

## Systematic Estimates of Initial-Value Decadal Predictability for Six AOGCMs

GRANT BRANSTATOR, HAIYAN TENG, AND GERALD A. MEEHL

*National Center for Atmospheric Research,\* Boulder, Colorado*

MASAHIDE KIMOTO

*Atmosphere and Ocean Research Institute, University of Tokyo, Kashiwa, Japan*

JEFF R. KNIGHT

*Met Office Hadley Centre, Exeter, United Kingdom*

MOJIB LATIF

*Leibniz Institute of Marine Sciences, University of Kiel, Kiel, Germany*

A. ROSATI

*Geophysical Fluid Dynamics Laboratory, Princeton University, Princeton, New Jersey*

(Manuscript received 25 April 2011, in final form 23 August 2011)

### ABSTRACT

Initial-value predictability measures the degree to which the initial state can influence predictions. In this paper, the initial-value predictability of six atmosphere–ocean general circulation models in the North Pacific and North Atlantic is quantified and contrasted by analyzing long control integrations with time invariant external conditions. Through the application of analog and multivariate linear regression methodologies, average predictability properties are estimated for forecasts initiated from every state on the control trajectories. For basinwide measures of predictability, the influence of the initial state tends to last for roughly a decade in both basins, but this limit varies widely among the models, especially in the North Atlantic. Within each basin, predictability varies regionally by as much as a factor of 10 for a given model, and the locations of highest predictability are different for each model. Model-to-model variations in predictability are also seen in the behavior of prominent intrinsic basin modes. Predictability is primarily determined by the mean of forecast distributions rather than the spread about the mean. Horizontal propagation plays a large role in the evolution of these signals and is therefore a key factor in differentiating the predictability of the various models.

### 1. Introduction

Techniques are being developed that enable forecasts of the evolution of climate over the next few decades to be initialized from the observed state. This new area of climate science is referred to as “decadal prediction”

(e.g., Meehl et al. 2009). However, the climate system, being chaotic, is sensitive to small perturbations to the initial state. Hence, the influence of any particular initial state on a forecast is discernable for only a finite time. When one considers how long an initial condition has a detectable impact, one is studying the initial-value predictability of the climate system. Our study focuses on the initial-value predictability properties of models used to make decadal predictions and how much those properties vary from one model to another. These attributes should guide the design of decadal prediction systems and the interpretation of the forecasts they produce and even help assess whether initializing these models with the observed state is potentially beneficial.

---

\* The National Center for Atmospheric Research is sponsored by the National Science Foundation.

---

*Corresponding author address:* Grant Branstator, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307.  
E-mail: branst@ucar.edu

Furthermore, model behavior provides an indication of the possible predictability properties of nature. Model-to-model variations in predictability represent the uncertainty in these model-based estimates of this important attribute of nature.

Starting with Griffies and Bryan (1997b,a), various studies have estimated the predictability of different models used for decadal forecasts (Collins 2002; Collins and Sinha 2003; Pohlmann et al. 2004; M. Collins et al. 2006; Msadek et al. 2010; Meehl et al. 2010). Most studies have concentrated on the meridional overturning circulation of the Atlantic, and the predictability characteristics they have documented have apparently been highly model dependent. However, careful comparison of their results is difficult because different metrics have often been employed. Furthermore, each study has used different initial conditions, and it is known that predictability properties can depend on the initial state (Pohlmann et al. 2004; Msadek et al. 2010). A few investigations (e.g., M. Collins et al. 2006) have applied the same metrics to more than one model but for only a few initial states, as the ensemble technique that has been employed for estimating predictability is computationally expensive. Boer (2000, 2004) has analyzed the statistics of control runs to produce so-called “diagnostic potential predictability.” This quantity has the desirable attribute of taking into account many states, but its connection to initial-value predictability is not well established. Alternatively, others (e.g., Newman 2007; Teng and Branstator 2011) have fitted simple models to control run behavior and used them to infer predictability properties.

In this study we build on and expand these previous efforts by applying uniform measures and methods to six different atmosphere–ocean global climate models (AOGCMs). Furthermore, rather than the traditional ensemble approach, we make use of methodologies that enable us to directly quantify predictability properties from long control integrations. These methods include the analog methodology that has been employed in some studies of atmospheric predictability (Lorenz 1969) and the use of multivariate regression operators (DelSole and Tippett 2009b). These techniques enable us to consider the average predictability of each model, where “average” refers to averaging over forecasts beginning from each initial state in the long control trajectory. Furthermore, the methods that we employ make it possible to systematically compare the predictability of many models without coordinating large, perturbed ensemble experiments at various institutions.

The six AOGCMs that we consider and some basic properties of their intrinsic variability are provided in section 2. In section 3 we describe the measures and methods we use to infer predictability from control

TABLE 1. Description of the six models.

Model*	Forcing	Length (analyzed/total)	Reference
CCSM3	Present day	700/1000	W. D. Collins et al. (2006)
CCSM4	Preindustrial	700/1300	Gent et al. (2011)
GFDL CM2.1	Preindustrial	2500/3000	Delworth et al. (2006)
HadCM3	Preindustrial	5400/5900	Gordon et al. (2000)
KCM	Present day	4200/6700	Park and Latif (2010)
MIROC3.2	Preindustrial	3600/6000	Nozawa et al. (2007)

\* CCSM3: Community Climate System Model, version 3

CCSM4: CCSM, version 4

GFDL CM2.1: Geophysical Fluid Dynamics Laboratory Climate Model version 2.1

HadCM3: Third climate configuration of the Met Office Unified Model

KCM: Kiel Climate Model

MIROC3.2: Model for Interdisciplinary Research on Climate 3.2

integrations. Basin and mode predictability characteristics of the various models are compared in sections 4 and 5, respectively. Section 6 describes three factors that contribute to the substantial model-to-model variations in predictability that we find, while conclusions that can be derived from our findings are presented in section 7.

## 2. Models and their intrinsic variability

The six AOGCMs in our investigation include CCSM3, CCSM4, GFDL CM2.1, HadCM3, KCM and MIROC3.2. All of these models, except CCSM4 and KCM, have been thoroughly evaluated in the Fourth Assessment Report (AR4) of the Intergovernmental Panel on Climate Change (IPCC) (Randall et al. 2007). Descriptions of the models can be found in the references given in Table 1, which also includes pertinent information about model control runs. In these experiments the external forcing is fixed at either preindustrial or present-day levels. Note that for some models we do not use the early parts of the control runs when the modeled climate is not close to equilibrium.

In analyzing the predictability of these models, we focus on the annual mean depth-averaged ocean temperature between 300 m and the surface, which we refer to as T0–300. We choose T0–300 because it is somewhat shielded from weather noise, so its statistics are more robust than those for sea surface temperature—yet, unlike deeper layers, it has enough communication with the overlying atmosphere to potentially influence it on decadal time scales (Branstator and Teng 2010; Teng et al. 2011).

When examining T0–300, we use fields that have been interpolated to a common T42 grid. At times we focus on those aspects of variability that are represented by a

model's leading empirical orthogonal functions (EOFs) and corresponding principal components (PCs) for a given basin. In this way we concentrate on those characteristics that are well sampled and have large enough amplitudes to be of interest. When calculating the EOFs, we use the control run for each model and incorporate area weighting. For PC-based calculations we employ values that have been normalized by their climatological standard deviation.

Before quantifying predictability, we analyze some properties of intrinsic variability in the various models. First, we consider the geographical distribution of low-pass T0–300, as produced by a 5-yr Lanczos filter. This filter emphasizes the variability one might expect to be important in our study. In general, the strongest low-pass variability is located in the North Pacific (20°–65°N, 120°E–110°W) and North Atlantic (20°–75°N, 80°W–0°) (Fig. 1). This is true not only in each model but also for an estimate based on the short observational record (Levitus et al. 2009) (Fig. 1, top). These two basins are the focus of our study. There is some commonality in low-pass variability in these basins among the models, most notably in the existence of variability centers in the vicinity of the Kuroshio Extension and the North Atlantic subpolar gyre. On the other hand, there are also substantial differences among the various models. For example, a strong southwest–northeast swath of variability in North Pacific midlatitudes is present in half of the models and absent in the other half as well as in the observations.

Another characteristic of intrinsic variability, which is more closely related to predictability, is its power spectrum. Averaged power spectra for T0–300 in the two basins of each of the six models are shown in Fig. 2. The distinctive red character of the spectra coincides with the Hasselmann (1976) idea that oceanic climate variability can be approximated as a white-noise-driven autoregressive process with memory. The longer the memory, the more highly predictable is such a process in the sense that distributions of predictions starting from the same initial conditions diverge less rapidly in systems with long memory. At sufficiently long time scales, the spectrum of a red noise process reaches a plateau. Leith (1975) and DelSole and Tippett (2009a) reasoned that within that plateau only the component of variability that exceeds the plateau value is predictable. Obvious departures from the plateau are most pronounced in the North Atlantic, but these departures vary greatly from model to model: HadCM3 and KCM do not have obvious plateaus; in MIROC3.2 it is reached only for very low frequencies; and CCSM3 and GFDL CM2.1 have clear peaks at the high frequency end of the plateau. Alternatively, one might argue that it is departures from a red spectrum at any point in the spectrum, not just from within the white noise plateau, that

should be considered to be potentially predictable. Again, such departures occur at different frequencies and to varying degrees in the various models.

The spectral information in Fig. 2 is averaged over large basins, but spectral characteristics can vary greatly within a basin. A simple indication of this is seen in Fig. 3, which shows the  $e$ -damping time for each model at each grid point as derived from fitting exponential curves to T0–300 autocorrelation values. This characteristic time is one way to encapsulate the redness of the spectrum with a single number. Strikingly, the regions with the most persistence vary from model to model. These substantial model-to-model variations, together with those in Figs. 1 and 2, indicate the potential for large contrasts in predictability from one model to another.

### 3. Methodology

#### a. Measures

Any complete climate prediction consists of an evolving distribution of states. Determining the predictability of the climate system consists of quantifying by how much and for how long such forecast distributions can be distinguished from a background distribution. In our study, when quantifying predictability we use a model's climatological distribution of states as that background.

One property of forecast distributions that is of interest is their spread, which we measure by mean square difference (MSD). MSD is simply one-half of the average squared Euclidean distance between all pairs of members of the forecast distribution divided by the number of state variables. When calculating MSD, states are represented by vectors of EOF coefficients. Because we use normalized coefficients, our MSD results would not change if the patterns used to represent states were linear combinations of EOFs rather than the EOFs themselves (Griffies and Bryan 1997b). Typically MSD is calculated using forecast pairs that are initially very similar (e.g., when taken from an ensemble of slightly perturbed states), but, for sufficiently long forecasts, this quantity approaches a value of one as the pairs become no more similar than random draws from the climatological distribution.

The second indicator of predictability in our study is relative entropy (Cover and Thomas 2006), which for Gaussian distributions is equal to

$$R = \frac{1}{2} \log_2(e) \left\{ \ln \left[ \frac{\det(\boldsymbol{\sigma}_b^2)}{\det(\boldsymbol{\sigma}_f^2)} \right] + \text{trace}(\boldsymbol{\sigma}_f^2 / \boldsymbol{\sigma}_b^2) + (\boldsymbol{\mu}_f - \boldsymbol{\mu}_b)^T (\boldsymbol{\sigma}_b^2)^{-1} (\boldsymbol{\mu}_f - \boldsymbol{\mu}_b) - n \right\}. \quad (1)$$

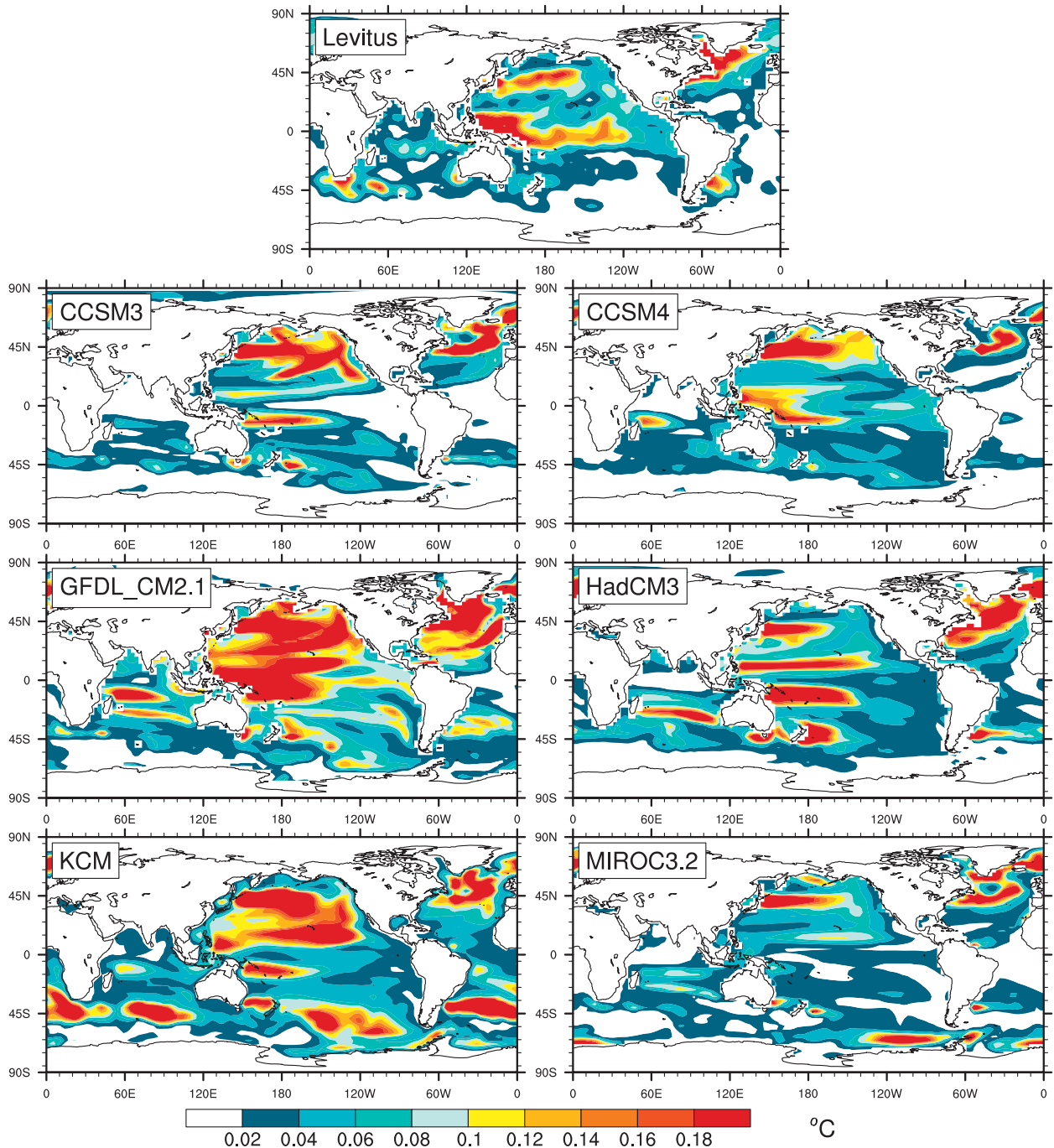


FIG. 1. Variance of 5-yr low-pass filtered T0-300 in Levitus et al. observations (during the period of 1955-2009) and in six AOGCM control run experiments.

In (1)  $\mu_b$  and  $\mu_f$  stand for the mean state vectors in  $n$ -dimensional background and forecast distributions  $P_b$  and  $P_f$ , respectively, while  $\sigma_b^2$  and  $\sigma_f^2$  correspond to covariance matrices representing relationships in these same distributions. As Kleeman (2002), Majda et al. (2005), and Branstator and Teng (2010) have pointed

out, this quantity is attractive because it can be used for multivariate forecasts and because it has a well-defined meaning in information theory. Extensive discussions of relative entropy are included in the aforementioned papers and references therein. For the purposes of our paper, it should be sufficient to understand that relative

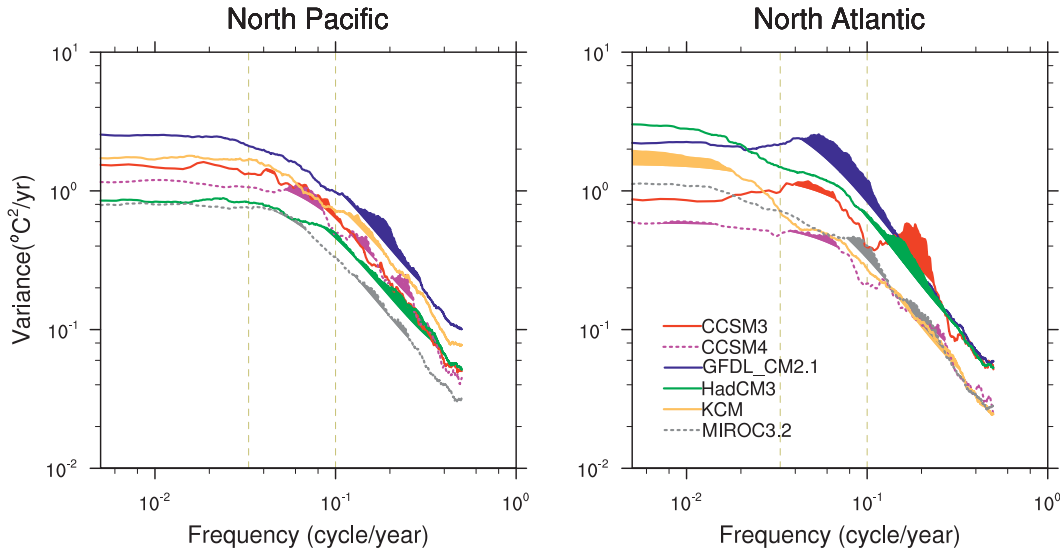


FIG. 2. Averages of T0–300 power spectra at each grid point in the (left) North Pacific and (right) North Atlantic with latitudinal weighting. A taper of 10% of the values at each end was applied, and the spectral estimates were produced by averaging over approximately 7% of the periodogram values. A Fisher transformation was applied to calculate the domain-averaged 1-yr lag autocorrelation values ([www.ncl.ucar.edu/Document/Functions/Shea\\_util/specx\\_ci.shtml](http://www.ncl.ucar.edu/Document/Functions/Shea_util/specx_ci.shtml)), and spectra exceeding the 95% significance level based on these values are shaded. The two vertical (gray) reference lines denote frequencies corresponding to 30- and 10-yr periods.

entropy measures the extra information one has about the state of the climate system by knowing that it is in the forecast distribution relative to the information one would have if one assumed it were a member of the climatological distribution. Here information is measured in terms of the reduced number of binary bits it takes on average to represent the state. Unless affected by sampling, relative entropy approaches zero as forecast distributions approach the climatological distribution.

As we have mentioned, in our study we are interested in the predictability of a system in terms of average properties, where averages are taken over predictions initiated from each state on a system’s attractor. We refer to such averages as “attractor averages,” denote them by a double overbar, and approximate them by averaging over all states from a model’s control integration. As explained in appendix A, for the regression method of estimating predictability statistics introduced later in this section,

$$\overline{\overline{\text{trace}(\sigma_f^2/\sigma_b^2) + (\mu_f - \mu_b)^T(\sigma_b^2)^{-1}(\mu_f - \mu_b) - n}} = 0. \tag{2}$$

Hence, we can find  $\overline{\overline{R}}$  by evaluating the attractor average of the first term in (1), which only involves covariances. Interestingly, the first term in (1) is equal to the difference in entropy between distributions  $P_b$  and  $P_f$ , which Schneider and Griffies (1999) refer to as “predictive

information.” Because of (2), in our application attractor averages of predictive information and relative entropy are the same. The equivalence of these two quantities for attractor averages has been demonstrated in a more general context by DelSole (2004).

Kleeman (2002) has proposed that it is useful to distinguish the contribution resulting from the term

$$\frac{1}{2} \log_2(e)[(\mu_f - \mu_b)^T(\sigma_b^2)^{-1}(\mu_f - \mu_b)]$$

in (1) (called the “signal relative entropy”) from the contribution of the remaining terms (called the “dispersion relative entropy”). When (2) is valid, attractor averages of these components can be found from the statistics of covariances of forecast distributions alone. This does not mean that ensemble means make no contribution to total entropy, but rather their contribution can be diagnosed from forecast distribution covariances.

When forecasts are composed of finite ensembles, predictability can be said to be lost when forecast distributions are indistinguishable from a background distribution in terms of some threshold of statistical significance (Branstator and Teng 2010; Teng and Branstator 2011). In this study, which does not use ensembles, such a concept is less meaningful. Instead, we compare MSD and relative entropy to reference values as an aid in comparing predictability. For MSD we use 0.60 and 0.90

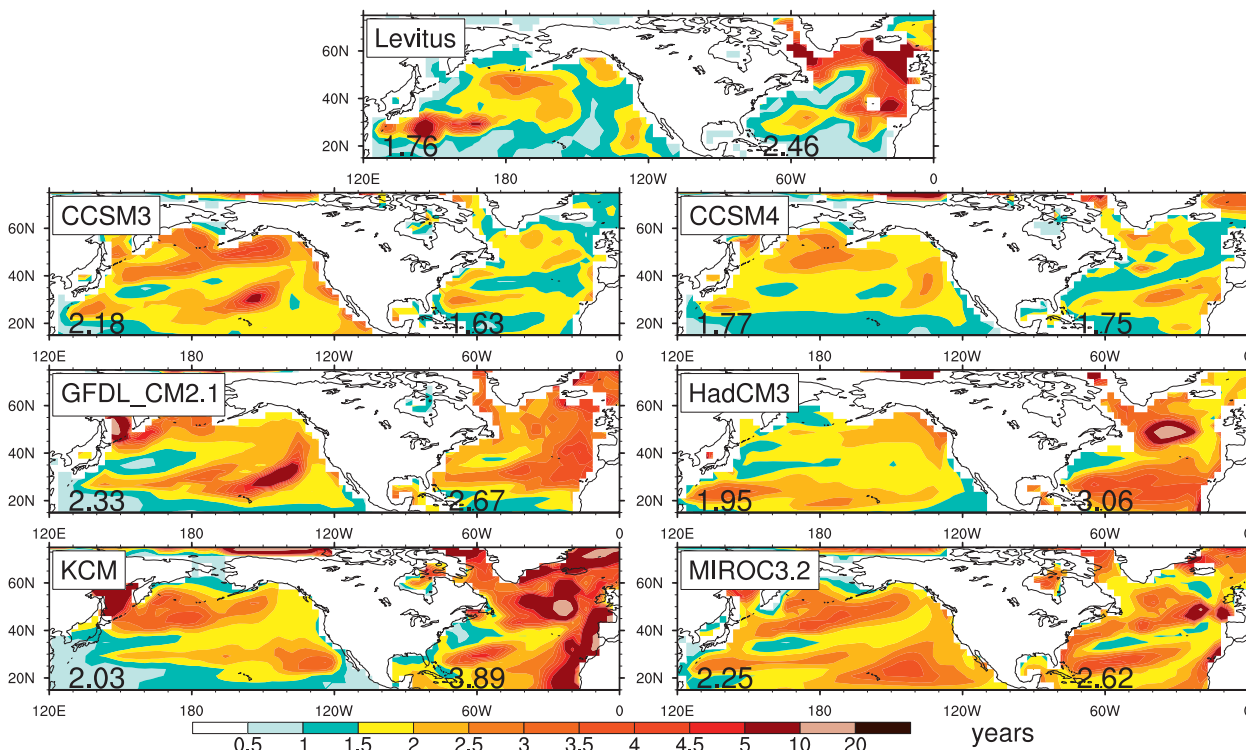


FIG. 3. The  $e$ -damping time of T0–300 at each grid point in the two northern ocean basins, derived by least squares fitting a line to the logarithm of autocorrelation coefficients. Values in the southwest corner of each basin are the area averages of the  $e$ -damping times in that basin.

as reference points. For relative entropy we use a value that is essentially equivalent to a MSD of 0.60. For an  $n$ -dimensional system this value is arrived at by finding the dispersion relative entropy of a forecast distribution for which each component has a normalized variance of 0.60 and vanishing covariances, namely,

$$R_n^{\text{nom}} = \frac{n}{2} \log_2(e) [\ln(1/0.60) + 0.60 - 1]. \quad (3)$$

We sometimes refer to this value as representing a “nominal” limit of predictability because it is, in fact, an arbitrary threshold.

#### b. Methods

The usual method for assessing the predictability of a climate model is to use Monte Carlo techniques in which forecast distributions are approximated by finite ensembles whose initial members are produced by randomly perturbing an initial state. As we have mentioned, for practical reasons this approach allows estimation of predictability for only a handful of initial states, which may not be representative of general system behavior. As an alternative, we use two techniques that only require long control integrations and give estimates of attractor average predictability.

One of the methods we employ uses analogs. Unlike our second method, which approximates a system with a simple model, this approach makes no assumptions about the dynamics of the system being examined. Rather, for every year  $t$  in the control run we find that year  $t'$  for which the corresponding control run state,  $\mathbf{s}(t')$ , is most similar to  $\mathbf{s}(t)$ . When doing this we only consider years  $t'$  that are separated from  $t$  by at least 50 years, and we measure similarity in terms of gridpoint Euclidean distance. By calculating the squared difference between  $\mathbf{s}(t + \tau)$  and  $\mathbf{s}(t' + \tau)$  for all such pairs, we are able to find the attractor average mean square difference between states  $\tau$  years after they were similar. Note that this procedure is similar to analyzing the spread at forecast year  $\tau$  of initially similar pairs of realizations from a Monte Carlo perturbation ensemble except that it takes into account spread rates from initial states over the entire attractor rather than for a single point. Of course, the degree to which the considered initial states represent the complete system attractor and the appropriateness of considering the resulting analog pairs to be small perturbations of each other depends on the length of the available control experiment.

To test and demonstrate the capabilities of the analog approach, we compare it to predictability estimates

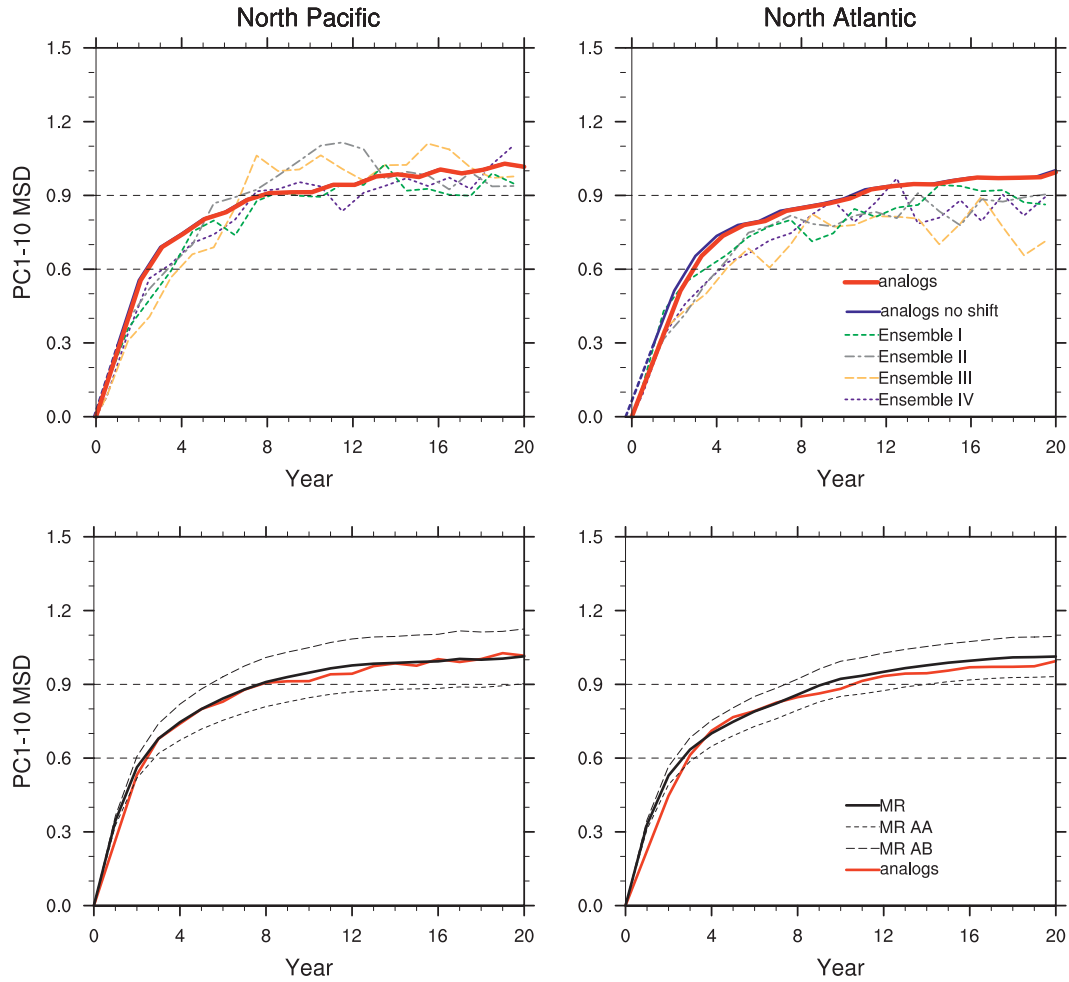


FIG. 4. Attractor average mean square difference (MSD) of the leading 10 normalized PCs of T0–300 in the (left) North Pacific and (right) North Atlantic as a function of forecast range. (top) Comparison of MSD from four 40-member CCSM3 ensemble experiments (dashed lines) and MSD based on analogs from a 700-yr CCSM3 control run (red and blue solid lines). The blue solid line assumes that the initial spread of the analog pairs roughly agrees with the perturbation growth after 1 forecast year. After extrapolating the blue curve (indicated by the blue dashed line) to locate the time range corresponding to zero MSD, we shift the analog MSD curves (blue solid lines) so that MSD becomes zero at time range zero and the shifted curve is denoted by the red line. (bottom) Comparison of MSD from the analog pairs [red solid lines, as in (top)] with MSD from multivariate linear regression (MLR) (black line). The black dotted curves show the attractor average MSD for the MLR operators applied to dependent data, and the black dashed curves pertain to when they are applied to independent data (appendix B).

derived from the traditional Monte Carlo method. Branstator and Teng (2010) reported on the predictability characteristics of four ensemble experiments produced by CCSM3. Each of these experiments started from a different ocean/land state and comprised 40 realizations that were identical except for having either different initial atmospheric states or tiny perturbations to the solar constant. The top row of Fig. 4 indicates the MSD for the leading 10 normalized PCs of T0–300 in the North Pacific and North Atlantic in each experiment. By this MSD measure, after less than a decade each of the ensembles has nearly reached an equilibrium

configuration and is thus equivalent to an ensemble of random draws. Recall that the MSD value of 0.90 marked with a dashed line in the figure is intended as a reference point and does not correspond to a particular significance level.

On that same plot we show the corresponding measure of attractor average forecast distribution spread based on analogs from the 700-yr CCSM3 control run. The blue line shows the MSD for all such pairs and for their evolution during the succeeding 20 years. To account for the fact that initially the pairs have finite separation, we extrapolate MSD curves to earlier times

using the Leith (1975) result that, during early stages of perturbation growth, perturbation variance tends to increase linearly (dashed blue line in Fig. 4). We then shift the analog MSD curves so that MSD is zero at a forecast range of zero years (red curve). A similar procedure is used for all analog results in our study.

Use of different metrics, domains, or variables to select analogs could modify our results, but the similarity between the analog results and ensemble results in Fig. 4 indicates that the procedure used here does give good estimates of the growth rates of perturbations in CCSM3. Furthermore, in results not shown here, when we have based our choice of analogs on leading PCs and larger domains, our results have not changed materially. There is, however, one prominent difference between the analog results and ensemble results in Fig. 4, namely, the slow saturation of each ensemble in the North Atlantic that is not reproduced by the analog method. But, this feature is forced by changes in greenhouse gas concentrations in the ensemble experiments (Branstator and Teng 2010) and thus should not be reflected in the estimates of initial-value predictability that the analog method produces.

The analog method as we have applied it has the drawback that it only measures dispersion, but the ensemble mean also contributes to predictability. For example, as is apparent from the definition of relative entropy (1), a forecast ensemble with the same covariances as the climatological distribution can still contain information if its mean is different from the climate mean. In fact, as measured by contrasting signal and dispersion relative entropy, it is not uncommon for the mean to make a larger contribution to decadal initial-value predictability than is made by the spread (Branstator and Teng 2010). Therefore, we employ a second technique for estimating predictability that enables us to take into account both contributions. We refer to it as the multivariate linear regression (MLR) method.

To understand the MLR method, let  $E_N$  be the leading  $N$  EOFs of T0–300 in the domain of interest and consider  $S$ , a set consisting of all states  $\mathbf{s}(t)$  from a control run that have the same projection onto  $E_N$ . If  $M \ll N$ , then it is reasonable to assume that the difference between the evolution in the control run of the first  $M$  PCs of a particular member of  $S$  and the mean evolution of all members of  $S$  is a result of the dispersion of states with perturbed initial conditions that occurs in a chaotic system. After all, the members of  $S$  differ only for PC components of an index greater than  $N$ , and these generally have small amplitudes relative to the first  $M$  PCs. Next, assume that for the first  $M$  PCs, the value  $\tau$  years into the future of the ensemble mean of any such  $S$  can

be estimated from an  $M \times N$  MLR operator  $\mathbf{L}^\tau$  that is computed from lag-0 and lag- $\tau$  covariances of the leading  $N$  PCs in the control integration. This is reasonable because regression is designed to predict ensemble mean behavior. From these two assumptions,  $\mathbf{L}^\tau \mathbf{s}(t) - \mathbf{s}(t + \tau)$  estimates the effect of initial perturbations to the evolution of the first  $M$  PCs of  $\mathbf{s}(t)$ . Hence, by calculating  $\mathbf{L}^\tau \mathbf{s}(t) - \mathbf{s}(t + \tau)$  for all  $t$ , the covariance statistics of perturbed predictions at range  $\tau$  can be estimated. DelSole and Tippett (2009b) have utilized a similar regression approach to study predictability. From (1) and (2), these covariances can be used to find attractor average relative entropy, as well as its signal and dispersion components.

In addition to being applicable when total relative entropy is required, the MLR approach is attractive because it should require less data than the analog method. The MLR method assumes a particular functional form for system dynamics, and one needs only enough data to determine the associated parameters through statistical analysis. By contrast, for the analog method to be effective, one must have enough data to find many similar pairs of states, which is known to be a very data intensive problem for high-dimensional systems (Van den Dool 1994). On the other hand, AOGCM evolution is likely to have some nonlinear behavior that cannot be captured by MLR. Hence, MLR model errors are not just a consequence of a system's sensitivity to initial perturbations, so the estimates of dispersion they provide are inflated. For this reason, we consider the MLR results to be fundamentally less reliable than those from the analog approach.

Finally, note that the MLR method bears some similarities to the use of linear inverse modeling in predictability studies (Newman 2007; Alexander et al. 2008), but one difference is that, in linear inverse modeling, a single propagation operator is assumed valid for all forecast ranges, whereas we construct a different MLR operator for each forecast range. Hence, unless the AOGCMs examined here are exactly first-order Markov processes, in which case the linear inverse modeling assumption is valid, the MLR operators we use are necessarily more accurate.

In the results shown in this paper  $N = 30$  and  $M = 10$ , but none of our results is affected in any important way if we use  $M = 5$  or  $M = 20$ . The value of  $N$  was chosen to maximize the variance predicted by the MLR model while not overfitting relationships in the finite control samples. As explained in appendix B, the sensitivity of our results to sampling has also been minimized by adapting a strategy suggested by Lorenz (1977).

When our MLR procedure is applied to CCSM3, we produce the attractor average estimates of MSD for



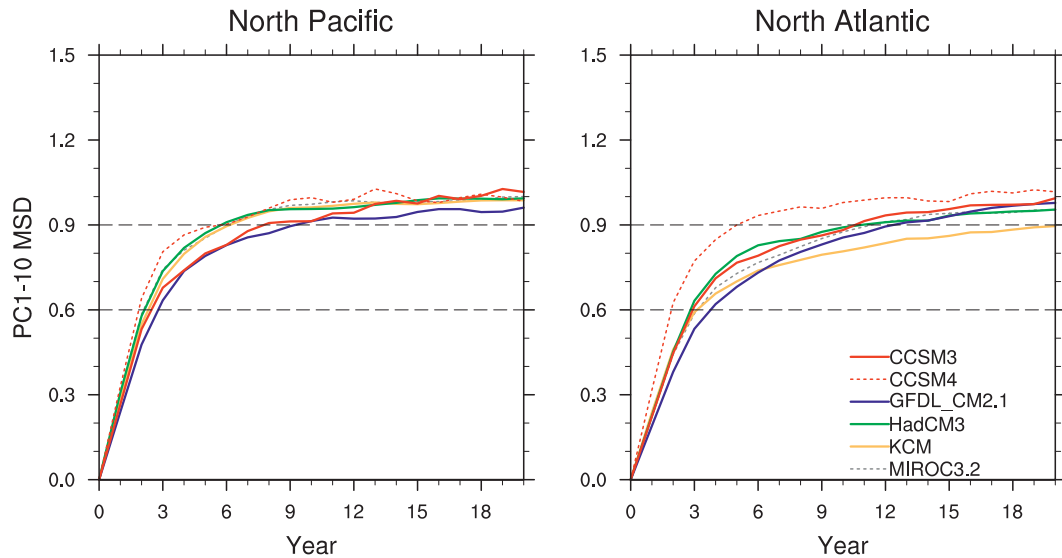


FIG. 5. For six AOGCMs, attractor average MSD for the leading 10 normalized PCs of T0–300 in the (left) North Pacific and (right) North Atlantic found by averaging across all analog pairs in a model's control run as a function of forecast range.

PC1–10 shown by the dark black curves in the bottom panels of Fig. 4. The MLR estimates are very similar to those from the analog approach, which are also shown in the bottom panels. Likewise, in results not shown here, when we have compared univariate attractor average MSD at each grid point in the North Pacific and North Atlantic as given by the analog and MLR approaches, we have found that they are also very similar.

From the similarity of analog and MLR regional and local MSD values, we draw two conclusions. First, since the arguments and results of appendix B suggest that our MLR results are not significantly affected by sampling errors, it appears that neither are our analog results. Second, despite its simplifying assumptions, the MLR method gives results that must be of similar quality to those from the analog approach. Taking these results into consideration, in the remainder of our study we use the analog approach, with its less stringent assumptions, when considering the MSD measure of predictability, and we use MLR when signal and total relative entropy are investigated.

#### 4. Basin predictability

As the first step in quantifying initial-value predictability of the six models in our study, we apply the analog method tested in the previous section to arrive at attractor average basin MSD values (Fig. 5). Focusing first on the North Pacific, in the first few years the differences among models are modest with differences of little more than a year for the range at which their spread

reaches 0.60 of the spread of random climatological draws. But, at longer ranges substantial model-to-model variations in the growth of MSD are apparent. For example, CCSM3 and GFDL CM2.1 cross the 0.90 threshold 2–3 yr after the other four models. Variations in the rates of spreading are even more distinctive in the North Atlantic. There is a factor of 2 difference in the range when the models reach 0.60; for CCSM4 this occurs in about 2 yr, while for GFDL CM2.1 it occurs in slightly less than 4 yr. Model-to-model variations in the range at which the 0.90 threshold is reached are even more pronounced. In the model with lowest predictability (CCSM4), it takes 5 yr to reach this threshold, while the model with highest predictability (KCM) requires 20 yr.

From our analysis of intrinsic variability (Figs. 1 and 3), we might expect the above basin average results to mask significant geographical variations in predictability. To check this possibility, we carry out a series of calculations in which for each model and for each forecast range we calculate MSD for all analog pairs at each grid point. We use these values to find the forecast range at which each gridpoint MSD reaches 0.60 and find that the statistics of Fig. 5 do indeed hide substantial variations within each basin and among the models (Fig. 6).

In the North Pacific, from a basin average perspective, CCSM3 and GFDL CM2.1 are the models with highest predictability (Fig. 5), but Fig. 6 shows that this high predictability is actually concentrated in just a few regions. Interestingly, the location of these regions is different in these two models, with the eastern and northern

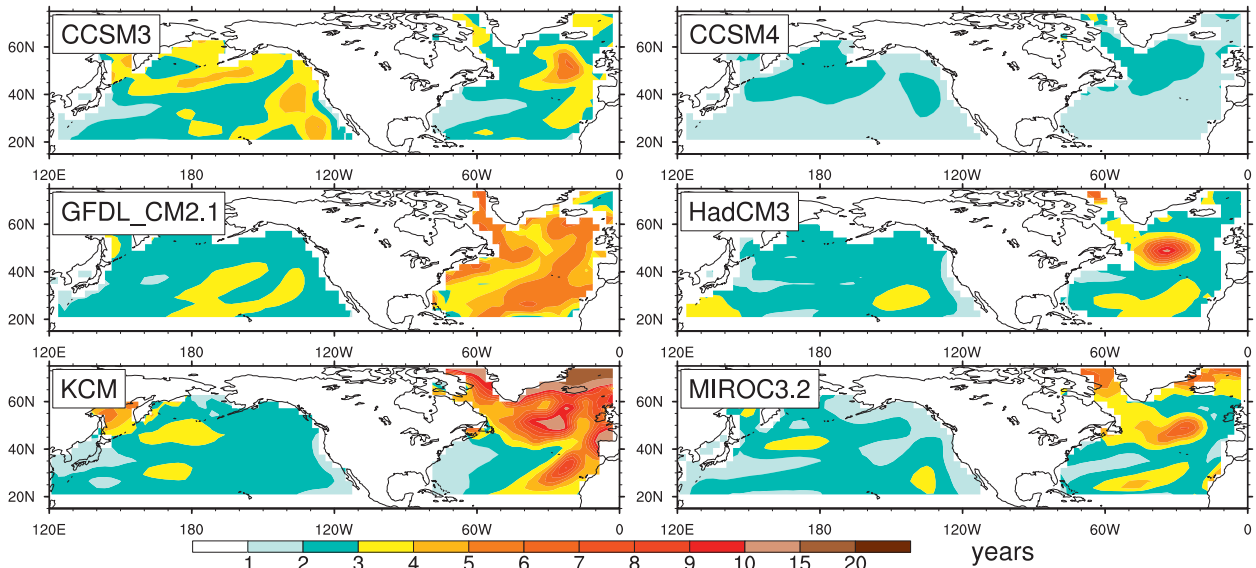


FIG. 6. As derived from analogs, the forecast years in which the attractor average MSD at each grid point reaches 0.6 of the spread of random control run draws.

basin perimeter having highest predictability in CCSM3, while the basin interior has highest predictability in the GFDL model. Though the North Pacific variations are subtler in the other models, for them it is also true that there is almost no commonality in locations of high predictability. In the North Atlantic, there are also huge geographical contrasts in predictability and the threshold years span a much larger range. In the most extreme cases the range varies from 1 yr in some locations to more than 10 yr in others. In this basin, however, the

structure of these geographical contrasts is somewhat more organized than in the North Pacific basin. Many of the models have one band of high predictability in the subtropics and another to the west of the British Isles. But even in these regions the year at which the 0.60 threshold is attained varies a great deal from model to model.

Next we turn to relative entropy as estimated from the MLR method to incorporate both mean signals and spread when quantifying predictability. Figure 7 displays the attractor average total relative entropy in each

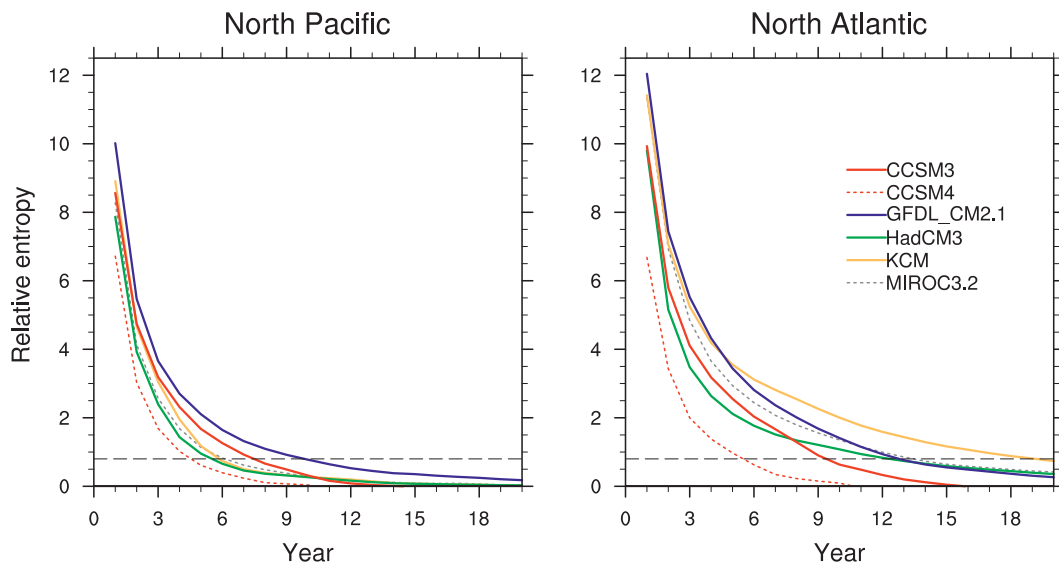


FIG. 7. Attractor average relative entropy as a function of forecast range in the (left) North Pacific and (right) North Atlantic using 10 leading PCs estimated from the MLR method. The horizontal dashed line indicates  $R_{10}^{\text{nom}}[(3)]$ , which corresponds to the nominal loss of predictability.

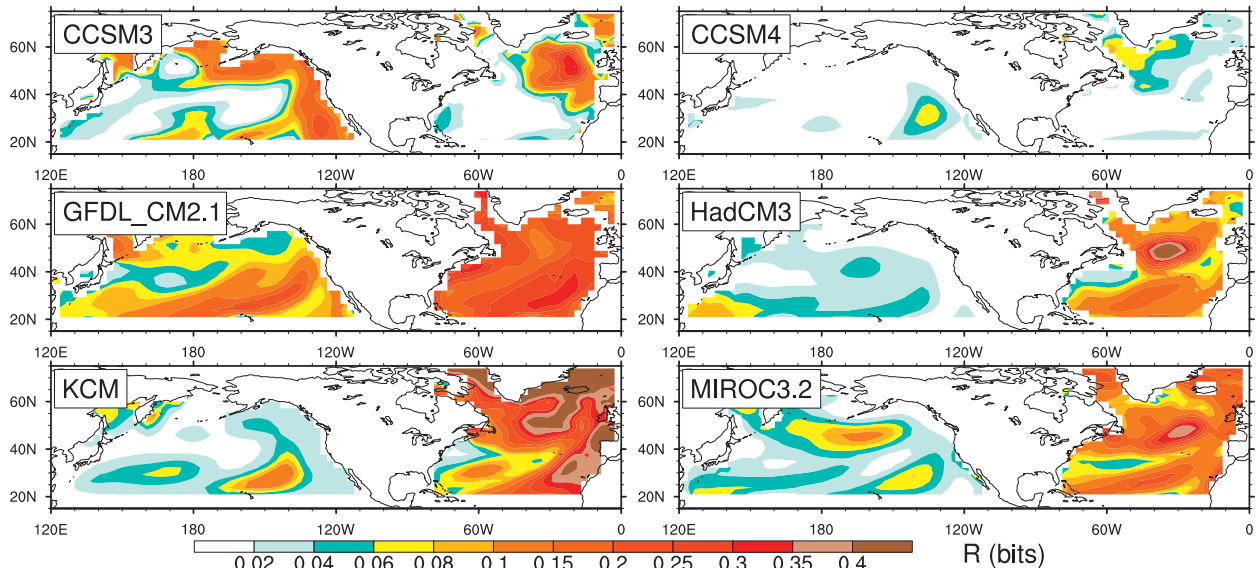


FIG. 8. Averages for forecast years 6–10 of attractor average univariate relative entropy for gridpoint values of T0–300. These are based on MLR forecasts of 10 EOFs, which are then transformed to gridpoint values.

basin for each AOGCM based on 10 PCs. Loosely speaking, each decrease of one unit of relative entropy corresponds to a halving of the accuracy with which we can forecast the state.

Initially information content is extremely large for each model because we have nearly exact knowledge of what state the system is in. During the first few years of the predictions, information rapidly decreases. Between years 1 and 3, on average, relative entropy is more than halved in both basins in all models, but the rate at which information is lost varies considerably from one model to another. Even after one year there are substantial differences, and these contrasts continue to expand, as indicated by the contrasting curvatures. In the North Pacific, at year 3 there is about a factor of 2 difference in relative entropy between the models with the most predictability (GFDL CM2.1) and the least predictability (CCSM4), corresponding to a difference of more than two units of relative entropy. In the North Atlantic at year 3, there is almost a factor of 3 difference between the models with highest and lowest predictability; considering the logarithmic character of relative entropy, the corresponding four unit difference is very substantial. For reference in Fig. 7, the dotted line signifies a value of  $R_{10}^{\text{nom}}$ . This is the information that corresponds to the 0.60 dashed threshold in Fig. 5 and the threshold used to construct Fig. 6 and also corresponds to our nominal limit of predictability. In the North Pacific this value is reached after 5–10 yr. In the North Atlantic some models reach this value in less than 6 yr, while others retain this much information in their forecasts for more than 15 yr.

To quantify regional variations in predictability in terms of information, we consider the geographical variation of relative entropy. We utilize the same MLR forecasts employed to construct Fig. 7 but transform the resulting 10 PC error vectors to gridpoint values and then calculate the attractor average error variances needed to evaluate univariate relative entropy. Figure 8 displays the average of these values for forecast years 6–10. Though near the end of the useful range of forecasts, we show this range because it is when contrasts between models are largest. This plot reinforces the conclusion drawn from Fig. 6 that model and basin differences in predictability are largely a reflection of local features. For example, it is not uncommon for there to be a factor of 10 difference in the relative entropy within a single basin in a given model. As far as intermodel variability is concerned, there is some contrast between the two basins. In the North Atlantic, as seen when we considered the forecast spread at intermediate ranges, there is some consistency among several of the models in that the subtropics and subpolar gyre tend to be regions of high predictability in many models. But in extreme cases, such as the difference between CCSM4 and KCM in the northern North Atlantic, even for these features there can be a factor of 10 contrast. In the North Pacific, relative entropy serves to reveal that model-to-model variations are even stronger than in the North Atlantic. Indeed, with the exception of regions northeast of Hawaii and south of the Aleutians, no one location has enhanced predictability in more than two models.

## 5. Mode predictability

Another useful perspective is to focus on the predictability of each AOGCM's prominent intrinsic modes. They are not necessarily the most predictable patterns (Teng and Branstator 2011; DelSole et al. 2011) as is sometimes hypothesized, but the fact that they have high amplitude makes it worthwhile to give them special attention. In the previous section, we did use an EOF representation of the T0–300 field, but this was done as a means of efficiently representing the data and not as a way to identify physical modes of each model. An exhaustive treatment of the predictability of the modes of our six models is beyond the scope of our study. Instead, we concentrate on a single T0–300 mode in each basin. To allow for modes whose structure may be time dependent, we use a complex EOF (CEOF) analysis (von Storch and Zwiers 1999) to identify them.

When the leading CEOF in the North Pacific is calculated, we see that it is structurally similar in the various models (Fig. 9). In each model one phase, consisting of a prominent zonally elongated feature extending eastward from the coast of Japan, and a second feature of opposite sign along the west coast of North America, resembles the Pacific decadal oscillation (Mantua et al. 1997). The second phase, which by construction tends to occur some years later, is made up of a zonally stretched feature corresponding to an eastward and northward shift of the zonal feature from the first phase together with the emergence of an oppositely signed anomaly to the south. To varying degrees the mode is largely a standing oscillation, but clearly there is also a propagating component in most models. These same features exist in CEOF1 from nature (top row of Fig. 9). In addition to these common features, there are notable variations in the structure from one model to another, primarily in the eastern portion of the high amplitude phase's east–west anomaly.

The situation in the North Atlantic is rather different in that there is more model-to-model variability in CEOF1 (Fig. 10) and propagation plays a much bigger role in the evolution of the mode, with both phases having comparable amplitude. In most models there is a suggestion of eastward propagation of anomalies originating off the coast of the Maritime Provinces of Canada and in several models westward propagation at higher latitudes. But GFDL CM2.1 appears to have features that propagate in the opposite sense, and there is no clear propagation in the pattern in nature (though with only about 50 years of data, the robustness of this pattern is uncertain). Taking into account CEOF2 in the North Atlantic does not increase the

similarity of patterns among models or between nature and models.

To assess the predictability of these leading modes, we use the same MLR approach used above except the resulting errors are projected onto the two patterns of CEOF1. Figure 11 shows the total relative entropy implied by the covariances of these errors for each model. With only 2 degrees of freedom it is not surprising that there is less information in these forecasts than in the 10 PC forecasts in Fig. 7. But, in terms of the forecast range at which relative entropy reaches the nominal predictability limit, when considered as a group the predictability of the modes is qualitatively similar to the predictability of 10 combined PCs. In the North Pacific, similar to what we saw for generic structures, the range at which forecasts of the modes reach  $R_2^{\text{nom}}$  is about 3–8 yr. In the North Atlantic, this threshold value is reached at a range of between 6 and 20 yr, depending on the model. We find that the large model-to-model variations in modal predictability is of similar magnitude to the variations DelSole et al. (2011) found when comparing the predictability of a single, highly predictable pattern in 14 AOGCMs. Note an implication of the small values in Fig. 11 compared to those in Fig. 7 is that basin predictability is not controlled by a single leading mode.

## 6. Three distinguishing factors

In this section we highlight three factors that can substantially enhance the predictability of a system and that may also help distinguish the predictability characteristics of the various models. The first factor is the mean of forecast distributions. Often predictability studies concentrate on distribution spread; however, as referred to above, earlier investigations have made the point that the distribution mean can be more important. Indeed, as appendix C shows, one expects the mean to dominate information content at long range. This is borne out when we decompose the attractor average relative entropy of Fig. 7 into signal and dispersion components as defined in section 3a (not shown). At year 1, in both basins for all models the signal contributes roughly half of the total relative entropy. By year 6, depending on the model, between 70% and 90% of the total relative entropy in the North Pacific resides in the signal, and between 60% and 70% in the North Atlantic comes from the signal. Beyond year 6 these percentages continue to grow.

To further quantify the importance of the mean signals, in Fig. 12 we plot the year that univariate values of attractor average total relative entropy exceed  $R_1^{\text{nom}}$  based on MLR forecasts. This is the same threshold used

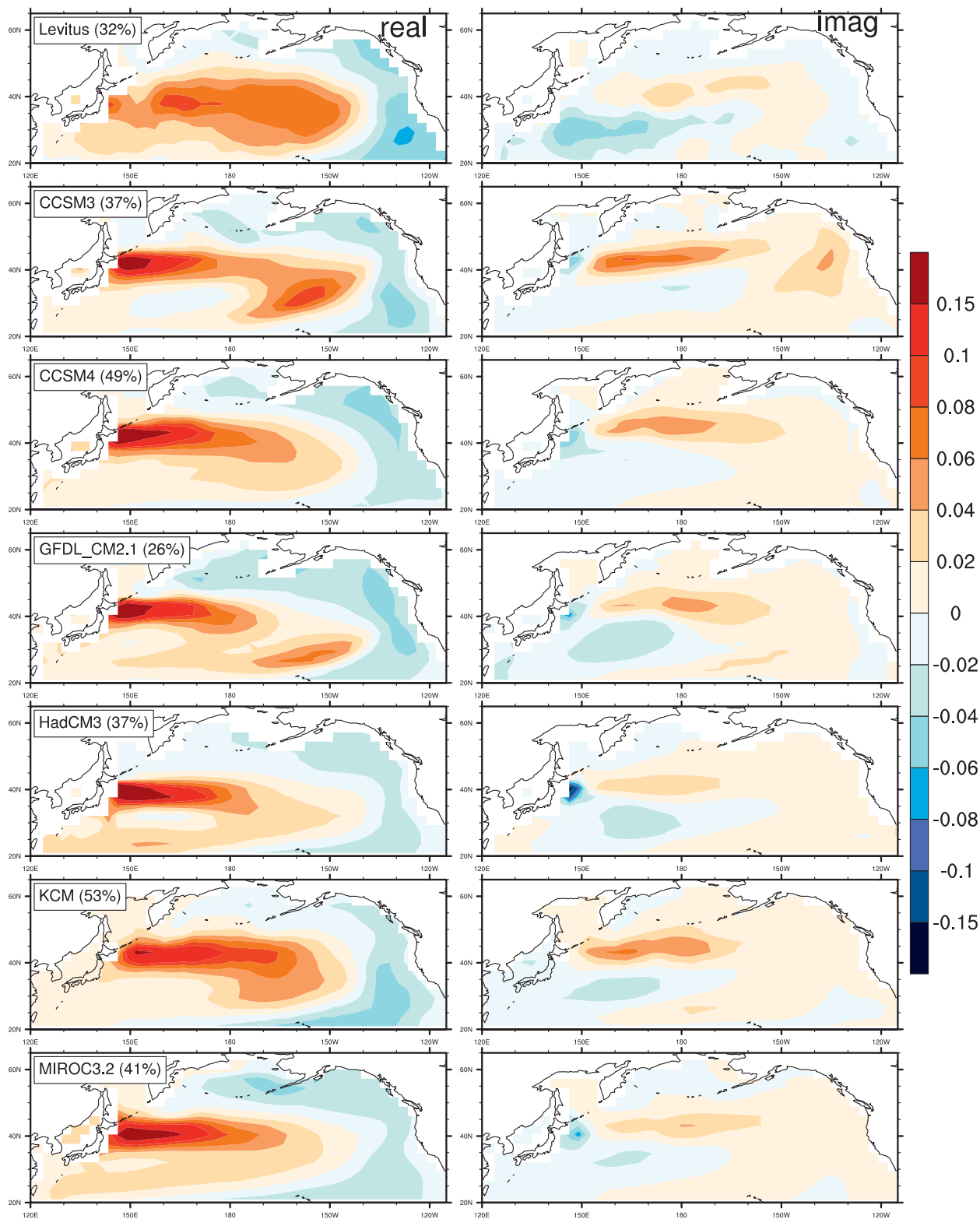


FIG. 9. CEOF1 of T0-300 in the North Pacific in Levitus et al. observations and six AOGCM control runs. The model CEOFs have been rotated to maximize their similarity to the pattern from the observations.

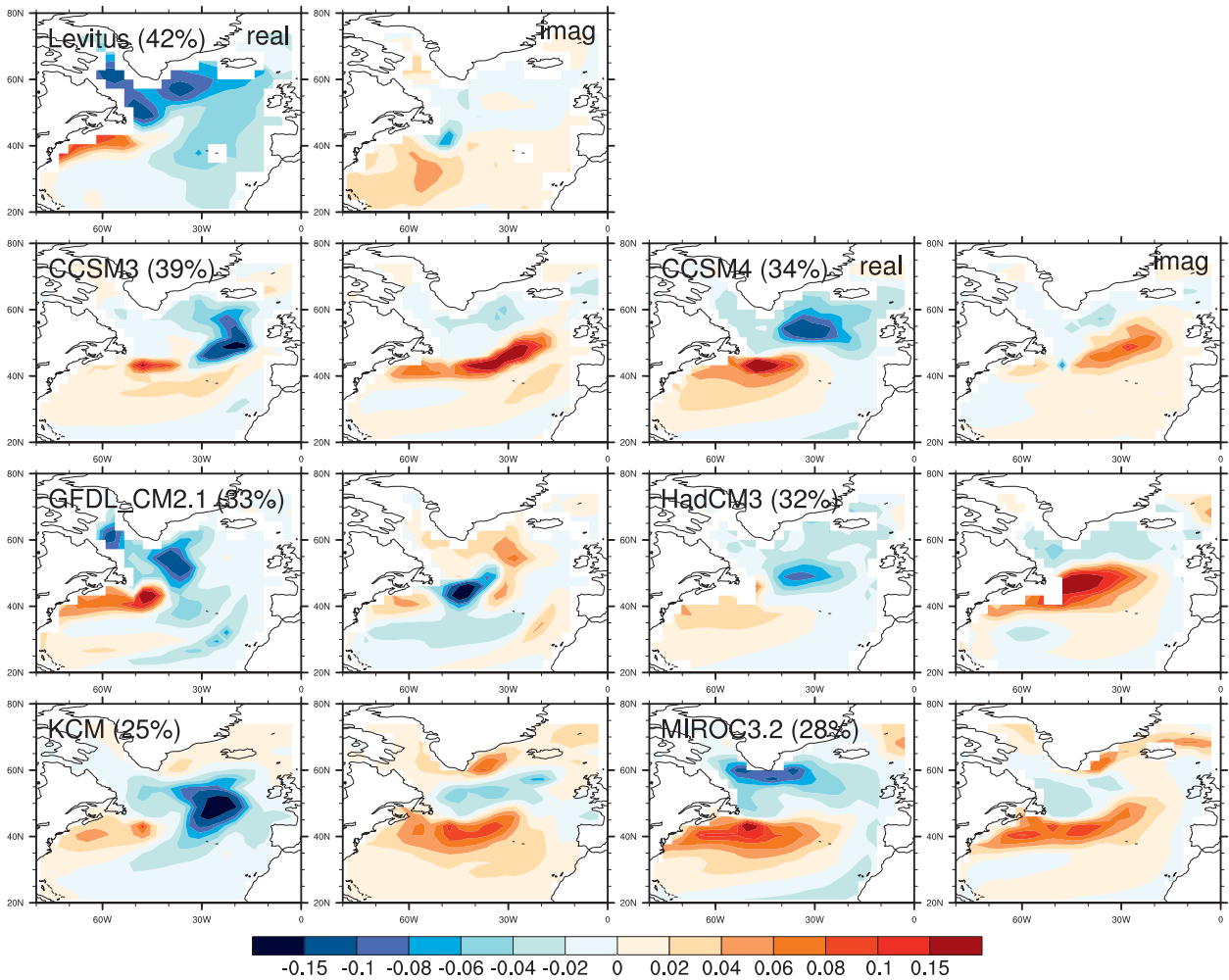


FIG. 10. As in Fig. 9 but for the CEOF1 of T0–300 in the North Atlantic.

in Fig. 6 when only spread was taken into account.<sup>1</sup> Comparing Fig. 12 with Fig. 6, we find that, when the signal component is included, it is 2–3 times longer before the nominal predictability limit is reached. Moreover, the geographical and intermodel variations become more substantial in Fig. 12, with the time range varying from 2 yr to as high as 20 yr and more. Hence, when using relative entropy to measure predictability properties, it is in the evolution of the mean forecast state that most of the predictability resides, and by implication it is this component of forecast distributions that controls differences in predictability properties among models.

<sup>1</sup> Note that, if we redo the calculation with covariances set to zero, we find little difference in the results; differences between Fig. 6 and Fig. 12 result from the signal, not from the inclusion of covariances when calculating relative entropy for Fig. 12.

The second factor affecting predictability is persistence. In the southwest corner of each basin in the Fig. 3 plots of T0–300  $e$ -damping times, the area-average damping time for that basin is given. The substantial model-to-model variations in these average measures of persistence reflect the different degrees of redness of the spectra in Fig. 2. Comparing these damping times with the basin predictability properties in Figs. 5 and 7, we see that there is a tendency for more persistent models to have higher predictability. For example, in the North Pacific the most persistent model (GFDL CM2.1) has the highest predictability and the least persistent model (CCSM4) has the lowest predictability. Comparison of prominent local characteristics of persistence in Fig. 3 with geographical distributions of predictability in Figs. 6, 8, and 12 suggests that some, but not all, model-to-model variations in local predictability are related to local persistence.

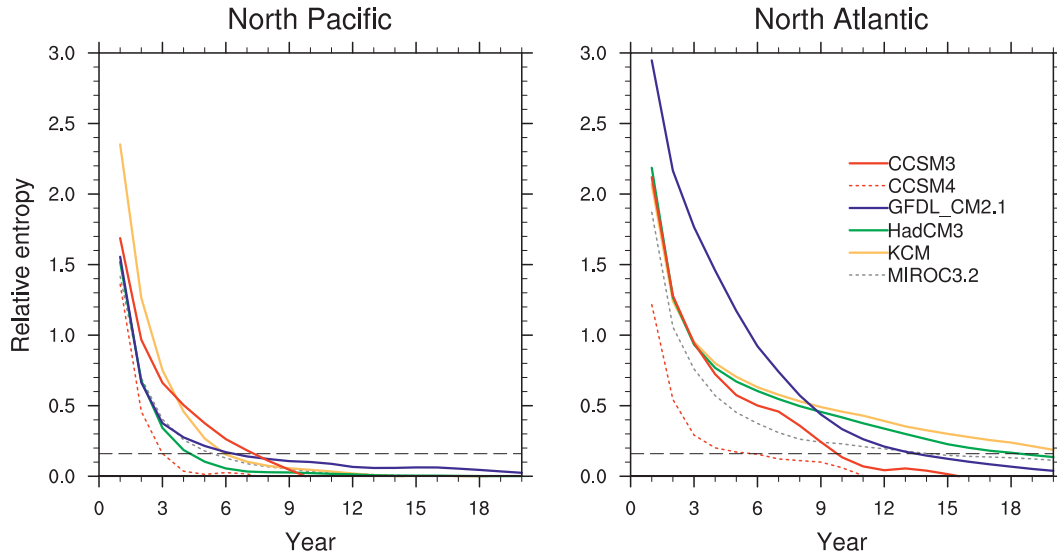


FIG. 11. Using the MLR method, attractor average bivariate relative entropy in the (left) North Pacific and (right) North Atlantic for 1–20-yr forecasts of projections onto CEOF1. The horizontal dashed lines indicate the relative entropy value  $R_2^{nom}$  [(3)], which is our nominal limit of predictability.

The third key factor affecting predictability is the importance of propagation to the evolution of ocean perturbations. Propagation has been considered an important mechanism for decadal variability in the ocean (Meehl et al. 1998; Saravanan and McWilliams 1998), and its contribution to predictability has been studied by Kleeman (2002), Power and Colman (2006), and Barlas et al. (2007). Teng and Branstator (2011) and Teng et al. (2011) quantified the importance of propagating modes

to the predictability that they found in CCSM3 large ensemble experiments.

To confirm that propagation is an important factor, we calculate the signal relative entropy of our models when propagation is and is not present. Propagation can also affect dispersion, but this influence is more difficult to assess with the tools of our study. To estimate attractor average signal relative entropy for the complete systems, we simply use (2) and MLR-derived forecast covariances.

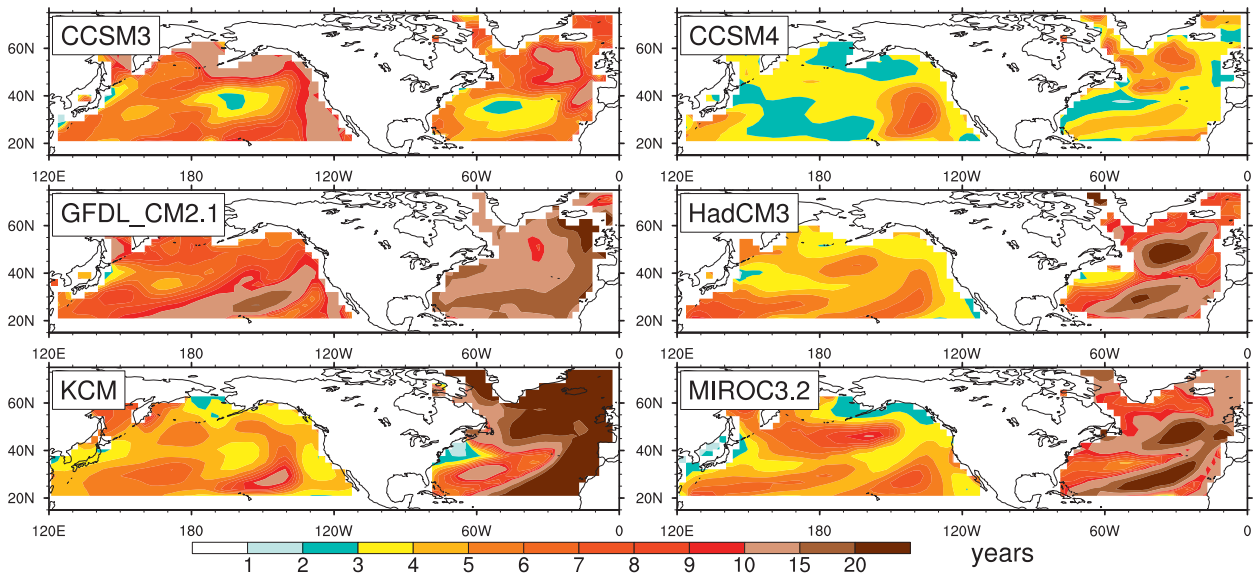


FIG. 12. Forecast range at which attractor-averaged univariate relative entropy at each grid point derived from the MLR method reaches  $R_1^{nom}$  [(3)].

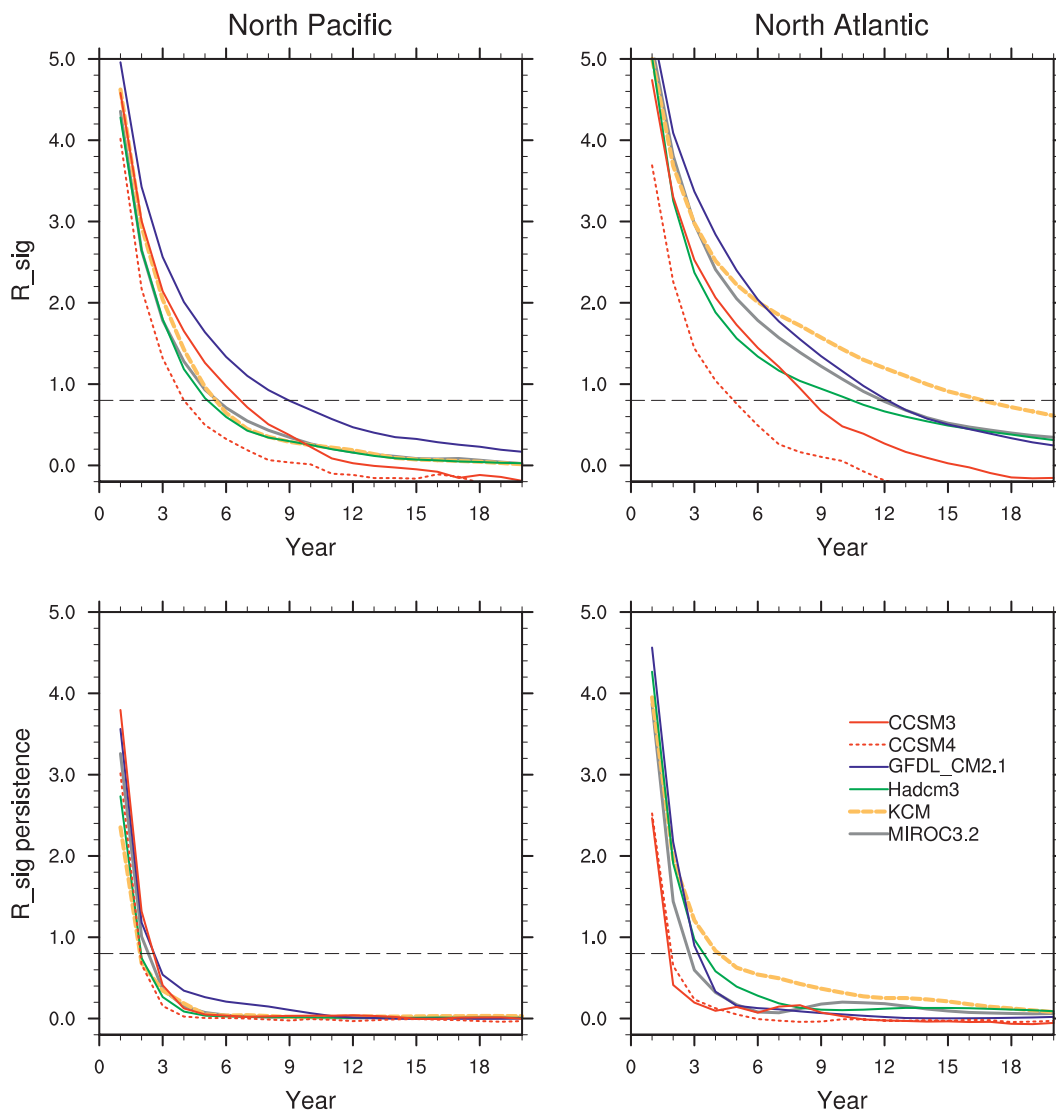


FIG. 13. (top) Attractor average signal component of relative entropy of the (left) North Pacific and (right) North Atlantic, using 1–20-yr forecasts of the 10 leading PCs from the MLR method. (bottom) As in the top row, but the MLR operator is replaced with a univariate regression operator. The horizontal dashed lines indicate  $R_{10}^{\text{nom}} [(3)]$ , which indicates the nominal loss of predictability.

To estimate the average signal when propagation is not present, we remove the effect of PC interactions by constructing univariate MLR operators for each grid point and for each forecast range and apply them to each state on a model's attractor. Given the arguments and results presented above, it is not surprising that, for the multivariate calculation, signal relative entropy (Fig. 13, top row) is very similar to our previous total relative entropy results (Fig. 7), especially after the first few years. (This similarity supports our decision to concentrate on the signal in our analysis of propagation.) By comparison (Fig. 13, bottom row), when the effects of

propagation are not included, signals lose their predictability 2–4 times more rapidly. The univariate operators are able to approximately differentiate the least from the most predictable models, as could be anticipated from the previous paragraph's discussion of persistence and the fact that the univariate operators represent nothing more than univariate lag correlations. But differences in predictability among the models are much less substantial when only damping times are affecting the results. Hence, it appears that propagation characteristics are key to quantifying model-to-model variations in predictability.



## 7. Conclusions

Through analysis of long control integrations with time invariant external conditions, we have quantified the predictability properties of six climate models to 1) help determine whether it is worthwhile to initialize decadal predictions with the observed state of the climate system, 2) assess how much model-to-model variability there is in the initial-value predictability limits of comprehensive AOGCMs, and 3) set a lower bound on the *uncertainty* of AOGCM estimates of the predictability properties of nature.

Our calculations indicate that by basinwide measures, for a typical initial condition and for a typical model, predictability in upper-ocean conditions resulting from initializing with a specific state nominally lasts for approximately a decade in the North Atlantic and somewhat less in the North Pacific. Of course these limits depend on how one defines “nominal.” Here we have arbitrarily defined it to mean that a forecast distribution has as much information as is contained in the spread of a distribution whose variance is 60% of the climatological distribution.

A second finding of our study is that for a given AOGCM the predictability properties of the upper ocean vary geographically by large amounts within the North Pacific and North Atlantic basins. For example, the relative entropy in forecasts at the decadal range varied by factors of 10 within each basin in five of the six models that we investigated. However, we found less variability among the predictability of prominent modes of a given model than one might expect in that the leading mode in a basin is not necessarily substantially more predictable than a generic pattern.

Given the presence of regional predictability at ranges well beyond a few years, even in the model with least predictability, our investigation serves to further substantiate the implication of other studies that there is merit in devoting resources to develop the capability to initiate decadal range forecasts with the observed state of the climate system.

On the other hand, our results indicate that quantification of the value in added skill that is potentially realizable as a result of initializing decadal forecasts is highly uncertain because there are large model-to-model variations in ocean predictability. For example, the typical nominal basinwide limits referred to above are actually midpoints in a range of values between 4 and 10 yr for the North Pacific and 5 and 19 yr for the North Atlantic, depending on the model chosen. Similarly, for some locations some models can differ from each other in the information content of their decadal forecasts by as much as a factor of 10. And there are also wide model-to-model

differences in the structure and predictability of the leading basin modes.

That initial-value predictability varies a great deal from model to model and from location to location was anticipated by our analysis of the intrinsic variability of each model in that we found very substantial model and geographical differences in the amplitude and temporal characteristics of decadal time-scale fluctuations. It turned out that the most important model-to-model predictability differences arose from differences in how ensemble means evolved in their predictions rather than in the spread about prediction means. To a certain extent, contrasts in the evolution of means were a reflection of differing degrees of persistence in the models. However, even more important in distinguishing model predictability traits was how they treated the propagation of mean signals.

Given the large model-to-model differences that we have quantified, a second conclusion of our study is that any comprehensive program aimed at the decadal prediction problem needs to carefully assess the predictability of the model it uses and take this assessment into account when designing its prediction system and interpreting the resulting predictions. For example, the maximum range of predictions that rely on information in the initial state for skill could be bound in this way. Furthermore, because the observational record is short, estimates of the decadal predictability properties of nature from observations alone are inadequate, so model predictability must be used as a guide. For this reason, an additional implication of the model-to-model differences documented here is that currently there is great uncertainty as to the predictability properties of nature on decadal time scales.

Since our study implies that it is important to assess the predictability of every model used for decadal forecasts, we think our demonstration that systematic estimates of predictability can be determined without recourse to computationally expensive ensemble experiments is useful. Either through use of analogs or multivariate linear regression, estimates can be made that represent the average initial-value predictability of a model for essentially all initial conditions on the model's attractor. Both methodologies make it possible to systematically compare model predictability properties without using highly coordinated projects, though for a particular application one or the other has specific advantages.

Finally, we note that though large model-to-model differences like those we have found are likely to also exist for aspects of the model state that we have not considered, the specific predictability limits found here may turn out to be very different for other variables,

depths, regions, and phenomena. Of particular interest is whether the ocean predictability that we have documented carries over to atmospheric predictability. For CCSM3, Teng et al. (2011) found that the imprint of Atlantic Ocean predictability on the atmosphere was weak, but of course this aspect of predictability may also vary from model to model and is poorly understood in the observed climate system.

*Acknowledgments.* The authors are grateful to Drs. Wonsun Park and Arai Miki for providing the model data. They also appreciate Tim DelSole for pointing out to them the relationship expressed in (2) as well as the helpful comments that he made on an earlier version of this paper. Portions of this study were supported by the Office of Science (BER), U.S. Department of Energy, Cooperative Agreement DE-FC02-97ER62402 and DOE Grant DE-SC0005355 and by the National Science Foundation. MK was supported by the Innovative Program of Climate Change Prediction for the 21st Century, conducted by the Ministry of Education, Culture, Sports, Science and Technology of the Japanese government. JRK was supported by the Joint DECC/Defra Met Office Hadley Centre Climate Programme (GA01101). The research leading to the results with the Kiel Climate Model has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under Grant 212643.

## APPENDIX A

### Attractor Average Relative Entropy

Suppose  $\mathbf{s}_i$ , at times  $i = 1, \dots, T$ , are  $n$ -dimensional reduced state vectors from an AOGCM's control integration of length  $T$ , and let  $\mathbf{L}^\tau$  be the linear regression operator that predicts  $\mathbf{s}_{i+\tau}$  from  $\mathbf{s}_i$ . Then, using symbols from section 3a and defining  $\mathbf{x}^2 = \mathbf{x}\mathbf{x}^\top$  for vector  $\mathbf{x}$ ,

$$\begin{aligned} n &= \text{tr}(\boldsymbol{\sigma}_b^2 \boldsymbol{\sigma}_b^{-2}) = \text{tr} \left\{ \left[ \frac{1}{T - \tau - 1} \sum_{i=1}^{T-\tau} (\mathbf{s}_{i+\tau} - \mathbf{L}^\tau \mathbf{s}_i \right. \right. \\ &\quad \left. \left. + \mathbf{L}^\tau \mathbf{s}_i - \boldsymbol{\mu}_b \right)^2 \right] \boldsymbol{\sigma}_b^{-2} \right\} \\ &= \text{tr} \left\{ \left[ \frac{1}{T - \tau - 1} \sum_{i=1}^{T-\tau} (\mathbf{s}_{i+\tau} - \mathbf{L}^\tau \mathbf{s}_i)^2 \right] \boldsymbol{\sigma}_b^{-2} \right\} \\ &\quad + \frac{1}{T - \tau - 1} \sum_{i=1}^{T-\tau} \text{tr}[(\boldsymbol{\mu}_f - \boldsymbol{\mu}_b)^2 \boldsymbol{\sigma}_b^{-2}] \end{aligned}$$

Here we have used the fact that regression errors are uncorrelated with predictands, as well as our assumption

that regression gives perfect forecasts of the evolution of AOGCM means. Recognizing the identity  $\text{tr}(\mathbf{x}^2 \mathbf{A}) = \mathbf{x}^\top \mathbf{A} \mathbf{x}$  for matrix  $\mathbf{A}$ , (2) follows in the limit of large  $T$ .

## APPENDIX B

### Multivariate Linear Regression Model Errors

We construct models that predict future T0–300 states of an AOGCM by using standard multivariate linear regression (MLR) formulas. Lorenz (1977) has considered the problem of estimating the statistics of errors from MLR predictions. Suppose one has a finite dataset produced by a process and subdivides it into partition  $A_1$  and partition  $A_2$  and then constructs a regression operator based on  $A_1$ . If that operator is applied to  $A_1$ , the resulting errors will underestimate the true error variance, while the errors resulting from its application to  $A_2$  will be an overestimate. Lorenz has demonstrated, however, that the geometric mean of these two estimates will be a good approximation to the true error variance. (By “true error” we mean the error variance that would result if we used an infinite dataset produced by the process.) Our tests suggest that, with one modification, the Lorenz approach can be adapted to covariance estimation. If the two estimates of error variance are not too different from the true error, then their geometric and arithmetic means will be similar. But, an arithmetic mean has the advantage that it can also be applied to estimates of covariance, which may be negative.

Using the Lorenz method, but with arithmetic means, we estimate the true error covariances of MLR operators as follows. We first divide a control dataset into five equal partitions  $A_1, \dots, A_5$ . Then for each  $k$ , we construct an operator using data from all partitions except  $A_k$  and evaluate its errors and their covariances when it is applied 1) to those same four partitions and 2) to  $A_k$ . Averaging these 10 error covariances gives us our estimate of the true error covariances.

As a test of this procedure we apply it to CCSM3. The dotted curves in the bottom panels of Fig. 4 show the attractor average MSD errors for the five operators applied to dependent data. The dashed curves pertain to when they are applied to independent data. It is when one takes the average of these two curves (solid black curve) that the regression errors match the analog estimates.

As a further test of whether our approach gives error covariance estimates with acceptable errors, we divide each model's control dataset into two halves and repeat our calculation of basin values of relative entropy in Fig. 7 for each half. In all cases (not shown), values based on the two halves are very similar to each other and to the values in Fig. 7. For example, the year at which most

individual models reach saturation differ by no more than about one-half year. Even the most sensitive saturations times—namely, in the North Pacific for the models with the shortest control integrations (CCSM3 and CCSM4)—differ by less than a year for the two dataset halves.

## APPENDIX C

### Attractor Average Dispersion versus Signal Relative Entropy

If one considers the definition of relative entropy (1) and its decomposition into signal relative entropy and dispersion relative entropy proposed by Kleeman (2002), then attractor averages of the signal component are generally expected to be larger than attractor averages of the dispersion component. For, to the extent that projections of forecast distributions onto two different EOFs have vanishing covariance (which they must at long forecast range for a transitive dynamical system), (2) implies that total, signal, and dispersive relative entropy are each a simple function of the ratios of forecast PC variances and their climatological variance. When (1) is evaluated under these assumptions, signal becomes the increasingly dominant component of relative entropy as the variance ratio approaches unity. For example, when the variance ratio is 0.30 for each PC, signal is 58% of the total relative entropy; when the ratio is 0.60, it is 78%; and when the ratio is 0.90, it is 95%. Since relative entropy is invariant to linear transformations of the state vector, this result does not depend on the use of an EOF basis.

## REFERENCES

- Alexander, M., L. Matrosova, C. Penland, J. D. Scott, and P. Chang, 2008: Forecast Pacific SSTs: Linear inverse model predictions of the PDO. *J. Climate*, **21**, 385–402.
- Barlas, N., K. Josic, S. Lapin, and I. Timofeyev, 2007: Non-uniform decay of predictability and return of skill in stochastic oscillatory models. *Physica D*, **232**, 116–127.
- Boer, G. J., 2000: A study of atmosphere-ocean predictability on long time scales. *Climate Dyn.*, **16**, 469–477.
- , 2004: Long time-scale potential predictability in an ensemble of coupled climate models. *Climate Dyn.*, **23**, 29–44.
- Branstator, G., and H. Teng, 2010: Two limits of initial-value decadal predictability in a CGCM. *J. Climate*, **23**, 6292–6310.
- Collins, M., 2002: Climate predictability on interannual to decadal time scales: The initial value problem. *Climate Dyn.*, **19**, 671–692.
- , and B. Sinha, 2003: Predictability of decadal variations in the thermohaline circulation and climate. *Geophys. Res. Lett.*, **30**, 1306, doi:10.1029/2002GL016504.
- , and Coauthors, 2006: Interannual to decadal climate predictability in the North Atlantic: A multimodel-ensemble study. *J. Climate*, **19**, 1195–1203.
- Collins, W. D., and Coauthors, 2006: The Community Climate System Model version 3 (CCSM3). *J. Climate*, **19**, 2122–2143.
- Cover, T., and J. Thomas, 2006: *Elements of Information Theory*. Wiley, 776 pp.
- DelSole, T., 2004: Predictability and information theory. Part I: Measures of predictability. *J. Atmos. Sci.*, **61**, 2425–2440.
- , and M. K. Tippett, 2009a: Average predictability time. Part I: Theory. *J. Atmos. Sci.*, **66**, 1172–1187.
- , and —, 2009b: Average predictability time. Part II: Seamless diagnosis of predictability on multiple time scales. *J. Atmos. Sci.*, **66**, 1172–1187.
- , —, and J. Shukla, 2011: A significant component of unforced multidecadal variability in the recent acceleration of global warming. *J. Atmos. Sci.*, **68**, 909–926.
- Delworth, T. L., and Coauthors, 2006: GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *J. Climate*, **19**, 643–674.
- Gent, P. R., and Coauthors, 2011: The Community Climate System Model version 4. *J. Climate*, **24**, 4973–4991.
- Gordon, C., C. Cooper, C. A. Senior, H. Banks, J. M. Gregory, T. C. Johns, J. F. B. Mitchell, and R. A. Wood, 2000: The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dyn.*, **16**, 147–168.
- Griffies, S. M., and K. Bryan, 1997a: A predictability study of simulated North Atlantic multidecadal variability. *Climate Dyn.*, **13**, 459–488.
- , and —, 1997b: Predictability of North Atlantic multidecadal climate variability. *Science*, **275**, 181–184.
- Hasselmann, K., 1976: Stochastic climate models part I. Theory. *Tellus*, **28**, 473–485.
- Kleeman, R., 2002: Measuring dynamical prediction utility using relative entropy. *J. Atmos. Sci.*, **59**, 2057–2072.
- Leith, C. E., 1975: Climate response and fluctuation dissipation. *J. Atmos. Sci.*, **32**, 2022–2026.
- Levitus, S., J. I. Antonov, T. P. Boyer, R. A. Locarnini, H. E. Garcia, and A. V. Mishonov, 2009: Global ocean heat content 1955–2008 in light of recently revealed instrumentation problems. *Geophys. Res. Lett.*, **36**, L07608, doi:10.1029/2008GL037155.
- Lorenz, E. N., 1969: Atmospheric predictability as revealed by naturally occurring analogues. *J. Atmos. Sci.*, **26**, 636–646.
- , 1977: An experiment in nonlinear statistical weather forecasting. *Mon. Wea. Rev.*, **105**, 590–602.
- Majda, A. J., R. V. Abramov, and M. J. Grote, 2005: *Information Theory and Stochastics for Multiscale Nonlinear Systems*. CRM Monogr., Vol. 25, American Mathematical Society, 133 pp.
- Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis, 1997: Pacific interdecadal climate oscillation with impacts on salmon production. *Bull. Amer. Meteor. Soc.*, **78**, 1069–1079.
- Meehl, G. A., J. M. Arblaster, and W. G. Strand, 1998: Global scale decadal climate variability. *Geophys. Res. Lett.*, **25**, 3983–3986.
- , and Coauthors, 2009: Decadal prediction: Can it be skillful? *Bull. Amer. Meteor. Soc.*, **90**, 1467–1485.
- , A. Hu, and C. Tebaldi, 2010: Decadal prediction in the Pacific region. *J. Climate*, **23**, 2959–2973.
- Msadek, R., K. W. Dixon, T. L. Delworth, and W. Hurlin, 2010: Assessing the predictability of the Atlantic meridional overturning circulation and associated fingerprints. *Geophys. Res. Lett.*, **37**, L19608, doi:10.1029/2010GL044517.
- Newman, M., 2007: Interannual to decadal predictability of tropical and North Pacific sea surface temperatures. *J. Climate*, **20**, 2333–2356.
- Nozawa, T., T. Nagashima, T. Ogura, T. Yokohata, N. Okada, and H. Shioyama, 2007: Climate change simulations with a coupled

- ocean-atmosphere GCM called the Model for Interdisciplinary Research on Climate: MIROC. CGER Rep. CGER-I073-2007, 93 pp. [Available online at <http://www.cger.nies.go.jp/publications/report/i073/I073.html>.]
- Park, W., and M. Latif, 2010: Pacific and Atlantic multidecadal variability in the Kiel Climate Model. *Geophys. Res. Lett.*, **37**, L24702, doi:10.1029/2010GL045560.
- Pohlmann, H., and Coauthors, 2004: Estimating the decadal predictability of a coupled AOGCM. *J. Climate*, **17**, 4463–4472.
- Power, S., and R. Colman, 2006: Multiyear predictability in a coupled general circulation model. *Climate Dyn.*, **26**, 247–272.
- Randall, D. A., and Coauthors, 2007: Climate models and their evaluation. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds., Cambridge University Press, 589–662.
- Saravanan, R., and J. C. McWilliams, 1998: Advective ocean-atmosphere interaction: An analytical stochastic model with implications for decadal variability. *J. Climate*, **11**, 165–188.
- Schneider, T., and S. M. Griffies, 1999: A conceptual framework for predictability studies. *J. Climate*, **12**, 3133–3155.
- Teng, H., and G. Branstator, 2011: Initial-value predictability of prominent modes of North Pacific subsurface temperature in a CGCM. *Climate Dyn.*, **36**, 1813–1834, doi:10.1007/s00382-010-0749-7.
- , —, and G. A. Meehl, 2011: Predictability of the Atlantic overturning circulation and associated surface patterns in two CCSM3 climate change ensemble experiments. *J. Climate*, **24**, 6054–6076.
- Van den Dool, H., 1994: Searching for analogues, how long must we wait? *Tellus*, **46A**, 314–324.
- von Storch, H., and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research*. Cambridge University Press, 484 pp.