Clim Dyn (2015) 44:2557–2569
DOI 10.1007/s00382-014-2398-8

On the potential of extratropical SST anomalies for improving climate predictions

Arun Kumar · Hui Wang

Received: 2 January 2014 / Accepted: 3 November 2014 / Published online: 19 November 2014

Abstract Skill for initialized decadal predictions for atmospheric and terrestrial variability is posited to reside in successful prediction of sea surface temperatures (SSTs) associated with the low-frequency modes of coupled ocean–atmosphere variability, for example, Pacific Decadal Oscillation (PDO) or Atlantic Multi-decadal Oscillation (AMO). So far, assessments of the skill of atmospheric and terrestrial variability in decadal predictions, however, have not been encouraging. Similarly, in the context of seasonal climate variability, teleconnections between SSTs associated with PDO and AMO and terrestrial climate have also been noted, but the same SST information used in predictive mode has failed to demonstrate convincing gains in skill. Are these results an artifact of model biases, or more a consequence of some fundamental property of coupled evolution of ocean–atmosphere system in extratropical latitudes, and the manner in which extratropical SST anomalies modulate (or constrain) atmospheric variability? Based on revisiting an analysis of a simple model that replicates the essential characteristics of coupled ocean–atmosphere interaction in extratropical latitudes, it is demonstrated that lack of additional skill in predicting atmospheric and terrestrial variability is more a consequence of fundamental characteristics of coupled evolution of ocean–atmosphere system. The results based on simple models are also substantiated following an analysis of a set of seasonal hindcasts with a fully coupled model.

Keywords Climate prediction · Seasonal prediction · Decadal prediction · Skill of climate predictions

1 Introduction

Recently, considerable efforts have been devoted to predicting the evolution of climate with a lead time of 1–30 years. These efforts are referred to as decadal climate predictions, and were included as part of the CMIP5 model evaluations (Taylor et al. 2012). The success of skillful decadal predictions relies on the premise in our ability to predict the low-frequency modes of coupled ocean–atmosphere variability (Soloman et al. 2011; Meehl et al. 2013). Examples of such low-frequency modes are Pacific Decadal Oscillation (PDO); Atlantic Multi-decadal Oscillation (AMO). The sea surface temperature (SST) signature of these modes often has extratropical fingerprint (Zhang and Delworth 2006; Meehl and Hu 2006; Wang et al. 2009; Sutton and Dong 2012). It is the expectation that skillful prediction of these SSTs, and subsequent response in atmospheric and terrestrial variability will lead to useful prediction of variables of societal relevance such as surface temperature and precipitation.

The ongoing assessment of skill of decadal prediction efforts in forecasting terrestrial surface temperature and precipitation so far has been disappointing. The skill of decadal predictions is generally assessed based on value added by initializing the current state of the ocean and comparing skill against the uninitialized predictions (Goddard et al. 2013). Of the analyses that do provide an assessment of skill over terrestrial regions (Teng et al. 2011; van Oldenborgh et al. 2012; MacLeod et al. 2012; Kim et al. 2012; Muller et al. 2012; Goddard et al. 2013; Doblas-Reyes et al. 2013), the results have not been encouraging. In some

A. Kumar (✉) · H. Wang
Climate Prediction Center, NOAA/NWS/NCEP, 5830 University Research Court, NCWCP, College Park, MD 20740, USA
e-mail: arun.kumar@noaa.gov

H. Wang
Innovim, Greenbelt, MD, USA

of the analyses it has also been noted that for ensemble mean prediction the observed associations between the SST fingerprint of low-frequency modes of SST variability and the atmospheric and terrestrial variability are not replicated (Teng et al. 2011; van Oldenborgh et al. 2012; Muller et al. 2012). Some specific examples from the published literature include: "…A1B and Commitment ensembles produce similar AMOC, subsurface temperature, and SST mean signals during the first decade. It is more difficult to find coherent signals in the atmospheric circulation…." (Teng et al. 2011); "…In spite of the fact that these teleconnections were also active over the period 1960–2009, the multi-model ensemble does not show skill in these regions, nor in other regions with AMO or decadal ENSO teleconnections…" (van Oldenborgh et al. 2012). In the ensemble prediction mode, is the inability to reproduce observed associations between the SST fingerprint of PDO, AMO and corresponding atmospheric and terrestrial variables an artifact of model biases or a consequence of some fundamental property of coupled evolution of ocean–atmosphere system in extratropical latitudes?

On the shorter lead time prediction of variability of seasonal time scales, a similar issue of predictive value of SST fingerprint associated with PDO and AMO is continued to be debated. Observational results based on the simultaneous analysis of SST, atmospheric, and terrestrial variability show distinct associations between them (Gershunov and Barnett 1998; Dai 2013; Mills and Walsh 2013). Such associations are then conjectured to add predictive value towards improving skill of seasonal predictions, for example, "…Considering the large consistency of the atmospheric variables associated with the PDO within a season and the persistence of the PDO index over several months, there is the potential for modest amounts of PDO-derived atmospheric predictability on monthly to seasonal time scales over the geographical regions highlighted here…" (Mills and Walsh 2013).

Against the backdrop of observational studies on seasonal time scale, a different set of observational studies analyzing the predictive value of SST associated with low-frequency modes such as PDO, done in forecast mode (in that only the past value of SSTs is used to predict the atmospheric and terrestrial variability in future) has not shown promising results (Davis 1976; Guztler et al. 2002). Inferences based on general circulation model simulations have also found little influence of extratropical SSTs in constraining atmospheric and terrestrial variability (Pierce 2002; Kumar et al. 2013). We note that a constraining influence of extratropical SST anomalies in modulating atmospheric and terrestrial anomalies is a prerequisite for the predictive value of those SSTs in climate predictions; the influence of SST anomalies associated with El Niño—Southern Oscillation (ENSO) being such an example (Trenberth et al. 1998; Hoerling et al. 2001). Once again, is the dichotomy between results based on simultaneous analysis versus done in a predictive mode a consequence of some fundamental property of coupled evolution of ocean–atmosphere system in extratropical latitudes?

In this paper, by revisiting a simple model of coupled ocean–atmosphere variability we attempt to demonstrate that the discrepancy noted in seasonal and decadal prediction efforts, viz, skillful prediction of ensemble mean SST anomalies does not translate into added predictive value of atmospheric and terrestrial variables, is related to a fundamental property of coupled evolution of ocean–atmosphere system in extratropical latitudes. Our analysis is based on the model of Barsugli and Battisti (1998) and Bretherton and Battisti (2000) (hereafter referred to as BB1998 and BB2000 respectively), and particularly the analysis of BB2000. We note that although part of our analysis may seem like merely a revisit of BB2000, it is different, and is of importance in two aspects: (i) the analysis establishes the connection between the analyses of BB2000 and series of recent results on the disappointing side of initialized decadal predictions in improving skill of terrestrial quantities, or the discrepancy between associations noted in observational studies and their inability for improving skill of seasonal predictions. We note that a connection with the analyses of BB200 has not been put forth as a possible explanation for the lack of predictive skill over land, and (ii) the analyses of BB2000 are extended in a predictive mode that was discussed only peripherally in the concluding section of BB2000.

2 The model and its validation

The simple model of BB1998 and BB2000 is for the coupled evolution of an atmospheric (*Ta*) and an oceanic (*To*) temperature index. The time evolution of two indices is given by the following equations:

$$\frac{\partial T_a}{\partial t} = -aT_a + bT_o + N(t) \tag{1}$$

$$\beta\frac{\partial T_o}{\partial t} = cT_a - dT_o \tag{2}$$

The nomenclature follows that of BB2000, and a (=1.12) and d (=1.08) are the damping terms, c (=1) and b (=0.5) provide the coupling between ocean and atmosphere, β (=40) is related to the heat capacity of the oceanic mixed layer, and N is the Gaussian white noise. Details about the model formulation are given in BB1998 and BB2000. The indices *Ta* can be interpreted as the index associated with an atmospheric signature of the variability, the PDO, for example, and *To* as the index for the corresponding fingerprint in the ocean SST variability.

We first demonstrate the applicability of the simple model by comparing it against results from a coupled ocean–atmosphere general circulation model (COAGCM). The COAGCM simulation used in this analysis is a 500-year simulation with the National Centers for Environmental Prediction (NCEP) Climate Forecast System (CFS) coupled model version 1 (Saha et al. 2006). Results from the same simulation were also reported in Kumar et al. (2013).

For the purpose of comparison between COAGCM and the simple model, we focus on the ocean and atmosphere variability related to the PDO. For COAGCM, Fig. 1 (top panels) shows the spatial structure of the atmospheric and oceanic components of the PDO variability, and the corresponding time series are shown in the bottom panels. The PDO in the COAGCM simulation was identified based on empirical orthogonal function (EOF) analysis. Further

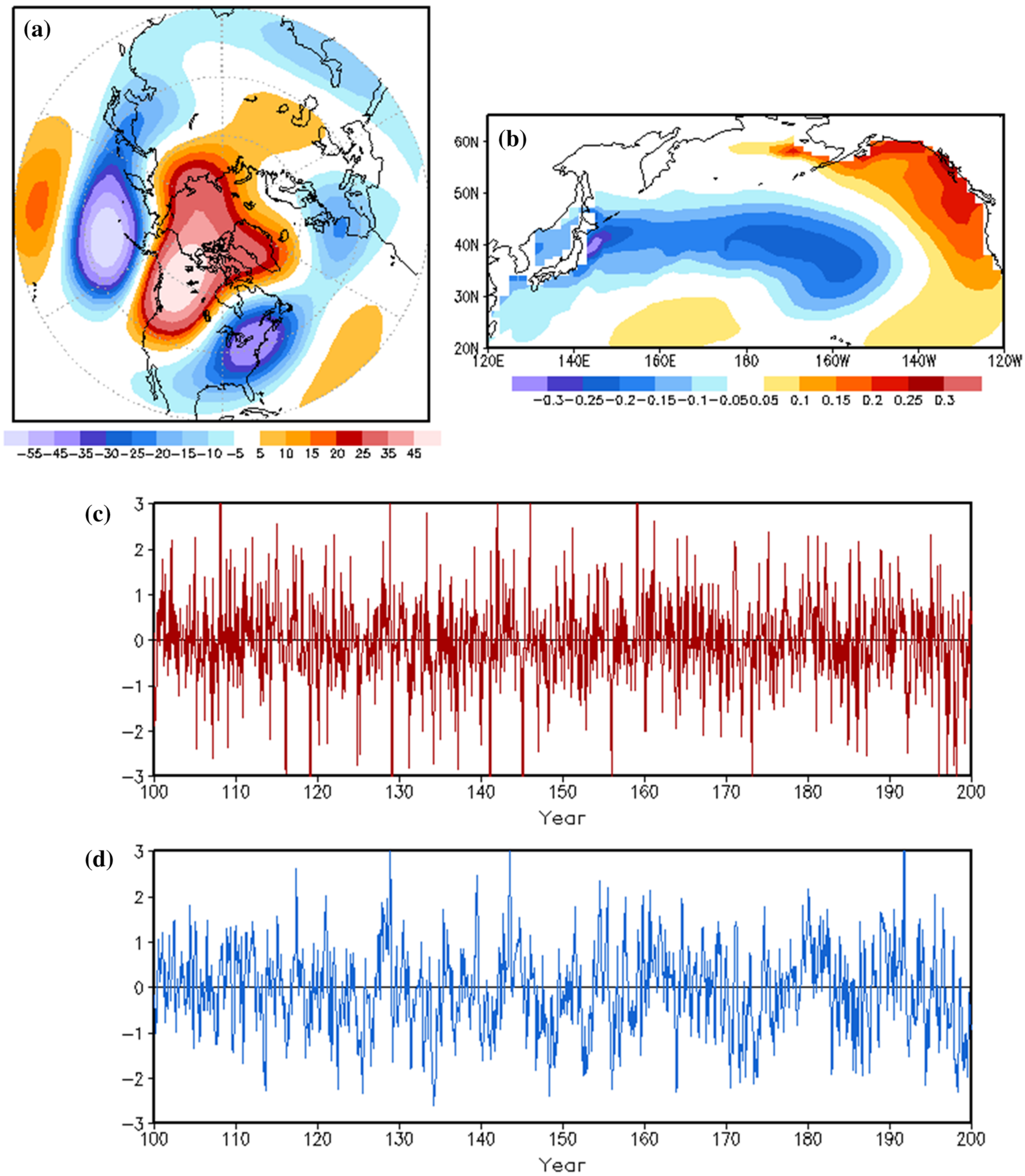

Fig. 1 **a** Spatial pattern of H200 anomalies (unit: gpm) associated with the leading EOF of monthly mean H200 computed over the Pacific–North American region (20°–90°N, 150°E–30°W), **b** SST anomalies (unit: K) associated with the EOF of H200 in (**a**), (**c**) normalized PC time series (*red*) of the first EOF of H200, and **d** normalized time series (*blue*) of projection coefficients for SST anomalies projecting onto the SST pattern in (**b**). The time series in (**c**) and (**d**) are only shown from year 100 to 200

details appear in Kumar et al. (2013) who also referred to the SST fingerprint of the PDO as the oceanic signature of the PDO (OPDO) and the atmospheric fingerprint as the atmospheric signature of the PDO (APDO).

Of relevance to our analysis are the characteristics of time series of the leading EOF of 200 hPa heights (H200) and the associated SST. The time series associated with H200 has high-frequency variations, while for SST, variations are on a longer time scale. This contrast is summarized succinctly by autocorrelations for the respective time series that are shown in Fig. 2 (top panel, black curves). The autocorrelation for the time series for monthly mean H200 has a very sharp drop with 1-month autocorrelation of ~0.2. In contrast, the autocorrelation for the time series for the SST pattern is on a longer time scale, and the value of ~0.2 autocorrelation is reached at about 6 months.

The lead-lag correlation between H200 and SST time series is shown in Fig. 2 (bottom panel, black curve) and illustrates the direction of forcing in the evolution of the coupled ocean–atmosphere system. The correlation is larger (smaller) when H200 leads (lags) the SST and indicates that it is the variability in H200 (the atmospheric component of the PDO or the APDO) that forces the variations in SST (the ocean component of the PDO, or the OPDO), while the influence of OPDO in constraining the future variations of the APDO is much smaller (Kumar et al. 2013).

The results shown in Fig. 2 capture the basic concepts of coupled ocean–atmosphere variability in the extratropical latitudes (Davis 1976; Frankignoul and Hasselmann 1977): atmospheric variability can be considered as a white noise, and hence has a sharp drop in autocorrelation; a larger thermal inertia of the ocean translates white noise atmospheric forcing into variability on a longer time scale, and as a consequence, the autocorrelation is also characterized by a longer time-scale; the lead-lag correlation being appreciably larger when the APDO index leads the OPDO index signifies that the atmospheric variability is the forcing for the ocean with little feedback from the ocean back to the atmosphere.

We next show that in the simple model if *To* and *Ta* are considered as a proxy for the APDO and OPDO index

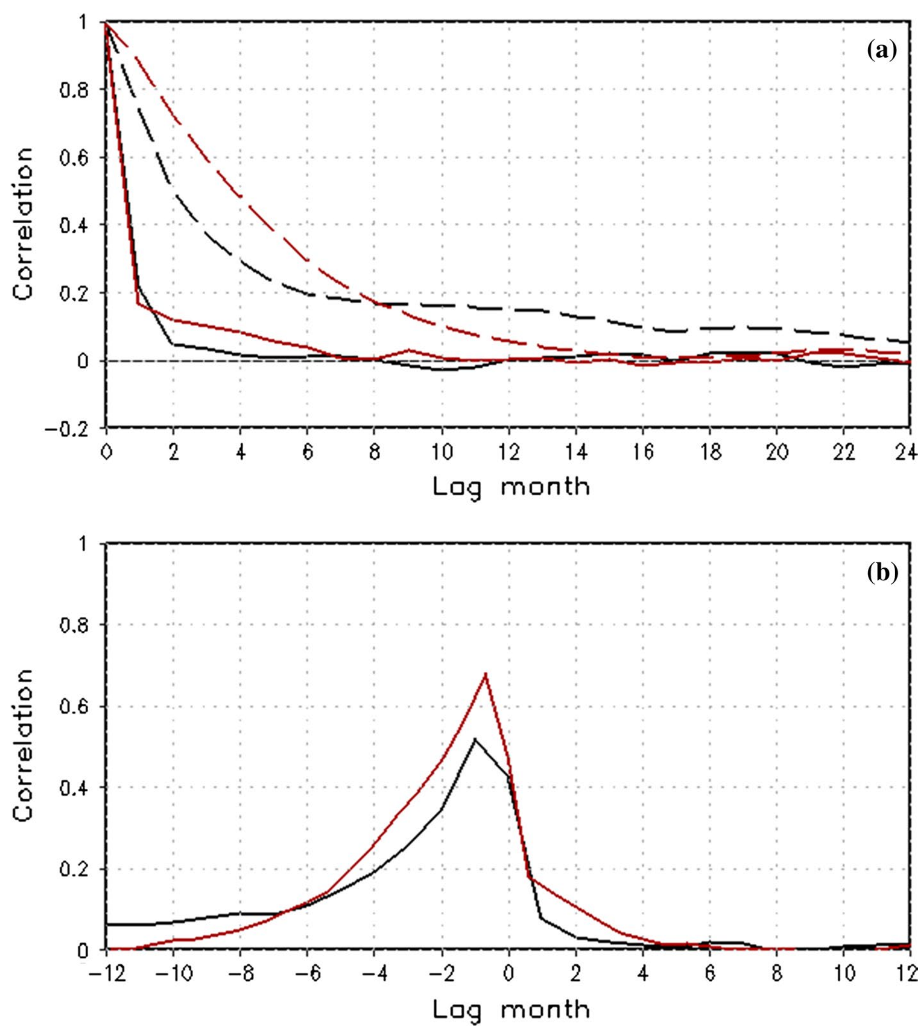

Fig. 2 **a** Autocorrelations for the 500-year-long monthly PC time series of the first EOF of H200 (*solid black line*) and the time series of SST projection coefficients (*dashed black line*) from the COAGCM simulation and autocorrelations for the monthly time series of atmospheric anomaly *Ta* (*solid red line*) and ocean anomaly *To* (*dashed red line*) from the simple model. **b** Lead–lag correlations between the 500-year monthly PC time series of the first EOF of H200 and the time series of SST projection coefficients (*black*) and between the time series of *Ta* and *To* (*red*) from the simple model. Negative lags in (**b**) mean that the H200 (*Ta*) leads the SST (*To*)

respectively, the autocorrelations and lead-lag correlations successfully replicate the results based on the COAGCM simulation. For a specific choice of parameters listed earlier, the simple model was integrated for one million time steps and the autocorrelations for *Ta* and *To* and lead-lag correlation between *Ta* and *To* are also shown in Fig. 2 top and bottom panels respectively (red curves). The autocorrelation for *Ta* has a sharp drop while *To* has a longer time scale. Lead-lag correlations are larger when *Ta* leads *To* and the correlation drops sharply when *To* leads *Ta*. The behavior of autocorrelations and lead-lag correlation based on the simple model captures the corresponding behavior for the COAGCM simulation. The simplified model, therefore, captures the essence of coupled ocean–atmosphere evolution in the extratropical latitudes simulated in the COAGCM. With this fact in hand, and similar to BB2000, we use the convenience of the simple model for further analysis.

3 Results

3.1 Atmospheric response to specified SSTs

We first revisit the results of BB2000 in the setting of an atmospheric general circulation model simulation forced with specified time evolution of ocean temperature (the so-called AMIP simulations). This analysis, although is the same as in BB2000, provides the necessary background and helps in understanding the discussion related to the extension of BB2000 in a setting of seasonal and decadal predictions.

The ensemble of AMIP simulations for the analysis is based on following steps:

- From a single integration of the simplified model for one million time steps we have the time series of *Ta* and *To*;
- The evolution of *To* time series is then specified in the integration of the atmospheric component (Eq. 1) (i.e., the AMIP mode) of the simplified model but with a different realization of the Gaussian white noise. The latter choice is justified on the basis that the atmospheric forcing is stochastic, and the same stochastic forcing that led to the generation of the original *Ta* and *To* is not expected to be replicated in initialized predictions;
- An ensemble of atmospheric realizations with 1,000 members was generated with different stochastic forcing, and the atmospheric time series thus generated is denoted by *T'ia* where prime differentiates the "replayed" atmospheric time series from the original time series *Ta*, and *i* indicates the ensemble member.

The process described above is equivalent to generating ensemble of initialized climate predictions with perturbed initial conditions based on COAGCMs.

The correlations between original atmospheric time series *Ta* and the replayed time series *T'ia*, for various time averages (achieved by band-pass filtering of time series with different frequencies) are shown in Fig. 3. As for BB2000 the correlations are done in two ways: (i) correlation of individual *T'ia* in the ensemble of 1,000 realizations with *Ta*, and (ii) correlation of ensemble mean *T'a* with *Ta*. The results reproduce those of BB2000:

- The average of the correlations between individual replayed realizations and original atmospheric time series increases as longer time scales are retained.

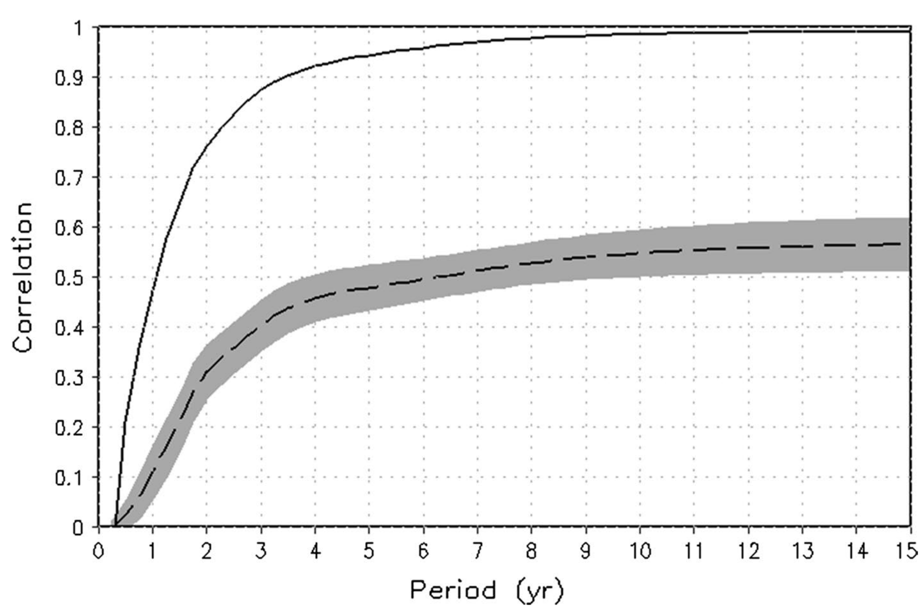

Fig. 3 Correlations (*solid line*) between band-pass filtered original atmospheric time series and band-pass filtered 1,000-member ensemble mean time series of the AMIP simulations and the averages (*dashed line*) of the correlations between the band-pass filtered original atmospheric time series and band-pass filtered time series of individual members of the AMIP simulations. The *grey shading* indicates ± one standard deviation of the spreads of the correlations with the 1,000 members around the mean values

At lower frequencies, the replayed atmospheric time series, therefore, captures larger and larger fraction of the original atmospheric time series *Ta*. For very long time-averages, the correlation between two time series asymptotes at ~0.5 or ~25 % variance of original *Ta* is captured;

- Due to sampling variability in individual realizations induced by different realizations of atmospheric noise, there is a spread in the correlations and is indicated by the grey area straddling the average correlation;
- The correlation of the ensemble mean of individual realizations *T'a* with *Ta* is substantially higher. For long time-averages, the correlation for the ensemble mean asymptotes to one indicating that all of low-frequency variations in the original atmospheric time series *Ta* can be captured by the replayed ensemble mean time series when *To* is specified. However, the amplitude of the ensemble mean time series is smaller (not shown). The reason that correlation with the ensemble mean is higher is because the noise component in the *T'a* is removed (Kumar and Hoerling 2000).

In the traditional setting of AMIP simulations where the observed time series of SST (or in our case, *To*) is specified, the component of the variance of the original atmospheric time series captured by the replayed atmospheric time series is interpreted as the atmospheric response to specified SSTs. Indeed, such interpretation is generally made in the context of assessment of potential predictability using AMIP simulations (Barnett et al. 1997; Phelps et al. 2004). This interpretation has been successfully used in the context of atmospheric response to tropical SST variability particularly related to ENSO. Further, this interpretation is justified in the context of ENSO where SSTs indeed represent an external forcing for the atmospheric variability (Wu et al. 2006; Chen et al. 2012), and the atmospheric response is potentially predictable if ENSO SSTs themselves can be predicted.

The interpretation of the replayed atmospheric variability as an atmospheric response to ocean forcing (*To*) for the extratropical ocean–atmosphere variability, and the use of the atmospheric response interpretation in a predictive sense, however, is problematic. This is so because (i) as the lead-lag correlation analysis indicates (Fig. 2) that the forcing for the ocean is primarily from the stochastic atmospheric variability, and (ii) the future state of the ocean (for it to subsequently replay *Ta*) can only be known if the evolution of atmospheric stochastic noise can be predicted in the first place. Therefore, interpretation of the replayed component of the atmospheric variability as an atmospheric response in a predictive sense cannot be made as it relies on a knowledge of future evolution of atmospheric stochastic (and unpredictable) noise. This was the essence of the brief concluding section of BB2000 stating that "Despite the excellent hindcast skill of an atmospheric ensemble forced by observed SSTs, less than 15 % of the variance in the seasonal atmospheric anomaly is predictable six months in advance…." In the next section we clarify this concept further using initialized prediction experiments based on the simple model, and this analysis is a more detailed articulation of BB2000 concluding remark. In Sect. 3.3 we validate prediction results based on a simple model using data from a seasonal prediction system.

3.2 Skill of *Ta* and *To* for initialized predictions

The fallacy of atmospheric response interpretation in the context of predictive utility is clarified by assessing the correlation skill of initialized predictions and how it differs from the correlation obtained in the AMIP simulations (Fig. 3). The set of initialized predictions is done in the following manner:

- From the original time series of *Ta* and *To* we start initialized predictions by selecting 1,000 initial values of *Ta* and *To* that are 1,000 time steps apart;
- For each pair of initial values of *Ta* and *To* an ensemble of predictions of size 100 based on Eqs. 1 and 2 is generated by using different realizations of atmospheric noise.

The average skill of initialized predictions is computed based on temporal correlation between the ensemble average of 100 predictions of *Ta* and *To* and the corresponding original *Ta* and *To* (which can now be interpreted as the verifying observations) over predictions from 1,000 different initial conditions. The difference between the analysis done in AMIP mode (discussed in Sect. 3.1) and the initialized predictions is that for the former, Eq. 1 is integrated with specified *To* to simulate subsequent values of *Ta*; for the initialized prediction both *Ta* and *To* are predicted using both Eqs. 1 and 2.

The average prediction skill for monthly mean *Ta* and *To* is shown in Fig. 4 (black curves). The skill for *Ta* has a much smaller value than for *To* from the very beginning of the forecast. This difference in prediction skill is consistent with the autocorrelations (or the persistence time scale) for respective time series (Fig. 2)—atmospheric anomalies lose their memory on a much faster time scale than the ocean anomalies do.

To give a better feel for initialized predictions, we illustrate two specific examples in Fig. 5. For the example in the top panel, integration starts from a state when both *Ta* and *To* have substantial initial anomalies. The example on the bottom panel starts from a state for which the initial values of *Ta* and *To* are closer to the climatological value of zero.

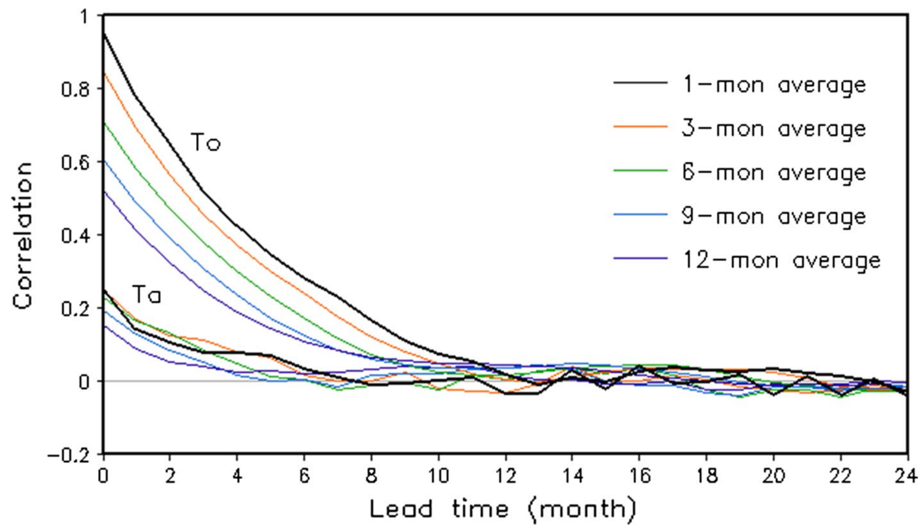

Fig. 4 Forecast skills of *Ta* and *To* with data averaged over different time averages from monthly mean to 12-month mean based on prediction using the simple model

For the first case, the ensemble mean prediction of *Ta* (red line) and of *To* (blue line) drifts towards the climatology, and is to be expected for initialized ensemble predictions, i.e., for longer lead predictions, the probability density function (PDF) of the ensemble of predictions will converge to the climatological PDF (Peng et al. 2011; Branstator and Teng 2010). Even though predictions started with the same initial conditions, the spread among predictions for *Ta* quickly reaches the climatological value (Fig. 5, orange curves). For *To*, one also sees the evolving divergence between initialized predictions within the ensemble members with lead time, and convergence to the climatological spread. For the case when the prediction starts from an initial state closer to the climatology, features similar to the previous case are also found. For completeness, Fig. 5 also shows the verifying time series of *Ta* and *To* (black curves).

One might argue that increasing time averages may lead to a better predictive skill. This, however, is not the case as is also illustrated in Fig. 4 where prediction skill for longer time-averages is also shown (color curves). This is because of two factors that have opposing influence on the prediction skill (Chen et al. 2013). The time-averaging, by reducing the contribution of stochastic noise, increases the signal and should lead to higher predictive skill due to an increase in signal-to-noise ratio (Kumar and Hoerling 2000). However, increasing the lead time of prediction that is necessary for longer time averages also leads to a reduction in skill countering the positive influence of increased time-averaging on prediction skill (Chen et al. 2013). Therefore, in contrast to increasing simulation skill for longer time averaging as was the case for the AMIP simulation (Fig. 3), such is not the case for initialized predictions.

3.3 Validation of simple model results based on seasonal prediction system

Results pertaining to forecast skill in Figs. 4 and 5 from the simple model are validated using the extensive hindcast data from a seasonal forecast system. Compared to the paucity of start dates and a small size of the ensemble for each start date in the current generation of decadal hindcasts (Taylor et al. 2012), seasonal forecast systems provide a much richer data set.

The seasonal forecast system used in this analysis is the NCEP Climate Forecast System version 2 (CFSv2). Of relevance for our discussion is the CFSv2 seasonal hindcasts for which four runs for nine target months were made every five days starting 1 January without considering 29 February in leap years. In this study, hindcasts from 1982–2010 are analyzed. For each month, an ensemble mean of 24 forecasts taken from the six available initial dates in the month is used to analyze forecasts for the subsequent months. For example, forecasts from 8, 13, 18, 23, 28 October, and 2 November, with four forecasts from each day, are used to compute ensemble mean forecast for nine subsequent months after October. Following this example, forecasts from October initial conditions for target months of November, December, January, etc., are referred to as forecasts at the lead time of 0, 1, 2 months, and so on. Further details about the CFSv2 hindcast configuration, and seasonal forecast system, in general, can be found in Kumar et al. (2012) and Saha et al. (2014).

Over the hindcast period, we analyze the skill for the ensemble mean of predicted ocean, atmosphere and terrestrial PDO indices against their observational counterparts. The ocean PDO index for observations and forecast are computed by projecting monthly mean SST anomalies

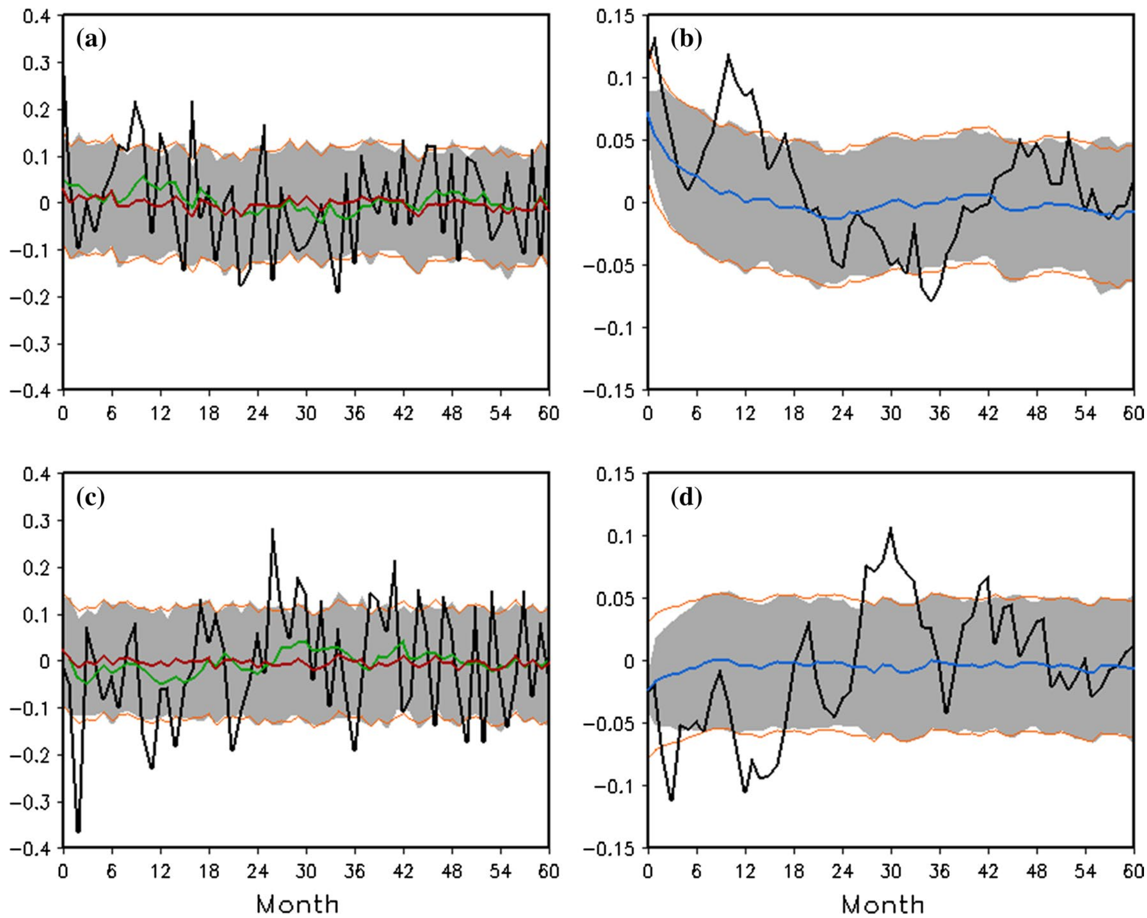

Fig. 5 Two examples of the original time series (*black*) of *Ta* (*left panels*) and *To* (*right panels*), the predicted time series of 100-member ensemble mean *Ta* (*red*) and *To* (*blue*) from month 0 to 60, and the time series (*green*) of 100-member ensemble mean *Ta* (*left panels*) from corresponding AMIP runs. *Grey shadings* denote the range of ± one standard deviation of the spreads of the 100 members around the mean values. The *orange lines* are ± one standard deviation of the climatological PDFs around the mean values

on the spatial pattern of SST associated with the observed PDO pattern (which is similar to Fig. 1, top right panel). Similarly, for the 2-m temperature (T2m) over land the corresponding observed and predicted PDO indices are computed by projecting monthly mean values over the North America domain (20°N–75°N, 55°W–170°W) on the spatial patterns of T2m associated with the PDO.

Figure 6 shows the time evolution of skill for the ensemble predicted PDO indices associated with SST over North Pacific with T2m over North America. Similar to that in Fig. 4 for results based on the simple model, the skill for the oceanic PDO index is initially much higher than skill for T2m index related to PDO. Differences in skill once again imply that after initializing the seasonal forecasts, although the fingerprint of the SSTs associated with the PDO can be predicted at longer leads (mainly due to larger thermal inertia of the oceans), it does not impart predictability on its terrestrial counterpart that is driven much more by stochastic variability.

Corresponding to Fig. 5 we also show examples of predictions for two specific years—forecasts from October 2001 and 2004 initial conditions. In Fig. 7, forecasts are shown for the evolution of projection coefficients of T2m and SST on the corresponding PDO patterns. The evolution of corresponding observed projection coefficients are shown in black. For T2m, the ensemble mean prediction (red line) converges quickly to zero, and further, the spread among 24 forecasts also reaches the climatological spread indicating a near complete loss of predictability. For SST, on the other hand, convergence of ensemble mean towards climatology, and further, the convergence of spread among individual forecasts towards the climatological spread are on a longer time scale. These results based on a comprehensive fully coupled seasonal prediction system are similar to those in Fig. 5 based on a simple model.

In the final analysis we demonstrate that for the seasonal forecast analyzed while the associations between oceanic and atmospheric or terrestrial variability are replicated on

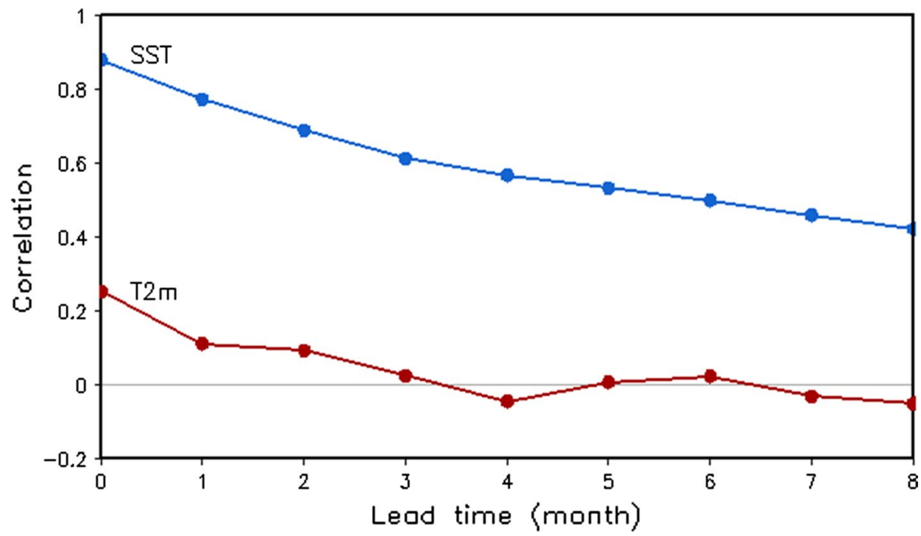

Fig. 6 Forecast skill of PDO-related SST (*blue*) in North Pacific and T2m (*red*) over North American land based on 24-member CFSv2 ensemble mean hindcasts with all initial conditions from January to December

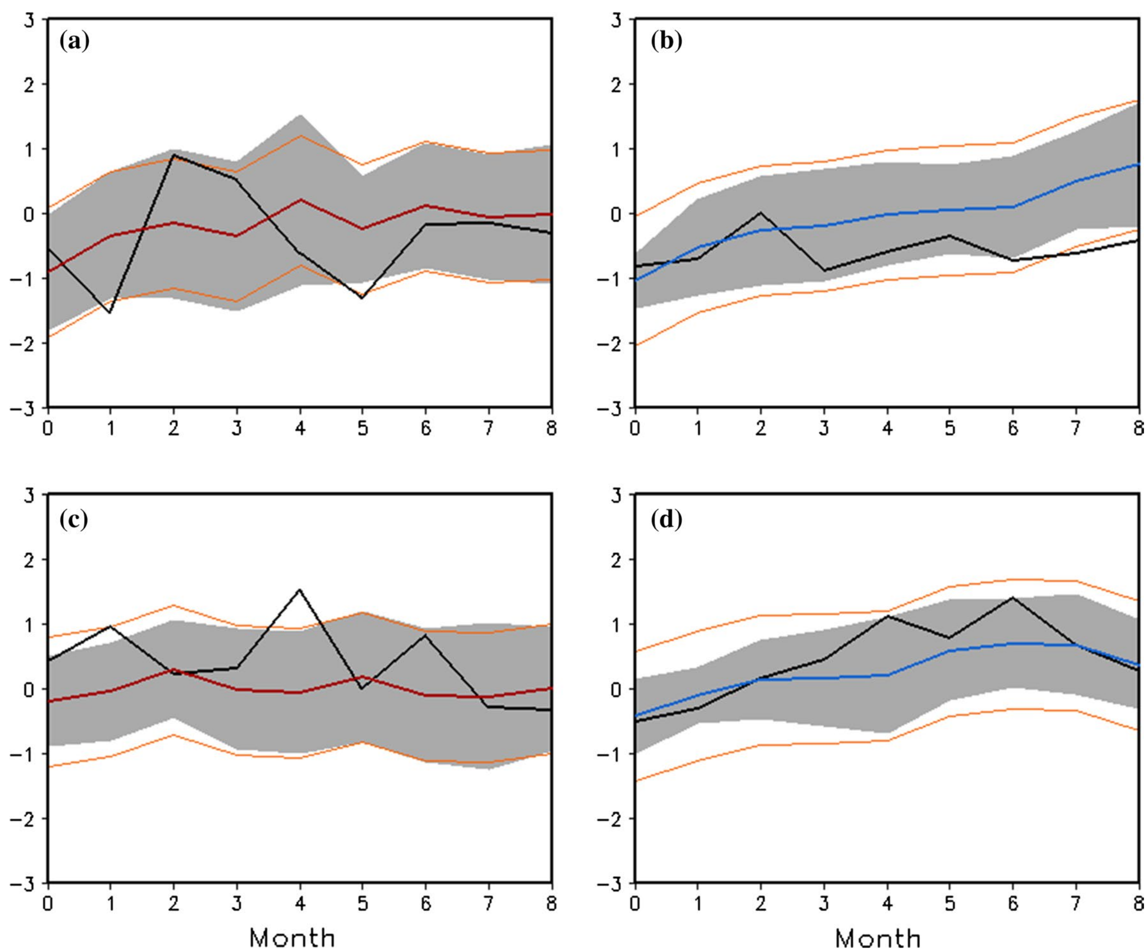

Fig. 7 *Top panels* CFSv2 9-month hindcasts with October 2001 initial conditions (24 members); *Bottom panels* CFSv2 9-month hindcasts with October 2004 initial conditions (24 members); T2m (*left*), SST (*right*); *Black lines* OBS; *Red* and *blue lines* 24-member ensemble mean forecasts for T2m and SST; The standard deviation of the spreads (*grey shading*) is normalized by the standard deviation of the climatological PDFs (*orange line*) to eliminate the seasonality

an individual forecast basis, the same does not happen for the analysis based on ensemble mean forecasts. Figure 8 shows the composite of SSTs (over the ocean), T2m (over land) and 500 hPa height based on selecting positive (negative) phases of OPDO index that are above (below) one standard deviation during each month of the forecast. We note that for each forecast month a pool of 696 (=24 initial conditions × 29 years) forecasts is available. The composites shown are defined as the difference between months with positive and negative phases of the OPDO index and divided by 2. This compositing procedure does not pay any attention to the initial value of the OPDO index or the forecast continuity.

Notable features in Fig. 8 are: (a) a smooth temporal progression of various spatial patterns that is reflection of slowly evolving associative relationship between different variables in the context of the PDO variability, and (b) almost a continuous transition from SST anomalies over the ocean to T2m over land reasons for which can be argued on the basis that both are driven by circulation anomalies as shown by 500 hPa height composites. The respective spatial patterns shown in Fig. 8 are very similar to those obtained based on simultaneous regression of SST, T2m and 500 hPa heights with the OPDO index (not shown).

For analyzing the ensemble mean forecasts we form composites based on the initial value of the OPDO index. We select positive (negative) phases of the OPDO index that are above (below) one sigma, and then follow the evolution of SST, T2m and 500 hPa height anomalies during subsequent forecast months. Based on this procedure 5 years (1982, 1987, 1992, 1993, 1997) with positive values of the OPDO index and 6 years (1998, 1999, 2001, 2005, 2007, 2010) with negative values of the OPDO index were selected. Forecast evolution in Fig. 9 is shown as one half the difference of composites between positive and negative OPDO years. We note that preponderance of positive (negative) phases of OPDO index before (after) 1998 is due to a low-frequency shift in the OPDO index from a positive to a negative phase around the same time.

In comparison of composites in Fig. 8, contrasting features to note in Fig. 9 following the evolution of ensemble mean composite of forecast are: (a) the spatial structure of SSTs in Fig. 9 over North Pacific is similar to that in Fig. 8. This is because years selected for the construction of composite in Fig. 9 are the years that had large projection of initial SST state on the oceanic fingerprint of the PDO. The results also show that the SSTs over the North Pacific slowly evolve towards the climatology; (b) the spatial continuity in anomalies from SSTs over the ocean to T2m over the land that is apparent in Fig. 8 is no longer maintained in Fig. 9. In fact, there is an abrupt

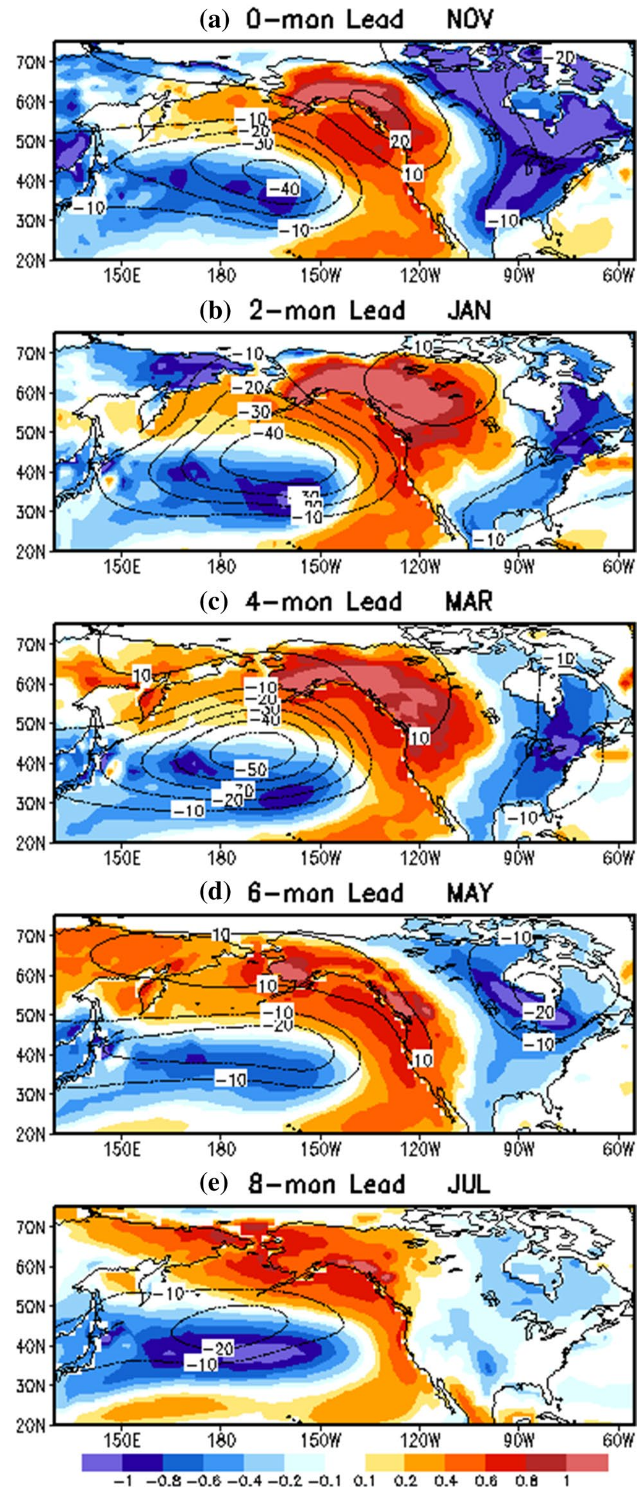

Fig. 8 Composite differences (+PDO minus −PDO)/2 of CFSv2 hindcasts with October initial conditions for SST (*shadings* over ocean, unit: K), T2m (*shadings* over land, unit: K), and 500-hPa height (contours) based on 24 individual members and corresponding PDO index. The contour interval is 10 gpm with zero contours omitted. Prior to the composite, ENSO-related anomalies are removed

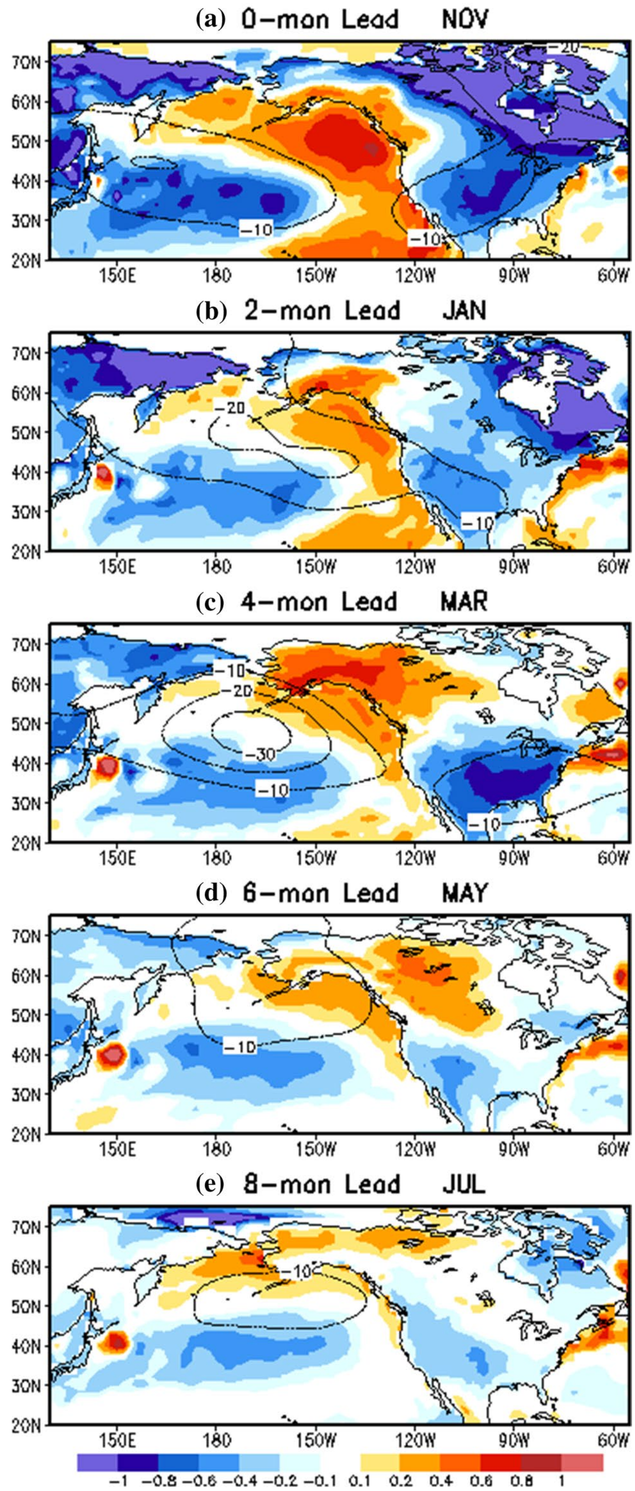

Fig. 9 As in Fig. 8, but for composites of 24-member ensemble mean anomalies based on initial observed PDO phases

transition from positive SST anomalies over the western coast of the U.S. to negative inland anomalies. The reason for this is lack of appreciable atmospheric circulation anomalies (also shown in Fig. 9) that provided the reason for an apparent connection between SST anomalies over the ocean and T2m anomalies over the land for the composites shown in Fig. 8.

The results shown in Figs. 8 and 9 are reminiscent of those described in earlier studies, i.e., although on an individual forecast basis apparent teleconnection between oceanic fingerprints of low-frequency modes of coupled variability and atmospheric and terrestrial variability exists, the same connection is missing when the analysis is based on the ensemble mean prediction. The reason being that once initialized, ensemble mean SST anomalies, because of larger thermal inertia, persist over a period. However, their lack of constraint on the atmospheric variability does not lead to appreciable (or consistent) atmospheric circulation anomalies that exist on an individual forecast basis. On an individual forecast basis, changes in SST (superimposed on the persistence) still occur and are forced by chaotic atmospheric variability, and are reflected as associations between SST and atmosphere as modes of coupled variability.

4 Summary and discussion

The results of BB2000 simple model run in the predictive mode, and evolution of skill with lead time, are fundamental in explaining the dichotomy seen in the results reported for decadal predictions and for resolving the debate about the potential utility of the SST fingerprint of PDO and AMO in enhancing seasonal prediction skill. Initialized decadal and seasonal predictions start from the observed state of SSTs, often with initial SSTs projecting strongly on the low-frequency modes of coupled ocean–atmosphere variability. Ensemble of these predictions essentially follows the experimental setup discussed in Sect. 3.2.

If SSTs in extratropical oceans do not constrain the atmospheric variability in the extratropical latitudes (as evidence by lead-lag correlations in Fig. 2), then as the ensemble mean of SST anomalies evolves towards the climatology following their typical persistence (and slower) time scale, the ensemble mean of atmospheric and terrestrial anomalies still converges to climatology on much faster time scale. The prediction skill of atmospheric and terrestrial quantities, assessed by comparing ensemble mean of initialized predictions against the corresponding observed single realization then leads to a low prediction skill. This is the case noticed in decadal and seasonal predictions in the context of having a demonstrable skill for extratropical SSTs without finding corresponding skill in terrestrial quantities.

The other characteristic of initialized predictions—associations that are found for individual forecasts but are not found between ensemble means of predicted SSTs and terrestrial quantities—also follows from the discussion above.

The associations (or teleconnections) between atmospheric and SST variability based on analyzing individual model simulations or forecasts capture the near-simultaneous relationship between the coupled evolution of ocean and atmosphere. Indeed, as was demonstrated in Sect. 3.3 and Fig. 8, for the individual forecasts such relationships are captured. For the evolution of the ensemble mean of forecasts, however, it is the atmospheric response to SST fingerprints related to extratropical modes of coupled variability that is of relevance, and while the SST fingerprint can be found in the early part of the forecast integration, corresponding atmospheric response consistent with associations inferred based on individual forecasts is non-existent.

Our analysis building upon the analysis of BB2000 provides an explanation for (a) why the skill of atmospheric and terrestrial quantities in initialized decadal predictions is not much better than their uninitialized counterpart, (b) why the observed teleconnection between SST and atmospheric and terrestrial quantities is not replicated in the ensemble of initialized decadal prediction runs, and (c) why similar teleconnection relationships on seasonal time scale have not translated to their application towards improving skill of seasonal predictions. The explanation is related to a fundamental aspect of coupled ocean–atmosphere variability in extratropical latitudes, and although, reductions in model biases will lead to improvements in prediction skill, the constraint of the coupled ocean–atmosphere variability will still be a basic limitation on prediction skill.

Acknowledgments Insigthful comments by Drs. Mingyue Chen and Peiteo Peng, and two anonymous reviewers are greatly appreciated. This study was supported by NOAA's Climate Program Office's Modeling, Analysis, Predictions, and Projections program.

References

Barnett TP, Arpe K, Bengtsson L, Ji M, Kumar A (1997) Potential predictability and AMIP implications of midlatitude climate variability in two general circulation models. J Clim 10:2321–2329

Barsugli JJ, Battisti DS (1998) The basic effects of atmosphere-ocean thermal coupling on midlatitude variability. J Atmos Sci 55:477–493

Branstator G, Teng H (2010) Two limits of initial-value decadal predictability in a CGCM. J Clim 23:6292–6311

Bretherton CS, Battisti DS (2000) An interpretation of the results from atmospheric general circulation models forced by the time history of the observed sea surface temperature distribution. Geophys Res Lett 27:767–770

Chen M, Wang W, Kumar A, Wang H, Jha B (2012) Ocean surface impacts on the seasonal precipitation over the tropical Indian Ocean. J Clim 25:3566–3582

Chen M, Wang W, Kumar A (2013) Lagged ensembles, forecast configuration, and seasonal predictions. Mon Weather Rev 141:3477–3497

Dai A (2013) The influence of the inter-decadal Pacific oscillation on US precipitation during 1923–2010. Clim Dyn. doi:10.1007/s00382-012-1446-5

Davis RE (1976) Predictability of sea surface temperature and sea level pressure anomalies over the North Pacific Ocean. J Phys Oceanogr 6:249–266

Doblas-Reyes FJ et al (2013) Initialized near-term regional climate change prediction. Nat Commun. doi:10.1038/ncomms2704

Frankignoul C, Hasselmann K (1977) Stochastic climate models. Part II: application to SST anomalies and thermocline variability. Tellus 29:289–305

Gershunov A, Barnett TP (1998) Interdecadal modulation of ENSO teleconnections. Bull Amer Meteor Soc 79:2715–2725

Goddard L et al (2013) A verification framework for interannual-to-decadal predictions experiments. Clim Dyn 40:245–272

Guztler DS, Kann DM, Thornbrugh C (2002) Modulation of ENSO-based long-lead outlooks of southwestern U.S. Winter Precipitation by the Pacific Decadal Oscillation. J Clim 17:1163–1172

Hoerling MP, Kumar A, Xu T-Y (2001) Robustness of the nonlinear climate response to ENSO's extreme cases. J Clim 14:1277–1293

Kim H-M, Webster PJ, Curry JA (2012) Evaluation of short-term climate change prediction in multi-model CMIP5 decadal hindcasts. Geophys Res Lett 39:L10701. doi:10.1029/2012GL051644

Kumar A, Hoerling MP (2000) Analysis of a conceptual model of seasonal climate variability and implications for seasonal predictions. Bull Amer Meteor Soc 81:255–264

Kumar A et al (2012) An analysis of the non-stationarity in the bias of sea surface temperature forecasts for the NCEP climate forecast system (CFS) version 2. Mon Weather Rev 140:3003–3016

Kumar A, Wang H, Wang W, Xue Y, Hu Z-Z (2013) Does knowing the oceanic PDO phase help predict the atmospheric anomalies in subsequent months? J Clim 26:1268–1285

MacLeod DA, Caminade C, Morse AP (2012) Useful decadal climate prediction at regional scales? A look at the ENSEMBLES stream 2 decadal hindcasts. Environ Res Lett 7:044012. doi:10.1088/1748-9326/7/4/044012

Meehl GA, Hu A (2006) Megadroughts in the Indian monsoon region and southwest North America and a mechanism for associated multidecadal Pacific sea surface temperature anomalies. J Clim 19:1605–1623

Meehl GA, et al (2013) Decadal climate prediction: an update from the trenches. Bull Amer Meteo Soc. doi:10.1175/BAMS-D-12-00241.1

Mills CM, Walsh JE (2013) Seasonal variation and spatial patterns of the atmospheric component of the pacific decadal oscillation. J Clim 26:1575–1594

Muller WA et al (2012) Forecast skill of multi-year seasonal means in the decadal prediction system of the Max Planck Institute for Meteorology. Geophys Rea Lett 39:L22707. doi:10.1029/2012GL053326

Peng P, Kumar A, Wang W (2011) An analysis of seasonal predictability in coupled model forecasts. Clim Dyn 36:419–430

Phelps MW, Kumar A, O'Brien JJ (2004) Potential predictability in the NCEP/CPC dynamical seasonal forecast system. J Clim 17:3775–3785

Pierce DW (2002) The role of sea surface temperatures in interactions between ENSO and the North Pacific Oscillation. J Clim 15:1295–1308

Saha S, et al (2014) The NCEP climate forecast system version 2. J Clim 27:2185–2208. doi:10.1175/JCLI-D-12-00823.1

Saha S et al (2006) The NCEP climate forecast system. J Clim 19:3483–3517

Soloman A et al (2011) Distinguishing the roles of natural and anthropogenically forced decadal climate variability: implications for prediction. Bull Amer Meteor Soc 92:141–155

Sutton RT, Dong B (2012) Atlantic Ocean influence on a shift in European climate in the 1990s. Nat Geosci 5:788–792. doi:10.1038/NGEO1595

Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. Bull Amer Meteor Soc 93:485–498. doi:10.1175/BAMS-D-11-00094.1

Teng H, Branstator G, Meehl GA (2011) Predictability of the Atlantic overturning circulation and associated surface patterns in two CCSM3 climate change ensemble experiments. J Clim 24: 6054–6076

Trenberth EK, 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 107(C7):14291–14324

van Oldenborgh GJ, Doblas-Reyes FJ, Wouters B, Hazeleger W (2012) Decadal prediction skill in a multi-model ensemble. Clim Dyn 38:1263–1280

Wang H, Schubert S, Suarez M, Chen J, Hoerling M, Kumar A, Pegion P (2009) Attribution of the seasonality and regionality in climate trends over the United States during 1950–2000. J Clim 22:2571–2590

Wu R, Kirtman BP, Pegion K (2006) Local air–sea relationship in observations and model simulations. J Clim 19:4914–4932

Zhang R, Delworth TL (2006) Impact of Atlantic multidecadal oscillation on India/Sahel rainfall and Atlantic hurricanes. Geophys Res Lett 33:L17712. doi:10.1029/2006GL026267