data points are seasonal, regional and multidecadal means, the Gaussian likelihood model fits the data without need for transformations. Many studies of impacts of climate change are predicated on the availability of joint projections of temperature and precipitation change. A warmer, wetter future is very different from a warmer, drier one if you are dealing with adaptation of agricultural practices, or water resources management (Groves et al., 2008). Here we propose a joint model for the two climate variables at the individual regional level that we first introduced in Tebaldi and Sansó (2008). Rather than simply extending the set up of Section 4.1, we consider decadal means covering the entire observed record and the entire simulation length. These additional data points will help estimate the correlation coefficients. We will model trends underlying these decadal mean time series, and estimate a correlation parameter between temperature and precipitation, once the trend is accounted for.

Here are the new assumptions:

- The vector of observed values O_t is a noisy version of the underlying temperature and precipitation process, with correlated Gaussian noise (we estimate the correlation from the data, through the estimation of the parameter β_{xo}).
- The true process is piecewise linear, for both temperature and precipitation. We fix the "elbow" at year 2000, which may allow for future trends steeper than the observed ones. Of course a slightly more general model could use a random change point approach, but given the coarse resolution of our time dimension and the limited amount of data at our disposal we choose to fix the change point.
- The model output X_{jt} is a biased and noisy version of the truth. We assume an additive bias and a bivariate Gaussian noise.
- We expect the model biases to be related across the population of models, i.e., we impose a common prior, and we estimate its mean parameter, so that we may determine an overall bias for the ensemble of model simulations, different from zero.

In the notation, superscripts T and P refer to the temperature and precipitation components of the vectors. Thus, the likelihood of the data is:

$$\begin{aligned} O_t^T &\sim N(\mu_t^T; \eta^T) \qquad \text{for } t = 1, \ldots, \tau_0 \\ O_t^P | O_t^T &\sim N(\mu_t^P + \beta_{xo}(O_t^T - \mu_t^T); \eta^P) \qquad \text{for } t = 1, \ldots, \tau_0 \\ \text{where } \beta_{xo} &\sim N(\beta_0, \lambda_o), \\ X_{jt}^T &\sim N(\mu_t^T + d_j^T; \xi_j^T) \qquad \text{for } t = 1, \ldots, \tau^* \text{ and } j = 1, \ldots, M \\ X_{jt}^P | X_{jt}^T &\sim N(\mu_t^P + \beta_{xj}(X_{jt}^T - \mu_t^T - d_j^T) + d_j^P; \xi_j^P) \\ &\qquad\qquad \text{for } t = 1, \ldots, \tau^* \text{ and } j = 1, \ldots, M. \end{aligned} \tag{11}$$

In Equations (11) we specify bivariate normal distributions for O_t and X_{jt} using conditionality. After accounting for the underlying trends and biases terms, $\beta_{x1}, \ldots, \beta_{xM}$ are used to model the correlation between temperature and precipitation in the climate model simulations, while β_{xo} is fixed at the value estimated through the observed record. Also in the likelihood of the observations, η^T and η^P are fixed to their empirical estimates.

The time evolution of the *true climate process* $\mu_t' = (\mu_t^T, \mu_t^P)$, consists of a piecewise linear trend in both components:

$$\begin{pmatrix} \mu_t^T \\ \mu_t^P \end{pmatrix} \equiv \begin{pmatrix} \alpha^T + \beta^T t + \gamma^T (t - \tau_0) \mathcal{I}_{\{t \geq \tau_0\}} \\ \alpha^P + \beta^P t + \gamma^P (t - \tau_0) \mathcal{I}_{\{t \geq \tau_0\}} \end{pmatrix}. \tag{12}$$

The priors for the parameters in Model (11) are specified hierarchically by assuming that

$$\beta_{xj} \sim N(\beta_0, \lambda_B),$$

$$d_j^T \sim N(a^T; \lambda_D^T),$$

$$d_j^P \sim N(a^P; \lambda_D^P)$$

for $j = 1, \ldots, M$

$$\xi_j^T \sim Ga(a_{\xi^T}, b_{\xi^T})$$

and

$$\xi_j^P \sim Ga(a_{\xi^P}, b_{\xi^P}).$$

λ_o is fixed to a value estimated on the basis of the observed record.

All the other quantities are assigned uninformative priors:

$$\beta_0, a^T, a^P \sim U(-\infty, +\infty)$$

and

$$\lambda_B, \lambda_D^T, \lambda_D^P, a_{\xi^T}, b_{\xi^T}, a_{\xi^P}, b_{\xi^P} \sim Ga(g, h),$$

where $g = h = 0.01$. Similarly, for the parameters in (12), we assume

$$\alpha^T, \beta^T, \gamma^T, \alpha^P, \beta^P, \gamma^P \sim U(-\infty, +\infty).$$

We are assuming that each climate model has its own precision in simulating the true temperature and precipitation time series, but we impose common priors to ξ_j^T and ξ_j^P $\forall j$, whose parameters are in turn estimated by the data. As we discussed in Section 4.1, this choice produces more robust estimates of the relative precisions of the different GCMs, not overly sensitive to small perturbations in the GCM trajectories.

The model-specific bias terms d_j^T, d_j^P are assumed constant over the length of the simulation. They model systematic errors in each GCM simulated variable. All the GCM biases for temperature, like all GCM biases for precipitation, are realization from a common Gaussian distribution, whose mean (a^T or a^P), as mentioned, may be different from zero, when the set of model trajectories is distributed around the truth non-symmetrically. We do not expect a systematic behavior across models when it comes to precipitation versus temperature biases, that is, we do not expect that models having relatively larger temperature biases would show relatively larger precipitation biases, so we do not model a correlation structure between d_j^T, d_j^P. In fact, this

correlation structure, if there at all, would not to be identifiable/separable from the correlation modeled through $\beta_{xo}, \beta_{x1}, \ldots, \beta_{xM}$, given the configuration of the present dataset. Notice that the correlation coefficients, β_{xo} and β_{xj} also have a common mean, β_0 possibly different from zero and that will be heavily influenced by the value of the observed correlation coefficient, β_{xo}. All the remaining parameters of the model have non-informative, conjugate distributions. Notice that we use improper priors for the location parameters of the Gaussian distributions and linear regression parameters in the correlation structure and in the trend structure, and proper but diffuse priors for the precision parameters and as hyper-priors of the $\xi_{\cdot}$ parameters. The likelihood and priors form a conjugate model, and as before a Gibbs sampler can be programmed to explore the posterior distributions for this model, with a Metropolis-Hastings step used to generate sample values for $a_{\xi^T}, b_{\xi^T}, a_{\xi^P}, b_{\xi^P}$. See Appendix B3 details of the Markov chain Monte Carlo implementation, or the original article Tebaldi and Sansó (2008).

5 Validating the statistical models

Climate predictions are a fairly safe business to be in: validation of the forecast comes about no sooner than every 10 or 20 years! We have already commented on how this situation sets climate forecasting well apart from weather forecasting, where statistical calibration of numerical models can be fine tuned continuously, using diagnostics of their performance coming at the steady and frequent rate of every six hours or so. The work by Adrian Raftery, Tilmann Gneiting and colleagues (Raftery et al. (2005) and references therein) has set the standard for Bayesian statistical methods brought to bear on ensemble weather forecast calibration and validation. In our application, we are also forecasting conditionally on a specific scenario of greenhouse gases emissions, which is always an idealized scenario. This discussion shows why we are not going to rely on observations to validate our statistical approaches. Rather, we validate our statistical assumptions by performing cross-validation. In all cases but for the simplest univariate model introduced first, we can estimate a posterior predictive distribution for a new model's trajectory or change. We can therefore compare the left-out model projections with their posterior-predictive distribution. In both cases we can do this exercise over all possible models, regions and – separately or jointly – both climate variables, temperature and precipitation averages. Statistical theory provides us with null hypotheses to test about the expected distribution of the values of $P(x*)$ where $x*$ is the left-out value and $P()$ is the posterior predictive cumulative distribution function of that quantity. We expect the values of $P(x*)$ obtained by performing the cross-validation exercise across models and across regions to be a sample from a uniform distribution. In all cases, after conducting the tests, we do not reject the hypothesis more frequently that what would be expected as a result of multiple testing. In the remainder of this section we detail the procedure of the cross-validation exercise, for each of the statistical models proposed. More details and actual results from the cross validation exercises can be found in Smith et al. (2009); Tebaldi and Sansó (2008).

5.1 Univariate model with hyperprior on precision parameters

The predictive distribution can be calculated under the assumption that the climate models are exchangeable. Conditionally on the hyperparameters $\mu,\ \nu,\ \beta,\ \theta,\ a_\lambda$ and b_λ, the distribution of $Y_{M+1} - X_{M+1}$ can be derived derived from conditioning on $\lambda_{M+1} \sim Ga(a_\lambda, b_\lambda)$, since, then, $Y_{M+1} - X_{M+1}|\lambda_{M+1} \sim N(\nu - \mu, \{(\beta-1)^2 + \theta^{-1}\}\lambda_{M+1}^{-1})$.

The conditional distribution should then be convolved with the joint posterior distribution of $(\mu, \nu, \beta, \theta, a_\lambda, b_\lambda)$, to obtain the full posterior predictive distribution. In practice we can carry out this integration within the Gibbs-Metropolis algorithm by

1. sampling at each step n the hyperparameter values $a_\lambda^{(n)}, b_\lambda^{(n)}, \nu^{(n)}, \mu^{(n)}, \beta^{(n)}, \theta^{(n)}$, corresponding to one draw from their joint posterior distribution. A draw of a random $\lambda_{j,n} \sim Ga(a_\lambda^{(n)}, b_\lambda^{(n)})$ can be generated and the statistic

$$U_j^{(n)} = \Phi\left\{\frac{Y_j - X_j - \nu^{(n)} + \mu^{(n)}}{\sqrt{\{(\beta_x^{(n)} - 1)^2 + {\theta^{(n)}}^{-1}\}(\lambda_{j,n})^{-1}}}\right\}$$

calculated.

2. Over all n iterations we can compute U_j, the mean value of $U_j^{(n)}$, representing an estimate of the predictive distribution function, evaluated at the true $Y_j - X_j$. If the statistical model is consistent with the data, U_j should have a uniform distribution on $(0, 1)$.
3. By computing U_j for each region, and each GCM we have a set of test statistics that we can evaluate for discrepancies, applying tests of fit to evaluate the hypothesis that the values are samples from a uniform distribution

5.2 Multivariate model, treating multiple regions at once

The procedure for cross validation in this model is very similar to what was just described. In this setting we do cross-validation of the variable $Y_{ij} - X_{ij}$, model j's projected change in region i. The full conditional, predictive distribution of this quantity is

$$Y_{ij} - X_{ij} \mid \text{rest} \sim N\left(\nu_0 - \mu_0 + \zeta_i' - \zeta_i + \alpha_j' - \alpha_j, \frac{1}{\eta_{ij}\lambda_j}\left\{\frac{(\beta_i - 1)^2}{\phi_i} + \frac{1}{\theta_i}\right\}\right).$$

The implementation of the cross validation exercise within the Gibbs-Metropolis algorithm is as follows:

1. We leave model j out and we run the Gibbs-Metropolis simulation.
2. After burn-in, at every step n, on the basis of the set of parameters currently sampled we generate corresponding values of $\lambda_j^{(n)}$, $\alpha_j^{(n)}$, $\alpha_j^{\prime(n)}$ and $\eta_{ij}^{(n)}$ as

$$\begin{aligned}
\lambda_j^{(n)} &\sim Ga\left(a_\lambda^{(n)}, b_\lambda^{(n)}\right) \\
\alpha_j^{(n)} &\sim N\left(0, \frac{1}{\psi_0^{(n)}}\right) \\
\alpha_j^{\prime(n)} &\sim N\left(\beta_0^{(n)}\alpha_j^{(n)}, \frac{1}{\psi_0^{(n)}\theta_0^{(n)}}\right), \\
\eta_{ij}^{(n)} &\sim Ga\left(c^{(n)}, c^{(n)}\right).
\end{aligned}$$

3. From these values we compute the statistic

$$U_{ij} = \frac{1}{N}\sum_{n=1}^{N}\Phi\left[\frac{Y_{ij} - X_{ij} - (\nu_0^{(n)} - \mu_0^{(n)}) - (\zeta_i^{\prime(n)} - \zeta_i^{(n)}) - (\alpha_j^{\prime(n)} - \alpha_j^{(n)})}{\sqrt{(\lambda_j^{(n)}\eta_{ij}^{(n)})^{-1}\left\{(\phi_i^{(n)})^{-1}(\beta_i^{(n)} - 1)^2 + (\theta_i^{(n)})^{-1}\right\}}}\right]. \qquad (13)$$

   by generating a sample value at each iteration and averaging them at the completion of the simulation.
4. As with the univariate analysis, we then perform various goodness of fit tests on the statistics U_{ij}. If the model is a good fit, they should be consistent with independent draws from the uniform distribution on $[0, 1]$.

5.3 Bivariate model of joint temperature and precipitation projections

After leaving each individual GCM's trajectories out in turn, we simplify the validation of the bivariate time series model by computing three marginal bivariate predictive distributions, one for current climate (defined as the bivariate distribution of average values of temperature and precipitation for the period 1981-2000), one for future climate (defined as the corresponding distribution for average values over the period 2081-2100) and one for climate change (defined as the joint distribution of the temperature and precipitation *differences* between the two same periods). We can then compute the two sets of pairs $(U_1 = P_T(X_*^T = x_*^T), U_2 = P_{P|T}(X_*^P = x_*^P))$ for both current and future time windows, and the pair $(U_1 = P_T(\Delta X_*^T = \Delta x_*^T), U_2 =$

$P_{P|T}(\Delta X_*^P = \Delta x_*^P))$ where the first univariate distribution function is simply the marginal predictive distribution of temperature change, while the second distribution function is the predictive distribution of precipitation change, conditional on the corresponding simulated temperature change. Finally we test the null hypothesis that the pairs of (z_{1j}, z_{2j}) for $j = 1, \ldots, M$ are independent and identically distributed random variates, sampled from uniform distributions on the $(0, 1)$ interval. The form of the posterior predictive and the way we simulate the U statistics is similar to what we described in the previous subsections, with all the parameters being sampled from the estimated value of the hyperparameters within the Gibbs sampler, and the conditional marginal predictive distributions having a manageable Gaussian form.

5.4 U-statistics and goodness-of-fit tests

Following Smith et al. (2009), in all cases described in Sections 5.1 through 5.3, the sets of U_{ij} (or Z_j) can be tested for goodness of fit with respect to a uniform distribution and independence by traditional tests like Kolmogorov-Smirnov, Cramér-von Mises and Anderson-Darling. Smith et al. (2009) also proposes the use of a correlation test where the ordered values of $U_{ij},\ j = 1, \ldots, M$ are correlated to the sequence $1/(M+1), 2/(M+1), \ldots, M/(M+1)$ and large values of the correlation coefficient (or small values of $1 - cor$) indicate a close fit. In most cases the goodness-of-fit tests result in acceptance of the null hypothesis that the U_{ij} or Z_j are independent within each region i.

6 Application: the latest model projections, and their synthesis through our Bayesian statistical models

The website of PCMDI provides instructions for the download of model output from all the GCMs that have contributed their simulations to the IPCC-AR4 effort. We choose a set of experiments run under a scenario of greenhouse gas emissions that can be thought of as a "business-as-usual" scenario, where concentrations of greenhouse gases increase over this century at a rate similar to what is being emitted currently, worldwide. We extract average temperature and precipitation projections from seventeen models. We area-average their output over standard subcontinental regions that have been used by IPCC and many studies in the literature (Giorgi and Francisco, 2000), see Figure 2. We consider two seasonal averages, December through February (DJF) and June through August (JJA). For the models treating temperature or precipitation separately we also average twenty-year periods, 1980-1999 as the current climate averages and 2080-2099 as the future. For the model that estimates the joint distribution of temperature and precipitation we consider time series of fifteen decadal averages covering the period 1950-2100. We compute the joint posterior distribution of all random parameters according to the univariate (one region at a time, with or without hyperprior over the precision parameters), multivariate (all regions at the same time) and bivariate (temperature and precipitation jointly) models. We then proceed to display and compare results for some quantities of interest. Most importantly, we consider changes in temperature (and precipitation) for all the regions, across the different statistical treatment. We also show posterior distributions of other parameters of interest, like model biases and interaction effects.

6.1 Changes in temperature, univariate and multivariate model

The eight panels of Figure 1 compare the posterior distribution of temperature change $\nu - \mu$ for the two univariate models (dashed and dotted curves) and $\nu_0 + \zeta_i' - \mu_0 - \zeta_i$ for the multivariate model (solid line) for a group of four regions in DJF and a different group of four regions in JJA. We have already mentioned in Section 4.1 that the first model, where the GCM-specific precision parameters are each a sample from a diffuse Gamma prior, suffers from an uneven distribution of "weight" among GCMs' projections. The multimodality of some of the PDFs in Figure 1 is an indication of this behavior. And even more generally, the fact that these PDFs are often shifted significantly from the location of the ensemble mean (indicated by a cross along the x-axis, whereas the individual GCMs are marked by dots along the basis of the curves) is the effect of

some of the GCMs "stealing the show", i.e., being attributed a much larger precision than others, where we would not expect such difference in relative importance among this family of state-of-the-art climate models. The univariate model with a common prior over the λ_js produces smoother, better centered PDFs that are not significantly different from those produced by the multivariate model (comparing the dashed and solid lines). The two series of boxplots (one for DJF, to the left and one for JJA, to the right) in Figure 3 offer a comparison of the three distributions for each region/season combination. For each region, the first boxplot from the bottom shows interquartile range, median and 5th-95th quantiles of the posterior distribution estimated by the unconstrained version of the univariate model, the second and third boxplot show the extent of the posterior for the univariate model with common prior for the λ_js and the multivariate model. These two sets of boxplots confirm the results of Figure 1, with the position of the first boxplot in each triplet often shifted away from the other two, which are more similar to one another. This display of all the 22 region is also indicative of the large differences in the amount of warming across regions. This justifies the choice of modeling region-specific parameters in the mean component of the likelihood of the multivariate model, ζ_i, ζ_i', with a Uniform prior over the real line that does not have a shrinkage effect towards zero.

6.2 Changes in temperature, multivariate and bivariate model

We now take the same eight region/season combinations of Figure 1 and compare, in Figure 4, the posterior distribution of temperature change for the multivariate model, $\nu_0 + \zeta_i' - \mu_0 - \zeta_i$, (solid line) to the marginal posterior distribution of the temperature change signal from the bivariate model (dotted line). We also show the posterior predictive distribution of a new unbiased GCM's projection of temperature change (dashed line) in the same panels. We note immediately that the posterior distribution from the bivariate model is much narrower than the posterior from the multivariate model. This is no surprise, given that we are using a much richer dataset (6 observed decades, 15 simulated decades for each of seventeen GCMs) and we are hypothesizing a piecewise linear function of time as the trajectory of mean temperature, thus we are in fact estimating only three parameters α^T, β^T and γ^T. Note that the joint posterior distribution from the bivariate model is a probabilistic representation of the entire trajectory of temperature (ad precipitation) over the decades (μ_t^T) but the availability of the Gibbs sample makes it straightforward to compute any deterministic function of it (in this case the difference between two bi-decadal means). The other obvious feature in the majority of these panels is that the posterior distributions from the bivariate model are significantly shifted from the center of the multivariate model's PDFs. Remember that the bivariate model is not simply "smoothing" sets of GCM "snapshots" in terms of multidecadal averages, but rather fitting a trend to their whole trajectories, anchoring its estimation to the observed series. We are thus comparing two different definitions of "temperature change", at least in terms of how the statistical model is estimating it. Likely, some of the models produce steeper trends than the underlying trend estimated from the entire ensemble, thus filling the distribution of the "snapshots" estimate to the right of the trend estimates', and shifting the mean accordingly. This feature may be scenario dependent, with lighter emission scenarios (forcing less of an acceleration in the trends) facilitating a better agreement between the two methods' results. The dashed curves in each figure are posterior predictive distribution of a new GCM's projection. The signal of temperature (change) underlying truth and model simulation is an abstract concept. Even the observations are a noisy representation of this signal. We could be justified then if we thought of a model trajectory as a possible future path for our climate, and accordingly we represented the uncertainty in this future projection by the posterior predictive distribution of a new GCM, whose width is of the same order of magnitude as the range of model projections, rather than being an inverse function of the square root of the number of data points, as the posterior distribution width is. In Figure 5 we complete the representation of probabilistic projections of climate change by showing contours of the posterior (tighter set of contours in each panel) and posterior predictive (wider set of contours in each panel) for the same set of region/season pairs of Figures 1 and 4.

6.3 Other uncertain quantities of interest, their full probabilistic characterization

Even the simple univariate treatment, after imposing the regularizing common prior over the λ_js, offers additional information besides the posterior PDFs of temperature change that may be used in applications of impact analysis. Posterior means of λ_js can be regarded as optimal estimates of model reliabilities, and utilized as "weights" for some further analysis of climate change impacts. For example, Fowler et al. (2007) shows how the normalized set of reliabilities can be used for a Monte Carlo simulation of weather scenarios at regional scales by stochastic weather generator models "driven" by different GCMs. The larger the GCM reliability, the larger the number of weather scenarios one wants to generate according to that GCM, and the total number of Monte Carlo simulations from the weather generator may be divided up proportionally to the normalized weights. In Figure 6 we show boxplots of the 16 GCM-specific λ_j's for the 8 region/season combinations of Figures 1 and 4, according to the univariate treatment, displaying them in each panel in increasing order from left to right according to their posterior means. From the figure we note how the distribution of "weight" among GCMs is very balanced, with no disproportionate attribution to a small subset of models. It is true however that different models gain different ranking depending on the region and season combination. This is consistent with the shared opinion among climate scientists and modelers that no model outperforms all others in every respect, different models showing different strengths in the simulation of regional climates, seasonal processes and so on.

We may be interested in evaluating the posterior estimates of the model-specific mean factors in the multivariate treatment, α_j and α'_j, as they may be interpreted as global average biases that a specific GCM imposes on its simulation of temperature. Figure 7 shows boxplots of the two sets of parameters for the simulation of temperature in DJF, as an example. The models have been ordered with respect to the value of the posterior mean of the respective parameter in the current climate simulation (α_j). As can be assessed by the two series of boxplots, there are models with biases that are significantly different from zero, mostly with negative (cold) biases in the left plot. Many more GCMs show significant biases in the future part of the simulation, on both sides of the zero line.

Another interesting question that can be answered through the computation of the marginal posterior distribution of a set of parameters in the multivariate model regards the form of the precision (or variance) components in the likelihood of the GCM simulated temperatures. The parameters η_{ij}s are let free to assume values significantly different from one, if an interaction between model-specific and region-specific behavior is deemed necessary to fit the dataset at hand. As Figure 8 demonstrates, however, by showing the distributions of $\eta_{i.}$ and $\eta_{.j}$, i.e. by combining all sample values in the Gibbs simulation across models and across regions, the distribution of these parameters is always centered around the value one, lending support to a model that factorizes the variance component into two terms without interaction (boxplots of all 22*16 region-and-model-specific parameters do not reveal any behavior inconsistent with this averages).

Last, we show posterior PDFs of the γ^T, γ^P trend parameters in the bivariate model. Recall that these parameters are introduced in order to estimate a possible change in the trend underlying the observed and modelled time series, after the year 2000. As can be seen from Figure 9 for the usual region/season combinations (listed along the vertical axis), all the parameters of the temperature trends are significantly different from zero (the zero line does not even appear in the plot). The trend parameters estimated for precipitation instead show a whole range of behaviors, with some of the distributions straddling the zero line, some indicating a significant decrease in the rate of change (with the distributions lying for the most part to the left of the zero line), some, at the contrary, estimating an increase in the trend.

7 Further Discussion

We have presented a series of statistical models aiming at combining multiple climate model simulations. The goal is to arrive at a rigorous characterization of the uncertainty in model projections, quantifying it into probability distributions of climate change. More precisely, we

determine PDFs of average temperature and precipitation change at seasonal, multi-decadal and regional scales, that can be used for further development as input to impact models in vulnerability and adaptation studies.

The challenge of synthesizing multiple model projections stems from the idiosyncratic nature of the data sample. GCMs are arguably not independent, they do not span the entire range of uncertainty, being fairly conservative guesses of the future trajectory of our climate system, and they may have systematic errors (biases) in their simulation of climate, that do not necessarily cancel out in an overall mean.

Given access to the recently organized, rich archives of GCM data, we are however able to start making sense of these ensembles of simulations. In this chapter we have shown how increasingly complex statistical models can account for features of the data like correlations between current and future simulation errors, biases, differential skill in different periods of the simulation or different regions of the globe by a given model. We have also shown how the difficulty of validating climate predictions may be side-stepped thanks to the calculation of posterior predictive distributions that open the way to cross validation exercises.

We have dedicated some time to discuss two ways of characterizing the uncertainty of these climate change experiments. One approach cares especially about the common central tendency of these models, and the uncertainty around this estimated central tendency. Thanks to the number of models available and the observed record that "anchors" our estimates of the true signal of climate, the central tendency is very precisely estimated, within a narrow band of uncertainty. The reality of climate change, however, may look more like one of these model trajectories than like this abstract climate signal. If we subscribe to this view, then a more consistent representation of our uncertain future is naturally found in the posterior predictive distribution, after integrating out all the uncertain parameters of the Bayesian model. The posterior predictive width is of the same order of magnitude as the range of model projections, much wider than the posterior estimate of the climate (change) signal's uncertainty and may look to expert eyes as a more realistic representation of the uncertainty at hands.

There are many ways by which the kind of analyses presented here may be carried forward. Two in particular seem germane to our approach. Furrer et al. (2007) chose to model the entire high resolution fields of variables that constitute the original output of these models, after registering them to a common grid. This approach has the advantage of providing a representation of projections that is very familiar to the climate science community, in the form of detailed maps of change, with the added dimension of a probabilistic characterization of their likely ranges. Rougier (2007) uses a set of experiments aimed at perturbing the parameterization of a specific climate model, the Hadley Center GCM. The work therefore represents a natural complement to ours, where intra-model uncertainties are fully characterized. Ideally, it could be incorporated in our framework, in order to treat within model and between model variability in concert.

We think of two main avenues for further development of multi-model analysis. The first is the representation of models' dependencies, which would protect us from the over-optimistic narrowing of the uncertainty with the increasing number of members in these ensembles. It seems clear that not all models add an independent piece of information to the ensemble, while the treatment of each one as independent of the other makes the estimate of the central tendency increasingly more precise with the increasing number of GCMs.

A second direction should lead to the exploration of models' performance within current climate simulations and how it correlates to models' reliability in the future part of their simulation. In our treatment, a coarse metric of performance was implicitly found in the difference between models' average projections in the part of the simulation that overlaps with observed records and the corresponding observed climate average. There are potentially many alternative metrics of performance that could be relevant for a model's ability to simulate future climate reliably, but the question is open, as is that of how to incorporate this information into our statistical models.

Appendices

A. Broader Context and Background

All of Bayesian statistics derives ultimately from the formula

$$\pi(\theta \mid x) \quad = \quad \frac{\pi(\theta) f(x \mid \theta)}{\int_\Theta \pi(\theta') f(x \mid \theta') d\theta'}. \tag{14}$$

Here:

- x is an observation vector, lying in some sample space $\mathcal{X}$;
- θ is a vector of unknown parameters, lying in a parameter space Θ;
- $\pi(\theta)$ represents the prior distribution. If Θ is discrete, this is a probability mass function; if Θ is continuous, this is a probability density function.
- $f(x \mid \theta)$ is the likelihood function; that is, the conditional distribution of x given θ. This is a probability mass function if $\mathcal{X}$ is discrete and a probability density function is $\mathcal{X}$ is continuous.

In the following discussion, we first review simple applications of (14) using *conjugate priors*; then we describe extensions to *hierarchical models*; finally we describe computational methods via the *Gibbs Sampler* and the *Hastings-Metropolis Sampler*, which the two best known examples of Markov chain Monte Carlo (MCMC) algorithms. We illustrate these concepts with reference to the models for climate change projections that form the bulk of this chapter.

A1. Conjugate Priors

The simplest examples of (14) have the following structure: both $\pi(\theta)$ and $f(x \mid \theta)$ have a parametric structure, which is such that $\pi(\theta \mid x)$ has the same parametric structure as that of $\pi(\theta)$. In that case, the prior is known as a *conjugate prior*.

A simple example is the Beta prior for a Binomial distribution. In this case, $\mathcal{X} = \{0, 1, 2, ..., n\}$ for known n, $\Theta = [0, 1]$, and $f(x \mid \theta) = \frac{n!}{x!(n-x)!}\theta^x(1-\theta)^{n-x}$ (the binomial distribution with n trials and probability of success θ). Consider the prior density

$$\pi(\theta) \quad = \quad \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}, \quad \theta \in [0, 1], \tag{15}$$

the Beta distribution with constants $a > 0,\ b > 0$. Then, after cancelling common terms in the numerator and denominator, (14) becomes

$$\pi(\theta \mid x) \quad = \quad \frac{\theta^{a+x-1}(1-\theta)^{b+n-x-1}}{\int_0^1 (\theta')^{a+x-1}(1-\theta')^{b+n-x-1} d\theta'}. \tag{16}$$

But the fact that (15) is a proper probability density for all $a > 0,\ b > 0$ implies

$$\int_0^1 (\theta')^{a+x-1}(1-\theta')^{b+n-x-1} d\theta' \quad = \quad \frac{\Gamma(a+x)\Gamma(b+n-x)}{\Gamma(a+b+n)}.$$

Hence (16) reduces to

$$\pi(\theta \mid x) \quad = \quad \frac{\Gamma(a+b+n)}{\Gamma(a+x)\Gamma(b+n-x)}\theta^{a+x-1}(1-\theta)^{b+n-x-1}.$$

Thus, the posterior distribution is of the same parametric form as the prior (15), but with a and b replaced by $a + x$ and $b + n - x$. This is the key idea of a conjugate prior.

A second example of a conjugate prior is the *gamma-normal prior*. Consider the sequence of conditional distributions

$$\lambda \quad \sim \quad Ga(a, b), \tag{17}$$

$$\mu|\lambda \quad \sim \quad N[m, (k\lambda)^{-1}], \tag{18}$$

$$x_1, ..., x_n|\mu, \lambda \quad \sim \quad N[\mu, \lambda^{-1}] \text{ (independent)}. \tag{19}$$

In this case the data equation (19) consists of normal observations with unknown mean and variance, but for notational convenience, we have written the variance of the normal distribution as λ^{-1} rather than the more conventional σ^2. The parameter λ is often called the precision.

In this case, equations (17) and (18) together define the joint prior of (λ, μ), with joint density

$$\frac{b^a \lambda^{a-1} e^{-b\lambda}}{\Gamma(a)} \cdot \sqrt{\frac{k\lambda}{2\pi}} \exp\left\{-\frac{k\lambda}{2}(\mu - m)^2\right\} \propto \lambda^{a-1/2} \exp\left[-\lambda\left\{b + \frac{k}{2}(\mu - m)^2\right\}\right]. \qquad (20)$$

Now consider the prior $\times$ likelihood

$$\frac{b^a \lambda^{a-1} e^{-b\lambda}}{\Gamma(a)} \cdot \sqrt{\frac{k\lambda}{2\pi}} \exp\left\{-\frac{k\lambda}{2}(\mu - m)^2\right\} \cdot \left(\frac{\lambda}{2\pi}\right)^{n/2} \exp\left\{-\frac{\lambda}{2}\sum(x_i - \mu)^2\right\}. \qquad (21)$$

Noting the side calculation

$$k(\mu - m)^2 + \sum(x_i - \mu)^2 \quad = \quad \frac{kn}{k+n}(m - \bar{x})^2 + \sum(x_i - \bar{x})^2 + (k+n)\left(\mu - \frac{km + n\bar{x}}{k+n}\right)^2$$

and defining $\bar{x} = \frac{1}{n}\sum x_i$, $\hat{m} = \frac{km+n\bar{x}}{k+n}$, (21) reduces to

$$\frac{\sqrt{k} b^a \lambda^{a+(n-1)/2}}{(2\pi)^{(n+1)/2}\Gamma(a)} \exp\left[-\lambda\left\{b + \frac{kn}{2(k+n)}(m - \bar{x})^2 + \frac{1}{2}\sum(x_i - \bar{x})^2 + \frac{k+n}{2}(\mu - \hat{m})^2\right\}\right]. \quad (22)$$

Comparing (22) with (20), we see that the posterior distribution is of the same form but with the constants a, b, m, k replaced by $\hat{a}, \hat{b}, \hat{m}, \hat{k}$, where

$$\begin{aligned}
\hat{a} &= a + \frac{n}{2}, \\
\hat{b} &= b + \frac{kn}{2(k+n)}(m - \bar{x})^2 + \frac{1}{2}\sum(x_i - \bar{x})^2, \\
\hat{m} &= \frac{km + n\bar{x}}{k+n}, \\
\hat{k} &= k + n.
\end{aligned}$$

In practice, we generally try to choose the prior to be as uninformative as possible. In (20), this is often achieved by setting $k = 0$ (in which case the improper conditional prior for density μ, derived from (18), is just a flat prior over $(-\infty, \infty)$) and setting a and b to very small constants, such as $a = b = 0.01$.

A2. Hierarchical Models

Although all Bayesian statistical models may ultimately be written in the form (14), this may not be the most convenient form either for conceptual model building or for mathematical solution. Often it is better to impose some structure on the unknown parameters that makes their interdependence clear. The most common way to do this is through a *hierarchical model*.

The simplest structure of hierarchical model consists of three layers:

- a top layer of parameters, that are common to all the individual units of the model,
- a middle layer of parameters that are specific to individual units,
- a bottom layer consisting of observations.

An example is the main univariate model considered in Smith et al. (2009), which is defined

by the equations

$$\begin{aligned}
\mu, \nu, \beta &\sim U(-\infty, \infty), && (23)\\
\theta &\sim Ga(a, b), && (24)\\
a_\lambda, b_\lambda &\sim Ga(a^*, b^*), && (25)\\
\lambda_1, ..., \lambda_M \mid a_\lambda, b_\lambda &\sim Ga(a_\lambda, b_\lambda), && (26)\\
X_0 \mid \mu &\sim N[\mu, \lambda_0^{-1}], \quad (\lambda_0 \text{ known}) && (27)\\
X_j \mid \mu, \lambda_j &\sim N[\mu, \lambda_j^{-1}], && (28)\\
Y_j \mid X_j, \mu, \nu, \theta, \lambda_j &\sim N[\nu + \beta(X_j - \mu), (\theta\lambda_j)^{-1}], && (29)
\end{aligned}$$

Here the top layer of parameters consists of $(\mu, \nu, \beta, \theta, a_\lambda, b_\lambda)$, with prior distributions specified by (23)–(25). The middle layer is (26), which defines the precisions $\lambda_1, ..., \lambda_M$ of the M climate models. The bottom layer is (27)–(29), which defines the distributions of the data X_0 (present-day observed climate mean) and X_j, Y_j (present-day and future projection under the jth climate model, $1 \leq j \leq M$). In this case, λ_0 is not considered a parameter because it is assumed known. The constants $a,\ b,\ a^*,\ b^*$ are not unknown parameters of the model but are fixed at the start of the analysis — in practice we set them all equal to 0.01. Sometimes they are called *hyperparameters* to distinguish them from parameters that are actually estimated during the course of the model fitting.

The more general "multivariate" model of Smith et al. (2009) may also be represented in this three-layer hierarchical form, where the model precisions λ_j, having the same prior distribution (26), and the interaction parameters η_{ij}, with independent prior distributions $Ga(c, c)$, are in the middle layer of the hierarchy to indicate their dependence on additional unknown parameters a_λ, b_λ and c. For this model, there are many more unknown parameters, but the basic structure of the model is not more complicated than the univariate version of the model. In general, hierarchical models may have more than three layers, or may be subject to more complex dependencies (e.g. the components of the model may be represented as the vertices of a graph, where the edges of the graph represent the dependence structure), but the simple three-layer structure is sufficient for the present application.

A3. MCMC

Any hierarchical model may be represented conceptually in the form (14) by representing all the unknown parameters together as a single long vector θ. However, in virtually all such cases, the model cannot be solved by simple analytic integration to compute the posterior density, while standard numerical methods of integration, such as the trapezoidal rule, are too slow or insufficiently accurate to yield practical results. *Markov chain Monte Carlo* (MCMC) methods are a class of Monte Carlo simulation techniques whose purpose is to simulate a random sample from the posterior distribution of θ. Although these methods are also used in non-hierarchical problems, in cases where a conjugate prior is not available or is inappropriate for a particular application, they really come into their own in solving hierarchical Bayesian problems. The name derives from the fact that they generate a random sample from a Markov chain, which can be proved to converge to the true posterior distribution as the size of the Monte Carlo sample tends to infinity.

The *Gibbs sampler* is based on first partitioning θ into components $\theta_1, ..., \theta_p$. Here, p may simply be the total number of unknown parameters in the model, though it is also possible for group parameters together when it is convenient to do so (for example, in the gamma-normal prior, we are effectively grouping λ and μ together as a single parameter θ: in a hierarchical model, the precision and mean for a single unit could together be one of $\theta_1, ..., \theta_p$). The key element is that it must be possible to generate a random variable from any one of $\theta_1, ..., \theta_p$ conditional on the other $p - 1$ held fixed: this is often achieved in practice through the use of conjugate priors. The algorithm is then as follows:

1. Choose arbitrary starting values $\theta_1, ..., \theta_p$, say $\theta_1^{(1)}, ..., \theta_p^{(1)}$. Set the counter $b = 1$.
2. Given current values $\theta_1^{(b)}, ..., \theta_p^{(b)}$, generate a Monte Carlo random variate from the distribution of θ_1, conditional on $\theta_j = \theta_j^{(b)},\ j = 2, ..., p$. Call the result $\theta_1^{(b+1)}$. Next, gen-

erate a Monte Carlo random variate from the distribution of θ_2, conditional on $\theta_1 = \theta_1^{(b+1)},\ \theta_j = \theta_j^{(b)},\ j = 3, ..., p$. Call the result $\theta_2^{(b+1)}$. Continue with $\theta_3, \theta_4, ...$, up to generating a Monte Carlo random variate $\theta_p^{(b+1)}$ from the distribution of θ_p, conditional on $\theta_j = \theta_j^{(b+1)},\ j = 1, ..., p-1$.

3. Set counter $b \to b+1$ and return to step 2.
4. Continue until $b = B$, the desired number of Monte Carlo iterations. The value of B depends on the complexity of the problem and available computer time, but a value in the range 5,000–100,000 is typical.
5. The first $B_0 < B$ samples are discarded as *burn-in* samples, reflecting that the Markov chain has not yet converged to its stationary distribution. Among the last $B - B_0$ iterations, it is usual to retain only every tth iteration, for some suitable $t > 1$. This reflects the fact that successive iterations from the Gibbs sampler are highly correlated, but by thinning in this way, the remaining iterations may be treated as approximately uncorrelated.
6. The resulting $(B - B_0)/t$ values of θ are treated as a random sample from the posterior distribution of θ. Posterior means and standard deviations, predictive distributions, etc., are calculated by averaging over these random samples.

As an example, let us consider the simpler form of (23)–(29) in which equation (25) is omitted and a_λ, b_λ are treated as known constants. As a notational point, θ in (24) is a single (scalar) parameter in a multi-parameter model, not to be confused with our notation of θ as the vector of all unknown parameters in (14). The context should make clear which of the two uses of θ is intended in any particular instance.

In this case, the joint density of all the unknown parameters and observations is proportional to

$$\theta^{a+M/2-1} e^{-b\theta} e^{-\frac{1}{2}\lambda_0(X_0-\mu)^2} \prod_{j=1}^{M} \left[\lambda_j^{a_\lambda} e^{-b_\lambda \lambda_j} \cdot e^{-\frac{1}{2}\lambda_j(X_j-\mu)^2 - \frac{1}{2}\theta\lambda_j\{Y_j-\nu-\beta(X_j-\mu)\}^2} \right]. \qquad (30)$$

It is not possible to integrate (30) analytically with respect to all the unknown parameters $\mu, \nu, \beta, \theta, \lambda_1, ..., \lambda_M$. However, for any single one of these parameters, we can integrate using standard forms based on the normal or gamma densities. For example, as a function of θ alone, conditional on all the others, (30) is proportional to

$$\theta^{a+M/2-1} e^{-\theta[b+\frac{1}{2}\sum_j \lambda_j\{Y_j-\nu-\beta(X_j-\mu)\}^2]},$$

which in turn is proportional to a gamma density with parameters $a + \frac{M}{2}$ and $b + \frac{1}{2}\sum_j \lambda_j\{Y_j - \nu - \beta(X_j - \mu)\}^2$. Therefore the updating step for θ is to generate a random variate with this gamma distribution, replacing the previous value of θ.

Similarly, the conditional density of λ_j (for fixed $j \in \{1, 2, ..., M\}$) is gamma with parameters $a_\lambda + 1$ and $b_\lambda + \frac{1}{2}[(X_j - \mu)^2 + \theta\{Y_j - \nu - \beta(X_j - \mu)\}^2]$. The remaining parameters μ, ν, β have conditional normal distributions that can be calculating by completing the square in the exponent. As an example, we give the explicit calculation for μ. Expressed as a function of μ, (30) is of the form $e^{-Q(\mu)/2}$, where

$$Q(\mu) \quad = \quad \lambda_0(X_0 - \mu)^2 + \sum_j \lambda_j(X_j - \mu)^2 + \theta \sum_j \lambda_j\{Y_j - \nu - \beta(X_j - \mu)\}^2.$$

Completing the square,

$$Q(\mu) \quad = \quad (\lambda_0 + \sum_j \lambda_j + \theta\beta^2 \sum_j \lambda_j) \left\{ \mu - \frac{\lambda_0 X_0 + \sum_j \lambda_j X_j - \theta\beta \sum_j \lambda_j (Y_j - \nu - \beta X_j)}{\lambda_0 + \sum_j \lambda_j + \theta\beta^2 \sum_j \lambda_j} \right\}^2 + \text{const}$$

where "const" contains terms that do not depend on μ. But based on this representation, we recognize that $e^{-Q(\mu)/2}$ has the form of a normal density with mean and variance

$$\frac{\lambda_0 X_0 + \sum_j \lambda_j X_j - \theta\beta \sum_j \lambda_j (Y_j - \nu - \beta X_j)}{\lambda_0 + \sum_j \lambda_j + \theta\beta^2 \sum_j \lambda_j}, \quad \frac{1}{\lambda_0 + \sum_j \lambda_j + \theta\beta^2 \sum_j \lambda_j},$$

so the Gibbs sampling step for μ is to generate a random variate from this normal distribution. The calculations for ν and β are similar so we omit the details.

In a typical application of this algorithm, the Gibbs sampler was run for an initial 12,500 iterations as a burn-in, followed by 50,000 iterations, with every 50th iteration being saved. Thus, we ended up with a Monte Carlo sample of size 1,000, which are approximately independent from the joint posterior distribution of the unknown parameters.

A4. The Metropolis-Hastings Algorithm

The Gibbs sampler still relies on the assumption that, after partitioning θ (the full vector of unknown parameters) into $\theta_1, ..., \theta_p$, the conditional distribution of each θ_j, given all the $\theta_k,\ k \neq j$, may be represented in a sufficiently explicit form that it is possible to generate a Monte Carlo variate directly. In cases where this is not possible, there are by now a wide variety of alternative MCMC algorithms, but the oldest and best known is the Metropolis-Hastings algorithm, which we now describe.

Its general form is as follows. We assume again that θ is an unknown parameter vector lying in some parameter space Θ. Suppose we want to generate Monte Carlo samples from θ, with the probability mass function or probability density function $g(\theta)$. We also assume we have some stochastic mechanism for generating an "update" θ', given the current value of θ, that is represented by a Markov transition kernel $q(\theta, \theta')$. In other words, for each $\theta \in \Theta$, we assume there exists a probability mass function or probability density function $q(\theta, \cdot)$, that represents the conditional density of θ' given θ. In principle, $q(\theta, \theta')$ is arbitrary subject to only mild restrictions (for example, the Markov chain generated by $q(\theta, \theta')$ must be irreducible, in the sense that samples generated by this Markov chain will eventually cover the whole of Θ). In practice, certain simple forms such as random walks are usually adopted.

The main steps of the algorithm are:

1. Start with an initial trial value of θ, call it $\theta^{(1)}$. Set counter $b = 1$.
2. Given current value $\theta^{(b)}$, generate a new trial θ' from the conditional density $q(\theta^{(b)}, \theta')$.
3. Calculate
$$\alpha \quad = \quad \min\left\{ \frac{g(\theta')q(\theta', \theta^{(b)})}{g(\theta^{(b)})q(\theta^{(b)}, \theta')}, 1 \right\}.$$
4. With an independent draw from the random number generator, determine whether we "accept" θ', where the probability of acceptance is α.
5. If the result of step 4 is to accept θ', set $\theta^{(b+1)} = \theta'$. Otherwise, $\theta^{(b+1)} = \theta^{(b)}$.
6. Set counter $b \rightarrow b + 1$ and return to step 2.
7. Continue until $b = B$, the desired number of Monte Carlo iterations.

As with the Gibbs sampler, it is usual to discard a large number of initial iterations as "burn-in", and then to retain only a thinned subset of the remaining iterations, to ensure approximate independence between Monte Carlo variates.

A key feature of the algorithm is that the acceptance probability α depends only on *ratios* of the density g — in other words, it is not necessary to specify g exactly, so long as it is known up to a normalizing constant. This is precisely the situation we face with (14), where the numerator $\pi(\theta)f(x \mid \theta)$ is known explicitly but the denominator requires an integral which is typically intractable. However, with the Metropolis-Hastings algorithm, it is not necessary to evaluate the denominator.

A simpler form of the algorithm arises if the kernel q is symmetric, i.e. $q(\theta', \theta) = q(\theta, \theta')$ for all $\theta \neq \theta'$. In that case, the formula for the acceptance probability reduces to

$$\alpha \quad = \quad \min\left\{ \frac{g(\theta')}{g(\theta^{(b)})}, 1 \right\}.$$

In this form, the algorithm is equivalent to the Monte Carlo sampling algorithm of Metropolis et al. (1953), which was used for statistical physics calculations for decades before the algorithm's rediscovery by Bayesian statisticians. The general form of the algorithm, and its justification in terms of the convergence theory of Markov chains, was due to Hastings (1970).

Although it is possible to apply the Metropolis-Hastings sampler entirely independently of the Gibbs sampler, in practice, the two ideas are often combined, where the Gibbs sampler is used to update all those components of θ for which explicit conditional densities are available in a form that can easily be simulated, and a Metropolis-Hastings step is used for the remaining components. This is exactly the way the algorithm was applied in Smith et al. (2009). To be explicit, let us return to the model defined by equations (23)–(29), where now we treat a_λ, b_λ as unknown parameters with prior density (25). Isolating those parts of the joint density that depend just on a_λ, b_λ, we have to generate Monte Carlo samples from a density proportional to

$$g(a_\lambda, b_\lambda) \quad = \quad (a_\lambda b_\lambda)^{a^*-1} e^{-b^*(a_\lambda+b_\lambda)} \prod_{j=1}^{M} \frac{b_\lambda^{a_\lambda} \lambda_j^{a_\lambda} e^{-b_\lambda \lambda_j}}{\Gamma(a_\lambda)}. \tag{31}$$

This is not of conjugate prior form so we use the Metropolis-Hastings algorithm to update a_λ and b_λ. The form of updating rule is that given the current a_λ and b_λ, we define $a'_\lambda = a_\lambda e^{\delta(U_1 - \frac{1}{2})}$, $b'_\lambda = b_\lambda e^{\delta(U_2 - \frac{1}{2})}$, for random U_1 and U_2 that are uniform on $[0,1]$ (independent of each other and all other random variates). Both the form of this updating rule and the choice of δ are arbitrary, but we have found it works well in practice, and have used $\delta = 1$ in most of our calculations. With this rule,

$$q\{(a_\lambda, b_\lambda), (a'_\lambda, b'_\lambda)\} \quad = \quad \frac{1}{\delta^2 a'_\lambda b'_\lambda}, \quad a'_\lambda \in \left(a_\lambda e^{-\delta/2}, a_\lambda e^{\delta/2}\right), b'_\lambda \in \left(b_\lambda e^{-\delta/2}, b_\lambda e^{\delta/2}\right).$$

Therefore $q\{(a'_\lambda, b'_\lambda), (a_\lambda, b_\lambda)\}/q\{(a_\lambda, b_\lambda), (a'_\lambda, b'_\lambda)\} = a'_\lambda b'_\lambda/(a_\lambda b_\lambda)$ and the acceptance probability α reduces to

$$\min\left\{\frac{g(a'_\lambda, b'_\lambda) a'_\lambda b'_\lambda}{g(a_\lambda, b_\lambda) a_\lambda b_\lambda}, 1\right\}.$$

A5. Diagnostics

In this section, we briefly review some of the available diagnostics.

With the Metropolis-Hastings algorithm, the kernel $q(\theta, \theta')$ often contains a tuning parameter (δ in the above example), which controls the size of the jumps. This raises the question of how to choose the tuning parameter. For a specific class of Metropolis sampling rules based on a multivariate normal target, Gelman et al. (1996) showed that there is an optimal acceptance probability α, which ranges from 0.44 in one dimension to 0.23 as the dimension tends to ∞. Using this result as a guideline, it is often recommended that for general Metropolis sampling rules, the acceptance probability should be tuned so that it lies between about 0.15 and 0.5 on average. There are many specific proposals for adaptively optimizing this choice; the paper by Pasarica and Gelman (2008) contains one recent proposal and reviews earlier literature.

The question of how many iterations of a MCMC algorithm are required to achieve adequate convergence to the stationary distribution has been the subject of much research. One well-regarded procedure was proposed by Gelman and Rubin (1992) and extended by Brooks and Gelman (1998). The Gelman-Rubin procedure requires running several samplers in parallel, using widely dispersed initial values for the parameters. Then, it calculates "potential scale reduction factor" R. This can be interpreted as an estimate of the possible reduction in variance of the posterior mean of a particular parameter, if the Markov chain were run to convergence, compared with the current iterations. Since R is itself an unknown parameter estimated from the sample output, it is common to quote both the median and some upper quantile (say, the .975 quantile) of the sampling distribution. The ideal value of R is 1; values much above 1 are taken to indicate non-convergence of the MCMC procedure.

Another issue is inference after sampling. Suppose we are trying to estimate the posterior mean of a particular parameter; in our climate application, ν, the estimated mean of future values of a climatic variable, is of particular interest. If we had a Monte Carlo sample of independent estimates of ν, the usual standard error calculation would give an estimate of the sampling variability of the posterior mean of ν. In an MCMC sample, it is desirable to take account of the fact that successive draws from the sample may be autocorrelated. There are many possible procedures; one method, due to Heidelberger and Welch (1981), constructs a

nonparametric estimate of the spectral density at low frequencies, and uses this to correct the standard error.

The Gelman-Rubin and Heidelberger-Welch procedures are included, among several others, in the CODA diagnostics package (Plummer et al., 2006), which is available as a downloadable package within R (R Development Core Team, 2007).

A6. Further reading

There are by now many books on the principles of Bayesian data analysis and MCMC algorithms; a small selection includes Gamerman and Lopes (2006); Gelman et al. (2003); Robert and Casella (2004). The book by Robert (2005) is a somewhat more theoretical treatment that makes the link between modern practices in Bayesian statistics and classical decision theory and inference.

B. Computational details

B1. Computation for the univariate model

We detail here the Gibbs/Metropolis algorithm to sample the posterior distribution of the model in Section 4.1, fitting one region at a time, with a hyperprior on the parameters of the prior for the precisions, λ_js. This algorithm is originally described in Smith et al. (2009).

Under the region-specific model, the joint density of θ, μ, ν, β, a_λ, b_λ, X_0 and $\lambda_j,\ X_j,\ Y_j,\ ,\ (j = 1, ..., M)$ is proportional to

$$\begin{aligned}&\theta^{a+M/2-1}e^{-b\theta}e^{-\frac{1}{2}\lambda_0(X_0-\mu)^2}a_\lambda^{a^*-1}e^{-b^*a_\lambda}b_\lambda^{a^*-1}e^{-b^*b_\lambda}\ .\\&\cdot\prod_{j=1}^{M}\left[\lambda_j^{a_\lambda}e^{-b_\lambda\lambda_j}\ .\ e^{-\frac{1}{2}\lambda_j(X_j-\mu)^2-\frac{1}{2}\theta\lambda_j\{Y_j-\nu-\beta(X_j-\mu)\}^2}\right].\end{aligned} \tag{32}$$

Define

$$\tilde{\mu} = \frac{\lambda_0 X_0 + \sum\lambda_j X_j - \theta\beta\sum\lambda_j(Y_j-\nu-\beta X_j)}{\lambda_0+\sum\lambda_j+\theta\beta^2\sum\lambda_j}, \tag{33}$$

$$\tilde{\nu} = \frac{\sum\lambda_j\{Y_j-\beta(X_j-\mu)\}}{\sum\lambda_j}, \tag{34}$$

$$\tilde{\beta} = \frac{\sum\lambda_j(Y_j-\nu)(X_j-\mu)}{\sum\lambda_j(X_j-\mu)^2}. \tag{35}$$

In a Monte Carlo sampling scheme, all the parameters in (32), with the exception of a_λ and b_λ, may be updated through Gibbs sampling steps, as follows:

$$\mu \mid \text{rest} \sim N\left[\tilde{\mu}, \frac{1}{\lambda_0+\sum\lambda_j+\theta\beta^2\sum\lambda_j}\right], \tag{36}$$

$$\nu \mid \text{rest} \sim N\left[\tilde{\nu}, \frac{1}{\theta\sum\lambda_j}\right], \tag{37}$$

$$\beta \mid \text{rest} \sim N\left[\tilde{\beta}, \frac{1}{\theta\sum\lambda_j(X_j-\mu)^2}\right], \tag{38}$$

$$\lambda_j \mid \text{rest} \sim G\left[a+1, b+\frac{1}{2}(X_j-\mu)^2+\frac{\theta}{2}\{Y_j-\nu-\beta(X_j-\mu)\}^2\right], \tag{39}$$

$$\theta \mid \text{rest} \sim G\left[a+\frac{M}{2}, b+\frac{1}{2}\sum\lambda_j\{Y_j-\nu-\beta(X_j-\mu)\}^2\right]. \tag{40}$$

For the parameters a_λ, b_λ, the following Metropolis updating step is proposed instead:

1. Generate U_1, U_2, U_3, independent uniform on $(0, 1)$.
2. Define new trial values $a'_\lambda = a_\lambda e^{\delta(U_1-1/2)}$, $b'_\lambda = b_\lambda e^{\delta(U_2-1/2)}$. The value of δ (step length) is arbitrary but $\delta = 1$ seems to work well in practice, and is therefore used here.

3. Compute

$$\begin{aligned}\ell_1 &= Ma_\lambda \log b_\lambda - M\log\Gamma(a_\lambda) + (a_\lambda - 1)\sum \log\lambda_j - b_\lambda \sum \lambda_j + a^* \log(a_\lambda b_\lambda) - b^*(a_\lambda + b_\lambda),\\ \ell_2 &= Ma'_\lambda \log b'_\lambda - M\log\Gamma(a'_\lambda) + (a'_\lambda - 1)\sum \log\lambda_j - b'_\lambda \sum \lambda_j + a^* \log(a'_\lambda b'_\lambda) - b^*(a'_\lambda + b'_\lambda).\end{aligned}$$

   This computes the log likelihood for both (a_λ, b_λ) and (a'_λ, b'_λ), allowing for the prior density and including a Jacobian term to allow for the fact that the updating is on a logarithmic scale.

4. If

$$\log U_3 < \ell_2 - \ell_1$$

   then we accept the new (a_λ, b_λ), otherwise keep the present values for the current iteration, as in a standard Metropolis accept-reject step.

This process is iterated many times to generate a random sample from the joint posterior distribution. In the case where a_λ, b_λ are treated as fixed, the Metropolis steps for these two parameters are omitted and in this case the method is a pure Gibbs sampler. An R program (REA.GM.r) to perform the sampling is available for download from `http://www.image.ucar.edu/~tebaldi/REA`.

B2. Computation for the multivariate model

The following computations are originally described in Smith et al. (2009). Omitting unnecessary constants, the joint density of all the random variables in the model that treats all regions at the same time is

$$\begin{aligned}&(ca_\lambda b_\lambda)^{a-1} e^{-b(c+a_\lambda+b_\lambda)} \cdot \left[\prod_{i=0}^{R} \theta_i^{a-1} e^{-b\theta_i}\right] \cdot \left[\prod_{i=1}^{R} \phi_i^{a-1} e^{-b\phi_i}\right] \cdot \left[\prod_{j=1}^{M} \lambda_j^{a_\lambda - 1} e^{-b_\lambda \lambda_j} \frac{b_\lambda}{\Gamma(a_\lambda)}\right] \cdot \left[\psi_0^{a-1} e^{-b\psi_0}\right] \cdot\\ &\cdot \left[\prod_{i=1}^{R}\prod_{j=1}^{M} \eta_{ij}^{c-1} e^{-c\eta_{ij}} \frac{c^c}{\Gamma(c)}\right] \cdot \left[\prod_{j=1}^{M} \sqrt{\psi_0} e^{-\frac{1}{2}\psi_0 \alpha_j^2}\right] \cdot \left[\prod_{j=1}^{M} \sqrt{\theta_0\psi_0} e^{-\frac{1}{2}\theta_0\psi_0(\alpha'_j - \beta_0\alpha_j)^2}\right] \cdot\\ &\cdot \left[\prod_{i=1}^{R} e^{-\frac{1}{2}\lambda_{0i}(X_{i0} - \mu_0 - \zeta_i)^2}\right] \cdot \left[\prod_{i=1}^{R}\prod_{j=1}^{M} \sqrt{\eta_{ij}\phi_i\lambda_j} e^{-\frac{1}{2}\eta_{ij}\phi_i\lambda_j(X_{ij} - \mu_0 - \zeta_i - \alpha_j)^2}\right] \cdot\\ &\cdot \left[\prod_{i=1}^{R}\prod_{j=1}^{M} \sqrt{\eta_{ij}\theta_i\lambda_j} e^{-\frac{1}{2}\eta_{ij}\theta_i\lambda_j\left\{Y_{ij} - \nu_0 - \zeta'_i - \alpha'_j - \beta_i(X_{ij} - \mu_0 - \zeta_i - \alpha_j)\right\}^2}\right] \qquad (41)\end{aligned}$$

Define

$$\tilde{\mu}_0 = \frac{\sum_i \lambda_{0i}(X_{i0}-\zeta_i) + \sum_i \phi_i \sum_j \eta_{ij}\lambda_j(X_{ij}-\zeta_i-\alpha_j) - \sum_i \beta_i\theta_i \sum_j \eta_{ij}\lambda_j\{Y_{ij}-\nu_0-\zeta_i-\alpha'_j-\beta_i(X_{ij}-\zeta_i-\alpha_j)\}}{\sum_i\{\lambda_{0i}+(\phi_i+\beta_i^2\theta_i)\sum_j \eta_{ij}\lambda_j\}} \quad (42$$

$$\tilde{\nu}_0 = \frac{\sum_i \theta_i \sum_j \eta_{ij}\lambda_j\{Y_{ij}-\zeta'_i-\alpha'_j-\beta_i(X_{ij}-\mu_0-\zeta_i-\alpha_j)\}}{\sum_i \theta_i\{\sum_j \eta_{ij}\lambda_j\}}, \quad (43$$

$$\tilde{\zeta}_i = \frac{\lambda_{0i}(X_{i0}-\mu_0) + \phi_i \sum_j \eta_{ij}\lambda_j(X_{ij}-\mu_0-\alpha_j) - \beta_i\theta_i \sum_j \eta_{ij}\lambda_j\{Y_{ij}-\nu_0-\zeta'_i-\alpha'_j-\beta_i(X_{ij}-\mu_0-\alpha_j)\}}{\lambda_{0i}+(\phi_i+\beta_i^2\theta_i)\sum_j \eta_{ij}\lambda_j}, \quad (44$$

$$\tilde{\zeta}'_i = \frac{\sum_j \eta_{ij}\lambda_j\{Y_{ij}-\nu_0-\alpha'_j-\beta_i(X_{ij}-\mu_0-\zeta_i-\alpha_j)\}}{\sum_j \eta_{ij}\lambda_j}, \quad (45$$

$$\tilde{\beta}_0 = \frac{\sum_j \alpha'_j\alpha_j}{\sum_j \alpha_j^2}, \quad (46$$

$$\tilde{\beta}_i = \frac{\sum_j \eta_{ij}\lambda_j(Y_{ij}-\nu_0-\zeta'_i-\alpha'_j)(X_{ij}-\mu_0-\zeta_i-\alpha_j)}{\sum_j \eta_{ij}\lambda_j(X_{ij}-\mu_0-\zeta_i-\alpha_j)^2}, \quad (i\neq 0), \quad (47$$

$$\tilde{\alpha}_j = \frac{\beta_0\theta_0\psi_0\alpha'_j + \lambda_j\sum_i \eta_{ij}\phi_i(X_{ij}-\mu_0-\zeta_i) - \lambda_j\sum_i \eta_{ij}\theta_i\beta_i\{Y_{ij}-\nu_0-\zeta'_i-\alpha'_j-\beta_i(X_{ij}-\mu_0-\zeta_i)\}}{\psi_0+\beta_0^2\theta_0\psi_0+\lambda_j\sum_i \eta_{ij}\phi_i+\lambda_j\sum_i \eta_{ij}\theta_i\beta_i^2}, \quad (48$$

$$\tilde{\alpha}'_j = \frac{\beta_0\theta_0\psi_0\alpha_j + \lambda_j\sum_i \eta_{ij}\theta_i\{Y_{ij}-\nu_0-\zeta'_i-\beta_i(X_{ij}-\mu_0-\zeta_i-\alpha_j)\}}{\theta_0\psi_0+\lambda_j\sum_i \eta_{ij}\theta_i}. \quad (49$$

The conditional distributions required for the Gibbs sampler are as follows:

$$\mu_0 \mid \text{rest} \sim N\left[\tilde{\mu}_0, \frac{1}{\sum_i\{\lambda_{0i} + (\phi_i + \beta_i^2\theta_i)\sum_j \eta_{ij}\lambda_j\}}\right], \tag{50}$$

$$\nu_0 \mid \text{rest} \sim N\left[\tilde{\nu}_0, \frac{1}{\sum_i\{\theta_i \sum_j \eta_{ij}\lambda_j\}}\right], \tag{51}$$

$$\zeta_i \mid \text{rest} \sim N\left[\tilde{\zeta}_i, \frac{1}{\lambda_i + (\phi_i + \beta_i^2\theta_i)\sum_j \eta_{ij}\lambda_j}\right], \tag{52}$$

$$\zeta_i' \mid \text{rest} \sim N\left[\tilde{\zeta}_i', \frac{1}{\theta_i \sum_j \eta_{ij}\lambda_j}\right], \tag{53}$$

$$\beta_0 \mid \text{rest} \sim N\left[\tilde{\beta}_0, \frac{1}{\theta_0\psi_0 \sum_j \alpha_j^2}\right], \tag{54}$$

$$\beta_i \mid \text{rest} \sim N\left[\tilde{\beta}_i, \frac{1}{\theta_i \sum_j \eta_{ij}\lambda_j (X_{ij} - \mu_0 - \zeta_i - \alpha_j)^2}\right], \quad (i \neq 0) \tag{55}$$

$$\alpha_j \mid \text{rest} \sim N\left[\tilde{\alpha}_j, \frac{1}{\psi_0 + \beta_0^2\theta_0\psi_0 + \lambda_j \sum_i \eta_{ij}\phi_i + \lambda_j \sum_i \eta_{ij}\theta_i\beta_i^2}\right], \tag{56}$$

$$\alpha_j' \mid \text{rest} \sim N\left[\tilde{\alpha}_j', \frac{1}{\theta_0\psi_0 + \lambda_j \sum_i \eta_{ij}\theta_i}, \right], \tag{57}$$

$$\theta_0 \mid \text{rest} \sim Gam\left[a + \frac{M}{2}, b + \frac{1}{2}\psi_0 \sum_j (\alpha_j' - \beta_0\alpha_j)^2\right], \tag{58}$$

$$\theta_i \mid \text{rest} \sim Gam\left[a + \frac{M}{2}, b + \frac{1}{2}\sum_j \eta_{ij}\lambda_j\{Y_{ij} - \nu_0 - \zeta_i' - \alpha_j' - \beta_i(X_{ij} - \mu_0 - \zeta_i - \alpha_j)\}^2\right], \quad (i \neq 0) \tag{59}$$

$$\phi_i \mid \text{rest} \sim Gam\left[a + \frac{M}{2}, b + \frac{1}{2}\sum_j \eta_{ij}\lambda_j (X_{ij} - \mu_0 - \zeta_i - \alpha_j)^2\right], \tag{60}$$

$$\lambda_j \mid \text{rest} \sim Gam\left[a_\lambda + R, b_\lambda + \frac{1}{2}\sum_i \eta_{ij}\phi_i (X_{ij} - \mu_0 - \zeta_i - \alpha_j)^2 + \frac{1}{2}\sum_i \eta_{ij}\theta_i\{Y_{ij} - \nu_0 - \zeta_i' - \alpha_j' - \beta_i(X_{ij} - \mu_0 - \zeta_i - \alpha_j)\}^2\right], \tag{61}$$

$$\psi_0 \mid \text{rest} \sim Gam\left[a + M, b + \frac{1}{2}\sum_j \alpha_j^2 + \frac{1}{2}\theta_0 \sum_j (\alpha_j' - \beta_0\alpha_j)^2\right], \tag{62}$$

$$\eta_{ij} \mid \text{rest} \sim Gam\left[c + 1, c + \frac{1}{2}\phi_i\lambda_j (X_{ij} - \mu_0 - \zeta_i - \alpha_j)^2 + \frac{1}{2}\theta_i\lambda_j\{Y_{ij} - \nu_0 - \zeta_i' - \alpha_j' - \beta_i(X_{ij} - \mu_0 - \zeta_i - \alpha_j)\}^2\right]. \tag{63}$$

Were a_λ, b_λ and c fixed, as in the univariate analysis, the iteration (50)–(63) could be repeated many times to generate a random sample from the joint posterior distribution. Having added a layer by making the three parameters random variates, two Metropolis steps are added to the iteration (50)–(63), as follows.

For the sampling of a_λ and b_λ jointly, define U_1, U_2 two independent random variables distributed uniformly over the interval $(0, 1)$, and the two candidate values $a_\lambda' = a_\lambda e^{(\delta(u_1 - \frac{1}{2}))}$ and $b_\lambda' = b_\lambda e^{(\delta(u_2 - \frac{1}{2}))}$, where δ is an arbitrary increment, chosen as $\delta = 1$ in the implementation to

follow. We then compute

$$l_1 = MRa_\lambda \log(b_\lambda) - MR\log(\Gamma(a_\lambda) + R(a_\lambda - 1)\sum_j \log(\lambda_j) - Rb_\lambda \sum_j \lambda_j + a\log(a_\lambda b_\lambda) - b(a_\lambda + b_\lambda), \quad (64)$$

$$l_2 = MRa'_\lambda \log(b'_\lambda) - MR\log(\Gamma(a'_\lambda) + R(a'_\lambda - 1)\sum_j \log(\lambda_j) - Rb'_\lambda \sum_j \lambda_j + a\log(a'_\lambda b'_\lambda) - b(a'_\lambda + b'_\lambda). \quad (65)$$

In (64) and (65) we are computing the log likelihoods of (a_λ, b_λ) and (a'_λ, b'_λ), allowing for the prior densities and including a jacobian term, allowing for the fact that the updating is taking place on a logarithmic scale. Then, within each iteration of the Gibbs/Metropolis simulation, the proposed values (a'_λ, b'_λ) are accepted with probability $e^{l_2 - l_1}$.

Similarly, the updating of c takes place by proposing $c\prime = ce^{(\delta(u_3 - \frac{1}{2}))}$, where u_3 is a draw from a uniform distribution on $(0, 1)$, and computing

$$k_1 = MRc\log(c) - MR\log(\Gamma(c)) + (c-1)\sum_i\sum_j \log(\eta_{ij}) - c\sum_i\sum_j \eta_{ij} + a\log(c) - bc \quad (66)$$

$$k_2 = MRc'\log(c') - MR\log(\Gamma(c')) + (c'-1)\sum_i\sum_j \log(\eta_{ij}) - c'\sum_i\sum_j \eta_{ij} + a\log(c') - bc'. \quad (67)$$

Then, within each iteration of the Gibbs/Metropolis simulation, the proposed value c' is accepted with probability $e^{k_2 - k_1}$.

The method is available as an R program (REAMV.GM.r) from
`http://www.image.ucar.edu/~tebaldi/REA/`.

B3. Computation for the bivariate model

The MCMC algorithm for the bivariate model is taken from Tebaldi and Sansó (2008). Note that in the following, the "prime" symbol denotes the operation of centering a variable (O_t or X_{jt}) by the respective climate signal $\mu_t = \alpha + \beta t + \gamma t \mathcal{I}_{\{t>=\tau_0\}}$.

Coefficients of the piecewise linear model:
Define

$$A = \tau_0\eta^T + \tau_0\eta^P\beta^2_{xo} + \tau^*\sum_j \xi^T_j + \tau^*\sum_j \xi^P_j\beta^2_{xj}$$

and

$$\begin{aligned} B = {} & \eta^T\sum_{t\le\tau_0}(O^T_t - \beta^T t) + \eta^P\sum_{t\le\tau_0}\beta^2_{xo}(O^T_t - \beta^T t - \beta_{xo}O^{P'}_t) + \sum_j \xi^T_j\sum_{t\le\tau_0}(X^T_{jt} - \beta^T t - d^T_j) + \\ & \sum_j \xi^T_j\sum_{t>\tau_0}(X^T_{jt} - \beta^T t - \gamma^T(t-\tau_0) - d^T_j) + \sum_j \xi^P_j\beta^2_{xj}\sum_{t\le\tau_0}(X^T_{jt} - \beta^T t - d^T_j - \beta_{xj}(X^{P'}_{jt} - d^P_j)) + \\ & \sum_j \xi^P_j\beta^2_{xj}\sum_{t>\tau_0}(X^T_{jt} - \beta^T t - \gamma^T(t-\tau_0) - d^T_j - \beta_{xj}(X^{P'}_{jt} - d^P_j)). \end{aligned}$$

Then

$$\alpha^T \sim \mathcal{N}(\frac{B}{A}, (A)^{-1}).$$

Define

$$A = \tau_0\eta^P + \tau^*\sum_j \xi^P_j$$

and

$$\begin{aligned} B = {} & \eta^P\sum_{t\le\tau_0}(O^P_t - \beta^P t - \beta_{xo}O^{T'}_t) + \sum_j \xi^P_j\sum_{t\le\tau_0}(X^P_{jt} - \beta^P t - d^P_j - \beta_{xj}(X^{T'}_{jt} - d^T_j)) + \\ & \sum_j \xi^P_j\sum_{t>\tau_0}(X^P_{jt} - \beta^P t - \gamma^P(t-\tau_0) - d^P_j - \beta_{xj}(X^{T'}_{jt} - d^T_j)). \end{aligned}$$

Then

$$\alpha^P \sim \mathcal{N}(\frac{B}{A}, (A)^{-1}).$$

Define

$$A = \eta^T \sum_{t \leq \tau_0} t^2 + \eta^P \beta_{xo}^2 \sum_{t \leq \tau_0} t^2 + \sum_j \xi_j^T \sum_{t \leq \tau^*} t^2 + \sum_j \xi_j^P \beta_{xj}^2 \sum_{t \leq \tau^*} t^2$$

and

$$\begin{aligned} B &= \eta^T \sum_{t \leq \tau_0} t(O_t^T - \alpha^T) + \eta^P \sum_{t \leq \tau_0} \beta_{xo}^2 t(O_t^T - \alpha^T) - \beta_{xo} t O_t^{P'}) + \sum_j \xi_j^T \sum_{t \leq \tau_0} t(X_{jt}^T - \alpha^T - d_j^T) + \\ & \sum_j \xi_j^T \sum_{t > \tau_0} t(X_{jt}^T - \alpha^T - \gamma^T(t - \tau_0) - d_j^T) + \sum_j \xi_j^P \sum_{t \leq \tau_0} t(\beta_{xj}^2 (X_{jt}^T - \alpha^T - d_j^T) - \beta_{xj}(X_{jt}^{P'} - d_j^P)) + \\ & \sum_j \xi_j^P \sum_{t > \tau_0} t(\beta_{xj}^2 (X_{jt}^T - \alpha^T - \gamma^T(t - \tau_0) - d_j^T) - \beta_{xj}(X_{jt}^{P'} - d_j^P)). \end{aligned}$$

Then

$$\beta^T \sim \mathcal{N}(\frac{B}{A}, (A)^{-1}).$$

Define

$$A = \eta^P \sum_{t \leq \tau_0} t^2 + \sum_j \xi_j^P \sum_{t \leq \tau^*} t^2$$

and

$$\begin{aligned} B &= \eta^P \sum_{t \leq \tau_0} t(O_t^P - \alpha^P - \beta_{xo} O_t^{T'}) + \sum_j \xi_j^P \sum_{t \leq \tau_0} t(X_{jt}^P - \alpha^P - d_j^P - \beta_{xj}(X_{jt}^{T'} - d_j^T)) + \\ & \sum_j \xi_j^P \sum_{t > \tau_0} t(X_{jt}^P - \alpha^P - \gamma^P(t - \tau_0) - d_j^P - \beta_{xj}(X_{jt}^{T'} - d_j^T)). \end{aligned}$$

Then

$$\beta^P \sim \mathcal{N}(\frac{B}{A}, (A)^{-1}).$$

Define

$$A = \sum_j \xi_j^T \sum_{t > \tau_0} (t - \tau_0)^2 + \sum_j \xi_j^P \beta_{xj}^2 \sum_{t > \tau_0} (t - \tau_0)^2$$

and

$$B = \sum_j \xi_j^T \sum_{t > \tau_0} (t - \tau_0)(X_{jt}^T - \alpha^T - \beta^T t - d_j^T) + \sum_j \xi_j^P \sum_{t > \tau_0} (t - \tau_0)(\beta_{xj}^2 (X_{jt}^T - \alpha^T - \beta^T t - d_j^T) - \beta_{xj}(X_{jt}^{P'} - d_j^P)).$$

Then

$$\gamma^T \sim \mathcal{N}(\frac{B}{A}, (A)^{-1}).$$

Define

$$A = \sum_j \xi_j^P \sum_{t > \tau_0} (t - \tau_0)^2$$

and

$$B = \sum_j \xi_j^P \sum_{t > \tau_0} (t - \tau_0)(X_{jt}^P - \alpha^P - \beta^P t - d_j^P - \beta_{xj}(X_{jt}^{T'} - d_j^T)).$$

Then

$$\gamma^P \sim \mathcal{N}(\frac{B}{A}, (A)^{-1}).$$

Bias terms and their priors' parameters:
Define

$$A = \tau^* \xi_j^T + \tau^* \xi_j^P \beta_{xj}^2 + \lambda_D^T$$

and

$$B = \xi_j^T \sum_{t \leq \tau^*} X_{jt}^{T'} + \xi_j^P \sum_{t \leq \tau^*} (\beta_{xj}^2 X_{jt}^{T'} - \beta_{xj}(X_{jt}^{P'} - d_j^P)) + \lambda_D^T a^T.$$

Then

$$d_j^T \sim \mathcal{N}(\frac{B}{A}, (A)^{-1}).$$

Define

$$A = \tau^* \xi_j^P + \lambda_D^P$$

and

$$B = \xi_j^P \sum_{t \leq \tau^*} (X_{jt}^{P'} - \beta_{xj}(X_{jt}^{T'} - d_j^T)) + \lambda_D^P a^P.$$

Then

$$d_j^P \sim \mathcal{N}(\frac{B}{A}, (A)^{-1}).$$

Define $A = M\lambda_D^T$ and $B = \lambda_D^T \sum_j d_j^T$, then

$$a^T \sim \mathcal{N}(\frac{B}{A}, (A)^{-1}).$$

Define $A = M\lambda_D^P$ and $B = \lambda_D^P \sum_j d_j^P$, then

$$a^P \sim \mathcal{N}(\frac{B}{A}, (A)^{-1}).$$

$$\lambda_D^T \sim \mathcal{G}(1 + \frac{M}{2}; 1 + \frac{\sum_j (d_j^T - a^T)^2}{2}).$$

$$\lambda_D^P \sim \mathcal{G}(1 + \frac{M}{2}; 1 + \frac{\sum_j (d_j^P - a^P)^2}{2}).$$

The correlation coefficients between temperature and precipitation in the models, and their prior parameters:
Define $A = \xi_j^P \sum_t (X_{jt}^{T'} - d_j^T)^2 + \lambda_B$ and $B = \xi_j^P \sum_t (X_{jt}^{T'} - d_j^T)(X_{jt}^{P'} - d_j^P) + \lambda_B \beta_0$, then

$$\beta_{xj} \sim \mathcal{N}(\frac{B}{A}, (A)^{-1}).$$

Define $A = M\lambda_B + \lambda_o$ and $B = \lambda_B \sum_{j>0} \beta_{xj} + \lambda_o \beta_{xo}$, then

$$\beta_0 \sim \mathcal{N}(\frac{B}{A}, (A)^{-1}).$$

$$\lambda_B \sim \mathcal{G}(0.01 + \frac{M}{2}; 0.01 + \frac{\sum_j (\beta_{xj} - \beta_0)^2}{2}).$$

Precision terms for the models:

$$\xi_j^T \sim \mathcal{G}(a_{\xi^T} + \frac{\tau^*}{2}; b_{\xi^T} + \frac{\sum_t (X_{jt}^{T'} - d_j^T)^2}{2}).$$

$$\xi_j^P \sim \mathcal{G}(a_{\xi^P} + \frac{\tau^*}{2}; b_{\xi^P} + \frac{\sum_t (X_{jt}^{P'} - d_j^P - \beta_{xj}(X_{jt}^{T'} - d_j^T))^2}{2}).$$

Only the full conditionals of the hyperparameters $a_{\xi^T}, b_{\xi^T}, a_{\xi^P}, b_{\xi^P}$ cannot be sampled directly, and a Metropolis step is needed. We follow the solution described in Smith et al. (2009). The algorithm works identically for the two pairs, and we describe it for a_{ξ^T} and b_{ξ^T} (the sampling is done jointly for the pair). We define U_1, U_2 as independent random variables, uniformly distributed over the interval $(0, 1)$, and we compute two proposal values $a'_{\xi^T} = a_{\xi^T} e^{(\delta(u_1 - \frac{1}{2}))}$

and $b'_{\xi^T} = b_{\xi^T} e^{(\delta(u_2 - \frac{1}{2}))}$, where δ is an arbitrary increment, that we choose as $\delta = 1$. We then compute

$$\begin{aligned} \ell_1 &= M a_{\xi^T} \log b_{\xi^T} - M \log \Gamma(a_{\xi^T}) + (a_{\xi^T} - 1) \sum_j \log \xi_j^T - b_{\xi^T} \sum_j \xi_j^T + \\ & \quad 0.01 \log(a_{\xi^T} b_{\xi^T}) - 0.01(a_{\xi^T} + b_{\xi^T}), \end{aligned} \tag{68}$$

$$\begin{aligned} \ell_2 &= M a'_{\xi^T} \log b'_{\xi^T} - M \log \Gamma(a'_{\xi^T}) + (a'_{\xi^T} - 1) \sum_j \log \xi_j^T - b'_{\xi^T} \sum_j \xi_j^T + \\ & \quad a \log(a'_{\xi^T} b'_{\xi^T}) - b(a'_{\xi^T} + b'_{\xi^T}). \end{aligned} \tag{69}$$

In (68) and (69) we are computing the log likelihoods of (a_{ξ^T}, b_{ξ^T}) and (a'_{ξ^T}, b'_{ξ^T}). Then, within each iteration of the Gibbs/Metropolis algorithm, the proposed values (a'_{ξ^T}, b'_{ξ^T}) are accepted with probability $e^{\ell_2 - \ell_1}$ if $\ell_2 < \ell_1$, or 1 if $\ell_2 \geq \ell_1$.

An R program that implements this simulation is available as REA.BV.r from `http://www.image.ucar.edu/~tebaldi/REA/`.

Acknowledgements

We thank Linda O. Mearns, Doug Nychka and Bruno Sanso' who have been working with us on various aspects of these analyses. Richard Smith is supported by NOAA grant NA05OAR4310020. Claudia Tebaldi is grateful to the Department of Global Ecology, Carnegie Institution of Washington, Stanford, and the National Center for Atmospheric Research, Boulder for hosting her at the time when this chapter was written.

References

Brooks, S. and A. Gelman (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics 7*, 434–455.

Furrer, R., S. Sain, D. Nychka, and G. Meehl (2007). Multivariate Bayesian analysis of atmosphere-ocean general circulation models. *Environmental and Ecological Statistics 14*(3), 249–266.

Gamerman, D. and H. Lopes (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall. Second Edition.

Gelman, A., J. Carlin, H. Stern, and D. Rubin (2003). *Bayesian Data Analysis*. Chapman & Hall, CRC Press. Second Edition.

Gelman, A., G. Roberts, and W. Gilks (1996). Efficient Metropolis jumping rules. In J. Bernardo, J. O. Berger, D. A.P., and S. A.F.M. (Eds.), *Bayesian Statistics 5*, pp. 599–607. Oxford University Press.

Gelman, A. and D. Rubin (1992). Inference from iterative simulation using multiple sequences. *Statistical Science 7*, 457–511.

Giorgi, F. and R. Francisco (2000). Evaluating uncertainties in the prediction of regional climate change. *Geophysical Research Letters 27*(9), 1295–1298.

Gleckler, P. J., K. Taylor, and C. Doutriaux (2007). Performance metrics for climate models. *Journal of Geophysical Research 113*(D06104). doi:10.1029/2007JD008972.

Groves, D. G., D. Yates, and C. Tebaldi (2008). Uncertain global climate change projections for regional water management planning. *Water Resources Research 44*(W12413). doi:10.1029/2008WR006964.

Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika 57*, 97–109.

Heidelberger, P. and P. Welch (1981). A spectral method for confidence interval generation and run length control in simulations. *Communications of the ACM 24*, 233–245.

IPCC (2007). *Climate Change 2007 – The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the IPCC*. Solomon, S. et al. (eds.), Cambridge University Press. 996 pp.

Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics 21*, 1087–1092.

Nakicenovic, N. (2000). *Special Report on Emissions Scenarios: A Special Report of Working Group III of the Intergovernmental Panel on Climate Change*. Cambridge University Press. 599 pp.

Pasarica, C. and A. Gelman (2008). Adaptively scaling the Metropolis algorithm using expected squared jumped distance. *Statistica Sinica*. to appear.

Plummer, M., N. Best, K. Cowles, and K. Vines (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC. *R NEWS 6*, 7–11.

R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Available from `http://www.R-project.org`.

Raftery, A. E., T. Gneiting, F. Balabdoui, and M. Polakowski (2005). Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review 133*, 1155–1174.

Robert, C. (2005). *The Bayesian Choice*. Springer Texts in Statistics. Second Edition.

Robert, C. and G. Casella (2004). *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Second Edition.

Rougier, J. (2007). Probabilistic inference for future climate using an ensemble of climate model evaluations. *Climatic Change 81*, 247–264. doi: 10.1007/s10584-006-9156-9.

Smith, R., C. Tebaldi, D. Nychka, and L. Mearns (2009). Bayesian modeling of uncertainty in ensembles of climate models. *Journal of the American Statistical Association*. to appear.

Tebaldi, C. and R. Knutti (2007). The use of the multi-model ensemble in probabilistic climate projections. *Philosophical Transactions of the Royal Society, Series A 1857*, 2053–2075.

Tebaldi, C. and D. Lobell (2008). Towards probabilistic projections of climate change impacts on global crop yields. *Geophysical Research Letters 35*(L08705), doi:10.1029/2008GL033423.

Tebaldi, C., L. Mearns, D. Nychka, and R. Smith (2004). Regional probabilities of precipitation change: A Bayesian analysis of multimodel simulations. *Geophysical Research Letters 31*(L24213), doi:10.1029/2004GL021276.

Tebaldi, C. and B. Sansó (2008). Joint projections of temperature and precipitation change from multiple climate models: A hierarchical Bayesian approach. *Journal of the Royal Statistical Society*. Series A, (forthcoming).

Tebaldi, C., R. Smith, D. Nychka, and L. Mearns (2005). Quantifying uncertainty in projections of regional climate change: A Bayesian approach to the analysis of multi-model ensembles. *Journal of Climate 18*(10), 1524–1540.

Washington, W. and C. L. Parkinson (2005). *Introduction to Three-dimensional Climate Modeling*. University Science Books. 368 pp.

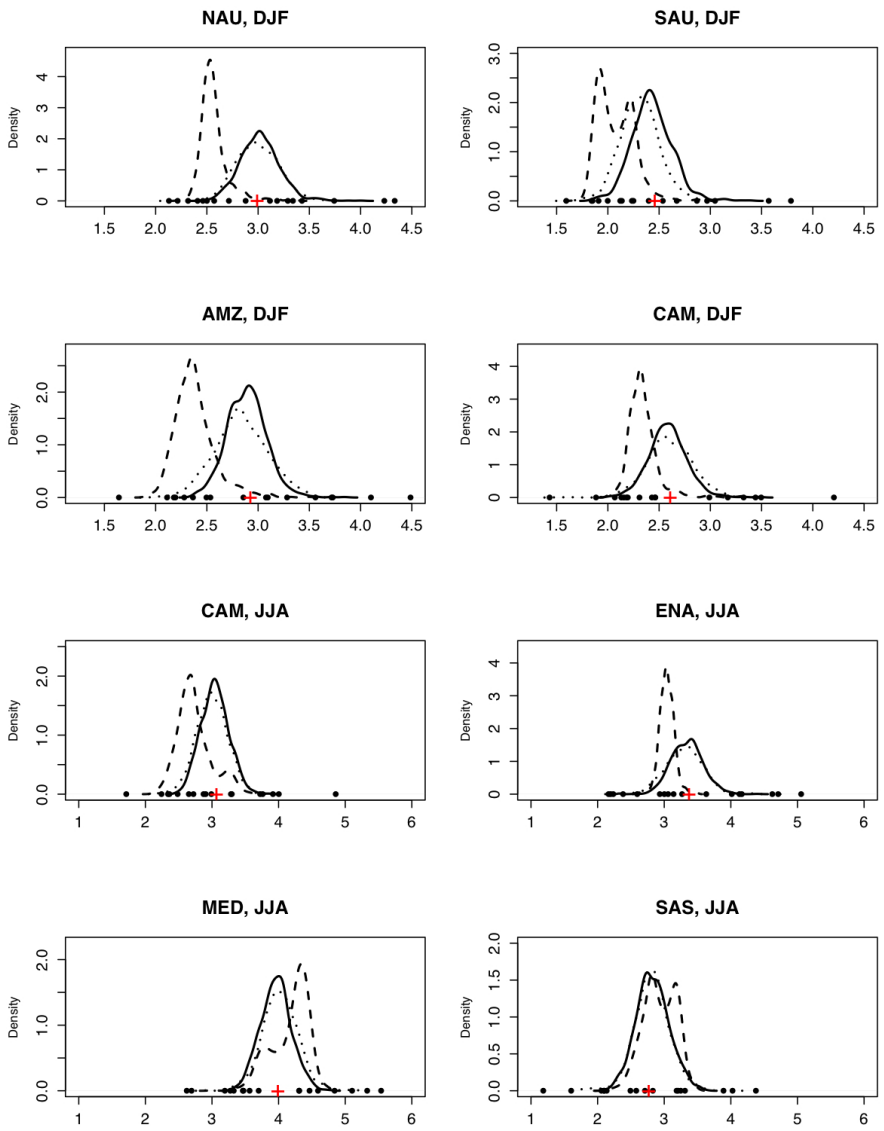

Figure 1: Posterior PDFs of temperature change for a set of four regions and two seasons (December through January, DJF and June through August, JJA). Dashed line: simplest univariate model; dotted line: univariate model with common prior for λ_js; solid line: multivariate model.

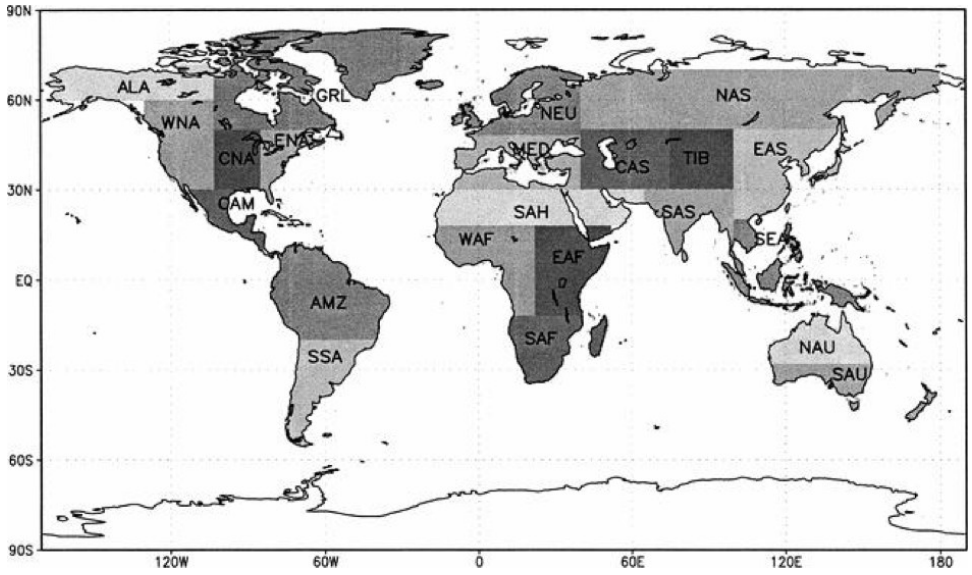

Figure 2: The 22 Giorgi regions. Output of temperature and precipitation for each GCM and observations is area-averaged over each of these large regions, for a given season and (multi-) decadal periods.

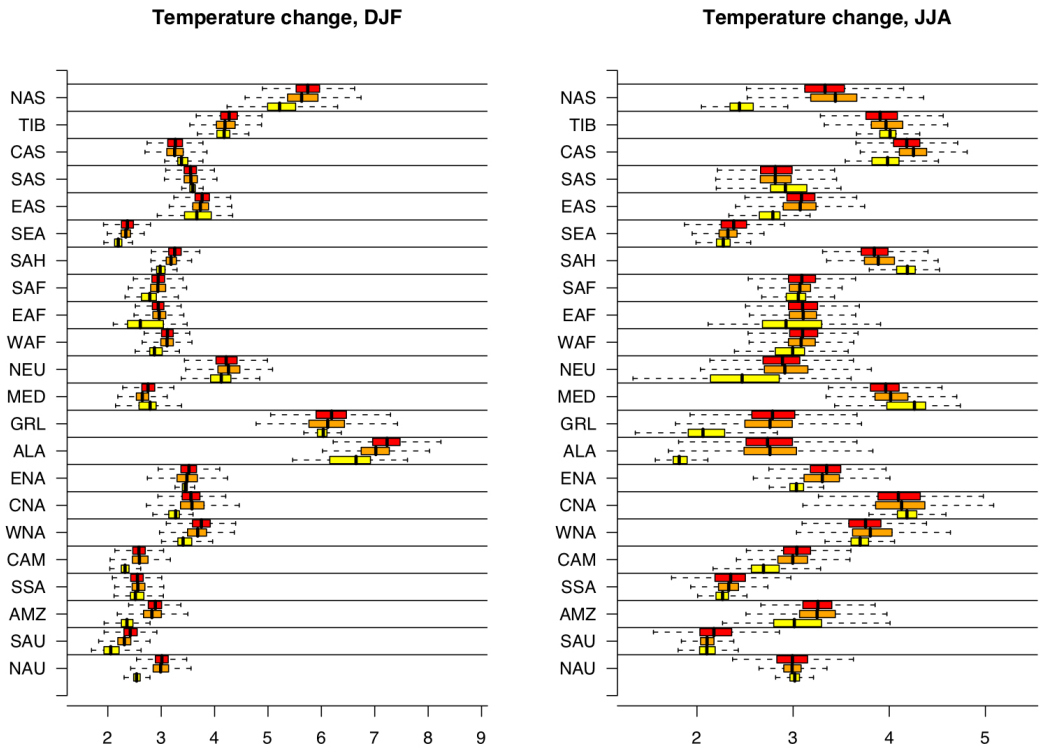

Figure 3: Posterior distributions of temperature change for the 22 regions of Figure 2 in DJF (left panel) and JJA (right panel). For each region we show three boxplots corresponding to the posterior distributions derived from the three statistical models represented also in Figure 1. The labeling along the vertical axis and the horizontal lines in the plot region identify each region by the acronyms of Figure 2. For each region, the lower boxplot, lightest in color, corresponds to the simplest univariate model. The middle boxplot corresponds to the univariate model with hyperprior on the reliability parameters. The third boxplot, darkest in color corresponds to the multivariate model.

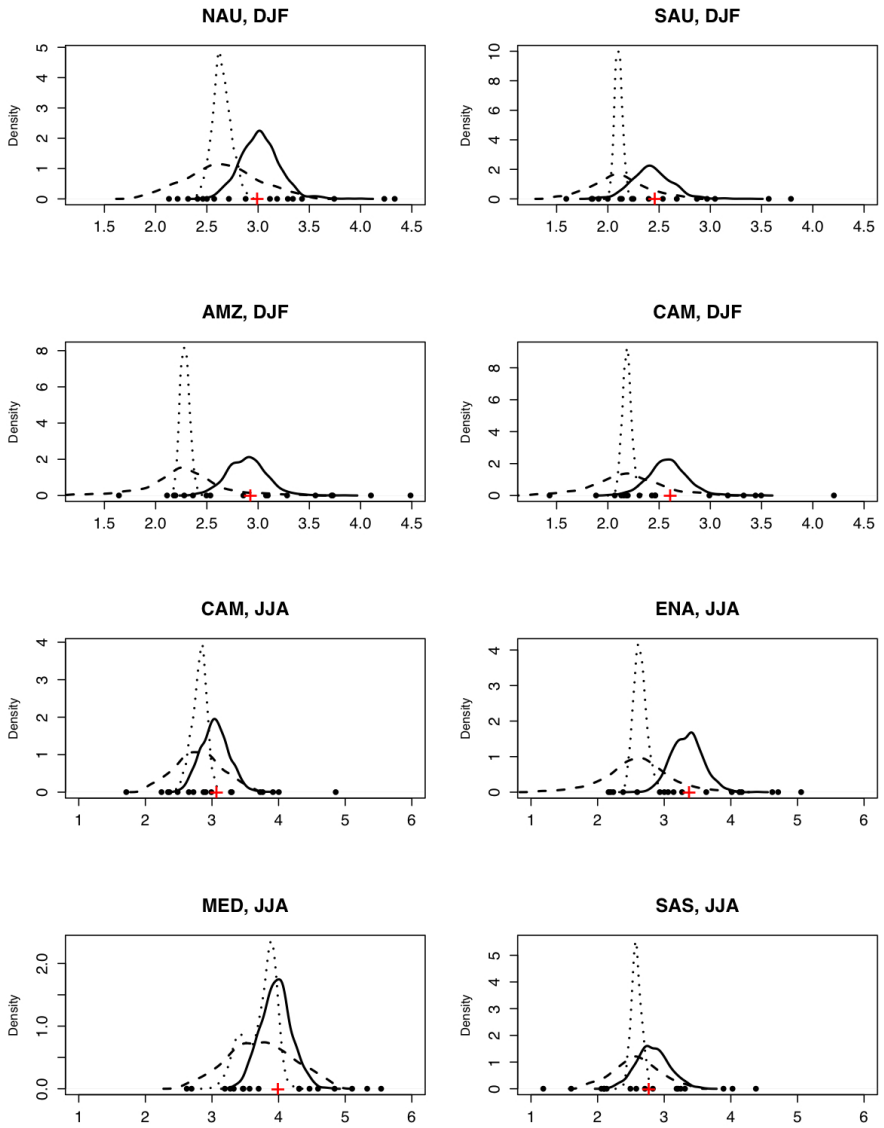

Figure 4: Similar to Figure 1, but now the three PDFs are the posterior from the multivariate model (solid line), the posterior from the bivariate model (dotted line) and the posterior predictive from the bivariate model (dashed line). The curves drawn as solid lines are the same as in Figure 1

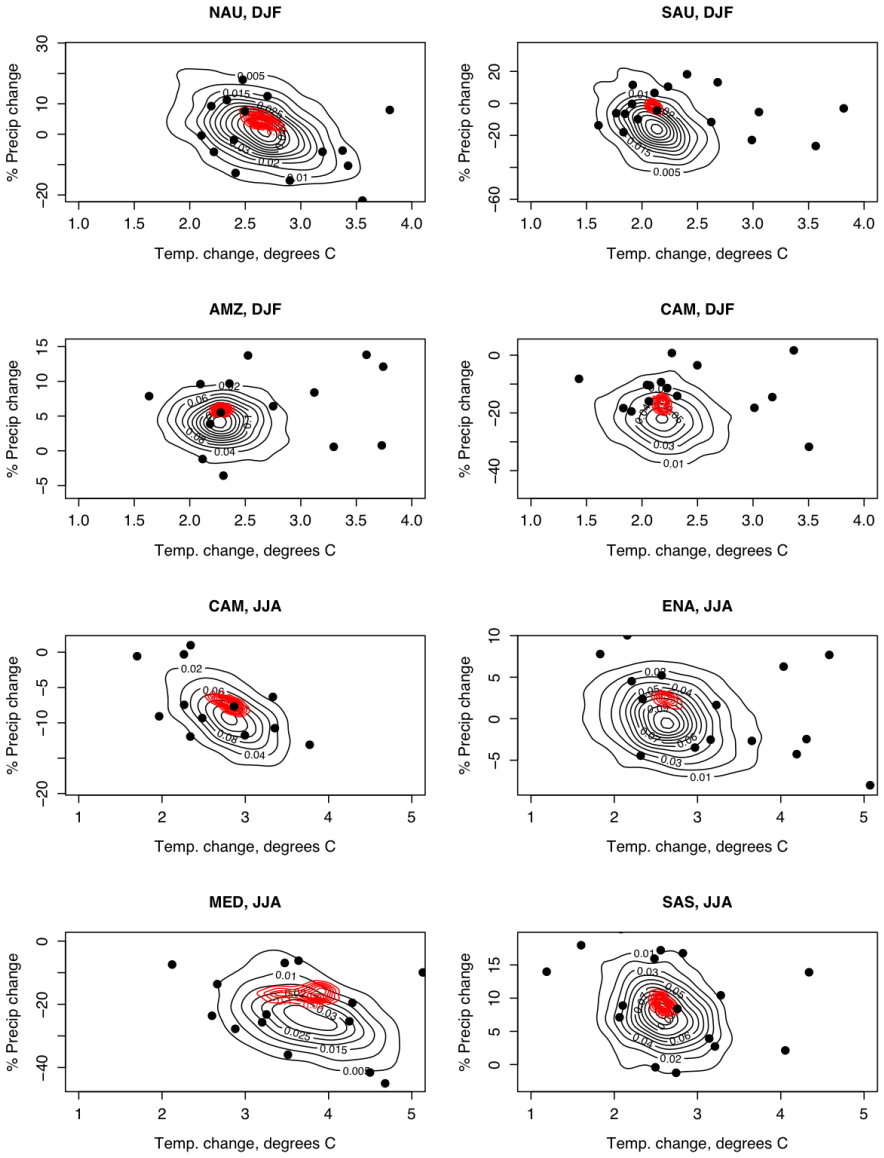

Figure 5: Contours of joint changes in temperature and precipitation (the latter as percentage of current average precipitation) for the same set of regions/seasons as in Figures 1 and 4. The tight contours correspond to the posterior PDF from the bivariate model. The wider contors represent the posterior predictive PDFs for the projections of a new GCM, unbiased.

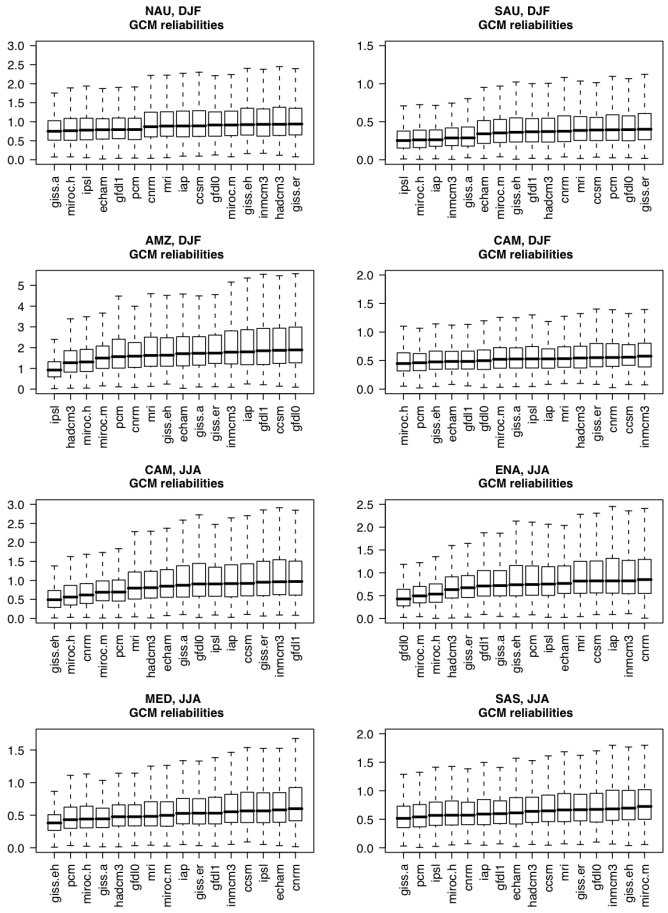

Figure 6: Posterior PDFs for the model reliability parameters λ_js derived from the univariate model with common prior on them. The region/season combinations are the same as in Figures 1 and 4. Each boxplot in a given panel corresponds to a GCM. Relatively larger values of λ_j are interpretable as better model reliability in the given region/season.

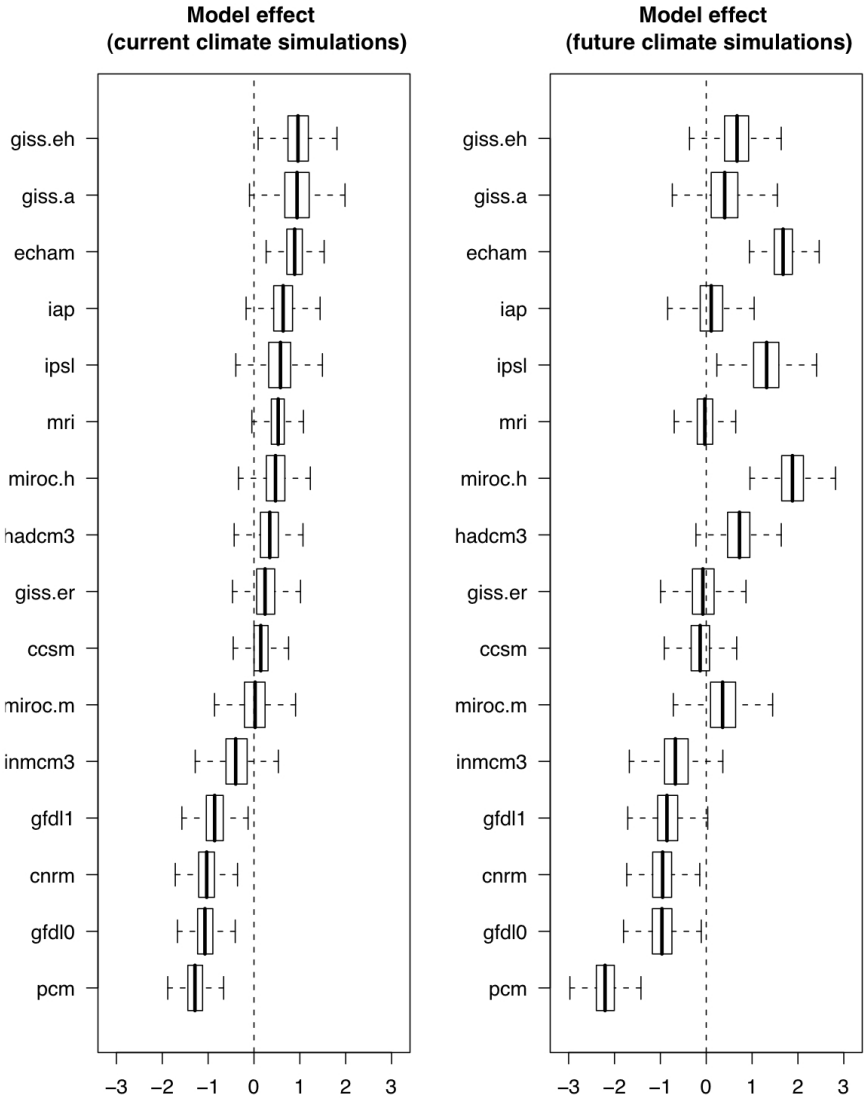

Figure 7: Posterior distributions of model-specific mean effects in the multivariate model. The left panel shows boxplots for the α_js, interpretable as global average biases in each model's current climate simulation. The corresponding distributions of the α'_js, interpretable as future simulation biases, are shown in the panel on the right.

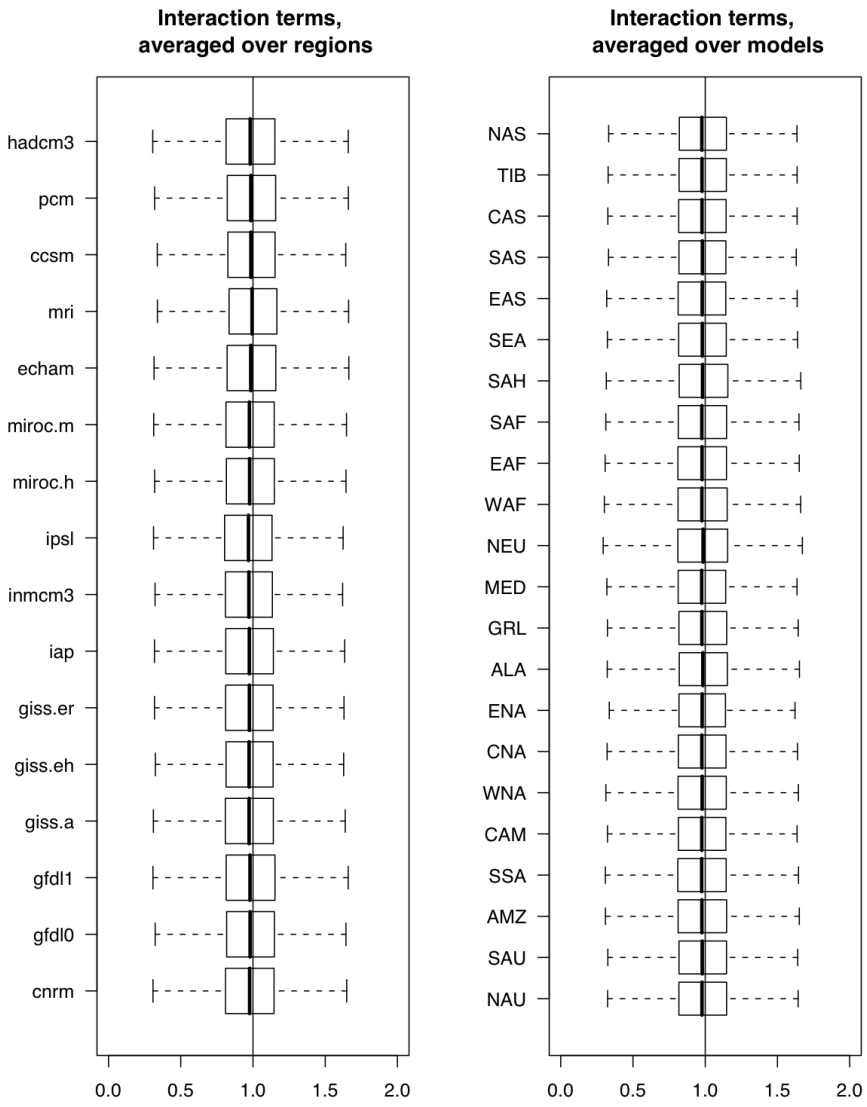

Figure 8: Interaction effect parameters (η_{ij}) of the precisions in the likelihood of the multivariate model. Left panel shows their posterior distributions integrated over regions ($\eta_{.j}$). Right panel shows the distributions integrated over models ($\eta_{i.}$).

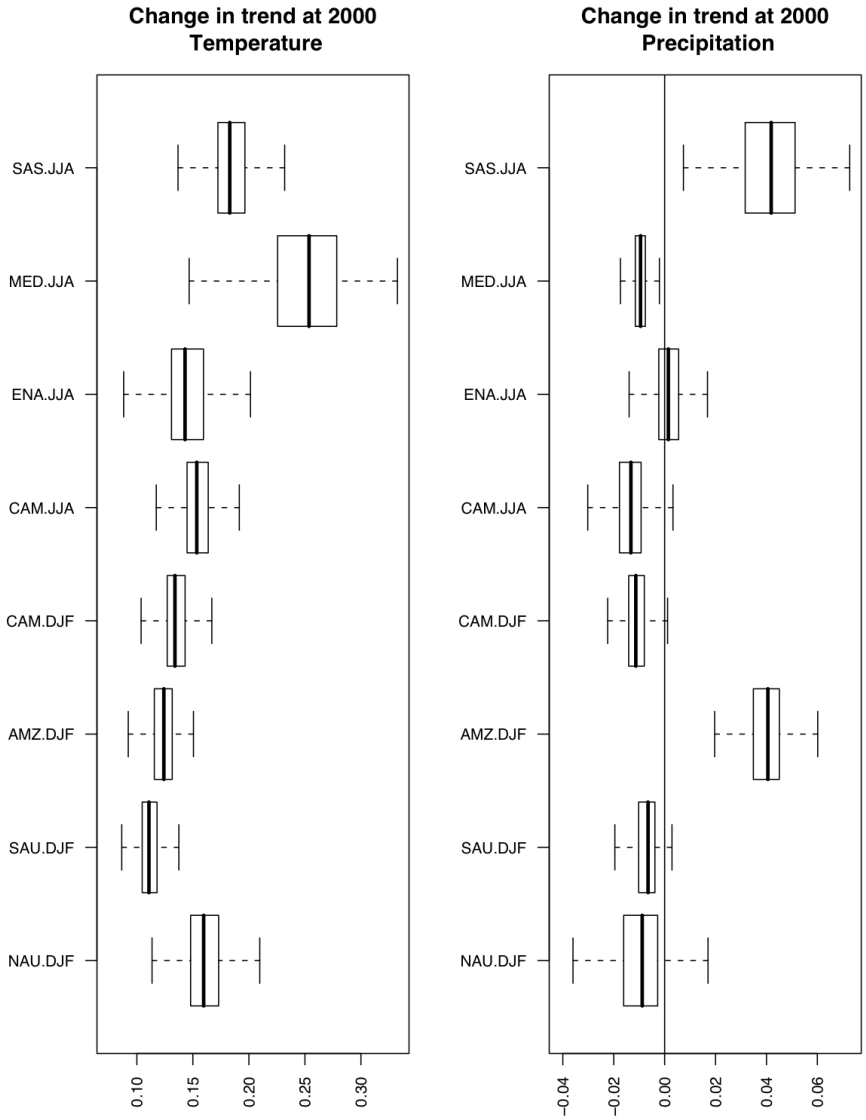

Figure 9: Incremental trend parameters γ^T (temperature time series) and γ^P (precipitation time series) in the bivariate model. When significantly different from zero they estimate a change in the slope of the linear trend after 2000.