

## Evidence of Decadal Climate Prediction Skill Resulting from Changes in Anthropogenic Forcing

TERRY C. K. LEE

*Department of Mathematics and Statistics, University of Victoria, Victoria, British Columbia, Canada*

FRANCIS W. ZWIERS

*Canadian Centre for Climate Modelling and Analysis, Climate Research Division, Environment Canada, University of Victoria, Victoria, British Columbia, Canada*

XUEBIN ZHANG

*Climate Monitoring and Data Interpretation Division, Climate Research Division, Environment Canada, Downsview, Ontario, Canada*

MIN TSAO

*Department of Mathematics and Statistics, University of Victoria, Victoria, British Columbia, Canada*

(Manuscript received 7 June 2005, in final form 17 February 2006)

### ABSTRACT

It is argued that simulations of the twentieth century performed with coupled global climate models with specified historical changes in external radiative forcing can be interpreted as climate hindcasts. A simple Bayesian method for postprocessing such simulations is described, which produces probabilistic hindcasts of interdecadal temperature changes on large spatial scales. Hindcasts produced for the last two decades of the twentieth century are shown to be skillful. The suggestion that skillful decadal forecasts can be produced on large regional scales by exploiting the response to anthropogenic forcing provides additional evidence that anthropogenic change in the composition of the atmosphere has influenced the climate. In the absence of large negative volcanic forcing on the climate system (which cannot presently be forecast), it is predicted that the global mean temperature for the decade 2000–09 will lie above the 1970–99 normal with a probability of 0.94. The global mean temperature anomaly for this decade relative to 1970–99 is predicted to be 0.35°C with a 5%–95% confidence range of 0.21°–0.48°C.

### 1. Introduction

There is now a large body of evidence (see, e.g., the reviews of Mitchell et al. 2001; International Ad Hoc Detection and Attribution Group 2005) that changes in the external radiative forcing of the climate system have had a substantial impact on its evolution since the industrial revolution. These forcing changes have been caused by the changing composition of the atmosphere, mainly as a result of anthropogenic emissions of green-

house gases and aerosol precursors from fossil fuel burning, and secondarily by natural decadal-scale variations in volcanic and solar forcing. Evidence of the effects of these forcing changes on the climate system has been detected in surface air temperature at global scales (Mitchell et al. 2001; International Ad Hoc Detection and Attribution Group 2005) and recently also at continental and subcontinental scales (Zwiers and Zhang 2003; Stott 2003; Braganza et al. 2004; Gillett et al. 2004a; Zhang et al. 2006; Karoly and Wu 2005). A consistent picture of change that is related to the change in external forcing is also emerging in several other aspects of the climate system, such as ocean heat content, snow and sea ice cover extent, growing season length, tropopause height, precipitation, and mean sea level pressure (see, e.g., International Ad Hoc Detec-

---

*Corresponding author address:* Dr. Francis Zwiers, Canadian Center for Climate Modelling and Analysis, Climate Research Division, Environment Canada, University of Victoria, P. O. Box 1700, STN CSC, Victoria, BC V8W 2Y2, Canada.  
E-mail: francis.zwiers@ec.gc.ca

tion and Attribution Group 2005, and references therein).

Given the strength of the evidence, it seems natural to ask whether forcing projections can be used to forecast large-scale climate change on decadal time scales. Given their potential applications, many of which would involve some type of hedging, it is desirable that any decadal-scale forecast should be probabilistic rather than deterministic. One approach for obtaining such forecasts would be to produce large ensembles of forced climate simulations that would then be interpreted in a probabilistic manner, either directly or after some type of postprocessing to adjust for model biases. Unfortunately, such an approach would be expensive to implement given the need for large ensembles and the complexity of climate system models that have been used to study the evolution of the climate of the twentieth century (see, e.g., Mitchell et al. 2001; International Ad Hoc Detection and Attribution Group 2005). However, recent methodological developments, notably the application of Bayesian techniques to climate change detection (Berliner et al. 2000; Min et al. 2004; Schnur and Hasselmann 2005; Lee et al. 2005), provide a means by which this possibility can be evaluated using small ensembles of simulations. The remainder of this paper briefly explains the technique and describes preliminary results obtained when an ensemble of simulations of the twentieth century using only the history of anthropogenic forcing change are evaluated as climate hindcasts on decadal time scales.

## 2. Methods

In this section we will briefly describe the methods that we use to make and verify decadal hindcasts of temperature change. The term hindcast refers to a prediction statement about the past that is made using only information (such as initial conditions) that was available prior to the prediction period, while forecast corresponds to a statement about the future.

### a. Forecasting model

Forecasts of decadal climate change have several potential sources of skill. These include the initial ocean state (e.g., Boer 2004), possibly the initial land surface state, the response to external forcing conditions during the forecast period, and the continued adjustment of the climate during the forecast period toward a new equilibrium consistent with forcing that continues to persist, such as that which results from previous changes in atmospheric composition. Transient climate simulations of the twentieth century with specified his-

torical changes in radiative forcing can be thought of as climate hindcasts that attempt to exploit the latter two sources of skill. The extension of such simulations into the twenty-first century using scenarios of future emission change can further be considered as forecasts of future climate change, at least for relatively short 1–2 decade periods into the future (e.g., Zwiers 2002) because at these forecast leads, the climate response appears not to be very sensitive to the details of the forcing scenario that must be specified.

To describe the statistical forecasting model, we must first introduce some notation and make certain assumptions. Thus, we let  $O_t$  represent an observed decadal mean field containing  $n$  spatial grid points for a decade  $t$ , such as that of the 1970s. We can then think of the departure of  $O_t$  from some base climatology  $\bar{O}_t$  as being the sum of the response to external forcing during decade  $t$  relative to the base period, plus the effects of internal natural variability. The latter might itself be predictable on decadal time scales as an initial value problem, with skill obtained from ocean and perhaps land surface initial conditions as noted above (e.g., Grotzner et al. 1999; Collins and Allen 2002; Pohlmann et al. 2004), but we will assume that this is not the case for the purposes of this study. With these assumptions, a reasonable statistical model for  $\mathbf{Y}_t = \mathbf{O}_t - \bar{\mathbf{O}}_t$  might be

$$\mathbf{Y}_t = \mathbf{X}_t \beta_t + \varepsilon_t, \quad (1)$$

where vector  $\mathbf{X}_t = \mathbf{S}_t - \bar{\mathbf{S}}_t$  contains the model-simulated response in decade  $t$  to past external forcing relative to the climatological base period,  $\varepsilon_t$  results from internal variability, and  $\beta_t$  is a scaling factor that accounts for errors in the magnitude of the simulated response to the specified external forcing.

Such a model can be used to make a decadal climate forecast by making a suitable choice of a climatological base period. Following the standard World Meteorological Organization convention for defining the current mean climate, we have chosen the base period for decade  $t$  to be the previous three decades  $t - 10$ ,  $t - 20$ , and  $t - 30$ . We made this choice because it is an approach often used in seasonal forecasting (see, e.g., Derome et al. 2001). However, it does impose some challenges, particularly if it is anticipated that at least some of the skill will derive from external forcing. For example, the forcing from anthropogenic greenhouse gas emissions has gradually strengthened over the twentieth century, and this has resulted in a response that has also strengthened over time (e.g., Mitchell et al. 2001) and that therefore would have increasingly expressed itself in the moving operational climatology. Therefore, subtracting the current operational climatol-

ogy from the current decadal mean will remove at least part of the response that results from anthropogenic greenhouse gas forcing. Consequently, forecast skill that is anticipated to derive from such external forcing may be somewhat reduced when anomalies are expressed relative to the moving operational climatology.

We indicate decades by their first year. Thus, for the 1970s, the observed anomaly field is computed as

$$\mathbf{Y}_{1970} = \mathbf{O}_{1970} - (\mathbf{O}_{1960} + \mathbf{O}_{1950} + \mathbf{O}_{1940})/3, \quad (2)$$

and the corresponding model simulated field of response anomalies is given by

$$\mathbf{X}_{1970} = \mathbf{S}_{1970} - (\mathbf{S}_{1960} + \mathbf{S}_{1950} + \mathbf{S}_{1940})/3, \quad (3)$$

where, in order to reduce the effects of internal variability on the simulated response pattern,  $\mathbf{S}_t$  is in fact the mean of an ensemble of simulations of the twentieth century, each using the same forcing prescription. With such a convention and given an appropriate scaling factor estimate  $\hat{\beta}_t$ , a hindcast (or forecast depending upon the choice of  $t$ ) can be made for decade  $t$  by calculating

$$\hat{\mathbf{O}}_t = \bar{\mathbf{O}}_t + \hat{\mathbf{X}}_t \hat{\beta}_t + \hat{\varepsilon}_t, \quad (4)$$

where  $\hat{\mathbf{X}}_t = \mathbf{X}_t$  is the model forecast anomaly response and  $\hat{\varepsilon}_t = \mathbf{0}$  is the forecast of the internal variability in decade  $t$ . The third term,  $\hat{\varepsilon}_t$ , is zero because we have assumed that there is no climate predictability from internal sources on decadal time scales. This is likely to be a suboptimal assumption (e.g., Boer 2004; Pohlmann et al. 2004).

Equation (4) constitutes a valid hindcast in the sense that it depends only on the evolution of the forcing specified to the climate model prior to and during the forecast decade  $t$ . This evolution is known from historical observations, albeit with uncertainty (Ramaswamy et al. 2001) to the end of the twentieth century, and subsequently can be specified from a forcing scenario, such as one from the Intergovernmental Panel on Cli-

mate Change (IPCC) Special Report on Emissions Scenarios (SRES; Nakicenovic et al. 2000). As noted above, the forecast for the 2000 or 2010 decade is not likely to be sensitive to the specific choice of the forcing scenario that is used to extend the simulations of the twentieth century into the future.

The point-value forecasts obtained in this way can be extended to probabilistic forecasts by noting (i) that the simple forecasting model described above is, in fact, the same statistical model that is used for climate change detection; and (ii) that this model can be given a Bayesian interpretation if we consider its parameters to be random variables that have their own statistical distributions (e.g., Lee et al. 2005).

### b. Bayesian extension and forecast updating procedure

Following Berliner et al. (2000) and Lee et al. (2005), we will assume that  $\beta_t$  has a Gaussian probability density function  $\phi(m_t, c_t)$  with mean  $m_t$  and variance  $c_t$ . The value of  $m_t$  and  $c_t$  will be chosen according to our subjective knowledge on  $\beta_t$  and information from past observations. Further, we assume that  $\varepsilon_t$  is independent of  $\beta_t$  and that  $\varepsilon_t$  has a multivariate Gaussian probability density function  $\phi_n(\mathbf{0}, \mathbf{\Sigma})$  of dimension  $n$ . With these assumptions, it follows from the forecast model (1) that the  $n$  dimensional hindcast distribution of  $\mathbf{Y}_t$ , conditional on the model simulated response anomaly  $\mathbf{X}_t$ , is given by

$$f(\mathbf{Y}_t|\mathbf{X}_t) = \phi_n(m_t\mathbf{X}_t, c_t\mathbf{X}_t\mathbf{X}_t^T + \mathbf{\Sigma}). \quad (5)$$

As discussed in Lee et al. (2005), the distribution on  $\beta_t$  can be chosen to reflect prior knowledge about the presence of the simulated response anomaly in the observations and the uncertainty of the response.

Having constructed this multivariate distribution, the hindcast distribution at point  $i$  is easily obtained by calculating the marginal distribution for  $\mathbf{y}_{it}$ , which is given by

$$f(\mathbf{y}_{it}|\mathbf{X}_t) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(\mathbf{Y}_t|\mathbf{X}_t) d\mathbf{y}_{t1} d\mathbf{y}_{t2} \dots d\mathbf{y}_{t(i-1)} d\mathbf{y}_{t(i+1)} \dots d\mathbf{y}_{tm} = \phi(m_t\mathbf{x}_{it}, c_t\mathbf{x}_{it}^2 + \mathbf{\Sigma}_{ii}). \quad (6)$$

Point and interval forecasts for point  $i$  (i.e., confidence intervals for the forecast temperature change in decade  $t$  relative to the current three-decade climatology) are easily obtained from this distribution. The probability hindcast for a particular event  $\mathbf{E}$  for point  $i$  is also easily obtained by integrating  $f(\mathbf{y}_{it}|\mathbf{X}_t)$  over the hindcast event of interest. That is, the hindcast probability of event  $\mathbf{E}$  at point  $i$  can be obtained by computing

$$\int_{\mathbf{E}} f(\mathbf{y}_{it}|\mathbf{X}_t) d\mathbf{y}_{it}. \quad (7)$$

Typically, we would take  $\mathbf{E}$  to be an event such as the occurrence of an above-normal decadal mean temperature where “normal” is defined as the operational three-decade climatological mean that is current for the decade for which the hindcast is issued.

Hindcast distributions for other quantities such as the global mean  $\mathbf{y}_t = w_t^T \mathbf{Y}_t$  are easily obtained as

$$f(\mathbf{y}_t | \mathbf{X}_t) = \phi(m_t w_t^T \mathbf{X}_t, c_t w_t^T \mathbf{X}_t \mathbf{X}_t^T w_t + w_t^T \Sigma w_t), \quad (8)$$

where, in the case of the global mean,  $w_t$  is a vector of weights proportional to area. Note that the dot in place of a subscript indicates that (weighted) averaging has been performed over that subscript.

Once a hindcast has been produced for an initial decade  $t$  and verified against observations for that period, hindcasts (or forecasts as the case may be) can be produced for subsequent decades by means of a simple Bayesian updating procedure. In particular, the hindcast for decade  $t + 10$  is obtained by deriving a probability distribution for  $\beta_{t+10}$  that is based on the outcome for decade  $t$ . The analysis that produced the initial forecast was based on observations prior to decade  $t$ , a simulation of the response to external forcing on the climate system through to the end of decade  $t$ , and a prior distribution  $\phi(m_t, c_t)$  on the scaling factor  $\beta_t$  that reflects knowledge about the true value of  $\beta_t$  before observing  $\mathbf{Y}_t$ . The latter also accounts for various sources of uncertainty (Lee et al. 2005). Once the observations for the initial decade  $t$  become available, knowledge regarding the scaling factor can be updated by calculating the posterior distribution on  $\beta_t$ . According to Bayes's theorem (e.g., West and Harrison 1997) this distribution is given by

$$f(\beta_t | \mathbf{X}_t, \mathbf{Y}_t) = \phi(m_{t+10}, c_{t+10}), \quad (9)$$

where  $c_{t+10} = (1/c_t + \mathbf{X}_t^T \Sigma^{-1} \mathbf{X}_t)^{-1}$  and  $m_{t+10} = c_{t+10}(m_t/c_t + \mathbf{X}_t^T \Sigma^{-1} \mathbf{Y}_t)$ . This updated version of the distribution on  $\beta_t$ , which represents a combination of the prior information that was available regarding  $\beta_t$  and information that was subsequently extracted from the observations for decade  $t$ , can now be used as the prior distribution for  $\beta_{t+10}$  to generate the hindcast for decade  $t + 10$ . In another words, the hindcast for decade  $t + 10$  is derived by using the posterior distribution on  $\beta_t$  as the prior distribution for  $\beta_{t+10}$  to obtain the hindcast distribution  $f(\mathbf{Y}_{t+10} | \mathbf{X}_{t+10})$  for the climate state in decade  $t + 10$ . Once observations become available for decade  $t + 10$ , a further updated posterior distribution  $f(\beta_{t+10} | \mathbf{X}_{t+10}, \mathbf{Y}_{t+10})$  can then be calculated for making the hindcast in decade  $t + 20$ , etc. Thus such a posterior-prior updating process allows us to improve our knowledge, over time, regarding the scaling factor  $\beta$  that is required to best match the model-simulated decadal temperature increments with observed temperature increments.

### c. Climate change hindcast and skill evaluation

By defining our problem in terms of anomalies from a moving climatological base period, we can easily define hindcast events that are related to climate change. In this study we consider only a two-category forecast system, that is, we consider either an increase in temperature in decade  $t$  relative to the base period (above normal) or conversely, a decrease (below normal). In contrast, seasonal forecasting systems (e.g., O'Lenic 1994; Mason et al. 1999; Derome et al. 2001) typically provide three-category forecasts of the likelihood of above, near, and below normal. Such an extension to the present two-category system would not be difficult, but would increase uncertainty somewhat because the event boundaries, as well as the forecasts themselves, would then become dependent upon our estimates of the internal climate variability. Using a two-category system avoids this source of uncertainty by allowing us to define events relative only to the current operational base climatology. Thus, we calculate hindcast probabilities at point  $i$  by defining the event  $\mathbf{E}$  in Eq. (7) to be either  $[0, \infty)$  for above normal, or  $(-\infty, 0)$  for below normal. If we predict that there will be no climate change (relative to the base period), the probabilities for both events should be equal to 1/2. Otherwise, the probabilities would differ from 1/2, depending on the strength and sign of the effect of the forcing.

Once the probability hindcasts are generated, one can evaluate the skill of these hindcast for each decade  $t$  over the  $n$  spatial grid points contained in the observation vector  $\mathbf{Y}_t$ . Such an evaluation allows us to assess whether knowledge of forcing change during the decades leading up to decade  $t$  can be translated into usable forecast skill on decadal time scales. Such skill, if present, would also provide additional, and very practical, supporting evidence for the attribution of observed twentieth-century climate change to external forcing change.

A widely used verification statistic for probability forecasts is the Brier score (Brier 1950; Wilks 1995). The Brier score, evaluated over  $n$  forecasts, is given by

$$B = \frac{1}{n} \sum_{i=1}^n (p_i - q_i)^2, \quad (10)$$

where  $p_i$  is the forecast probability of an event  $\mathbf{E}$  at point  $i$  and  $q_i$  is an indicator variable that is set to 1 or 0 depending upon whether or not the event occurred. To assess the skill of the forecast,  $B$  is converted to the Brier skill score (BSS), which is defined as

$$\text{BSS} = 1 - B/B_{\text{cli}}, \quad (11)$$

where  $B_{\text{cli}}$  is the climatologically expected Brier score of the event. It is easily shown that  $B_{\text{cli}} = 0.25$  in the case of a system using two equally likely climatological categories. The BSS is equal to 1 for a perfect probability hindcast, 0 for a hindcast that performs the same as the no-climate-change hindcast and negative for a hindcast that performs worse than the no-climate-change hindcast.

#### d. Additional hindcasts

To investigate whether Bayesian procedure described above improves or diminishes forecast skill from the models, or whether indeed, the climate models contribute skill beyond a simple straw man approach to forecasting, we also consider three additional hindcast variants.

The first, called the raw model hindcast, is produced by not updating the mean of the prior distribution in the posterior-prior updating process. That is, we specify that  $m_t = 1$  at each time point  $t$  so that the mean of the hindcast distribution is the ensemble mean. However, we still allow the width of the prior to vary between time periods according to Eq. (9). The effect is that the variance of the forecast distribution [see (5)] is inflated by the factor  $c_t \mathbf{X}_t \mathbf{X}_t^T$ . This is roughly equivalent to the usual practice of adding a factor  $1/m$  to the forecast variance, where  $m$  is the ensemble size, to account for sampling variability in the ensemble mean. Note the details of the treatment of the prior variance have almost no influence on our results.

The second, called the blended hindcast, uses a prior distribution where the mean  $\omega \times m_t + (1 - \omega) \times 1$  is a blend of the posterior mean obtained from the updating process at time  $t - 10$ , and the mean  $m_t = 1$  that would be appropriate if the model always responded correctly to external forcing. The variance of the prior distribution continues to be updated as in Eq. (9). The weighting factor  $\omega$  can be varied between 1 and 0, with  $\omega = 1$  producing the Bayesian hindcast described previously, and  $\omega = 0$  producing the raw model hindcast described above. We used  $\omega = 0.5$ , thereby allowing the update process to learn partially from previous success by the model in reproducing observed large-scale climate variations, but also allowing for the possibility that previous performance may have a detrimental effect on future forecast skill and lead to underestimation of the scaling factor  $\beta_t$ . Underestimation of  $\beta_t$  in a given decade might occur for a variety of reasons. These would include (i) small ensemble sizes, which would lead to contamination of the model ensemble response  $\mathbf{X}_t$  by sampling errors and thus underestimation of  $\beta_t$ ; (ii) poor response to a short time scale forcing such as a volcanic event; or (iii) the occurrence of unusual natu-

ral internal variability, such as a strong El Niño event, during a given decade that one would not expect a model to reproduce.

A third, called the persistence hindcast, is produced by using  $\mathbf{Y}_{t-1}$  as  $\mathbf{X}_t$  and then generating the hindcast using the full Bayesian mechanism describe previously. We anticipate that the persistence hindcast will be difficult to beat. While it does not benefit from a sophisticated formulation of the anticipated response to external forcing, it does implicitly benefit from knowledge of the state of the climate system at the start of the forecast period, including the true response to external forcing up to that point. In addition, the Bayesian hindcasting process should be able to learn from aspects of internal climate variability that persist from one decade to the next.

### 3. Application

#### a. Data

The observational dataset used in this study is the same as that used in Lee et al. (2005), namely, the Hadley Centre–Climate Research Unit variance adjusted temperature dataset (HadCRUTv; Jones et al. 2001). This is a combined dataset of monthly surface air and sea temperature anomalies relative to 1961–90 on a  $5^\circ \times 5^\circ$  latitude–longitude grid for the period 1870–1999. The various versions of this dataset have been used extensively in previous climate change detection and attribution studies.

The climate simulations of the twentieth century used in this study are from the Canadian Centre for Climate Modeling and Analysis (CCCma) second-generation Coupled Model (CGCM2; Flato and Boer 2001), the Second and Third Hadley Centre Coupled GCMs (HadCM2 and HadCM3) and simulations from six models in the IPCC Fourth Assessment Report (AR4) model archives that are driven with estimates of historical forcings for the twentieth century [the Community Climate System Model version 3 (CCSM3.0), the Geophysical Fluid Dynamics Laboratory model versions 2 and 2.1 (GFDL2.0 and GFDL2.1), the Model for Interdisciplinary Research on Climate version 3.2 (MIROC3.2), the Meteorological Research Institute (MRI) and Parallel Climate Model (PCM)]. A summary of the simulations used in this study is displayed in Table 1. Ensemble sizes for individual models range from three to eight simulations of the twentieth century. Earlier ensembles available for this study include only anthropogenic forcing, with sulfate aerosol forcing limited to the direct effect in some instances, while the IPCC AR4 simulations all include anthropogenic and natural external forcing, and generally include indirect

TABLE 1. Summary of simulations used in this study. “Anthro” indicates that the simulation is driven by anthropogenic forcing consisting at least of greenhouse gas and direct sulfate aerosol forcing in the case of CGCM2 and HadCM2, but also including forcing from other sources, such as the indirect effects of sulphate aerosols, other nonsulfate aerosols, ozone, and land use in some of the more complete IPCC AR4 models. “Anthro+Nat” indicates that the simulation is also driven by reconstructions of historical natural forcings such as solar and volcanic. The control length column is filled in only when a control simulation is available from that model. Horizontal atmospheric resolution is indicated either by the model’s spectral resolution, or by gridbox size expressed in degrees of lat  $\times$  degrees of lon. Further details of the models are available online (CGCM2: <http://www.cccma.ec.gc.ca>; HadCM2 and HadCM3: <http://www.metoffice.com/research/hadleycentre/models/modeltypes.html>; IPCC AR4 models: [http://www-pcmdi.llnl.gov/ipcc/model\\_documentation/ipcc\\_model\\_documentation.php](http://www-pcmdi.llnl.gov/ipcc/model_documentation/ipcc_model_documentation.php)).

Model	Atmospheric resolution	Runs	Forcings	Control length (yr)
CGCM2	T32 $\times$ L10	3	Anthro	1000
HadCM2	2.5° $\times$ 3.75° $\times$ L19	4	Anthro	1019
HadCM3	2.5° $\times$ 3.75° $\times$ L19	4	Anthro	1640
CCSM3.0	T85 $\times$ L26	8	Anthro+Nat	
GFDL2.0	2.0° $\times$ 2.5° $\times$ L24	3	Anthro+Nat	
GFDL2.1	2.0° $\times$ 2.5° $\times$ L24	3	Anthro+Nat	
MIROC3.2	T42 $\times$ L20	3	Anthro+Nat	
MRI	T42 $\times$ L30	5	Anthro+Nat	
PCM	T42 $\times$ L26	4	Anthro+Nat	

aerosol effects. Three long control simulations from CGCM2, HadCM2, and HadCM3 are also used in this study. All simulations, which are available in a variety of grid sizes (Table 1), were interpolated onto the 5°  $\times$  5° grid of the observations and subsequently averaged into regional decadal means (details to be described below). An analysis is conducted for each individual model and for the ensemble mean of the simulations from the six IPCC AR4 models.

We conduct our analysis of decadal predictability on regional decadal means calculated over 30°  $\times$  40° latitude–longitude grid boxes as in Lee et al. (2005). Monthly means in 30°  $\times$  40° regions were calculated by averaging all available observed, or simulated, 5°  $\times$  5° monthly means in the region. Observed annual means are treated as missing if even 1 month within the year is missing. Decadal means are treated as missing if fewer than 6 of the 10 yr are present. The base period temperature is treated as missing only if all three decadal means are missing. To avoid systematic bias, missing data are not filled in. Instead, model output is flagged as missing whenever the corresponding observations are missing. In the absence of any missing data, the observational vector  $\mathbf{Y}_t$  in a given decade would have length  $n = 6 \times 9 = 54$ , where 54 is the number of

30°  $\times$  40° grid boxes that cover the globe. Missing data reduces the dimension length to a number ranging from 44 for the decade of the 1930s to 51 for the decade of the 1990s.

#### b. Covariance matrix estimation and dimension reduction

To generate our probability hindcasts, it is necessary to have an estimate of the natural internal variability of the surface temperature on the decadal and regional scales that are retained in the observation vector  $\mathbf{Y}_t$ . That is, we require an estimate of the variance–covariance matrix  $\mathbf{\Sigma}$  of the term  $\epsilon_t$  that appears in Eq. (1). As in our previous detection study (Lee et al. 2005), we estimate this matrix from long control simulations because the available instrumental record is not long enough to provide a reliable estimate of decadal-scale variability.

To avoid bias in our probability hindcasts, we produce two independent estimates of  $\mathbf{\Sigma}$ . The first estimate of  $\mathbf{\Sigma}$  is used in an empirical orthogonal function (EOF) dimension reduction step that retains only the large-scale variability in the observational vector  $\mathbf{Y}_t$ , while the second estimate provides an estimate of the internal variability covariance structure in the subspace that was determined from the first estimate.

The details of the covariance matrix estimation are as follows. Three control simulations are available in this study. To avoid bias when estimating the covariance matrix, we used only the last 1000 yr of HadCM2 and HadCM3 control simulations so that their length matches with that of the CGCM2 control simulation. Each control simulation is divided into two 500-yr subsets and each 500-yr control run subset is formed into 99 10-yr chunks, each overlapped by 5 yr. Decadal means are computed from these chunks and the prior three-decade mean is subtracted from each chunk. This results in 93 decadal anomalies from their respective prior three-decade climatologies within each 500-yr subset of each control run. This approach of calculating overlapping decadal anomalies provides somewhat more information for calculating the covariance matrix than would be available from only the 47 nonoverlapping decadal anomalies that can be computed from years 31–40, 41–50, . . . , 491–500 of the control run. By overlapping decades, an additional 46 decadal anomalies are obtained from years 36–45, 46–55, . . . , 486–495. A sample covariance matrix is calculated for each control run from the first collection of 93 anomalies using the standard formula. The average  $\hat{\mathbf{\Sigma}}$  of the three resulting covariance matrices is then used as an estimate of  $\mathbf{\Sigma}$ . A second estimate  $\hat{\mathbf{\Sigma}}$  is similarly calculated from the collections of 93 anomalies obtained from the second 500-yr segment of each of the three control runs.

Note that these calculations are repeated for each hindcast because the masking that reflects whether observations are missing varies from one hindcast period to the next.

The estimated covariance matrix  $\hat{\Sigma}$  is used to define a dimension reduction that retains only the large spatial scales of variation in decadal anomalies. This is done by projecting all observed and simulated decadal anomalies onto the  $k$  gravest EOFs of  $\hat{\Sigma}$ . The choice of  $k$  is determined by two criteria. First, the retained scales should well represent the model-simulated response anomaly  $\mathbf{X}_t$ . More importantly, the model simulated internal variability should be consistent with observed variability (Allen and Tett 1999) on these scales. Previous detection work at global and regional scales suggests that moderate values of  $k$  can be used. As in Zwiers and Zhang (2003) and Lee et al. (2005), we find that our results are not sensitive to the choice of  $k$  for  $5 \leq k \leq 25$ .

Given such a dimension reduction, a natural estimate of the variance structure of the internal variability in the reduced space is  $\mathbf{P}^{(k)T} \tilde{\Sigma} \mathbf{P}^{(k)}$ , where the columns of  $\mathbf{P}^{(k)}$  are the first  $k$  EOFs of  $\hat{\Sigma}$ . We use  $\tilde{\Sigma}$  rather than  $\hat{\Sigma}$  to estimate the internal variability in the reduced space to avoid biases that would creep into the analysis from using only one control run sample to estimate both the EOFs and the variability in the reduced space. Such biases arise because the EOF basis vectors inevitably “adapt” to the specific variations present in the part of the control simulation from which they are estimated (Allen and Tett 1999; Allen and Stott 2003).

The variance and mean of the posterior distribution of the Bayesian analysis in the reduced space can be estimated using  $\tilde{\Sigma}$  as

$$\begin{aligned} c_{t+10} &= [1/c_t + \mathbf{X}_t^T \mathbf{P}^{(k)} (\mathbf{P}^{(k)T} \tilde{\Sigma} \mathbf{P}^{(k)})^{-1} \mathbf{P}^{(k)T} \mathbf{X}_t]^{-1} \\ m_{t+10} &= c_{t+10} \{ [\omega m_t + (1 - \omega)] c_t \\ &\quad + \mathbf{X}_t^T \mathbf{P}^{(k)} (\mathbf{P}^{(k)T} \tilde{\Sigma} \mathbf{P}^{(k)})^{-1} \mathbf{P}^{(k)T} \mathbf{Y}_t \}. \end{aligned} \quad (12)$$

Similarly, the hindcasting distribution [cf. Eq. (5)] becomes

$$\begin{aligned} f(\mathbf{Y}_t | \mathbf{X}_t) &= \phi_n([\omega m_t + (1 - \omega)] \mathbf{P}^{(k)} \mathbf{P}^{(k)T} \mathbf{X}_t, \\ &\quad c_t \mathbf{P}^{(k)} \mathbf{P}^{(k)T} \mathbf{X}_t \mathbf{X}_t^T \mathbf{P}^{(k)} \mathbf{P}^{(k)T} + \tilde{\Sigma}). \end{aligned} \quad (13)$$

Recall that  $\omega = 1$  for the full Bayesian hindcast,  $\omega = 0.5$  for the blended hindcast, and  $\omega = 0$  for the raw hindcast. Also, as noted above, the matrices  $\tilde{\Sigma}$  and  $\hat{\Sigma}$  vary slightly from one hindcast period to the next because the masking of the observations varies. Thus, the individual EOFs that are used for the dimension reduction also vary somewhat from one hindcast period to the next.

### c. Determining the significance of the BSS

The BSS that we use to evaluate our forecasts is affected by the specific realization  $\varepsilon_t$  of the climate’s internal variability that is present during the hindcast period. Thus, one would expect the BSS of an unskilled hindcast or forecast to vary about zero as a result of sampling variability. It is therefore necessary to construct an upper critical bound for the BSS in order to identify a skill threshold above which one can reject the null hypothesis that the forecast is not skillful. Such a critical level can be estimated by verifying our hindcast against decadal anomalies calculated from the three available 1000-yr control simulations. A total of  $97 \times 3 = 291$  such anomalies can be obtained by using non-overlapping 10-yr chunks. The resulting sample of BSS’s reflects the range of skill scores one would obtain if the verification dataset consists of only internal climate variability. An upper 5% critical level for the BSS is then easily estimated by calculating the 95th percentile of the sample of 291 BSS. This critical value varies slightly from one decade to the next because the observational masking changes in time.

### d. Hindcast results

Temperature anomaly hindcasts for each model and the AR4 ensemble were produced using the methods described above for seven decades (1930–39, 1940–49, . . . , 1990–99). In addition, we also produced a forecast for the decade 2000–09 using the CGCM2 model that will be discussed in the concluding section of this paper.

To produce our first hindcast for the decade 1930–39, it was necessary to specify a prior distribution on the scaling factor  $\beta_{1930}$  that appears in Eqs. (1) and (4). That is, we chose values for the parameters  $m_{1930}$  and  $c_{1930}$ . Prior distributions for the subsequent hindcasts were then obtained by using the posterior-prior updating process described above. We subjectively chose the prior distribution on  $\beta_{1930}$  to be  $\phi(1, 0.25)$ , thereby producing an initial four standard deviation uncertainty on the scaling factor that ranges from 0 to +2. Lee et al. (2005) discuss the suitability of this choice in the context of the uncertainties that affect the climate model-simulated external forcing response anomalies  $\mathbf{X}_t$ . The robustness of our results to the initial choice of prior is discussed in the latter part of this section.

Figure 1 displays the BSS for the above-normal event as a function of decade with 15 EOFs retained for the full Bayesian hindcast (with  $\omega = 1$ ), together with corresponding 5% critical values for rejecting the null hypothesis of an unskillful hindcast for the CGCM2 hindcast (thick horizontal line segments). The critical values for other models are very similar and thus not shown.

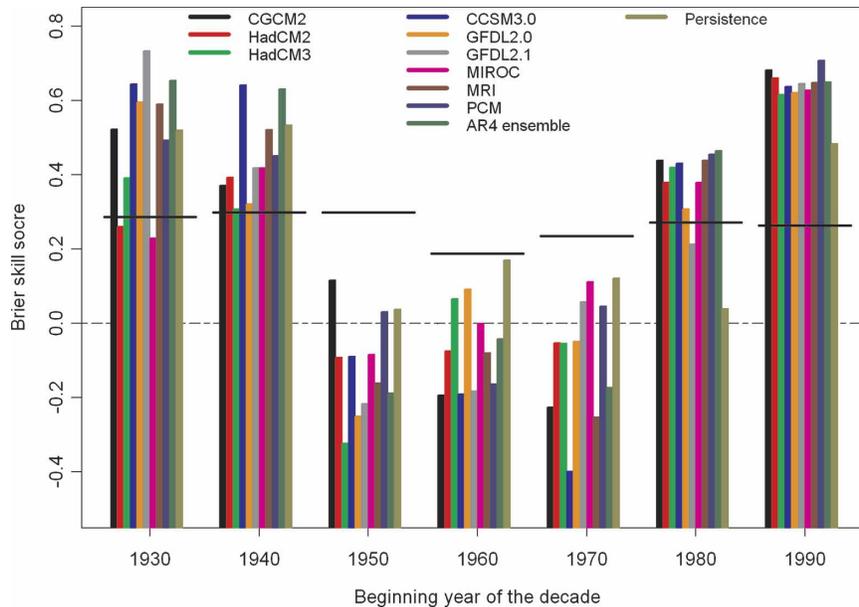


FIG. 1. BSSs for the above-normal event with 15 EOFs retained. The thick horizontal line segments indicate the critical skill threshold for rejecting the no-skill null hypothesis. Skill scores are shown only for the full Bayesian analysis with  $\omega = 1$ . Results for the blended and raw hindcasts are very similar.

We have also repeated our analysis by retaining 5, 10, 20, and 25 EOFs (not shown) and find that the BSS is not very sensitive to the number of EOFs retained. With 15 EOFs retained, the BSS for CGCM2 lies above the critical value for the decades 1930–39, 1940–49, 1980–89, and 1990–99, suggesting that the temperature anomaly hindcasts for those decades may have significant skill. Specifically, the BSSs for these periods are 0.523, 0.373, 0.439, and 0.684, respectively. Similar results are obtained for HadCM2 and HadCM3, for the hindcasts produced from the AR4 model simulations and for their ensemble mean, where the BSS is above the critical value for the early and late decades of the twentieth century. This suggests that the inclusion of natural forcing in the simulations is not the solution to the apparent lack of skill in the 1950s–70s.

The skill obtained for the blended and raw hindcasts (not shown) is very similar to that obtained with the full Bayesian hindcast, with minor variations in skill (both increases and decreases) depending upon the model that is used. This indicates that variation in the details of the prior updating process does not have a large influence on forecast skill, at least as measured by the BSS. This is a reasonable result given that the BSS measures the agreement between the spatial pattern of the hindcasts of above-normal probability, and the verifying pattern of above-normal events. The prior updating process would have some influence on the ampli-

tude of the pattern of hindcast probabilities, but it does not affect its shape, and thus has little influence on the BSSs.

A concern with respect to the skill scores is that they may be sensitive to the choice of prior distribution on the 1930s scaling factor  $\beta_{1930}$ . Thus, we also conducted our full Bayesian analysis using two other classes of priors to evaluate that possibility. The first type of prior is identical to that used above except that the initial mean ( $m_{1930}$ ) was varied between  $-1$  and  $3$ . The second type of prior has  $m_{1930} = 1$  but uses variances  $c_{1930}$  that range from  $0.1$  to  $1.1$ . The BSS for CGCM2 obtained by using these priors ranges from  $-1.09$  to  $0.59$  for the decade 1930–39 and from  $0.184$  to  $0.460$  for the decade 1940–49. However, BSSs in subsequent decades, after the Bayesian updating process commences, are very similar to those shown in Fig. 1. The hindcast probability, hindcast anomaly, and posterior distribution were found to be insensitive to the initial choice of prior after the first three decades, with only a minor impact on the third decade. Similar behavior was also exhibited by the other models that we have considered here. The robustness of our results after the initial two to three decades is mainly due to the calibration of the distribution of  $\beta_i$  from the posterior-prior updating process. While we will continue to show results for all decades in the remainder of this paper based on our initial choice of prior, the sensitivity to the choice of prior during the

first two hindcast decades indicates that results for those two decades should be downweighted.

Figure 2 displays the global mean hindcast probability for the above-normal event as a function of decade with 15 EOFs retained together with the observed proportion of such events and its confidence bound. An approximate 95% confidence bound for the observed proportion  $\hat{p}^o$  can be defined as  $\hat{p}^o \pm 2\sqrt{\hat{p}^o(1-\hat{p}^o)/n}$ , where  $n$  is the number of spatial points in the analysis. This bound accounts for sampling variations in the observed proportion of above-normal events that would be expected under similar conditions, and under the assumption of spatial independence. The actual confidence bound is likely to be wider because of dependence between the regions. Figure 2a shows that the CGCM2 hindcast probabilities significantly underestimate the proportion of observed above-normal events for the decades 1930–39 and 1940–49, but are within the uncertainty range for 1950s, 1960s, 1980s, and 1990s. Considering all 10 model hindcasts together, we see that the hindcast performance is generally poor in the 1930s and 1940s, but starting from the 1950s, a majority of the models “correctly” hindcast the observed proportion in each decade. As noted previously, hindcasts of the first two decades are sensitive to the choice of initial prior, and thus these results should be discounted. Also, as with the results for the BSS, the details of the prior updating process have a relatively small effect, although there is some evidence (cf. Figs. 2b,c with Fig. 2a) that either the blended or raw hindcast performs slightly better than the full Bayesian hindcast, perhaps because the latter allows the scaling factor  $\beta_i$  to be too heavily influenced by past forecast errors.

Figure 3 shows the corresponding hindcasts as derived from Eq. (8) for the actual global mean decadal temperature anomalies together with their corresponding 5%–95% hindcast confidence intervals and the observed anomalies. This graph also provides a forecast for the 2000–09 decade using the CGCM2 model. The information provided is essentially the same as that provided in Fig. 2. In particular, we see that there is good agreement between the observed and hindcast anomalies during the last four decades of the twentieth century. Specifically, for the full Bayesian procedure, the CGCM2, HadCM3, CCSM3.0, MRI, and PCM hindcasts were able to capture the observed anomalies throughout 1960s–90s. In addition, the AR4 ensemble hindcast was able to capture six out of seven observed anomalies. In contrast, the hindcast anomalies of the earlier skillful hindcasts, for the decade of the 1930s and 1940s, were less promising for all the models, except for the CCSM3.0 and AR4 ensembles. Based on the antic-

ipated response to anthropogenic forcing and assuming that there will not be a substantial change in natural external forcing, the global mean temperature anomaly for the current decade (2000–09) is predicted to be 0.35°C with a 5%–95% confidence range 0.21° to 0.48°C using the CGCM2 model.

Again, as above, there is some evidence that the blended or raw hindcast (Figs. 3b,c) has a slightly better performance than the full Bayesian hindcast (Fig. 3a). The AR4 ensemble, for example, is able to capture all of the observed anomalies in the former two cases. There is some evidence that the full Bayesian hindcast overly constrains the model forecasts (cf., e.g., the 1990s in Fig. 3a with the same period in Figs. 3b,c), suggesting that the scaling factor  $\beta_i$  has been underestimated. However, the blended and raw hindcasts also impose weaker constraints on models that warm quickly, such as CGCM2, with the result that greater warming is forecast for the first decade of the third millennium. In particular, the global mean temperature anomaly for 2000–09 is predicted to be 0.48°C (5%–95% confidence range: 0.34°–0.61°C) for the blended hindcast and 0.52°C (5%–95% confidence range: 0.39°–0.66°C) for the raw hindcast using the CGCM2 model.

Figure 4 displays 5%–95% for the posterior distribution of the scaling factor  $\beta_i$  for the full Bayesian hindcast as in Eq. (12) with 15 EOFs retained. The mean of the posterior distribution has a general downward trend for all the models. For example, for the CGCM2 model, 5%–95% of the posterior confidence interval lies below 1 for the last five decades. This may be partly due to negative bias in the estimate of  $\beta_i$  due to weak decadal-scale signals (see Allen and Stott 2003 for a discussion). However, it may also reflect an oversimulated response to twentieth-century forcing, perhaps because of missing forcings in the case of this model (Lee et al. 2005). In contrast, 5%–95% ranges for the AR4 ensemble include the possibility that  $\beta_i = 1$  for the 1960s, 1970s, 1980s, and 1990s, providing evidence toward attribution of natural and anthropogenic influence on climate for these four decades in the context of optimal fingerprinting. Bayesian climate change assessment focusing on the use of the posterior distribution of the scaling factor has been studied by Lee et al. (2005) and Berliner et al. (2000). Note that the  $\beta_i$  for the AR4 ensemble is larger than those obtained using the individual AR4 models. This is possibly due to the reduction of sampling uncertainty in the simulated response pattern for the AR4 ensemble hindcast as more ensemble members are used. As would be expected, the scaling factors obtained for the other two hindcasts (not shown) are much more tightly constrained to  $\beta_i = 1$ .

Even though our hindcasts are skillful for part of the

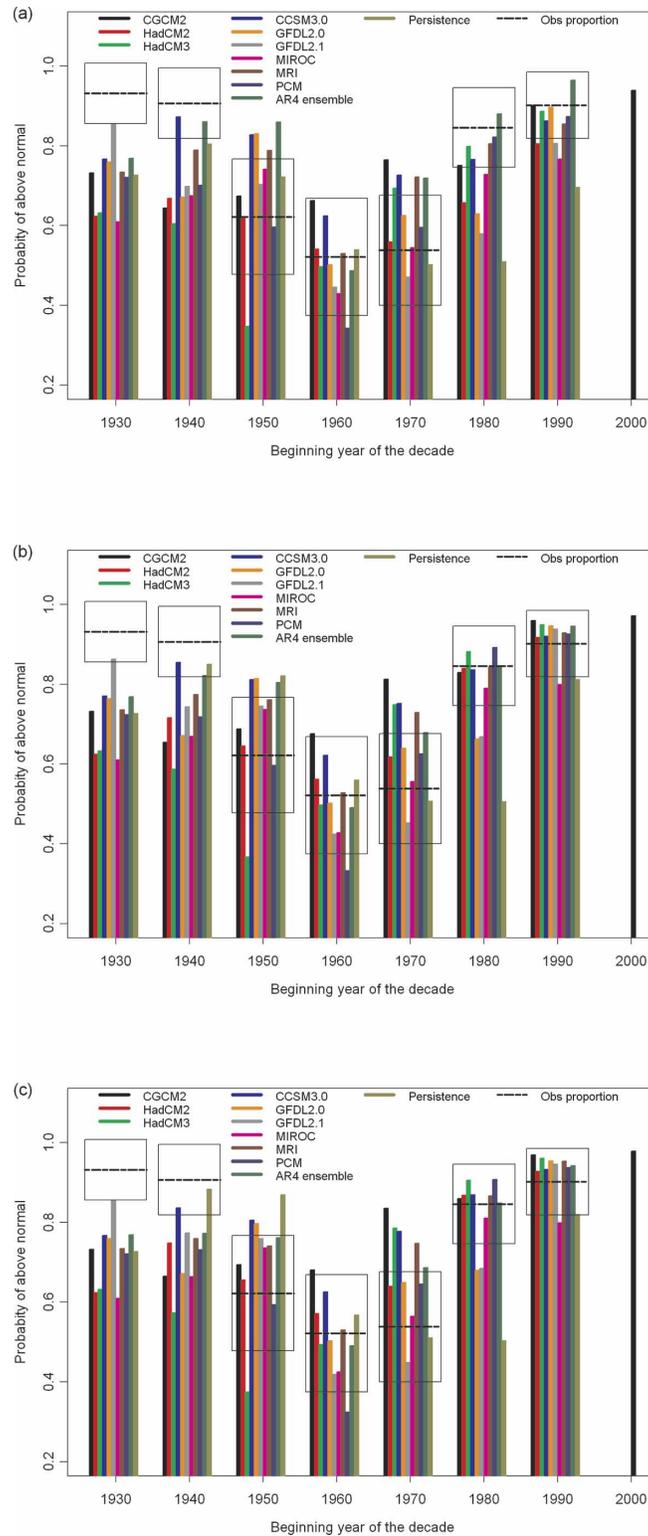


FIG. 2. Global mean hindcast probabilities of above-normal decadal mean temperatures in  $30^{\circ}$  lat  $\times$   $40^{\circ}$  lon regions (bars) and the observed proportion of regions with above-normal temperatures (thick horizontal dashed line segments) together with its estimated uncertainty (box). The probability scale is indicated on the left. A regional decadal mean temperature is considered to be above normal in a given decade when it is greater than the corresponding climatological mean temperature in the region for the preceding three decades. (a) Full Bayesian hindcast, (b) blended hindcast, and (c) raw hindcast.

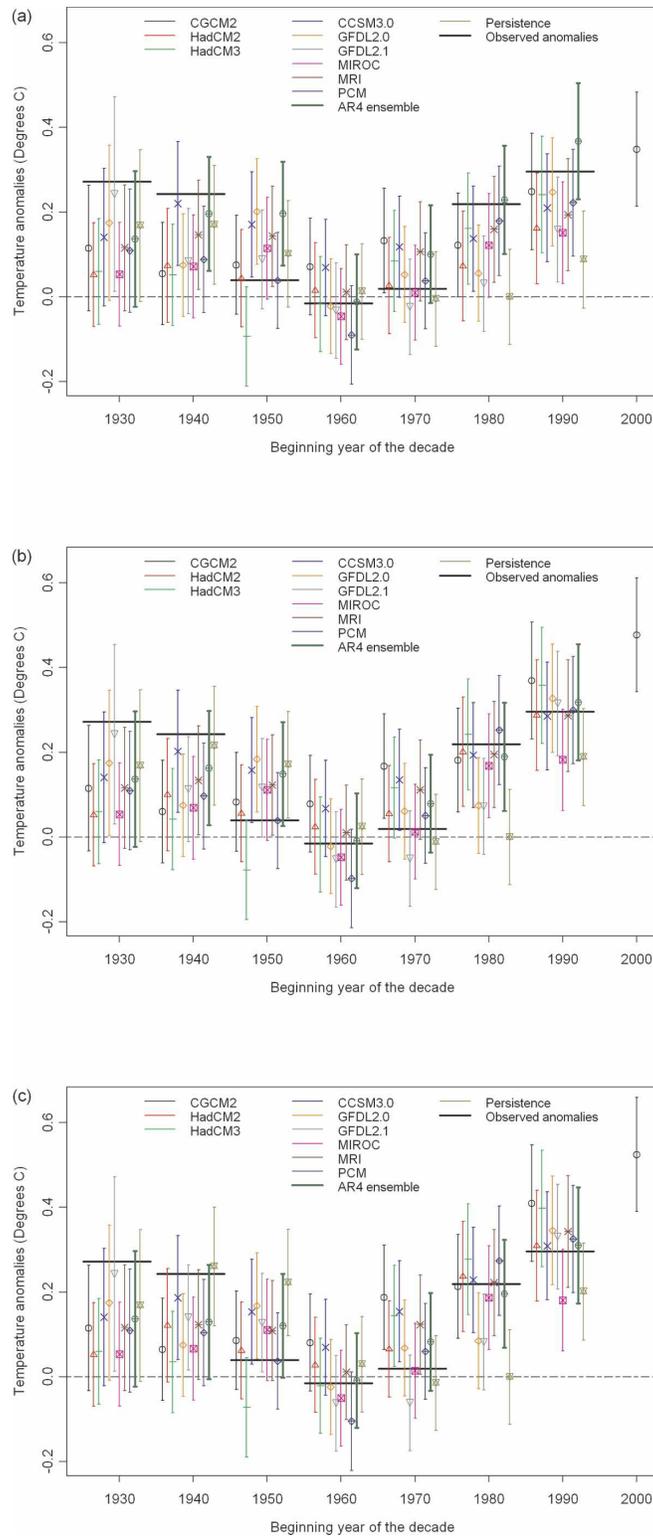


FIG. 3. Hindcasts of global decadal mean surface temperature anomalies and their 5%–95% confidence bounds based on Eq. (8), together with the observed global mean anomalies (thick horizontal line segments), with 15 EOFs retained. A forecast for the decadal global mean anomaly for the decade 2000–09, relative to the 1970–99 climatology, is also displayed ( $^{\circ}\text{C}$ ). (a) Full Bayesian hindcast, (b) blended hindcast, and (c) raw hindcast.

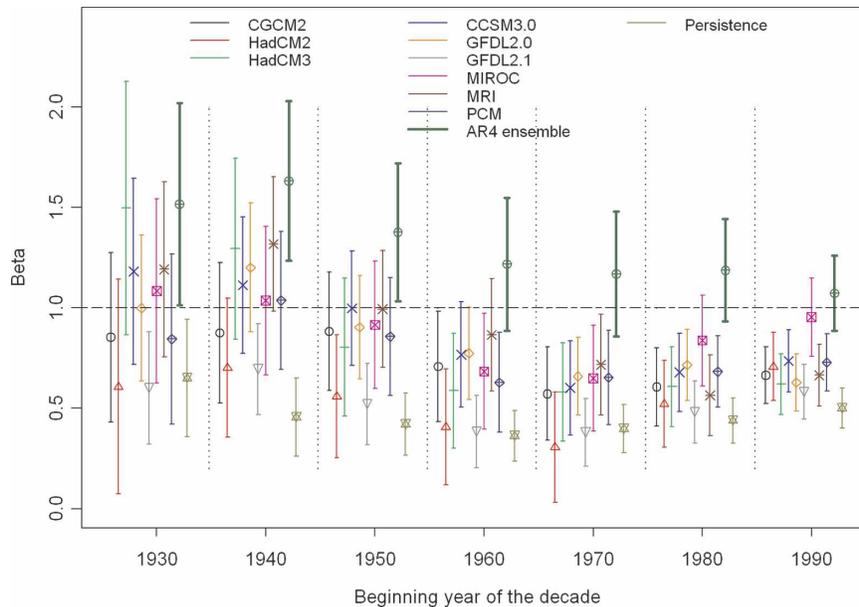


FIG. 4. Mean and 5%–95% of the posterior distribution of the scaling factor  $\beta$ , from the full Bayesian hindcast with 15 EOFs retained.

twentieth century, there remains the question of whether this reflects the capability of the climate models to predict the decadal response to external forcing, or whether the skill is really just an artifact arising from natural persistence. Such persistence might arise from two sources—either low-frequency natural internal variability or a persistent forcing disequilibrium resulting in a continuing response of the climate system to that disequilibrium.

The former possibility is, in fact, taken into account in the estimation of the 5% critical value for the BSS (see section 3c). BSSs for the decades of the 1980s and 1990s (Fig. 1), which are not affected by the choice of initial prior, are significantly greater than the estimated critical value, suggesting an external, rather than internal source of skill. This assessment is, of course, subject to the caveat that the control runs used to estimate these critical values are assumed to correctly simulate the natural internal low-frequency variability of the climate system on the space and time scales retained in the hindcasting analysis.

The latter possibility is considered by evaluating the performance of the persistence forecast, the results of which are indicated by the olive green bars in Figs. 1–4. As discussed above, persistence is a very tough forecast to beat at decadal leads given the apparent linearity of the response to forcing (e.g., Gillett et al. 2004b) and the fact that the response during any given decade is more reflective of a continuing response to historical forcing, than to forcing change during the hindcast de-

cade. The BSSs for the persistence hindcast are similar to those obtained using climate model hindcasts, except in the last two decades, where the BSS is significantly lower for the persistence hindcast (Fig. 1). This holds regardless of whether one uses the full Bayesian hindcast procedure (Fig. 1) or the blended or raw hindcast procedures (not shown). Similarly, the persistence hindcast underhindcasts the global mean hindcast probability of above normal during the last two decades, regardless of the hindcast procedure used (Fig. 2), and it underpredicts the global mean temperature anomaly (Fig. 3). The latter problem is particularly evident in the full Bayesian hindcast (Fig. 3a), perhaps because the “signal” used in persistence hindcast is heavily contaminated by noise from internal variability (which leads to negative bias in the estimate of  $\beta_i$ ). Thus, despite the expectation that the persistence hindcasts would be difficult to beat, it would appear that the model-based hindcasts outperform the persistence hindcast during the last two decades when anthropogenic forcing is largest.

#### 4. Conclusions

In this paper, we have put forward another approach to climate change detection analysis that is based on the skill of probabilistic decadal hindcasts that are produced from simulations of the climate of the twentieth century with a Bayesian technique. Specifically, we consider hindcasts of decadal temperature anomalies on

large spatial scales relative to the three-decade operational climatologies that are current at the time of the hindcast. Consistent with other detection studies, our Bayesian analysis indicates that the combined effect of greenhouse gas and sulfate aerosols is detectable in the latter part of the twentieth century. Statistical characteristics of the hindcasts, such as the global mean hindcast of the probability of above normal and the hindcast of the global mean temperature, are consistent with the characteristics of the verifying observations from the 1950s onward. The BSSs indicate that the model-based hindcasts demonstrate significant skill during the last two decades of the twentieth century. Comparison between the model-based hindcast and a persistence hindcast suggest that the models add value during this period relative to simple persistence. There is also some evidence of decadal hindcast skill during the first two decades considered, those of the 1930s and 1940s, but we consider those results to be somewhat more tenuous because these results are sensitive to the initial choice of prior distribution. As in other studies (Derome et al. 2001; Kharin and Zwiers 2002; Gillett et al. 2002; Zhang et al. 2006), there is also some evidence that the ensemble model mean approach performs more consistently than do individual models. The inclusion of natural external forcing does not appear to significantly improve short-term (i.e., decadal) hindcast skill, perhaps because the response associated with natural forcing is small relative to that associated with anthropogenic forcing.

Further work will be required to more clearly identify the factors that contribute to skill on the decadal time scale, to make more sophisticated use of multimodel ensembles as in seasonal forecasting (e.g., Derome et al. 2001; Kharin and Zwiers 2002), and by assimilating observed ocean, and perhaps land surface, state information into the model. A multisignal analysis that attempts to tease out the effect on skill of the different external forcing factors may be feasible, but requires the specification of a multivariate prior distribution. The corresponding hindcast distribution and posterior distribution can then be derived from Eq. (1) by using Bayes's theorem based on the prior distribution.

Forecasts for events in the future can be generated using the same methodology. However, one cannot carry out the posterior-prior updating process since observations are not available. Also, simulations of the twentieth century must then be extended into the future using a scenario of future emissions. While the simulated response one–two decades into the future is not likely to be sensitive to scenario details (e.g., Zwiers 2002), there are forcing uncertainties, such as the possibility of unforeseen volcanic activity, that must be

taken into account. Using the CGCM2 simulated anthropogenic signal as  $\mathbf{X}_t$  and using the posterior distribution from 1990 to 1999 as the prior distribution, we predict that in the absence of large negative volcanic forcing on the climate system (which cannot presently be forecast) the global mean temperature anomaly for the decade 2000–09 will be above the 1970–99 normal with a probability of 0.94. The global mean temperature increment for this decade is correspondingly predicted to be 0.35°C with a 5%–95% confidence range 0.21°–0.48°C (Fig. 3a). The suggestion that such decadal forecasts are now apparently skillful on large regional scales based only on anthropogenic forcing, and that they can be regularly updated and verified, provides additional evidence for the influence that anthropogenic forcing is having on our climate.

*Acknowledgments.* We thank Tim Palmer for the discussion on the links between prediction and detection that prompted us to undertake this project. We gratefully acknowledge that Terry Lee was supported by the Canadian Foundation for Climate and Atmospheric Science through the Canadian CLIVAR Research Network. Work by Min Tsao was supported by the Natural Sciences and Engineering Research Council through a Discovery Grant. We also gratefully acknowledge two anonymous reviewers whose comments helped to improve the paper.

#### REFERENCES

- Allen, M. R., and S. F. B. Tett, 1999: Checking for model consistency in optimal fingerprinting. *Climate Dyn.*, **15**, 419–434.
- , and P. A. Stott, 2003: Estimating signal amplitudes in optimal fingerprinting. Part I: Theory. *Climate Dyn.*, **21**, 477–491.
- Berliner, L. M., R. A. Levine, and D. J. Shea, 2000: Bayesian climate change assessment. *J. Climate*, **13**, 3805–3820.
- Boer, G. B., 2004: Long time-scale potential predictability in an ensemble of climate models. *Climate Dyn.*, **23**, 29–44.
- Braganza, K., D. J. Karoly, A. C. Hirst, P. Stott, R. J. Stouffer, and S. F. B. Tett, 2004: Simple indices of global climate variability and change. Part II: Attribution of climate change during the twentieth century. *Climate Dyn.*, **22**, 823–838.
- Brier, G. W., 1950: Verification of forecasts expressed in terms of probabilities. *Mon. Wea. Rev.*, **78**, 1–3.
- Collins, M., and M. R. Allen, 2002: Assessing the relative roles of initial and boundary conditions in interannual to decadal climate predictability. *J. Climate*, **15**, 3104–3109.
- Derome, J., and Coauthors, 2001: Seasonal predictions based on two dynamical models. *Atmos.–Ocean*, **39**, 485–501.
- Flato, G. M., and G. J. Boer, 2001: Warming asymmetry in climate change simulations. *Geophys. Res. Lett.*, **28**, 195–198.
- 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.
- , A. J. Weaver, F. W. Zwiers, and M. D. Flannigan, 2004a: Detecting the effect of human induced climate change on

- Canadian forest fires. *Geophys. Res. Lett.*, **31**, L18211, doi:10.1029/2004GL020876.
- , M. F. Wehner, S. F. B. Tett, and A. J. Weaver, 2004b: Testing the linearity of the response to combined greenhouse gas and sulfate aerosol forcing. *Geophys. Res. Lett.*, **31**, L14201, doi:10.1029/2004GL020111.
- Grotzner, A., M. Latif, A. Timmermann, and R. Voss, 1999: Interannual to decadal predictability in a coupled ocean-atmosphere general circulation model. *J. Climate*, **12**, 2607–2624.
- International Ad Hoc Detection and Attribution Group, 2005: Detecting and attributing external influences on the climate system: A review of recent advances. *J. Climate*, **18**, 1291–1314.
- Jones, P. D., T. J. Osborn, K. R. Briffa, C. K. Folland, E. B. Horton, L. V. Alexander, D. E. Parker, and N. A. Rayner, 2001: Adjusting for sampling density in grid box land and ocean surface temperature time series. *J. Geophys. Res.*, **106**, 3371–3380.
- Karoly, D. J., and Q. Wu, 2005: Detection of regional surface temperature trends. *J. Climate*, **18**, 4337–4343.
- Kharin, V. V., and F. Zwiers, 2002: Climate predictions with multimodel ensembles. *J. Climate*, **15**, 793–799.
- Lee, T. C. K., F. Zwiers, X. Zhang, G. Hegerl, and M. Tsao, 2005: A Bayesian approach to climate change detection and attribution. *J. Climate*, **18**, 2429–2440.
- Mason, S. J., L. Goddard, N. E. Graham, E. Yulaeva, I. Sun, and P. A. Arkin, 1999: The IRI seasonal climate prediction system and the 1997/98 El Niño event. *Bull. Amer. Meteor. Soc.*, **80**, 1853–1873.
- Min, S.-K., A. Hense, H. Paeth, and W.-T. Kwon, 2004: A Bayesian decision method for climate change signal analysis. *Meteor. Z.*, **13**, 421–436.
- Mitchell, J. F. B., D. J. Karoly, G. C. Hegerl, F. W. Zwiers, M. R. Allen, and J. Marengo, 2001: Detection of climate change and attribution of causes. *Climate Change 2001: The Scientific Basis*, Cambridge University Press, J. T. Houghton et al., Eds., 695–738.
- Nakicenovic, N., and Coauthors, 2000: *IPCC Special Report on Emissions Scenarios*. Cambridge University Press, 599 pp.
- O'Lenic, E., 1994: Operational long-lead forecast for the climate outlook. Tech. Procedures Bull. 418, NOAA/NWS/CPC, 30 pp. [Available from NOAA/CPC, 5200 Auth Rd., Camp Springs, MD 20746.]
- Pohlmann, H., M. Botzet, M. Latif, A. Roesch, M. Wild, and P. Tschuck, 2004: Estimating the decadal predictability of a coupled AOGCM. *J. Climate*, **17**, 4463–4472.
- Ramaswamy, V., and Coauthors, 2001: Radiative forcing of climate change. *Climate Change 2001: The Scientific Basis*, J. T. Houghton et al., Eds., Cambridge University Press, 349–416.
- Schnur, R., and K. Hasselmann, 2005: Optimal filtering for Bayesian detection and attribution of climate change. *Climate Dyn.*, **24**, 45–55.
- Stott, P. A., 2003: Attribution of regional-scale temperature changes to anthropogenic and natural causes. *Geophys. Res. Lett.*, **30**, 1728, doi:10.1029/2003GL017324.
- West, M., and J. Harrison, 1997: *Bayesian Forecasting and Dynamic Models*. Springer-Verlag, 670 pp.
- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp.
- Zhang, X., F. W. Zwiers, and P. A. Stott, 2006: Multimodel multisignal climate change detection at regional scale. *J. Climate*, **19**, 4294–4307.
- Zwiers, F. W., 2002: The 20-year forecast. *Nature*, **416**, 690–691.
- , and X. Zhang, 2003: Towards regional scale climate change detection. *J. Climate*, **16**, 793–797.