

## A Predictability Measure Applied to Seasonal Predictions of the Arctic Oscillation

YOU MIN TANG

*Environmental Science and Engineering, University of Northern British Columbia, Prince George, British Columbia, Canada*

HAI LIN

*RPN, Meteorological Service of Canada, Dorval, Quebec, Canada*

JACQUES DEROME

*Department of Atmospheric and Oceanic Sciences, McGill University, Montreal, Quebec, Canada*

MICHAEL K. TIPPETT

*IRI, Columbia University, Palisades, New York*

(Manuscript received 8 May 2006, in final form 7 February 2007)

### ABSTRACT

In this study, ensemble seasonal predictions of the Arctic Oscillation (AO) were conducted for 51 winters (1948–98) using a simple global atmospheric general circulation model. A means of estimating a priori the predictive skill of the AO ensemble predictions was developed based on the relative entropy ( $R$ ) of information theory, which is a measure of the difference between the forecast and climatology probability density functions (PDFs). Several important issues related to the AO predictability, such as the dominant precursors of forecast skill and the degree of confidence that can be placed in an individual forecast, were addressed. It was found that  $R$  is a useful measure of the confidence that can be placed on dynamical predictions of the AO. When  $R$  is large, the prediction is likely to have a high confidence level whereas when  $R$  is small, the prediction skill is more variable. A small  $R$  is often accompanied by a relatively weak AO index. The value of  $R$  is dominated by the predicted ensemble mean. The relationship identified here, between model skills and the  $R$  of an ensemble prediction, offers a practical means of estimating the confidence level of a seasonal forecast of the AO using the dynamical model.

Through an analysis of the global sea surface temperature (SST) forcing, it was found that the winter AO-related  $R$  is correlated significantly with the amplitude of the SST anomalies over the tropical central Pacific and the North Pacific during the previous October. A large value of  $R$  is usually associated with strong SST anomalies in the two regions, whereas a poor prediction with a small  $R$  indicates that SST anomalies are likely weak in these two regions and the observed AO anomaly in the specific winter is likely caused by atmospheric internal dynamics.

### 1. Introduction

The Arctic Oscillation (AO) is the dominant mode of monthly mean sea level pressure variability over the Northern Hemisphere with an out-of-phase relation between the sea level pressure over the Arctic Basin and at the midlatitudes (Thompson and Wallace 1998). The AO has a close association with the North Atlantic Os-

cillation (NAO) due to its strong manifestation over the Atlantic sector. The interannual and longer-term changes in the wintertime AO have an enormous impact on the climate of the Northern Hemisphere (e.g., Thompson and Wallace 2001). The NAO has long been recognized as the major circulation pattern influencing the weather from the eastern North America to Europe (e.g., Greatbatch 2000). The seasonal climate prediction skill in the Northern Hemisphere to a great extent relies on the predictive capability of major atmospheric modes of monthly mean variability, for example, the AO and the Pacific–North American (PNA; Derome et al. 2005). To explore the AO predictability is thus a

---

*Corresponding author address:* Youmin Tang, Environmental Science and Engineering, University of Northern British Columbia, 3333 University Way, Prince George, BC V2N 4Z9, Canada.  
E-mail: ytang@unbc.ca

critical step toward fully understanding the predictability of the climate seasonal prediction.

An important issue in predictability studies is the uncertainty of predictions. Climate seasonal predictions apply for several months and the decisions made in response to them often are economically significant. Therefore estimates of the uncertainty of prediction are highly desirable. On the other hand, climate seasonal prediction is still at its early stage compared with the numerical weather prediction (NWP) and the El Niño–Southern Oscillation (ENSO) prediction, and its skill is relatively low. Therefore, a study of uncertainties in seasonal climate predictions is especially important at present.

The technique used in predictability studies in NWP has primarily been ensemble prediction, in which a priori likely skill (or usefulness) for an individual prediction might be estimated by the ensemble spread (e.g., Buizza and Palmer 1998). However, little connection was found between the ensemble spread and the prediction skill in some dynamical models (Kumar et al. 2000; Tang et al. 2005). Instead, in some studies, an alternate criterion that has been used as a predictor of forecast skill is the leading eigenmode amplitude (signal size) of the forecast initial conditions (Kleeman and Moore 1999; Tang et al. 2005), which essentially represents the contribution of persistence to the predictive skill (von Storch and Xu 1990; von Storch and Baumhefner 1991). When climate variability modes are present with larger amplitudes, they are more likely to be able to “resist” dissipation by the chaotic or stochastic components of the system, making them more predictable.

Recently, a new theoretical framework for measuring the uncertainty of predictions has been developed and applied to examine ENSO and seasonal climate predictability (Schneider and Griffies 1999; Kleeman 2002; Tippett et al. 2004; Tang et al. 2005; DelSole 2004, 2005; DelSole and Tippett 2007). The approach is built on information theory (Cover and Thomas 1991). It has been argued that the relative entropy ( $R$ ), defined by the differences between the climatological probability density functions (PDFs) and the prediction PDF, can explain well why the two reliability measures discussed above are central to predictability studies (Kleeman 2002; Tang et al. 2005). In particular, when the PDFs are Gaussian,  $R$  consists of two components: one is the dispersion component associated with the ensemble spread and the other one is the signal component related to the leading eigenmode amplitudes present in the initial conditions or forced by the boundary conditions.

In this paper, we will apply the relative entropy

method to estimate the degree of confidence of the AO predictions performed by a reasonably skillful atmospheric general circulation model. Of special interest in this paper are the appropriate measures of the confidence of AO dynamical predictions and the dominant precursors that control variations in the measures.

## 2. Model and ensemble prediction

The model used in this study is the simple global atmospheric circulation model (SGCM), initially designed by Hoskins and Simmons (1975), and then further developed by Hall (2000). It is a primitive equation dry atmospheric model, and has a global domain with horizontal resolution of T21 and five levels in the vertical. A detailed description of the model may be found in Hall (2000) and Hall et al. (2001a,b). An important feature of this model is that it uses an empirical forcing calculated from observed daily data. By computing the dynamical terms of the model, together with a linear damping, with daily global analyses and averaging in time, the residual term for each time tendency equation is obtained as the forcing. The collective effect of these forcing terms represents all processes that are not resolved by the model’s dynamics such as diabatic heating (including latent heat release related to the transient eddies) and the deviation of dissipative processes from linear damping. This atmospheric model has been used for seasonal predictions, and was found to be similar in prediction skill to a more complex GCM (Derome et al. 2005).

Global ensemble forecasts were made for the 51 boreal winters [December–January–February (DJF)] from 1948/49 to 1998/99 with an ensemble size of 70. The initial conditions for the seasonal forecasts were the 0000 UTC 1 December analyses from the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR; Kalnay et al. 1996). Each ensemble run was constructed by adding to the initial conditions a small-amplitude perturbation pattern, which is the scaled down anomaly (with respect to the 51-yr winter climatology) of a random winter day in the 51-yr NCEP–NCAR dataset (excluding the winter being predicted).

For each winter, a time-independent forcing is used, that is obtained from the daily data of the NCEP–NCAR reanalysis. The approach of a persistent forcing anomaly is applied. For a given winter, the forcing was obtained as the November-mean-forcing anomaly of that year added to the DJF-mean climatological forcing. The calculation of the climatology was done in the framework of cross validation (i.e., the winter to be predicted was excluded). A more detailed description

of the forcing specification may be found in Derome et al. (2005).

The skill of the ensemble mean prediction has been evaluated in detail by Derome et al. (2005). It was found that the SGCM has a statistically significant skill in forecasting the AO variability, actually even better than that of a more complex GCM (e.g., the Canadian GCM2). In the present study, the AO is defined as the leading empirical orthogonal function (EOF) mode of the wintertime (DJF) mean sea level pressure anomalies (MSLPA) north of 20°N from the NCEP–NCAR reanalysis. The observed and each individual forecast DJF MSLPA field over the 51 winters are projected onto the AO pattern to obtain the corresponding observed and SGCM-predicted principal component time series (i.e., AO indices). These indices are used in the following discussion.

### 3. Relative entropy and predictability

Suppose that the AO index is modeled as a random variable  $T$  whose climatological or equilibrium PDF is  $q(T)$ . In many practical situations there is considerable knowledge of the climatological PDF from long-term historical observations. We will use a perfect model approach (i.e., we will assume that the “observed” state of the atmosphere can be any one of the ensemble members predicted by the SGCM). The climatological PDF is then obtained from the model forecasts over all the 51 winters. The ensemble prediction of the SGCM for a given winter produces a forecast PDF, denoted  $p(T)$ . The extent to which the forecast and climatological distributions differ is an indication of a potential predictability. There is, of course, no predictability when the forecast and climatological distributions are identical. A useful measure of the difference between  $q(T)$  and  $p(T)$  from information theory (Cover and Thomas 1991) is the relative entropy  $R$  or Kullback–Leibler *distance* between the two PDFs, defined as

$$R = \int P(T) \log \frac{p(T)}{q(T)} dT. \quad (1)$$

The quantity  $R$  measures the informational inefficiency of using the climatological rather than the forecast PDF and  $R \geq 0$  with equality if and only if  $p = q$  (Cover and Thomas 1991). The relative entropy can be used as an indicator of predictability, or prediction utility, in that it measures the additional utility of the ensemble prediction as compared with a climatological prediction. Larger values of  $R$  indicate that potentially more useful information is being supplied by a prediction.

The notion that  $R$  is a measure of the predictability

can also be interpreted in the Bayesian framework. From the Bayesian perspective, the climatological distribution is a *prior* distribution derived from previous observations. A prediction augments this prior information, and the additional information provided by the prediction constitutes the prediction PDF, which should be referred to as a posterior distribution in the Bayesian terminology. The  $R$  quantifies the amount of information that  $p$  provides beyond  $q$ . In other words, in a perfect model framework, the extent to which this prediction PDF *differs* from the original prior is a measure of the usefulness of the prediction. In practice,  $p$  and  $q$  can be approximated using kernel density estimation, and the integral in (1) approximated by a discrete sum.<sup>1</sup>

If the PDFs are Gaussian, the relative entropy  $R$  may be expressed exactly in terms of the prediction variance  $\sigma_p^2$ , the model climatological variance  $\sigma_q^2$  (perfect model framework), and the difference  $\mu^p - \mu^q$  of the ensemble and the climatological means (Kleeman 2002):

$$R = \frac{1}{2} \left[ \log \frac{\sigma_q^2}{\sigma_p^2} + \frac{\sigma_p^2}{\sigma_q^2} + \frac{(\mu^p - \mu^q)^2}{\sigma_q^2} - 1 \right]. \quad (3)$$

The first two terms on the rhs of (3) are determined by the climatological variance and prediction variance, and represent the contribution of the dispersion or spread of the ensemble to  $R$ . The third term on the rhs of (3) is governed by the amplitude of the predicted ensemble mean and measures the contribution of the predicted signal size to  $R$ . The first two terms minus 1 is referred to as the dispersion component (DC) and the third term as the signal component (SC; Kleeman 2002; Tang et al. 2005). Therefore, for Gaussian distributions  $R = DC + SC$  both ensemble spread and signal size are incorporated into the relative entropy  $R$ . The DC contributes to  $R$  when the prediction variance is different from the climatological variance, and the SC contributes when the mean of the prediction distribution differs from that of the climatological distribution.

In this study, we use (3) to calculate  $R$  since the Gaussian assumption holds reasonably well for all prediction cases (see the next section). For a non-Gaussian system,  $R$  should be computed directly from (1). Estimating the PDF for a non-Gaussian system is an inter-

<sup>1</sup> If  $T$  is divided into  $n$  bins,  $R$  could be approximated by

$$R = \sum_{i=1}^n f_i^p \log \frac{f_i^p}{f_i^q}, \quad (2)$$

where  $f_i^p$  and  $f_i^q$  are the prediction frequency and the climatological frequency in bin  $i$ .

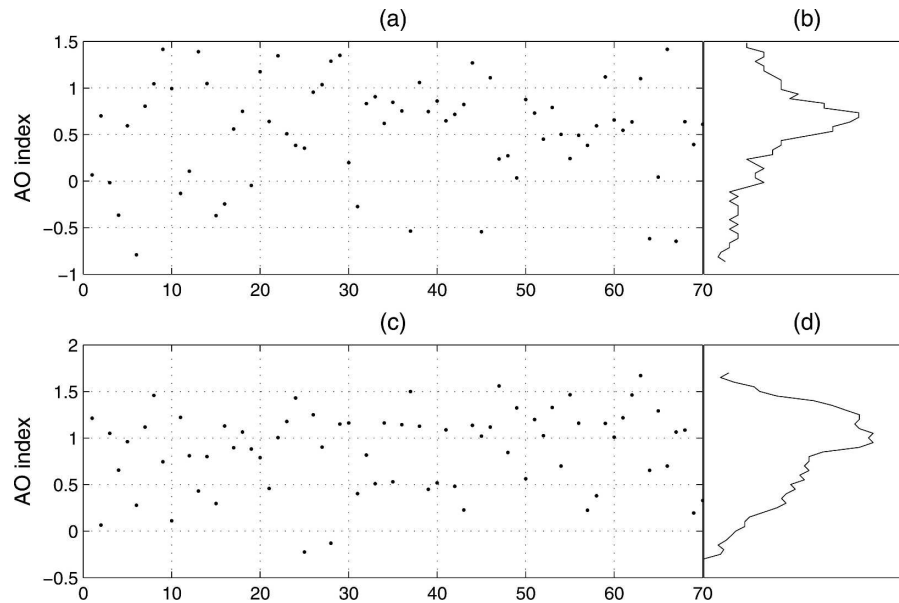


FIG. 1. (a), (c) Two arbitrarily chosen ensemble prediction cases, each with 70 members. (b), (d) The estimated probability distribution of the two ensemble predictions.

esting problem in its own right, especially when the number of ensemble members is small (Kleeman and Majda 2005).

#### 4. The relationship between prediction utility and prediction skill

The Gaussian assumption is first examined in order to use (3). Figure 1 is an estimate of the PDF for two ensemble predictions of the AO index, from randomly chosen winters. Figure 1 indicates that the Gaussian assumption roughly holds for both cases. An examination of all ensemble predictions produced similar results (not shown). The Kolmogorov–Smirnov normality test (DeGroot 1991) shows that all ensemble predictions pass the test at the significance level of 0.1.

Displayed in Fig. 2a are the variations of the AO prediction utility  $R$  for 51 winters during 1948–98, as a function of time. The climatological mean  $\mu_q$  and variance  $\sigma_q^2$  are estimated from all ensemble members and years (sample size is 3570) as in Tippett et al. (2004).<sup>2</sup> The prediction mean  $\mu_p$  and variance  $\sigma_p^2$  are estimated each winter from the 70-member ensemble. As can be seen, it is apparent that a large prediction utility  $R$  is found in a few predictions such as those of 1955, 1959, 1975, 1983, and 1994. For many other predictions,  $R$  is small.

<sup>2</sup> Alternatively they were also estimated from the observation, leading to similar results.

When prediction and climatology distributions are identical, the relative entropy  $R$  is zero from (1). In theory, a nonzero value of  $R$  indicates predictability. However, in practice, a finite sample size introduces sampling errors that lead to a nonzero  $R$  even though there is no extra information supplied by the prediction. Therefore the statistical significance level should exceed the extent of uncertainty due to the finite sample size. We quantify the extent of uncertainty using a Monte Carlo method as in Tippett et al. (2004). A sample with 70 members is randomly drawn from the climatology distribution and its relative entropy  $R$  is computed with respect to the climatology distribution. This process is repeated 10 000 times, and the value above 95% of 10 000  $R$  is considered to be the significant level as shown in Fig. 2a (solid line). During the 51 winter AO predictions, 44 predictions have a significant relative entropy, accounting for 86% of all predictions.

Figure 2b shows the absolute error of each ensemble mean prediction. A comparison of Fig. 2b with 2a reveals that a large  $R$  is often associated with a good prediction skill (i.e., small absolute error) whereas when  $R$  is small, the skill tends to be lower. This is very similar to a so-called triangular relation that was used to characterize the relationship between ensemble spread and skill in ensemble NWP (e.g., Buizza and Palmer 1998) and in ENSO models (e.g., Xue et al. 1997; Moore and Kleeman 1998); namely, when the ensemble spread is small, the skill is good whereas when it is large, the skill is much more variable. Thus, we also use the “triangular relation” to describe the

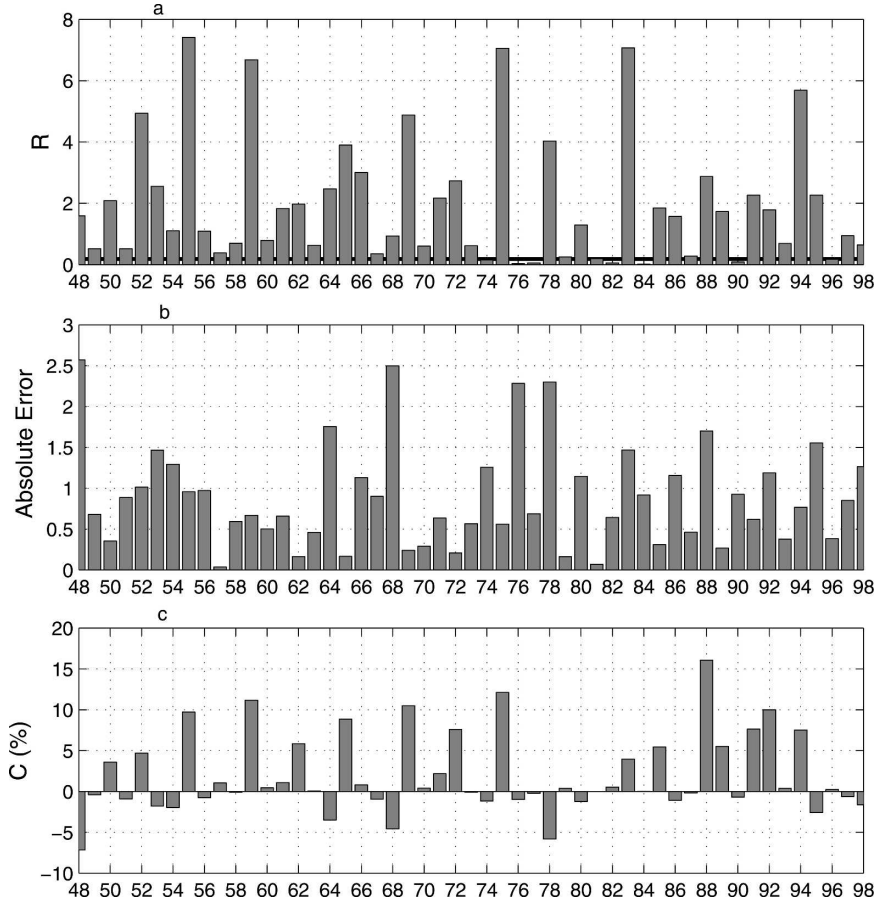


FIG. 2. (a) Relative entropy  $R$  as a function of time for the 51 winters from 1948 to 1998; (b) same as (a), but for absolute error; (c) same as (a), but for the contribution of each prediction  $C$  to the correlation skill. Solid line in (a) shows the Monte Carlo test at 95% confidence level.

relationship between  $R$  and the predictive skill. It should be noticed that we examined the ensemble spread and the absolute error for this SGCM, and did not find a significant relationship between them as shown in Fig. 3. In fact, Fig. 3 shows a sometime inverse relationship between the ensemble spread and the absolute error (i.e., when the ensemble spread is large the absolute error might be small, and vice versa, indicating that the ensemble spread is not a good indicator of the AO prediction skill for the SGCM).

From Figs. 2a,b, one can find that some small  $R$ s are associated with small absolute errors. This is interesting since a small  $R$  suggests little extra information to be provided by the prediction. To explore this, we examined all predictions with absolute errors smaller than 1.0 and with  $R$  smaller than 1.0 (16 cases altogether). It was found that all of these cases have relatively weak AO anomalies in the observations and ensemble mean, as shown in Fig. 4, leading to small absolute errors. On the other hand, a weak ensemble mean AO anomaly

suggests its status approaching its climatology, leading to a small  $R$  based on (2).

To further explore the relation between  $R$  and the prediction skill, we examine the contribution of each prediction to the correlation skill  $r$ , traditionally defined as

$$r = \frac{1}{N} \sum_i T_i^p T_i^o, \quad (4)$$

where  $T$  denotes the normalized AO index with zero mean,  $i$  is for year,  $p$  is for predictions,  $o$  is for observations, and  $N$  is the number of samples used to calculate  $r$ . In the case of the predictions,  $T^p$  refers to the ensemble mean.

The contribution of each prediction to  $r$ , denoted as  $C_i$ , can be measured by

$$C_i = \frac{\frac{1}{N} T_i^p T_i^o}{r} \times 100\%. \quad (5)$$

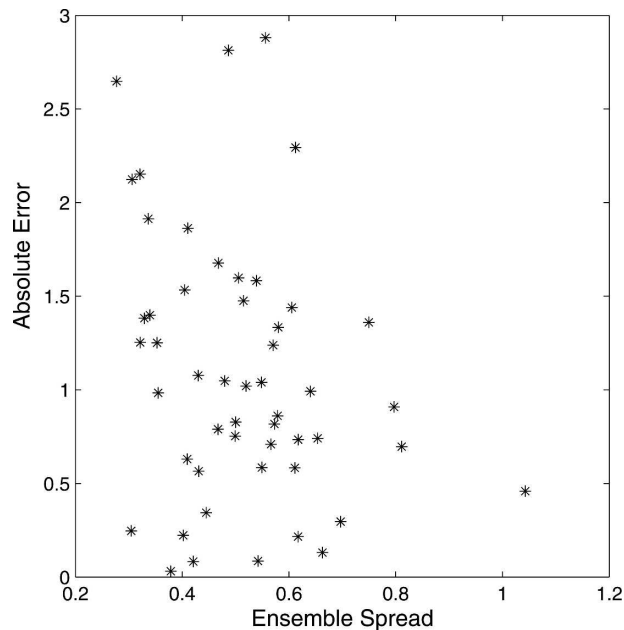


FIG. 3. Scatterplot of ensemble spread and the absolute error of the ensemble mean prediction.

Figure 2c shows variations of  $C$  with time. A feature shown in Fig. 2c is that there is a large variation of  $C$  among the predictions. While some winters have good predictions that account for significant contributions to

$r$ , others have a very small  $C$ . A comparison of Fig. 2c with 2a reveals that a large  $C$  generally corresponds to a large  $R$  except the 1988 case. The year 1988 has the strongest AO activity during the 51 winters with an AO index [i.e.,  $T^{\circ}$  in (5)] of 2.85, leading to a very large  $C$  but only a moderate  $R$ . The correlation coefficient between  $C$  and  $R$  over the 51 predictions is 0.61, which is statistically significant at a confidence level of 1%. Such a good relation between  $C$  and  $R$  is especially obvious for large  $R$ . We calculated an accumulated  $C$  over predictions with the five largest  $R$ s (i.e.,  $R > 5.0$ ), and found that 44% of the correlation skill  $r$  came from the contribution of the five predictions. Table 1 shows correlation skills between predicted and observed AO indices, obtained using different samples classified by  $R$ . As can be seen, the predictions with a larger  $R$  lead to better skill than those with a smaller  $R$ , with a correlation of 0.8 and 0.7 for  $R$  greater than 3 and 2.5, respectively. It should be pointed out that it may be misleading to compute the correlation from subsets of data, in particular when the subset is made up of high-amplitude cases (von Storch and Zwiers 1999). Also as there are fewer samples used as  $R$  increases, it is possible that the change in sample size is responsible for this increase in skill with  $R$ . To evaluate this, we used a bootstrap method to measure the extent of the uncertainty in the computed correlation due to the finite

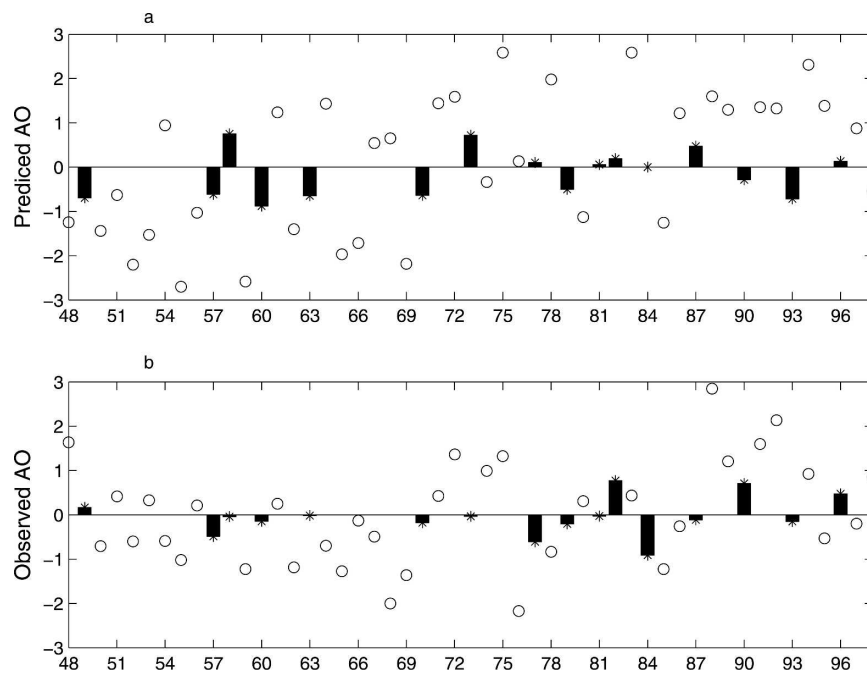


FIG. 4. (a) The bars with the asterisks are the predicted ensemble mean AO index with absolute error smaller than 1.0 and  $R$  smaller than 1.0. The circles denote the predicted AO index for the remaining cases; (b) Same as (a), but for observed AO index.

TABLE 1. Correlation skill between predicted ensemble mean and observed AO indices as a function of  $R$  (the number shown in parentheses is the number of samples used).

$R_0$	$R > R_0$	$R \leq R_0$
0.1	0.41 (51)	—
1.0	0.55 (30)	-0.30 (21)
1.5	0.60 (25)	-0.35 (26)
2.0	0.66 (18)	-0.42 (33)
2.5	0.70 (13)	0.16 (38)
3.0	0.80 (10)	0.24 (41)

sample size.<sup>3</sup> The results show that the increase in the correlation skill in Table 1 results from the contribution of more skillful predictions with larger  $R$ , rather than from the uncertainty of the finite sample size (Fig. 5).

### 5. The dominant component controlling $R$

We have explored the relationship between the prediction utility and the model prediction skill. Our results show that the prediction utility  $R$  is a good indicator for the AO prediction skill. A “triangular” relationship can be suggested between  $R$  and the model skill. When  $R$  is large, the prediction is typically good, whereas when  $R$  is small, the prediction skill is much more variable. A small  $R$  is often accompanied with a relatively weak AO. Next, we will examine what determines the variations in  $R$ .

As discussed in section 3,  $R$  is the sum of a DC and an SC. Figures 6a,b depict the scatterplots of  $R$  with SC and DC for the period of 51 winters. The figures show that SC is significantly larger than DC, and dominates  $R$ . As can be seen,  $R$  and SC vary linearly with a slope of unity. The correlation coefficient between  $R$  and SC is 0.99. In contrast to the good relation between SC and  $R$ , however, the relation between DC and  $R$  is much less significant.

A further examination of SC and DC reveals that SC and DC are highly related with two widely used variables in ensemble prediction: the ensemble mean and the ensemble spread (i.e., ensemble variance). Shown in Figs. 6c,d are the scatterplots of SC with the ensemble mean squared, and of DC with the ensemble spread. Figure 6c shows a near-perfect relationship between SC and the square of the ensemble mean, with the correlation of 0.99. A significant negative correlation also exists between DC and ensemble spread in

<sup>3</sup> A 1000-member ensemble correlation was computed. Each correlation was obtained using randomly taken sample pairs of predicted and observed AO indices with the same sample size as that used in Table 1. The standard deviation of ensemble correlation was used to represent the extent of the uncertainty.

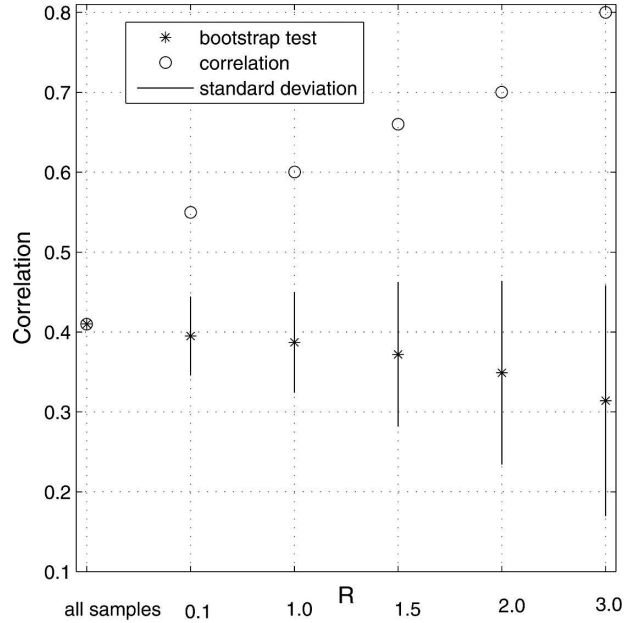


FIG. 5. The circles denote the correlation between the predicted and observed AO indices for the predictions classified by  $R$  (see Table 1); the asterisks denote the 1000-member ensemble correlation mean (see text); and the vertical lines are the uncertainty extent of the correlation estimate (std dev) due to the finite sample size estimated by the bootstrap method.

Fig. 6d. Here DC is composed of two terms on the rhs of (3), that is,  $\log(\sigma_q^2/\sigma_p^2)$  and  $\sigma_p^2/\sigma_q^2$ ; DC is inversely proportional to the ensemble spread in the first term but positively proportional to it in the second term. Thus, a negative correlation between DC and the ensemble spread is mainly due to much more contribution of the first term to DC. This can typically occur when the ensemble spread is very small. An examination of all ensemble predictions showed that out of the 51 ensemble predictions, 50 have the ensemble spread smaller than 1, and 43 smaller than 0.4. There are most likely two factors to cause such small ensemble spreads here: 1) the nonlinearity of the SGCM is relatively weaker compared with observation and more complex GCMs, especially since all ensemble members (for a given winter) use the same forcing; and 2) the perturbation patterns used to construct the ensemble predictions are not optimal like singular vectors or breeding vectors that can lead to the fastest growth of model errors.

Since  $R$  is mainly controlled by SC, Fig. 6c suggests that the prediction utility  $R$  contains information from both the ensemble mean and ensemble spread, but is dominated by the prediction ensemble mean. Therefore, the prediction skill is highly associated with the predicted ensemble mean amplitude of the AO index. When the predictive mean signals are large (due to

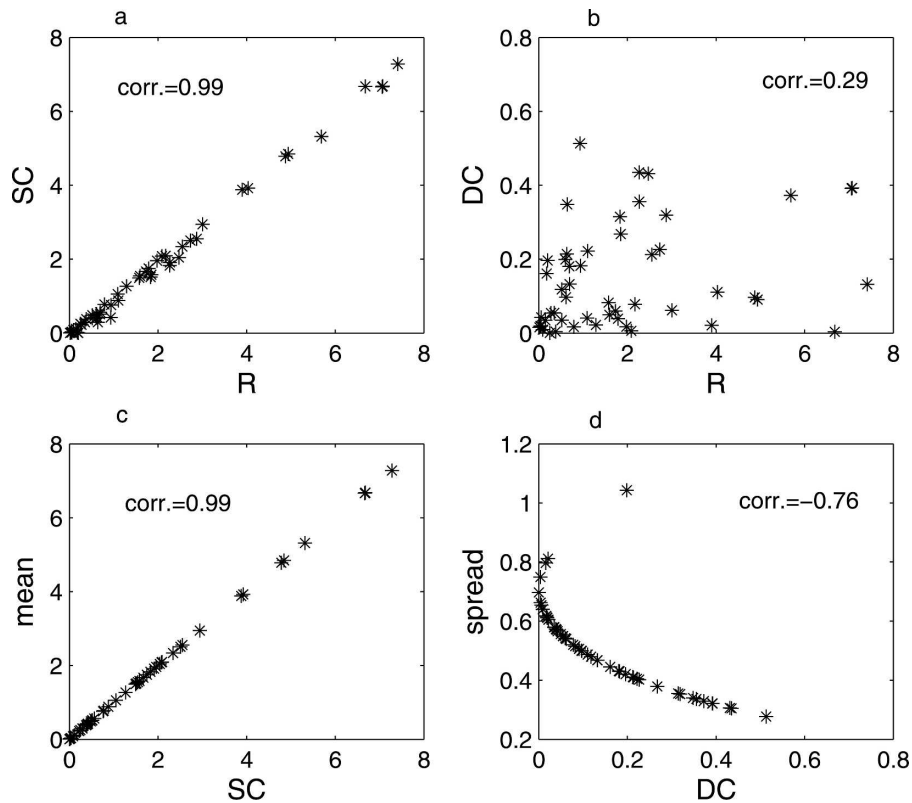


FIG. 6. Scatterplots of relative entropy  $R$  with (a) signal component  $SC$  and (b) dispersion component  $DC$ . Scatterplot of (c)  $SC$  with the square of ensemble mean and (d)  $DC$  with ensemble spread (standard deviation).

large-amplitude forcings),  $R$  is also large, suggesting that such predictions are more reliable than for small mean signals (weak forcing). Note that this result might be model dependent, and related to the fact that the SGCM has a small ensemble spread. Whitaker and Loughé (1998) found that the relation between spread and skill is strong when the variability of the ensemble variance is large. However the strong relation between the ensemble mean and the model skill was also found in some complex GCMs (Kumar et al. 2000; Tippett et al. 2004; Tang et al. 2005). The results are also consistent with the fact that mean winter forecasts over North America tend to be better for winters with strong ENSO forcing (e.g., Derome et al. 2005).

When variations in  $R$  are mainly due to variations in the ensemble mean, the correlation  $r$  between the ensemble mean and the observation is related to the expected correlation  $\rho_c$  between the correlation contribution  $C$  and the relative entropy  $R$  for normally distributed variables with constant variance. A theoretical relationship between  $r$  and  $\rho_c$  is (see the appendix)

$$\rho_c = \sqrt{\frac{2r^2}{1+r^2}}. \quad (6)$$

For the AO index,  $r = 0.41$  and the value of  $\rho_c$  predicted by (6) is 0.53, which is reasonably close to the observed correlation between the contribution  $C$  and relative entropy  $R$  of 0.61. Since the expected correlation is built on the assumptions that the variables follow a Gaussian distribution and  $R$  is proportional to the square of ensemble mean, the consistency between the expected correlation and the actual value supports the assumptions and the above analyses of relative entropy.

As discussed above,  $R$ , dominated by the signal component, is a good indicator of prediction reliability for strong AO events. This suggests that some possible relationship exists between  $R$  and the persistence of the AO index, since strong AO modes in the initial condition are more likely to be able to resist dissipation by the chaotic or stochastic components of the system, leading to better persistence and prediction skill. Thus, it is of interest to compare  $R$  with a simpler measure of predictability related to persistence. For simplicity, we define the square of the amplitude of the November AO index as a simple measure (SM) to quantify the reliability of a winter AO prediction. Figure 7a shows the variation of SM as a function of each prediction during the 51 winters from 1948 to 1998. Comparing



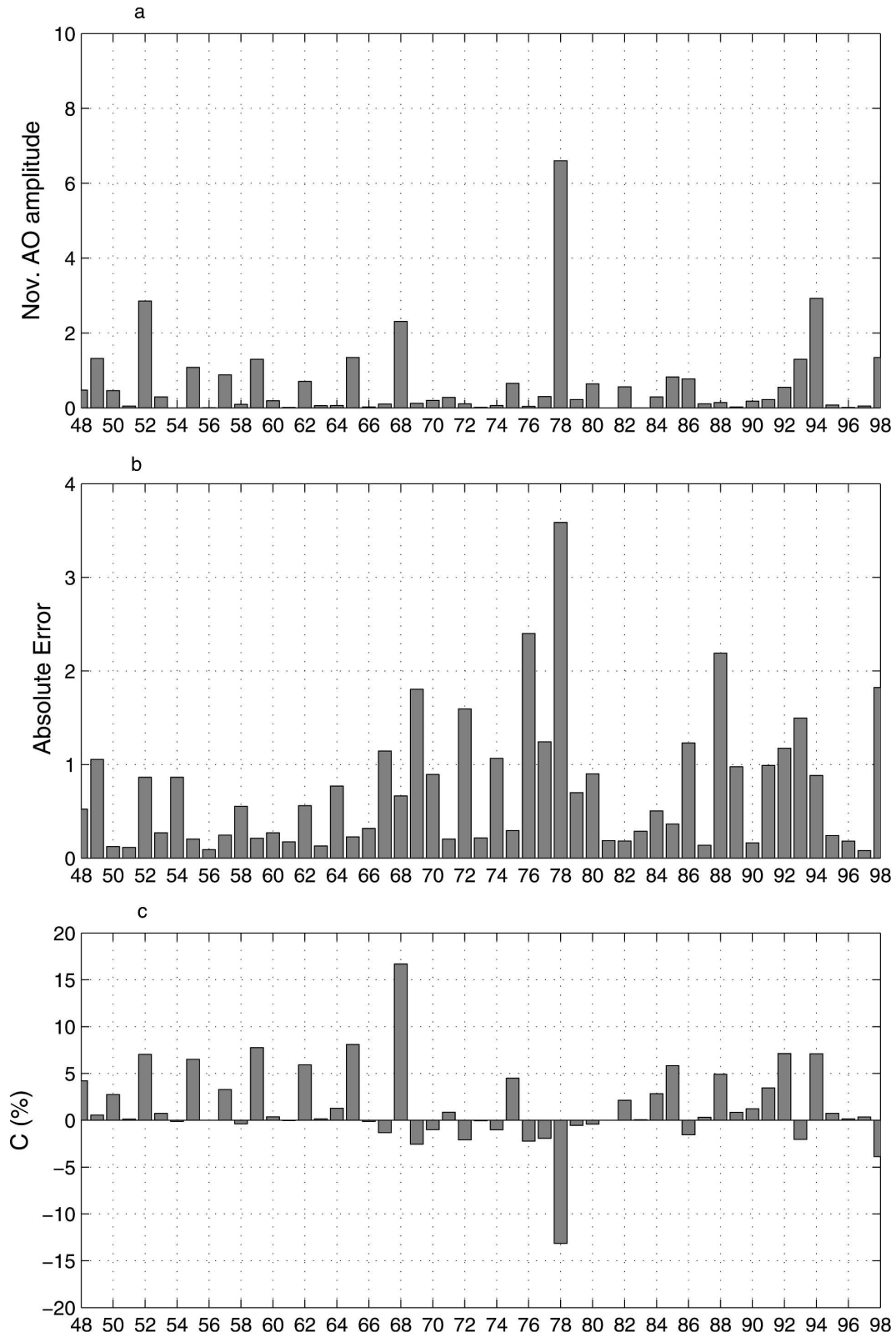


FIG. 7. (a) Same as Fig. 2a, but for the simple measure SM (i.e., the amplitude of November AO index). (b), (c) Same as in Figs. 2b,c, respectively, but for persistence prediction using the November AO index as the predicted value.

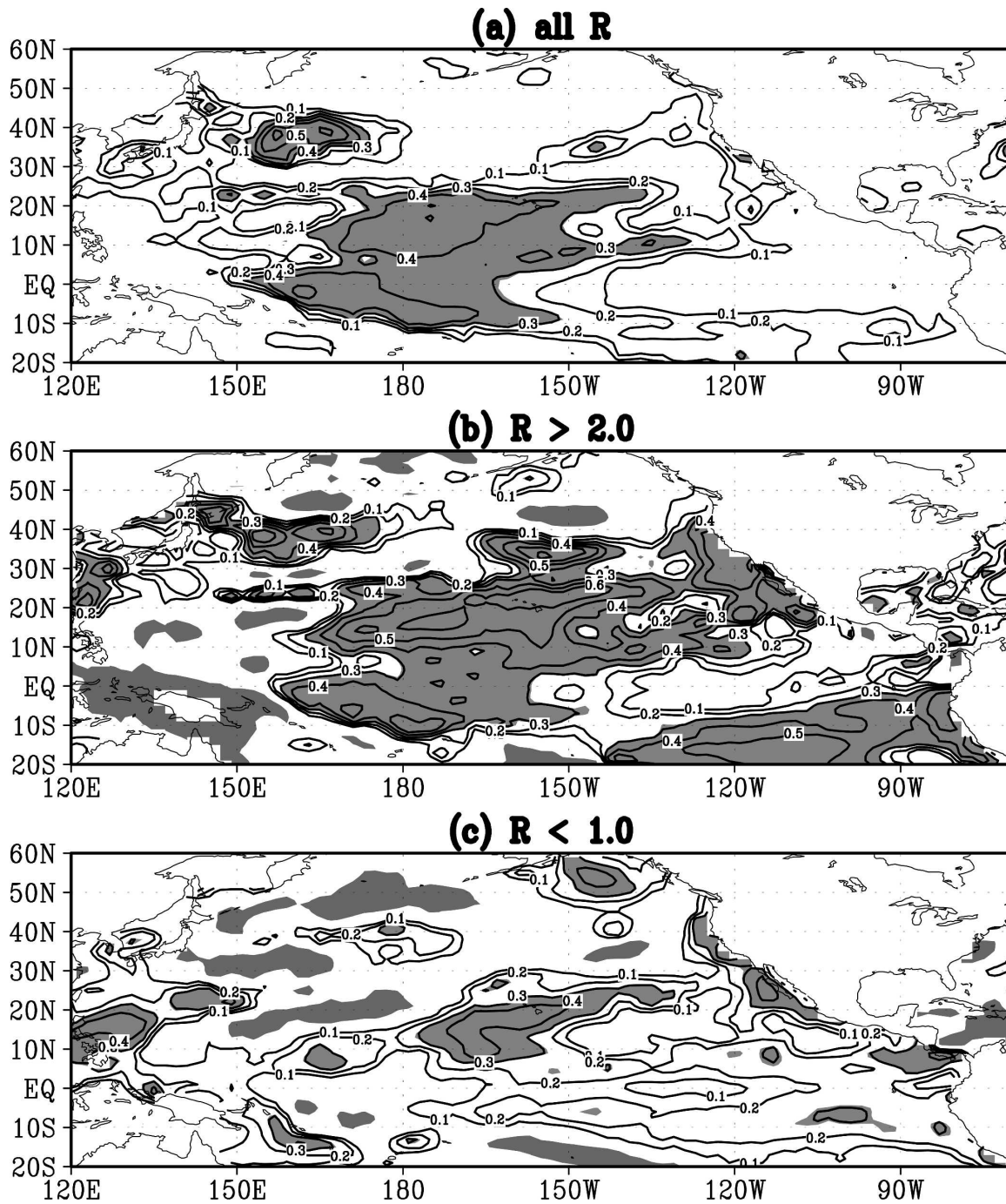


FIG. 8. (a) Correlation between prediction utility  $R$  and the October SST over all predictions. (b) Same as (a), but using the predictions with  $R > 2.0$ . (c) Same as (a), but  $R < 1.0$ . The shaded area is the correlation over 0.3.

Fig. 7a with Figs. 2b,c reveals that SM is unable to well quantify the reliability of winter AO prediction. The correlation between SM and  $C$  is only  $-0.06$ .

Figures 7b,c are skill scores of the persistence prediction using the November AO index. As can be seen, SM is not an effective measure of reliability for

persistence prediction. This is probably because the signal component dominating  $R$  is more related to the strength of the persisted model forcings than to the November AO itself. For example, Tang et al. (2005) found that the amplitude of subsurface ocean heat content, rather than the SST itself, is the best substitute of

$R$  to measure the reliability of ENSO predictions. Also, SM may be too simple to properly represent the persistence of the AO index. A more refined definition might be needed to quantify the persistence capability, such as the time series of leading Principal Oscillation Pattern (POP) modes (von Storch and Xu 1990; von Storch and Baumh€ofner 1991).

It should be noticed that persistence produces a significant (but low) prediction skill with a correlation of 0.37 between the predicted and observed AO index for the period 1948–98. In the next section, we will see that part of model prediction skill is likely due to this persistence.

## 6. The relationship between $R$ and SST forcing

As discussed in the preceding section, the prediction utility  $R$  is a good indicator of the AO prediction skill through quantitatively measuring the extra information of predictions. It is of interest to further explore the possible source of the extra information on the predictions in this SGCM. We thus turn to analyze the forcing that is related to the AO signal. As mentioned in section 2, the model forcing is expressed not only in the model's thermal equation, but also in vorticity, divergence, and surface pressure equations. To avoid the complexity, we look at the SST anomaly instead. The SST anomaly constitutes the most important signal in the atmospheric boundary conditions, and its variability is likely well associated with the model forcing that is empirically calculated using the observed data under such a boundary condition.

We calculated the correlation between  $R$  and the SST anomaly (SSTA) of each grid point for each month, from January to November before the prediction, over a global domain for the period during 1950–98. The observed SST from the Comprehensive Ocean–Atmosphere Data Set (COADS) dataset (Smith et al. 1996) was used. The prediction utility  $R$  in Eq. (2) is dominated by the square of the ensemble prediction mean, so that  $R$  is independent of the sign of the ensemble prediction mean. So, the square of SSTA ( $SSTA^2$ ) was used instead of the SST anomaly itself for computing the correlation to address the strength of the SST signal in influencing the reliability of the AO prediction. The results show that significant correlations appear in the tropical central Pacific and the North Pacific (NP), with the maximum in October, as shown in Fig. 8a. As can be seen, a large region of significant correlation coefficients resides in the tropical central Pacific. In the North Pacific around 40°N, there is another smaller but stronger region of significant correlations, appearing to the west of the date line. The cor-

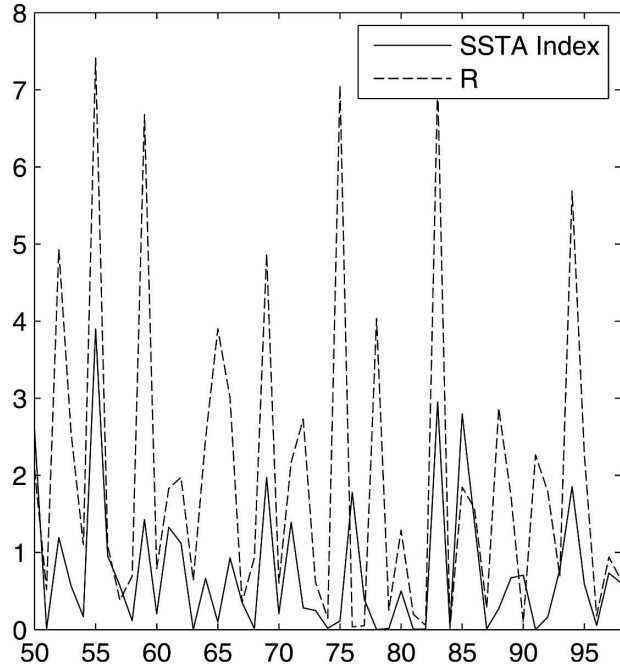


FIG. 9. Variation of  $SSTA^2$  index in the NP region (solid line) and prediction utility  $R$  (dashed line).

relation of  $R$  to the Niño-4  $SSTA^2$  index (averaged  $SSTA^2$  over 5°–5°N, 160°E–150°W) and to NP  $SSTA^2$  index (averaged  $SSTA^2$  over 35°–45°N, 150°E–180°) during 1950–98 is 0.41 and 0.56, respectively, both being statistically significant at a confidence level of 0.01. Figure 9 compares the variations of  $R$  and NP  $SSTA^2$  index, showing a good agreement between them.

Figures 8b,c are similar to Fig. 8a but with two different sample groups classified by  $R$  in calculating the correlation: one for  $R > 2.0$  and the other for  $R < 1.0$ . They indicate that the correlation coefficients depend on the prediction utility  $R$ . When  $R$  is large, the correlation is high whereas when  $R$  is small, the correlation is less significant. This suggests that a large prediction utility  $R$  that is likely to lead to a reliable prediction is well linked to a strong SST forcing in the tropical central and the North Pacific Ocean, whereas a poor prediction with a small  $R$  might often accompany a weak SST forcing in the two regions.

A composite of the square of the October SSTA of the 51 winters is shown in Figs. 10a,b, for  $R > 2.0$  and  $R < 1.0$ , respectively. Apparently a large (small)  $R$  corresponds to strong (weak) SST forcing in the tropical central Pacific Ocean and in the NP, which is consistent with the above correlation analysis. From these results, it can be suggested that the  $R$  is well linked to the signal of SST forcing in the tropical Pacific and in the NP.

It is of interest to directly correlate the global SSTA

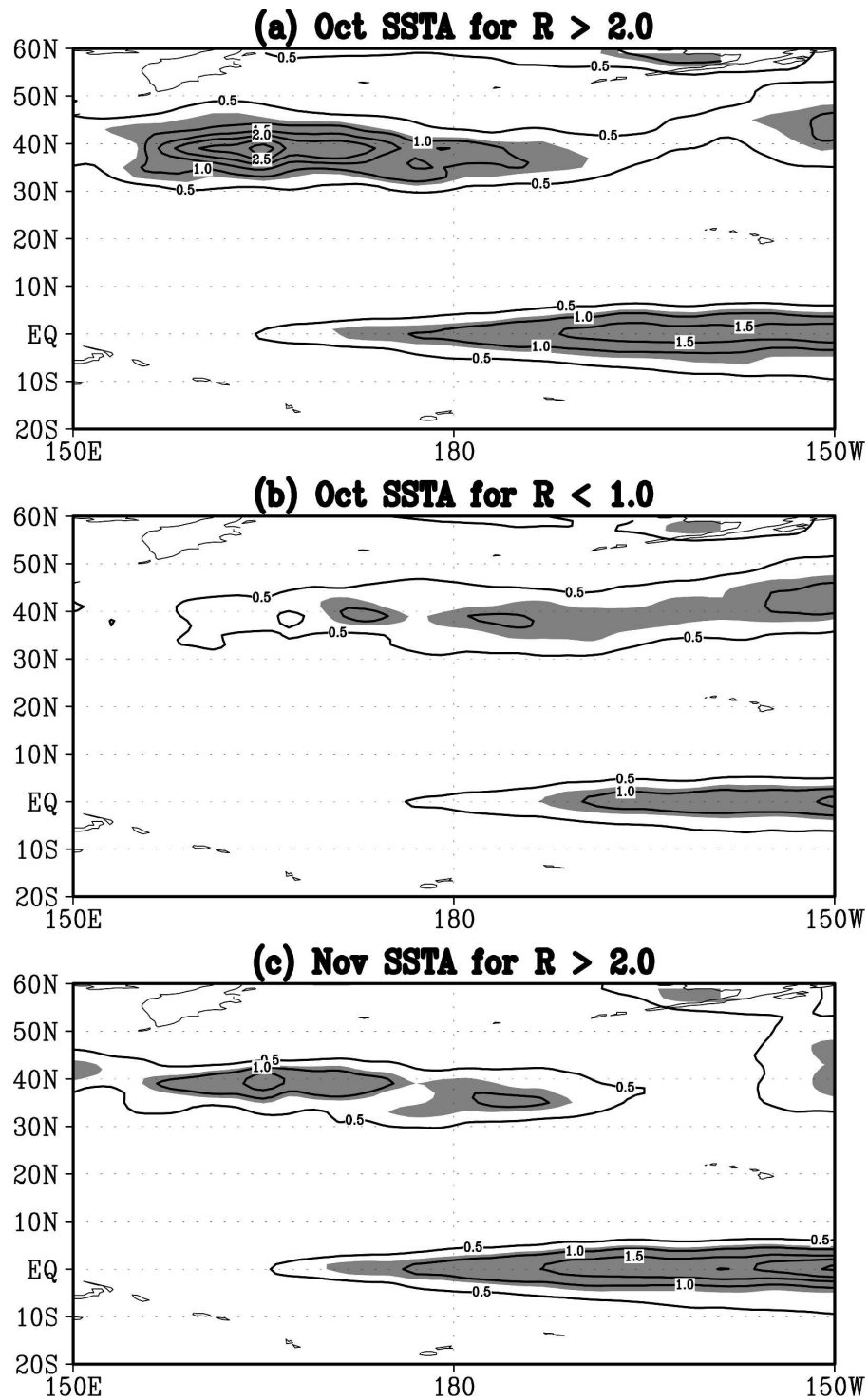


FIG. 10. Composite of mean SSTA<sup>2</sup> of October for years of (a)  $R > 2.0$ , (b)  $R < 1.0$ , and (c) same as (a), but for November.

in October with the model AO index. The result shows that there is no statistically significant correlation between them for the period of 1950–98, suggesting that if the source of model skill for the winter AO comes from

a lagged response to an SST forcing, such a response is likely nonlinear. Wu and Hsieh (2004) found a quadratic response of 500-mb atmospheric variability to the tropical Pacific SSTA by a neural network analysis.

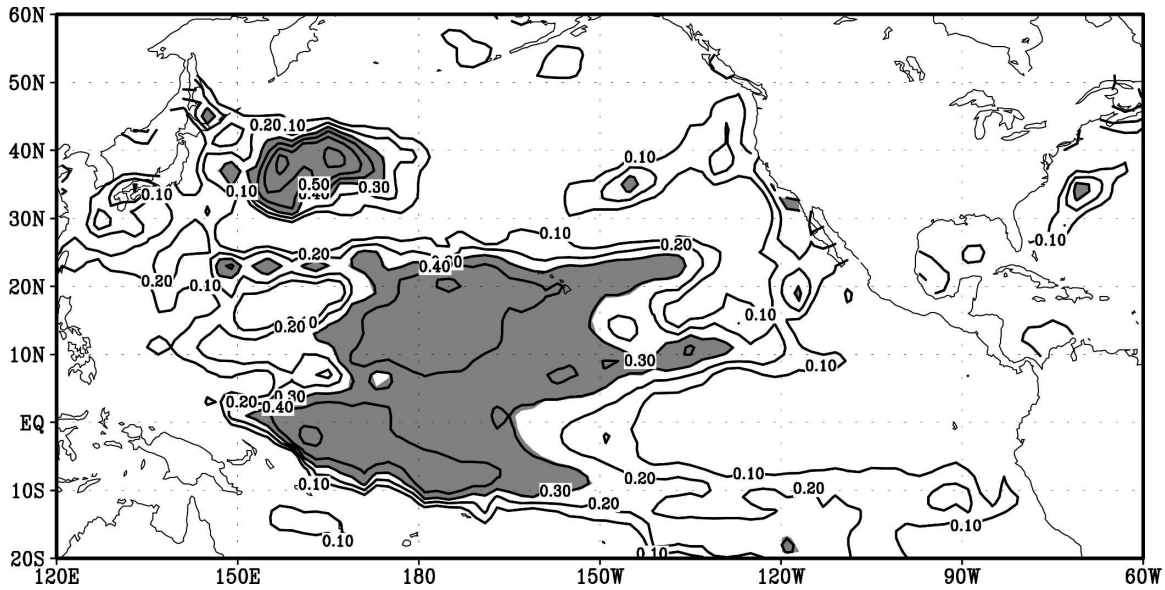


FIG. 11. Correlation between the square of the model DJF AO index and the square of the October SSTA over all predictions.

Figure 11 shows the correlation between the square of the October SSTA with the square of the model AO index for DJF, indicating that such a nonlinear response exists. It is hardly surprising to find a good agreement between Figs. 11 and 8a since the relative entropy  $R$  is dominated by the amplitude of the predicted AO index.

The link between the AO and tropical forcing has been suggested in previous studies (e.g., Lin et al. 2002; Greatbatch et al. 2003; Lin et al. 2005a,b). Derome et al. (2005) found that the predicted ensemble mean AO index is significantly correlated with the time series of an EOF pattern of diabatic heating that is characterized by larger variances in the equatorial central Pacific and the Indonesian region. Our analyses presented above further explore the link of the AO to Pacific SST forcing in terms of the  $R$ . In particular, we found that the SST forcing in the NP region also remarkably correlates with AO predictability. It was observed in previous studies that the SSTA in the tropical Pacific often covaries with that in the North Pacific. For example, in an El Niño event, a warm tropical Pacific SST is accompanied by a cold SST anomaly in the North Pacific. Lau and Nath (1996) attributed the link between the SSTA in these two regions to the “atmospheric bridge” mechanism.

An interesting result found here is that the  $R$  is highly related to the October tropical SST signal, rather than to the SST signal in November, which appears to be the most relevant to the initial conditions and model forcing. This is most probably because 1) the October SST anomaly is much stronger than the November SST

anomaly, as shown in Fig. 10c. As discussed above, the strength of the SSTA signal plays an important role in influencing the  $R$ . 2) We used the observed atmospheric circulation data in November to calculate the model forcing. It has been found that the mid-high-latitude atmospheric circulation has a lagged response to the tropical SST anomaly by around 1 month (Jin and Hoskins 1995; Hall and Derome 2000). The nature and cause of the delayed atmospheric response to El Niño was investigated in Kumar and Hoerling (2003). The October SST anomaly is therefore likely to be the best representation of the November forcing that we used in the seasonal predictions. By a simple correlation analysis, Mo et al. (1998) also found that the major modes of interannual variability in the Northern Hemisphere mean-winter 500-hPa field [such as the PNA and Western Pacific (WP)] in December–March are correlated most significantly with the Pacific SST anomalies in the previous October.

Thus, the source of the model skill appears to be attributable to the forcing of the tropical SSTA in October. This could be interpreted in two aspects: 1) the tropical SSTA in October has a significant impact on the model atmospheric forcings in November due to their around 1-month lag relationship. As the initial conditions, the November forcing plays an important role in model predictions. 2) The SSTA in October impacts the November AO itself, leading to some prediction skill through persistence as mentioned in section 3. To explore the possible response of the November AO to the October SSTA, we first correlated the November AO with the October SSTA over the global

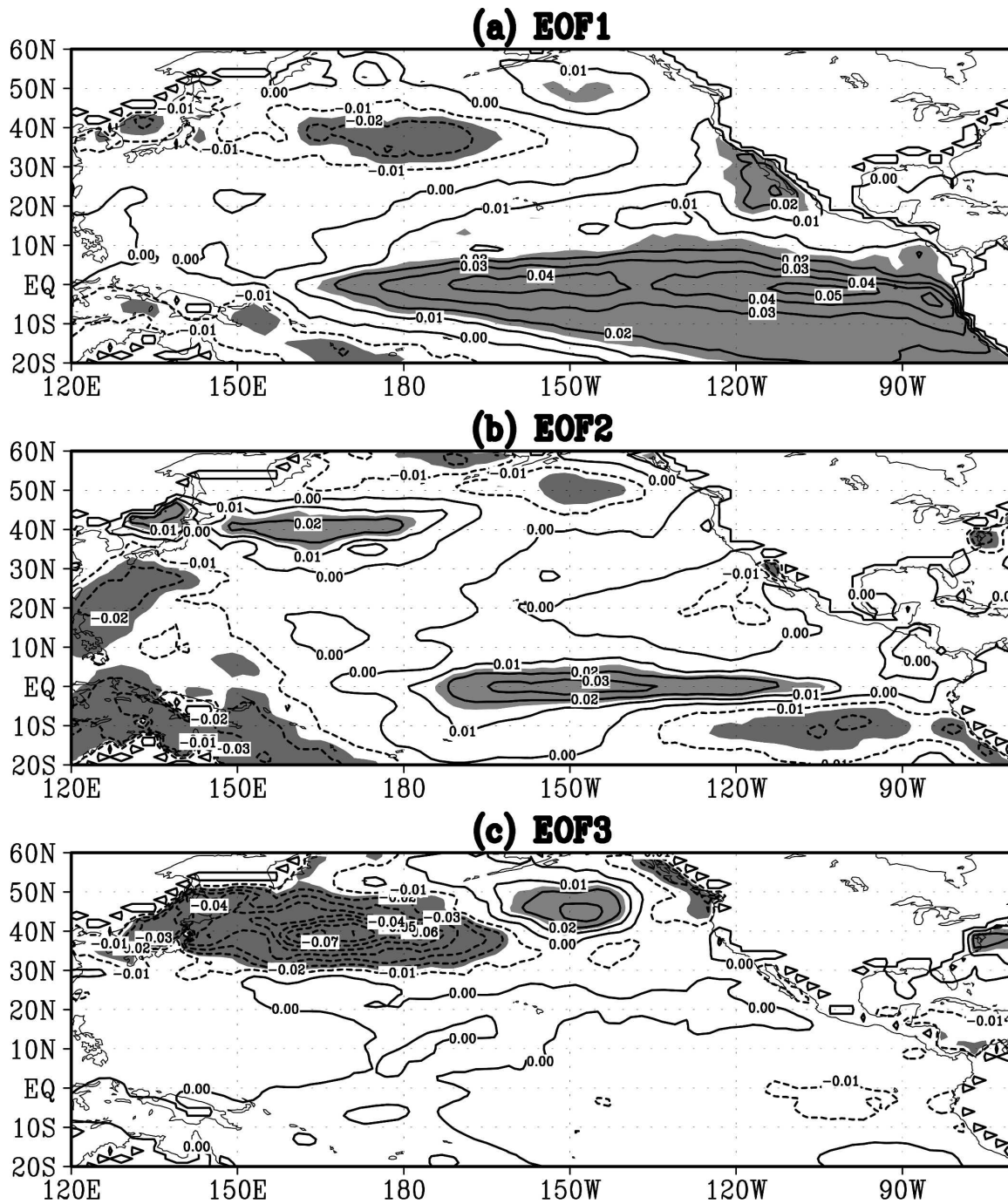


FIG. 12. The first three EOF modes of the tropical SSTA in the October, derived from 1950–98: EOF (a) 1, (b) 2, and (c) 3.

domain, and found very small correlation coefficients everywhere between them, suggesting that the lag response, if it exists, should be nonlinear. To examine this possibility, we performed an EOF analysis on the October SSTA over the Pacific domain of 20°S–60°N and 120°E–90°W. The first three modes are plotted in Fig. 12, which explain 60% of the total variance. The three

EOF modes have relatively much larger variances in the tropical Pacific and in the NP, similar to Figs. 8a,b. A nonlinear response of the November AO to the October SSTA was identified using a nonlinear regression by the neural network (Tang et al. 2001) with the time series of the three EOF modes as the inputs and the November AO index as the output. The cross-

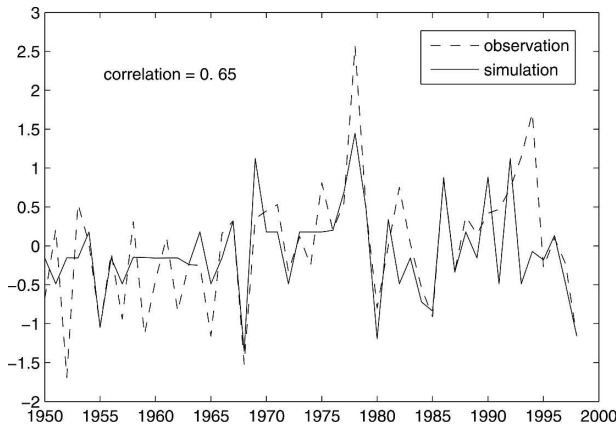


FIG. 13. Variation of the November AO index from 1950–98. The dashed line is the observed values, and the solid line is the modeled values using the Pacific October SSTA and a nonlinear regression via a Neural Network method.

validation correlation of the simulated AO index from the constructed nonlinear regression against the observed counterpart is 0.65 for the period of 1948–98, which is statistically significant at the confidence level of 99%. Figure 13 illustrates the simulated and observed AO indices, indicating that most of the variance of the November AO variability can be explained by the October SSTA forcing via a nonlinear regression of the neural network.

## 7. Discussion and summary

An important task of predictability studies is to measure the reliability of the prediction, and to determine dominant factors that affect the prediction accuracy. By applying the information theory, we have explored the AO predictability using a simple global atmospheric general circulation model. It was found that the  $R$ , defined by the relative entropy, can measure reasonably well the reliability of the AO predictions of the SGCM. In general, when  $R$  is large, the corresponding AO prediction is found to be more reliable than when  $R$  is small. Such a “triangular” relationship between the  $R$  and model skill is different from the ENSO predictability where the ENSO prediction skill is more likely to be a monotonic function of  $R$  (Tang et al. 2005). Like the ENSO predictability, the  $R$  of AO prediction also has the following property that it is dominated by the predictive ensemble mean (i.e., the signal component has a much more important contribution to  $R$  than the ensemble spread). We also examined the model skill and ensemble spread, and did not find any significant relationship between them. This indicates that the ensemble spread is not an effective indication of predic-

tion skill in this SGCM. This is most probably due to the weak ensemble spread in the SGCM. A weak ensemble spread might be related to model dynamics, the ensemble perturbation method, and the lack of variability of the model forcing from member to member. It should be noted that while this result may be model dependent, some more complex GCM models also displayed a strong relationship between the ensemble mean and the model skill (e.g., Kumar et al. 2000; Tippet et al. 2004; Tang et al. 2005).

A practical significance of the above conclusion relates to the possible use of the SGCM to issue probabilistic forecasts operationally. Since only the ensemble mean is responsible for the  $R$  and the skill, and the ensemble spread is weak, a large size of ensemble does not seem required to generate a prediction and measure the reliability of predictions. Usually one can estimate well the mean of the distribution with a few samples if the distribution has a very small variance. To explore the impact of the ensemble size on the model skill and  $R$ , we repeated all calculations performed in section 3 using different ensemble sizes, as shown in Fig. 14. The correlation skills of the predicted AO index against the observed AO index are all 0.41 when the ensemble size is changed to 50, 30, and 10, respectively, which is same as the original skill with the ensemble size of 70. The correlation coefficients between the  $R$  and correlation contribution  $C$  are 0.61, 0.60, and 0.59, respectively, almost unchanged compared with original value of 0.61. These results suggest that as few as 6–10 members may be appropriate to approach probabilistic forecasts operationally using the particular SGCM.

Using global SST observations, we found that the  $R$  of the winter AO prediction is significantly correlated with the amplitude of the SST anomaly in the tropical central Pacific and the North Pacific in the previous October. A large  $R$  that is likely to lead to a reliable prediction is usually linked to a strong SST forcing in the two regions whereas a poor prediction with small  $R$  is associated with a weak SST forcing in the two regions. The primary contributor is likely the tropical Pacific SSTA, while the SST anomaly in the North Pacific is probably a result of the “atmospheric bridge” mechanism. The tropical link of the AO is in agreement with previous studies (Lin et al. 2002; Greatbatch et al. 2003; Lin et al. 2005a). Using the tropical Pacific SST signal, Lin et al. (2005b) developed a correction scheme for seasonal predictions. It was found that this scheme significantly increases the predictive skill of the NAO in the seasonal predictions of two GCMs.

It has been recognized that model initial conditions exert a strong influence on ENSO model prediction skill (e.g., Kleeman and Moore 1997; Tang et al. 2005).

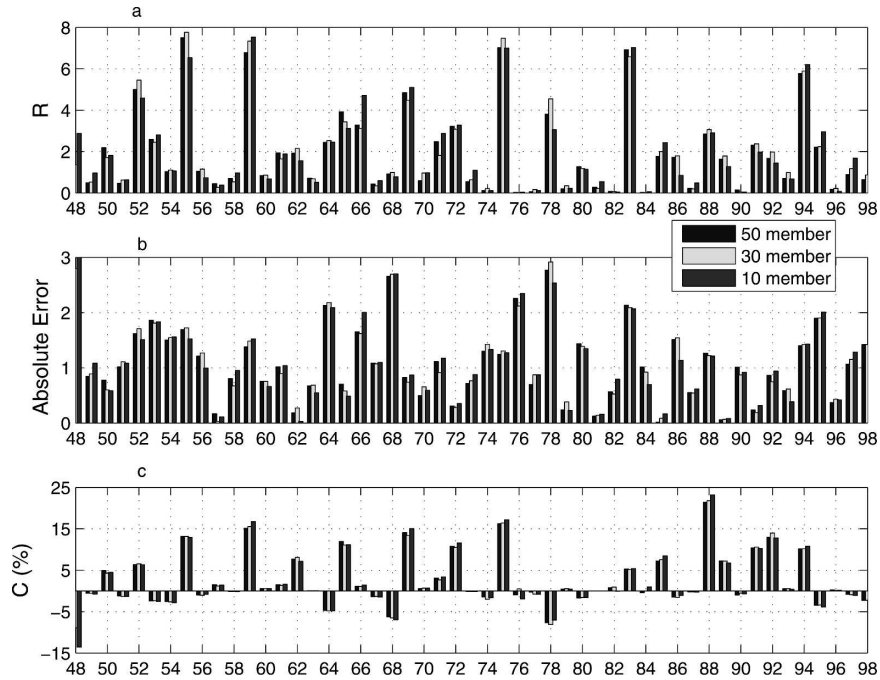


FIG. 14. Same as Fig. 2, but for different ensemble sizes.

In contrast to the ENSO predictability, the SST anomaly in the previous October is most significantly correlated with the winter AO prediction skill. This is probably due to the fact that the November model forcing anomaly that is persisted throughout the prediction period and that determines the forecast skill is significantly related to the October SST anomaly.

*Acknowledgments.* We wish to thank Hans von Storch and two other anonymous reviewers for their valuable comments. This work is supported by the Canadian Foundation for Climate and Atmospheric Sciences (CFCAS) through Grant GR-523 to YT and by the Canadian Climate Variability Research Network (for HL and JD), funded by CFCAS and the Natural Sciences and Engineering Research Council of Canada. MT is funded by a grant/cooperative agreement from the National Oceanic and Atmospheric Administration (NA05OAR4311004). The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or any of its subagencies.

## APPENDIX

### A Theoretical Relationship between Predictability and Prediction Skill

The relation in (6) builds an important connection between the measure of prediction uncertainty using the relative entropy  $R$  and the prediction skill. It is

obtained by noting that for normally distributed variables with constant variance, the  $R$  is proportional to  $\mu^2 + \text{const.}$ , and  $C$  is proportional to  $\mu(\mu + \varepsilon)$  where  $\mu$  is the ensemble mean and the quantity  $\mu + \varepsilon$  is the observation. The observation is the ensemble mean plus a noise term with mean zero,  $\langle \varepsilon \rangle = 0$ . The variance  $\langle \varepsilon^2 \rangle$  of the noise term determines the correlation between observation and ensemble mean. The square of the correlation between the  $R$  and correlation contribution  $C$  is

$$\begin{aligned}
 \rho_C^2 &= \frac{\langle (R - \langle R \rangle)(C - \langle C \rangle) \rangle^2}{\langle (R - \langle R \rangle)^2 \rangle \langle (C - \langle C \rangle)^2 \rangle} \\
 &= \frac{\langle (\mu^2 - \langle \mu^2 \rangle)(\mu(\mu + \varepsilon) - \langle \mu^2 \rangle) \rangle^2}{\langle (\mu^2 - \langle \mu^2 \rangle)^2 \rangle \langle (\mu(\mu + \varepsilon) - \langle \mu^2 \rangle)^2 \rangle} \\
 &= \frac{\langle (\mu^2 - \langle \mu^2 \rangle)^2 \rangle^2}{\langle (\mu^2 - \langle \mu^2 \rangle)^2 \rangle \langle \mu^2 + \varepsilon^2 \mu^2 - 2\langle \mu^2 \rangle \mu^2 + \mu^4 \rangle} \\
 &= \frac{\langle (\mu^2 - \langle \mu^2 \rangle)^2 \rangle}{\langle \mu^2 + \varepsilon^2 \mu^2 - 2\langle \mu^2 \rangle \mu^2 + \mu^4 \rangle} \\
 &= \frac{\langle \mu^4 \rangle - \langle \mu^2 \rangle^2}{\langle \mu^4 \rangle - \langle \mu^2 \rangle^2 + \langle \varepsilon^2 \rangle \langle \mu^2 \rangle} \\
 &= \frac{1}{1 + \frac{\langle \varepsilon^2 \rangle \langle \mu^2 \rangle}{\langle \mu^4 \rangle - \langle \mu^2 \rangle^2}} = \frac{1}{1 + \frac{1}{2} \frac{\langle \varepsilon^2 \rangle}{\langle \mu^2 \rangle}},
 \end{aligned}$$



where  $\langle \dots \rangle$  denotes the expectation, and we use the fact that  $\langle \mu^4 \rangle = 3\langle \mu^2 \rangle^2$  for normally distributed variables. A similar calculation shows that the correlation  $r$  is related to the signal-to-noise ratio  $\langle \mu^2 \rangle / \langle \varepsilon^2 \rangle$  by Kleeman and Moore (1999):

$$r^2 = \frac{1}{1 + \frac{\langle \varepsilon^2 \rangle}{\langle \mu^2 \rangle}}$$

since

$$\frac{\langle \varepsilon^2 \rangle}{\langle \mu^2 \rangle} = \frac{1}{r^2} - 1,$$

$$\rho c^2 = \frac{1}{1 + \frac{1}{2} \left( \frac{1}{r^2} - 1 \right)} = \frac{2r^2}{2r^2 + 1 - r^2} = \frac{2r^2}{r^2 + 1}.$$

#### REFERENCES

- Buizza, R., and T. N. Palmer, 1998: Impact of ensemble size on ensemble prediction. *Mon. Wea. Rev.*, **126**, 2503–2518.
- Cover, T. M., and J. A. Thomas, 1991: *Elements of Information Theory*. Wiley, 576 pp.
- DeGroot, M. H., 1991: *Probability and Statistics*. 3d ed. Addison-Wesley, 816 pp.
- DeSole, T., 2004: Predictability and information theory. Part I: Measures of predictability. *J. Atmos. Sci.*, **61**, 2425–2440.
- , 2005: Predictability and information theory. Part II: Imperfect forecasts. *J. Atmos. Sci.*, **62**, 3368–3381.
- , and M. K. Tippett, 2007: Predictability: Recent insights from information theory. *Rev. Geophys.*, in press.
- Derome, J., H. Lin, and G. Brunet, 2005: Seasonal forecasting with a simple general circulation model: Predictive skill in the AO and PNA. *J. Climate*, **18**, 597–609.
- Greatbatch, R. J., 2000: The North Atlantic Oscillation. *Stochastic Environ. Res. Risk Assess.*, **14**, 213–242.
- , H. Lin, J. Lu, K. A. Peterson, and J. Derome, 2003: Tropical/extratropical forcing of the AO/NAO, 2003: A corrigendum. *Geophys. Res. Lett.*, **30**, 1738, doi:10.1029/2003GL017406.
- Hall, N. M. J., 2000: A simple GCM based on dry dynamics and constant forcing. *J. Atmos. Sci.*, **57**, 1557–1572.
- , and J. Derome, 2000: Transients, nonlinearity, and eddy feedback in the remote response to El Niño. *J. Atmos. Sci.*, **57**, 3992–4007.
- , —, and H. Lin, 2001a: The extratropical signal generated by a midlatitude SST anomaly. Part I: Sensitivity at equilibrium. *J. Climate*, **14**, 2035–2053.
- , H. Lin, and J. Derome, 2001b: The extratropical signal generated by a midlatitude SST anomaly. Part II: Influence on seasonal forecasts. *J. Climate*, **14**, 2696–2709.
- Hoskins, B. J., and A. J. Simmons, 1975: A multi-layer spectral model and the semi-implicit method. *Quart. J. Roy. Meteor. Soc.*, **101**, 637–655.
- Jin, F., and B. J. Hoskins, 1995: The direct response to tropical heating in a baroclinic atmosphere. *J. Atmos. Sci.*, **52**, 307–319.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
- Kleeman, R., 2002: Measuring dynamical prediction utility using relative entropy. *J. Atmos. Sci.*, **59**, 2057–2072.
- , and A. M. Moore, 1997: A theory for the limitation of ENSO predictability due to stochastic atmospheric transients. *J. Atmos. Sci.*, **54**, 753–767.
- , and —, 1999: A new method for determining the reliability of dynamical ENSO predictions. *Mon. Wea. Rev.*, **127**, 694–705.
- , and A. J. Majda, 2005: Predictability in a model of geostrophic turbulence. *J. Atmos. Sci.*, **62**, 2864–2879.
- Kumar, A., and M. Hoerling, 2003: The nature and causes for the delayed atmospheric response to El Niño. *J. Climate*, **16**, 1391–1403.
- , A. B. Barnston, P. Peng, M. P. Hoerling, and L. Goddard, 2000: Changes in the spread of the variability of the seasonal mean atmospheric states associated with ENSO. *J. Climate*, **13**, 3139–3151.
- Lau, N.-C., and M. Nath, 1996: The role of the “atmospheric bridge” in linking tropical Pacific ENSO events to extratropical SST anomalies. *J. Climate*, **9**, 2036–2057.
- Lin, H., J. Derome, R. J. Greatbatch, K. A. Peterson, and J. Lu, 2002: Tropical links of the Arctic Oscillation. *Geophys. Res. Lett.*, **29**, 1943, doi:10.1029/2002GL015822.
- , —, and G. Brunet, 2005a: Tropical Pacific link to the two dominant patterns of atmospheric variability. *Geophys. Res. Lett.*, **32**, L03801, doi:10.1029/2004GL021495.
- , —, and —, 2005b: Correction of atmospheric dynamical seasonal forecasts using the leading ocean-forced spatial patterns. *Geophys. Res. Lett.*, **32**, L14804, doi:10.1029/2005GL023060.
- Mo, R. P., J. Fyfe, and J. Derome, 1998: Phase-locked and asymmetric correlations of the wintertime atmospheric patterns with the ENSO. *Atmos.–Ocean*, **36**, 213–239.
- Moore, A. M., and R. Kleeman, 1998: Skill assessment for ENSO using ensemble prediction. *Quart. J. Roy. Meteor. Soc.*, **124**, 557–584.
- Schneider, T., and S. M. Griffies, 1999: A conceptual framework for predictability studies. *J. Climate*, **12**, 3133–3155.
- Smith, T. M., R. W. Reynolds, R. E. Livezey, and D. C. Stokes, 1996: Reconstruction of historical sea surface temperatures using empirical orthogonal functions. *J. Climate*, **9**, 1403–1420.
- Tang, Y., W. W. Hsieh, B. Tang, and K. Haines, 2001: A neural network atmospheric model for hybrid coupled modeling. *Climate Dyn.*, **17**, 445–455.
- , R. Kleeman, and A. M. Moore, 2005: On the reliability of ENSO dynamical predictions. *J. Atmos. Sci.*, **62**, 1770–1791.
- Thompson, D. W. J., and J. M. Wallace, 1998: The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophys. Res. Lett.*, **25**, 1297–1300.
- , and —, 2001: Regional climate impacts of the Northern Hemisphere annular mode. *Science*, **293**, 85–89.
- Tippett, M. K., R. Kleeman, and Y. Tang, 2004: Measuring the potential utility of seasonal climate predictions. *Geophys. Res. Lett.*, **31**, L22201, doi:10.1029/2004GL021575.
- von Storch, H., and J. S. Xu, 1990: Principal oscillation pattern analysis of the tropical 30- to 60- day oscillation. Part I: Defi-

- , and D. P. Baumhefner, 1991: Principal oscillation pattern analysis of the tropical 30- to 60-day oscillation. Part II: The prediction of equatorial velocity potential and its skill. *Climate Dyn.*, **5**, 1–12.
- , and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research*. Cambridge University Press, 494 pp.
- Whitaker, J. S., and A. F. Loughe, 1998: The relationship between ensemble spread and ensemble mean skill. *Mon. Wea. Rev.*, **126**, 3292–3302.
- Wu, A., and W. W. Hsieh, 2004: The nonlinear Northern Hemisphere atmospheric response to ENSO. *Geophys. Res. Lett.*, **31**, L02203, doi:10.1029/2003GL018885.
- Xue, Y., M. A. Cane, S. E. Zebiak, and T. N. Palmer, 1997: Predictability of a coupled model of ENSO using singular vector analysis. Part II: Optimal growth and forecast skill. *Mon. Wea. Rev.*, **125**, 2057–2073.

Copyright of *Journal of Climate* is the property of *American Meteorological Society* and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.