

Two Approaches to Quantifying Uncertainty in Global Temperature Changes

ANA LOPEZ

Climate Research Laboratory, Centre for the Environment, Oxford University, Oxford, United Kingdom

CLAUDIA TEBALDI

National Center for Atmospheric Research, Boulder, Colorado

MARK NEW

Climate Research Laboratory, Centre for the Environment, Oxford University, Oxford, United Kingdom

DAVE STAINFORTH AND MYLES ALLEN

Atmospheric, Oceanic and Planetary Physics, University of Oxford, Oxford, United Kingdom

JAMIE KETTLEBOROUGH

Climate Research Data Applications, Met Office, Exeter, United Kingdom

(Manuscript received 6 June 2005, in final form 15 December 2005)

ABSTRACT

A Bayesian statistical model developed to produce probabilistic projections of regional climate change using observations and ensembles of general circulation models (GCMs) is applied to evaluate the probability distribution of global mean temperature change under different forcing scenarios. The results are compared to probabilistic projections obtained using optimal fingerprinting techniques that constrain GCM projections by observations. It is found that, due to the different assumptions underlying these statistical approaches, the predicted distributions differ significantly in particular in their uncertainty ranges. Results presented herein demonstrate that probabilistic projections of future climate are strongly dependent on the assumptions of the underlying methodologies.

1. Introduction

Two different approaches, based on fundamentally different assumptions, have recently been developed to estimate the uncertainty in climate model predictions of future global and regional mean surface temperature change, under different forcing scenarios.

The method introduced by Allen and coworkers (Allen et al. 2000, 2003; Stott and Kettleborough 2002, hereafter referred to as ASK) assumes that since general circulation models (GCMs) are designed and tuned to reproduce the observed climate, any robust prediction of uncertainty in future climate should be as model

independent as possible, and be constrained by the only *objective* information available, that is, the observed climate and recent climate change.

On the other hand, the statistical model developed by Tebaldi and coworkers (Tebaldi et al. 2004, 2005) uses Bayesian statistics to estimate a distribution of future climatologies, from the combination of past observed and the corresponding GCM-simulated climatologies. The Bayesian approach is motivated by observing that the earth's climate is a *nonrepeatable* experiment, and probabilities cannot be determined through a frequentist approach, as that would require a sample from a large number of planets earth and their corresponding climates. Therefore, a GCM ensemble is assumed to be a sample of the full potential climate model space compatible with the observed climate, and a statistical model is developed that combines the information from these GCMs with the single observed

Corresponding author address: Ana Lopez, Centre for the Environment, Oxford University, South Parks Rd., Oxford OX13QY, United Kingdom.
E-mail: alopez@ouce.ox.ac.uk

climate realization. The characteristics of the predicted distribution of climate change will of course depend on the assumptions built into the statistical model that describes the data.

The goal of this work is to compare predictions of global temperature change using these two methodologies in order to better understand their characteristics and limitations, and the dependence of their predictions on the underlying assumptions. We note that Tebaldi's method has been conceived and used to estimate probability density functions (PDFs) of climate change specifically for regionally aggregated signals, deriving its conceptual framework from the work by Giorgi and Mearns (2002), whose criteria to weight the ensemble members were formulated for the specific regional scales analyzed therein. On the other hand, the ASK approach only recently has been applied at the regional scale (Stott 2003; Stott et al. 2006a). Therefore only by applying Tebaldi's approach at the global scale can a comparison of the two methodologies be carried out. Thus, after positing this caveat about the different original conceptualization, we proceed with the comparison believing that it might nonetheless provide insights into the assumptions and results of both approaches.

In what follows, we apply the Bayesian model to a set of 16 coupled GCMs, from the second Coupled Model Intercomparison Project (CMIP-2; Covey et al. 2003), including the NCAR–Washington–Meehl (NCARWM) coupled model (Meehl et al. 1996). We use the control and forced runs of these models to compute the probability distribution for the transient climate response (TCR; the mean temperature change at the time of CO₂ doubling under a 1% annual increase of CO₂ concentration). We compare our results with those obtained using ASK (Allen et al. 2003). We also apply this model to five of the above GCMs for which simulations under different future forcing scenarios are available at the Intergovernmental Panel on Climate Change Data Distribution Center (IPCC DDC; information online at <http://www.ipcc-ddc.cru.uea.ac.uk>). In this case we compute the distribution of global temperature change under alternative forcing scenarios in the twenty-first century for three different decades and compare our results with the ones obtained using the ASK methodology (Stott and Kettleborough 2002).

The paper is organized as follows. In section 2 we summarize the basic assumptions of the ASK approach following Kettleborough et al. (2006), outline the Bayesian model from Tebaldi et al. (2005), and describe the data used in our analysis. In section 3 we use the Bayesian model to compute PDFs for the transient climate response and global temperature change under

different Special Report on Emissions Scenarios (SRES) forcing scenarios in the twenty-first century. We compare the ASK approach with our results and explore how changes to the assumptions of the Bayesian model modify them. We conclude in section 4 with a discussion of our results and possible modifications to the Bayesian model that would take into account some of the aspects of GCM datasets not accounted for in its present form, such as model tuning to observations and intermodel dependence.

2. Statistical models and climate data

a. ASK approach

The ASK approach evolved from climate change attribution studies, in which past warming detected in climate observations was attributed to different possible forcing factors applying optimal fingerprinting techniques (Allen and Stott 2003; Hasselmann 1997). Using spatiotemporal linear regression, the observed warming (y) over the twentieth century is written as a linear combination of the modeled response patterns (x_i) to different forcings (labeled by the index i) as follows:

$$y = \sum_{i=1}^m (x_i - v_i)\beta_i + v_0, \quad (1)$$

where the β_i are the linear regression coefficients. These coefficients can be interpreted as the numbers by which the model's simulated response to each forcing can be scaled up or down and still remain consistent with the observations. The sources of uncertainty in Eq. (1) are the errors in the response patterns (v_i) and the error that causes the observations to be different from the linear fit (v_0), which in turn are due to internal variability of the climate system. Moreover, in Eq. (1) the observations and response patterns can be space and time dependent. In this way, that is, by keeping information about spatial patterns (even for global studies), the signal-to-noise ratio is maximized, and signals that are not isotropically distributed in the atmosphere (such as aerosols) can still be detected (Stott et al. 2006b). Notice that since the β_i are space independent, this approach only allows for errors in the GCMs simulations that change the magnitude of the response, but not its shape (Kettleborough et al. 2006).

Allen et al. (2000) extended optimal fingerprinting to quantify uncertainty in predictions of global mean temperature change in the following way. Assuming that a model that over- or underestimates the climate response to past forcing will continue to do so by a similar fraction in the future, the same scaling factors [β_i computed in Eq. (1)] were applied to the model predictions

(x_i^{for}) to produce a best estimate of future warming (y_i^{for}) ,

$$y^{\text{for}} = \sum_{i=1}^m (x_i^{\text{for}} - v_i^{\text{for}})\beta_i + v_0^{\text{for}}, \quad (2)$$

where v_i^{for} represents uncertainty in the model response (due to internal variability, model error, and/or uncertainty in the future forcing), and v_0^{for} represents any reason why the future climate is different from the scaled version of the modeled forced response (including internal variability and errors resulting from assuming that the optimal fingerprinting scaling factors can be used to predict future response). The distribution for future climate change, y_i^{for} , can be calculated from Eq. (2) given the distributions for β_i , v_i^{for} , and v_0^{for} . The distributions of the β_i are Gaussians with mean and variance determined by the total least squares solution of the regression Eq. (1), provided that the noise in the observations and the model responses are Gaussian and independent, and all the possible values of β_i are equally probable a priori (Kettleborough et al. 2006).

Stott and Kettleborough (2002) used this approach to compute the future global mean temperature rise under a range of SRES emissions scenarios, and Allen et al. (2003) applied it to the data from the coupled model of the Hadley Centre (HadCM3) to estimate a probability distribution for TCR.

A crucial property of this technique is that, for the projected global warming over the twenty-first century under a given forcing scenario, the predictions are primarily constrained by the observations and depend only secondarily on the GCM used to produce them (Allen et al. 2000; Kettleborough et al. 2006; Stott et al. 2006b). It is reasonable to expect that for this particular forecast (i.e., global warming over the twenty-first century), the consistency of the results for different GCMs can be justified by the fact that all properly formulated climate models must satisfy a set of basic physical constraints that regulate global mean temperature change at the time scales considered in this analysis (Allen et al. 2003; Kettleborough et al. 2006).

However, this method can be applied to other variables and spatial scales, provided that in Eq. (1) the strength of the anthropogenic signal in the given variable is large compared to the internal variability. Otherwise, the predicted change will be poorly constrained by observations and its corresponding uncertainty very large (Stott et al. 2006a). For some variables, such as global precipitation, the signal of attributable change is delayed with respect to the signal in global temperature, and the method will not be applicable until the signal appears in the variable of interest (Allen and Ingram 2002).

Another important characteristic of this approach is that the uncertainty limits in the predictions are determined by a combination of the error in the linear fit between a given GCM and observations, and a GCM-based estimate of the noise introduced by the natural variability [v_0 and v_0^{for} in Eqs. (1) and (2), respectively]. This provides a natural limit to how much the uncertainty can be reduced by improving the fit.

b. Tebaldi et al. approach

Tebaldi et al. (2005) developed a Bayesian statistical model that combines data from observations and a multimodel ensemble of GCMs to compute PDFs of future temperature and precipitation change over large regions (Giorgi and Francisco 2000b) under different forcing scenarios. This model constitutes a formal Bayesian implementation and extension of the reliability ensemble averaging (REA) approach (Giorgi and Mearns 2002, 2003). A basic assumption of the REA approach is that the ability of a GCM to reproduce current mean climate (its bias) constitutes a measure of its reliability. However, smaller bias in the current-climate model data does not necessarily imply greater accuracy in the future; therefore, the criteria of convergence is introduced, where a GCM is also considered to be more reliable if its future projection is closer to the weighted ensemble mean. In the Tebaldi et al. approach, these two criteria inform the shape of the posterior distribution as a consequence of the assumptions formulated in the statistical model, without being explicitly formalized in it.

Within the framework of Bayesian statistics, all uncertain variables of interest are assumed to be random quantities. Thus, given these variables (for instance present and future temperature) and the data (observed and GCM-simulated present temperature and GCM-simulated future temperature), Bayes's theorem states that the probability of the variables given the data (the posterior distribution) is proportional to the probability of the data given the variables (the likelihood) times the probability of the variables (the prior distribution). In the rest of this section, we briefly describe the likelihoods, and the prior and posterior distributions of the Tebaldi et al. model. For a more complete discussion, including the methodology used to compute the full posterior distribution, we refer the reader to Tebaldi et al. (2005).

As a first approximation, which does not account for issues such as tuning models to observations or inter-model correlations, models and observed climate are treated as statistically independent, and are assigned Gaussian likelihoods. Thus, the likelihoods for the observations of current global mean temperature (X_0),

and simulations of present (X_i) and future (Y_i) mean temperature by the i th model, are normal distributions given by

$$\begin{aligned} X_0 &\sim N[\mu, (\lambda_0)^{-1}] \\ X_i &\sim N[\mu, (\lambda_i)^{-1}] \\ Y_i &\sim N[\nu + \beta(X_i - \mu), (\theta\lambda_i)^{-1}], \end{aligned} \quad (3)$$

where, μ and ν are random variables representing the (unknown) true present and future mean temperatures, respectively. The random variables $\sigma_i^2 = \lambda_i^{-1}$ can be thought of as a measure of the i th GCM precision, and $\sigma_0^2 = \lambda_0^{-1}$ is a number that estimates the natural variability specific to the time average applied to the observations. The random variable θ allows for the possibility of the future and present temperatures having different variances by a multiplicative factor common to all GCMs. Finally, the random variable β introduces a correlation between present and future temperature responses.

The likelihoods described above are defined in terms of the random variables μ , ν , λ_i , θ , and β whose prior distributions need to be specified. At times prior knowledge, independent of the data at hand, is available, either as expert judgment or as a result of previous studies that produced information about the random variables of interest. In such cases, the prior distribution can be assigned a specific, “informative” shape. In Tebaldi et al. (2005) all the prior distributions are chosen to be uninformative, to make the analysis as objective as possible. In particular, the prior distributions for μ , ν , and β are uniform densities over the real line; those for λ_i and θ are gamma distributions whose parameters (a , b and c , d , respectively) are chosen such that the distribution has mean one and a large variance over the positive real line.

Using Bayes’s theorem, the joint posterior distribution for the parameters μ , ν , θ , β , λ_i is obtained as the product of the likelihoods and the prior distributions specified above, and is given by (up to a normalization constant)

$$\prod_{i=1, N} \left(\lambda_i^{a-1} e^{-b\lambda_i} \lambda_i^{\theta^{1/2}} \exp \left\langle -\frac{\lambda_i}{2} \{ (X_i - \mu)^2 + \theta [Y_i - \nu - \beta(X_i - \mu)]^2 \} \right\rangle \right) \theta^{c-1} e^{-d\theta} \exp \left[-\frac{\lambda_0}{2} (X_0 - \mu)^2 \right]. \quad (4)$$

In obtaining this distribution, the likelihood of the data given the unknown variables has been written as a product of the likelihoods of each individual GCM and the observations, the underlying assumption being that model-simulated temperatures and observations are statistically independent.

From the joint posterior distribution, we can derive the conditional posterior distribution for the present temperature for instance, by fixing the rest of the variables and considering the resulting distribution as a function of present temperature only.

The conditional distribution for μ is a Gaussian with mean and variance

$$\begin{aligned} \tilde{\mu} &= \frac{\sum \lambda_i X_i - \theta \beta \sum \lambda_i (Y_i - \nu - \beta X_i) + \lambda_0 X_0}{\sum \lambda_i + \theta \beta^2 \sum \lambda_i + \lambda_0}, \\ \sigma_\mu &= \frac{1}{\sum \lambda_i + \theta \beta^2 \sum \lambda_i + \lambda_0}, \end{aligned} \quad (5)$$

respectively. Since it was assumed that the observations (X_0) and the GCM climatologies (X_i) are statistically independent, their roles in the predicted posterior of present temperature are similar. However, while the Bayesian model self-consistently computes the model’s weights (λ_i), the observation’s weight (λ_0) is a fixed parameter.

Analogously, the conditional distribution for ν is a Gaussian with mean and variance

$$\begin{aligned} \tilde{\nu} &= \frac{\sum \lambda_i [Y_i - \beta(X_i - \mu)]}{\sum \lambda_i}, \\ \sigma_\nu &= \frac{1}{\theta \sum \lambda_i}. \end{aligned} \quad (6)$$

Notice that in the determination of the mean present and future true temperatures, the precision parameter λ_i acts as a weight for the data points of its corresponding GCM. Equations (5) and (6) show that the variances of μ and ν scale as $1/N$ for large N , where N is the number of GCMs in the ensemble. Thus, there is no lower bound to the uncertainty associated to these quantities.

Similarly derived, the conditional posterior distribution of λ_i is a Gamma function with mean

$$\tilde{\lambda}_i = \frac{a + 1}{b + \frac{(X_i - \mu)^2}{2} + \frac{\theta [Y_i - \nu - \beta(X_i - \mu)]^2}{2}}. \quad (7)$$

This expression shows how the bias and convergence criteria are implicitly built into the statistical model, since the individual precision parameter or weight λ_i for

each GCM is large provided that bias $|X_i - \mu|$ and convergence $|Y_i - \nu|$ are small. Within this framework, due to the fact that the main criteria to weight a GCM are its bias and convergence, the weighted ensemble mean has a strong influence on the predictions. In other words, the results are constrained by their convergence to a future climate essentially determined by the weighted ensemble mean. Note, however, that these weights or precision parameters are formulated as random variables themselves, and their distribution is estimated jointly with the other random quantities' distributions. In other words, these weights are not fixed quantities assigned to each GCM but rather they are random quantities, self-consistently determined by the Bayesian model once the initial choices for likelihoods and prior distributions are made, incorporating the uncertainty inherent in the evaluation of the relative skills of the different GCMs. In fact, talking about "weights" for the Bayesian model is a conceptual, but useful, simplification of the role of the parameters λ_i in the analysis.

Given the predicted posterior distributions for μ and ν , the Bayesian model can estimate the signal of temperature change as the posterior distribution of $\Delta T = \nu - \mu$, which by definition does not include the uncertainty related to the natural variability. This fact alone will likely make the PDFs from this approach tighter than the ones from ASK in which the noise associated with the natural variability is explicitly taken into account in the determination of the scaling factors (see section 2a).

c. Data

To compute the PDF for transient climate response (TCR), we apply the Bayesian model to data from 16 CMIP2 GCMs, for which there are constant forcing (control run) and perturbed ($1\% \text{ yr}^{-1}$ increasing atmospheric CO_2) simulations. We take the mean of the control run for each model as the simulated present climate (X_i), because the CO_2 concentrations in the control runs are comparable to the observed ones during the period 1960–90. Since the time of CO_2 doubling in the perturbed runs is year 70, the simulated future climate (Y_i) is computed as the 20-yr mean centered at year 70 of these runs, except for the NCARWM model where data are available only up to year 75. In this case, we consider the average of the last 15 yr of the forced run.

The observed temperature for the period 1960–90 is taken as $X_0 = 287.25 \text{ K}$ (Jones et al. 1999) and its natural variability as $\lambda_0 = 56.33 \text{ K}^{-2}$ ($\sigma_0 = 0.13 \text{ K}$), estimated, as in Tebaldi et al. (2005), by computing 30-yr moving averages of observed, detrended global mean

temperatures over the twentieth century and taking the differences between maximum and minimum values.

For the computation of global temperature change in the twenty-first century under different SRES scenarios, the available data compose time series for global surface temperature for the period 1900–2100 for five different GCMs (ECHAM4, HadCM3, CGCM2, CSIR, CCSR) with twenty-first century runs under A2 and B2 forcing scenarios.

We take as model data for the present climate (X_i) the 1960–90 mean for each model. We wish to compare our results against those of Stott and Kettleborough (2002), who used the ASK approach to calculate PDFs for global temperature change for the decades 2020–30, 2050–60, and 2090–00 relative to the 1990–2000 decade, under A2, B2, A1FI, and B1 SRES forcing scenarios. We therefore apply the Bayesian model to different sets of Y_i calculated as the 30-yr mean centered at 2025 and 2055, respectively. We ignore the last decade of the twenty-first century because most of the data are available only up to 2099, and a comparison with a 30-yr mean centered at 2085 introduces artificial discrepancies between the two results. We take a 30-yr mean instead of the 10-yr mean used by Stott and Kettleborough so that we can compare analogous climatologies within the Bayesian approach (which uses a 30-yr mean for the present and future climate).

3. Results

a. TCR

Figure 1 shows the predicted posterior distributions for present (μ , left curve in top panel) and future temperature (ν , right curve in top panel), and for the global temperature change $\Delta T = \nu - \mu$ (bottom panel). It is clear that the Bayesian model PDF for the global temperature change is much narrower than the one obtained using ASK (Allen et al. 2003, their Fig. 1, repeated in Fig. 2 below); the 95% interval for the Bayesian PDF is 1.66–1.97 K, while that for ASK is 1.12–3.43 K. The bimodality in the Bayesian PDF is due to the fact that there are two clusters of climate models converging to two different points in $\Delta T = \nu - \mu$.

It is not surprising that the two methods predict different uncertainty limits, since they are based on fundamentally different assumptions. However, since probabilistic predictions of climate change are increasingly being required for climate change impacts analysis, it is important to be able to provide explanations on how the different assumptions of the statistical models modify their predictions. There are various characteristics of Tebaldi's approach that may be responsible for the relatively narrower width of its PDF compared to

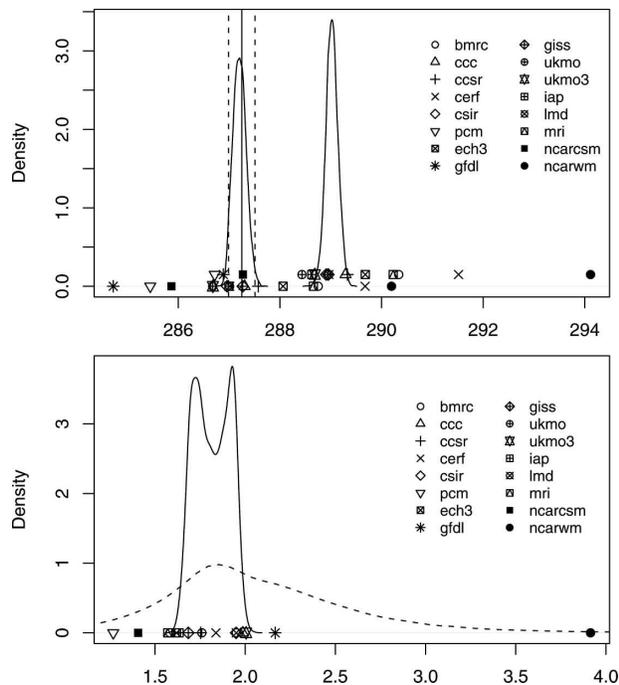


FIG. 1. (top) Predicted posterior distributions for present (μ , plotted on the left) and future global temperature (K) (ν , plotted on the right) using the Bayesian approach. The vertical solid line is the observed temperature X_0 , and the vertical dotted lines are the limits of the observed temperature's natural variability ($\pm 2\sigma_0$). The marks along the horizontal axis correspond to present and future temperatures simulated by the 16 GCMs. (bottom) Predicted global temperature change $\Delta T = \nu - \mu$ (K) using the Bayesian (solid line) and the ASK (dashed line) approaches.

that of ASK. Since in this work we restrict ourselves to an exploration of the current formulation of this Bayesian model, without reformulating its basic assumptions, we analyze to what extent the relaxation of the bias and convergence criteria, by varying the settings of the model priors, changes the uncertainty ranges of its predictions. Other possible causes of the discrepancies and suggestions for modifications of the basic statistical model to account for them will be discussed in section 4.

In the Bayesian approach, the variables that control bias and convergence are λ_0 , θ , and λ_i . In what follows, we will only focus on how λ_0 and the prior distribution for θ change the predictions of the statistical model. In principle, we could also explore the effect of modifying the priors for λ_i . However, it is difficult to justify a particular choice of λ_i priors for each different GCM, since there is no objective way to determine by how much each model should be deemed more or less precise.

The parameter λ_0 could be given a full Bayesian treatment if we had a sufficiently long record of observations (a multi-30-yr time series) that could be used

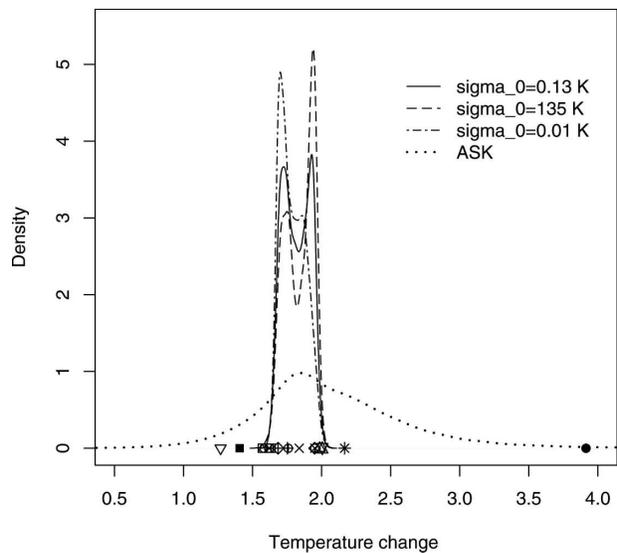


FIG. 2. Global temperature change (K) for different values of λ_0 . The values of λ_0 chosen correspond to $\sigma_0 = 135$ K (dashed line), $\sigma_0 = 0.01$ K (dashed-dotted line), and $\sigma_0 = 0.13$ K (solid line). The dotted line corresponds to the ASK approach. The marks along the horizontal axis indicate the temperature changes simulated by the 16 analyzed GCMs.

for its estimation. In that case, we would obtain a posterior distribution for λ_0 consistent with the data analyzed. However, we do not follow that approach here because our goal is simply to explore whether very different values of λ_0 could substantially change the predictions of the Bayesian model. It is clear from Fig. 1 that with sufficiently small λ_0 (large σ_0^2), all present-day GCM temperatures would fall within the range of the X_0 -postulated likelihood. However, Eq. (5) shows that, for small enough λ_0 , observations have very little influence on the predictions and the bias simply measures the distance between the GCM-simulated temperature and the weighted ensemble mean (Tebaldi et al. 2005). On the other hand, increasing λ_0 (decreasing σ_0^2) ties $\tilde{\mu}$ closely to the observations [Eq. (5)], but leaves more GCMs as outliers of the X_0 -postulated likelihood (top panel in Fig. 1). Therefore, changing λ_0 between extreme values gives us information about how the observations are weighted in the model predictions and how stringent the bias criterion is, since large λ_0 imposes a stricter constraint for the models to agree with the observations.

Figure 2 shows the PDFs for global temperature change for different values of λ_0 , compared to that obtained using ASK's approach. It is clear from this plot that allowing for the likelihood of the observations X_0 to be more inclusive (small λ_0 , large σ_0) or less inclusive (large λ_0 , small σ_0) of the estimations for present climate of the different GCMs does not affect the width of

the PDF. The distributions of climate change are still bimodal, again due to the fact that there are two clusters of climate models converging to two different points in $\Delta T = \nu - \mu$, but the amplitude of the two modes depends on the value of λ_0 since, as can be seen in Eq. (5), the values of λ_0 influence the estimate of μ . For the dataset analyzed here, larger posterior probabilities are attributed to smaller or larger values of the predicted present temperature for very small or very large λ_0 , respectively. This causes the shift in the mass of the distribution of $\Delta T = \nu - \mu$ as a function of the value of λ_0 observed in Fig. 2.

Figure 2 indicates that changing the estimated uncertainty in the observations in this Bayesian model does not alter significantly the uncertainty in the predictions for TCR. This is in marked contrast to the TCR distribution computed by ASK where the estimate of the uncertainty in observations (natural variability) effectively determines the width of the final distribution (Allen et al. 2003).

We next explore the variable θ , which was introduced to allow for the models to have different present and future temperature variances. A characteristic property of Bayesian models is that in the presence of a large dataset the influence of the prior on the posterior distribution becomes negligible. However, for small datasets, such as those used here, the priors' assumptions influence the results, which are then conditional on the assumptions made in defining the priors. Tebaldi et al. (2005) chose an uninformative prior for θ based on the fact that this choice ensures that the data points have maximum impact on the results in the absence of any a priori opinion. Alternatively, we can assume that the choice of prior for θ can be thought of as a tuning parameter for loosening the convergence criterion. Thus, a narrow prior with a small value for θ implies that we do not expect the future simulations to be as accurate as the present ones, so we make the convergence criteria less stringent.

Figure 3 shows PDFs of global temperature change for different priors for θ . The range of temperature changes predicted by the Bayesian model increases as the prior distribution for θ becomes more concentrated over very small values on the real line, that is, if both the mean value and the variance of this prior are very small. By inspecting Fig. 3, we see that $\langle \theta \rangle_{\text{prior}}$ somewhere between 0.1 and 0.01 produces a PDF of global temperature change whose uncertainty range is similar to the one produced using the ASK approach.

The effect of choosing small values for θ on the posterior distributions can be seen in Eq. (6), where it is evident that smaller θ values decrease the accuracy of the predicted future temperature. Moreover, Eq. (7)

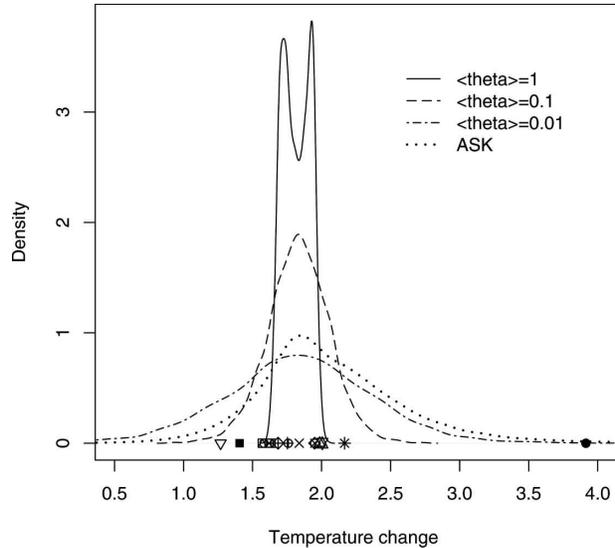


FIG. 3. Global temperature change (K) for different priors for θ . The dotted line corresponds to the ASK approach. The marks along the horizontal axis indicate the temperature changes simulated by the 16 analyzed GCMs.

shows that small θ values reduce the impact of the convergence criterion in the determination of the GCMs' precision parameters λ_i . Notice that, if one would like to drop the convergence criterion completely, a model with entirely different likelihoods has to be proposed. To set $\theta = 0$ is not a viable option, since that would turn the likelihoods of future climate (Y_i) into improper distributions.

We have seen so far how the choice of prior for θ affects the posterior distributions for global temperature change. It is natural to ask what will be its effect on regional predictions, in particular when choosing the informative prior for θ (with small $\langle \theta \rangle_{\text{prior}}$) that produces a PDF of similar spread to the ASK distribution. It is possible therefore to use this "best fit" prior for θ at the global scale as an informative prior to compute PDFs at regional levels. However, there is no physical or mathematical justification as to why the prior for a global analysis should be similar to a regional one. The question may simply be posed as an issue of the robustness of the regional results, that is, in terms of how much the regional PDFs will change when adopting different priors for θ . Tebaldi et al. (2004) explored this issue for changes in precipitation and found a consistent widening of the PDFs for small values of θ . However, the sensitivity of the results was different for different regions depending on the degree of clustering of the individual GCMs. Moreover in all cases, while the tails of the distributions changed significantly, the location of the central mass was basically unchanged. In Fig. 4 we present the results for two regional temperature

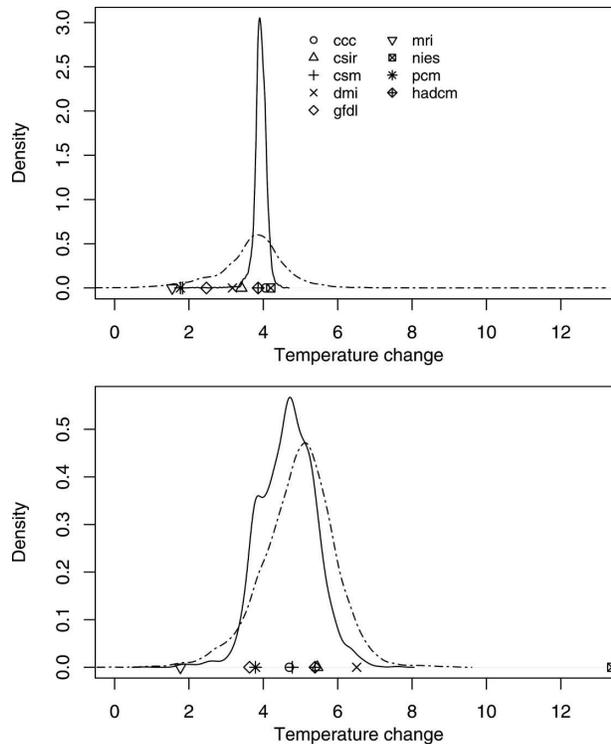


FIG. 4. Regional temperature change (K) for the uninformative prior for θ ($\langle\theta\rangle_{\text{prior}} = 1$, solid line), and for the informative prior ($\langle\theta\rangle_{\text{prior}} = 0.01$, dashed-dotted line). The two regions illustrated in the figure are (top) southern South America and (bottom) northern Europe. The marks along the horizontal axis correspond to regional temperature changes simulated by the 9 GCMs analyzed by Tebaldi et al. (2004).

change PDFs obtained with the uninformative prior θ , and with a prior that makes the global predictions with the Tebaldi model comparable with the ASK approach prediction for the A2 scenario. The two regions illustrated [southern South America (SSA) and northern Europe (NEU) as defined by Giorgi and Francisco (2000a)] are representative of two extreme cases. Using the standard prior for θ , in the SSA region five out of nine of the GCMs cluster around values of ΔT between 3 K and 4 K, thus the central mass of the distribution falls in that region. In the case of NEU, seven out of nine models are more evenly distributed between 3 K and 6 K, producing a distribution that has roughly that range. Figure 4 shows that for regions with many outliers the posterior PDF is very sensitive to changes in the priors, but it is far more stable when there are no outliers. However, in both cases the tails of the distributions change significantly when changing to the informative θ prior, indicating that care should be taken if these distributions are going to be used for impact analysis. We should also remark that our choice of $\langle\theta\rangle_{\text{prior}}$ implies a very small value for its variance as well,

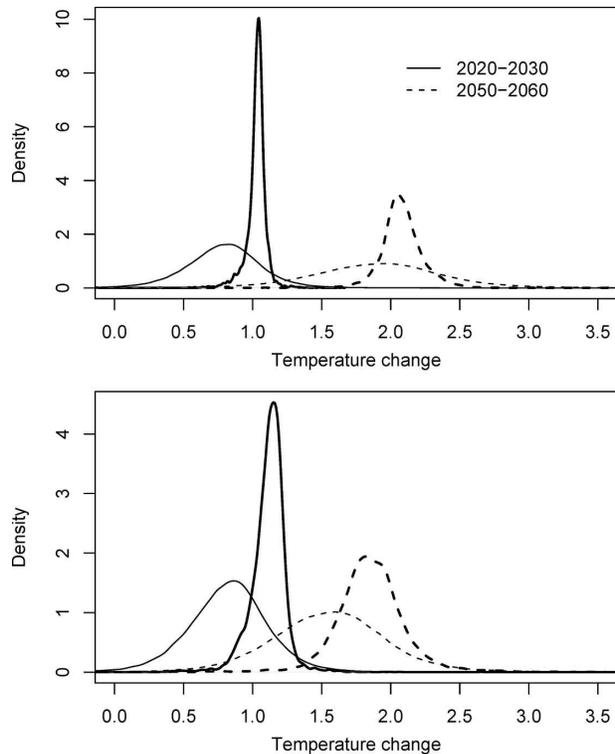


FIG. 5. Distributions of temperature change (K) under (top) A2 and (bottom) B2 forcing scenarios for 2020–30 and 2050–60, computed using the Bayesian approach (thicker lines) and that of Stott and Kettleborough (2002) (thinner lines).

effectively constraining the estimation of θ to the value chosen for $\langle\theta\rangle_{\text{prior}}$. In this case it is reasonable to expect that it will have a strong impact in the inferences.

b. SRES scenarios

In Fig. 5, we present the results obtained by analyzing five GCMs for which twenty-first century runs under A2 and B2 forcing scenarios are available from the IPCC DDC. These PDFs are calculated assuming an uninformative prior for θ . The Bayesian method predicts higher temperature changes for the two decades (2020–30 and 2050–60) than Stott and Kettleborough (2002). This is due to the fact that the mean of the distribution for the Bayesian prediction is the weighted ensemble mean for ΔT , which turns out to be larger than the scaled temperature change for the HadCM3 model, that is, the mean of Stott and Kettleborough's distribution.

We also observe that the range of uncertainty of the Bayesian PDF is closer (compared to the TCR calculation) to the one predicted by ASK without any tuning of the parameters. A possible explanation of this result is that the sample we are analyzing is very sparse and roughly evenly distributed within its range. Therefore,

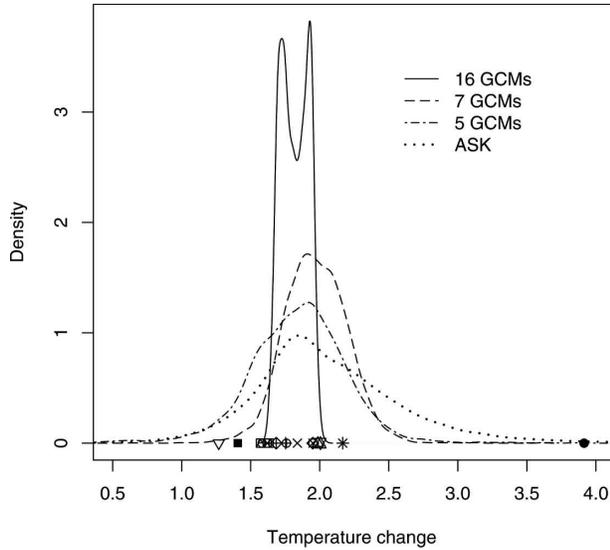


FIG. 6. TCR (K) for different sampling strategies, including 7 models (bmrc, pcm, gfdl, ukmo, ukmo3, ncarcs, and ncarwm; dashed line) and 5 models (cerf, pcm, ech3, ukmo3, and ncarwm; dashed-dotted line) from the original 16-model ensemble. The dotted line corresponds to the ASK approach. The marks along the horizontal axis indicate the temperature changes simulated by the 16 analyzed GCMs.

there is no clustering of models around any value of $\Delta T = \nu - \mu$; in particular, there is not clustering of models around an ensemble mean. This results in wider distributions due to the fact that all the data points contribute with a similar weight to the final distribution, in contrast with the case in which outliers do not contribute at all and clustered data points have a much higher weight.

To test the plausibility of this argument, in Fig. 6 we repeated the TCR calculation, but for samples of 7 and 5 models chosen to sample as uniformly as possible the range of the 16 original models. Within this sampling strategy, the wider uncertainty range is obtained for the five-model sample due to the fact that it samples the range of the full ensemble more uniformly. The sample of seven GCMs considerably widens the distribution as well. Note that this strategy of sampling models to span the range of uncertainty is in accord with Allen et al. (2003).

Besides the influence of the convergence criterion in giving large weight to models that cluster around one another and downweighting outliers, Fig. 6 shows that the reduction in the uncertainty range of the Bayesian posterior PDF is also due to the fact that the uncertainty decreases with sample size. To confirm this affirmation in a more general way, we computed TCR by applying the Bayesian model to ensembles of different sizes (n), randomly taken from the 16 GCMs. In Fig. 7

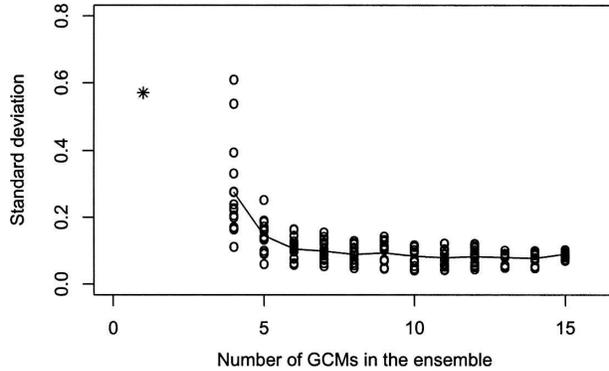


FIG. 7. Standard deviation of the posterior for $\Delta T = \nu - \mu$ (K) as a function of the number of samples n , for 15 different random choices for each n . The asterisk for $n = 1$ corresponds to the result obtained using the ASK approach. The black solid line connects the means of the plotted standard deviations.

we plot the standard deviation of the posterior for $\Delta T = \nu - \mu$ as a function of the ensemble size n ($n > 3$, otherwise the statistical model is overparameterized) for different random samples within each choice of n . The asterisk for $n = 1$ corresponds to the results obtained using the ASK approach. Since the sampling is random, the result will depend, of course, on the way the samples are distributed within the range of the 16 GCMs, and in particular, on whether the randomly sampled models are outliers or close to the weighted ensemble mean. Thus, for small values of n there are wide variations in the standard deviation. For larger n , the bias and convergence criteria embedded in the Bayesian model work more effectively at discarding outliers, reducing the dependence of the standard deviation of $\Delta T = \nu - \mu$ on the sample size.

Figure 7 shows (black solid line) that the uncertainty range predicted by the Bayesian model depends not only on the bias and convergence criteria as shown before, but also on the size of the sample, simply reflecting the fact that the variance of a sample of N random variables should be roughly proportional to $1/N$, if only statistical errors (i.e., those that scale with sample size) are taken into account.

On the other hand, we may ask how the distributions produced by the ASK approach might change, if the method was generalized to allow for the treatment of more than a single GCM. Central to the ASK approach is that at the spatial scales relevant to global studies, model-simulated response patterns are correct to within a scaling factor, so in this sense there are no independent models.

In an attribution study using the optimal fingerprinting approach, Gillett et al. (2002) choose to undertake linear regression between the observations and the

multimodel mean and produce a unique distribution in this way, determining the uncertainty range not only by the error in the linear fit but also by the natural variability of the climate system as simulated by the control runs of the analyzed GCMs. Gillett's approach suggests that even if one could assume that the GCMs are statistically independent (as Gillett et al. do when taking the simple mean), the uncertainty range cannot simply scale with sample size. In a more recent study, Stott et al. (2006b) apply the ASK approach to obtain global predictions of climate change for three GCMs. Even though there are differences between the means and uncertainty ranges of the PDFs for the different GCMs, the predicted means are closer together than for the unscaled model simulations, and the uncertainty ranges are comparable. Therefore, a simple mean of the three distributions (which assumes that they are statistically independent) has an uncertainty range similar to those of the individual distributions. These results suggest that the natural variability imposes a lower bound to the uncertainty range that is violated if all the errors are treated as statistical errors scaling with sample size.

4. Discussion

We have compared two different approaches (ASK and Tebaldi) for estimating uncertainty (formalized through PDFs) for global mean temperature change predicted by small ensembles of global climate models. We found that the uncertainty ranges predicted by the Bayesian method of Tebaldi et al. (2005) are substantially narrower than those obtained using the ASK approach.

Due to the fact that the statistical approaches are based on completely different assumptions, there is neither reason to expect that both methods will produce the same uncertainty range, nor the same expected value of temperature change. However, since impacts assessment communities are increasingly requiring probabilistic projections of climate change, it is important to provide clear explanations about how the assumptions made in formulating the statistical model affect its predictions.

There are several characteristics of the Bayesian model that influence the width of its posterior PDFs, and that are possibly relevant in determining the difference with respect to the results obtained using ASK's approach. In what follows, we summarize our findings regarding the effect of bias and convergence on Tebaldi et al.'s predictions, and discuss possible modifications of this model to account for the other factors that determine their uncertainty range.

a. Bias and convergence criteria

The Tebaldi et al. model formalizes the REA approach (Giorgi and Mearns 2002), giving higher weight to predictions from those GCMs that are better at reproducing present *mean* climate (small bias) and that remain close to the weighted ensemble mean in the future (strong convergence). In the previous section, we explored how these criteria can be relaxed by modifying the prior parameters of the model. In particular, we found that by relaxing the convergence criteria, that is, by assuming an informative small prior for θ , the uncertainty range can be increased. A possible choice of informative θ prior could be the one that gives a hindcast for attributable warming consistent with current expert opinion in the attribution community (Goldstein and Rougier 2004; Kennedy and O'Hagan 2001).

While the influence of bias and convergence in the predicted uncertainty range was analyzed in this paper without changing the basic assumptions of Tebaldi's model, the investigation of the following factors requires a reformulation of the statistical assumptions that goes beyond the scope of this work.

b. Types of uncertainties

In the previous section, we have shown that besides the influence of the bias and convergence criteria in determining the uncertainty range predicted by the Bayesian model, for large sample sizes the uncertainty will scale with $1/N$, consistent with the fact that all the sources of errors are treated as statistically independent errors. However, in this problem there are other uncertainties that do not scale with sample size, such as those introduced by the presence of internal variability in observations and GCM-simulated patterns, and systematic biases of the GCMs, due to the fact that they have been tuned to reproduce the observations. If these errors were taken into account in the formulation of the statistical model, they would presumably impose a lower bound on the estimated uncertainty range in a way analogous to that done by the natural variability in the ASK approach, and could potentially induce correlations between different GCMs.

c. Statistical independence of the GCMs

The Tebaldi et al. model is built on the assumption that the GCMs and observations are statistically independent. This was originally justified by the fact that even though the climate models are tuned to reproduce patterns of observed climate at global scales, at regional scales the assumption of independence was a reasonable first-order approximation. However, due to the fact that GCMs are constructed to reproduce observa-

tions, a more natural way of writing the likelihood of models and observations would be through a joint multivariate distribution, whose specification is, however, problematic because of the difficulty of quantifying intermodel dependence. Perhaps a simpler road to the same end is the analysis of specifically designed multi-model experiments, or perturbed physics ensembles, where the systematic sampling could admit an easier formalization of such an issue.

d. Spatial patterns and time trends

While the ASK approach exploits the GCM information about time trends and spatial patterns, comparing them to the observed ones, the Tebaldi et al. model has as inputs 20- or 30-yr averages of global temperatures, missing potentially valuable information about the ability of the models to reproduce temporal and spatial patterns. The generalization of Tebaldi's model to the analysis of time trends instead of climatology is rather straightforward, for example, by including a trend term in the mean component of the Gaussian likelihoods. However, preliminary results show that the uncertainty range of the predictions is not substantially changed by this reformulation. The treatment of climate model output at high resolution rather than aggregate regional measures requires a nonstraightforward extension by way of modeling spatial processes on the sphere, in order to incorporate the spatial correlation structure between grid locations. At present, there are no results about uncertainty characterization at regional scales comparable to the regionally aggregated results.

In this paper, we have compared predictions of global climate change computed using two different approaches. We found that Tebaldi et al.'s Bayesian model predicts very tight uncertainty ranges at the global level. Furthermore, the tails of the computed PDFs are sensitive to the choice of models within the GCM ensemble, and to the prior for the parameter θ , which controls the models' convergence. This suggests that when applied to regional predictions, this method is useful for illustrating the relative uncertainty in the predictions from region to region. However, its use in providing quantitative information such as confidence intervals on regional climate change is potentially more problematic, since the predicted PDFs are highly conditional on the formulation of the statistical model, the priors for the model variables, and the GCM sample.

On the other hand, even though the larger uncertainty range predicted by the ASK approach is consistent with increasingly large uncertainty ranges found for climate sensitivity by different authors (Forest et al. 2002; Frame et al. 2005; Gregory et al. 2002; Knutti et al. 2002; Wigley and Raper 2001), this result is also

dependent on the assumptions made to analyze the data. In particular, it relies on a strictly linear extrapolation that is well justified for global temperature change but might not be straightforwardly applicable to small regional scales or different climate variables (Allen and Ingram 2002). Moreover, even though the method is designed to depend primarily on observations and only secondarily on the GCM analyzed, further work is needed in order to strictly quantify this dependence and its influence on the fine details of the predicted distributions (Stott et al. 2006a).

We conclude that probabilistic predictions of climate change appear to be at this point highly dependent on the assumptions built into the statistical model developed to analyze the data. In this sense, their application to impacts or risks studies can only be made subject to a clear understanding of the influence of those assumptions on the final results.

Acknowledgments. Ana Lopez holds a Daphne Jackson Fellowship funded by the U.K. Natural Environment Research Council. Claudia Tebaldi's research is funded by NCAR's Weather and Climate Impacts Assessment Science Program. NCAR is sponsored by the National Science Foundation. MRA received partial support from the NOAA-DoE International Detection and Attribution Group.

REFERENCES

- Allen, M. R., P. A. Stott, J. F. B. Mitchell, R. Schnur, and T. L. Delworth, 2000: Quantifying the uncertainty in forecasts of anthropogenic climate change. *Nature*, **407**, 617–620.
- , and W. J. Ingram, 2002: Constraints on future changes in climate and the hydrologic cycle. *Nature*, **419**, 224–232.
- , and P. A. Stott, 2003: Estimating signal amplitudes in optimal fingerprinting. Part I: Theory. *Climate Dyn.*, **21**, 477–491.
- , J. A. Kettleborough, and D. Staniforth, 2003: Model error in weather and climate forecast. *Proc. Predictability Seminar*, Reading, United Kingdom, ECMWF, 275–294.
- Covey, C., K. M. AchutaRao, U. Cubasch, P. Jones, S. J. Lambert, M. E. Mann, T. J. Phillips, and K. E. Taylor, 2003: An overview of results from the Coupled Model Intercomparison Project. *Global Planet. Change*, **37**, 103–133.
- Forest, C. E., C. E. Forest, P. H. Stone, A. P. Sokolov, M. R. Allen, and M. D. Webster, 2002: Quantifying uncertainties in climate system properties with the use of recent climate observations. *Science*, **295**, 113–117.
- Frame, D. J., B. B. Booth, J. A. Kettleborough, D. A. Staniforth, J. M. Gregory, M. Collins, and M. R. Allen, 2005: Constraining climate forecasts: The role of prior assumptions. *Geophys. Res. Lett.*, **32**, L09702, doi:10.1029/2004GL022241.
- Gillett, N. P., F. W. Zwiers, A. J. Weaver, G. C. Hegerl, M. R. Allen, and P. A. Stott, 2002: Detecting anthropogenic influence with a multi-model ensemble. *Geophys. Res. Lett.*, **29**, 1970, doi:10.1029/2002GL015836.
- Giorgi, F., and R. Francisco, 2000a: Evaluating uncertainties in the prediction of regional climate change. *Geophys. Res. Lett.*, **27**, 1295–1298.

- , and —, 2000b: Uncertainties in regional climate change prediction: A regional analysis of ensemble simulations with the HADCM2 coupled AOGCM. *Climate Dyn.*, **16**, 169–182.
- , 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**, 1141–1158.
- , and —, 2003: Probability of regional climate change based on the reliability ensemble averaging (REA) method. *Geophys. Res. Lett.*, **30**, 1629, doi:10.1029/2003GL017130.
- Goldstein, M., and J. Rougier, 2004: Probabilistic formulations for transferring inferences from mathematical models to physical systems. *SIAM J. Sci. Comput.*, **26**, 467–487.
- Gregory, J. M., R. J. Stouffer, S. C. B. Raper, P. A. Stott, and N. A. Rayner, 2002: An observationally based estimate of the climate sensitivity. *J. Climate*, **15**, 3117–3121.
- Hasselmann, K., 1997: Multi-pattern fingerprint method for detection and attribution of climate change. *Climate Dyn.*, **13**, 601–611.
- Jones, P. D., M. New, D. E. Parker, S. Martin, and I. G. Rigor, 1999: Surface air temperature and its changes over the past 150 years. *Rev. Geophys.*, **37**, 173–199.
- Kennedy, M. C., and A. O’Hagan, 2001: Bayesian calibration of computer models. *J. Roy. Stat. Soc.*, **63B**, 425–450.
- Kettleborough, J. A., B. B. Booth, P. A. Stott, and M. R. Allen, 2006: Estimates of uncertainty in predictions of global mean surface temperature. *J. Climate*, in press.
- Knutti, R., T. F. Stocker, F. Joos, and G.-K. Plattner, 2002: Constraints on radiative forcing and future climate change from observations and climate model ensembles. *Nature*, **416**, 719–723.
- Meehl, G. A., W. M. Washington, D. J. Erickson III, and B. P. Briegleb, 1996: Climate change from increased CO₂ and direct and indirect effects of sulfate aerosols. *Geophys. Res. Lett.*, **23**, 3755–3758.
- Stott, P. A., 2003: Attribution of regional-scale temperature changes to anthropogenic and natural causes. *Geophys. Res. Lett.*, **30**, 1728, doi:10.1029/2003GL017324.
- , and J. A. Kettleborough, 2002: Origins and estimates of uncertainty in predictions of twenty-first century temperature rise. *Nature*, **416**, 723–726.
- , —, and M. R. Allen, 2006a: Uncertainty in continental-scale temperature predictions. *Geophys. Res. Lett.*, **33**, L02708, doi:10.1029/2005GL024423.
- , J. F. B. Mitchell, M. R. Allen, T. L. Delworth, J. M. Gregory, G. A. Meehl, and B. D. Santer, 2006b: Observational constraints on past attributable warming and predictions of future global warming. *J. Climate*, **19**, 3055–3069.
- Tebaldi, C., L. O. Mearns, D. Nychka, and R. L. Smith, 2004: Regional probabilities of precipitation change: A Bayesian analysis of multimodel simulations. *Geophys. Res. Lett.*, **31**, L24213, doi:10.1029/2004GL021276.
- , R. L. Smith, D. Nychka, and L. O. Mearns, 2005: Quantifying uncertainty in projections of regional climate change: A Bayesian approach to the analysis of multimodel ensembles. *J. Climate*, **18**, 1524–1540.
- Wigley, T. M. L., and S. C. B. Raper, 2001: Interpretation of high projections for global-mean warming. *Science*, **293**, 451–454.