PNA Predictability at Various Time Scales

WAQAR YOUNAS

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

YOUMIN TANG

Environmental Science and Engineering, University of Northern British Columbia, Prince George, British Columbia, Canada, and State Key Laboratory of Satellite Ocean Environment Dynamics, Hangzhou, China

(Manuscript received 4 August 2012, in final form 17 June 2013)

ABSTRACT

In this study, the predictability of the Pacific–North American (PNA) pattern is evaluated on time scales from days to months using state-of-the-art dynamical multiple-model ensembles including the Canadian Historical Forecast Project (HFP2) ensemble, the Development of a European Multimodel Ensemble System for Seasonal-to-Interannual Prediction (DEMETER) ensemble, and the Ensemble-Based Predictions of Climate Changes and their Impacts (ENSEMBLES). Some interesting findings in this study include (i) multiple-model ensemble (MME) skill was better than most of the individual models; (ii) both actual prediction skill and potential predictability increased as the averaging time scale increased from days to months; (iii) there is no significant difference in actual skill between coupled and uncoupled models, in contrast with the potential predictability where coupled models performed better than uncoupled models; (iv) relative entropy (RE_A) is an effective measure in characterizing the potential predictability of individual prediction, whereas the mutual information (MI) is a reliable indicator of overall prediction skill; and (v) compared with conventional potential predictability measures of the signal-to-noise ratio, the MI-based measures characterized more potential predictability when the ensemble spread varied over initial conditions.

Further analysis found that the signal component dominated the dispersion component in RE_A for PNA potential predictability from days to seasons. Also, the PNA predictability is highly related to the signal of the tropical sea surface temperature (SST), and SST–PNA correlation patterns resemble the typical ENSO structure, suggesting that ENSO is the main source of PNA seasonal predictability. The predictable component analysis (PrCA) of atmospheric variability further confirmed the above conclusion; that is, PNA is one of the most predictable patterns in the climate variability over the Northern Hemisphere, which originates mainly from the ENSO forcing.

1. Introduction

The Pacific–North American (PNA) pattern is one of the most prominent modes of low-frequency climate variability in the Northern Hemisphere during winter (e.g., Wallace and Gutzler 1981; Woodhouse 1997). It significantly affects the weather and climate anomalies over North America. For example, during the PNA positive phase, above-normal temperatures can be witnessed in the western United States. In its negative phase, dry and warm conditions are detected in the eastern United States while relatively dry and cold conditions may be experienced in the west (Yarnal and Diaz 1986; Leathers et al. 1991). It has been argued that, in the midlatitude region, the seasonal climate prediction skill is mainly determined by the prediction skill of the PNA and the North Atlantic Oscillation (NAO), another important mode in mid to high latitudes (e.g., Hurrell 1995; Hurrell et al. 2003; Doblas-Reyes et al. 2003; Vitart 2004).

Past studies on PNA predictability covered several different aspects using models and observations, including model skill evaluation and analysis of the sources of predictability. While it has been well reported that PNA predictability, on monthly to seasonal time scales, is mainly from slowly varying external forcing, [e.g., the El Niño–Southern Oscillation (ENSO) forcing] (e.g., Horel and Wallace 1981; Hoskins and Karoly 1981; Simmons 1982; Sardeshmukh and Hoskins 1988; Kumar

Corresponding author address: Dr. 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

DOI: 10.1175/JCLI-D-12-00609.1

et al. 1996; Straus and Shukla 1997; Shukla et al. 2000), some other works also suggested that a considerable part of PNA predictability comes from the internal dynamics of the midlatitude atmosphere inherent to atmospheric instabilities (baroclinic nature in the extratropics) and nonlinear interactions between large-scale and synoptic-scale atmospheric process (e.g., Lau 1981; Wallace and Blackmon 1983; Chervin 1986; Palmer 1993; Straus and Shukla 2002).

In the field of statistical predictability of seasonal climate, it has been challenging to detect the predictability source from internal variability, which often is treated as stochastic forcing compared with large-scale slow-varying signals. Recently, information-based potential predictability measures were applied to study statistical predictability, providing a convenient way to explore the relative role of internal and external forcing to predictability. For example, Abramov et al. (2005) found, using a highly simplified model, that the ensemble spread associated with the internal dynamics is the main contributor to the PNA potential predictability. On the other hand, Kleeman (2008) demonstrated in another simplified model that the variation in midlatitude atmospheric predictability with respect to initial conditions is mainly determined by an ensemble signal related to external forcing. Thus, the source of PNA predictability is still an open question. A reasonable answer may be expected by exploring the relative contributions of external forcing and internal dynamics to the PNA potential predictability at different time scales using realistic atmospheric or climate models, which seems absent in the literature.

Another issue in the past studies of PNA predictability is the lack of a comprehensive evaluation of actual and potential predictability with time averaging. Usually the PNA predictability was evaluated using either actual skill (Renwick and Wallace 1995; Lin and Derome 1996; Nakaegawa and Kanamitsu 2006) or potential predictability (e.g., Barnett et al. 1997; Phelps et al. 2004; Abramov et al. 2005) separately, mainly at monthly to seasonal scales.[1] Very few studies have focused on the evaluation of the PNA predictability at various time scales from days to months, in particular for potential predictability. One exception is the work of Johansson (2007) where he investigated the actual correlation skill of PNA and NAO from daily to seasonal time scale using five years of winter predictions by the NCEP Climate Forecast System from 2001 to 2005. The PNA actual and potential predictability study at an integral framework at various time scales is absent in the literature. These gaps will be main focuses of this study.

In this study, the PNA predictability will be systematically explored from days to months using actual and potential predictability measures. The purpose of this study is to comprehensively analyze the PNA predictability and its variation at different time scales, using long-term ensemble predictions from multiple global coupled and uncoupled climate models, where the coupled models mean coupled atmospheric and oceanic general circulation models, also called a one-tier forecast in the seasonal prediction community, whereas the uncoupled models mean the atmospheric circulation models driven by persistent boundary (ocean) forcing, that is, a two-tier forecast. This allows us to derive statistically robust, general, and realistic conclusions. To achieve these, the analysis is done using ensemble predictions of 500-mb geopotential height from multiple-model ensembles including the Canadian Centre for Climate Modelling and Analysis (CCCma) Historical Forecast Project (HFP2) product of four different global models, the ECMWF DEMETER product from five global coupled models, and the European Centre for Medium-Range Weather Forecasts (ECMWF) Ensemble-Based Predictions of Climate Changes and their Impacts (ENSEMBLES) product from three global coupled models. The multiple-model ensemble (MME), or superensemble, of each product was used in this study. It has been argued that MME is usually better than a single model ensemble (SME) since uncertainties associated with the different model frameworks can offset each other and can be relieved by large size of ensemble members (e.g., Krishnamurti et al. 1999, 2000; Palmer et al. 2004; Yan and Tang 2013).

This study is organized as follows. In section 2, a brief description of the ensemble prediction products and analysis data is presented. Section 3 discusses different measures of prediction skill and predictability. Results and discussion for actual skill and potential predictability at various time scales are presented in sections 4 and 5 respectively. Section 6 explores the predictability in coupled and uncoupled models, and the temporal variations in PNA predictability and its source are presented in section 7. Predictable component analysis is presented in section 8, and we end with a broad discussion and conclusions in section 9.

2. Ensemble prediction products and observed data

Three ensemble prediction datasets were used in this study including HFP2, Development of a European

[1] For a clarification, we define the potential predictability used in this study as the prediction skill of a ''perfect'' forecast system (model), which does not make use of observation to measure whereas the actual prediction skill is the prediction skill evaluated using the observation. The predictability, used as a general expression, means both the potential predictability and actual prediction skill in this study.

Multimodel Ensemble System for Seasonal-to-Interannual Prediction (DEMETER), and ENSEMBLES stream 2. HFP2 is a collaborative project among some Canadian universities and government laboratories, whose objective is to test the extent to which the potential predictability of mean seasonal conditions could be achieved (Kharin et al. 2009; Derome et al. 2001). This product includes an ensemble of four global atmospheric models, the second- and third-generation atmospheric general circulation models (AGCM2 and AGCM3) (McFarlane et al. 1992), a reduced resolution version of medium-range weather forecast global spectral model (SEF) (Ritchie 1991), and the Global Environmental Multiscale (GEM) model. Each model produces an ensemble of 10 parallel integrations of four-month duration for the period of 1969–2003 from the beginning of each month. The integrations are initialized from the National Centers for Environmental Prediction (NCEP)–National Center for Atmospheric Research (NCAR) reanalysis (Kalnay et al. 1996) lagged at 12-h intervals prior to the initial time of prediction (ITF). That is, the first member is initialized 12 h before the ITF, the second member is initialized 24 h prior to the ITF etc., and the 10th member is initialized 5 days prior to the ITF. The oceanic forcing used in the atmospheric prediction is from the persistent prediction of the global sea surface temperature anomaly (SSTA), namely, that the SSTA of the initial month of the prediction was persisted during the entire prediction period, superimposed onto the climatological SST of the target month of prediction.

DEMETER includes seven state-of-the-art global coupled ocean–atmosphere models that produce a 6-month forecast starting from 1 February, 1 May, 1 August, and 1 November of each year, over a common period 1980–2001 (Palmer et al. 2004). Four SST perturbations are added to and subtracted from initial conditions of the hindcast to represent the uncertainty in the SST. These SST perturbations are based on two quasi-independent SST analyses. Each hindcast has been integrated for six months and comprises an ensemble of nine members. Two models from INVG and Max Planck Institute (MPI), which have very poor skill over the PNA region, were excluded in this study. Thus, the multimodel of DEMETER is formed by merging five models having 45 ensemble members.

The ENSEMBLES seasonal forecasts used in this study are from its stream 2 experiment. In this experiment, five globally coupled general circulation models were run, including the ECMWF, the Leibniz Institute of Marine Sciences at Kiel University (IFM-GEOMAR), Météo-France (MF), the Met Office (UKMO), and the Euro-Mediterranean Centre for Climate Change in Bologna (CMCC)-Istituto Nazionale di Geofisica e Vulcanologia (INGV) models. Each ensemble comprises nine runs initialized with different sets of ocean reanalysis generated from wind stress and SST perturbations. For each year, 7-month-long seasonal forecasts starting on the first day of February, May, August, and November have been initialized, covering the 46 years from 1960 to 2005 (Weisheimer et al. 2009). The November initial condition is extended to a 14-month duration of forecast except for CMCC-INGV. Among five models, IFM-GEOMAR and CMCC-INGV were not used owing to their relatively poor performance than the other models.

The current study focuses on the 500-mb geopotential height ensemble prediction for the common 21-yr period from 1980 to 2001. The hindcasts are initialized, respectively, on 1 February, 1 May, 1 August, and 1 November, and last four months. For validation purpose, the NCEP–NCAR reanalysis dataset (Kalnay et al. 1996), at horizontal resolution of $2.5° \times 2.5°$, was used.

Originally the PNA was defined by the four-point teleconnection patterns (Wallace and Gutzler 1981). However, the classic definition seems insufficient in describing the large-scale spatial–temporal structure of PNA patterns; thus, many researchers prefer to use the leading modes of principal component analysis (PCA) of the Northern Hemisphere 500-mb geopotential height anomalies to define the PNA, such as Thompson and Wallace (1998, 2000). In this study we follow this strategy, that is, applying for the rotated PCA onto the observed (NCEP) monthly mean 500-mb geopotential height anomalies over the domain between the equator and 87.5°N, from 1950 to 2000, to define the PNA. The definition is the same as one that has been operationally used in the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (Chen and Van den Dool 2003).

Shown in Fig. 1 are the first two modes of the PCA, clearly indicating that they well characterize the North Atlantic Oscillation (NAO) and PNA structure, respectively. The PNA pattern derived from monthly data (mode 2) is very similar to those from the data at other time scales such as weekly or daily data (Johansson 2007); thus for simplicity we use the monthly PNA pattern for all time scales in this study. The predicted PNA index is obtained by projecting the ensemble prediction of 500-mb geopotential height anomalies on the PNA pattern for all time scales from a day to a month.

3. Measures of prediction skill and predictability

a. Actual measures of predictability

In general, there are two groups of predictability measures: one is actual measures that make use of observations

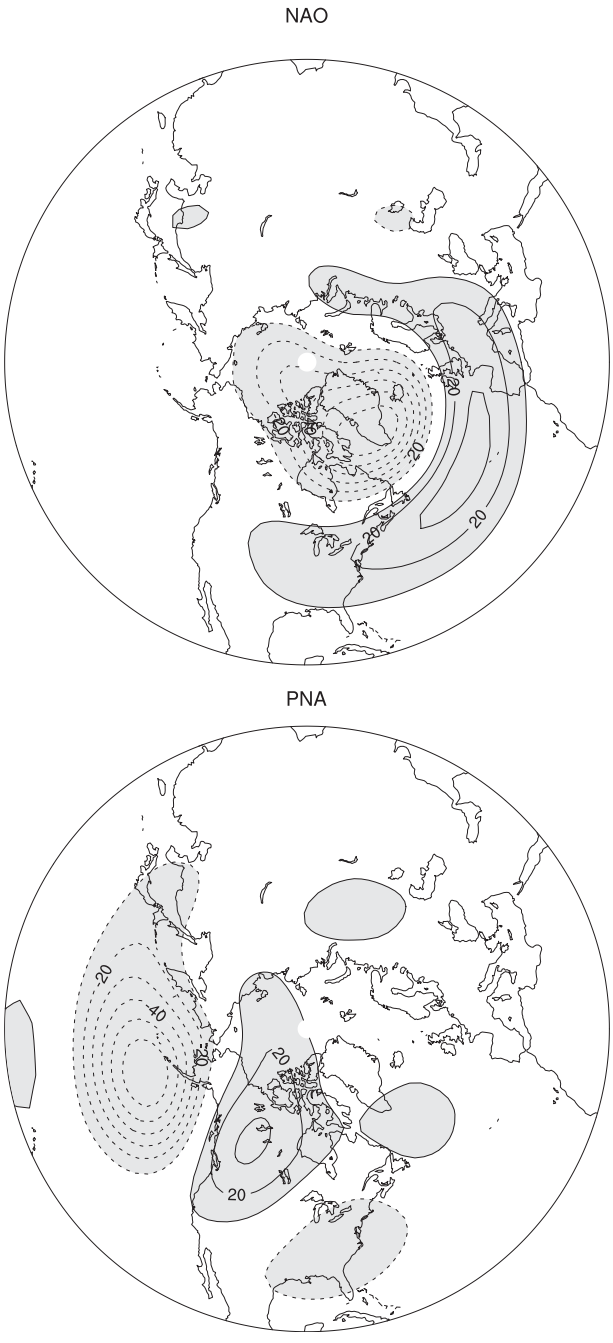

FIG. 1. NAO and PNA patterns obtained using rotated EOF analysis of 500-mb monthly NCEP geopotential height data. Contour interval is 10 m with negative contours dashed. Dark (light) shading indicates values <10 m (>10 m).

and the other one is potential predictability measures that do not make use of observations. The actual skill of ensemble-mean prediction over 22 years is measured by anomaly correlation (r) and rms error (RMSE), defined as

$$r(t) = \frac{\sum_{i=1}^{N} [T_i^p(t) - \mu^p][T_i^o(t) - \mu^o]}{\sqrt{\sum_{i=1}^{N} [T_i^p(t) - \mu^p]^2}\sqrt{\sum_{i=1}^{N} [T_i^o(t) - \mu^o]^2}}, \quad (1)$$

$$\mathrm{RMSE}(t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} [T_i^p(t) - T_i^o(t)]^2} \quad (2)$$

in which T is the index of interest, t is the lead time of prediction (varies with time scale), superscript p is the predicted index, and o is the observed counterpart. The μ^p is the mean of the forecasts, μ^o is the mean of observations, and N is the number of initial conditions.

RMSE indicates the mean ''distance'' between forecasts and observations over the verification time period. It usually increases with lead time and asymptotically approaches a ''saturation'' value. The saturation value is equivalent to the mean difference between two randomly chosen fields from the system (e.g., DelSole 2004).

b. Potential predictability measures

The measures of potential predictability include traditional metrics such as signal-to-noise ratio (SNR), ensemble mean, ensemble spread, and information-based metrics. As it is straightforward to calculate the ensemble mean and ensemble spread from ensemble predictions, here only information-based potential predictability measures will be introduced. The idea behind the information-based measures is to use the difference between two entropies, climatological entropy and predictive entropy, to quantitatively measure extra information that the prediction brings, as expressed by relative entropy (RE). The average of RE over all initial conditions is called mutual information (MI), quantifying the overall predictability of dynamic systems.

1) RELATIVE ENTROPY

Relative entropy, or Kullback–Leibler divergence, is a measure used to calculate the difference between two entropies (uncertainties), such as the difference between climatological entropy and the predictive entropy, as used in this study. If $q(v)$ denotes the climatological distribution for a random variable v and $p(v \mid o)$ denotes for forecast given the initial or boundary condition of ''o'', then for a continuous set of states a relative entropy (RE_A) is defined as

$$\mathrm{RE}_A \equiv \int P(v \mid o) \ln\left[\frac{P(v \mid o)}{q(v)}\right] dv. \quad (3)$$

Also, $q(v)$ is interpreted as prior distribution (climatological distribution), and $P(v|o)$ is described as posterior distribution (forecast distribution). For a Gaussian PDF approximation, RE_A can be calculated exactly in terms of predictive and climatological variances and a difference between their means. In that case, the expression for relative entropy is given by (Kleeman 2002)

$$\mathrm{RE}_A = \frac{1}{2}\left\{\ln\left[\frac{\det(\sigma_q^2)}{\det(\sigma_p^2)}\right] + \mathrm{tr}[(\sigma_p^2)(\sigma_q^{-2})] - n + (\mu^p - \mu^q)(\sigma_q^2)^{-1}(\mu^p - \mu^q)\right\}. \quad (4)$$

Here, σ_q and σ_p are the climatological and ensemble variances, respectively, while μ^p and μ^q are the ensemble and climatological mean of the system, and n is the number of degrees of freedom. As shown in (4), RE_A is composed of two components: (i) a reduction in climatological uncertainty by the prediction [first three terms of (4), also called the dispersion component] and (ii) a difference in the predictive and climatological means [last term of (4), which is also called the signal component]. These components are interpreted as components of utility of prediction (Kleeman 2002). A large value of RE_A indicates that more information that is different from the climatological distribution is being provided by the prediction, which can be interpreted to be more reliable.

Another measure is mutual information (MI), defined as the average of RE_A, over all initial conditions (e.g., DelSole 2004; Yang et al. 2012). Interestingly, there is a theoretical relationship between MI-based anomaly correlation skill ($\mathrm{AC}_{\mathrm{MI}}$ hereafter) and the conventional signal-to-total variance ratio (STR)-based counterpart (AC_p hereafter) if the prediction and climatological PDFs are Gaussian, as (Yang et al. 2012; Cheng et al. 2011) (also see the appendix)

$$\mathrm{AC}_p \le \mathrm{AC}_{\mathrm{MI}}. \quad (5)$$

The equality holds in (5) when ensemble variance is not a function of initial conditions; otherwise, the $\mathrm{AC}_{\mathrm{MI}}$ measures more predictability than AC_p. This is because the AC_p measures a linear relationship between prediction (ensemble mean) and "observation" (an ensemble member), which underestimates the true potential predictability that is statistically defined as the coherence between prediction and initial (boundary) conditions. On the other hand, $\mathrm{AC}_{\mathrm{MI}}$ measures the statistical dependence, both linear and nonlinear, between prediction and observation (Yang et al. 2012; Tang et al. 2013, manuscript submitted to *J. Climate*).

2) PREDICTABLE COMPONENT ANALYSIS

Predictable component analysis (PrCA) is analogous to the traditional principal component analysis (PCA), which decomposes the total variance into different structures (eigenvectors), whereas PrCA decomposes the total predictability into various patterns that explain different contributions to the total predictability. PrCA is especially useful if predictability is dominated by a few patterns, which allows us to focus on a few predictable structures instead of various structures that are not predictable.

In practice, the PrCA analysis is performed in a state space of truncated PCA modes so that the covariance matrix that is used to solve eigenvalue equation is full rank. This problem arises because, practically, the number of grid points should be much larger than the number of total samples in climate studies. We used the first 30 PCA modes for truncation in this study. A detailed discussion on PrCA algorithm can be referred to DelSole and Tippett (2007) and Tang et al. (2013, manuscript submitted to *J. Climate*).

4. Model verification and multiple-model ensemble approach

a. Properties of ensemble systems

As a starting point, the ensemble spread and the rms error of the ensemble mean prediction (ensemble RMSE hereafter) are examined for all of the individual models and the multiple-model ensemble (MME). Ensemble spread and the ensemble RMSE are considered to be the most basic quantities used to evaluate an ensemble prediction system. The former, to some extent, diagnoses the sensitivity of model error growth to initial uncertainty, whereas the latter directly measures the accuracy of the ensemble mean prediction against observations. The forecast error growth from initial uncertainty would saturate after some time owing to the nonlinear nature of the atmospheric system, which, on average, is about two weeks on daily time scales (Lorenz 1969). One of the motivations behind the ensemble forecasting is to estimate the forecast uncertainty using ensemble-derived variables such as ensemble spread, an often used quantitative in weather forecast. Stensrud et al. (1999) argued that forecasts with larger ensemble spread are probably less certain than forecasts with smaller ensemble spread. This is based on the assumption of a perfect model, which is usually made for potential predictability study. Under this assumption, one arbitrary realization of the forecast distribution is used as the hypothetical observation instead of real observation. Under this scenario, the actual observation is statistically indistinguishable from members

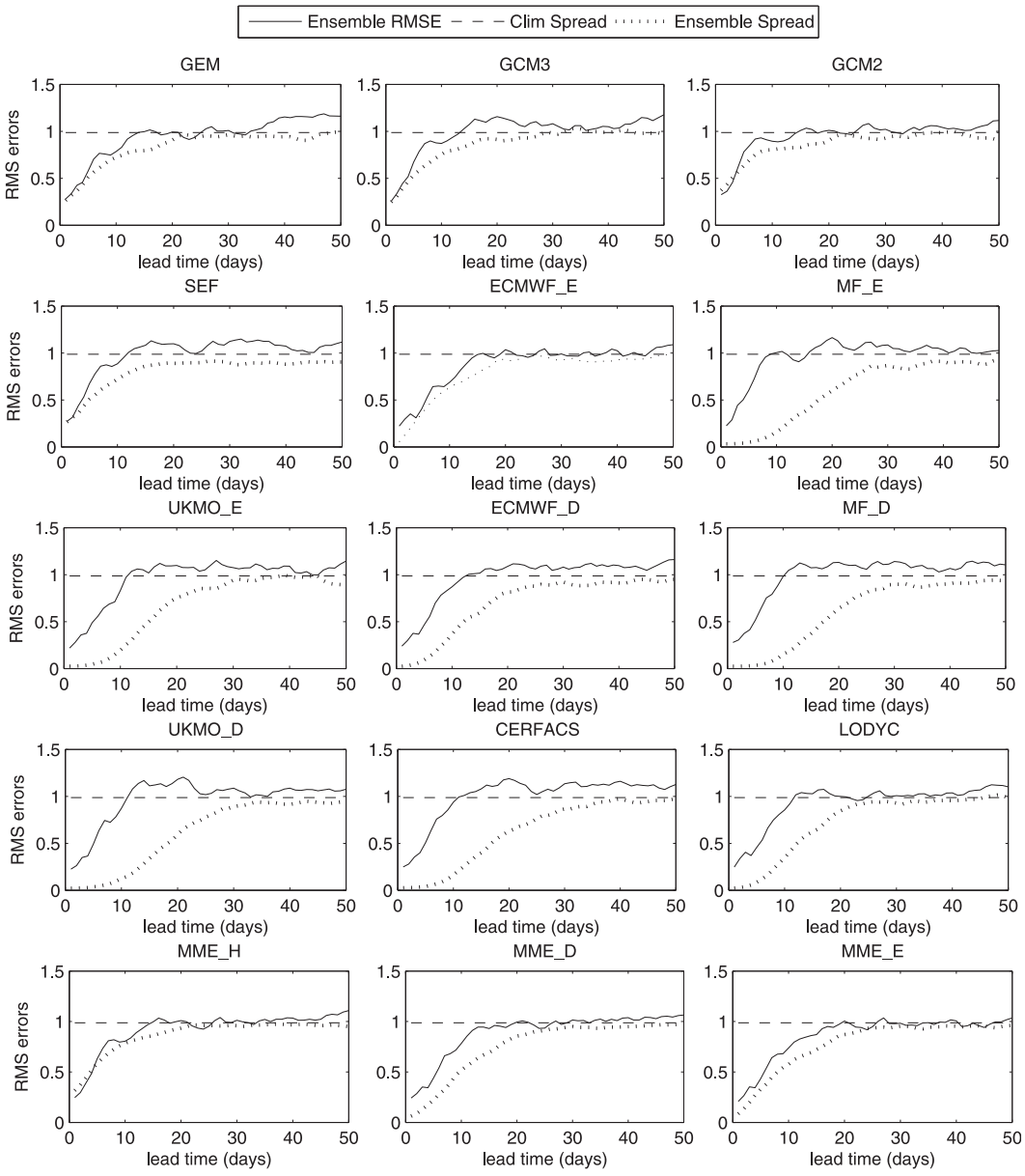

FIG. 2. RMSE and ensemble spread as a function of lead time over daily time scale, for different models and their MME. Here, the ENSEMBLES models are shown with "_E"; DEMETER models with "_D." Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique (CERFACS) and Laboratoire d'Océanographie Dynamique et de Climatologie (LODYC) are DEMETER models.

of the forecast ensemble, and the ensemble spread is equivalent to the RMSE (Begntsson et al. 2008; Cheng et al. 2011). Thus, for daily time scales, the upper limit of weather predictability defined by Lorenz (1969) depends on the saturation of the RMSE that in turn is determined by mean climatology variance (Shukla and Kinter 2006).

The variations of the ensemble spread and ensemble RMSE for all individual models and the MME of each ensemble product, as a function of lead time, for the daily time scale are shown in Fig. 2. The climatology variance obtained using observations during the period 1980–2001 is also presented as a reference. Among four uncoupled models of HFP2, the Global Environmental Multiscale Model (GEM), the third-generation General Circulation Model (GCM3), and second-generation General Circulation Model (GCM2) have sufficient spread at initial lead times as their ensemble spread and ensemble RMSE lie close to each other for most of the lead times. For SEF, ensemble spread is far from the ensemble RMSE, suggesting large biases in SEF as found

in Kharin et al. (2009). Regarding coupled models, only ECMWF from ENSEMBLES has ensemble spread close to ensemble RMSE. In all the other coupled models, either from ENSEMBLES or DEMETER, the ensemble spread is considerably smaller than the ensemble RMSE at shorter lead times. This is probably due to the insufficient initial perturbation (Bengtsson et al. 2008). A typical behavior in many current ensemble forecast systems is the underestimation of the ensemble spread since many possible sources of model-related uncertainty, such as parameter uncertainty, are often not well considered (Jolliffe and Stephenson 2003).

The MME of HFP2 (MME_H hereafter), DEMETER (MME_D hereafter), and ENSEMBLES (MME_E hereafter) apparently has a better relationship of ensemble RMSE–ensemble spread than individual models. The central argument for the MME superiority over the single model ensemble has two points: 1) the former takes a holistic consideration of uncertainties from both the initial conditions and the model uncertainties (e.g., Palmer and Shukla 2000; Palmer et al. 2004) and 2) lack of understanding of atmospheric behavior could possibly be offset by different assumptions of the model framework. Thus, most of the analyses in this study focus on MME, except that individual models are required for the purpose of skill validation.

b. PNA actual skill at different time scales

In this section, the actual prediction skill of individual models at different time scales will be evaluated. The bootstrap method is used to perform the statistical significance test of correlation skill, instead of the Student's *t* test that requires the effective number of degrees of freedom, a difficult quantity to estimate for a sample with a temporal average. The bootstrap experiment was designed as follows: 1) given the leading time of prediction, the observation and prediction are paired based on the same target time of prediction, that is, constructing the sample of observation–prediction pair for the entire period 1980–2001; 2) randomly choosing 95% of the samples of the prediction–observation pair, and calculating their correlation coefficient; and 3) repeating 2) 1000 times to obtain 1000 correlation coefficients, whose standard deviation is used as a threshold value at a given confidence level (i.e., the error bar in the figures).

The skill in predicting the PNA index as a function of lead time is shown in Fig. 3 for different models at different time scales: persistence skill is also provided as a reference. As can be seen in this figure, most of the models performed better than persistence, with modest improvement at the weekly time scale, compared to the mean of daily correlation coefficients. The increase in skill is significant at the biweekly time scale, compared to the mean of daily values (e.g., first lead time at the biweekly time scale is compared with mean of daily correlation coefficients from 1 to 14 days, and so on (also see discussion of Fig. 5). For the monthly time scale, the correlation skill at lead time of 1 (month) ranges from 0.46 to 0.65 among all models, which is meaningfully higher than the mean correlation skill of daily prediction over the first 30 days (Fig. 3g). Most monthly predictions have significant correlation beyond one season. These results are consistent with a widely recognized concept that predictability can be enhanced by taking the spatial or temporal average (e.g., Lorenz 1969; Van den Dool and Saha 1990). This is because large-scale variability does not change rapidly and the growth of initial errors is relatively slow for low-frequency components.

Figure 3 shows that most of the models have RMSE smaller than persistence at all lead times, over multiple time scales. The model prediction skill is considered to be useful only if it is better than climatology prediction; thus, the bottom limit of predictability occurs when the RMSE approaches the climatological spread (Kimoto et al. 1991). As can be seen in Fig. 3b, most of the models on daily time scales reached this limit around day 12. If we define correlation less than 0.2, a statistically significant criteria value at 95% confidence level from the bootstrap experiment, as the limit of predictability, a notable point is that the predictability limit in terms of RMSE is reduced as compared with the corresponding limit using correlation. This difference might be attributed to the significance level used for the correlation skill, that is, 95%. By changing the level to 99% for correlation, the RMSE and correlation have the almost equivalent predictability limit. For a weekly time scale, the predictability limit remained comparable to the daily time scale limit (Fig. 3d) whereas improvement was substantial at the biweekly time scale, where the predictability limit increased to 4–6 weeks (Fig. 3f). The impact of slowly varying boundary forcing was prominent at monthly time scales, as the skill in terms of RMSE lingered until two months (Fig. 3h).

The prediction skill for the MME of the three ensemble products, HFP2, DEMETER, and ENSEMBLES, is shown in Fig. 4. The overall features of MME skill can be stated as follows: all MMEs beat persistence skill at almost all lead times over all time scales, demonstrating the advantage of multiple models since some individual models performed worse than persistence at longer lead times, as shown in Fig. 3. Using bootstrap experiments, the prediction skill of the MME was found to be significant for the first 20 days at daily scale, 3–4 weeks for weekly time scale, 1–2 months for biweekly, and 3 months for monthly scale, indicating that the forecast skill increased with increasing time average.

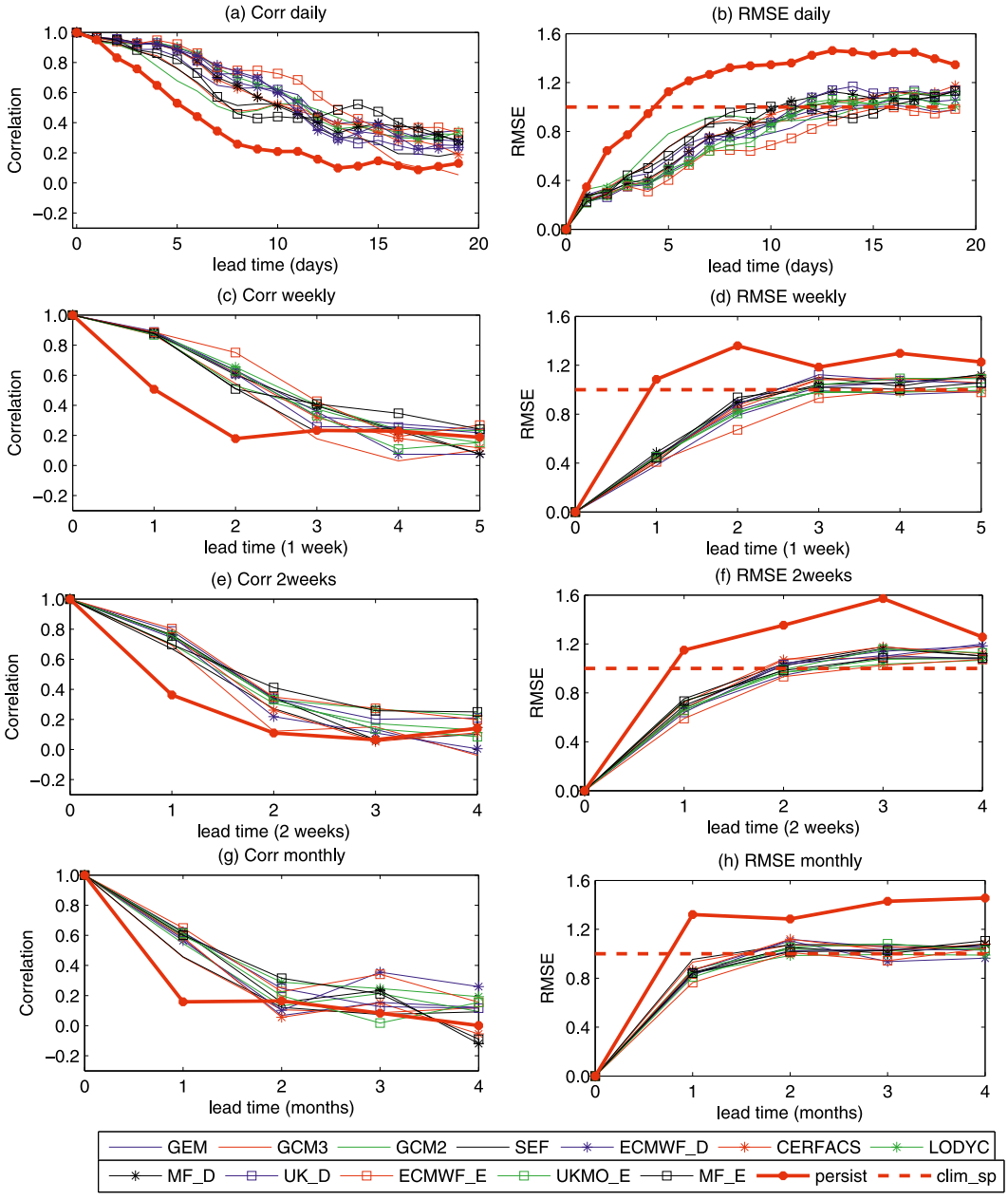

FIG. 3. Correlation and RMSE as a function of lead time for different time scales and models. Persistence skill and climatological spread is also shown for each time scale. Models are identified by color and symbols in the legend.

The increase in predictability with time scales can be further demonstrated by comparing correlation coefficients of two different scales at an equivalent lead time. For example, in Fig. 5, the average correlation coefficients of daily scale over day 1–7 (or day 8–14) are compared with the correlation skill at the 1-lead time (or 2-lead times) of weekly scale. It can be observed that the weekly skill is always larger than, or at least equivalent to, the value of the weekly average (the average of daily skill over 7 days) weekly scale. For biweekly scale, the enhancement in prediction skill is substantial even for 1-lead time (e.g., the correlation coefficient was over 0.73–0.8 for all three MMEs) as compared to the mean of corresponding daily values of the first 14 days (the correlation coefficient was 0.73 for MME_H, 0.8 for MME_D, and 0.8 for MME_E). A similar approach is applied to compare monthly time-scale correlation at different lead times, with mean of daily correlation coefficients. The enhancement in correlation skill continued at the monthly time scale where the

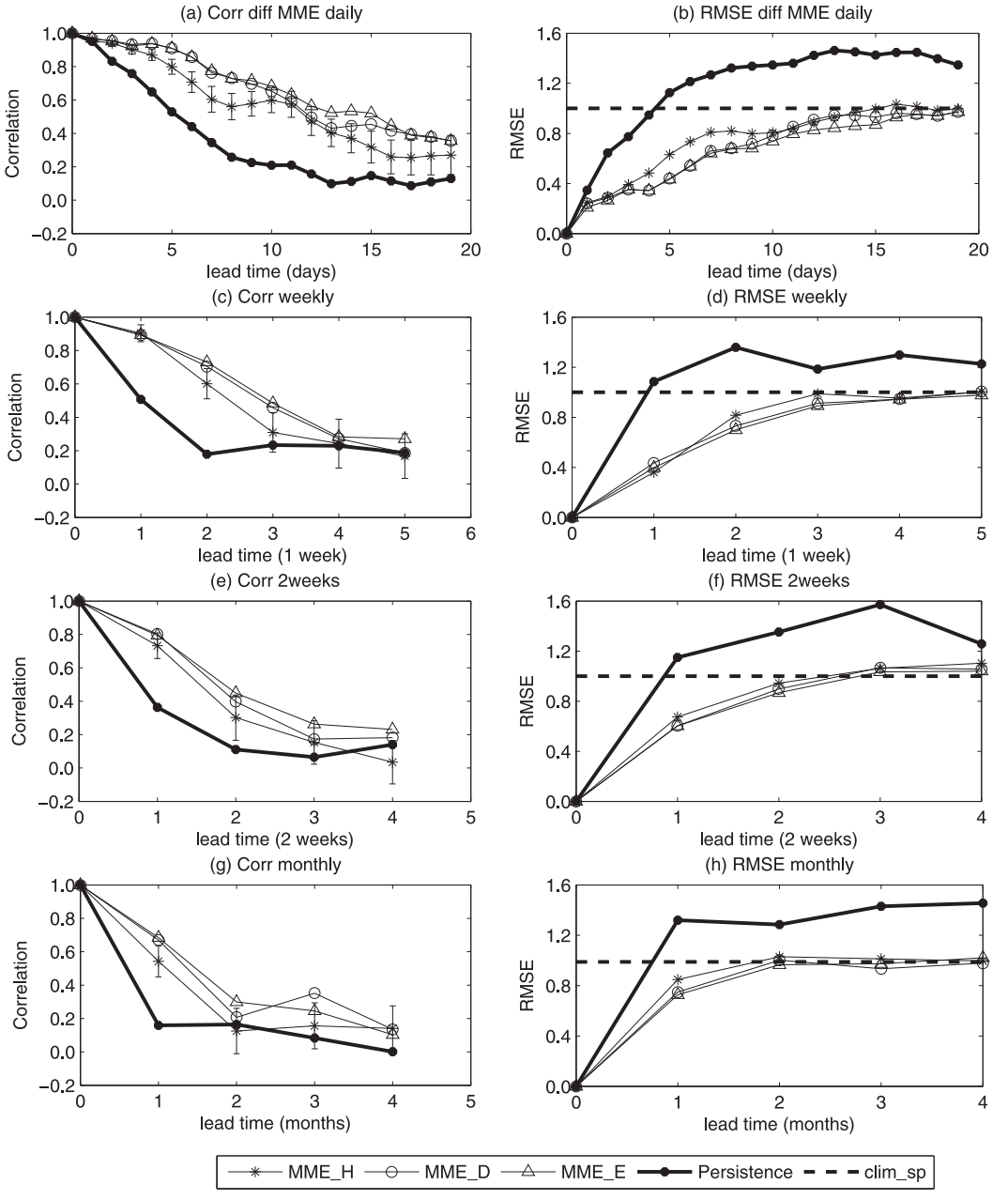

FIG. 4. As in Fig. 3, but using only MME_H, MME_D and MME_E. Vertical error bars are the sample standard deviation calculated using the bootstrap experiment at each lead time.

correlation was significant for lead times up to three months.

The RMSE skill of the MME is presented in Figs. 4b,d,f,h over different time scales. Compared with individual models, the RMSE of the MME reached error saturation (i.e., RMSE = climatological spread) at longer lead times. Similar to the correlation skill, the RMSE skill reflects the increase in predictability with time average, as indicated by the lead time at which error growth saturation is reached. For example, the time limit of predictability, in terms of the error growth saturation, increased from 14 days for the daily scale to 2 months for the monthly scale.

Thus, the actual prediction skill from the three ensemble products show that the skill increases with time averaging. Tribbia and Baumhefner (1988) argued that time averaging impacts predictability in two ways. First, it reduces the phase decorrelation rates and alleviates high-frequency noise (noise variance), which improves predictability; and second, time averaging reduces the

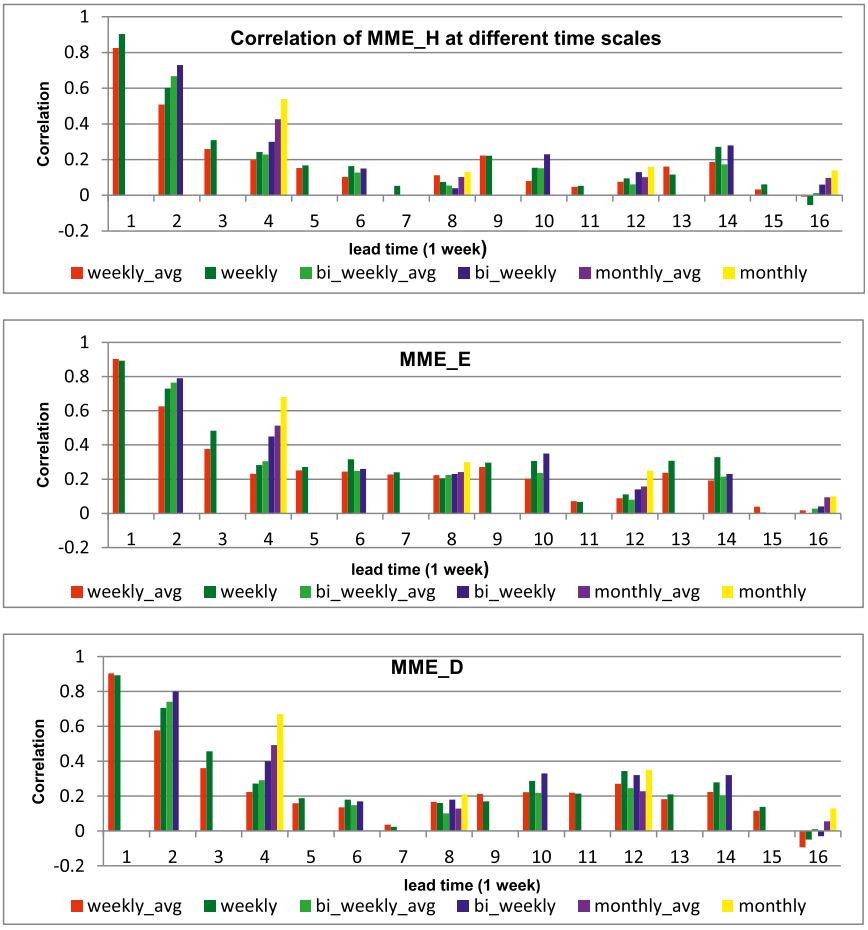

FIG. 5. Correlation coefficients of different MME over different time scales. Daily correlation values are averaged over one week (weekly_avg), 2 weeks (bi-weekly_avg), and month (monthly_avg) to compare with corresponding weekly, biweekly, and monthly time scale correlation.

climatological variance (e.g., total variance), which opposes the greater predictability trend. Thus, a net effect of time averaging is the competition between both contributions to predictability. Following Tribbia and Baumhefner (1988), the decrease in climatological variance and the variation of decorrelation with time averaging are estimated and compared with the correlation skill at different time scales (cf. Fig. 6). Here, only MME_H is randomly selected for comparison owing to the similarity of prediction skill of these models. The impact of time averaging is clearly observable in Fig. 6. For example, weekly, biweekly, and monthly time-scale forecasts are more predictable than daily forecast skill in Fig. 6a, whereas the decrease in climatological variance and autocorrelation (persistence) increase with time average as shown in Figs. 6b and 6c, respectively. A comparison among Figs. 6a–c reveals that the increase in the length of decorrelation (persistence skill) overweighs the decrease in climatological variance so that the former dominates the predictability. Thus, the increase in predictability with averaging time scale, even in the presence of a decrease in climatology variance, is due to two reasons: (i) the averaging lessens the noise, making the signal at initial conditions less dissipated, and (ii) the role of slowly varying boundary forcing becomes more and more influential with time average.

5. PNA potential predictability

a. MI-based predictability

Figure 7 shows AC_{MI} and AC_p for ENSEMBLES and HFP2 for different time scales. Two obvious features can be observed in this figure. First, the coupled model ENSEMBLES has significantly better skill than the uncoupled model HFP2 at multiple time scales, in contrast with the comparison of actual skills (Fig. 4). Second, the AC_{MI} is always larger than AC_P, as indicated by

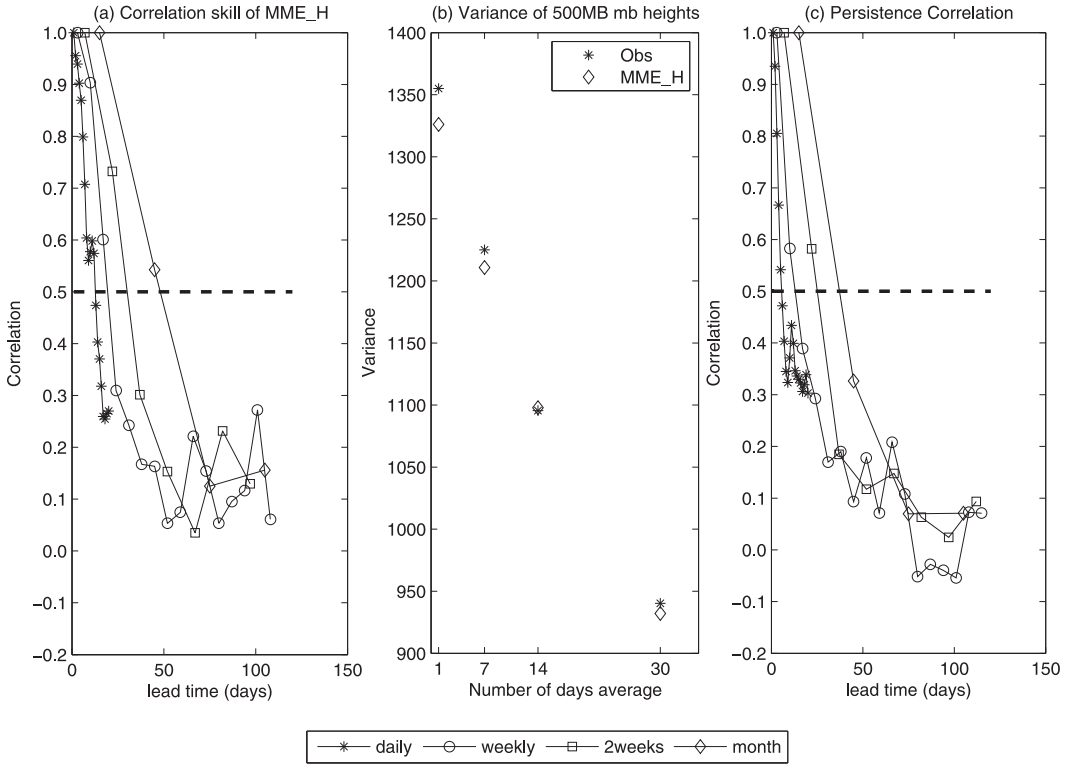

FIG. 6. (a) Anomaly correlation for MME_H over different time scales as a function of lead time, the daily correlation for first 20 days only. (b) Climatological variance of averaging time using observation (asterisks) and MME_H (open diamonds), and (c) persistence correlation at different time scales. The legend for (a),(c) is shown at the bottom.

(5), suggesting that the conventional predictability measure of signal-to-noise ratio (SNR) underestimates the potential predictability. These two features will be discussed further after we examine the variation in potential predictability with the averaging time scale, where ENSEBLES and AC_{MI} as the target of analysis are used owing to their better representation than HFP2 and AC_P.

Similar to actual skill, the potential predictability of the PNA pattern also increased with the time scale, as shown in Fig. 7. The improvement in potential predictability over weekly time scale, (Fig. 7c) is significant compared to daily time scale (Fig. 7a). The enhancement in predictability persisted at the biweekly time scale (Fig. 7e) and monthly time scale (Fig. 7g), where the potential predictability is meaningful for all four months. Table 1 shows the maximum lead time that remains at a correlation skill of 0.5 for all time scales.[2] The table shows that AC_{MI} remained significant until 40 days for MME_E and 28 days for MME_H at the daily time scale. In contrast with actual skill, the improvement in skill is also quite substantial even at weekly time scales (16 weeks for MME_E and 5 weeks for MME_H).

b. Comparison between MI and SNR

We also explored predictability using the SNR, where signal and noise are estimated using Rowell's scheme (Rowell 1998). The results show that the predictability by SNR has features similar to those by MI; namely, that the SNR increased with time averaging.

Figure 7 shows the AC_{MI} and AC_p as a function of lead time over different time scales. The most prominent features here are that the skill decreases with lead time and the AC_{MI} is higher than AC_p. The former is in reasonable agreement with the general conclusion that predictability declines with lead time in chaotic or stochastic dynamical systems whereas the latter is consistent with the theoretical in (5) as mentioned above. As shown in Fig. 7, at daily time scale, both AC_{MI} and AC_p were close to each other for the first few lead times (Figs. 7a,b), suggesting that the ensemble variance does not differ much from one forecast to the other for these lead times.

[2] The correlation value 0.5 is partially arbitrarily set, but is often used as a threshold of useful prediction skill in the seasonal prediction community.

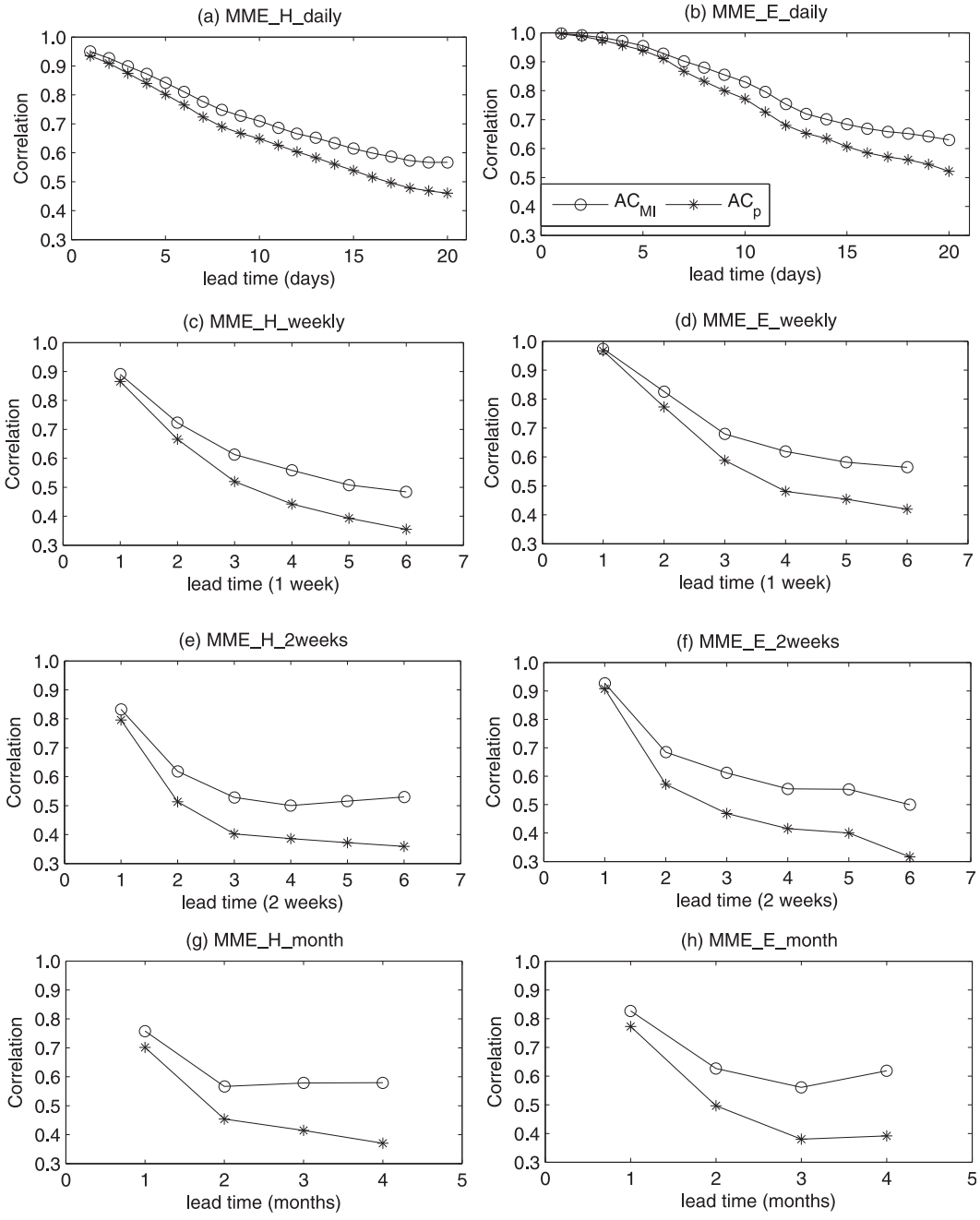

FIG. 7. Comparison of AC_{MI} with AC_p over different time scales for MME_H and MME_E.

For subsequent lead times, AC_{MI} was significantly larger than AC_p, simply because the ensemble variance was no longer approximately constant after sufficient development with lead times. The same phenomena can be observed at weekly and biweekly time scale. Over monthly time scale, even at 1-month lead time, AC_{MI} was better than AC_p (Figs. 7g,h).

The comparison of actual and potential skill using MME_E and MME_H at different time scales is shown in Fig. 8. Here, potential skill is measured using AC_{MI}. It can be observed that the potential prediction skill is higher than actual skill. Table 1 also shows the skill limit for both actual and potential predictability. It is clear that the potential predictability limit is quite large compared to actual skill at multiple time scales for both MME_H and MME_E. It is not surprising since the potential predictability represents an upper limit of prediction skill that a perfect model can achieve. That

TABLE 1. Actual and potential prediction skill (AC_{MI}) over different time scales using MME_H and MME_E (the correlation value 0.5 is chosen as the threshold for significance).

Time scale	AC_{MI}_ MME_H	AC_{MI}_ MME_E	Actual_skill_ MME_H	Actual_skill_ MME_E
Daily	28	40	11	15
Weekly	5	16	2	2
Biweekly	8	8	1	1
Monthly	4	4	1	1

the potential predictability is higher than actual skill suggests that there may be a lot of room to improve present prediction skill of PNA.

It should be emphasized here that the potential predictability may be smaller than the actual skill if the model is imperfect in estimating noise variance. For example, Batté and Déqué (2011) found that, if the ensemble is overdispersive and ensemble variance is overestimated, the potential predictability could be underestimated compared with the actual skill. On the other hand, if model error only exists in the ensemble mean (signal) as random error, the potential predictability is greater than the actual skill, as proven by Sardeshmukh et al. (2000). Overall, there are three kinds of errors impacting atmospheric predictability: nonlinear chaos, model random error, and model systematic error. The difference between actual and potential predictability is due to inclusion of these three factors in estimating the actual skill.

Another possible reason for the difference between actual and potential skill shown in Fig. 8 is due to the error of Gaussian assumption used in calculating potential predictability. To examine this, we estimated the probability distribution function (PDF) for all lead times and all time scales for HFP2 and ECMWF. The results show that the Gaussian assumption is always approximately held for any case. Shown in Fig. 9 are the PDFs of different randomly chosen lead times for the weekly time scale. Figure 9 also shows that there is no essential difference of Gaussian approximation between coupled and uncoupled models, indicating that the difference of potential skill between coupled and uncoupled models is not due to their disparity in assuming a Gaussian approximation. We will go back this issue in next subsection.

6. Predictability of coupled and uncoupled models

It is intriguing to explore the difference in predictability between coupled and uncoupled models. First, the difference of their actual prediction skill is examined. Using a bootstrap test, it is found that their difference in actual skill, shown in Fig. 4, is not statistically significant as indicated by the sampling error bars exceeding these differences. This is consistent with the findings of Johansson (2007), where he mentioned that coupling is not relevant in enhancing the prediction skill of PNA.

Figure 10 compares the difference of the potential predictability AC_{MI} between coupled model (MME_E) and uncoupled model (MME_H) for different time averaging. Again, the bootstrap method is used to estimate the extent of uncertainty owing to sampling error, as shown by the bars in this figure. A significant difference beyond the sampling uncertainty is witnessed in potential skill between MME_E and MME_H for all time scales, as shown in this figure, in contrast with actual skill results. Thus, one compelling question is why coupling does not lead to a better PNA prediction skill although it has higher potential predictability? To shed light on this issue, the prediction skill of SSTA (70°S–70°N) is calculated from MME_H and MME_E at the monthly time scale. The SSTA prediction is provided by the persistent scheme for uncoupled models (MME_H) and the oceanic component of the coupled models (MME_E). Figure 11 shows that the MME_E SST prediction skill is better than MME_H (persistence) in many areas, especially in the tropics and over North America (PNA index). With the increase in lead time, MME_E has higher skill for SSTA prediction than MME_H. Thus, the reason that coupled models do not lead to better PNA prediction than uncoupled models is probably because of the bias of atmospheric models, which can cause the coupling to be inconsistent with the observation, resulting in the skill in coupled model to not necessarily be better than that of the uncoupled model. One may understand the prediction of the MME_E and MME_H as two atmospheric model runs: one forced with persistent SST and the other forced with coupled model SST. The coupled model SST is better than the persistent SST (uncoupled model). If the atmospheric model was perfect, the atmosphere model would have more realistic forcing (prediction) while forced by the coupled model SST. However if the atmospheric model has bias, this conclusion may not hold. It should be noted that the coupled model SST beating the persistent SST prediction (uncoupled model) does not conflict with the above argument since the SST prediction forced by the biased atmosphere may still be better than the persistent SST prediction.

7. Temporal variation in PNA predictability and its source

In the last section, the overall potential predictability is discussed using mutual information. In this section, the emphasis will be put on the temporal (e.g., interannual or

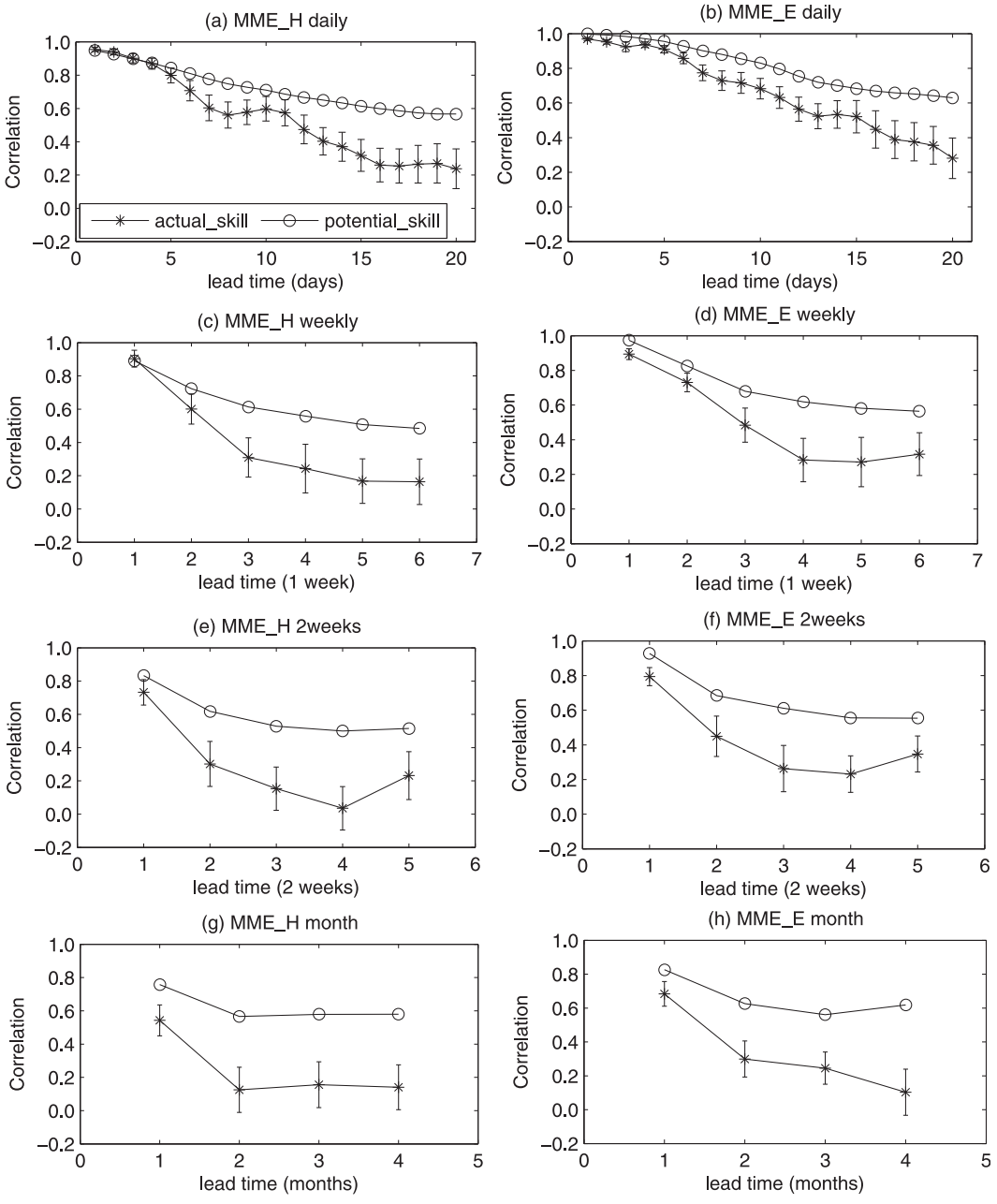

FIG. 8. Actual and potential skill for MME_H and MME_E; vertical error bars are the sample standard deviation calculated using the bootstrap experiment (see context).

decadal) variation in potential predictability and its source. Shown in Fig. 12 is the averaged RE_A over the effective lead times for all individual predictions by ENSEMBLES. Here, the effective lead time is equivalent to the maximum lead time beyond which the prediction skill is not significant and is determined by bootstrap experiment. A striking feature in Fig. 12 is that there are temporal variations in potential predictability of PNA in terms of RE_A, where large RE_A are mainly located in fewer predictions, such as 1982–83 and 1997–98 (strong ENSO events). For most of the other predictions, RE_A is small or exhibits small variations with initial time. Next, we will examine what determined these variations in RE_A at different time scales.

As described in section 3, RE_A is the sum of dispersion and signal components [cf. (4)]. The variations in both signal and dispersion components for MME_E over different time scales are shown in Fig. 13 for the period

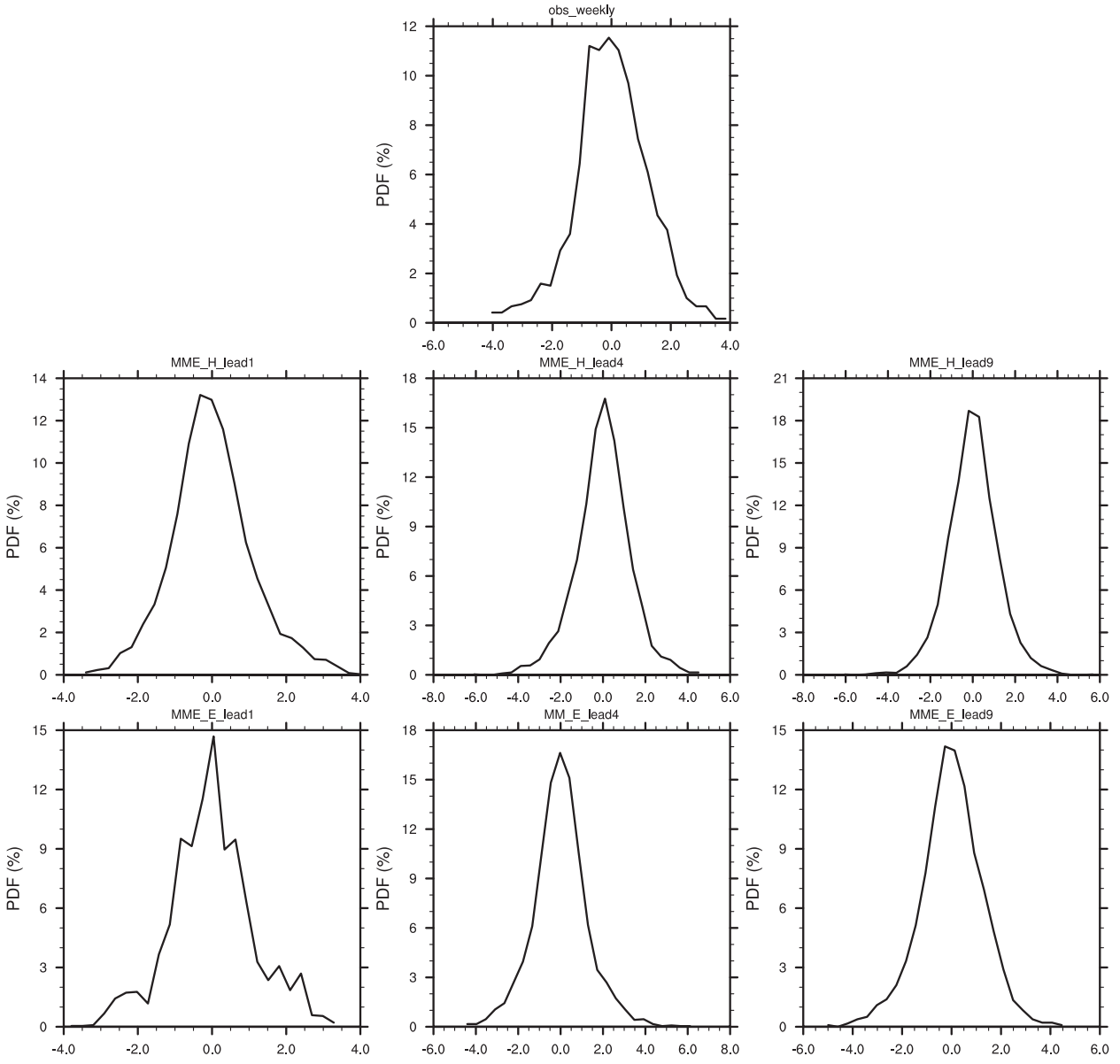

FIG. 9. The PDF of prediction for different randomly chosen lead times for the weekly time scale in the ECMWF and HFP2 models: the observed PDF for weekly time scale is also shown as reference at the top.

1980–2001. As can be seen in Fig. 13, the signal component is large in the predictions for 1980–85 and 1995–2001. In between, the signal is low (1985–95). The variation feature might be related to variation in the tropical Pacific SST forcing. For example, the two periods with large signal components have the strongest El Niño events (1982/83 and 1997/98). To explore the relationship between SST forcing and the potential predictability, lead–lag temporal correlation between PNA index and SSTA is calculated for each grid at monthly time scale over the region 30°N–30°S, 150°E–90°W using the periods of high signal component, that is, 1980–85 and 1996–2001, as shown in Fig. 14.

Figures 14a,c,e,g show the pattern correlation between the PNA index and predicted SST (ensemble mean prediction of SST) for ENSEMBLES. It can be observed that significant correlation mainly appears in the tropical central eastern Pacific (ENSO region). The correlation patterns resemble a typical ENSO pattern, implying that the ENSO is a major source of PNA potential predictability. This result also favors the notion that the role of tropical SST forcing is to amplify atmospheric variability, such as the PNA, by impacting its signal (ensemble mean) (e.g., when forcing is high, the PNA signal is strong and the prediction skill is high, and

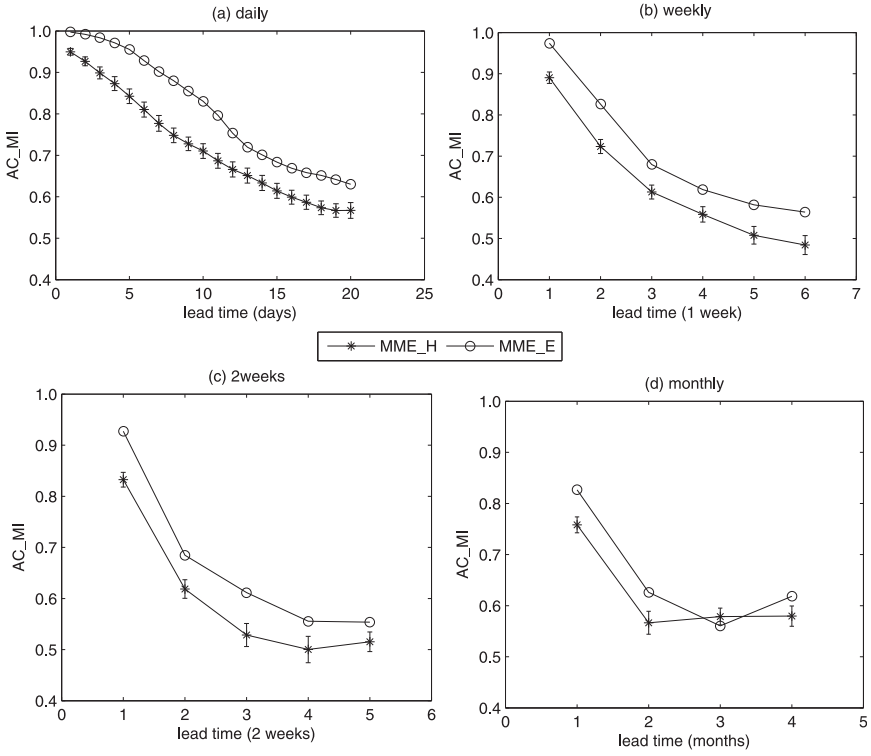

FIG. 10. AC_{MI} as function of lead time over different time scales for MME_E and MME_H is presented: vertical error bars are the sample standard deviation calculated using bootstrap experiment (see context).

vice versa). In particular, maximum correlation is found when SST leads by two and three months (Figs. 14e,g) (e.g., 0.6–0.8 in ENSO region, significant at the 95% confidence level). A high correlation is also found when the PNA observed index is lagged-correlated with observed SSTA by 2–3 months (not shown). The lagged correlation of 2–3 months of ENSO and PNA has also been reported by Munoz et al. (2010). A similar correlation pattern is also obtained for MME_H where the persistence prediction of SSTA was correlated with PNA index as shown in Figs. 14b,d,f,h.

8. Predictable component analysis

In preceding sections, we discussed the PNA predictability at various time scales and its possible sources related to ENSO forcing. In these analyses the PNA index, defined by the time series of the leading PCA mode, has been used. This simplifies the analysis and easily captures the main feature of PNA predictability. However, such a one-dimensional index may not well characterize some features of the multiple space of atmospheric circulation, especially the atmospheric response to SST forcing. Thus, in this section, we will explore the predictability of atmospheric variability over the Northern Hemisphere, especially the position and role of the PNA predictability in the total predictability of atmospheric variability.

To achieve the above goal, we performed the predictable component analysis (PrCA) using monthly anomaly data of 500-mb geopotential height of MME_E and MME_H over the Northern Hemisphere for each lead time of prediction. As discussed in section 3b(2), we obtained multiple PrCA modes for each lead time of prediction-like multiple modes in PCA. In this study, we only consider the modes that can characterize the PNA. Shown in Figs. 15 and 16 are the PNA-like modes of PrCA using MME_E and MME_H at different lead months, chosen from the first four modes, as higher modes have very small and negligible contribution toward total predictability. As can be seen in these figures, the PNA-like pattern usually appears as the second mode in the PrCA except as the first mode in the prediction of lead time of 1 month for MME_E.[3] This indicates that the PNA is one of the most predictable structures among low-frequency atmospheric variability. It should be noted that

[3] A further examination found that most first modes of the PrCA characterize a NAO-like pattern, which we will not discuss here.

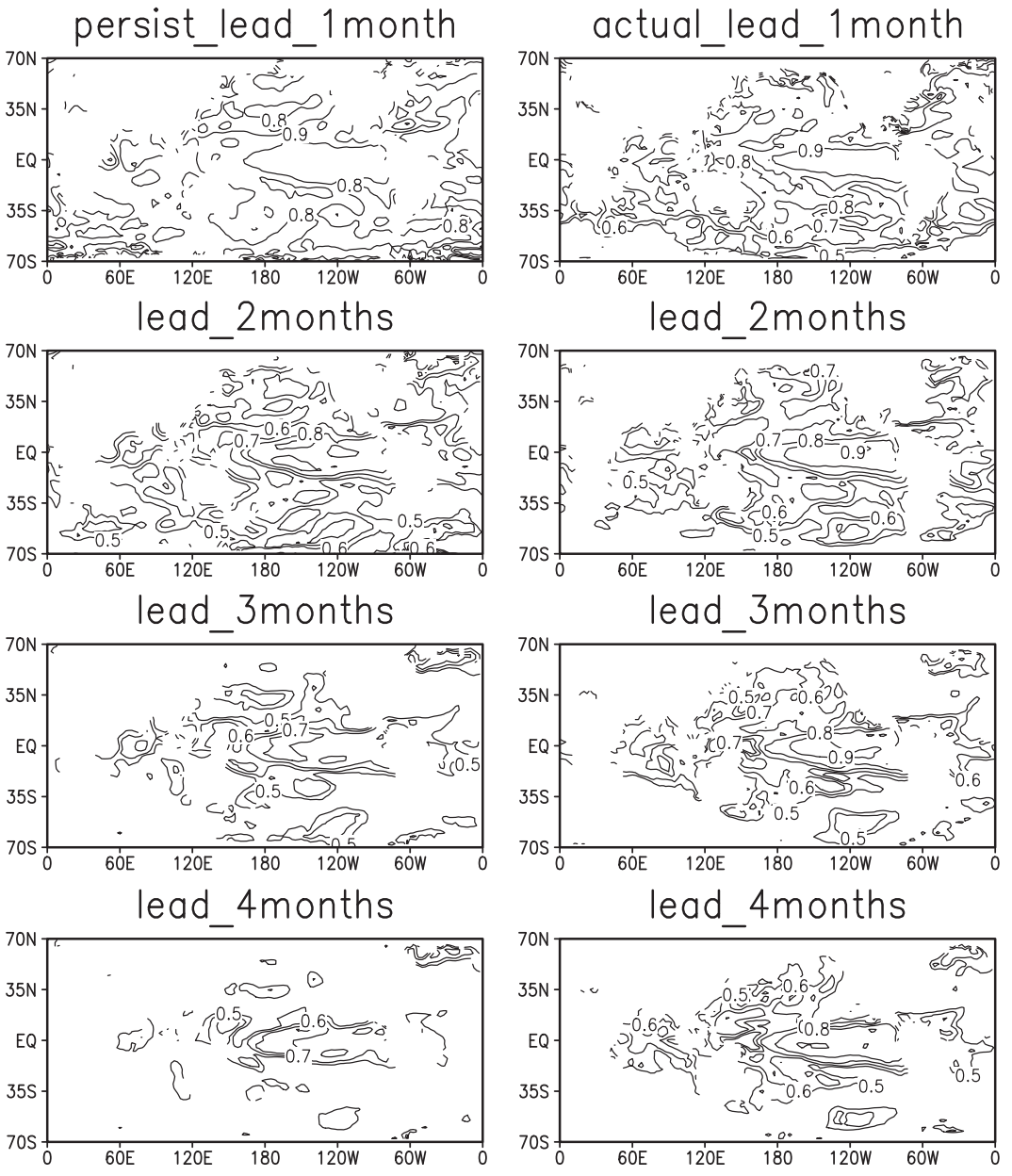

FIG. 11. Monthly correlation skill of SST using (left) persistence and (right) MME_E at different lead months.

the first pattern in PrCA, unlike PCA analysis, does not necessarily explain the most variability, as shown within parentheses above each panel of Figs. 15 and 16. This is justified because the PrCA seeks optimal modes based on predictability rather than on explained variance (variability).

To gain further insight into the source of PNA predictability, we projected the time series of PrCA patterns in Figs. 15 and 16 onto the observed SSTA of the tropical Pacific region. Shown in Fig. 17 are the projected patterns, which resemble the warm phase of ENSO. Figure 17 builds a direct bridge between the predictable components of PNA with ENSO, namely that the most predictable component of PNA is mainly due to the tropical SST forcing.

9. Discussion and conclusions

In this study, the PNA predictability has been explored using actual and potential predictability measures for individual and a multiple-model ensemble of three ensemble prediction datasets, HFP2, DEMETER,

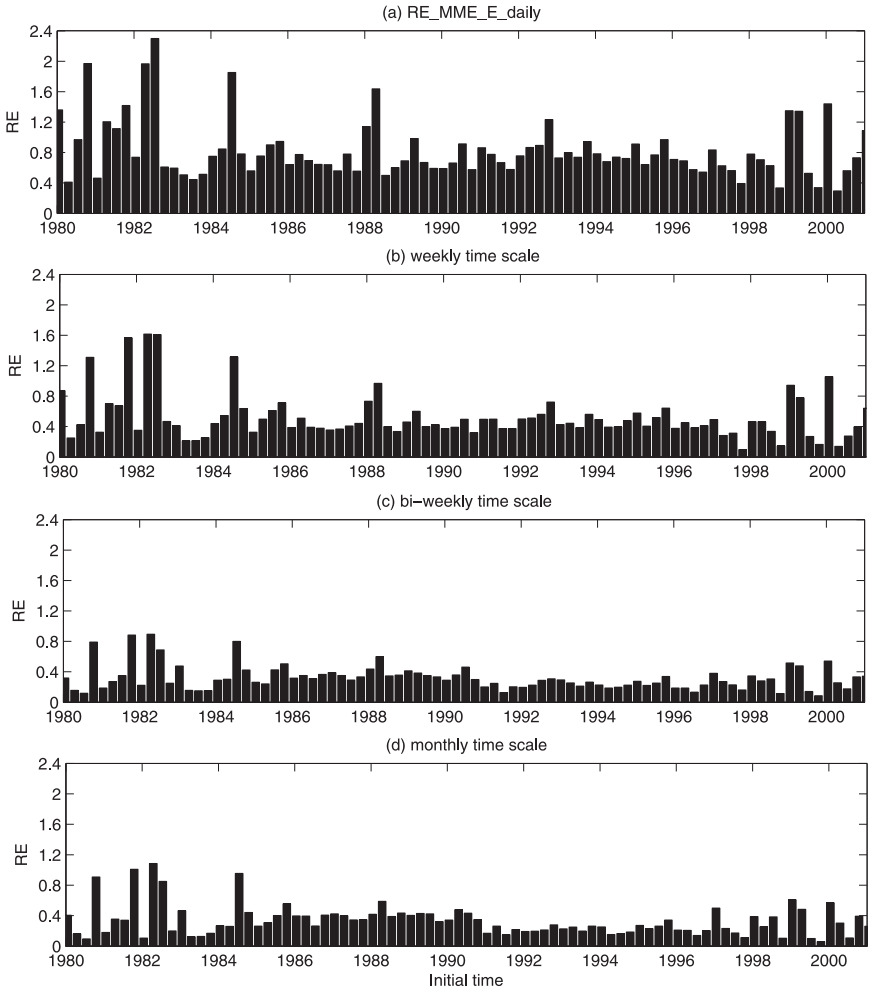

FIG. 12. Averaged relative entropy as a function of initial time is shown at different time scales for MME_E.

and ENSEMBLES stream 2. The first one is an ensemble of uncoupled models, whereas the latter two are the products of fully coupled models. The primary purpose of this study is to examine the efficacy of time averaging with regard to extracting a predictable signal from day-to-day weather fluctuations. The emphasis was put on actual skill and potential predictability over different time scales from days to seasons and the intercomparison of different scales and different models. The actual skill measures include correlation and RMSE, and the potential skill measures include signal to noise ratio and information-based metrics.

The comparison of MME skill with individual models in terms of actual skill is first evaluated. It was found that the MME prediction skill was better than most of the individual models. As a result, the MME is used to study PNA predictability. A comprehensive analysis reveals that, on the daily time scale, the PNA predictability limit is around 20 days. A modest improvement can be observed on the weekly time scale. The effect of time averaging was more pronounced at the biweekly time scale, and persevered in the monthly time scale to 3 months. Overall, predictability is seen to benefit with time averaging. The phase decorrelation rate decreased with time averaging, which resulted in an increase in predictability, but a reduction in climatological variability with the time average, on the other hand, restricted further improvements of time averaging to predictability. Thus, the predictability of filtered fields is a trade-off between these two factors.

For the first time, the PNA potential predictability is evaluated at multiple time scales. The relative entropy (RE_A) and MI-based anomaly correlation (AC_{MI}) served the purpose of measuring potential skill at multiple time scales using MME_H and MME_E. These measures were found to be consistent with the notion

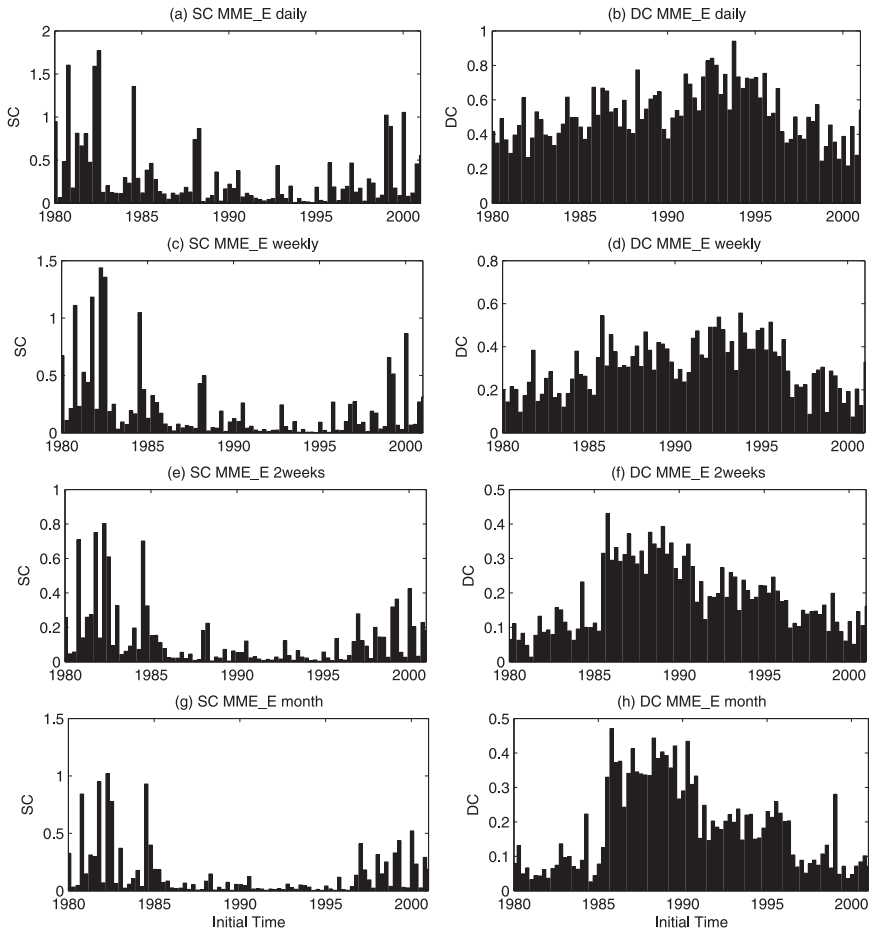

FIG. 13. Variations in averaged (left) signal and (right) dispersion components as a function of initial time for MME_E over different time scales.

that model prediction skill and predictability generally decreased with lead time (monotonicity). A practical comparison between the mutual-information-based AC_{MI} with SNR-based AC_p with averaging time affirm their theoretical relationship [cf. (5)] that former can measure more potential predictability than the latter. One advantage of using AC_{MI} is that it can measure the statistical dependence, linear or nonlinear, between ensemble mean and an ensemble member (hypothetical observation), whereas AC_p measures only linear dependence. When the prediction and climatological distribution are Gaussian and ensemble spread is invariant with forecast, the two measures are equivalent.

The RE_A is an effective measure in distinguishing the individual forecasts from each other where some forecasts are more predictable than others (it is not explicitly confirmed in this study but using RE_A as measure potential predictability the large RE_A corresponds to more prediction skill than low RE_A). It has been observed that RE can detect the signals related to SST forcing in terms of its inherent association with signal component, whereas other information-based metrics are lacking this property. This property of RE_A makes it more preferable to be used, as compared to other potential predictability measures. In this study, this association has been reconfirmed for the PNA region, irrespective of time scale of interest, by observing temporal variabilities in RE_A and signal component. The high amplitude of the signal component in some particular initial conditions, especially at monthly to seasonal time scale, was related to SST anomalies in the tropical Pacific (ENSO) region.

Comparison of predictability between coupled and uncoupled models reveals that the potential predictability of the PNA index measured using coupled models is larger than for uncoupled models, at multiple scales, in contrast with the actual skills, which show no significant difference between coupled and uncoupled models. To shed light on a possible reason, the SSTA prediction skill is evaluated for coupled and uncoupled models on monthly time scale at different lead times. A

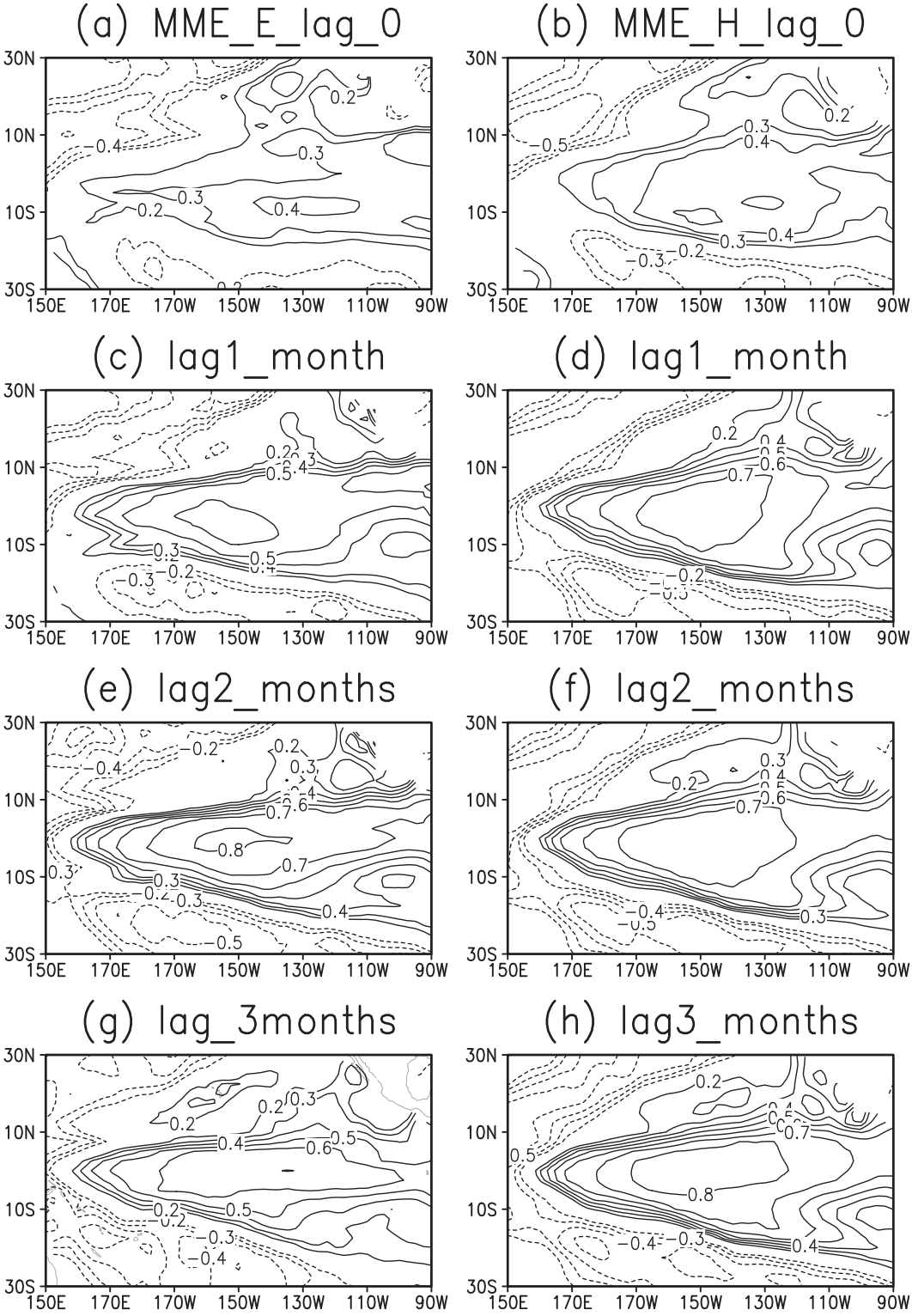

FIG. 14. Spatial pattern correlation between (left) predicted ENSEMBLES SST and PNA for MME_E at different time lags (based on selected initial conditions in which signal component was high), SST leading and PNA lagging behind; (right) as for predicted, but using observed SST and PNA of MME_H.

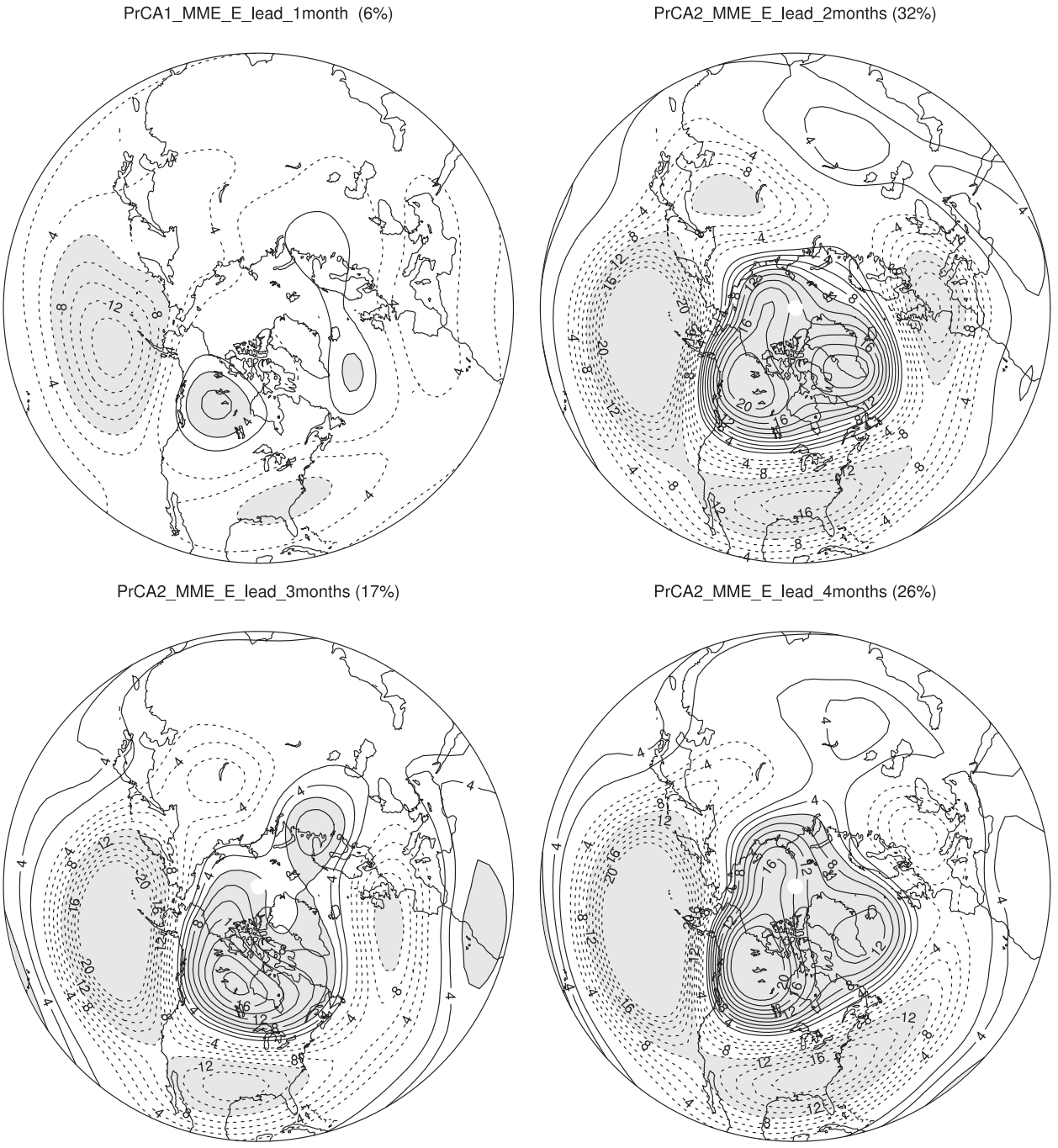

FIG. 15. PrCA patterns at different lead months for MME_E and MME_H that capture the PNA-related variability. For each pattern, the variance explained based on total predictability is shown (within parentheses) at the top.

persistent SSTA is used for uncoupled models, whereas a predicted SSTA from coupled models is analyzed. It was found that the SST prediction skill in coupled models is much better, even at a lead time of four months, especially in the tropical and PNA regions. Thus, atmospheric bias is the most probable reason responsible for weakening, or offsetting, the advantage of a coupled model in predicting PNA.

The source of PNA potential predictability from the monthly to seasonal time scale has also been investigated in this study. It is well known that variations in the North Pacific region are teleconnected with variations in the

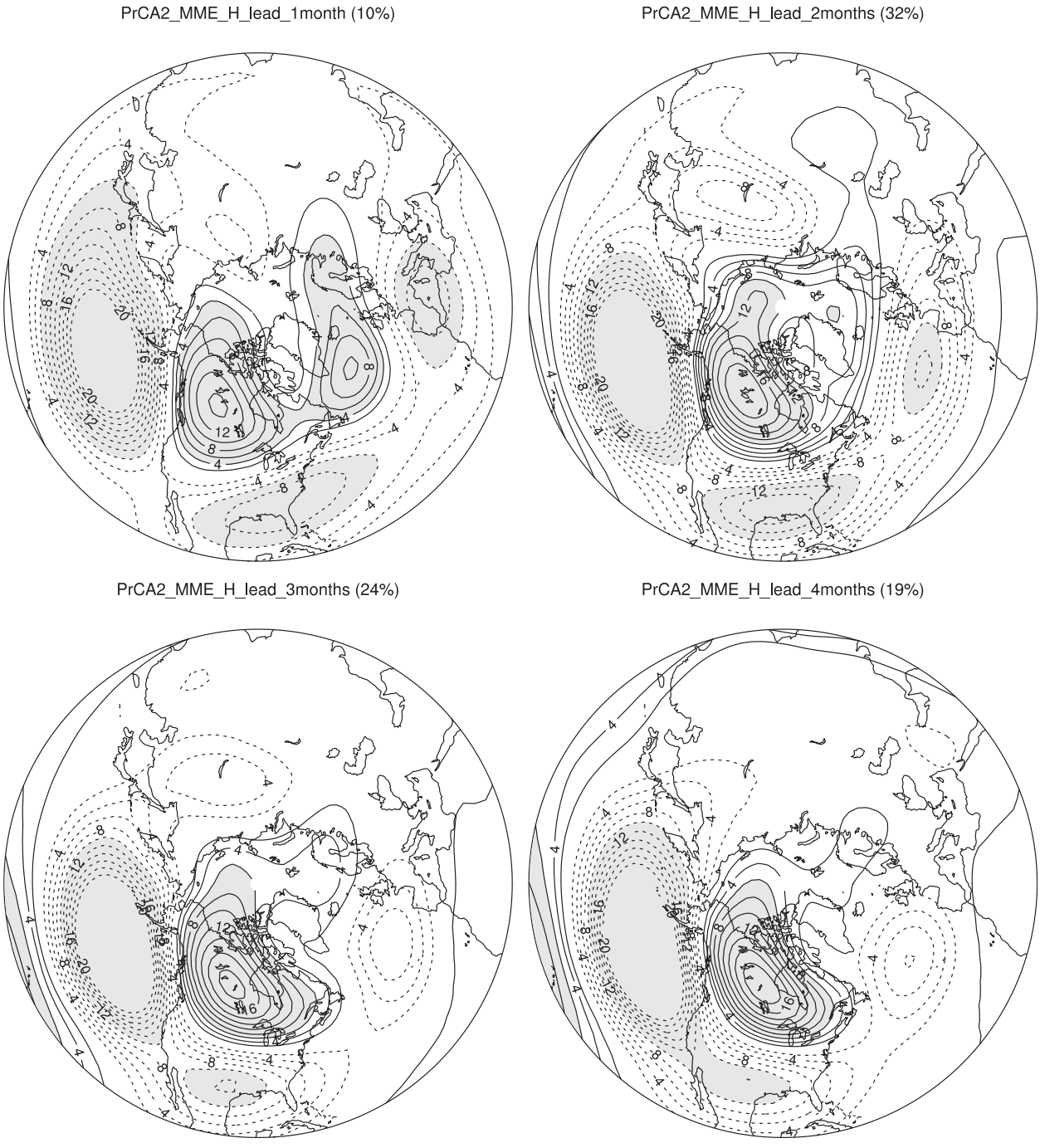

FIG. 16. As in Fig. 15, but using MME_H.

tropical SST forcing (Horel and Wallace 1981; Hoskins and Karoly 1981; Simmons 1982; Sardeshmukh and Hoskins 1988; Straus and Shukla 1997, 2002). Thus, emphasis was placed on the role of ENSO in PNA predictability, which was explored using lead–lag correlation between the PNA index and the SST signal. It was found that the tropical SST forcing impacts the PNA potential predictability mainly by changing the amplitude of the ensemble mean. The SSTA–PNA correlation patterns in the models resemble the typical ENSO pattern, suggesting that the ENSO is the main source of PNA seasonal predictability. Their maximum correlation occurred at a lag of 2–3 months, which is consistent with some past studies.

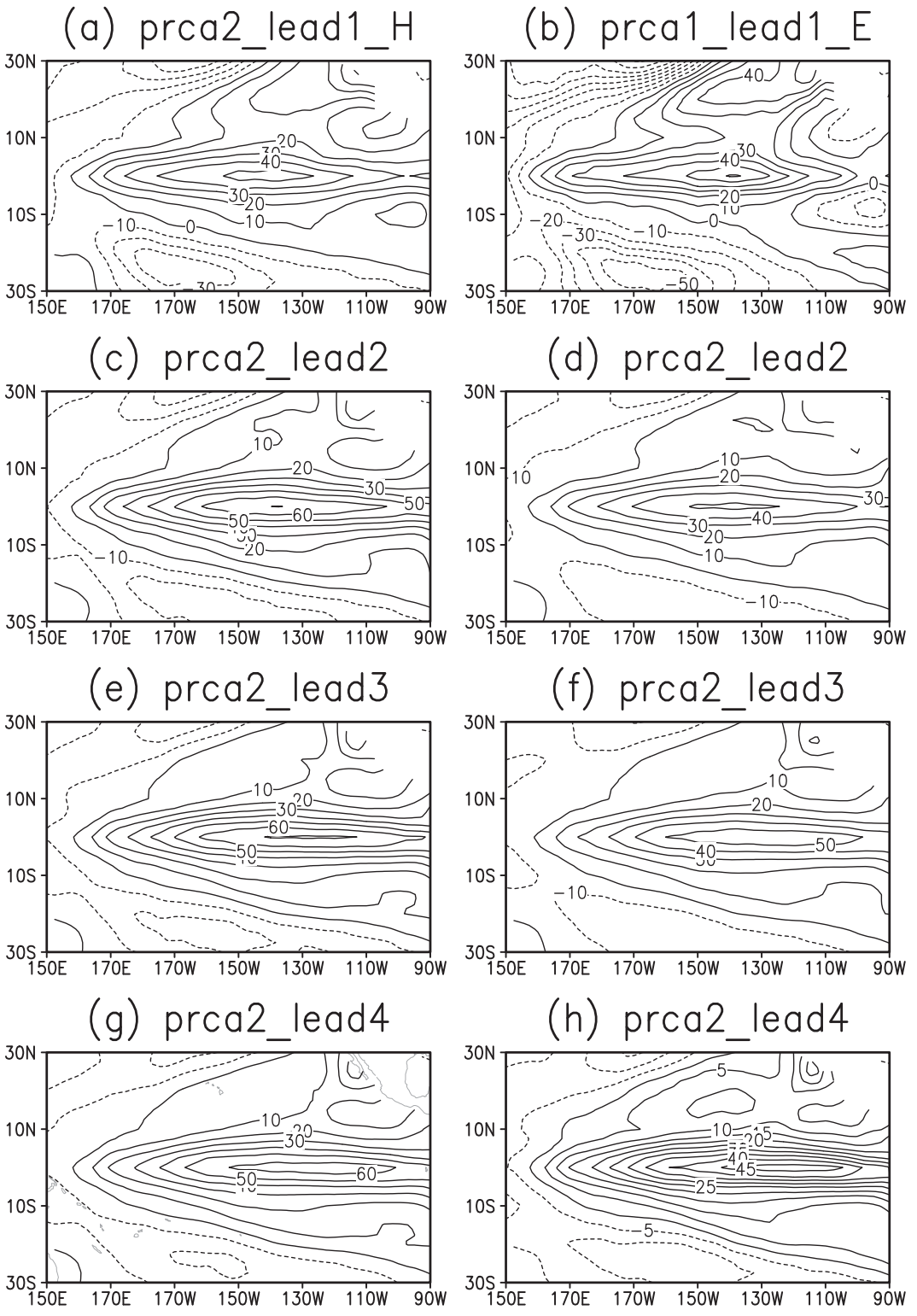

FIG. 17. As in Fig. 14, but for the projection of the PrCA patterns on observed SST.

PrCA analysis explores the PNA predictability and its source in the data space with temporal and spatial variation. It was found that the PNA is one of two most predictable patterns among low-frequency atmospheric variability. Its main source of predictability is attributed to the tropical SSTA forcing, that is, ENSO.

Acknowledgments. This work is supported by NSERC Discovery Grant and UNBC Graduate Entrance Scholarship, as well as the research grant from NSF of China (41276029).

APPENDIX

Relationship between AC_{MI} and AC_p

If the prediction and climatological PDFs are Gaussian, the mutual information (MI) can be written as (e.g., Yang et al. 2012; Tang et al. 2013, manuscript submitted to *J. Climate*)

$$\mathrm{MI} = \frac{1}{2}(\ln\sigma_q^2 - \langle \ln\sigma_p^2 \rangle) \geq \frac{1}{2}(\ln\sigma_q^2 - \ln\langle\sigma_p^2\rangle) = -\frac{1}{2}\ln\left(\frac{\langle\sigma_p^2\rangle}{\sigma_q^2}\right) = -\frac{1}{2}\ln(1-\mathrm{STR}) \quad \text{(A1)}$$

in which

$$\mathrm{STR} = \sigma_p^2/\sigma_q^2, \quad \text{(A2)}$$

where σ_q^2 and σ_p^2 have the same meaning as in (4). The inequality in (A1) is due to the fact that arithmetic mean is larger than or equal to the geometric mean. A detailed derivation can be found in Yang et al. (2012). The STR can be interpreted as the perfect correlation skill AC_p that is defined by the correlation between ensemble mean and a random ensemble member (e.g., Tang et al. 2013, manuscript submitted to *J. Climate*):

$$\mathrm{AC}_p = \mathrm{sqrt}(\mathrm{STR}). \quad \text{(A3)}$$

Using (A3), (A1) can be rewritten as

$$\mathrm{MI} \geq -\frac{1}{2}\ln(1-\mathrm{AC}_p^2). \quad \text{(A4)}$$

Namely,

$$\mathrm{AC}_p \leq \mathrm{sqrt}(1-e^{-2\mathrm{MI}}). \quad \text{(A5)}$$

The MI-based correlation AC_{MI} is defined by (e.g., Delsole 2004; Yang et al. 2012)

$$\mathrm{AC}_{\mathrm{MI}} \equiv \sqrt{1-e^{-2\mathrm{MI}}}. \quad \text{(A6)}$$

Thus, we can have the relationship between AC_{MI} and AC_p as expressed by (5).

REFERENCES

Abramov, R., A. Majda, and R. Kleeman, 2005: Information theory and predictability for low-frequency variability. *J. Atmos. Sci.,* **62,** 65–87.

Barnett, T. P., K. Arpe, L. Bengtsston, M. Ji, and A. Kumar, 1997: Potential predictability and AMIP implications of midlatitude climate variability in two general circulation models. *J. Climate,* **10,** 2321–2329.

Batté, L., and M. Déqué, 2011: Seasonal predictions of precipitation over Africa using coupled ocean–atmosphere general circulation models: Skill of the ENSEMBLES project multimodel ensemble forecasts. *Tellus,* **63A,** 283–299.

Bengtsson, L. K., L. Magnusson, and E. Kallen, 2008: Independent estimates of asymptotic variability in an ensemble forecast system. *Mon. Wea. Rev.,* **136,** 4105–4112.

Chen, W. Y., and H. Van den Dool, 2003: Sensitivity of teleconnection patterns to the sign of their primary action center. *Mon. Wea. Rev.,* **131,** 2885–2899.

Cheng, Y., Y. Tang, and Y. D. Chen, 2011: Relationship between predictability and forecast skill of ENSO at various time scales. *J. Geophys. Res.,* **116,** C12006, doi:10.1029/2011JC007249.

Chervin, R. M., 1986: Interannual variability and seasonal climate predictability. *J. Atmos. Sci.,* **43,** 233–251.

DelSole, T., 2004: Predictability and information theory. Part I: Measures of predictability. *J. Atmos. Sci.,* **61,** 2425–2440.

——, and M. K. Tippett, 2007: Predictability: Recent insights from information theory. *Rev. Geophys.,* **45,** RG4002, doi:10.1029/2006RG000202.

Derome, J., and Coauthors, 2001: Seasonal predictions based on two dynamical models. *Atmos.–Ocean,* **39,** 485–501.

Doblas-Reyes, F. J., V. Pavan, and D. B. Stephenson, 2003: The skill of multimodel seasonal forecasts of the wintertime North Atlantic Oscillation. *Climate Dyn.,* **21,** 501–514.

Horel, J. D., and J. M. Wallace, 1981: Planetary-scale phenomena associated with Southern Oscillation. *Mon. Wea. Rev.,* **109,** 813–829.

Hoskins, B. J., and D. J. Karoly, 1981: The steady linear response of the spherical atmosphere to thermal and orographic forcing. *J. Atmos. Sci.,* **38,** 1179–1196.

Hurrell, J. W., 1995: Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science,* **269,** 676–679.

——, Y. Kushnir, G. Ottersen, and M. Visbeck, 2003: An overview of the North Atlantic Oscillation. *The North Atlantic Oscillation: Climate Significance and Environmental Impact, Geophys. Monogr.,* Vol. 134, Amer. Geophys. Union, 1–35.

Johansson, A., 2007: Prediction skill of NAO and PNA from daily to seasonal time scales. *J. Climate,* **20,** 1957–1975.

Jolliffe, I. T., and D. B. Stephenson, 2003: *Forecast Verification—A Practitioner's Guide in Atmospheric Science*. Wiley, 240 pp.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–472.

Kharin, V., Q. Teng, F. W. Zwiers, G. J. Boer, J. Derome, and J. S. Fontecilla, 2009: Skill assessment of seasonal hindcasts from the Canadian historical forecast project. *Atmos.–Ocean,* **47,** 204–223.

Kimoto, M., H. Mokougawa, and S. Yoden, 1991: Medium-range forecast skill variation and blocking transition: A case study. *Mon. Wea. Rev.,* **120,** 1616–1627.

Kleeman, R., 2002: Measuring dynamical prediction utility using relative entropy. *J. Atmos. Sci.,* **59,** 2057–2072.

——, 2008: Limits, variability, and general behavior of statistical predictability of the midlatitude atmosphere. *J. Atmos. Sci.,* **65,** 263–275.

Krishnamurti, T. N., and Coauthors, 1999: Improved weather and seasonal climate forecasts from multimodel superensembles. *Science,* **258,** 1548–1550.

——, and Coauthors, 2000: Coupled atmosphere–ocean modeling of El Niño of 1997–98. *J. Climate,* **13,** 2428–2459.

Kumar, A., M. Hoeling, M. Ji, A. Leetmaa, and P. Sardeshmukh, 1996: Assessing a GCM's suitability for making seasonal predictions. *J. Climate,* **9,** 115–129.

Lau, N.-C., 1981: A diagnostic study of recurrent meteorological anomalies appearing in a 15-yr simulation with a GFDL general circulation model. *Mon. Wea. Rev.,* **109,** 2287–2311.

Leathers, D. J., B. Yarnal, and M. A. Palecki, 1991: The Pacific–North American teleconnection pattern and United States climate. Part I: Regional temperature and precipitation associations. *J. Climate,* **4,** 517–528.

Lin, H., and J. Derome, 1996: Changes in PNA predictability associated with PNA pattern. *Tellus,* **48A,** 553–571.

Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus,* **21,** 289–307.

McFarlane, N. A., G. J. Boer, J.-P. Blanchet, and M. Lazare, 1992: The Canadian Climate Centre second generation general circulation model and its equilibrium climate. *J. Climate,* **5,** 1013–1044.

Munoz, E., C. Wang, and D. Enfield, 2010: The intra-America springtime sea surface temperature anomaly dipole as fingerprint of remote influences. *J. Climate,* **23,** 43–56.

Nakaegawa, T., and M. Kanamitsu, 2006: Cluster analysis of the seasonal forecast skill of the NCEP SFM over the Pacific–North America sector. *J. Climate,* **9,** 123–138.

Palmer, T. N., 1993: Extended-range atmospheric prediction and the Lorenz model. *Bull. Amer. Meteor. Soc.,* **74,** 49–65.

——, and J. Shukla, 2000: Editorial to DSP/PROVOST special issue. *Quart. J. Roy. Meteor. Soc.,* **126,** 1989–1990.

——, and Coauthors, 2004: Development of a European Multimodel Ensemble System for Seasonal-to-Interannual Prediction (DEMETER). *Bull. Amer. Meteor. Soc.,* **85,** 853–872.

Phelps, M. W., A. Kumar, and J. J. O'Brien, 2004: Potential predictability in the NCEP CPC dynamical seasonal forecast system. *J. Climate,* **17,** 3775–3785.

Renwick, J. A., and J. M. Wallace, 1995: Predictable anomaly patterns and forecast skill of Northern Hemisphere wintertime 500-mb-height fields. *Mon. Wea. Rev.,* **123,** 2114–2131.

Ritchie, H., 1991: Application of the semi-Lagrangian method to a multi-level spectral primitive-equations model. *Quart. J. Roy. Meteor. Soc.,* **117,** 91–106.

Rowell, P. D., 1998: Assessing potential seasonal predictability with an ensemble of multi-decadal GCM simulations. *J. Climate,* **11,** 109–120.

Sardeshmukh, P. D., and B. J. Hoskins, 1988: The generation of global rotational flow by steady idealized tropical divergence. *J. Atmos. Sci.,* **45,** 1228–1268.

——, G. P. Compo, and C. Penland, 2000: Changes of probability associated with El Niño. *J. Climate,* **13,** 4268–4286.

Shukla, J., and J. L. Kinter, III, 2006: Predictability of seasonal climate variation: A pedagogical review. *Predictability in Weather and Climate,* T. Palmer and R. Hagedorn, Eds., Cambridge University Press, 306–341.

——, and Coauthors, 2000: Dynamical seasonal prediction. *Bull. Amer. Meteor. Soc.,* **81,** 2593–2606.

Simmons, A. J., 1982: The forcing of stationary wave motion by tropical diabetic heating. *Quart. J. Roy. Meteor. Soc.,* **108,** 503–514.

Stensrud, D. J., H. E. Brooks, M. S. Tracton, and E. Rogers, 1999: Using ensembles for short-range forecasting. *Mon. Wea. Rev.,* **127,** 433–446.

Straus, D. M., and J. Shukla, 1997: Variations of midlatitude transient dynamics associated with ENSO. *J. Atmos. Sci.,* **54,** 777–790.

——, and ——, 2002: Does ENSO force PNA? *J. Climate,* **15,** 2340–2358.

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 ——, 2000: Annular modes in the extratropical circulation. Part I: Month-to-month variability. *J. Climate,* **13,** 1000–1016.

Tribbia, J. J., and D. P. Baumhefner, 1988: Estimates of the predictability of low frequency variability with a spectral general circulation model. *J. Atmos. Sci.,* **45,** 2306–2317.

Van den Dool, H. M., and S. Saha, 1990: Frequency dependence in forecast skill. *Mon. Wea. Rev.,* **118,** 128–137.

Vitart, F., 2004: Monthly forecasting system at ECMWF. *Mon. Wea. Rev.,* **132,** 2761–2779.

Wallace, J. M., and D. S. Gutzler, 1981: Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.,* **109,** 784–812.

——, and M. L. Blackmon, 1983: Observations of low-frequency atmospheric variability. *Large-Scale Dynamical Processes in the Atmosphere,* B. J. Hoskins and R. P. Pearce, Eds., Academic Press, 54–95.

Weisheimer, A., and Coauthors, 2009: ENSEMBLES: A new multimodel ensemble for seasonal-to-annual predictions—Skill and progress beyond DEMETER in forecasting tropical Pacific SSTs. *Geophys. Res. Lett.,* **36,** L21711, doi:10.1029/2009GL040896.

Woodhouse, C. A., 1997: Tree-ring reconstruction of circulation indices. *Climate Res.,* **8,** 117–127.

Yan, X., and Y. Tang, 2013: An analysis of multimodel ensemble for seasonal climate predictions. *Quart. J. Roy. Meteor. Soc.,* **139,** 1389–1401, doi:10.1002/qj.2019.

Yang, D., Y. Tang, Y. Zhang, and X. Yang, 2012: Information-based potential predictability of the Asian summer monsoon in a coupled model. *J. Geophys. Res.,* **117,** D03119, doi:10.1029/2011JD016775.

Yarnal, B., and H. F. Diaz, 1986: Relationships between extremes of Southern Oscillation and the winter climate of the Anglo–American Pacific Coast. *J. Climate,* **6,** 197–219.