Quantifying Uncertainty in Projections of Regional Climate Change: A Bayesian Approach to the Analysis of Multi-model Ensembles

Claudia Tebaldi [1], Richard L. Smith [2] , Doug Nychka[3]
and Linda O. Mearns[3]

June 9, 2004

[1]Project Scientist, National Center for Atmospheric Research, Boulder, CO
[2]Professor, University of North Carolina, Chapel Hill, NC
[3]Senior Scientist, National Center for Atmospheric Research, Boulder, CO

Abstract

A Bayesian statistical model is proposed that combines information from a multi-model ensemble of atmosphere-ocean general circulation models and observations to determine probability distributions of future temperature change on a regional scale. The posterior distributions derived from the statistical assumptions incorporate the criteria of bias and convergence in the relative weights implicitly assigned to the ensemble members. This approach can be considered an extension and elaboration of the Reliability Ensemble Averaging method. For illustration, we consider output of mean surface temperature from 9 AOGCMs, run under the A2 SRES scenario, for Boreal winter and summer, aggregated over 22 land regions. The shapes of the final probability density functions of temperature change vary widely, from unimodal curves for regions where model results agree, or outlying projections are discounted due to bias, to multimodal curves where models that cannot be discounted on the basis of bias give diverging projections. Besides the basic statistical model, we consider including correlation between present and future temperature responses, and test alternative forms of probability distributions for the model error terms. We suggest that a probabilistic approach, particularly in the form of a Bayesian model, is a useful platform from which to synthesize the information from an ensemble of simulations, at regional scales. The probability distributions of temperature change reveal features such as multi-modality and long tails that could not otherwise be easily discerned. Furthermore, the Bayesian model can serve as an interdisciplinary tool through which climate modelers, climatologists, and statisticians can work more closely. For example, climate modelers, through their expert judgment, could contribute to the formulations of prior distributions in the statistical model.

1 Introduction

Numerical experiments based on Atmosphere-Ocean General Circulation Models (AOGCMs) are one of the primary tools used in deriving projections for future climate change. However, the strengths and weaknesses that individual AOGCMs display have led authors of model evaluation studies to state that *"no single model can be considered 'best' and it is important to utilize results from a range of coupled models"* (McAvaney *et al.*, 2001). In this paper we propose a probabilistic approach to the synthesis of climate projections from different AOGCMs, in order to produce probabilistic forecasts of climate change.

Determining probabilities of future global temperature change has flourished as a research topic in recent years (Schneider, 2001; Wigley and Raper, 2001; Forest *et al.*, 2002; Allen *et al.* 2000). This work has focused on energy balance or reduced climate system models that facilitate many different model integrations and hence allow the estimation of a distribution of climate projections. However, the focus on low-dimensional models prevents a straightforward extension of this work to regional climate change analyses, which are the indispensable input for impacts research and decision making. Thus, we draw on more conventional AOGCM experiments in order to address specifically regional climate change. Probabilistic forecasts at the level of resolution of the typical AOGCM is a feat of enormous complexity. The high dimensionality of the datasets, the scarcity of observations in many regions of the globe and the limited length of the observational records have required ad hoc solutions in detection and attribution studies, and it is hard to structure a full statistical approach, able to encompass these methods and add the further dimension of the super-ensemble dataset. At present, we offer a first step towards a formal statistical treatment of the problem, by examining regionally averaged temperature signals. Thus, we greatly simplify the dimensionality of the problem, while still addressing the need for regionally differentiated analyses.

1.1 Model bias and model convergence

In the absence of overt tuning, the ability of an AOGCM to reproduce current regional climate is an important factor of its reliability. But a small model *bias* does not necessarily imply accurate future projections. The criterion of model *convergence* has been advocated by some authors (Raisanen, 1997; Giorgi and Francisco, 2000) as an additional means of assessing model reliability. Based on convergence, a member of a multimodel ensemble is given more weight if its future projection or climate change signal agrees with the majority of other ensemble members. Conversely, a member is downweighted if its projection differs substantially from the ensemble majority. This criterion can be theoretically derived from an assumption that the observed AOGCMs form a random sample from some superpopulation of AOGCMs. This viewpoint arguably ignores some practical realities of climate modeling — for example, the possibility that one model may include some vital feedback mechanism that all the others ignore — but the statistical framework that we develop is flexible enough to allow different definitions of what are "consensus" and "outliers", and thereby test the sensitivity of the predictions to the definitions we adopt.

Recently, Giorgi and Mearns' Reliability Ensemble Average (REA) method (Giorgi and Mearns, 2002) quantified the two criteria of bias and convergence for multimodel evaluation, and produced estimates of regional climate change, their uncertainty bounds and model reliabilities through a weighted average of the individual AOGCM results. The REA weights contain a measure of model bias with respect to current climate and a measure of model convergence (applied to the models' projected change), defined as the deviation of the individual projection with respect to the central tendency of the ensemble (i.e., the final weighted average). In a subsequent note Nychka and Tebaldi (2003) show that the REA method is in fact a conventional statistical estimate for the center of a distribution that departs from a Gaussian shape by showing heavier tails. Thus, although the REA estimates were proposed by Giorgi and Mearns as a reasonable quantification of heuristic criteria, there exists a formal statistical model that can justify them as an optimal procedure. The statistical model presented here will also produce estimators qualitatively similar to the REA method.

1.2 A Bayesian approach to AOGCM uncertainty

There is interest in the geosciences in moving from single value predictions to probabilistic forecasts, in so far as presenting a probability distribution of future climate is a more flexible quantification for drawing inferences and serving decision making (Reilly *et al.* 2001; Webster, 2003; Dessai and Hulme, 2003). Bayesian methods are not the only option when the goal is a probabilistic representation of uncertainty, but are a natural way to do so in the context of climate change projections, where the asymptotic arguments for the distributional properties of the estimates produced by the maximum likelihood approach are not applicable.

Within the Bayesian approach, a crucial step is the choice of prior distributions for the quantities of interest. Within the climate community, different viewpoints have been expressed regarding the use of expert judgement in quantifying uncertainty. For example, Wigley and

Raper (2001) explicitly chose prior distributions to be consistent with their own expert judgement of the uncertainty of key parameters such as climate sensitivity, and Reilly *et al.* (2001) argued that such assessments should be part of the IPCC process for quantifying uncertainty in climate projections. Opposing this view, Allen *et al.* (2001) argued that "no method of assigning probabilities to a 100-year climate forecast is sufficiently widely accepted and documented in the refereed literature to pass the extensive IPCC review process." (They suggested that there might be better prospects of success with a 50-year forecast, since over this time frame there is better agreement among models with respect to both key physical parameters and the sensitivity to different emissions scenarios.) Following this philosophy, Forest *et al.* (2001) stated "an objective means of quantifying uncertainty...is clearly desirable" and argued that this was achievable by choosing parameters "that produce simulations consistent with 20th-century climate change". However, their analysis did not use multi-model ensembles. We view the present approach as an extension of the philosophy reflected in the last two quotations, where we use uninformative prior distributions but use both model-generated and observational data to calculate meaningful posterior distributions. Another point in favor of uninformative prior distributions is that they typically lead to parameter estimates similar to non-Bayesian approaches such as maximum likelihood. However, Bayesian methods are more flexible when combining different sources of uncertainty, such as those derived from present-day and future climate model runs, and we regard this as a practical justification for adopting a Bayesian approach.

1.3 Outline

Section 2 contains the description of the basic statistical model and its extensions. Prior distributions for the parameters of interest, and distributional assumptions for the data, conditional on these parameters, are described. Some approximations to the posterior distribution are also presented in order to gain insight into the nature of the statistical model and its results. In Section 3 the model is applied to the same suite of AOGCM experiments as in Giorgi and Mearns (2002), and some findings from the posterior distributions are presented. In Section 4 we conclude with a discussion of our model assumptions and what we consider promising directions for extending this work. Appendix 1 is an introduction to Bayesian statistical modeling tailored to this study with an example of a prior-posterior update that has a closed-form solution. The Markov chain Monte Carlo (MCMC) method used to estimate the posterior distributions is described in Appendix 2.

2 Statistical models for AOGCM projections

In the Bayesian framework all parameters of interest, such as the value of temperature for present and future climate or the variability of a particular AOGCM response, are regarded as random quantities, in so far as they are uncertain. The analysis has three components: prior, likelihood and posterior, and we introduce them below. Subsequent sections give the analytical details for the basic model and two alternatives.

Prior choice

Prior distributions for all random quantities are elicited independently of the data at hand. In cases where there is not sufficient agreement among experts to determine a prior distribution, and no data from previous studies may be incorporated in the prior specification, *uninformative* priors are typically chosen, allowing a wide range of possible values, and having the minimal effect, relative to the data, on the final inference (Bernardo and Smith, 1994, Ch.4, Sect.5). Following the arguments presented in Section 1.2, uninformative priors are chosen for all the random parameters in our model.

Likelihood

The second component of a Bayesian model is the distribution (likelihood) of the data, as function of the random parameters. We assume that the AOGCM responses have a symmetric distribution, whose center is the "true value" of temperature, but with an individual variability, to be regarded as a measure of how well each AOGCM approximates the climate response to the given set of natural and anthropogenic forcings. In the presence of an ensemble of realizations for a given AOGCM we could address the issue of natural internal variability of the model as well, with a simple modification of our statistical assumptions. The assumption of a symmetric distribution around the "true value" of temperature for the suite of multi-model responses has been implicitly supported by CMIP studies (Meehl *et al.*, 2000) where better validation properties have been demonstrated for the mean of a super-ensemble rather than the individual members. The additional assumption in our model that the single AOGCM's realizations are centered around the "true value" could be easily modified in the presence of reliable data on systematic biases, through a straightforward correction of the models' response, or by incorporating an AOGCM-specific random effect in the mean response if single-model ensembles were available.

Posterior distribution

Through Bayes' Theorem prior and likelihood are combined into the joint posterior distribution of all random variables in the model. We refer to Gilks *et al.* (1996) for an introduction to Bayesian modeling, and examples of data analyses, but we also present in Appendix 1 a simple example of prior-posterior update, where the posterior can be computed analytically. When the complexity of the model precludes a closed-form solution, which is the case in our application, an empirical estimate of the posterior distribution can be obtained through Markov chain Monte Carlo (MCMC) simulation. MCMC techniques are efficient ways of simulating samples from the posterior distribution, by-passing the need of computing it analytically. Thus, approximations to the form of the posterior distribution can be made through smoothed histograms and other numerical summaries based on the MCMC samples. We describe the implementation of the MCMC simulation for our analysis in Appendix 2.

2.1 The model for a single region

In this section we present the analytical form of our basic statistical model. We list first the distributional assumptions for the data (likelihoods), then the priors for all random parameters. We then present conditional approximations to the posterior that will help to interpret our results. In the following, the index i refers to the set of 9 AOGCMs (see Table 1). Throughout, let X_i and Y_i denote the temperature simulated for the present and future by the i^{th} model, seasonally and regionally averaged.

Likelihoods

$$X_i \sim N(\mu, \lambda_i^{-1}), \tag{1}$$
$$Y_i \sim N(\nu, (\theta\lambda_i)^{-1}). \tag{2}$$

Here μ and ν are the true values of present and future temperature in a specific region and season. A key parameter of interest will be $\Delta T \equiv \nu - \mu$, representing the expected temperature change. For X_i, the parameter λ_i is the reciprocal of the variance and will be referred to as the precision of model i. To allow for the possibility that Y_i has different precision from X_i, we write its precision $\theta\lambda_i$ where θ is a parameter to be estimated. A more general model might allow for θ to be different in different models, but that is not possible in the present set up given the limited number of data points from which we have to estimate these parameters.

The model (1)–(2) could be expressed alternatively through the linear equations:

$$X_i = \mu + \epsilon_i \quad (3)$$
$$Y_i = \nu + \epsilon'_i/\sqrt{\theta} \quad (4)$$

with the assumptions that the error terms have a common Gaussian distribution, and that ϵ_i is independent of ϵ_j for $i \neq j$, and ϵ_i is independent of ϵ'_i.

We model the likelihood of the observations of current climate as

$$X_0 \sim N(\mu, \lambda_0). \quad (5)$$

or equivalently

$$X_0 = \mu + \epsilon_0 \quad (6)$$

Here μ is the same as in (3), but λ_0 is of a different nature from $\lambda_1, \ldots, \lambda_9$. While the latter are measures of model-specific precision, and depend on the numerical approximations, parameterizations, grid resolutions of each AOGCM, λ_0 is a function of the natural variability specific to the season, region and time-average applied to the observations. We use estimates of regional natural variability from Giorgi and Mearns (2002), interpretable as the standard deviation of the three-decadal average that X_0 represents (see Table 2 for the specific values of natural variability for DJF and JJA in the 22 regions). Thus, we fix the value of the parameter λ_0 to the reciprocal of the squared value of these estimates. The parameter λ_0 could be given a full Bayesian treatment if we had a sufficiently long record of observations that we could use for its estimation.

Prior distributions

The model described by (1)-(6) is formulated as a function of the random parameters $\mu, \nu, \theta, \lambda_1, \ldots, \lambda_9$. Prior distributions are chosen for each of these parameters as follows:

1. λ_i, $i = 1, \ldots, 9$ have Gamma prior densities (indicated by $\mathrm{Ga}(a, b)$), of the form

$$\frac{b^a}{\Gamma(a)} \lambda_i^{a-1} \exp -bx$$

with a, b known and chosen so that the distribution will have a large variance over the positive real line. Similarly, $\theta \sim \mathrm{Ga}(c, d)$, with c, d known. These are standard prior choices for the precision parameters of Gaussian distributions. In particular they are *conjugate*, in the sense of having the same general form as the likelihood considered as a function of λ_i, thus ensuring the computational tractability of the model. We choose $a = b = c = d = 0.001$, so that the Gamma distributions, even if proper, approximate the non-informative characteristic that we require in our analysis.

2. True climate means, μ and ν for present and future temperature respectively, have Uniform prior densities on the real line. Even if these priors are improper (i.e. do not integrate to one), the form of the likelihood model ensures that the posterior is a proper density function. Alternative analyses in which the prior distribution is restricted to a finite, but sufficiently large, interval, do not in practice produce different results.

Posterior distributions
We apply Bayes' Theorem to the likelihood and priors specified above. The resulting joint posterior density for the parameters μ, ν, θ, $\lambda_1, \ldots, \lambda_9$ is given by, up to a normalizing constant,

$$\prod_{i=1}^{9} \left[\lambda_i^{a-1} e^{-b\lambda_i} \cdot \lambda_i \theta^{1/2} \exp\left\{ -\frac{\lambda_i}{2}((X_i - \mu)^2 + \theta(Y_i - \nu)^2) \right\} \right] \cdot \theta^{c-1} e^{-d\theta} \cdot \exp\left\{ -\frac{\lambda_0}{2}(X_0 - \mu)^2 \right\}. \tag{7}$$

The distribution in equation (7) is not a member of any known parametric family and we cannot draw inference from its analytical form. The same is true for the marginal posterior distributions of the individual parameters, that we cannot compute from (7) through closed-form integrals. Therefore, Markov chain Monte Carlo simulation is used to generate a large number of sample values from (7) for all parameters, and approximate all the summaries of interest from sample statistics. We implement the MCMC simulation through a Gibbs sampler, and details of the algorithm are given in Appendix 2.

One can gain some insight as to how the posterior distribution synthesizes the data and the prior assumptions by fixing some groups of parameters and considering the conditional posterior for the others. For example, the distribution of μ fixing all other parameters is a Gaussian distribution with mean

$$\tilde{\mu} = (\sum_{i=0}^{9} \lambda_i X_i)/(\sum_{i=0}^{9} \lambda_i) \tag{8}$$

and variance

$$\left(\sum_{i=0}^{9} \lambda_i \right)^{-1}. \tag{9}$$

In a similar fashion, the conditional distribution of ν is Gaussian with mean

$$\tilde{\nu} = (\sum_{i=1}^{9} \lambda_i Y_i)/(\sum_{i=1}^{9} \lambda_i) \tag{10}$$

and variance

$$\left(\theta \sum_{i=1}^{9} \lambda_i \right)^{-1}. \tag{11}$$

The forms (8) and (10) are analogous to the REA results, being weighted means of the 9 AOGCMs, and the observation, with weights $\lambda_1, \ldots, \lambda_9, \lambda_0$, respectively. As in the case of the REA method, these weights will depend on the data, but a fundamental difference is that in our analysis they are random quantities, and thus we account for the uncertainty in their estimation.

Such uncertainty will inflate the width of the posterior distributions of ν, μ, and thus also ΔT. An approximation to the mean of the posterior distribution of the λ_i's, for $i = 1, \ldots, 9$, is:

$$E(\lambda_i | \{X_0, \ldots, X_9, Y_1, \ldots, Y_9\}) \approx \frac{a+1}{b + \frac{1}{2}((X_i - \tilde{\mu})^2 + \theta(Y_i - \tilde{\nu})^2)}. \tag{12}$$

The form of (12) shows that the individual weight λ_i is large provided both $|X_i - \tilde{\mu}|$ and $|Y_i - \tilde{\nu}|$ are small. These two quantities correspond to the *bias* and *convergence* criteria respectively. $|Y_i - \tilde{\nu}|$ measures the distance of the i^{th} model future response from the overall average response, and so has characteristics similar to the convergence measure in Giorgi and Mearns (2002). The important difference from REA for this model is that the distance is based on the future projection (Y_i) rather than the temperature change ($Y_i - X_i$). As for the bias term, notice that in the limit, if we let $\lambda_0 \to \infty$ (we let the observation X_0 become an extremely precise estimate of the true temperature μ), $\tilde{\mu} \to X_0$, and the bias term becomes in the limit $|X_i - X_0|$, the same definition of bias as in the REA analysis. Finally, by having chosen $a = b = 0.001$ we ensure that the contribution of the prior assumption to (12) is negligible.

2.2 Introducing correlation between present and future climate responses within each AOGCM, and robustifying the model

Although the model presented in Section 2.1 reproduces the basic features of the REA method, in this section we modify our model in two important directions. First, we relax the assumption of independence between X_i and Y_i, for a given climate model. We do so by linking X_i and Y_i through a linear regression equation; this is equivalent to assuming that (X_i, Y_i) are jointly normal, given the model parameters, with an unknown correlation coefficient which is left free to vary, a-priori, between $+1$ and -1. The form of the correlated model is:

$$X_i = \mu + \eta_i \tag{13}$$

and, given X_i, μ and ν,

$$Y_i = \nu + \beta_x(X_i - \mu) + \xi_i/\sqrt{\theta}. \tag{14}$$

The parameter β_x, when not zero, will thus introduce a direct (if positive) or inverse (if negative) relation between $X_i - \mu$ and $Y_i - \nu$. By using a common parameter we force the sign of the correlation to be the same for all AOGCMs, a constraint that is required because only two data points (current and future temperature response) per model are available for a given region and season. The value of the β_x parameter has also implications in terms of the correlation between $Y_i - X_i$, the signal of temperature change produced by the i^{th} AOGCM, and the quantity $X_i - \mu$, the model bias for current temperature. A value of β_x equal to 1 translates into conditional independence of these two quantities, while values greater than or less than 1 would imply positive or negative correlation between them. (In the REA analysis, Giorgi and Mearns treat these two quantities as independent, after computing empirical estimates of their correlations and finding that they are for the most part low in absolute value and not statistically significant.) But the most important feature of the model with correlation has to do with the form of the posterior distribution for the λ_i parameters. By an approximation similar to what we use in Section 2.1 for the basic model, we derive that the posterior mean has now the approximate form:

$$E(\lambda_i|\{X_0,\ldots,X_9,Y_1,\ldots,Y_9\}) \approx \frac{a+1}{b+\frac{1}{2}((X_i-\tilde{\mu})^2+\theta(Y_i-\tilde{\nu}-\beta_x(X_i-\tilde{\mu}))^2)}. \qquad (15)$$

if $\beta_x \approx 1$ the form in (15) uses a convergence term now similar the REA method's, measuring the distance between $\Delta T_i \approx Y_i - X_i$ and the consensus estimate of temperature change $\tilde{\nu}-\tilde{\mu}$. We will compare the results from the basic model and this modified version, in order to assess how differently the two definitions of convergence reflect on the final posterior distributions of regional temperature changes.

A second concern is the sensitivity of our model to outlying data points. This is a consequence of hypothesizing normal distributions for X_i and Y_i, given the model parameters. Traditionally, robustness is achieved by choosing distributions with heavier tails than the Gaussian's, and Student-t distributions with few degrees of freedom are often a convenient choice because of flexibility (the degrees of freedom parameter allowing to vary the degree of their departure from the Gaussian) and computational tractability. Thus, we model η_i and ξ_i in equations (13) and (14) above as independently distributed according to a Student-t distribution, with ϕ degrees of freedom. We perform separate analyses for ϕ varying in the set $\{1,2,4,8,16,32,64\}$, the lower end of the range being associated with heavier tail distributions, the higher being equivalent to a Gaussian model. We are not a-priori labeling some of the AOGCMs as outliers, but we examine the difference in the posterior distributions that ensues as a function of the statistical model assumed for the error terms. Varying the degrees of freedom can accommodate extreme projections to a larger or smaller degree and is equivalent to the convergence criterion being less or more stringent, respectively. This achieves at least qualitatively the same effect as the exponents in the REA's weights, by which the relative importance of the bias and convergence terms could be modified.

3 The analysis: results from the Markov chain Monte Carlo simulation

3.1 Data

We analyse temperature responses under the A2 emission scenario (Nakicenovic *et al.* 2000), just a portion of the dataset analyzed by the REA method in Giorgi and Mearns. Output from 9 different AOGCMs' (see Table 1 and references therein) consists of present and future average surface temperatures, aggregated over the 22 regions in Figure 1, two seasons (DJF and JJA) and two 30-year periods (1961-1990; 2071-2100). Thus, we will separately treat (X_i, Y_i), $i = 1,\ldots,9$, for each of the 22 regions and 2 seasons. Observed temperature means for the 22 regions and the 1961-1990 period (X_0) were also taken from Giorgi and Mearns (2002), together with estimates of regional natural variability, that we list in Table 2 and use to determine the values of the parameter λ_0. The entire dataset can be downloaded from `www.cgd.ucar.edu/~nychka/REA`.

3.2 Results for a group of representative regions

We present posterior distributions derived from the basic model (Section 2.1) and the model introducing correlation between X_i and Y_i (Section 2.2), both models assuming Gaussian distributions for the likelihoods. We choose to focus on a representative group of six regions (Alaska

(ALA), East North America (ENA), Southern South America (SSA), Northern Europe (NEU), East Africa (EAF), Northern Australia (NAU)), for both winter (DJF) and summer (JJA) season.

3.2.1 Temperature change

Figures 2 and 3 show posterior distributions of temperature change $\Delta T \equiv \nu - \mu$ in the six regions, for winter and summer season respectively, for the basic model (solid lines) and for the model with correlation (dashed lines). For reference, the 9 models' individual responses $Y_i - X_i$, $i = 1, \ldots, 9$ are plotted along the x axis (as diamonds), together with Giorgi and Mearns' REA estimates of $\tilde{\Delta T}$ plus or minus 2 standard deviations (as a triangle superimposed on a segment). From the relative position of the diamonds one can qualitatively assess a measure of convergence for each model (defined in terms of the individual AOGCMs' temperature change response), and discriminate between models that behave as outliers and models that reinforce each other by predicting similar values of temperature change. The bias measure is not immediately recoverable from this representation, so we list in table 3 values of the bias for each of the 9 AOGCMs in the 6 regions, for summer and winter. A comparison of the densities in Figures 2 and 3 with the bias values listed in the table reveals how models having relatively smaller bias receive relatively larger weight. Models that perform well with respect to both criteria are the ones where the probability density function is concentrated. The REA results are consistent overall with mean and range of the dashed curves. For most of the regions (especially in winter) the solid and dashed lines are in agreement, with the solid lines showing a slightly wider probability distribution. This is expected because of the introduction, in (14) of the covariate X_i, that has the effect of making the distribution of Y_i more concentrated about ν. As a consequence the distribution of the difference $\nu - \mu$ is also less uncertain. For these region-season combinations, and for a majority of those not shown, it is the case that the two models, with or without correlation, result in almost identical posterior distributions of temperature change. They represent cases in which the 9 AOGCMs maintain the same behavior, when judged in terms of their absolute projections of future climate, or in terms of their temperature change signals. Those that can be defined as extreme are so in both respects, those that agree with each other, and form the consensus, do so in both respects.

For a few region-season combination (here we show three representative, EAF, NEU and NAU in summer) the posterior densities are dramatically different for the two statistical models. This is consequence of applying two different measures of convergence, and AOGCM results that are "outliers" with respect to one but are not so with respect to the other. In these instances, it matters which criteria of convergence we apply, and so the REA method is in agreement only with the dashed curves.

The shapes of the densities in Figure 2 and 3 vary significantly. Regular, unimodal curves are estimated in regions where there is obvious agreement among models, or where outlying models are downweighted due to large biases. Multimodal curves characterize regions where AOGCMs give disparate predictions, none of which can be discounted on the basis of model bias. These are also most often the cases when the results are sensitive to whether the basic model or the one with correlation is adopted. The value of presenting results in terms of a posterior distribution is obvious in the case of multimodal curves. For these densities the mean and median would not be good summaries of the distribution. Although we do not necessarily see these multimodal results as having an obvious physical interpretation, they definitely highlight the problematic nature of predictions in specific areas of the globe. For these regions our analysis gives more insight than traditional measures of central tendency and confidence intervals, and considering the sensitivity of the results to the statistical model assumptions is extremely relevant.

Thus, the introduction of the parameter β_x influences heavily the outcome of the analysis, for some region-season combinations, and thus deserves a careful consideration. For most of the regions its posterior concentrates to the right of zero, indicating a significant positive correlation between present and future temperature responses of all AOGCMs. Figure 7 shows the range of distributions of β_x for the six regions and two seasons. It is also interesting to notice that most of the mass of these posterior densities is concentrated around one, supporting the REA finding of weak correlation between temperature *change* signal and bias.

3.2.2 Precision parameters

Figures 4 and 5 summarize the posterior distributions for the precision parameters λ_i, in the form of boxplots (shown on a logarithmic scale, because of their high degree of skeweness). In our analysis λ_i's are the analogue to the model weights in REA's final estimates. However, λ_i is for us a random variable, and the scoring of the 9 AOGCMs should here be assessed through the relative position of the 9 boxplots, rather than by comparing point estimates. Comparing these distributions across models for a single region and season may suggest an ordering of the 9 AOGCMs: large λ_i's indicate that the distribution of the AOGCM response is more tightly concentrated around the true climate response. Thus, posterior distributions that are relatively shifted to the right indicate an AOGCM's better performance than distributions shifted to the left. It is significant that there is large overlap among these distributions, a sign that there is substantial uncertainty in the relative weighing of the models. So, caution should be exercised in using this criterion to rank models. Substantiating further this finding, we have calculated the posterior mean for each λ_i and standardized them to percentages. They are listed in Table 4 and 5, in the rows labeled as $\lambda_i / \sum \lambda_i (\%)$. The tables show clearly that the 9 AOGCMs are weighted differently in different regions, and seasons. This suggests differential skill in reproducing regional present day climate and a different degree of consensus among models for different regional signals of temperature change.

The REA method provided an overall measure of *reliability* for each of the 9 AOGCMs by summing up the values of the weights that each model had obtained over all 22 regions and ranking the outcomes. Table 6 presents our summary measure of the 9 models' relative weighting, by listing the median rank of each model over the 22 regions, separately for the winter and summer seasons. The table confirms that there is a difference in the relative weighting of the 9 AOGCMs depending on the season. All the results presented here are from the model with correlation, where the posterior means of the λ_i parameters have closer resemblance to the REA weights.

3.2.3 Inflation/deflation parameter

The parameter θ represents the inflation/deflation factor in the AOGCMs' precision when comparing simulations of present-day to future climate. When adopting the model with correlation the posterior for θ is always concentrated over a range of values greater than one, as a consequence of the tightening effect of the form (14) described earlier. As for the posterior distributions derived from the basic model, we present them for each of the six regions in Figure 6, in the form of boxplots. For some of the regions, here, the range is concentrated over values less than one, and a deterioration in the degree of precision of the 9 AOGCMs is suggested. As can be assessed by the boxplots in Figure 6, which are again representative of the whole set of results, each region tells a different story, and so does each season. For East North America, the simulations (i.e., their agreement) deteriorate in the future, and this holds true for both seasons. For Northern Australia, especially in the summer, and for Southern South America in both seasons the contrary appears to be true. For Alaska the behavior is different in winter (where a loss of precision is indicated) than in summer. For the two other regions, East Africa and Northern Europe, the inference is not as clear cut. We cannot offer any insight into the reason why this is so, our expectation being that the projections in the future, in the absence of the regression structure (14), would be at least as variable around the central mean as the current temperatures simulations are, likely more so. Considering in detail the properties of the AOGCMs for these regions may help to interpret these statistical results. However, we point out that the assumption of a common parameter for all AOGCMs may be too strict, and these results may change if a richer dataset, with single-model ensembles, allowed us to estimate model-specific θ factors.

3.2.4 Introducing heavy-tail distributions

We estimated alternative statistical models, with heavier tail characteristics, by adopting Student-t distributions for X_i and Y_i, as explained in Section 2.2.

Different values of degrees of freedom (ϕ in the notation of Section 2.2) were tested, from 1 (corresponding to heavy-tail distributions, more accommodating of outliers) to 64 (corresponding to a distribution that is nearly indistinguishable from a Gaussian). By comparing the posterior densities so derived, we can assess how robust the final results are to varying statistical assumptions. It is the case for all regions that the overall range and shape of the temperature change distributions are insensitive to the degrees of freedom, and this holds true when testing either the basic model, or the model with correlation. Only for a small number of regions do 1/2 degrees of freedom produce posterior distributions of climate change that spread their mass towards the more isolated values among the 9 AOGCM responses. The difference from the basic Gaussian model is hardly detectable in a graph comparing the posterior densities from different model formulations. We conclude that the posterior estimates are not substantially affected by the distributional assumptions on the tails for X_i and Y_i.

4 Discussion, extensions and conclusions

4.1 Overview of temperature change distributions

Figure 8 and 9 look at differences in the estimates of temperature change and uncertainty among regions, in the form of a series of boxplots, sorted by their median values. We use the model with correlation but no heavy tail modification. Boxplots represent the relative position of a few quantiles of the distributions (median, inter-quartile range and 0.05, 0.95 quantiles), and are useful for assessing different magnitudes of warming, and different degrees of uncertainty (variability) across regions and between seasons. Notice that all distributions are limited to the positive range, making the case for global warming. Warming in winter is on average higher than in summer, for all regions, and high latitudes of the northern hemisphere are the regions with a more pronounced winter climate change. These all are by now undisputed results from many different studies of climate change (Cubasch *et al.* 2001). The variability of the distributions is widely different among regions, supporting the notion that for some regions the signal of climate change is stronger and less uncertain than for others. Regions such as East North America (ENA), Southern South America (SSA), Eastern Asia (EAS), Mediterranean (MED) in winter, West Africa (WAF), Southern Australia (SAU) in summer show extremely tight distributions, predicting the number of degrees of warming with relative certainty. On the other hand, regions like Northern Asia (NAS) and Central Asia (CAS), in winter, East Africa (EAF) and West North America (WNA) in summer and Amazons (AMZ), Central North America (CNA) and Northern Europe (NEU) in both seasons show a wide range of uncertainty in the degrees of warming. As we already mentioned, the width and shape of these distributions may be taken as a signal of the degree of agreement among AOGCMs over the temperature projection in the regions. Problematic regions, characterized by multimodal or simply diffuse distributions may suggest areas that merit special attention by the climate modeling community.

4.2 Advantages of a Bayesian approach

The results in this paper demonstrate how the quantitative information from a multi-model experiment can be organized in a coherent statistical framework based on a limited number of explicit assumptions. These models are constrained in their assumptions and form by the available number and nature of the data points. Within these limits, heuristic criteria can be formalized and subsumed within an estimation procedure that is based on well understood statistical properties.

Our Bayesian analysis yields posterior probability distributions of all the uncertain quantities of interest. These posterior distributions for regional temperature change and for a suite of other parameters provide a wealth of information about AOGCM's reliability and temporal (present to future) correlations. We deem this distributional representation more useful than a point estimate with error bars, because important features such as multi-modality or long tails become evident. Also, the uncertainty quantified by the posterior distributions is an important product of this analysis and it is not easily recovered using non-Bayesian methods. Recent work (Raisanen and Palmer, 2001; Giorgi and Mearns, 2003) has addressed the need for probability forecasts, but offered only an assessment of the probability of exceeding thresholds, identified as the (weighted) fraction of the ensemble members for which the thresholds in question were exceeded. Moreover, in the case where each ensemble member is equally weighted, issues of model validation are ignored. A differential weighting of the members on the basis of the REA method is more appropriate but depends on a subjective heuristic formulation that is difficult to evaluate. In contrast, we think that the Bayesian approach is not only flexible but facilitates an open debate of the assumptions that generate probabilistic forecasts.

4.3 Sensitivity analysis of the statistical assumptions

The posterior distributions of regional temperature change in many region-season combinations may differ both in variance and shape, depending on the statistical model adopted, i.e., when introducing the correlation structure between present and future model projections. These sensitivities suggest the need to enrich the experimental setting, thus formulating the statistical model as free as possible from arbitrary constraints. For example, imposing a common correlation parameter β_x among models is a strong assumption that could be relaxed if more information were available. Similarly, single model ensembles would make it possible to estimate the internal variability of the AOGCM and allows for the separation of intra-model versus inter-model variability. In addition, climatological information could be incorporated into the prior. The range of our posteriors for the present and future temperature means would remain the same in the presence of proper, but still uninformative priors (i.e. either Uniform distributions limited to a physically reasonable range, or Gaussian distributions concentrating most of their mass over such ranges). However, a modeling group may have confidence that results in a particular region are positively biased, or relatively more accurate than in other regions, on the basis of separate experiments. This kind of information could be applied to the formulation of a different prior assumption for that AOGCM's precision parameter. In our opinion, analyzing the robustness of results to model assumptions is as valuable as analysis of the results themselves, highlighting the weak parts of the statistical formulation, and pointing to the need of closer communication between modelers, climatologists and statisticians in order to circumscribe the range of sensible assumptions. In this regard the Bayesian model presented here may be viewed as a device to foster more effective collaborations.

A natural extension to the current region-specific model is a model that introduces AOGCM-specific correlation between regions and thus borrows strength from all the regional temperature signals in order to estimate model variability and biases. As we discussed in Section 1, working with higher resolution AOGCM output, both spatially and temporally, will require a much more complex effort. Richer information on AOGCM performance as linked to specific regional/seasonal climate reproductions will have to be incorporated in the likelihood of the AOGCM's responses and/or the priors on the precision parameters. In addition, other climate responses may be considered, jointly with temperature. Precipitation is of course the obvious candidate for such an extension, and a bivariate distribution of temperature and precipitation is a natural extension.

4.4 More collaboration

We think that the results in their distributional form uncover some aspects of climate modeling easily overlooked using descriptive statistics. The differential behavior of single AOGCMs when comparing regions, seasons and alternative convergence criteria may be highlighted by the form of the posterior distributions, while difficult to infer from simpler analyses.

It should be noted that we used the criteria established by Giorgi and Mearns (2002, 2003) without reconsidering their validity. They used these criteria as ones that are commonly discussed in the climate change literature as relevant to evaluating climate change projections. One can call into question whether these two criteria are the most relevant for evaluating the reliability of projections of climate change. The ability of a model to reproduce the current climate is usually regarded as a necessary, but insufficient reason for considering the model's response to future forcings as reliable (McAveny *et al.*,2001). The convergence criterion can be questioned because models can produce similar responses due to similarities in model structures, not because the models are converging on the 'true' response. Yet, agreement among many models has been viewed as strengthening the likelihood of climate change (Cubasch *et al.*, 2001; Giorgi *et al.*, 2001a,b). As we noted in Section 2.2 both the form of the REA weights (with the use of two different exponents for the bias and convergence terms) and the specification of the likelihood in our statistical model allow the final result to depend to a lesser degree on the convergence criterion. Besides changing the likelihood assumptions one could impose a prior distribution for the θ parameter that concentrates most of its mass on values less than one. In this way, one posits a lower level of confidence in the precision of the future simulation than of the present, implicitly accepting a less stringent criterion of convergence for the future trajectories. The form of (12) shows that a value of θ less than one would downweight large deviations in the convergence term $(Y_i - \nu)^2$, thus achieving a similar effect to the exponent in the REA weight.

From a more general perspective, however, our goal was to extend the REA method, not to reevaluate its criteria. It is expected that other criteria, and more complex combinations of criteria will evolve over time as this statistical work is refined. Much work is being devoted to more sophisticated and extensive methods of model performance evaluation and inter-model comparison (Hegerl *et al.* 2000, Meehl *et al.* 2000) and future analyses may modify or extend the two criteria of bias and convergence to account for them.

In conclusion we view our results as an illustration of the power of bringing statistical modeling to experiments where quantifying uncertainty is an intrinsic concern. Moreover we hope this work may foster more deliberate analysis that incorporates scientific knowledge of climate modeling at global and regional scales.

Acknowledgements

This research was supported through the National Center for Atmospheric Research Inititiative on Weather and Climate Impact Assessment Science, which is funded by the National Science Foundation. Additional support was provided through NSF grants DMS-0084375 and DMS-9815344. We wish to thank Filippo Giorgi for making very helpful comments from his perspective as a climate modeler on the short comings of a purely statistical presentation. Also, we thank Art Dempster for his cogent remarks on an early draft.

Appendix 1: Example of a Bayesian statistical model, with closed-form solution to the prior-posterior update

Consider the problem of estimating a probability distribution function of future mean temperature, ν, in a given region, on the basis of 9 climate models' output $(Y_1, Y_2, \ldots, Y_9)$. It is the goal of Bayesian analysis to derive a *posterior* distribution for ν that combines any *prior* information available, for example about regional climate, and the reliability of the 9 AOGCMs, with the *likelihood* of the data $(Y_1, Y_2, \ldots, Y_9)$.

Formally, the posterior is $P(\nu|Y_1, Y_2, \ldots, Y_9)$, the probability distribution of ν *conditional* on the data. Everything to the right of the conditioning bar, $|$, is fixed, observed. Bayes' theorem states that given two events A and B, the conditional probability of the first, given the second is

$$P(A|B) = \frac{P(A,B)}{P(B)} = \frac{P(B|A) \cdot P(A)}{P(B)}.$$

Typically the problem is structured so that it is conceptually easy to model $P(B|A)$ (usually the *likelihood* of the data), and $P(A)$ (the *prior*). Then Bayes' theorem recovers $P(A|B)$. The final result is a reshaping of the prior beliefs through the information contained in the likelihood of the data. The specific case of computing $P(\nu|Y_1, Y_2, \ldots, Y_9)$ is now described, in a simple model framework where the posterior can be computed in closed form.

- $P(B|A) \equiv P(\{Y_1, \ldots, Y_9\}|\nu)$ is the *likelihood* for the data, "conditioning on the true temperature". We assume that if ν is known then each climate model response is normally distributed about this true value but with a possibly different precision (the inverse of the variance) λ_i. We also make the assumption that model biases are statistically independent: knowledge of the bias for one model gives no information about the bias in another model. In notation:

$$\begin{aligned} P(\{Y_1, \ldots, Y_9\}|\nu) &= \prod_{i=1}^{9} P(Y_i|\nu) \\ P(Y_1|\nu) &\equiv \mathcal{N}(\nu, \lambda_1^{-1}) \\ &\cdots \\ P(Y_9|\nu) &\equiv \mathcal{N}(\nu, \lambda_9^{-1}) \end{aligned}$$

  where $\mathcal{N}(\nu, \lambda^{-1})$ denotes a Gaussian distribution with mean ν and variance λ^{-1}.

- $P(A) \equiv P(\nu)$ is the *prior* and summarizes all the knowledge available about the true temperature in the region, separate from what the data tell us. For example, a climatologist may indicate a sensible range of temperatures, expected to contain the mass of the probability distribution, say up to 10^{-3} residual. The climatologist must do so without considering $Y_1, \ldots, Y_9$, because that would imply using the data twice in the analysis. Here we assume a Gaussian prior

$$P(\nu) = \mathcal{N}(\nu_{\text{prior}}, \lambda_{\text{prior}})$$

  where ν_{prior} might be chosen based on current climate and λ_{prior} could be very small indicating very little prior information about ν for future climate.

- $P(B) \equiv P(\{Y_1, \ldots, Y_9\})$ is not a function of ν, and its role is as a normalization constant for the posterior to integrate to one.

- $P(A|B) \equiv P(\nu|\{Y_1, \ldots, Y_9\})$. Determining the posterior density for many practical problems involves complicated algebra, numerical integration or Monte Carlo methods. We choose this example because it admits an analytical solution for $P(\nu|\{Y_1, \ldots, Y_9\})$, whose form has interpretable properties. We give the derivation below, from which one finds that the posterior is a Gaussian distribution with mean and variance that depend on the data $\{Y_1, \ldots, Y_9\}$ and the prior parameters $\{\lambda_1, \ldots, \lambda_9, \nu_{\text{prior}}, \lambda_{\text{prior}}\}$.

Multiplying likelihood and prior at the numerator of Bayes' formula one obtains

$$P(\nu|\{Y_1, \ldots, Y_9\}) = \frac{1}{C} \prod_{i=1}^{9} \frac{\lambda_i^{1/2}}{(2\pi)^{1/2}} \exp\{\frac{\lambda_i}{2}(Y_i - \nu)^2\} \cdot \frac{\lambda_{\text{prior}}^{1/2}}{(2\pi)^{1/2}} \exp\{\frac{\lambda_{\text{prior}}}{2}(\nu - \nu_{\text{prior}})^2\} \quad (16)$$

where C is a normalizing constant. By a series of algebraic manipulations the expression becomes

$$P(\nu|\{Y_1,\ldots,Y_9\}) = \frac{\sqrt{\Lambda}}{\sqrt{2\pi}} \exp\left\{\frac{\Lambda(\nu - (\sum_{i=1}^9 \lambda_i Y_i + \lambda_{\text{prior}}\nu_{\text{prior}})/\Lambda)^2}{2}\right\} \tag{17}$$

where $\Lambda = \sum_{i=1}^9 \lambda_i + \lambda_{\text{prior}}$. This is the form of a Gaussian distribution, with mean $(\sum_{i=1}^9 \lambda_i Y_i + \lambda_{\text{prior}}\nu_{\text{prior}})/\Lambda$ and variance $1/\Lambda$.

Thus, we have derived that under the prior and likelihood assumptions listed above the posterior distribution of ν is a Gaussian, with mean given by a weighted average of the prior mean and the data points. Intuitively, Bayes' procedure weighs the data points by a measure of how tightly they are distributed around ν (measured by the precisions of the likelihood functions, λ_i), and weighs the prior guess ν_{prior} on the basis of the degree of accuracy we attribute to our prior belief (measured by the precision of the prior distribution, λ_{prior}). If the prior variance becomes very large, then λ_{prior} is correspondingly small and the posterior mean is close to the weighted average of the climate model results with little modification by the prior mean. Finally, we note that the weighted average converges to the maximum likelihood estimate for ν, as $\lambda_{\text{prior}} \to 0$, that is for a prior that becomes more and more uncertain, i.e. uninformative.

Appendix 2: Prior-posterior update through Markov chain Monte Carlo simulation

The joint posterior distributions derived from the models in Section 2 are not members of any known parametric family. However, the distributional forms (Gaussian, Uniform and Gamma) chosen for the likelihoods and priors are conjugate, thus allowing for closed-form derivation of all full conditional distributions (the distributions of each parameter, as a function of the remaining parameters assuming fixed deterministic values). We list here such distributions, for the robust model that includes correlation between X_i and Y_i in the form of regression equation (14). The variables s_i, t_i $i = 1,\ldots 9$ are introduced here as an auxiliary randomization device, in order to efficiently simulate from Student-t distributions within the Gibbs sampler. They are not essential parts of the statistical model. Fixing $s_i = t_i = 1$, $\beta_x = 0$ allows the recovery of the full conditionals for $\lambda_1,\ldots,\lambda_9$, μ, ν and θ of the basic univariate model as a special case.

$$\lambda_i|\ldots \sim \text{Ga}\left(a+1, b + \frac{s_i}{2}(X_i-\mu)^2 + \frac{\theta t_i}{2}\{Y_i - \nu - \beta_x(X_i-\mu)\}^2\right), \tag{18}$$

$$s_i|\ldots \sim \text{Ga}\left(\frac{\phi+1}{2}, \frac{\phi + \lambda_i(X_i-\mu)^2}{2}\right), \tag{19}$$

$$t_i|\ldots \sim \text{Ga}\left(\frac{\phi+1}{2}, \frac{\phi + \theta\lambda_i\{Y_i - \nu - \beta_x(X_i-\mu)\}^2}{2}\right), \tag{20}$$

$$\mu|\ldots \sim N\left(\tilde{\mu}, \left(\sum s_i\lambda_i + \theta\beta_x^2\sum t_i\lambda_i + \lambda_0\right)^{-1}\right), \tag{21}$$

$$\nu|\ldots \sim N\left(\tilde{\nu}, \left(\theta\sum t_i\lambda_i\right)^{-1}\right), \tag{22}$$

$$\beta_x|\ldots \sim N\left(\tilde{\beta}_x, \left(\theta\sum t_i\lambda_i(X_i-\mu)^2\right)^{-1}\right), \tag{23}$$

$$\theta|\ldots \sim \mathrm{Ga}\left(c+\frac{N}{2}, d+\frac{1}{2}\sum t_i\lambda_i\{Y_i-\nu-\beta_x(X_i-\mu)\}^2\right). \tag{24}$$

Above, we have used the following shorthand notation:

$$\tilde{\mu} = \frac{\sum s_i\lambda_i X_i - \theta\beta_x \sum \lambda_i t_i(Y_i-\nu-\beta_x X_i)+\lambda_0 X_0}{\sum s_i\lambda_i + \theta\beta_x^2\sum\lambda_i t_i+\lambda_0}, \tag{25}$$

$$\tilde{\nu} = \frac{\sum t_i\lambda_i\{Y_i-\beta_x(X_i-\mu)}{\sum t_i\lambda_i}, \tag{26}$$

$$\tilde{\beta_x} = \frac{\sum t_i\lambda_i(Y_i-\nu)(X_i-\mu)}{\sum t_i\lambda_i(X_i-\mu)^2}. \tag{27}$$

The Gibbs sampler can be easily coded so as to simulate iteratively from this sequence of full conditional distributions. After a series of random drawings during which the Markov Chain process forgets about the arbitrary set of initial values for the parameters (the burn-in period), the values sampled at each iteration represent a draw from the joint posterior distribution of interest, and any summary statistic can be computed to a degree of approximation that is direct function of the number of sampled values available, and inverse function of the correlation between successive samples. In order to minimize the latter, we save only one iteration result every 50, after running the sampler for a total of $500,000$ iterations, and discarding the first half as a burn-in period. These many iterations are probably not needed for this particular application but by providing them we are eliminating any possibility of bias resulting from too few MCMC iterations. The convergence of the Markov chain to its stationary distribution (the joint posterior of interest) is verified by standard diagnostic tools (Best *et al.* 1995). A self contained version of the MCMC algorithm, implemented in the free software package R (R development core team, 2004), is available at `www.cgd.ucar.edu/~nychka/REA`.

References

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

Allen, M., Raper, S. and Mitchell, J. (2001), uncertainty in the IPCC's Third Assessment Report. *Science* **293**, Policy Forum, pp. 430–433.

Best N.G., M. K. Cowles, and S. K. Vines, 1995: *CODA Convergence Diagnosis and Output Analysis software for Gibbs Sampler output: Version 0.3.*, available from `http://www.mrc-bsu.cam.ac.uk/bugs/classic/coda04/readme.shtml`).

Bernardo, J. M., and A. F. Smith, 1994: *Bayesian Theory.*, John Wiley & Sons.

Cubasch, U., G.A. Meehl, G.J. Boer, R.J.Stouffer, M. Dix, A. Noda, C.A. Senior, S. Raper, and K.S. Yap, 2001: Projections of future climate change. *Climate Change 2001: The Scientific Basis*, J. T. Houghton *et al.*, Eds., Cambridge University Press, 525-582.

Dai, A., T. M. L. Wigley, B. Boville, J. T. Kiehl, and L. Buja, 2001: Climates of the twentieth and twenty-first centuries simulated by the NCAR climate system model. *J. Climate,* **14**, 485-519.

Dessai, S. and M. Hulme (2003) Does climate policy need probabilities? *Tyndall Centre Working Paper*, **34**. Available from `http://www.tyndall.ac.uk/publications/publications.shtml`

Emori, S., T. Nozawa, A. Abe-Ouchi, A. Numaguti, M. Kimoto, and T. Nakajima (1999): Coupled ocean-atmosphere model experiments of future climate change with an explicit representation of sulfate aerosol scattering. *J. Meteorological Society of Japan*, **77**, 1299-1307.

Flato, G.M. and G. J. Boer, 2001: Warming asymmetry in climate change simulations. *Geophysical Research Letters*, **28**, 195-198.

Forest, C.E., Stone, P.H., Sokolov, A.P., Allen, M.R. and M.D. Webster, 2002: Quantifying Uncertainties in Climate System Properties with the Use of Recent Climate Observations. *Science.*, **295**, 113-117.

Gelman, A., Carlin, J.B., Stern, H.S. and D. Rubin, 1994: *Bayesian Data Analysis.* Chapman & Hall.

Gilks, W.R., S. Richardson, and D. J. Spiegelhalter, 1996: *Markov Chain Monte Carlo Methods in Practice.*,Chapman & Hall.

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

Giorgi F., and L.O. Mearns, 2002: Calculation of average, uncertainty range and reliability of regional climate changes from AOGCM simulations via the "reliability ensemble averaging" (REA) method. *J. Climate,* **15**, n. 10, 1141-1158.

Giorgi F., and L.O. Mearns, 2003: Probability of Regional Climate Change Calculated using the Reliability Ensemble Averaging (REA) Method. *Geophysical Research Letters*, **30**, n.12, 311-314.

Giorgi F., and co-authors, 2001a: Regional climate information: Evaluation and projections. In J. T. Houghton *et al.* (eds.) *Climate Change 2001: The Scientific Basis.* Contribution of Working Group I to the Third Assessment Report of the Intergovenmental Panel on Climate Change. Chapter 10. Cambridge: Cambridge University Press. pp. 583–638.

Giorgi, F., and co-authors, 2001b: Emerging patterns of simulated regional climatic changes for the 21st century due to anthropogenic forcings. *Geophys. Res. Lett.* **28**, 3317-3321.

Gordon, C., C. Cooper, C.A. Senior, H. T. Banks, J. M. Gregory, T. C. Johns, J.F. B. Mitchell and R.A. Wood, 2000: The simulation of SST, sea ice extents and ocean heat transport in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dynamics*, **16**, 147-168.

Gordon, H. B., and S. P. O'Farrell, 1997: Transient climate change in the CSIRO coupled model with dynamic sea ice. *Mon. Wea. Rev.,* **125**, 875-907.

Hegerl, G.C., P.A. Stott, M.R. Allen, J.F.B. Mitchell, S.F.B. Tett and U. Cubasch, 2000: Optimal detection and attribution of climate change: sensitivity of results to climate model differences. *Climate Dynamics,* **16**, 737-754.

Knutson, T. R., T. L. Delworth, K. W. Dixon, and R. J. Stouffer, 1999: Model assessment of regional surface temperature trends (1949-97). *J. Geophys. Res.,* **104**, 30981-30996.

McAvaney, B.J, C.Covey, S. Joussaume, V. Kattsov, A. Kitoh, W. Ogana, A.J. Pittman, A.J. Weaver, R.A. Wood, and Z.-C. Zhao, 2001: Model evaluation. *Climate Change 2001: The Scientific Basis*, J. T. Houghton *et al.*, Eds., Cambridge University Press, 471-524.

Mearns, L.O., M. Hulme, T. R. Carter, R. Leemans, M.Lal, and P. Whetton, 2001: Climate scenario development. *Climate Change 2001: The Scientific Basis*, J. T. Houghton *et al.*, Eds., Cambridge University Press, 739-768.

Meehl, G.A., G.J. Boer, C. Covey, M. Latif, and R.J.Stouffer, 2000: The Coupled Model Intercomparison Project (CMIP). *Bull. Am. Met. Soc.*, **81**, 313-318.

Nakicenovic, N., J. Alcamo, G. Davis, B. de Vries, J. Fenhann, S. Gaffin, K. Gregory, A. Grbler, T. Y. Jung, T. Kram, E. L. La Rovere, L. Michaelis, S. Mori, T. Morita, W. Pepper, H. Pitcher, L. Price, K. Raihi, A. Roehrl, H-H. Rogner, A. Sankovski, M. Schlesinger, P. Shukla, S. Smith, R. Swart, S. van Rooijen, N. Victor, Z. Dadi, 2000: IPCC Special Report on Emissions Scenarios, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 599 pp.

Noda, A., K. Yoshimatsu, S. Yukimoto, K. Yamaguchi, and S. Yamaki, 1999: Relationship between natural variability and CO_2-induced warming pattern: MRI AOGCM experiment. Preprints, 10^{th} *Symp. on Global Change Studies,* Dallas, TX, Amer. Meteor. Soc., **126**, 2013-2033.

Nychka, D., and C. Tebaldi, 2003: Comment on 'Calculation of average, uncertainty range and reliability of regional climate changes from AOGCM simulations via the "reliability ensemble averaging" (REA) method'. *J. Climate,* **16**, 883-884.

R Development Core Team, 2004, *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3, `http://www.R-project.org`.

Raisanen, J., 1997: Objective comparison of patterns of CO_2 induced climate change in coupled GCM experiments. *Climate Dynamics*, **13**, 197-211.

Raisanen, J and T. N. Palmer, 2001: A Probability and Decision Model Analysis of a Multi-model Ensemble of Climate Change Simulations. *J. Climate,* **14**, 3212-3226.

Reilly, J., P. H. Stone, C. E. Forest, M. D. Webster, H. D. Jacoby, R. G. Prinn, 2001, Uncertainty in climate change assessments. *Science*, **293**, 5529,430-433.

Schneider, S. H., 2001: What is dangerous climate change?. *Nature*, **411**, 17-19.

Stendel, M., T. Schmidt, E. Roeckner, and U. Cubasch, 2000: The climate of the 21st century: Transient simulations with a coupled atmosphere-ocean general circulation model. Danmarks Klimacenter Rep. 00-6.

Washington, W. M., J. W. Weatherly, G.A. Meehl, A.J. Semtner, Jr., T. W. Bettge, A.P. Craig, W. G. Strand, Jr., J.M. Arblaster, V.B. Wayland, R. James and Y. Zhang, 2000: Parallel Climate Model (PCM) control and transient simulations. *Climate Dynamics*, **16**, 755-774.

Webster, M., 2003: Communicating climate change uncertainty to policy-makers and the public. *Climatic Change* **61**, 1-8.

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

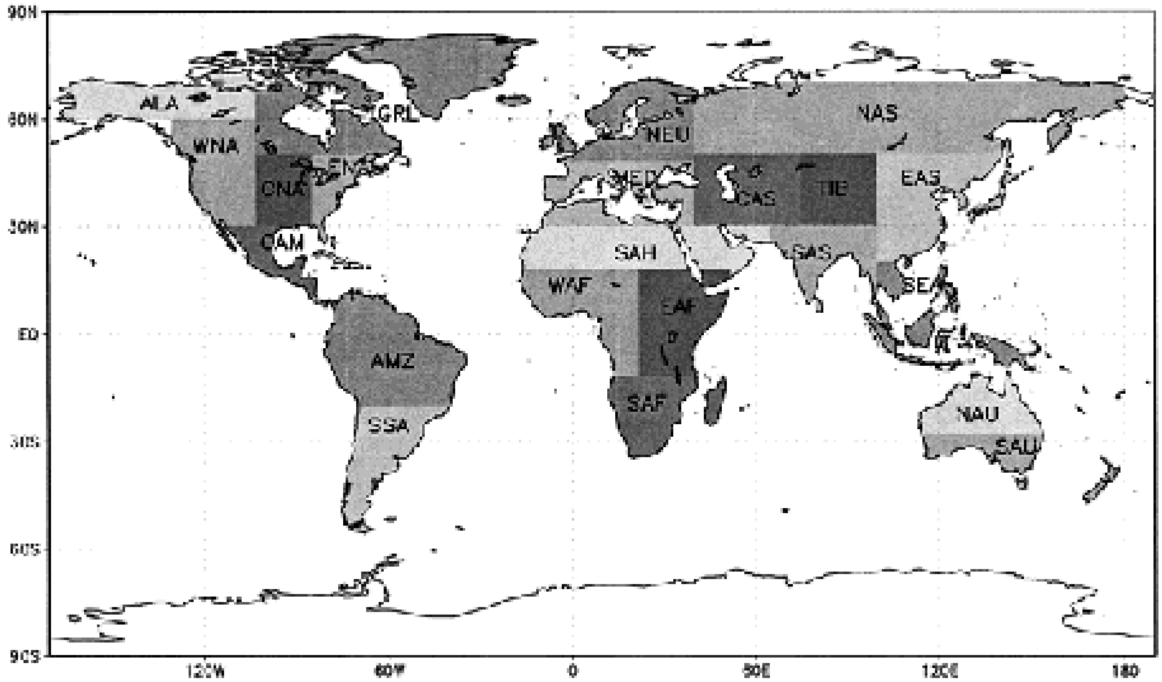

Figure 1: The 22 regions into which the land masses were discretized for both the REA analysis and our analysis.

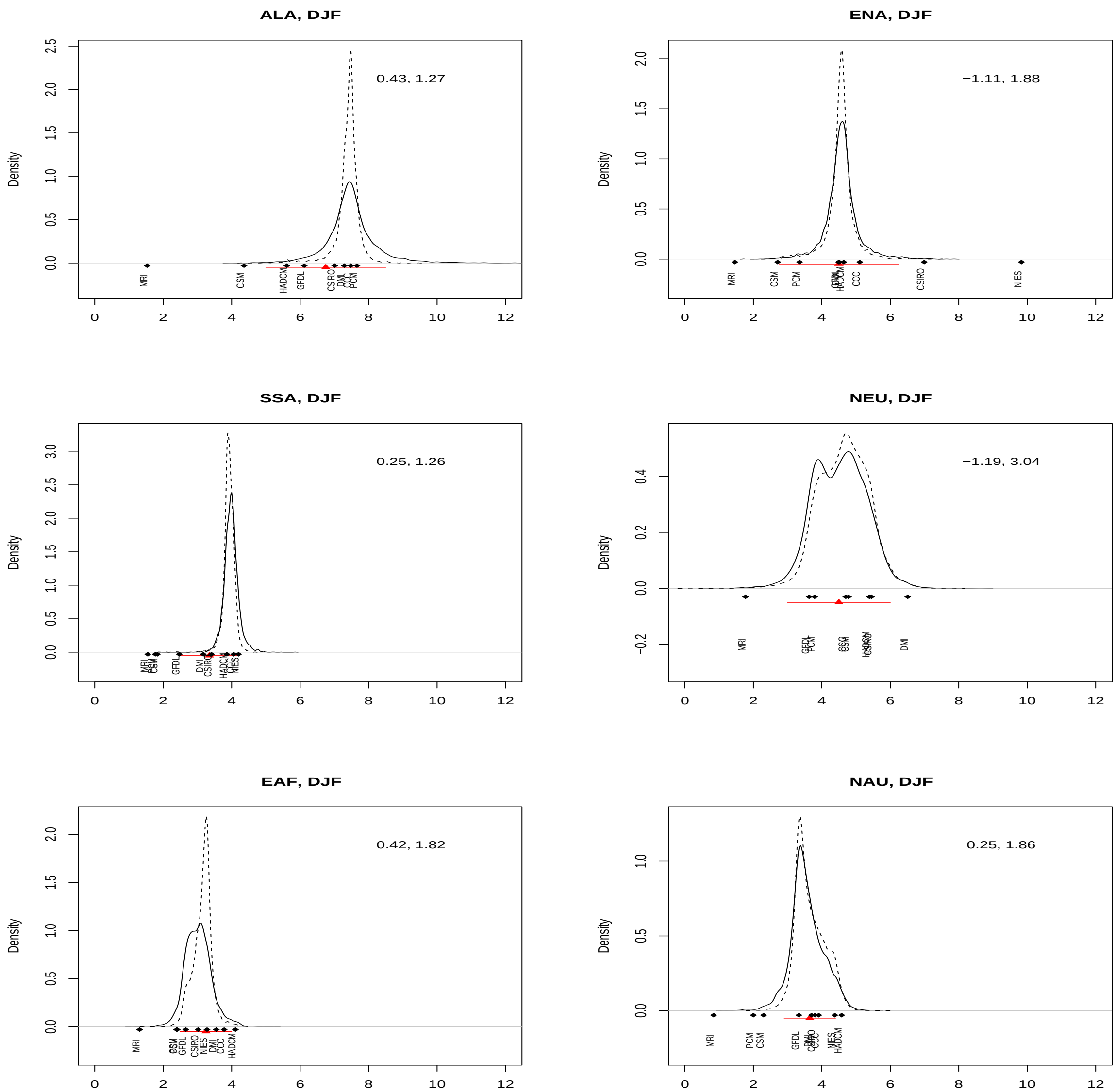

Figure 2: Posterior distributions of $\Delta T \equiv \nu - \mu$ for six chosen regions, for winter (DJF) season. Solid lines: basic model. Dashed lines: model with correlation between X_i and Y_i. The points along the base of the densities mark the 9 AOGCMs temperature change predictions. The triangle and segment indicate the REA estimate of mean change plus/minus a measure of natural variability. The two numbers in the upper right corner of each plot are the limits of the 95% posterior probability region for β_x.

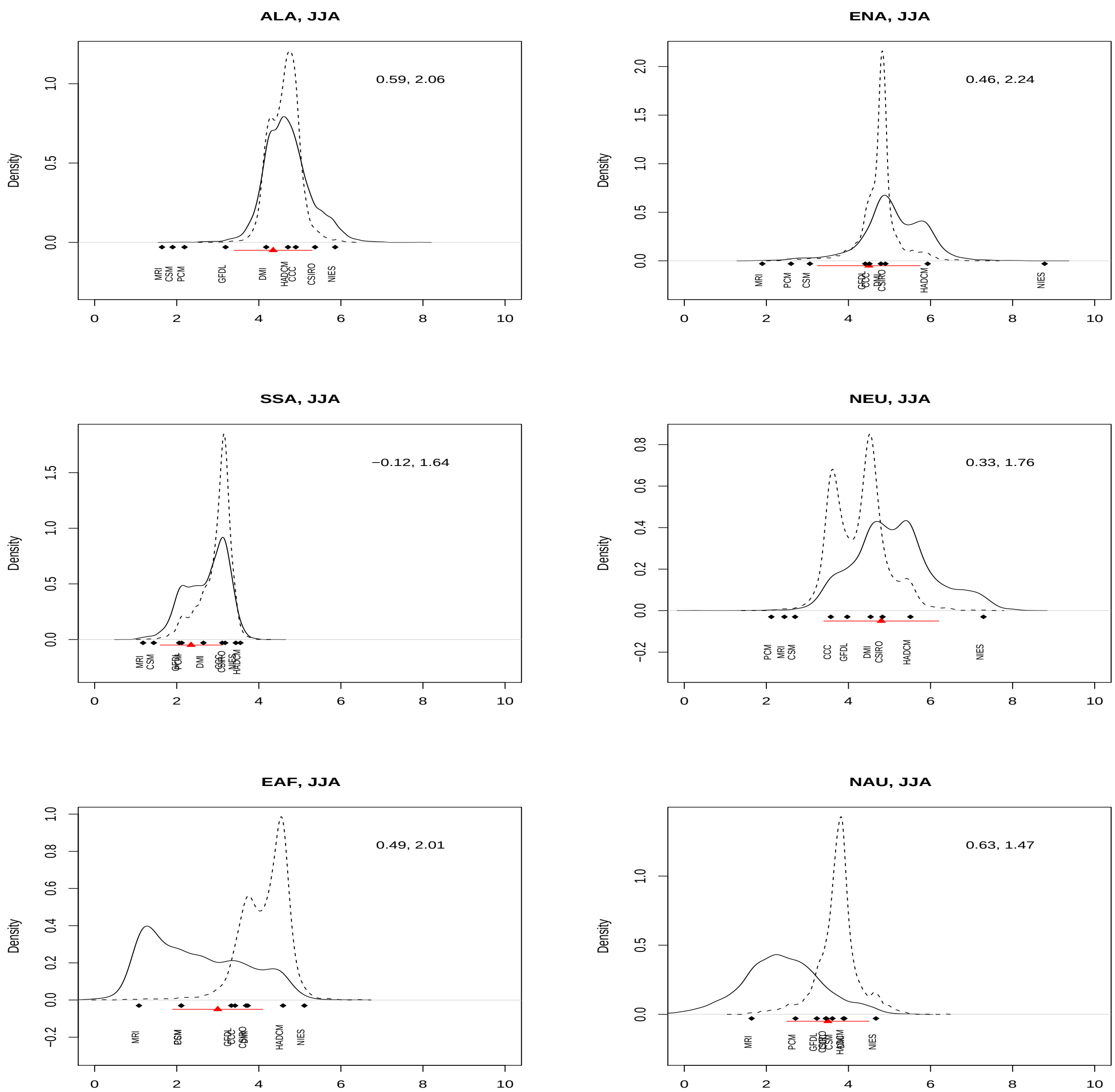

Figure 3: Posterior distributions of $\Delta T \equiv \nu - \mu$ for six chosen regions, for summer (JJA) season. All details as in Figure 2. In particular, solid lines correspond to posterior densities obtained from the basic model, dashed lines from the model with correlation between X_i and Y_i.

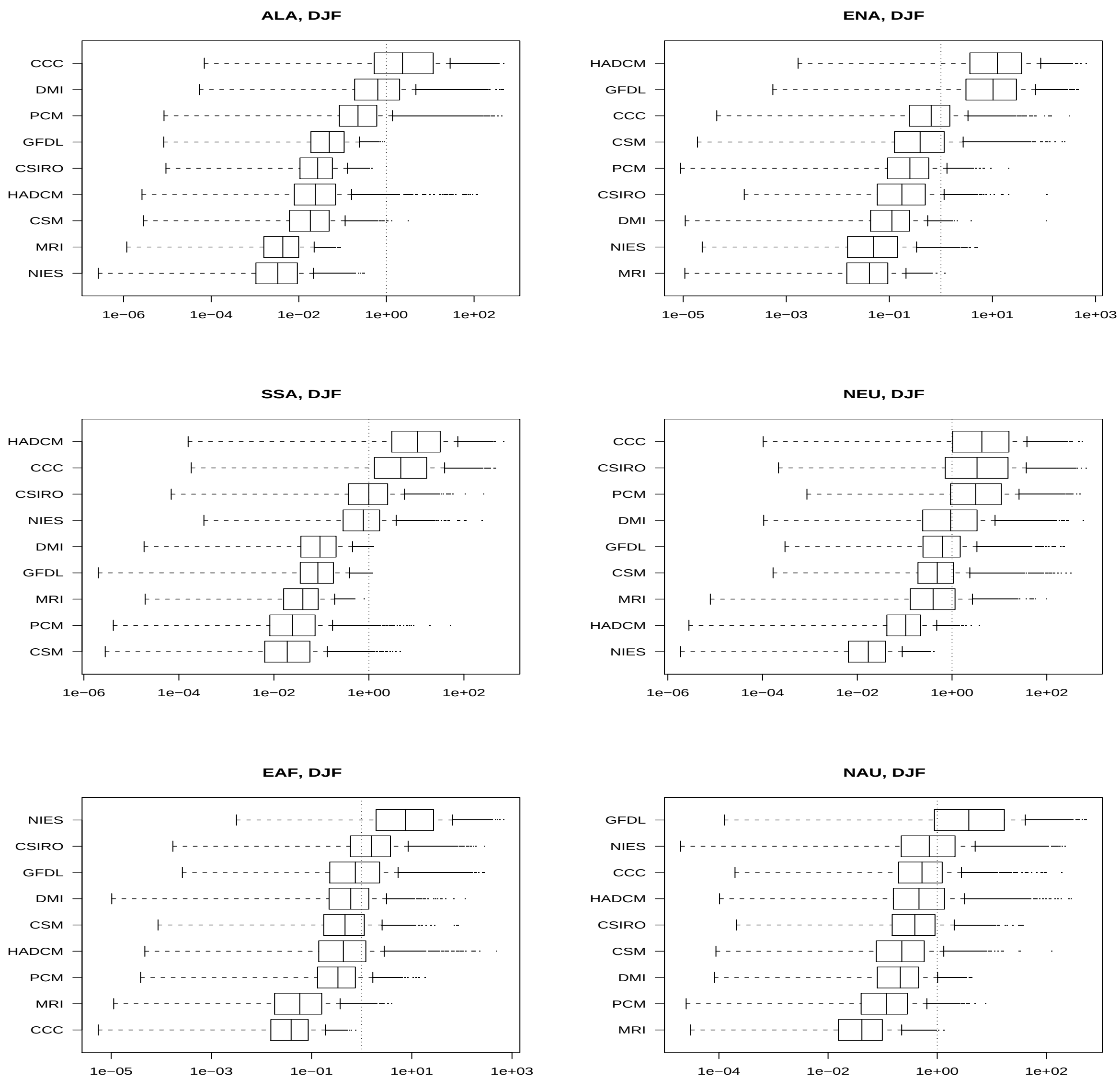

Figure 4: Posterior distributions of λ_i, the model-specific precision parameter, for six chosen regions, for winter season.

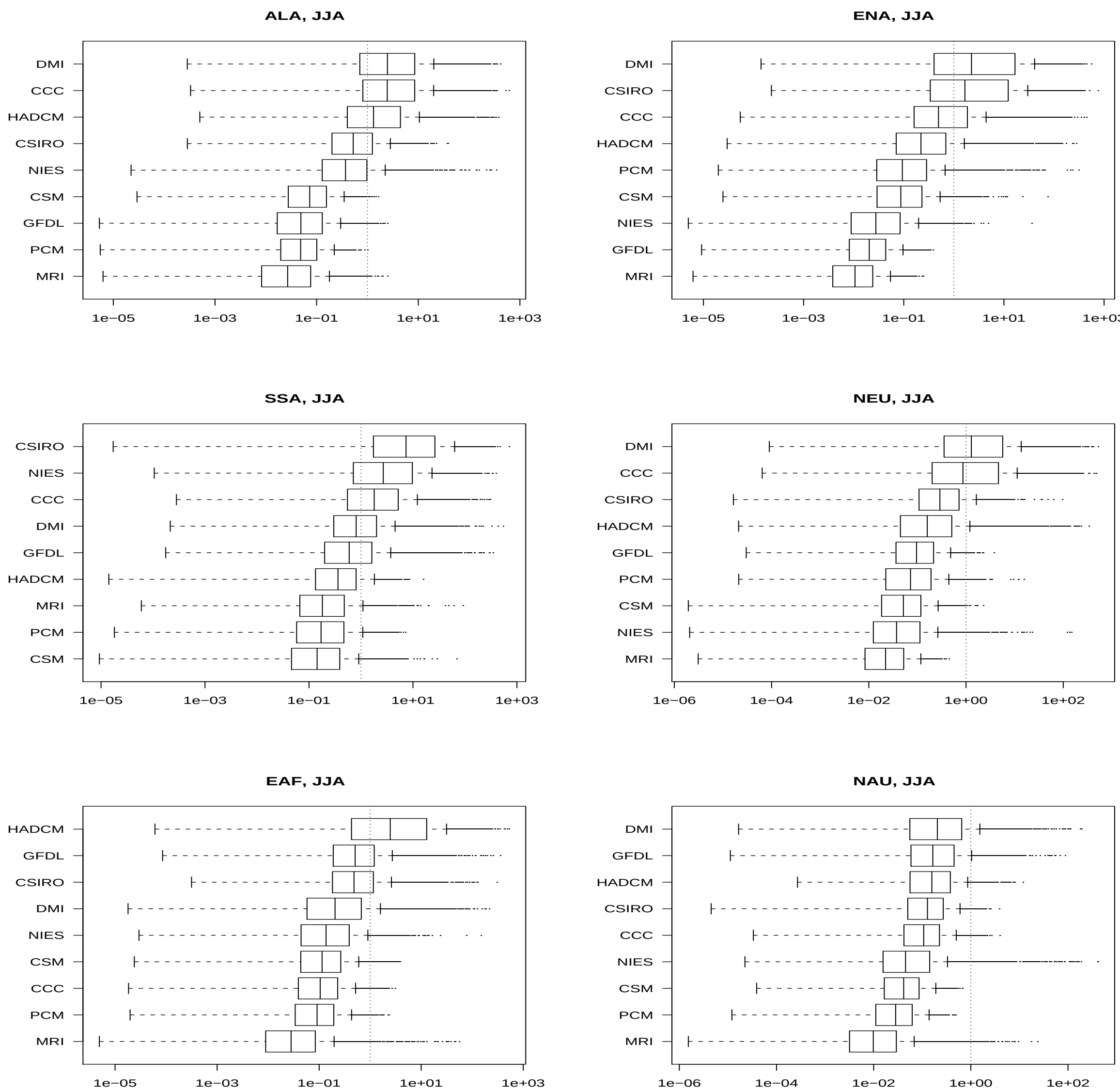

Figure 5: Posterior distributions of λ_i, the model-specific precision parameter, for six chosen regions, for summer season.

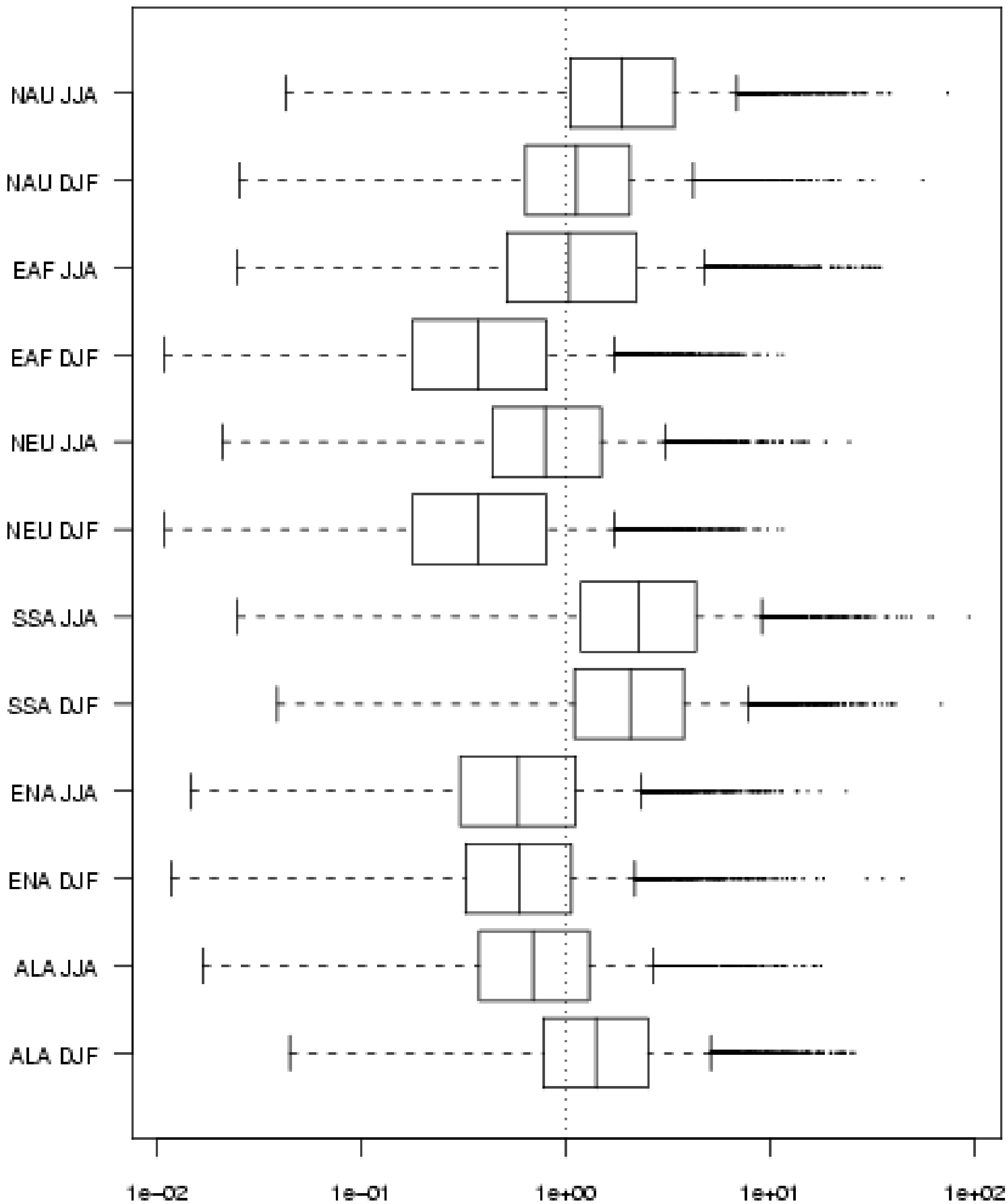

Figure 6: Posterior distributions of θ, the inflation/deflation factor for the precision parameters when simulating future climate, common to all 9 AOGCMs, for six chosen regions, for winter and summer season.

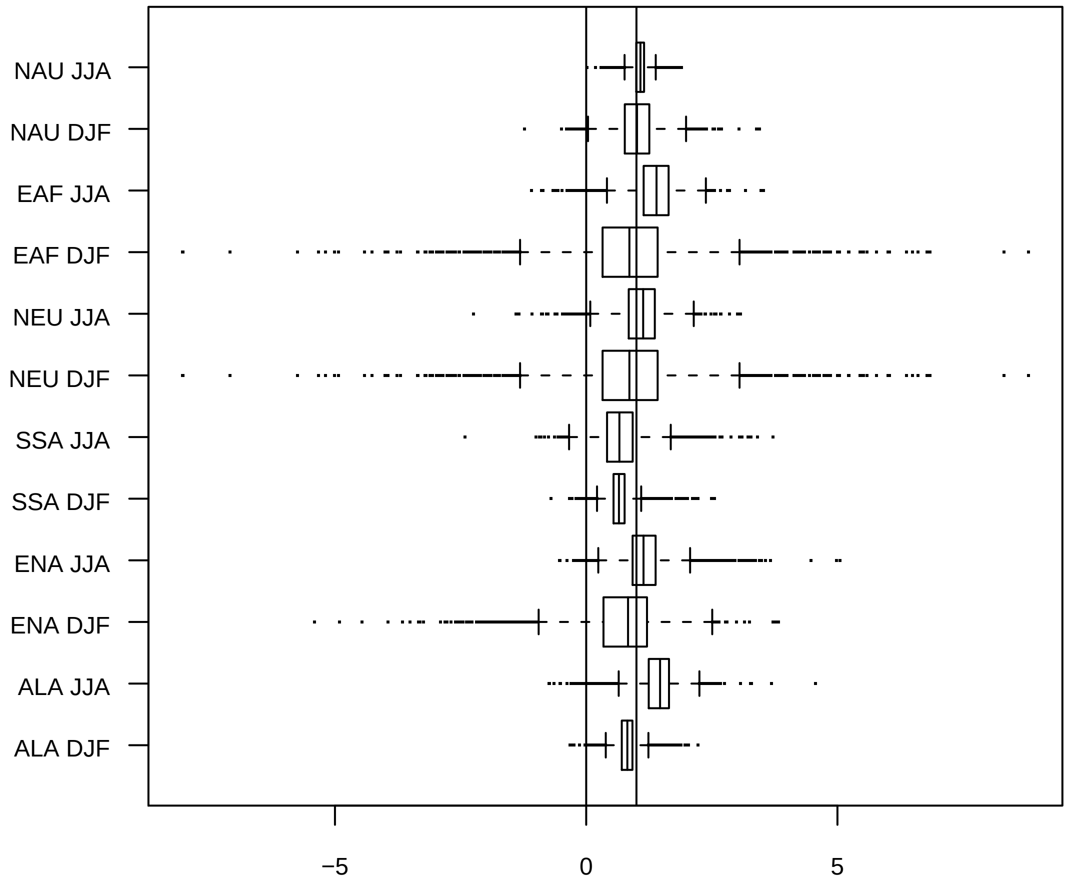

Figure 7: Posterior distributions of β_x, the regression parameter that introduces correlation between X_i and Y_i, common to all 9 AOGCMs, for six chosen regions, for winter and summer season. For reference, we draw a vertical line at zero – useful to assess the significance of the parameter magnitude – and a vertical line at one – to assess the consistence of our results to the REA analysis' assumption of independence between $Y_i - X_i$ and $X_i - \mu$.

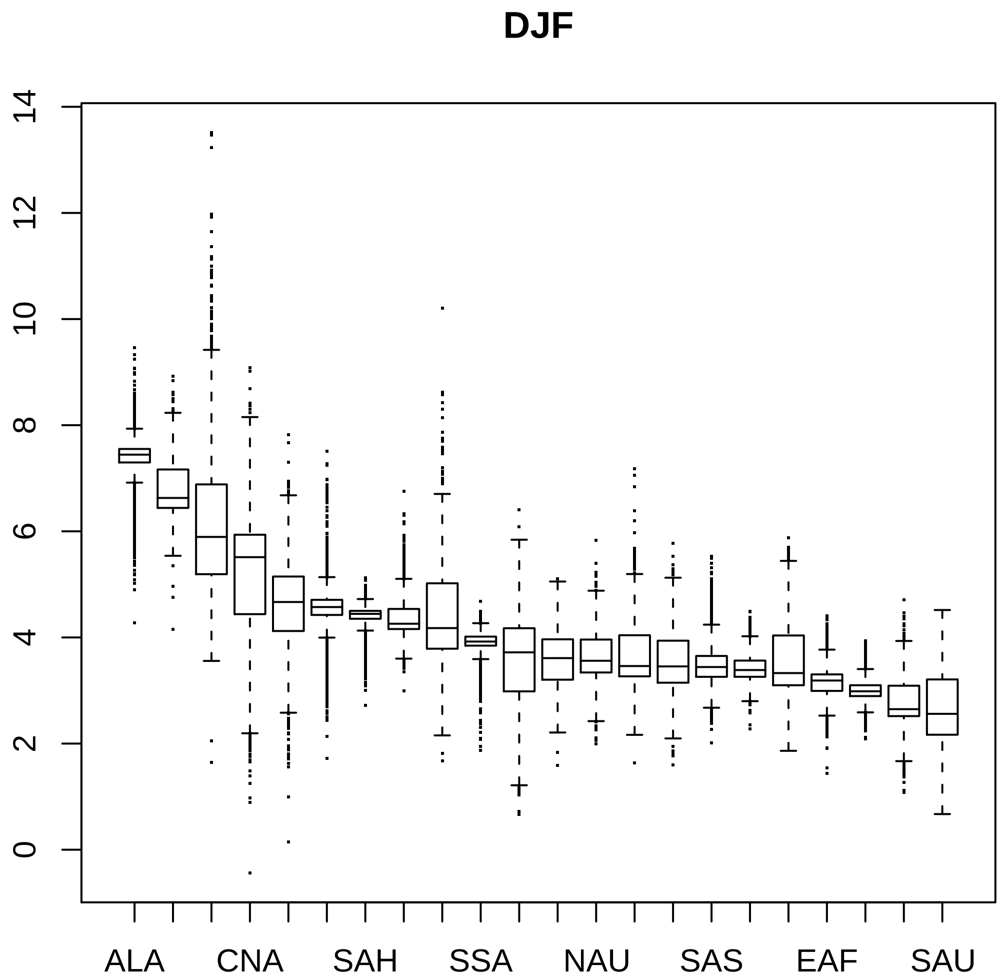

Figure 8: Posterior distributions of $\Delta T \equiv \nu - \mu$ for all 22 regions, for winter season.

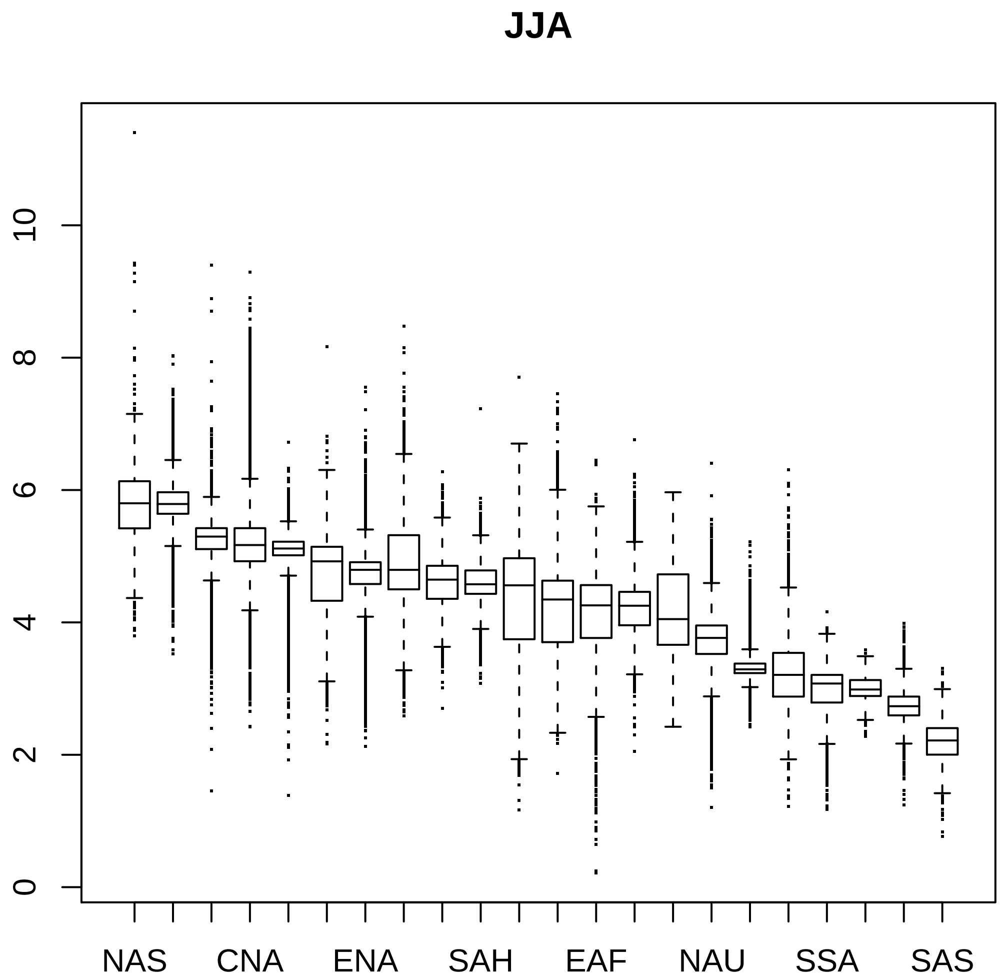

Figure 9: Posterior distributions of $\Delta T \equiv \nu - \mu$ for all 22 regions for summer season.

Table 1: The 9 Atmosphere-Ocean General Circulation Models whose output constitutes the data in our analysis.

AOGCM	reference	climate sensitivity
CCSR-NIES/version 2	Emori *et al.* (1999)	4.53
MRI/version 2	Noda *et al.* (1999)	1.25
CCC/GCM2	Flato and Boer (2001)	3.59
CSIRO/Mk2	Gordon and O'Farrell (1997)	3.50
NCAR/CSM	Dai *et al.* (2001)	2.29
DOE-NCAR/PCM	Washington *et al.* (2000)	2.35
GFDL/R30-c	Knutson *et al.* (1999)	2.87
MPI-DMI/ECHAM4-OPYC	Stendel *et al.* (2000)	3.11
UKMO/HADCM3	Gordon *et al.* (2000)	3.38

Table 2: Natural variability (degrees C) of observed temperature. These are taken from Giorgi and Mearns (2002). They were estimated by computing 30-year moving averages of observed, detrended, regional mean temperatures over the 20$^{\text{th}}$ century and taking the difference between maximum and minimum values.

region	DJF	JJA
NAU	0.40	0.50
SAU	0.50	0.25
AMZ	0.30	0.25
SSA	0.35	0.75
CAM	0.65	0.35
WNA	0.75	0.50
CNA	1.40	0.92
ENA	1.30	0.80
ALA	1.75	0.80
GRL	0.85	0.50
MED	0.40	0.60
NEU	1.00	0.90
WAF	0.45	0.50
EAF	0.42	0.50
SAF	0.40	0.50
SAH	0.45	0.75
SEA	0.40	0.25
EAS	0.85	0.75
SAS	0.40	0.25
CAS	1.00	0.75
TIB	0.85	0.75
NAS	1.00	0.70

Table 3: Model bias, for the 9 AOGCMs temperature response in the 6 regions in winter (DJF) and summer (JJA). Biases are computed as the deviation of the single AOGCM's response, X_i, from the mean of the posterior distribution for μ derived by our analysis.

	CCC	CSIRO	CSM	DMI	GFDL	MRI	NIES	PCM	HADCM
ALA, DJF	−0.23	5.83	2.95	0.98	4.29	11.88	2.15	−2.40	−2.62
ALA, JJA	0.50	1.37	−3.57	−0.32	2.26	1.82	0.57	−4.83	−0.64
ENA, DJF	1.15	−1.37	0.18	3.03	−0.17	4.44	−1.32	−1.56	0.10
ENA, JJA	−1.03	1.21	−2.48	1.04	7.26	8.31	1.16	−1.01	−0.62
SSA, DJF	−0.00	0.69	−1.82	3.16	3.46	4.85	−1.20	−0.97	0.09
SSA, JJA	−1.03	−0.48	−0.98	0.72	0.74	1.28	−0.39	−1.74	−2.01
NEU, DJF	0.47	0.34	−1.43	0.30	−1.04	0.57	−6.30	0.03	−3.25
NEU, JJA	0.10	2.18	−3.65	1.05	3.52	5.96	−0.63	−2.38	−0.83
EAF, DJF	−4.76	−1.06	−1.39	0.91	0.30	1.70	−0.28	−1.90	0.10
EAF, JJA	−3.21	−1.58	−2.49	1.12	−1.48	0.57	−1.29	−3.49	−0.05
NAU, DJF	−1.31	−1.58	−1.16	2.17	0.23	3.32	0.60	−2.05	−0.71
NAU, JJA	−3.31	−3.04	−5.15	1.62	−1.70	1.57	−0.95	−5.58	−2.15

Table 4: Relative weighting of the 9 AOGCMs across six regions chosen as examples. The values are computed as $100 \times (\lambda_i^* / \sum_{i=1}^{9} \lambda_i^*)$, where the 9 λ_i^*'s are the means of the posterior distributions derived by MCMC simulation. Winter temperature change analysis.

	CCC	CSIRO	CSM	DMI	GFDL	MRI	NIES	PCM	HADCM
ALA:	63.80	0.19	0.19	21.71	0.35	0.03	0.04	12.28	1.40
ENA:	2.84	0.74	2.99	0.34	42.26	0.11	0.24	0.75	49.73
SSA:	36.08	4.40	0.13	0.30	0.26	0.13	3.20	0.22	55.28
NEU:	28.21	30.93	3.08	8.71	3.46	2.15	0.05	23.12	0.29
EAF:	0.17	10.50	2.60	3.17	11.29	0.39	65.49	1.58	4.82
NAU:	4.81	2.96	2.09	1.29	66.55	0.29	13.18	0.86	7.97

Table 5: As in Table 4, but for summer temperature change analysis.

	CCC	CSIRO	CSM	DMI	GFDL	MRI	NIES	PCM	HADCM
ALA:	34.88	3.12	0.34	34.41	0.33	0.21	4.59	0.22	21.91
ENA:	14.07	34.38	0.58	43.57	0.08	0.04	0.22	1.78	5.27
SSA:	12.35	47.90	0.85	5.87	6.69	1.03	23.21	0.80	1.29
NEU:	40.50	2.92	0.39	43.88	0.69	0.16	1.09	0.70	9.67
EAF:	0.84	7.93	1.03	7.98	9.52	0.94	1.98	0.70	69.08
NAU:	4.59	5.67	1.72	31.09	15.68	2.04	29.30	1.23	8.67

Table 6: Median rank for each model over the 22 regions, for winter and summer temperature change predictions. For each region and season the models were ranked from first (1) to last (9) on the basis of the sorted values (from largest to smallest) of the posterior means of λ_i, $i = 1, ...9$. Then, the median value of the ranking for each model over the 22 regions was computed. As an example, if the median rank for model i is k, model i was ranked k^{th} or better in at least half the regions.

season	CCC	CSIRO	CSM	DMI	GFDL	MRI	NIES	PCM	HADCM
DJF	6.5	5.0	5.0	5.0	3.0	9.0	6.0	5.0	3.0
JJA	3.5	3.0	6.0	3.0	6.0	9.0	5.0	6.5	3.0