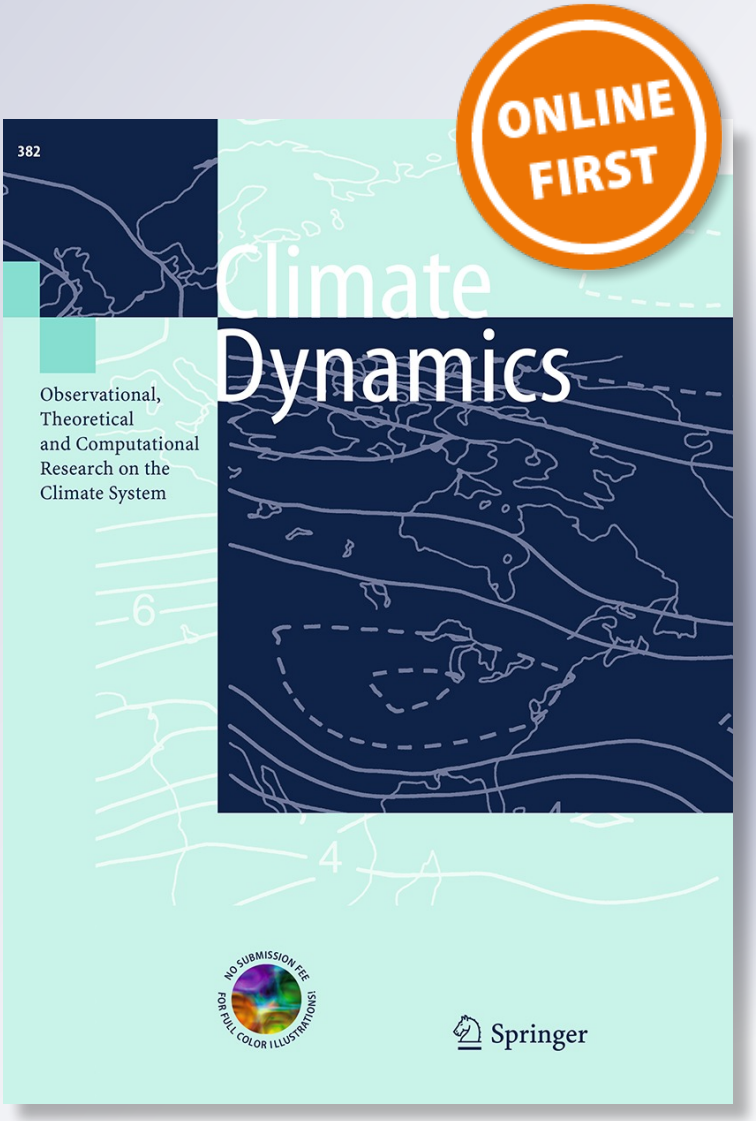
Potential predictability of Northern America surface temperature in AGCMs and CGCMs

Youmin Tang, Dake Chen & Xiaoqin Yan

Climate Dynamics
Observational, Theoretical and Computational Research on the Climate System

ISSN 0930-7575

Clim Dyn
DOI 10.1007/s00382-014-2335-x



Springer

Clim Dyn
DOI 10.1007/s00382-014-2335-x

Potential predictability of Northern America surface temperature in AGCMs and CGCMs

Youmin Tang · Dake Chen · Xiaoqin Yan

Received: 17 March 2014 / Accepted: 10 September 2014

Abstract In this study, the potential predictability of the Northern America (NA) surface air temperature (SAT) was explored using an information-based predictability framework and two multiple model ensemble products: a one-tier prediction by coupled models (T1), and a two-tier prediction by atmospheric models only (T2). Furthermore, the potential predictability was optimally decomposed into different modes for both T1 and T2, by extracting the most predictable structures. Emphasis was placed on the comparison of the predictability between T1 and T2. It was found that the potential predictability of the NA SAT is seasonal and spatially dependent in both T1 and T2. Higher predictability occurs in spring and winter and over the southeastern US and northwestern Canada. There is no significant difference of potential predictability between T1 and T2 for most areas of NA, although T1 has higher potential predictability than T2 in the southeastern US. Both T1 and T2 display similar most predictable components (PrCs) for the NA SAT, characterized by the inter-annual variability mode and the long-term trend mode. The first one is inherent to the tropical Pacific sea surface temperature forcing, such as the El Nino-Southern Oscillation, whereas the second one is closely associated with global warming. In general, the PrC modes can better characterize the predictability in T1 than in T2, in particular for the inter-annual variability mode in the fall. The prediction skill against observations is better measured by the PrC analysis than by principal component analysis for all seasons, indicating the stronger capability of PrCA in extracting prediction targets.

Keywords Seasonal climate prediction · Potential predictability · Ensemble predictions

1 Introduction

Surface air temperature (SAT) is the most widely used variable in the study of seasonal climate prediction, not only for its significance in practical applications but also for its relatively high potential predictability compared with other atmospheric variables. Many earlier seasonal climate predictions targeted SAT using statistical methods such as canonical correlation analysis (CCA) and multichannel singular spectrum analysis (MSSA) (e.g., Barnett and Preisendorfer 1987; Shabbar and Barnston 1996; Barnston and Smith 1996; Vautard et al. 1999). In the last decade, a great deal of effort has been made to use dynamical models to predict SAT at the seasonal time scale. In general, there are two types of dynamical models for seasonal climate predictions: atmospheric general circulation models (AGCMs) and atmosphere–ocean coupled GCMs (CGCMs). The former ignores the feedback between ocean and atmosphere, and the predicted sea

Y. Tang (✉) · X. Yan
Environmental Science and Engineering, University of Northern British Columbia, 3333 University Way, Prince George, BC V2N 4Z9, Canada
e-mail: ytang@unbc.ca

Y. Tang · D. Chen
State Key Laboratory of Satellite Ocean Environment Dynamics, Hangzhou, China

Present Address:
X. Yan
Department of Earth and Atmospheric Sciences, George Mason University, Fairfax, VA, USA

Published online: 25 September 2014

surface temperature (SST)[1] drives the atmospheric model. Thus, the seasonal prediction by AGCMs has two separate steps, ocean prediction and atmosphere prediction, and is often referred to as a two-tier (T2) forecast. In contrast, CGCMs integrate forward to generate ocean and atmosphere predictions simultaneously, which is thus called a one-tier (T1) forecast.

Theoretically, a T1 forecast is expected to be more skillful than a T2 forecast, as supported by some experiments. For example, Kug et al. (2008) found that the prediction skills of the seasonal mean precipitation were much improved in T1 experiments, although the forecasted SST in T1 were not accurate as those in T2. However, different examples exist. For instance, Jha and Kumar (2009) found that the atmospheric response to the El Niño-Southern Oscillation (ENSO) has small departures in both T2 and T1 simulations in winter. Peng et al. (2011) also reported that the seasonal predictability was consistent between T1 and T2, although some differences cannot be completely neglected, especially in regions with strong air-sea interactions, such as the warm-pool and monsoon region. Misra et al. (2013) recently found that T2 can achieve a more skillful seasonal prediction than T1 if the SST over the equatorial Pacific Ocean can be better predicted in T2 than in T1.

Seasonal climate predictability using both T1 and T2 forecasts has been studied intensively (e.g., Palmer et al. 2004; Luo et al. 2005; Kug et al. 2008; Wang et al. 2009; Kirtman 2003; Kirtman and Min 2009; Doblas-Reyes et al. 2009; Misra et al. 2013). Currently, many operational centers around the world use T1 for seasonal prediction. However, with limited computational resources, two options in developing seasonal climate models have been debated. One is to couple the climate components to develop T1 by sacrificing the resolution; the other is to raise the AGCM resolution to levels where parameterization of the sub-grid scale processes can be avoided (e.g., Palmer et al. 2009; Misra et al. 2013). One key issue in this debating is to compare the potential predictability between T1 and T2, namely, quantifying the upper limit of the predictability of T1 and T2. Compared with the analysis of actual prediction skills, there were few comparisons of potential predictability between T1 and T2 in literature, most of which used signal-to-noise measures that underestimate the potential predictability (e.g., Tang et al. 2014).

In this work, we apply information-based metrics to explore NA SAT potential predictability in T1 and T2 as an effort to complement recent works in exploring the actual prediction skills in T1 and T2. The information-based metrics of potential predictability have advantage over the conventional signal-to-noise ratio (SNR)-based metrics because the latter often underestimate the true potential predictability (e.g., Yang et al. 2012; Tang et al. 2014). Emphasis is placed on the potential predictability using information-based metrics, including relative entropy, mutual information and the most predictable component analysis (PrCA), and using multiple model ensembles, which has not been well addressed in the literature. The seasonal mean of the SAT over North America (referred to as NA-SAT hereafter) is used as the prediction target. The NA-SAT has been widely used in studying the seasonal climate predictability inherent to the tropical SST forcing. As argued in the literature, the extra-tropical atmospheric seasonal variability over the Pacific-North America (PNA) region has a pronounced response to the interannual variability of the tropical SST, especially in winter (e.g., Barnett and Preisendorfer 1987; Shabbar and Barnston 1996; Barnston and Smith 1996; Kumar and Hoerling 1998; Shukla et al. 1998, 2000; Schlosser and Kirtman 2005; Tang et al. 2014). It has been found that the seasonal climate is more predictable in North America than in other extra-tropical regions, and significant prediction skill can be achieved, especially when the tropical SST anomalies are large (e.g., Shukla et al. 2000; Derome et al. 2001; Kumar et al. 1998).

In this study, two ensemble products are chosen to investigate the T1 and T2 predictability. The first is the ensemble product HFP2 (phase 2 of the Canadian Historical Seasonal Forecasting Project, Environment Canada), composed of four AGCMs. The second is the seasonal prediction of ENSEMBLES (European Commission Framework Project 7), composed of five CGCMs. Four calendar seasons were targeted: March–April–May (MAM), June–July–August (JJA), September–October–November (SON) and December–January–February (DJF). The central questions that this study attempts to answer include the following: (1) What are the similarities and disparities in potential predictability between T1 and T2 and are these features geophysically and seasonally dependent? (2) Which and how are physical processes responsible for these similarities and disparities? For simplicity, the seasonal mean is used as the prediction target, namely, the average of the 3-month prediction starting from the beginning of March, June, Sep. and Dec., is analyzed, respectively.

This paper is structured as follows: Sect. 2 briefly describes the data and methods used in this study. Section 3 displays the potential predictability of T1 and T2 measured by information-based metrics. Section 4 extracts the PrCA for the two ensemble products and analyzes their leading modes. In Sect. 5, the actual prediction skill of the ensemble mean against the observation is evaluated for T1 and T2, followed by the conclusions in Sect. 6.

[1] The SST prediction is made using either a persistent model or another forecast model (e.g., a statistical model).

2 Data and methods

2.1 Data

The HFP2 is a collaborative project among some Canadian universities and government laboratories to test the potential predictability of mean-seasonal conditions. In HFP2, ten-member ensembles of seasonal hindcasts are produced with four general circulation models: the Canadian Centre for Climate Modeling and Analysis (CCCma) second and third generation atmospheric general circulation models (GCM2 and GCM3), a reduced-resolution version of the medium-range weather forecast global spectral model (SEF) (Ritchie 1991) and the global environmental multiscale model (GEM) developed at Recherche en prévision numérique (RPN).

In the HFP2 project, a set of ensemble prediction is conducted at the first day of the calendar month during the period from 1969 to 2002, with the lead time of 4 months (one season), producing the hindcasts for 12 seasons for each year in global scope. The initial conditions for the ensemble predictions are directly from the NCEP/NCAR (National Centers for Environmental Prediction/National Center for Atmospheric Research) reanalysis data. The ensemble was constructed by the lagged averaged method (LAF) using 12-h time lag, namely, the initial conditions are lagged at 12-h intervals prior to the forecast period without perturbations (the perturbations are instead fulfilled through the different atmospheric states). The first member is initialized at 12 h before the forecast period, the second member is initialized at 24 h prior to the forecast period, etc., and the 10th member is initialized at 5 days prior to the forecast period. The boundary conditions, sea surface temperature (SST) and sea ice (ICE) data were taken from the Seasonal Prediction Model Intercomparison Project-2 (SMIP-2) boundary data. The global SST anomaly of the month prior to the forecast period is persisted during the 4-month forecast period. The ice extents were initialized with the analysis and relaxed to the climatology within the first 15 days of integration. The NCEP weekly observations were used for the snow cover initialization. A detailed instruction on the HFP2 can be found in Lin et al. (2008) and Kharin et al. (2009).

The ENSEMBLES (http://www.ecmwf.int/research/EU_projects/ENSEMBLES/exp_setup/index.html) assembles ensemble forecasts from 5 global coupled atmosphere- ocean models, developed in UK Met Office (UKMO), Météo France (MF), the European Centre for Medium-Range Weather Forecasts (ECMWF), the Leibniz Institute of Marine Sciences at Kiel University (IFM-GEOMAR), and the Euro-Mediterranean Centre for Climate Change (CMCC-INGV) in Bologna. Each ensemble has nine runs of 7-month duration (except 14 in November) generated using different sets of ocean reanalysis generated from wind stress and SST perturbations, initiated separately from February, May, August and November. The hindcast period of the ENSEMBLES covers the 46 years of 1960–2005. The ENSEMBLES has been widely used in the literature (e.g., Weisheimer et al. 2009; Tang et al. 2014).

For verification purpose, the TS (Time-Series) 2.1 dataset of the monthly observed land surface air temperature from the Climatic Research Unit (CRU; http://www.cru.uea.ac.uk/cru/data/hrg.htm) was used. An interpolation was performed to convert the observation data grid resolution of $0.5^\circ \times 0.5^\circ$ to the model grid resolution of $2.5^\circ \times 2.5^\circ$.

2.2 Methods

2.2.1 Information-based potential predictability

The entropy is a quantitative measure of uncertainty for a random variable. The information-based potential predictability framework has been well introduced in recent literature (e.g., Kleeman 2002; DelSole and Tippett 2007; Brunsell and Wilson 2013; Tang et al. 2014). For the convenience of readers, the main ideas and formulas of this framework are briefly introduced as follows.

Suppose that the future state of a climate variable is predicted/modeled as a random variable denoted by v with a climatological distribution $p(v)$. One ensemble prediction produces a forecast distribution that is the conditional distribution $p(v|\Theta)$ given the initial condition Θ. The extent to which the forecast entropy defined by the distribution $p(v|\Theta)$ and the climatological entropy defined by the distribution $p(v)$ differ is an indication of the decrease of uncertainty, i.e., the potential predictability. One widely used "difference (distance)" of two entropies is the relative entropy (RE), also referred to as the Kullback–Leibler divergence, defined as:

$$RE = \int p(v|\Theta) \ln\left(\frac{p(v|\Theta)}{p(v)}\right) dv \tag{1}$$

For a special case when the climatology and prediction distributions are Gaussian, the RE can be expressed by the ensemble (or forecast) variance σ_p^2, the climatological variance σ_q^2, the ensemble (or forecast) mean μ_p and the climatological mean μ_q, namely, (Kleeman 2002)

$$RE = \frac{1}{2}\left[\underbrace{ln\left(\frac{\sigma_q^2}{\sigma_p^2}\right) + \frac{\sigma_p^2}{\sigma_q^2} - 1}_{Dispersion} + \underbrace{\frac{{\mu_p}^2}{\sigma_q^2}}_{Signal}\right] = PI + \frac{1}{2}\left[\frac{\sigma_p^2}{\sigma_q^2} - 1 + \frac{{\mu_p}^2}{\sigma_q^2}\right] \tag{2}$$

The RE can be divided into two items: one dominated by prediction and climatology variances, referred to as Dispersion; the other controlled by the ensemble mean and climatological variance, called Signal. The further interpretation of RE can be found in literature (e.g., Kleeman 2002).

In this study, the RE is calculated using (2), since Gaussian assumption is usually held for climate prediction of monthly or seasonal mean (e.g., Tang et al. 2005). The σ_p^2 and μ_p are calculated using the ensemble samples for each initial condition, and the μ_q and σ_q^2 are from all ensembles of initial conditions over the entire period. We also test the observed data in calculating the μ_q and σ_q^2, and obtained similar RE.

The RE measures the potential predictability of an individual prediction as a function of the initial condition and lead time. This is different from the conventional potential predictability measures based on the signal-to-noise ratio (SNR), which evaluates the overall predictability of the prediction system. The potential predictability of an individual prediction offers a practical means of estimating the confidence level of a seasonal forecast using the dynamical model.

Corresponding to the SNR-based overall predictability, the average of the REs over all initial conditions reflects the average predictability and was proved to be equal to the mutual information (MI), another quantity from information theory (DelSole 2004). In this study, we calculate MI by averaging the REs of all initial conditions over the entire period.Furthermore, the below quantity is defined by MI, namely,

AC_{MI} is an MI-based potential predictability measure,

$$AC_{MI} = \sqrt{1 - \exp(-2MI)} \tag{3}$$

equivalent to correlation coefficient (DelSole and Tippett 2007). It has been proved that AC_{MI} is better than the SNR-based measure such as perfect correlation in characterizing the 'true' potential predictability (e.g., DelSole 2004; Yang et al. 2012; Tang et al. 2014). In this study, we use the AC_{MI} to explore the potential predictability in the T1 and T2 ensemble predictions.

The signal (S) and noise (N) variance are calculated using the definition of Rowell (1998), i.e.,

$$\begin{aligned} Var(S) &= \frac{1}{M}\sum_{i=1}^{M}\left(\overline{X}_i - \overline{\overline{X}}\right)^2, \\ Var(S) &= \frac{1}{MK}\sum_{i=1}^{M}\sum_{j=1}^{K}\left(X_{i,j} - \overline{X}_i\right)^2. \end{aligned} \tag{4}$$

where $X_{i,j}$ is the j-th member of the ensemble prediction starting from the i-th initial condition. The K is the ensemble size and M is the total number of initial conditions (predictions), and $\overline{X}_i = \frac{1}{K}\sum_{j=1}^{K} X_{i,j}$; $\overline{\overline{X}} = \frac{1}{M}\sum_{i=1}^{M}\overline{X}_i$.

2.2.2 *Principal prediction component analysis (PrCA) and Maximum SNR*

The PrCA analysis is an approach to minimizing σ_p^2/σ_q^2 to derive the most predictable component. The minimization of σ_p^2/σ_q^2 is equivalent to the maximization of $1-\sigma_p^2/\sigma_q^2$, i.e., the SNR. Theoretically, the signal and noise are independent when the ensemble size is infinite. However, the ensemble size is always finite in reality; thus, the estimation of the signal is often contaminated by the noise. An optimal estimate for the largest potential predictability should maximize the SNR, from which the resultant signal component is the most predictable. The maximum SNR and the PrCA are often alternatively used due to their complete equivalence under the assumption of a Gaussian distribution. A detailed algorithm of the PrCA can be found in a broad scope of the literature (e.g., Allen and Smith 1997; Venzke and Allen 1999; Schneider and Griffies (1999); Sutton et al. 2000; DelSole and Tippett 2007).

2.3 Analysis strategy

In this study, the analyses were carried out for the period from January 1969 to December 2001 (33 years) for the HFP2 and ENSEMBLES. The analysis area covers all land grid cells in NA (210°E–300°E, 20°N–80°N). The model climatology is obtained by averaging all its ensemble members. The multi-model ensemble (MME) of the ENSEMBLES and the HFP2 are used to represent two types of prediction, T1 and T2, respectively. It has been argued that the MME is usually better than a single model ensemble (SME) because the uncertainties associated with the different model frameworks can offset each other and be relieved by the large number of ensemble members (Krishnamurti et al. 1999, 2000; Palmer et al. 2004). The MME mean was constructed here in the simplest way, namely, averaging each model ensemble mean (i.e., four means for HFP2 and five means for ENSEMBLES). The ensemble members of the HFP2 and ENSEMBLES are pooled together with equal weights to form super ensembles of T1 and T2 forecasts for the PrCA analyses (i.e., 40 members for the HFP2 and 45 members for the ENSEMBLES MME). To explore the seasonal dependence of predictability, we performed all analyses for each season.

3 The potential predictability in T1 (Tier1) and T2 (Tier2)

3.1 General features of MI-based skill

The potential predictabilityAC_{MI}, defined in Eq. (3), is shown in Fig. 1. The common features of the potential

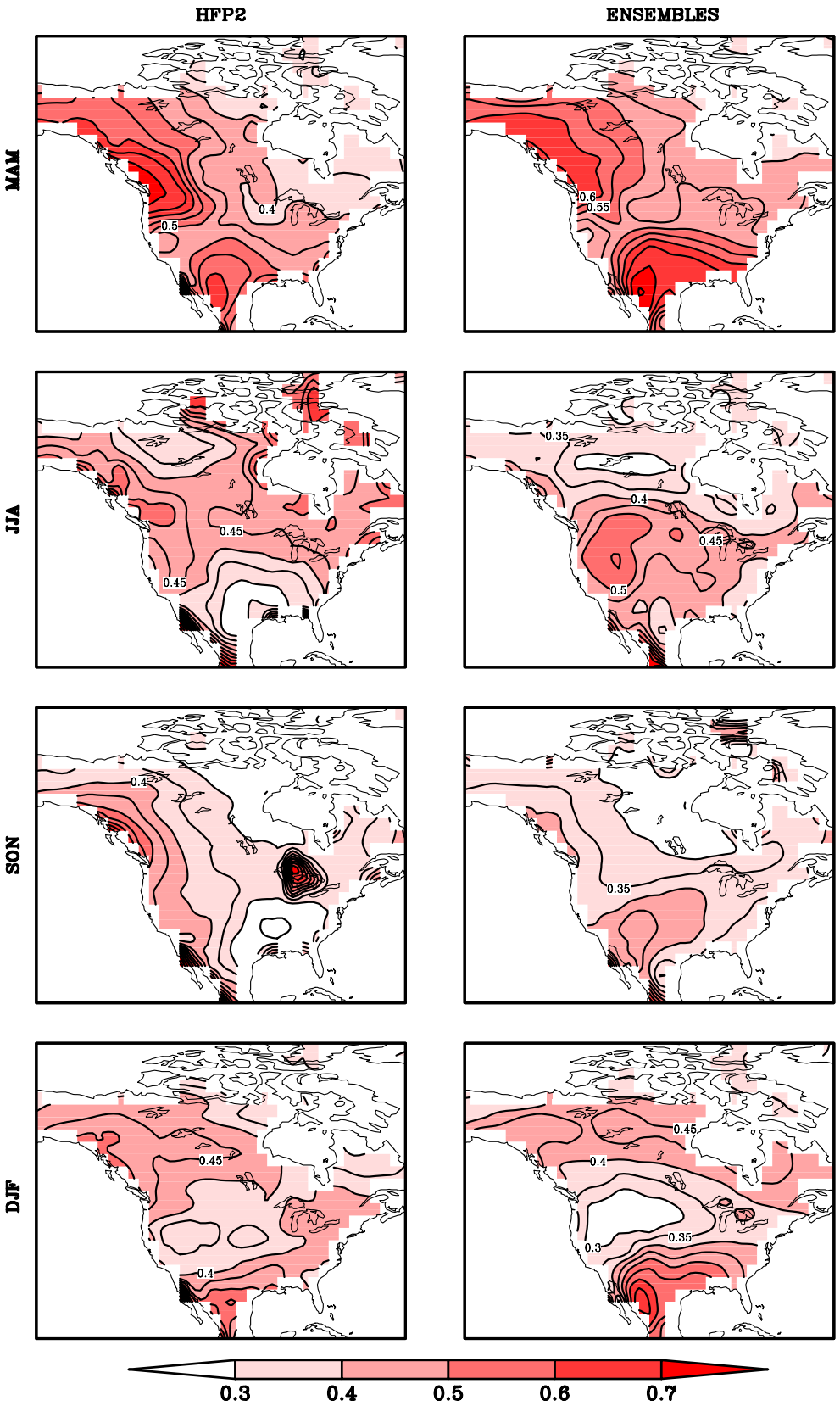


Fig. 1 MI-based potential predictability for T2 (*left*) and T1 (*right*) for four seasons (MAM, JJA, SON and DJF)

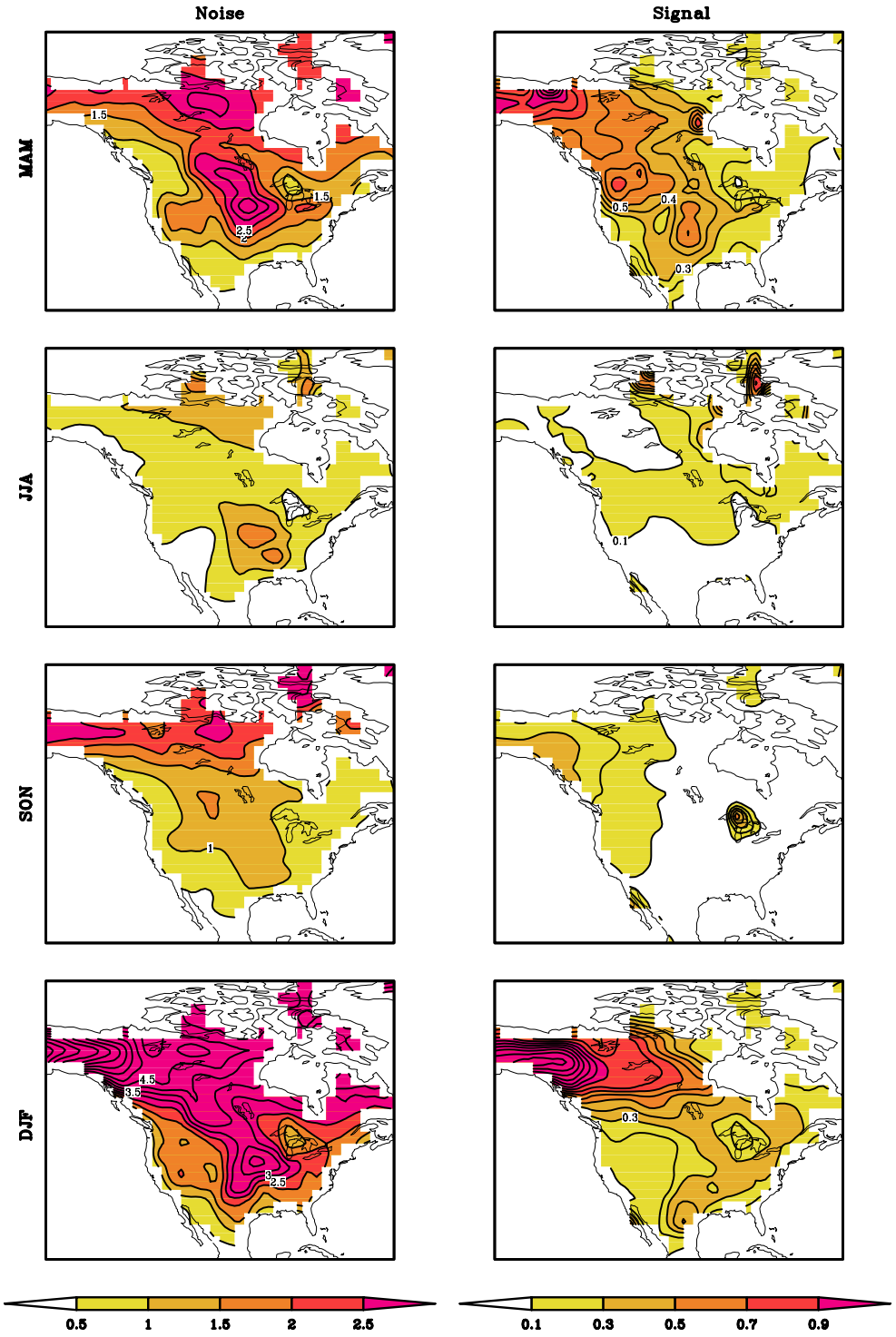


Fig. 2 The signal and noise of HFP2

predictability in both T1 and T2 include the following: (1) AC_{MI} is location and season dependent; (2) the relatively strong AC_{MI} occurs in winter and spring, followed by summer and fall, in both types of models; (3) spatially, the AC_{MI} shows larger amplitude over northwestern Canada and southeastern US in winter and spring in both types of models.

As argued in Tang et al. (2014), the variation in AC_{MI} is determined by the signal and noise variance. Shown in Fig. 2 is the signal and noise variance for T2. The spatial

pattern of the signal and noise variance for T1 is very similar to Fig. 2 (Tang et al. 2014). As seen in Fig. 2, the noise variance is much larger than the signal counterpart, suggesting that the internal variability is much larger than the external forcing over NA, most likely explaining why the seasonal climate prediction skill is usually lower in the middle latitudes than in the tropical regions, as reported in the literature (e.g., Yan and Tang 2012). In contrast, the seasonal and geographical variations are more significant in the signal strength than in the noise strength, as characterized by a much larger signal variance in winter and spring, followed by summer and fall, and in the areas over northwestern Canada and southeastern US. Comparison between Figs. 1 and 2 reveals that the signal dominates the spatial distribution and seasonal dependence of the potential predictability.

Shown in Fig. 3 are the regression coefficients of the NA SAT against the Nino3.4 SSTA, clearly displaying a strong response of the SAT over northwestern Canada and southeastern US to ENSO in spring and winter and even in summer. Thus, the large signal amplitude over northwestern Canada and southeastern US and in spring and winter is most likely due to two reasons. First, the tropical Pacific SSTA has a strong interannual variability in winter. Second, the response of NA to ENSO has a time lag of several months, i.e., a strong time lagged teleconnection between the NA SAT and ENSO. In the discussion of PrCA in Sect. 4, we will also see that Fig. 3 is similar to the most predictable mode of NA SAT.

3.2 Potential predictability of T1 and T2 models

The above analyses reveal many similar features of the potential predictability between T1 and T2 for most regions and seasons. The exception mainly occurs in some regions over the southern US where T1 has relatively high predictability in winter and spring. Shown in Fig. 4 is the difference in the potential predictability between T1 and T2. To examine whether the difference is statistically significant or due to sampling errors, we performed a boot-strap experiment test. First, we calculated the AC_{MI} of T1 by randomly drawing 85 % of the original data. This process is repeated 1,000 times. The difference between the mean over the 1,000 AC_{MI} and the original AC_{MI} is used as a criterion of sampling error. When the difference in AC_{MI} between T1 and T2 is beyond the criterion, the statistical significance is referred and shaded in Fig. 4. It should be noted that the criterion does not change much when the size of the randomly drawn sample is in the range of 85–95 %.

Figure 5 shows the difference of signal and noise between T1 and T2, indicating that when T1 (coupled models) can capture a strong signal, it also has a correspondingly strong noise. The relatively high potential predictability in the coupled models occurs only when the signal due to the coupling is strong enough to offset the strong noise, which occurs at specific seasons and locations, as suggested in Fig. 4. As argued in the literature (e.g., Kumar et al. 2003; Misra et al. 2013; Tang et al. 2014), the predictable signal at seasonal time scales primarily stems from the slowly varying SST forcing, suggesting that the atmosphere response to the SST is stronger in the coupled models than in the uncoupled models. This can be evidenced by Fig. 3, which indicates that the response of the NA SAT to ENSO is much stronger in T1 than in T2. It should be noted that a strong response of the SAT to ENSO does not necessarily require a more realistic SST forecast because atmospheric variables themselves have prediction errors. That may explain why the coupled models can have better predictability than the uncoupled models in some cases, even though the coupled models have worse prediction skill in SST (e.g., Kug et al. 2008). However, the potential predictability derived from the biased SST prediction may have a large difference from the actual model skill against the observations. To further explore this issue, we calculated the actual prediction skill of the SST in T1 and T2, respectively. Shown in Fig. 6 are the correlation skill of the SSTA prediction at one season lead in T1 (left panel) and its difference with the T2 counterpart (T1–T2, right panel). As shown, the SSTA prediction is very skillful at one season lead in the tropical oceans, in particular in the tropical Pacific, in T1 due to its high persistence. The difference of prediction skill between T1 and T2 is not significant in the tropical Pacific except in a narrow belt of the equatorial eastern Pacific. This may explain why the potential predictability of T1 is slightly higher than that of T2, primarily over the southern US, because the SST variability of the equatorial eastern Pacific has a significant impact on the climate of the southern US (e.g., Latif and Barnett 1994). For most regions of the extra-tropic Pacific, the skill of T2 (persistent skill) is better than that of T1 (coupled model skill).

4 The most predictable components in T1 and T2

Figures 7 and 8 show the spatial pattern of the first two PrCA modes for T1 and T2. As shown in Table 1, the first two modes in both T1 and T2 explain over 50 % of the signal variance for all seasons, and the other modes have negligible variance contributions. A comparison between T1 and T2 in Table 1 reveals that the first mode in T2 explains more variance than that in T1. This is apparent particularly in fall and winter, when the second mode has more explained variance than the first mode in T1. Theoretically, unlike Principle Component Analysis (PCA), the PrCA leading mode does not necessarily account for more

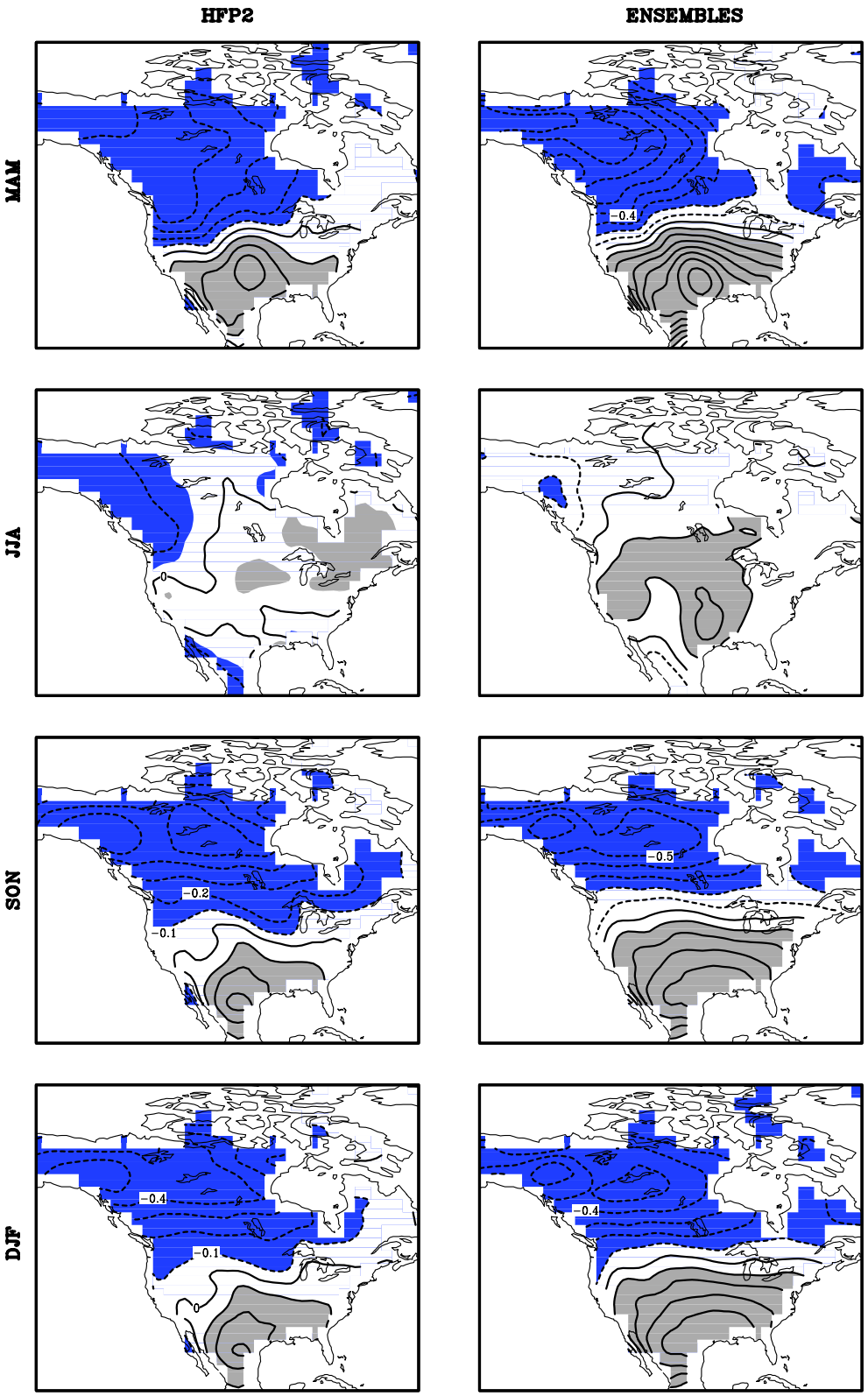


Fig. 3 Regression coefficient of the NA SAT to the first PC of tropical SST. The *shaded area* is statistically significant at the 99 % confidence level with a F-distribution test

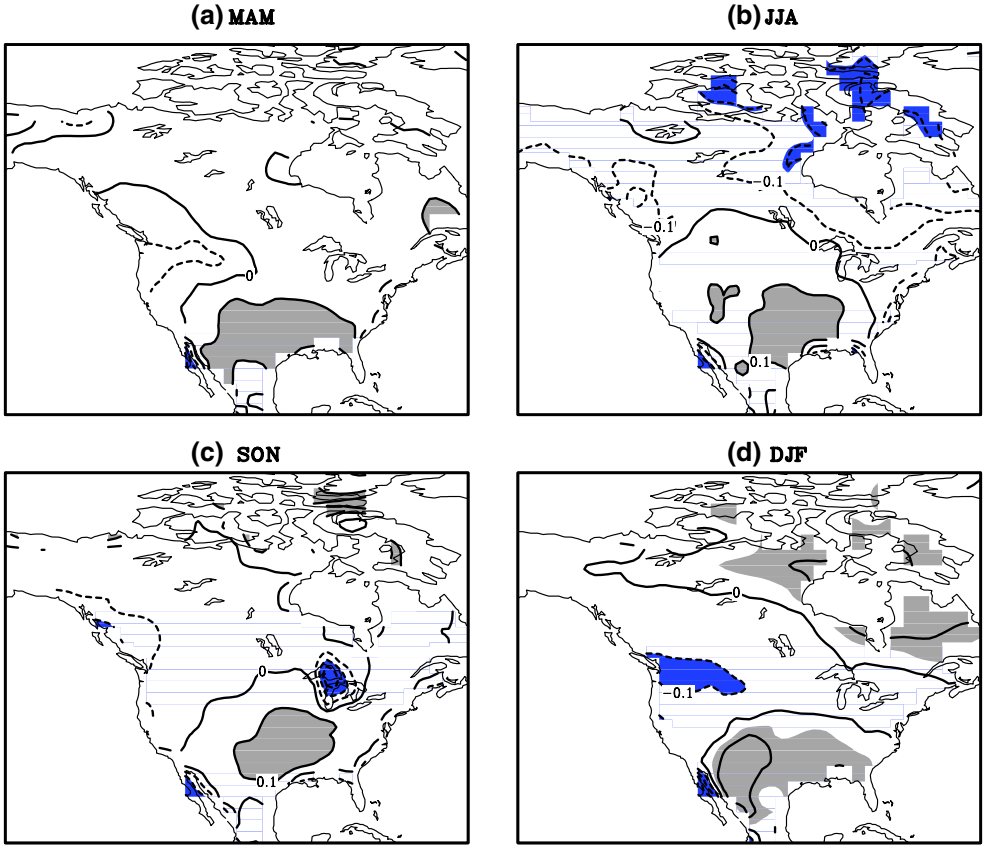


Fig. 4 The difference of AC_MI between ENSEMBLE and HFP2 (T1–T2). The *shaded area* is where the difference is statistically significant based on the boot-strap experiment (see context)

variance because the PrCA seeks the optimal modes based on the predictable component (predictability) rather than on the explained variance (variability) (e.g., Tang et al. 2014). Thus, Table 1 suggests that the most predictable component should be more directly associated with the SAT variability itself in T2, whereas the noise has more impact on the most predictable component in T1.

We also performed the PCA analysis with the ensemble mean of T1 and T2, respectively. The ensemble mean of the seasonal climate prediction primarily reflects the signal, although it may be still contaminated by some noise. Theoretically, if the number of ensemble members is large enough (e.g., infinite), the PCA of the ensemble mean is equivalent to the PrCA, both distinguishing the signal and noise. Shown in Figs. 9 and 10 are the spatial patterns of the first two PCA modes. A comparison between PrCA and PCA modes reveals that the PrCA mode 1 is very similar to the PCA counterpart, as shown in Figs. 7 and 9. For the second mode, one can find the similarity between Figs. 8 and 10 for JJA and SON, which are characterized by PrCA inter-annual variability mode in Fig. 8, if ignoring the sign that is kind of arbitrary. These indicate that the signal components extracted from PCA also dominate the most predictable modes, especially in PrCA inter-annual variability modes, in both T1 and T2.

Comparison between PrCA and PCA may reveal separate contributions of signal and noise to the most predictable component. Table 2 shows the correlation of the time series between PCA and PrCA for T1 and T2. As shown in this table, the correlation is much larger in T2 than in T1, with large differences occurring in fall and winter. This indicates that the impact of noise on the predictability is more effective in T1, particularly in fall and winter, which is also consistent with the relatively large noise strength in T1, especially in the two seasons, as shown in Figs. 2 and 5.

Figure 11 shows the time series of the first two PrCA modes. As seen in this figure, the first two most predictable patterns characterize the predictability of the inter-annual variability and long-term trend in T1 and T2. When the first mode represents the predictability of the inter-annual variability, the second mode primarily describes the long-term trend modulated by the interannual variability, and vice versa. For example, the first PrCA mode characterizes the predictability of the inter-annual variability in spring and winter and the predictability of the long-term trend in

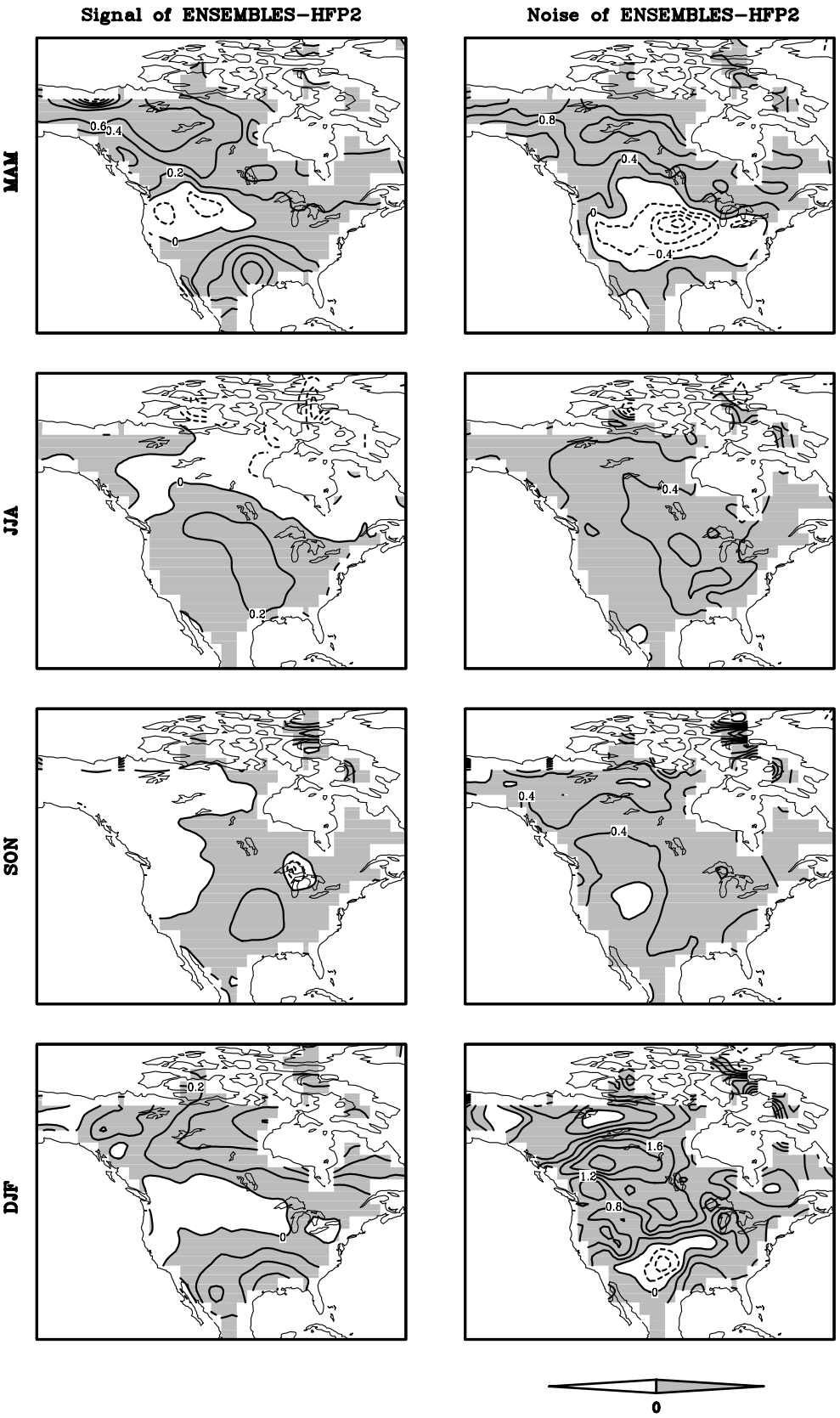


Fig. 5 Difference of signal and noise between T1 and T2 (T1–T2). The positive value of the difference is *shaded*

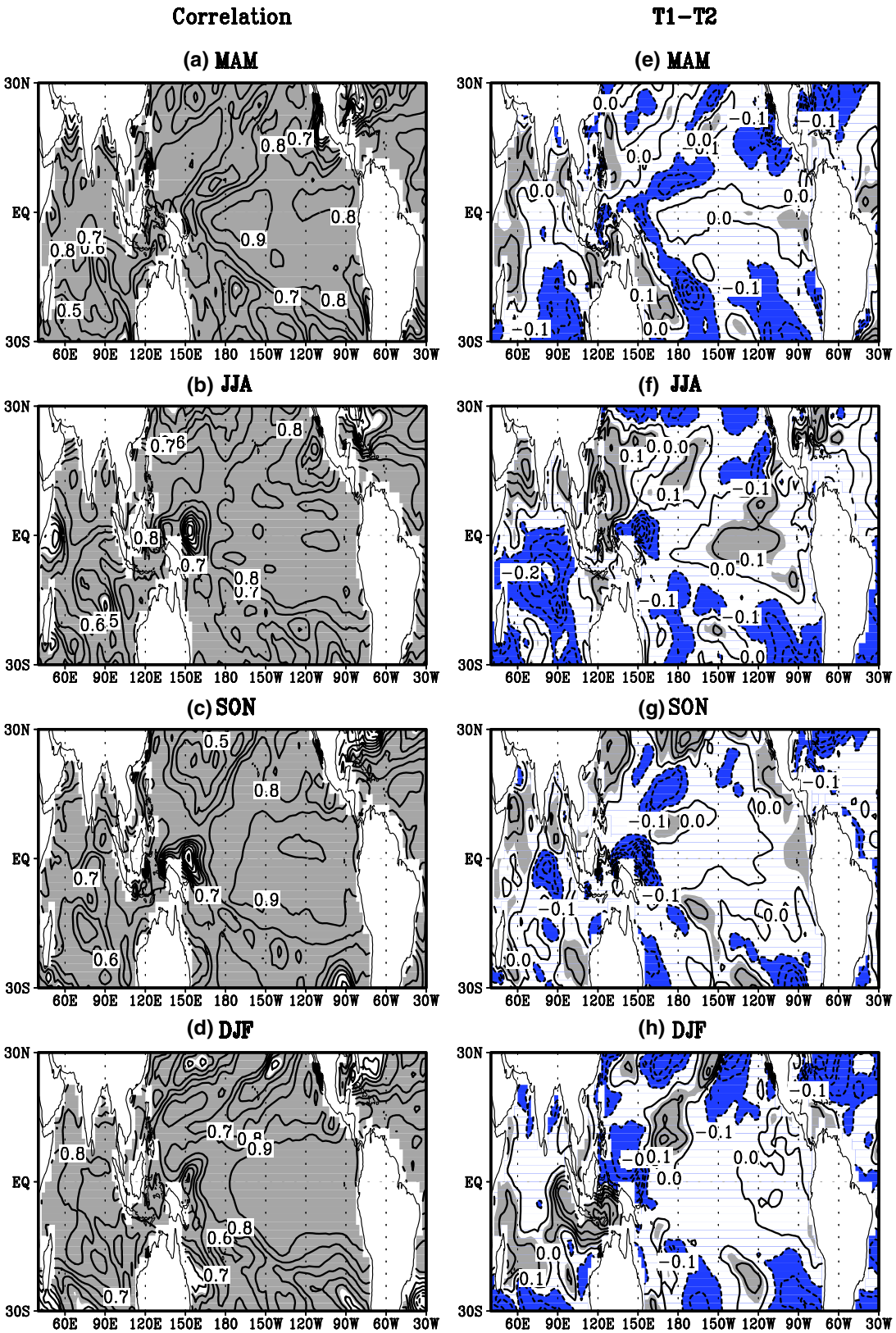


Fig. 6 The correlation skill of ENSEMBLE predicted SSTA against the observation (*left panel*) and the difference of correlation skill between ENSEMBLE and HFP2 (persistent skill) (*right panel*). The *gray shading* in (**a**–**d**) refers to statistically significant correlation, using student's-*t* test. The *shading* in (**e**–**h**) indicates the significant difference of correlation between T1 and T2, using boot-strap experiment as in Fig. 4. Here the significance confidence level is always set to 95 %

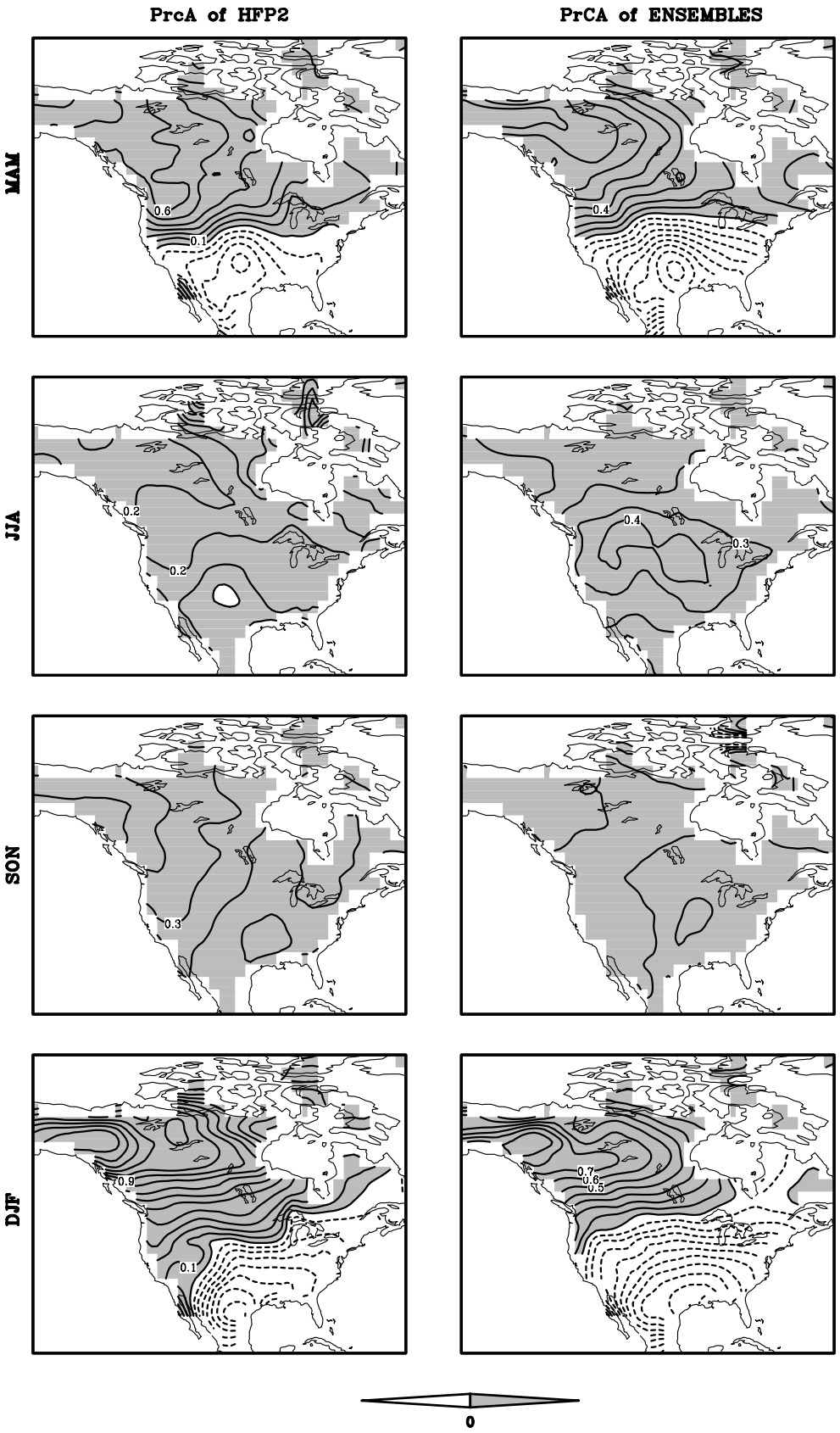


Fig. 7 The first mode of PrCA of the NA SAT prediction at one-season lead

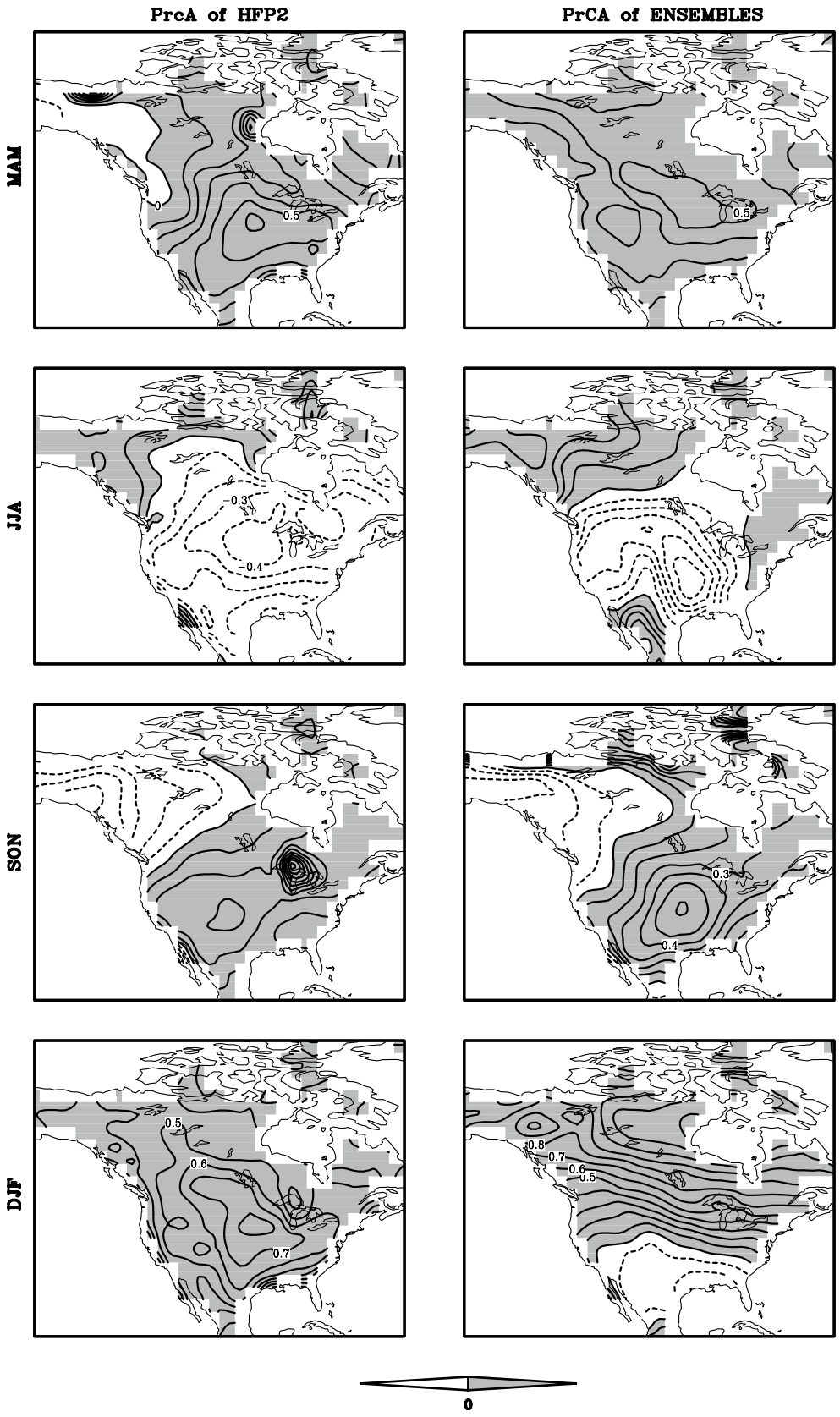


Fig. 8 Same as Fig. 7 but for mode 2

Table 1 Signal variance accounted for by each PrCA mode (%)

	MAM		JJA		SON		DJF	
	PCA	PrCA	PCA	PrCA	PCA	PrCA	PCA	PrCA
Mode1								
T1	67	55	50	43	37	26	63	29
T2	62	57	58	54	54	46	60	51
Mode2								
T1	19	24	19	18	27	30	21	45
T2	18	15	19	15	16	11	19	17

summer and fall, whereas the second mode is complementary with the first mode in characterizing the predictability of the two time scales. Because the prediction target here is the seasonal mean, the intra-seasonal variability has been filtered prior to PrCA analysis. It may be expected that the PrCA can also include the predictability of the intra-seasonal variability if daily or weekly data are used.

Comparing the spatial patterns of Figs. 7 and 8 with Fig. 11 reveals that the PrCA mode of inter-annual variability shows a dipole spatial structure between northwestern Canada and southeastern US. Such a dipole structure is very significant in spring and winter, as shown in PrCA mode 1, but less visible in summer and fall, as shown in PrCA mode 2, which is due to the seasonal variation of the strength of the SST forcing, as argued in the above section. However, the predictability of the long-term trend that is probably inherent to the signal of global warming in initial conditions is nearly uniformly distributed in the entire domain over NA, although its strength has some geographical variations in different seasons.

5 Actual prediction skills in T1 and T2

In the proceeding sections, we analyzed the potential predictability for T1 and T2 using information theory. The potential predictability assumes the prediction system to be perfect, and the prediction errors are primarily determined by the initial uncertainties. This potential predictability is also often called the first type of predictability, which characterizes the growth of the prediction error due to the initial errors excited by the nonlinear dynamics and stochastic forcing.

In reality, the prediction system is always imperfect. In this section, we will explore the actual prediction skill for T1 and T2. Shown in Fig. 12 is the correlation skill of the prediction against observations for T1 and T2, as well as their difference. As shown, the actual prediction skill has a spatial–temporal distribution similar to that of the potential predictability discussed above. For example, the actual skill is also best in spring and winter and over southeastern US and northwestern Canada.

The difference of the actual prediction skill between T1 and T2 is seasonally dependent, as shown in the right panel of Fig. 12. This is unlike their difference of potential predictability shown in Fig. 4, which is almost seasonally invariant, characterized by a comparable skill in most regions except the southeastern US, where T1 has a slightly higher skill than T2. This is especially apparent for summer and fall, when T1 has a better actual skill than T2, although both have similar potential predictability. This result is most likely due to a large systematic error existing in the HFP2 in summer and fall that has a large impact on the seasonal climate forecast (Jia et al. 2010). In addition, the ENSO forcing is quite different between T1 and T2 in summer and fall, as shown in Fig. 13, which may be another possible reason for the significant difference of actual prediction skill between T1 and T2 in the two seasons.

The overall prediction skill of the NA SAT is first evaluated using the PrCA components. Table 3 shows the prediction skill of the time series of the first two PrCA modes against the observed counterpart. The observed PrCA modes are obtained by projecting the observed SAT onto the optimal filters derived from the PrCA analysis. As shown in this table, the PrCA modes have in general higher prediction skills in T1 than in T2 for most seasons, indicating that the most predictable modes in T1 can better characterize reality. This is apparent in fall and winter, when the noise is much larger in T1, as discussed above. One possible reason is that the realistic noise that impacts real skill is not resolved in T2 as well as in T1, especially in fall and winter. In other words, the noise in T2 is most likely underestimated, thereby overestimating the most predictable components in T2, especially in fall and winter.

For the purpose of comparison, we also calculated the prediction skill of the time series of the first two PCA modes against the observed counterpart, as shown in Table 3. Apparently, the skill of the PrCAs is better than that of PCAs for all seasons, indicating the stronger capability of the PrCA in extracting the predictable component. This may offer a possible way to improve our operational prediction using PrCA.

A further examination of Table 3 reveals that the long-term trend predictable mode has higher prediction skill

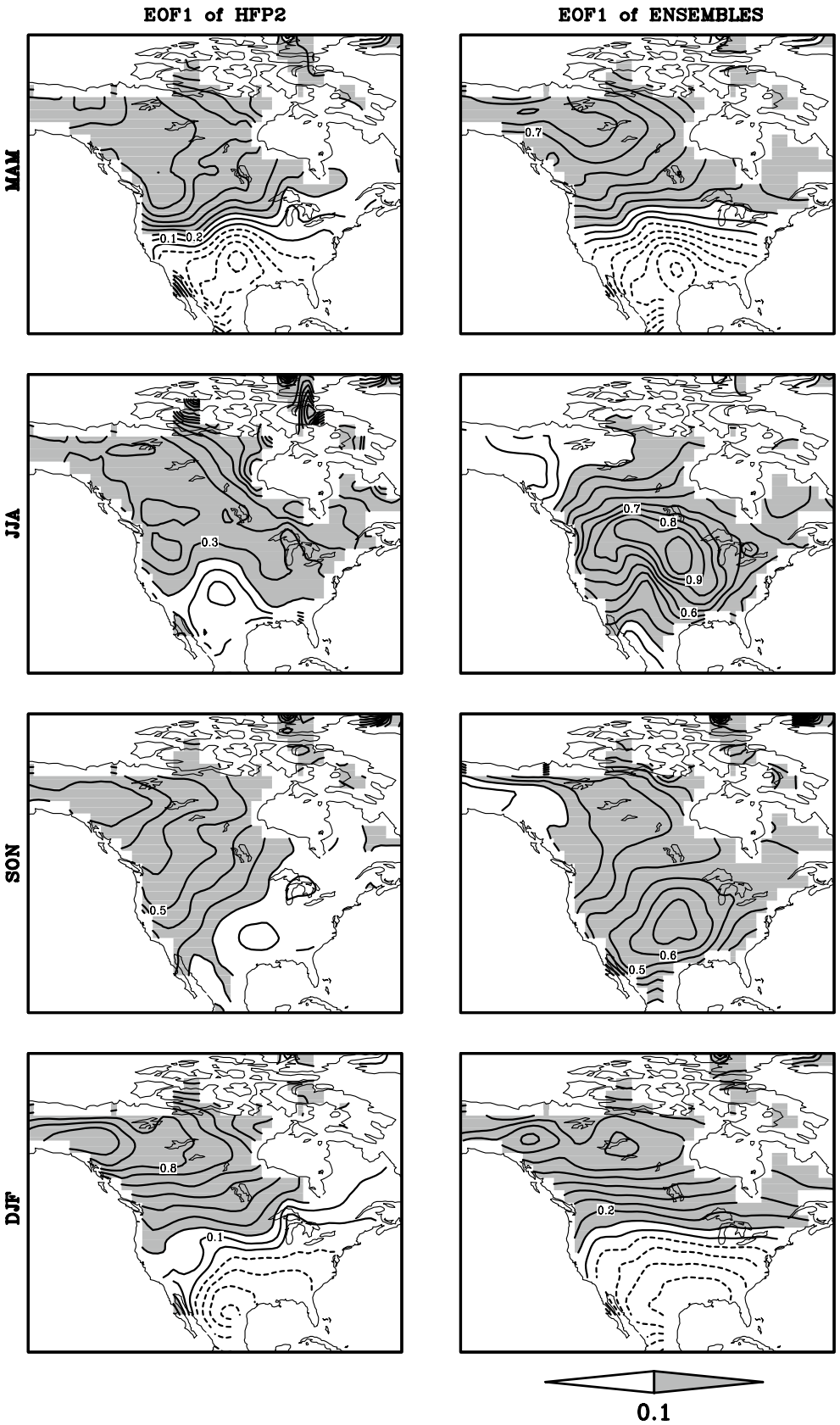


Fig. 9 The spatial pattern of the PCA first mode

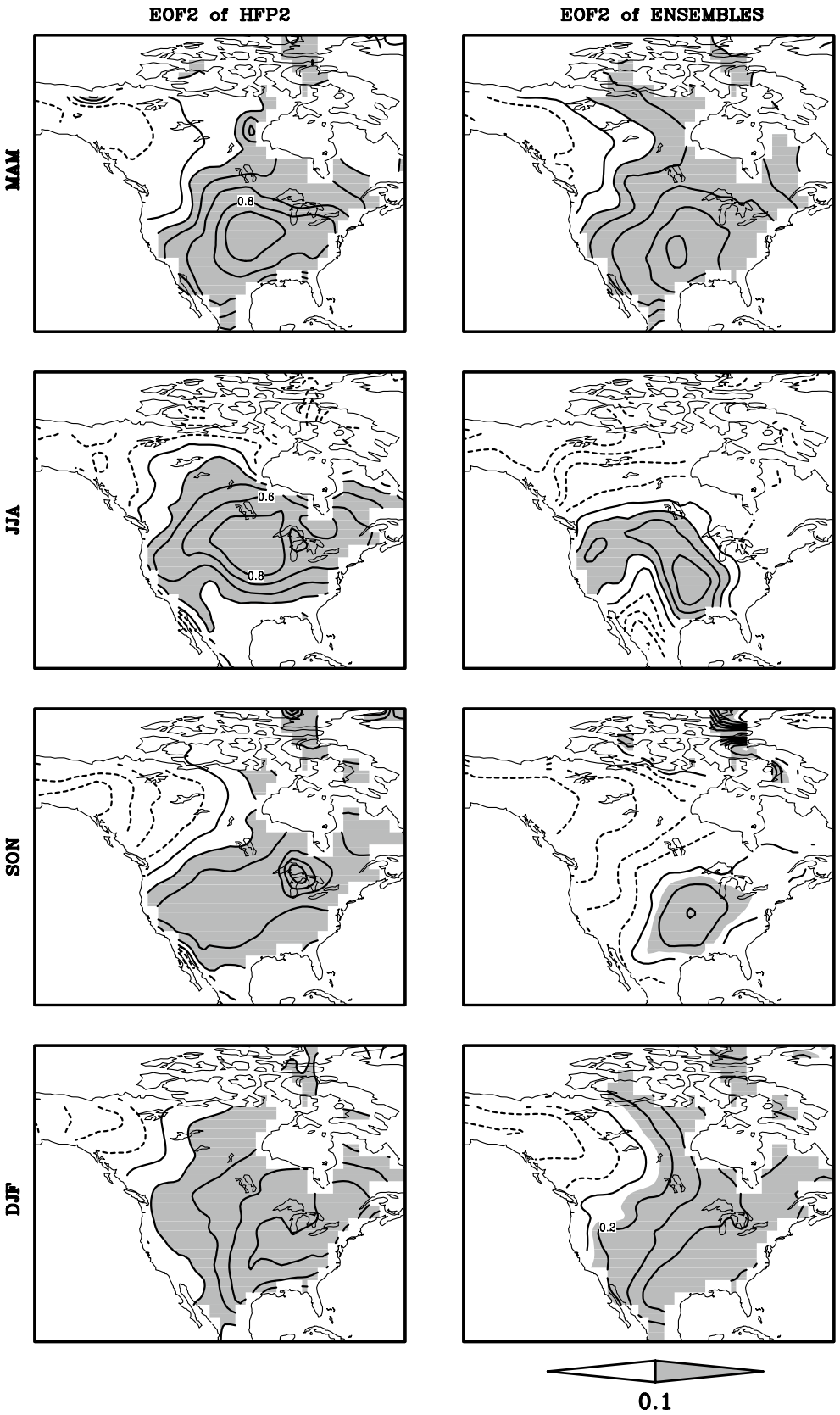


Fig. 10 The spatial pattern of the PCA second mode

than the inter-annual variability for all seasons. For inter-annual variability, the most predictable skill has significant seasonal dependence, being much higher in spring and winter than in summer and fall, which is most likely due to the seasonal variation of the ENSO forcing on the NA SAT. It has been recognized that the seasonal climate predictability in the mid-latitudes primarily originates from the ENSO predictability. This presumption is supported by the fact that there exists a strong teleconnection between ENSO and the extra-tropical climate variability in both observations and modeling, as discussed in the introduction.

Table 2 Correlation coefficient of time series between PrCA and PCA

	MAM	JJA	SON	DJF
Mode1				
T1	0.97	0.96	0.39	0.85
T2	0.99	0.99	0.99	0.99
Mode2				
T1	0.61	−0.92	0.70	−0.18
T2	0.95	−0.98	0.97	0.65

To build a direct link between the seasonal climate predictability and ENSO, we derived the pattern of the tropical surface temperature anomalies (SSTA) associated with the NA-SAT PrCA modes by projecting the time series of the PrCA modes onto the tropical SSTA. This projection is equivalent to the regression coefficient of the time series of the PrCA mode to the SSTA. Figure 13 shows the SSTA spatial patterns associated with the PrCA mode of inter-annual variability for T1 and T2, where the observed SSTA is used. We also projected the PrCA modes onto the predicted SSTA by T1 and T2, respectively, and obtained associated patterns similar to Fig. 12 (not shown). Apparently, the PrCA mode of inter-annual variability is closely associated with ENSO. For example, the first PrCA mode of the NA SAT has strong inter-annual variability (dipole pattern) in spring and winter in T1 and T2, which corresponds with strong El Nino events in the same seasons, as shown in

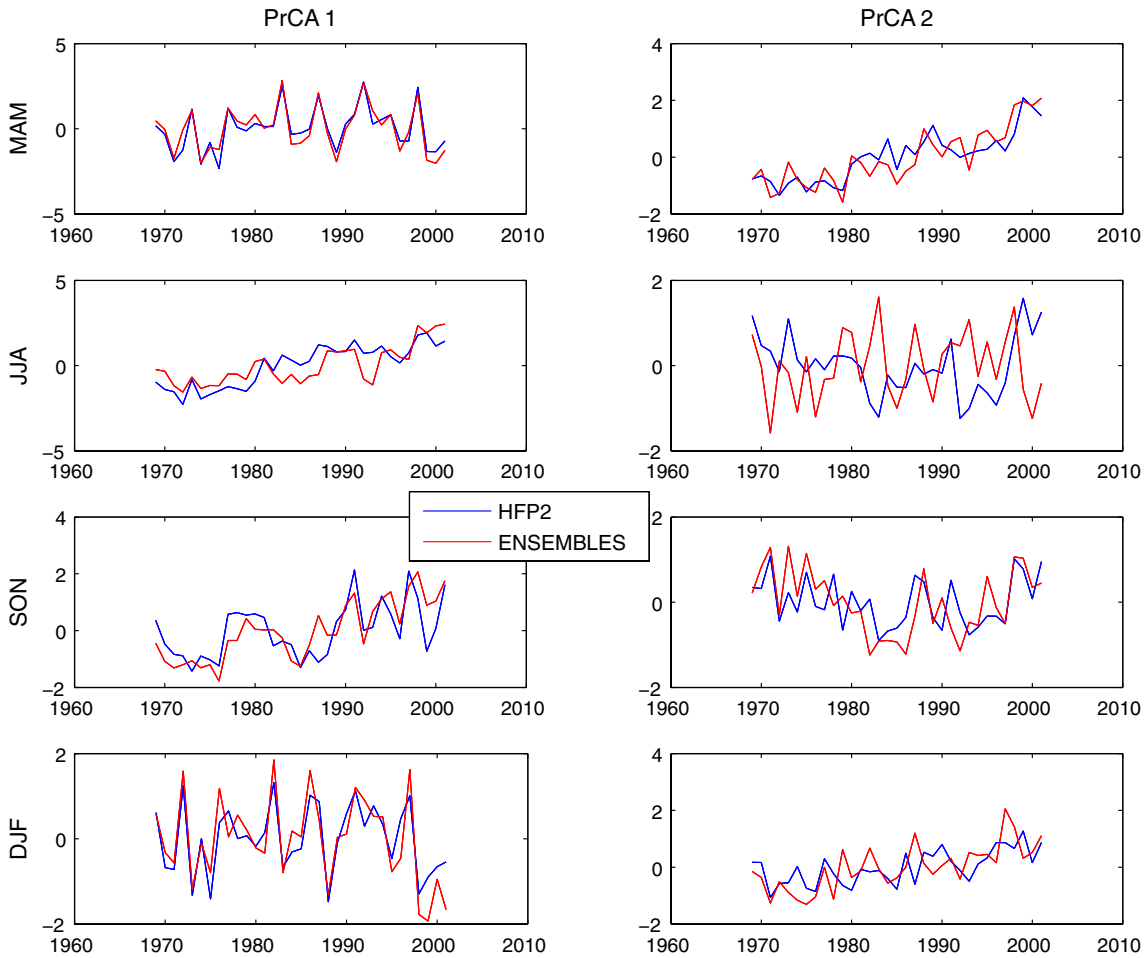


Fig. 11 Time series of PrCA for T1 and T2. Here the *red line* is for ENSEMBLES and the *blue* is for HFP2

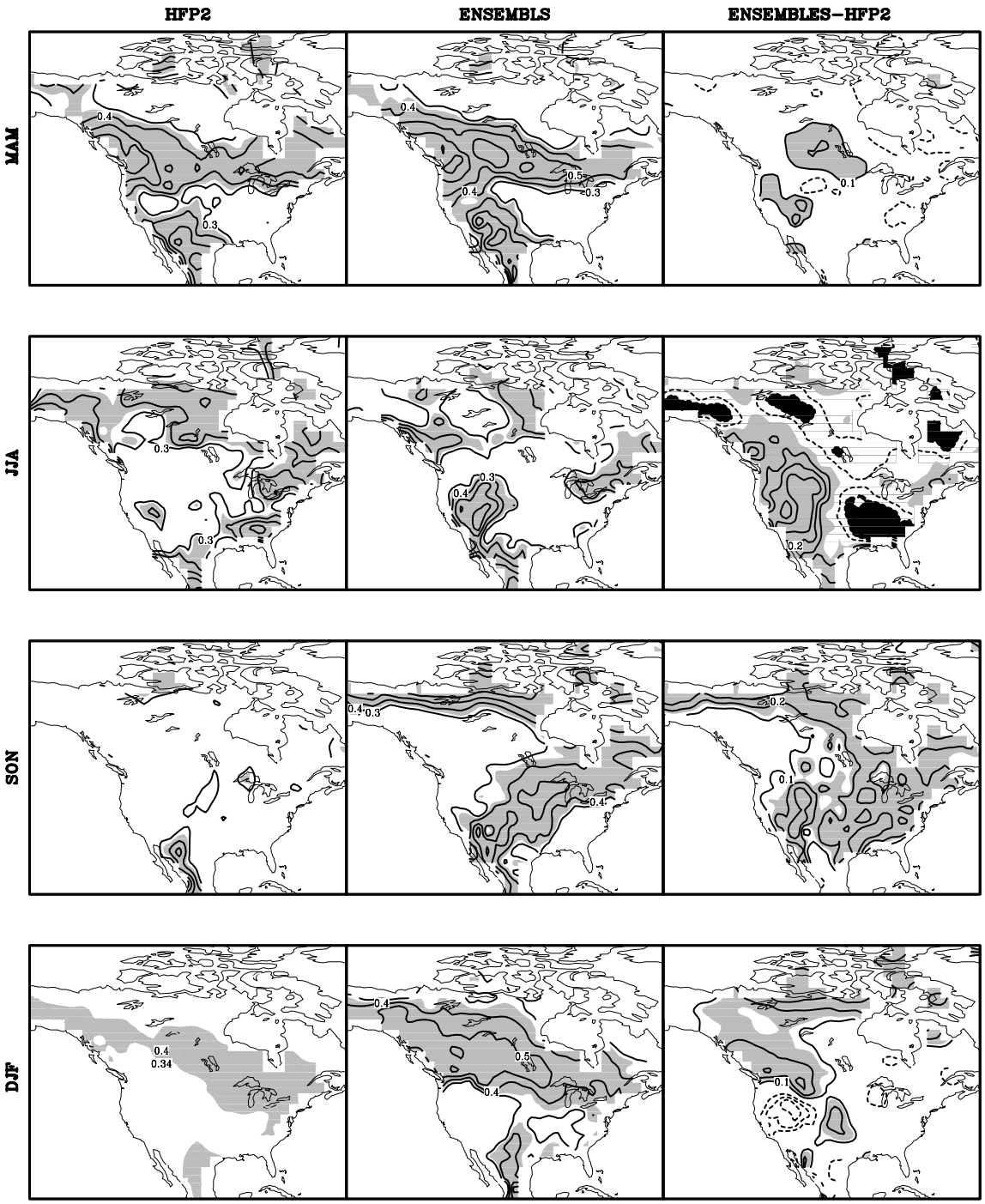


Fig. 12 Temporal anomaly correlations (CORR) of the NA SAT of HFP2 (*left panels*) and ENSEMBLES (*middle panels*) against CRU and their differences (*right panels*) in four seasons. CORR and their difference above the 95 % significance level are *shaded*

Fig. 13. A similar argument can be made for the association of the PrCA modes of the NA SAT with ENSO for summer and fall, when both the PrCA and the SSTA show relatively weak variability in T1 and T2. Thus, the most predictable component of the inter-annual variability of the NA SAT is clearly associated with the ENSO forcing.

As addressed above, the PrCA modes of interannual variability are characterized by a dipole structure. A further comparison between Figs. 7, 8 and 13 reveals that the El Nino events correspond with warming in northwestern Canada and cooling in the southeastern US. This is particularly obvious in winter and spring. However, the La Nina

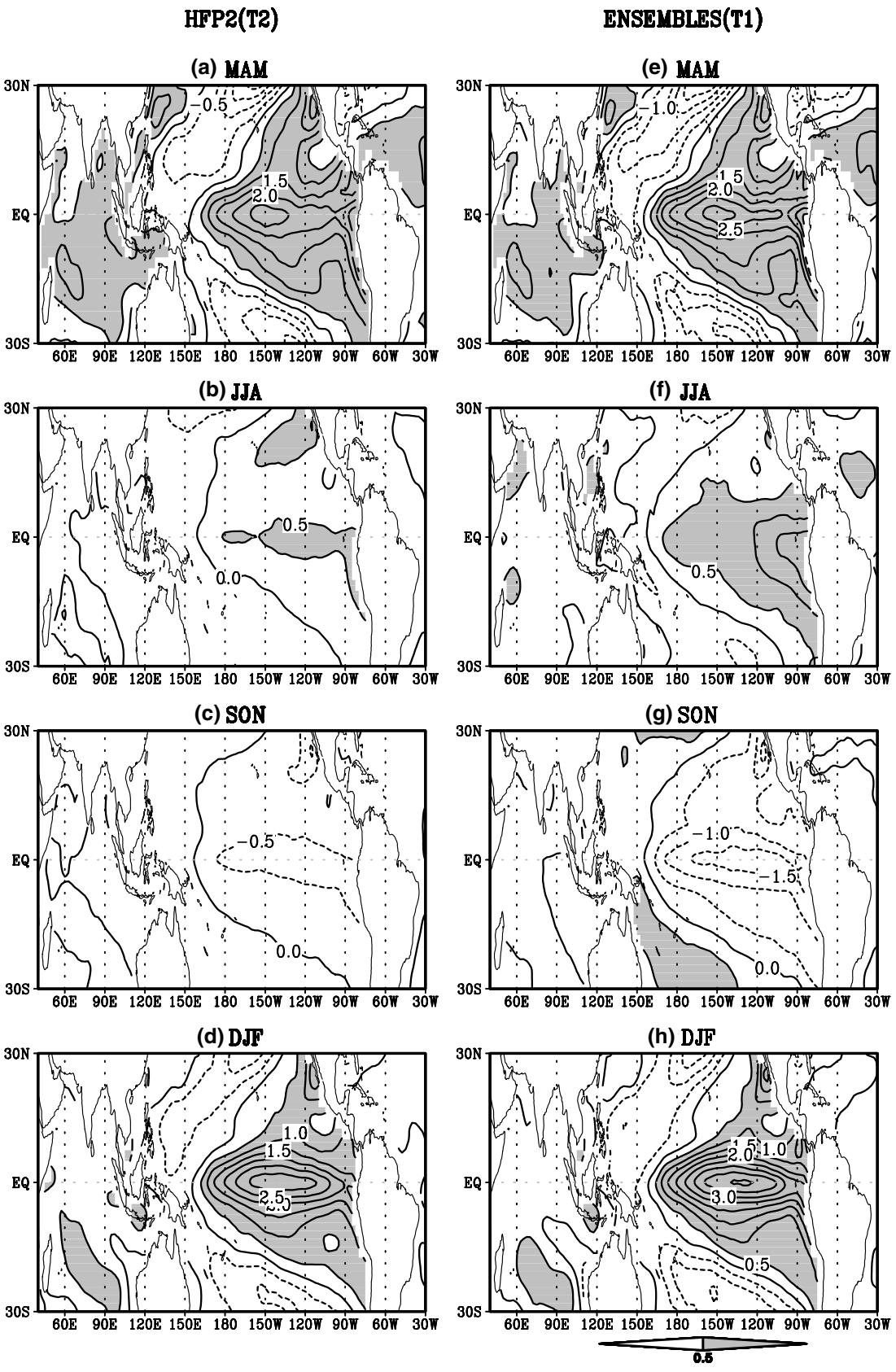


Fig. 13 SST pattern associated with inter-annual variability of the most predictable modes, which is obtained using predicted SSTA projected on the PrCA mode of the NA SAT

Table 3 Prediction skill (correlation) of each PrCA and PCA mode against observed counterpart

	MAM		JJA		SON		DJF	
	PrCA	PCA	PrCA	PCA	PrCA	PCA	PrCA	PCA
Mode1								
T1	0.73	0.45	0.75	0.57	0.71	0.69	0.63	0.50
T2	0.81	0.47	0.68	0.59	0.46	0.28	0.58	0.35
Mode2								
T1	0.77	0.54	0.47	0.32	0.40	0.19	0.67	0.45
T2	0.69	0.49	0.41	0.23	0.20	0.18	0.41	0.46

events lead to a cold climate in northwestern Canada and a warm climate in the southeastern US, as shown in fall. Such a teleconnection between the dipole structure of the PrCA of the NA SAT with ENSO can be bridged by PNA (e.g., Trenberth et al. 1998; Quan et al. 2006; Hoerling and Kumar 2002; Tang et al. 2014). For example, the positive PNA phase associated with the warming phase of ENSO features a stronger than usual height over the intermountain region of North America, denoting the advection of maritime air into the western and northern parts of North America and warming from subsidence directly under the high pressure ridge. Also during the positive phase of the PNA, deepening troughs of low pressure exist over the southern United States, bringing colder air farther into the South from the northern latitudes, increasing the winter storm activity for these regions, and causing below average temperatures across south-central and southeastern US.

A noticeable feature in Fig. 13 is that the SSTA forcing associated with the PrCA mode of inter-annual variability is much stronger in T1 than in T2 for summer and fall, with the magnitude of SSTA forcing in T1 approximately three times higher than that in T2. The reason why the summer and fall have a much stronger SSTA forcing in T1 than in T2 is not clear. However, this result seems to explain why T1 has higher skill than T2 in summer and particularly in fall as shown in Fig. 12.

6 Summary and discussion

In this study, the potential predictability of the seasonal mean anomalies of the NA SAT was examined using information-based metrics and multiple model ensembles. Emphasis is placed on the comparison of the potential predictability between two ensemble products: one from atmospheric models only (T2) and the other from fully coupled models (T1). It was found that the potential predictability of the NA SAT is seasonally and spatially dependent in both T1 and T2. The high predictability occurs in spring and winter and over the southeastern US and northwestern Canada. A further analysis of the signal and noise components indicates that there is a larger noise variance than signal variance in all seasons and over the entire domain of NA in both T1 and T2. This may explain why NA has a relatively lower potential predictability than the tropical region, as reported in the literature (e.g., Kug et al. 2008; Kumar et al. 2003; Livezey et al. 1996; Yang et al. 2012). While the signal strength shows strong seasonal and geographical dependence, the noise strength seems less varied in space and in season. Thus, the seasonal and geographical distribution of the NA SAT potential predictability is determined by the signal component, which is highly associated with the ENSO forcing.

Comparison of both T1 and T2 reveals that T1 has a larger signal variance than T2 in all seasons for most areas of NA. This is also true for the noise variance. Consequently, there is no significant difference of potential predictability between T1 and T2 for most areas of NA. The exception occurs in the southern US, where the difference of signal variance between T1 and T2 is larger than their difference of noise variance, resulting in a higher potential predictability in T1 than in T2.

The most predictable analysis (PrCA) was performed for both T1 and T2. It was found that the first two most predictable components of the NA SAT characterize the predictability of the inter-annual variability and long-term trend in both T1 and T2. For the long-term trend, its most predictable component is spatially less varied, suggesting uniform warming over the entire domain of NA, whereas for the inter-annual variability, the most predictable component displays a dipole spatial structure, characterized by the out-of-phase variation between northwestern Canada and southeastern US. Such a dipole structure has a seasonal dependence, with the most significance in spring and winter, followed by summer and fall. This is most likely due to the seasonal variation of the strength of the SST forcing.

A comparison of the PrCA between T1 and T2 reveals that the first mode explains more variances in T2 than in T1. This is because the signal is most likely more dominant in T2, so that its leading PrCA mode explains the largest variance contribution. It should be noted that the PrCA leading mode does not necessarily account for more variance because the PrCA seeks the optimal modes based on the predictable component (predictability) rather than on

the explained variance. In other words, the T2 may underestimate the realistic noise.

To explore the ability of PrCA to characterize the predictability, we calculated the correlation skill of the predicted PrCA mode against its observation counterpart. Comparison between T1 and T2 shows that the most predictable modes in T1 can better characterize the predictability. This is particularly apparent for the inter-annual variability mode in the fall, when the T1 PrCA mode has a much higher skill than the T2 counterpart. This is most likely because the SSTA forcing associated with the PrCA mode of inter-annual variability is much stronger in T1 than in T2 in fall, as discussed in Sect. 4 and in other literature (Kumar et al. 1998). This may explain why there is a significant difference of actual predictable skill between T1 and T2 in fall.

Several hypotheses should be mentioned. First, an ideal comparison of predictability between T1 and T2 should use the same atmospheric models so that their difference mainly reflects the contribution of coupling to predictability. In this study, the T1 and T2 use different atmospheric models. Thus, one should take caution to interpret the difference of the actual prediction skill between T1 and T2 reported in study. However, this work addresses the potential predictability that assumes the prediction systems to be perfect, thus allowing us to use two ensemble products with different atmospheric components. Namely, the difference of the potential predictability between T1 and T2 reported here should be interpreted as a different response of two perfect prediction systems (coupling and non-coupling) to initial errors, controlled by their nonlinear and stochastic dynamics, as in predictability study using the optimal error growth (e.g., Duan et al. 2009; Duan and Wei 2012). On the other hand, the multiple model ensembles used in this study can also greatly alleviate the prediction difference due to different atmospheric models. In addition, all analyses in this work are based on a Gaussian assumption, which allows for the analytic solutions of RE and MI. Fortunately, the Gaussian assumption usually holds well for the climate predictions of monthly or seasonal mean although some exceptions may exist. Nevertheless this work addresses theoretical merits and differences between T1 and T2, providing some bases of the development of T1 and T2 for the seasonal climate prediction. Also, the finding that the skill of the PrCAs is better than that of PCAs for all seasons offers a way to extract the better predictable component, which can help us improve the operational prediction for corresponding prediction target.

Acknowledgments This work was supported by the Canada NSERC Discovery Grant, NSF of China (41276029), and the National Key Basic Research Program of China (Grant No. 2013CB430302).

References

Allen MR, Smith LA (1997) Optimal filtering in singular spectrumanalysis. Phys Lett 234:419–428

Barnett TP, Preisendorfer R (1987) Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis. Mon Weather Rev 115:1825–1850

Barnston AG, Smith TM (1996) Specification and prediction of global surface temperature and precipitation from global SST using CCA. J Clim 9:2660–2697

Brunsell NA, Wilson CJ (2013) Multiscale interactions between water and carbon fluxes and environmental variables in a central U.S. Grassland. Entropy 15:1324–1341. doi:10.3390/e15041324

DelSole T (2004) Predictability and information theory. Part I: measures of predictability. J Atmos Sci 61:2425–2440

DelSole T, Tippett MK (2007) Predictability: recent insights from information theory. Rev Geophys 45:RG4002. doi:10.1029/2006RG000202

Derome J, Brunet G, Plante A, Gagnon N, Boer GJ, Zwiers FW, Lambert S, Ritchie H (2001) Seasonal predictions based on two dynamical models. Atmos Ocean 39:485–501

Doblas-Reyes FJ et al (2009) Addressing model uncertainty in seasonal and annual dynamical ensemble forecasts. doi:10.1002/qj.464

Duan W, Wei C (2012) The 'spring predictability barrier' for ENSO predictions and its possible mechanism: results from a fully coupled model. Int J Climatol. doi:10.1002/joc.3513

Duan W, Liu X, Zhu K, Mu M (2009) Exploring the initial errors that cause a significant "spring predictability barrier" for El Niño events. J Geophys Res Ocean. doi:10.1029/2008JC004925

Hoerling MP, Kumar A (2002) Atmospheric response patterns associated with tropical forcing. J Clim 15:2184–2203

Jha B, Kumar A (2009) A comparison of the atmospheric response to ENSO in coupled and uncoupled model simulations. Mon Weather Rev 137:479–487

Jia X, Lin H, Derome J (2010) Improving seasonal forecast skill of North American surface air temperature in fall using a post-processing method. Mon Weather Rev 138:1843–1857

Kharin VV, Teng Q, Zwiers FW, Boer GJ, Derome J, Fontecilla JS (2009) Skill assessment of seasonal hindcasts from the Canadian Historical Forecast Project. Atmos Ocean 47:204–223

Kirtman BP (2003) The COLA anomaly coupled model: ensemble ENSO prediction. Mon Weather Rev 10:2324–2341

Kirtman BP, Min D (2009) Multimodel ensemble ENSO prediction with CCSM and CFS. Mon Weather Rev 137:2908–2930

Kleeman (2002) Measuring dynamical prediction utility using relative entropy. J Atmos Sci 59:2057–2072

Krishnamurti TN, Kishtawal CM, LaRow TE, Bachiochi DR, Zhang Z, Willford CE, Gadgil S, Surendran S (1999) Improved weather and seasonal climate prediction forecasts from multimodel superensemble. Science 285:1548–1550

Krishnamurti TN, Kishtawal CM, Zhang Z, LaRow T, Bachiochi D, Williford E, Gadgil S, Surendran S (2000) Multimodel ensemble forecasts for weather and seasonal climate. J Clim 13:4196–4216

Kug J-S, Kang I-S, Choi D-H (2008) Seasonal climate predictability with tier-one and tier-two prediction systems. Clim Dyn 31:403–416

Kumar A, Hoerling MP (1998) Annual cycle of pacific-North American seasonal predictability associated with different phases of ENSO. J Clim 11:3295–3308

Kumar A, Hoerling MP, Schubert SD, Suarez MS (2003) Variability and predictability of 200-mb seasonal mean height during summer and winter. J Geophys Res 108:4169. doi:10.1029/2002JD002728

Latif M, Barnett TP (1994) Causes of decadal climate variability over the north pacific and North America. Science 266:634–637

Lin H, Brunet G, Derome J (2008) Seasonal forecasts of Canadian winter precipitation by postprocessing GCM integrations. Mon Weather Rev 136:769–783

Livezey RE, Masutani M, Ji M (1996) SST-forced seasonal simulation and prediction skill for versions of the NCEP/MRF model. Bull Am Meteorol Soc 77:507–517

Luo J, Masson S, Behera S, Shingu S, Yamagata T (2005) Seasonal climate predictability in a coupled OAGCM using a different approach for ensemble forecasts. J Clim 18:4474–4496

Misra V, Li H, Wu Z, DiNapoli S (2013) Global seasonal climate predictability in a two tiered forecast system: part I: boreal summer and fall seasons. Clim Dyn. doi:10.1007/s00382-013-1812-y

Palmer T et al (2004) Development of a European multi-model ensemble system for seasonal to inter-annual prediction (DEMETER). Bull Am Meteorol Soc 85:853–872

Palmer TN, Doblas-Reyes FJ, Weisheimer A, Rodwell MJ (2009) Toward seamless prediction: calibration of climate change projections using seasonal forecasts reply. Bull Am Meteorol Soc 90:1551–1554

Peng P, Kumar A, Wang W (2011) An analysis of seasonal predictability in coupled model forecasts. Clim Dyn. doi:10.1007/s00382-009-0711-8

Quan X, Hoerling M, Whitaker J, Bates G, Xu T (2006) Diagnosing sources of U.S. seasonal forecast skill. J Clim 19:3279–3293

Ritchie H (1991) Application of the semi-Lagrangian method to a multilevel spectral primitive equations model. Q J R Meteorol Soc 117:91–106

Rowell D (1998) Assessing potential seasonal predictability with an ensemble of multidecadal GCM simulations. J Clim 11:109–120

Schlosser CA, Kirtman BP (2005) Predictable skill and its association to sea-surface temperature variations in an ensemble climate simulation. J Geophys Res 110:D19107. doi:10.1029/2005JD005835

Schneider T, Griffies SM (1999) A conceptual framework for predictability studies. J Clim 12:3133–3155

Shabbar A, Barnston AG (1996) Skill of seasonal climate forecasts in Canada using canonical correlation analysis. Mon Weather Rev 124:2370–2385

Shukla J et al (1998) Predictability in the midst of chaos: a scientific basis for climate forecasting. Science 282:728–731

Shukla J et al (2000) Dynamical seasonal prediction. Bull Am Meteorol Soc 81:2593–2606

Sutton RT, Jewson SP, Rowell DP (2000) The elements of climate variability in the tropical Atlantic region. J Clim 13:3261–3284

Tang Y, Kleeman R, Moore A (2005) On the reliability of ENSO dynamical predictions. J Atmos Sci 62(6):1770–1791

Tang Y, Chen D, Yan X (2014) Potential predictability of Northern America surface temperature part I: information-based vs signal-to-noise based metrics. J Clim 27:1578–1599

Trenberth KE, Branstrator GW, Karoly D, Kumar A, Lau N-C, Ropelewski C (1998) Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. J Geophys Res 103(C7):14 291–14 324

Vautard R, Plaut G, Wang R, Brunet G (1999) Seasonal prediction of North American surface air temperatures using space-time principal components. J Clim 12:380–394

Venzke S, Allen MR (1999) The atmospheric response over the North Atlantic to decadal changes in sea surface temperature. J Clim 12:2562–2584

Wang B et al (2009) Advance and prospectus of seasonal prediction: assessment of the APCC/CliPAS 14-model ensemble retrospective seasonal prediction (1980–2004). Clim Dyn 33:93–117

Weisheimer A et al (2009) ENSEMBLES: a new multi-model ensemble for seasonal-to-annual predictions-skill and progress beyond DEMETER in forecasting tropical PacificSSTs. Geophys Res Lett 36(L21):711. doi:10.1029/2009GL040896

Yan X, Tang Y (2012) An analysis of multi-model ensemble for seasonal climate predictions. Q J R Meteorol Soc. doi:10.1002/qj.2019

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