CrossMark

Week 3–4 predictability over the United States assessed from two operational ensemble prediction systems

Lei Wang[1,2] · Andrew W. Robertson[3]

Abstract

The subseasonal predictability of surface temperature and precipitation is examined using two global ensemble prediction systems (ECMWF VarEPS and NCEP CFSv2), with an emphasis on the week 3–4 lead (i.e. 15–28 days ahead) fortnight-average anomaly correlation skill over the United States, in each calendar season. Although the ECMWF system exhibits slightly higher skill for both temperature and precipitation in general, these two systems show similar geographical variations in the week 3–4 skill in all seasons and encouraging skill in certain regions. The regions of skill are then interpreted in terms of large-scale teleconnection patterns. Over the southwest US in summer, the North American monsoon system leads to higher skill in precipitation and surface temperature, while high skill over northern California in spring is found to be associated with the seasonal variability of the Arctic Oscillation (AO). During winter, in particular, week 3–4 predictability is found to be higher during extreme phases of the El Niño–Southern Oscillation, Pacific-North American (PNA)/Tropical-Northern Hemisphere mode, and AO/North Atlantic Oscillation (NAO). Both forecast systems are found to predict these teleconnection indices quite skillfully, with the anomaly correlation of the wintertime NAO and PNA exceeding 0.5 for both models. In both models, the subseasonal contribution to the PNA skill is found to be larger than for the NAO, where the seasonal component is large.

1 Introduction

The prospect of substantial subseasonal to seasonal (S2S) climate predictability (2 weeks to a season ahead) has drawn increasing attention in recent years (Brunet et al. 2010), especially for the week 3–4 forecast range (i.e. 15–28 days ahead) where many operational weather forecast centers have started to issue experimental forecasts (Vitart et al. 2017). Great advances have been achieved in the past few decades in medium-range (i.e., 4–10 days) weather forecasts (Bauer et al. 2015) and seasonal (3–6 months) climate predictions (Kirtman et al. 2014), whereas the subseasonal forecast

range that lies in between has, until recently, received much less attention. A recent joint initiative of the World Weather and World Climate Research Programmes (WWRP/WCRP), the Subseasonal to Seasonal (S2S) Prediction Project, aims to bridge this gap (Vitart et al. 2012) and has organized a database of subseasonal forecasts from 11 operational centers around the world (Vitart et al. 2017).

Short and medium-range weather forecast skill comes mostly from atmospheric initial conditions, while seasonal forecast skill stems from slow-varying boundary conditions such as sea surface temperature (SST), snow and sea ice (e.g., Cohen nd Entekhabi 1999; García-Serrano et al. 2015), and soil moisture (e.g., Koster et al. 2010), as well as persistent stratospheric circulation features (e.g., Baldwin and Dunkerton 2001; Scaife et al. 2016). The subseasonal time scale is, however, influenced by both initial and boundary conditions, which makes it challenging to predict. On one hand, information in the atmospheric initial conditions is mostly lost beyond about 10 days and play a role only occasionally, as observational and systematic errors may amplify to a pronounced level at this lead time in numerical models. On the other hand, the stronger coupling between the atmosphere and boundary conditions takes place on longer

✉ Lei Wang
wanglei_ias@fudan.edu.cn

[1] Present Address: Department of Atmospheric and Oceanic Sciences, Institute of Atmospheric Sciences, Fudan University, 2005 Songhu Road, Yangpu, Shanghai 200438, China

[2] Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA

[3] International Research Institute for Climate and Society (IRI), Columbia University, Palisades, NY, USA

timescales and subseasonal forecasts may only marginally benefit from such processes. For example, the coupling at time scales of 2–3 weeks appears as high-frequency stochastic forcing by the atmosphere on the ocean mixed layer, while the ocean feedback on the atmosphere is still fairly weak, at least in midlatitudes (e.g., Deser and Timlin 1997). Li and Robertson (2015) examined week 1–4 forecast skill of boreal summer precipitation from three global prediction systems and found very little skill outside the tropics beyond 2 weeks of lead time. However, this could be due to the relatively small ensemble size in the reforecasts they analyzed, together with the limited number of reforecast start times. In this study, we revisit the subseasonal forecast skill for all calendar seasons over the United States from an updated set of reforecasts with more frequent starts and larger ensemble size (or using lagged ensembles); increased ensemble size often leads to better skill (e.g., Murphy 1988; Kumar 2009).

Subseasonal predictability in the extratropics is expected to be strongly modulated by planetary-scale teleconnection modes, and in particular the El Niño–Southern Oscillation (ENSO), North Atlantic Oscillation (NAO), Arctic Oscillation (AO), Pacific-North America (PNA) mode, Tropical-Northern Hemisphere (TNH) mode, and the Madden–Julian Oscillation (MJO); see the recent review of Stan et al. (2017). For example, the subseasonal variability of wintertime surface air temperature anomalies over North America has been found to be more predictable during certain phases of strong MJO events (Zhou et al. 2012; Rodney et al. 2013). Certain combinations of ENSO and MJO phases may boost subseasonal predictability over North America creating time windows of opportunity for more skillful forecasts (Johnson et al. 2014). By examining weekly boreal summer precipitation anomalies over a part of the Maritime Continent (Borneo Island), Li and Robertson (2015) found clear MJO modulation of precipitation during ENSO-neutral years, but that this may be overridden by strong ENSO events. Here we attempt to explore links between subseasonal predictability over the contiguous United States (CONUS) domain and these teleconnection modes, with a focus on week 3–4 predictability.

2 Data and methods

The subseasonal predictability of four key meteorological variables, surface (2 m) temperature (T2m), total precipitation (Prcp), mean sea level pressure (SLP), and 500 hPa geopotential height (Z500), are evaluated for the European Centre for Medium-Range Weather Forecasts (ECMWF) Variable Resolution Ensemble Prediction System monthly forecast system (VarEPS-monthly; Vitart et al. 2008) and National Centers for Environmental Prediction (NCEP) Climate Forecast System, version 2 (CFSv2; Saha et al. 2010)

reforecasts prepared for the WWRP/WCRP S2S project database (Vitart et al. 2017). The atmospheric component of the ECMWF model (version CY41R1) has 91 vertical levels and a horizontal resolution of TCO639 (~ 16 km) up to day 10 and TCO319 (~ 32 km) after day 10. Semiweekly reforecasts of the ECMWF model over the 20-year (1995/6–2014/5) reforecast period are analyzed, corresponding to real-time forecast start dates every Monday and Thursday from June 2015 to May 2016 (105 reforecast cycles). For example, a real-time forecast is initialized on Monday June 1, 2015 (a Monday start date), and 20 reforecasts are initialized for all June 1's between 1995 and 2014 (which do not necessarily all fall on Mondays). The ECMWF reforecasts consist of one control and 10 perturbed forecast ensemble members on each start date. More model details are given in Vitart et al. (2017).

The resolution of the NCEP model is 64 vertical levels and T126 (~ 100 km) in the horizontal. The NCEP model provides reforecasts with a small ensemble size (4) that are made every day at 00/06/12/18Z. Lagged ensembles of NCEP reforecasts are thus composed using three daily starts on and prior to the date of each ECMWF semi-weekly reforecast start, to match the ensemble sizes (11 for ECMWF and 12 for NCEP). Taking the above June 1 example, the 12 NCEP reforecasts from May 30 to June 1 (3 days and 4 realizations per day) are used to generate a lagged ensemble mean to compare with the ECMWF ensemble mean reforecast of June 1. The lagged ensemble skill is usually positively influenced by increasing ensemble size and negatively influenced by the drop in longer lead skill (e.g., Chen et al. 2013). DelSole et al. (2017) showed that the lagged ensemble skill for the large-scale components of NCEP T2m saturates around 4 days. Therefore, the 3-day lagged ensemble seems a reasonable choice for all four variables to be analyzed here.

The Global Precipitation Climatology Project (GPCP) version 1.2 (Huffman and Bolvin 2012) daily precipitation estimates on a 1° grid are used as the observational precipitation dataset, and ERA-Interim reanalysis on a 1.5° grid (Dee et al. 2011) is used for observational estimates of T2m, SLP, and Z500. The common period for the two models and observations is 1999–2010 (12 years in total), constrained by the NCEP reforecast availability; the results for ECMWF over the longer 1995–2014 period are also included in some comparisons to illustrate temporal robustness. Since the model data are provided on a 1.5° grid (~ 150 km) in the S2S database, the total precipitation reforecast is interpolated from 1.5° into 1° resolution so as to compare with the GPCP dataset, whereas the other variables are validated with ERA-Interim reanalysis at the 1.5° resolution.

The skill metrics include the temporal correlation of anomalies (CORA) and spatial correlation of anomalous patterns (CORP), each computed using the ensemble means

of fortnight week 3–4 (i.e., day 15–28) averages. The probabilistic evaluation of the skill for these two models has been documented by Vigaud et al. (2017), whereas the deterministic skill is primarily used in this study to explore possible sources of predictability. The anomalies are calculated by removing the observational climatological mean, and the models' week 3–4 reforecast climatology, in order to exclude the mean bias and model drift. Note that the skill is not cross-validated as the main purpose is not to evaluate the forecast accuracy but to explore the subseasonal predictability of the physical system. The forecast skill is evaluated in four seasons separately: winter (DJF), spring (MAM), summer (JJA), and fall (SON), to examine possible seasonal dependence of subseasonal predictability (e.g., Becker and Van Den Dool 2016; Vigaud et al. 2017; DelSole et al. 2017). More specifically, ECMWF reforecasts from all the semi-weekly start dates that fall in each 3-month season are lumped together as a time series to compare with observed time series. For example, the JJA time series are formed as 06/01/1999, 06/04/1999, 06/08/1999, …, 08/24/1999, 08/27/1999, 08/31/1999, 06/01/2000, 06/04/2000, …, 08/27/2010, 08/31/2010. DJF and JJA have 27 start dates each, while MAM and SON have 26 each. The seasonal forecast skill is evaluated with start dates that fall in each season, whereas some of the week 3–4 forecasts may fall in the first month of the next season. For example, the DJF week 3–4 skill actually refers to the skill of the forecasts ranges from mid-December to late-March.

Caution needs to be taken in estimating the statistical significance of the CORAs. The actual number of degrees of freedom (DOF) for the week 3–4 CORA is much smaller than the total number of starts for each season in 12 years (324 or 312 starts, except for 264 starts in 11 years for DJF), because of overlaps between adjacent fortnight averages. The number of independent reforecasts can be estimated as the total number of days in 12 seasons divided by 14, which is about 77 (71 for DJF). The equivalent DOF for the week 3–4 CORA is thus 75 (69 for DJF). Any week 3–4 CORA value greater than 0.2 (0.3) is considered statistically significant at the 5% (1%) level by a one-tailed t-test (as only positive CORA is considered skillful). Since all the variables examined here are serially and spatially correlated, the above estimation might be overly simplified and may not apply to all regions. Therefore, a permutation method (as in DelSole et al. 2017) is used to determine the significance levels and to compare with the t-test based estimation. A permuted ensemble is constructed by shuffling the yearly blocks in order to account for subseasonal serial correlation in the data. For example, the reference time series is organized as a series of 1-year blocks as follows: 1999/2000/2001/2002/2003/2004/2005/2006/2007/2008/2009/2010 and each block consists of a sub time series with the same number of start dates (see also the JJA example shown above). One example

of the shuffled time series appears as 2007/2006/2008/2001/2010/2003/2009/2002/2004/1999/2000/2005. The above permutation or shuffling is repeated 10,000 times to form an empirical distribution of CORA and the 99(95)th percentile of the resulting samples then defines the 1(5)% significance threshold value for CORA (corresponding to the 99(95)% confidence interval). The 1% threshold for CORA over the CONUS is mostly between 0.2 and 0.3 for all variables and all seasons (not shown), which is smaller than the estimated value of 0.3 based on the t-test with the equivalent DOF. In the following, we thus make the conservative assumption that CORA values greater than or equal to 0.3 are statistically significant, while the permutation significance levels are indicated by stippling the figures.

In order to better capture large-scale features, the model and observed precipitation fields are both spatially smoothed in spectral space, using a Fourier transform and an exponential weighting function $\exp[-(k/40)^2]$, where $k = \mathrm{sqrt}(k_x^2 + k_y^2)$ is the total wave number, prior to calculating the CORA and CORP. The truncation wavenumber 40 is chosen to remove small-scale noise while retaining physically relevant scales, where the equivalent wavelength is about 1000 km in the tropics and 700 km in the mid-latitudes. The spatial smoothing does not significantly improve the overall significant threshold of CORA but rather mainly extract the large-scale pattern of the CORA for the total precipitation (not shown).

Teleconnection indices are calculated from ERA-Interim reanalysis and the two models to examine their predictability. The AO index is calculated by projecting the first EOF mode of the ERA-Interim northern hemisphere (20–90°N) SLP field onto the reanalysis and reforecast data. The NAO index is defined as the difference of SLP between the domains (50°W–10°E, 25–55°N) and (40°W–20°E, 55–85°N), following Wang et al. (2017), which is very close to other definitions and is convenient to use. The PNA index is computed using the NOAA Climate Prediction Center (CPC) modified pointwise definition, an update of Wallace and Gutzler (1981). The TNH mode was originally defined pointwise as the difference between (55–65°N, 80–100°W) and (50–60°N, 145–165°W) on the 700 hPa pressure level (Robertson and Ghil 1999) and is calculated using the 500 hPa geopotential height because of the availability of the data. A sensitivity test using ERA-Interim data shows very little difference between the TNH indices computed from the two levels, as their temporal correlation is about 0.97 for 1995–2014. The statistical significance of these teleconnection indices is also estimated using the two methods above and again the one-tailed t-test gives stricter thresholds: week 3–4 CORA value greater than 0.2 (0.3) is significant at the 5% (1%) level, while the permutation method gives slight lower thresholds.

3 Results

3.1 Temporal correlation

The week 3–4 precipitation CORA shows fairly pronounced seasonal variations over the CONUS for both models, with highest overall skill in winter (DJF) and poorest overall skill in fall (SON) for the ECMWF model and in spring (MAM) for the NCEP model (Fig. 1). The spatial distribution of CORA shows qualitative similarity between the two models except for slightly higher skill for the ECMWF model. The highest winter precipitation
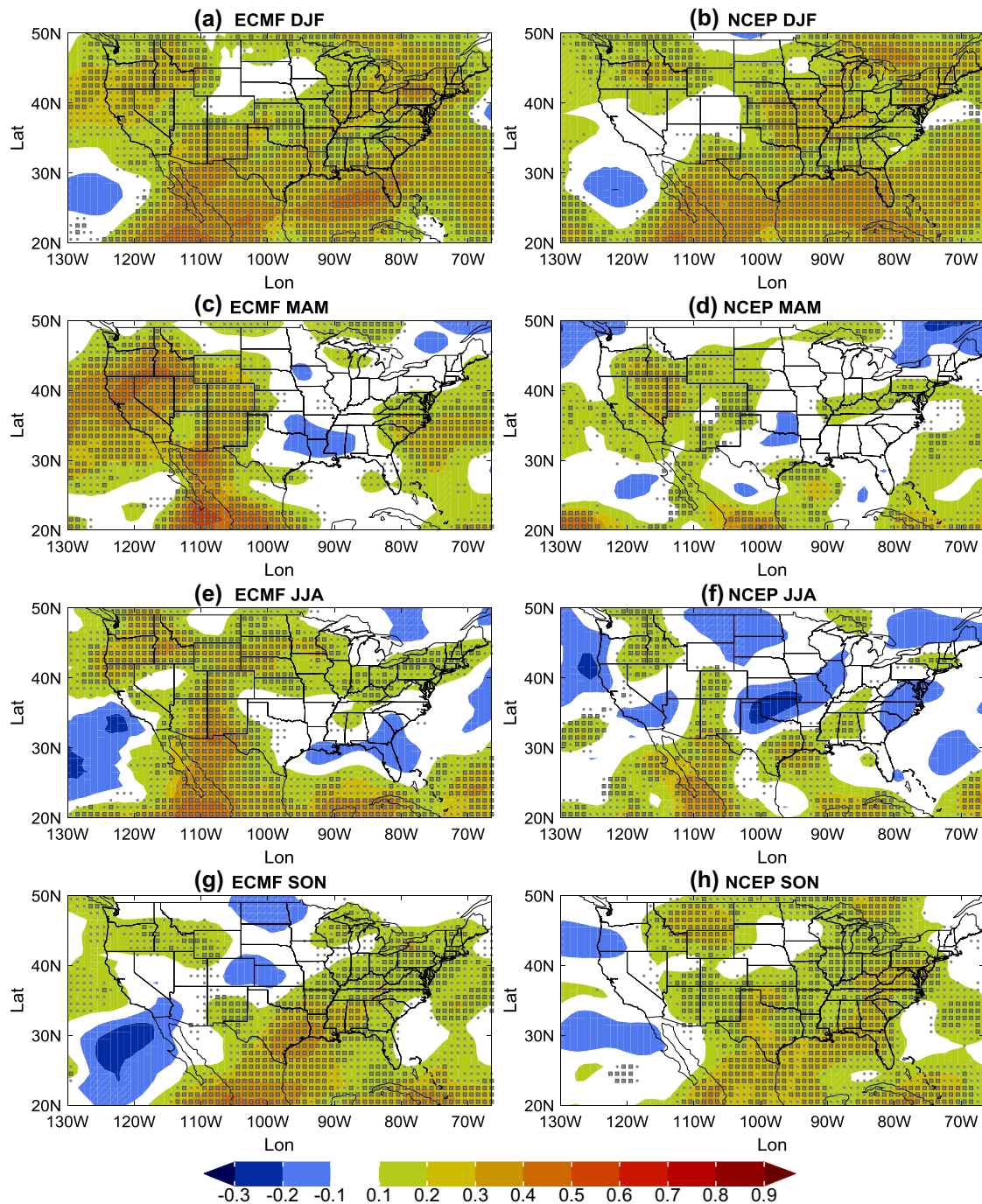
Fig. 1 Week 3–4 correlation of anomaly (CORA) skill maps for total precipitation, for ECMWF (left) and NCEP (right) models. Contour interval is 0.1 and any CORA marked with a dot/square symbol is significant at the 5/1% level based on a permutation method repeated for 10,000 times (hereafter for all CORA maps)
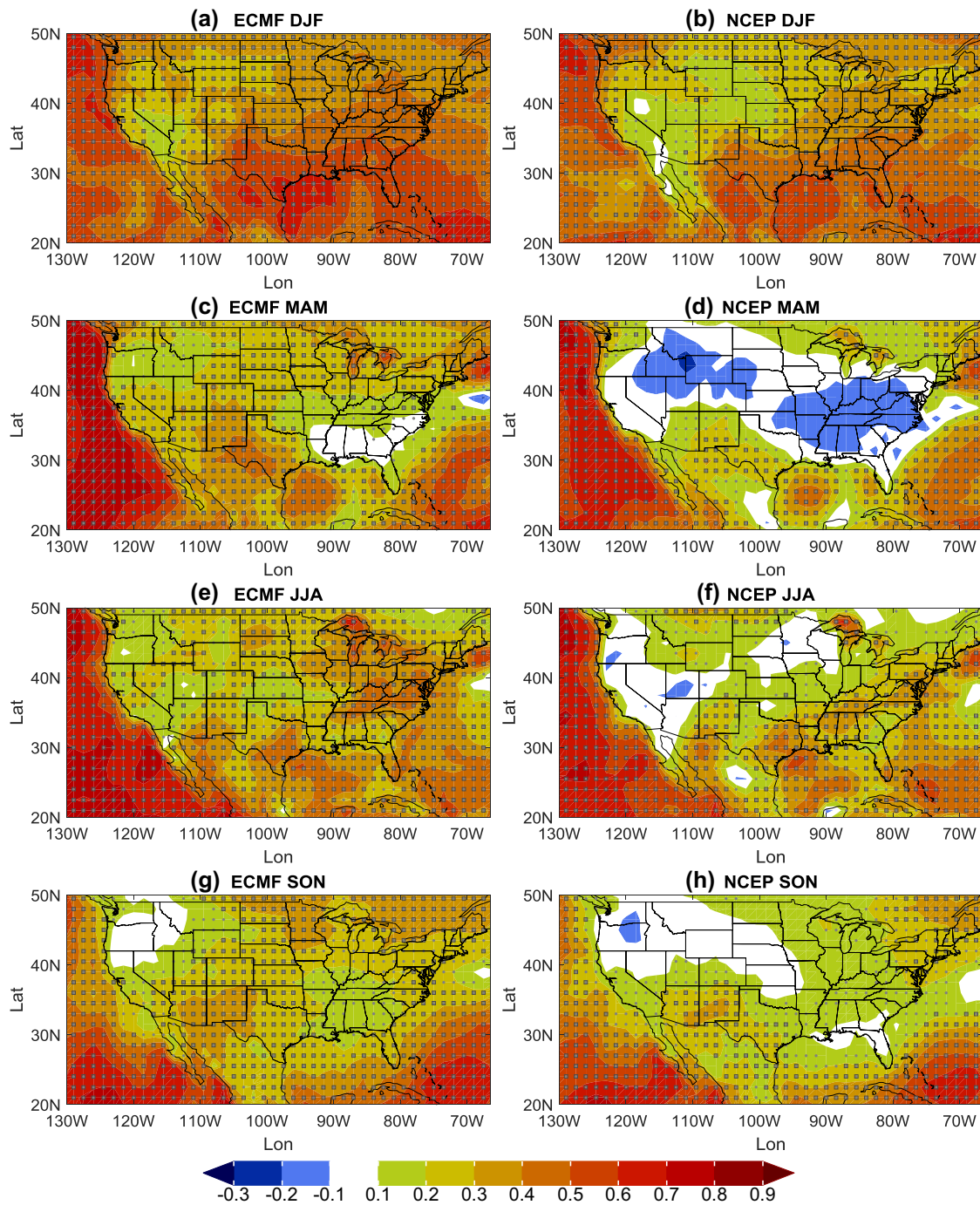
Fig. 2 Week 3–4 CORA skill maps for 2 m temperature

skill (CORA > 0.3 for ECMWF) in the US resides in the northeast US and Florida (Fig. 1a). The spring precipitation is difficult to predict in most regions of the US, especially for the NCEP model (Fig. 1d). However, the ECMWF model shows promising skill in the western US, around the borders between Idaho, Oregon, California and Nevada (Fig. 1c). This region of relatively high skill also appears in winter and summer, albeit with reduced skill

and spatial coverage (Fig. 1a, e). The southern part of Arizona and New Mexico shows relatively high skill that increases from winter to summer, shifting to SE Texas and Louisiana in fall (Fig. 1a, c, e, g), which corresponds with the seasonal evolution of the North American Monsoon (e.g., Higgins et al. 1997).

The CORA of week 3–4 T2m is higher than for precipitation and is generally larger in the eastern and extreme

southern US than in the western part for both models (Fig. 2). The spatial distribution of T2m CORA is in good agreement between the two models, while again the ECMWF model shows slightly higher skill. The T2m skill is much higher over the oceans (and often over the Great Lakes) than over the land, consistent with the contrast in heat content. Higher skill can be found in the Pacific basin in all seasons compared with that in the Atlantic basin, suggesting upstream influence from teleconnection modes such as the ENSO and PNA. This difference in skill may also be related to the different inherent persistence of ocean heat content due to differences in the mixed-layer depth in the coastal waters of these two ocean basins offshore of the United States (e.g., Deser et al. 2003). Both models exhibit highest skill in winter (DJF) and lowest skill in transitional seasons (MAM and SON). Similar to the precipitation, T2m forecast skill also suggests the influence of North American Monsoon on temperatures and shows a local maximum in southern Arizona and New Mexico in summer (Fig. 2e).

The above large-scale spatial structures of near-surface temperature and precipitation skill are mostly dynamically driven as they are generally consistent with the Z500 CORA maps in the corresponding seasons (Fig. 3). For example, the winter Z500 CORA maps show a southeast-northwest dipole structure over the US with high values (> 0.4) in the southern and eastern US (Fig. 3a, b), while the winter Prcp and T2m CORA maps show a similar structure to first order (Figs. 1a, b, 2a, b). In the other seasons (spring to fall), most regions of the US exhibit relatively low skill (Z500 CORA < 0.3) except for a tongue of high skill in the North American Monsoon regions extended from the tropics where high skill associated with tropical moisture surges and associated evaporative and cloud-induced cooling (e.g., Hales 1972; Adams and Comrie 1997). The NCEP winter and summer large-scale spatial structures (Fig. 2a, b, e, f) are consistent with those extracted using Laplacian eigenvectors by DelSole et al. (2017), albeit with different sampling methods. The very low skill of the NCEP model in MAM stands out in both the Z500 and T2m fields.

3.2 Sources of week 3–4 predictability

Since the high skill regions are usually of large spatial scale, the skill may be associated with large-scale low-frequency modes of variability in the atmosphere. At this lead time, the climate and variability of model forecasts is quite similar to that in the observations, where the models' teleconnection modes and their connection with the variables studied here are faithfully reproduced (not shown). As a result, the high skill regions in the model variables (Prcp/T2m/SLP/Z500) often show relatively high correlations between the variables and teleconnection indices in observations. For example, the observed winter precipitation in the southern and

eastern US is well correlated with the Nino3.4, NAO, and PNA indices (Fig. 4), where the modeled winter precipitation skill is relatively high (Fig. 1a). In particular, the high skill over Florida is an extension of high skill over the Gulf of Mexico that is primarily driven by these three teleconnection modes. The high skill of the winter temperature in the models is also consistent with high correlations between the observed temperature and the Nino3.4, NAO, and PNA indices in the southern and eastern US (Fig. 5). And again, the high skill in these surface variables is dynamically consistent with the high skill in Z500 in the southern and eastern US where the observed correlations are high (Fig. 6). Note that these teleconnection modes are not entirely independent, for example, both PNA and NAO are believed to be strongly influenced by ENSO (e.g., Straus and Shukla 2002; Moron and Gouirand 2003).

Skill levels in precipitation are relatively lower than those in temperature, and geopotential height (Figs. 1 vs. 2, 3). An exception is the high week 3–4 skill over the northwest US, in the region of the borders between California, Nevada, Oregon, and Idaho in spring (MAM; Fig. 7a), where T2m and Z500 skill is relatively low (Fig. 7c, e). Among the Nino3.4, AO, NAO, and PNA indices, only the AO index shows significant correlation with the total precipitation in observations, but poor correlation with T2m and Z500 (Fig. 7b, d, f). The AO index is also well correlated with SLP in this region (Fig. 7h) where the ECMWF model shows fairly good skill in SLP (Fig. 7g). Therefore, the high skill of precipitation and SLP in this region is likely associated with the AO (it is shown in Fig. 9; Table 1 below that the modeled AO skill is also high in this season). Interestingly, the precipitation skill over the northwest US peaks in spring for both models (Fig. 8), unlike the skill of other variables that usually peaks in winter (Figs. 1, 2, 3). The subseasonal component of the skill, computed by subtracting the seasonal average of each year separately before calculating the CORA, is less than half of the total in spring but dominates in the other seasons (Fig. 8a). Less than 1/3 of the total variance is explained by the subseasonal component in MAM (Fig. 8b). These features also hold true for a longer period, i.e., 1995–2014 for the ECMWF model (grey bars in Fig. 8), and indicates a sizeable interannual component of the week 3–4 skill in spring precipitation over the western US associated with the AO. This is consistent with a previously reported good correlation between the JFM-mean AO and the FMA-mean precipitation over the western US (McAfee and Russell 2008).

A natural follow-up question to ask is how well the models predict the teleconnection modes themselves, including the AO, NAO, PNA, and TNH. The TNH is included here because it has been argued to represent the ENSO response in the NH extratropics better than the PNA, whose timescales are more intraseasonal (Barnston and Livezey 1987;
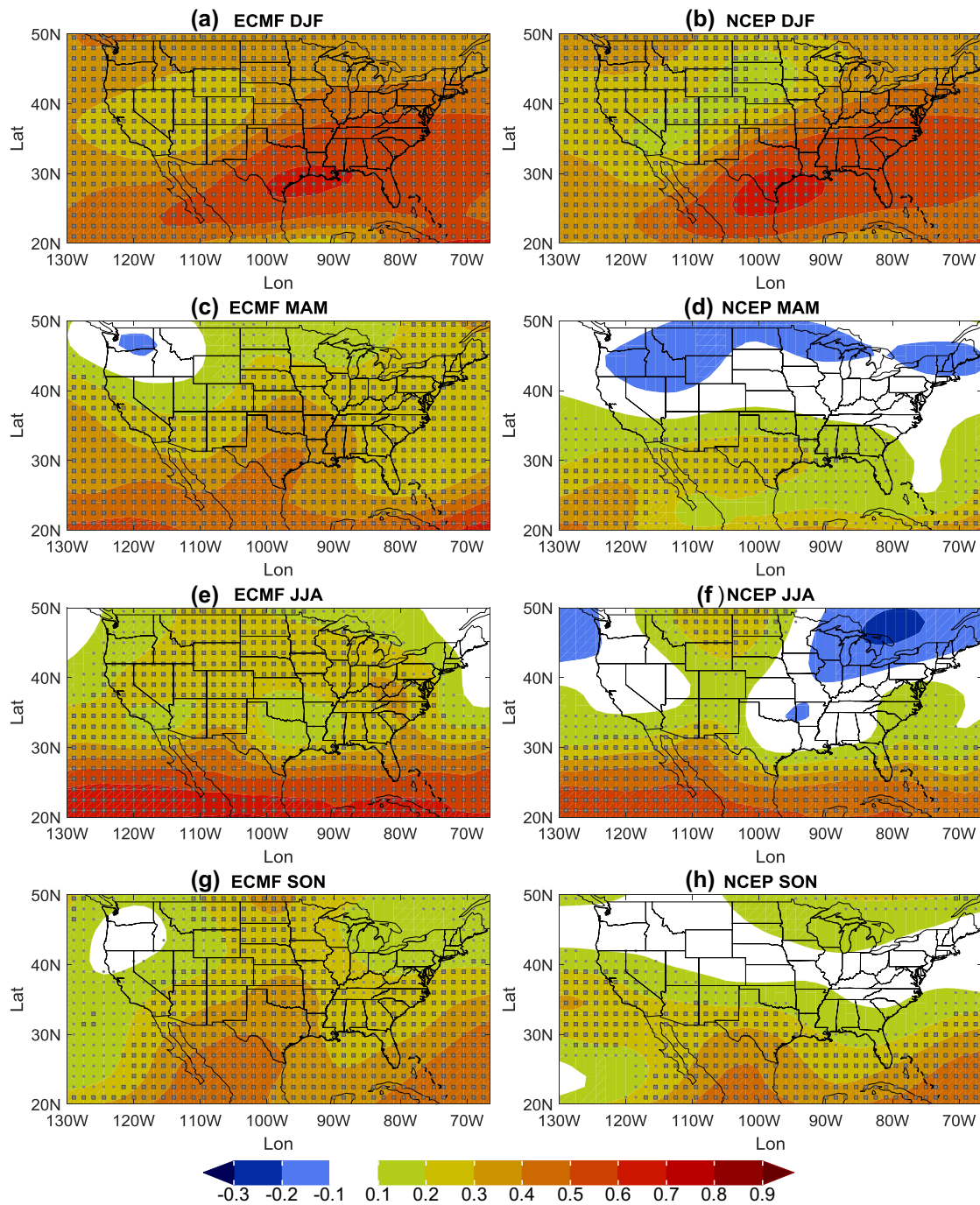
Fig. 3 Week 3–4 CORA skill maps for 500 hPa geopotential height

Robertson and Ghil 1999). The week 1 forecast skill is very high (about 0.9 or higher), for all four indices and for both models in all seasons, with decreases in week 2 and beyond that are larger in the transition and summer seasons. However, it is remarkable that the week 3–4 fortnight-average skill mostly exceeds 0.4 in winter in both models (Table 1), as this skill is higher than the 1% significance threshold value and the persistence skill (generally between 0.2 and
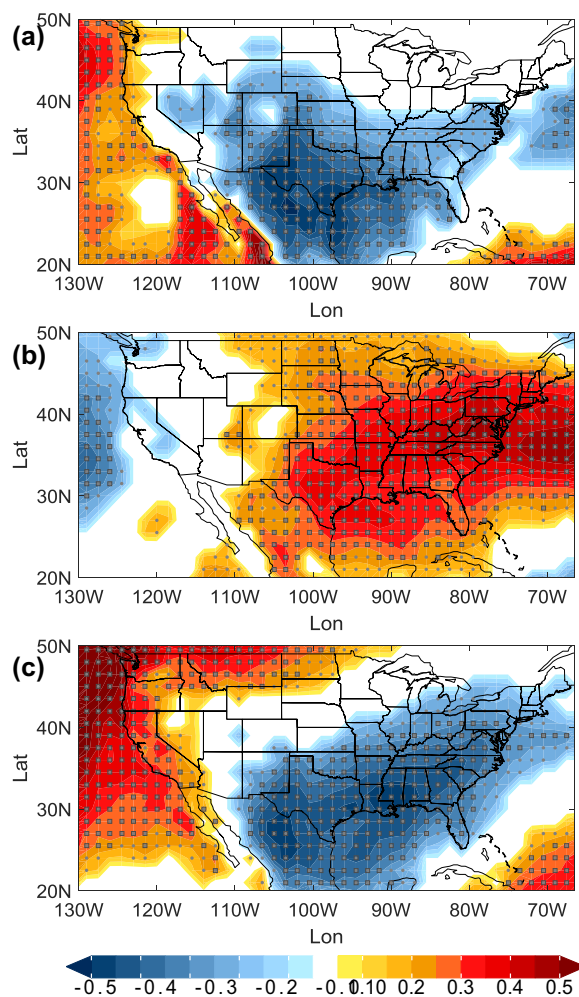
0.3 for different seasons; not shown). All four indices show highest week 3–4 skill in winter and lowest in summer (Fig. 9a). In general, Table 1 indicates that the skill in the week 3–4 fortnight average of the teleconnection indices derives from the week 3, but with week 4 skill significant in winter. Using earlier versions of the ECMWF and NCEP models, Johansson (2007) found only marginal skill in NAO and PNA at week 3 and week 4 (< 0.3). Younas and Tang
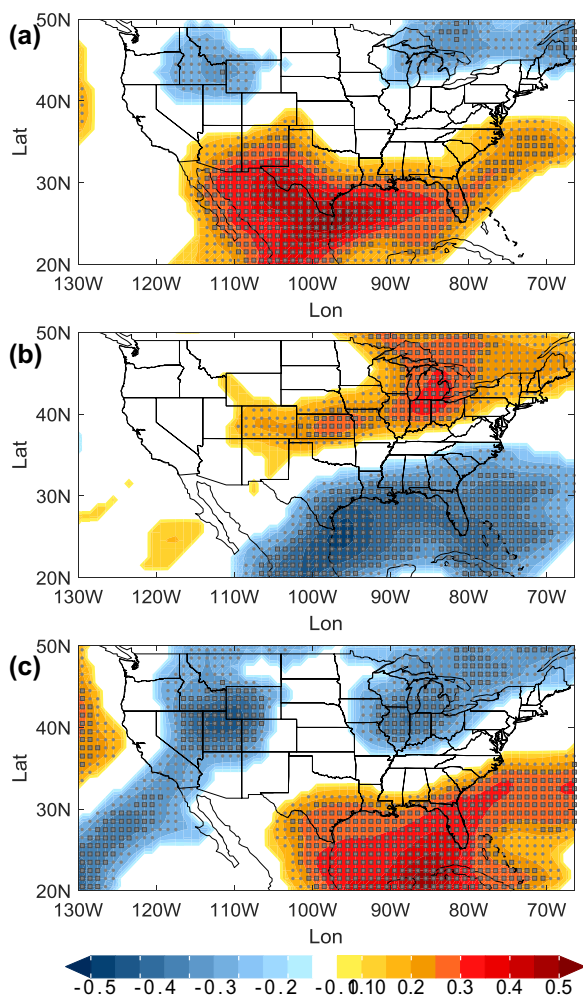
Fig. 4 Observational correlation maps between DJF dekadal (10-day mean) GPCP total precipitation and **a** Nino3.4 index, **b** NAO index, and **c** PNA index. Only the correlations significant at the 10% level by a two-tailed *t*-test are shown and any correlation marked with a dot/square symbol is significant at the 5/1% level (hereafter for all dekadal correlation maps)

Fig. 5 ERA-Interim correlation maps between DJF dekadal 2 m temperature and **a** Nino3.4 index, **b** NAO index, and **c** PNA index

(2013) showed that the PNA week 3–4 forecast skill is fairly modest (about 0.4 in CORA) in the ensemble mean of a suite of state-of-the-art seasonal forecast models but did not separate the seasons due to limited samples. The ECMWF and NCEP models exhibit similar week 3–4 skill in most cases except for higher fall NAO and spring PNA skill for the ECMWF model (Fig. 9).

Removing the seasonal contribution to the skill (by calculating CORA with respect to the seasonal averages of individual years), we find that the high winter skill for the AO and NAO indices contains an important *seasonal* component that contributes about 2/3 of the total variance explained (Table 2a, b vs. 1a, b; Fig. 9b). This is consistent with recent model findings of significant seasonal forecast skill of winter NAO and AO (e.g., Scaife et al.

2014; Dunstone et al. 2016; Wang et al. 2017). In contrast, the PNA winter skill mostly comes from its subseasonal components (Table 2c vs. 1c; Fig. 9b). Perhaps surprisingly, about half of the variance explained by the TNH winter skill can also be attributed to subseasonal scales (Table 2d vs. 1d; Fig. 9b). The contrast between the AO/NAO and PNA/TNH predictability indicates that the seasonal time scale sources of week 3–4 predictability, such as ENSO and the Quasi-Biennial Oscillation (QBO), are more substantial in the Atlantic sector in winter, than over the Pacific. On the other hand, in the transition and summer seasons the skill for all four indices is dominated by the subseasonal components. These conclusions are fairly robust against the period tested, as the longer period ECMWF 1995–2014 skill shows very similar attribution between seasonal and subseasonal components (bracketed numbers in Tables 1, 2; grey bars in Fig. 9).
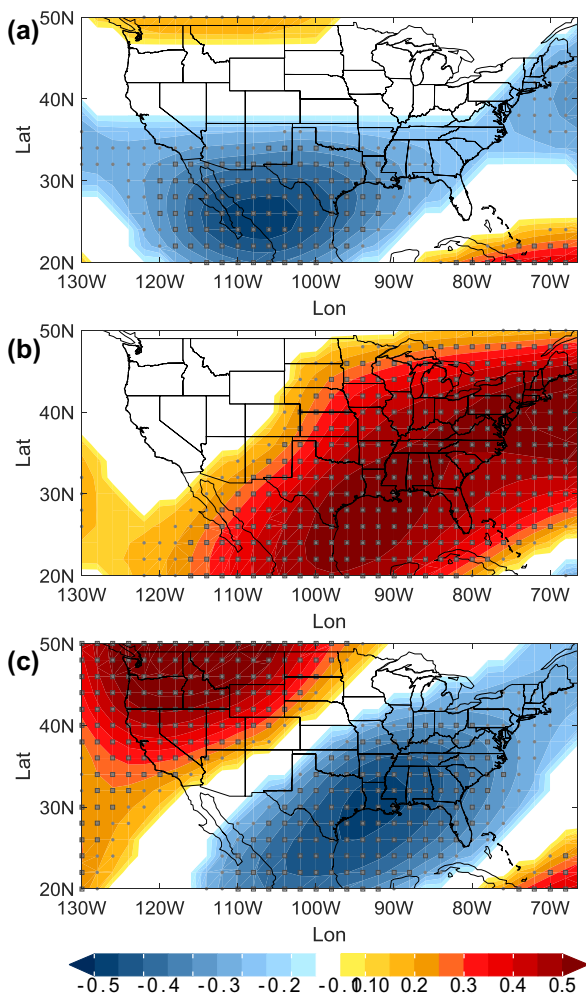
Fig. 6 ERA-Interim correlation maps between DJF dekadal 500 hPa geopotential height and **a** Nino3.4 index, **b** NAO index, and **c** PNA index

3.3 Pattern correlation

Here we address subseasonal predictability in terms of pattern correlations (CORP) between the fields of observed and forecasted week 3–4 averaged anomalies. This is performed for the US domain (70–130°W, 20–50°N) to investigate the role of ENSO teleconnections in week 3–4 forecast skill (Fig. 10), before examining the role of the atmospheric teleconnections in Fig. 11. Composites of monthly-mean week 3–4 CORP from the semi-weekly reforecasts for the negative (Nino3.4 index < −1), neutral (between −0.5 and 0.5), and positive (>1) phase of ENSO are plotted in Fig. 10 for each variable. As in the CORA, the ECMWF model (blue bars) generally has slightly higher CORP than NCEP (red bars). The composites on the negative and positive ENSO phases are visibly higher than that on the neutral phase for all four variables, especially for the surface temperature skill where both models show statistically significant differences

between neutral and extreme phases (Fig. 10b). The contrast between the negative and neutral phases is statistically significant for both models and all four variables, whereas the contrast between the positive and neutral phases is much weaker. The small number and relatively weak amplitude of El Niño events in the 1999–2010 period may be partly responsible for this asymmetry between ENSO phases. Analogous histograms of pattern correlations calculated for extreme phases of the AO/NAO/PNA/TNH yield less clear results (not shown). Nevertheless, relatively high skill in the US domain can be found in particular years during extreme phases of these teleconnection modes (see the highlighted periods in Fig. 11). For example, 2009/2010 winter was a particularly extreme case for all indices: Nino3.4 SST ≈ 1, AO $\approx −5$, NAO $\approx −3$, and PNA/TNH ≈ 4 (standardized values), when the T2m CORP was also quite high for the US domain. The high CORP during 1997/1998 winter also corresponds to high values in the Nino3.4 and PNA/TNH indices. On the other hand, the high CORP during 2006/2007 winter is primarily accompanied with extreme AO/NAO phases. The predictability is usually low when teleconnection modes are quiet (in neutral phases; e.g., 2003/2004 winter).

4 Summary and discussion

The week 3–4 forecast skill of surface temperature and precipitation is evaluated in the domain of the United States for the ECMWF VarEPS and NCEP CFSv2 global ensemble prediction systems for the period of 1999–2010, firstly in terms of temporal correlation of anomalies (CORA) for standard meteorological seasons (Figs. 1, 2). The ECMWF model shows slightly higher skill in general, which might be due to finer horizontal resolution than the NCEP model and improved physical parametrizations from earlier versions (Wheeler et al. 2017). The difference in skill might also be a result of the lead-time difference inherent in the lagged ensemble construction for the NCEP model. Nevertheless, the large-scale patterns of the anomaly correlation are fairly similar between these two systems, especially for relatively high skill regions over the northeast US and Florida in DJF, the western US in MAM precipitation, and the southwest in JJA. Most of these high skill regions are associated with the high skill of mid-tropospheric (500 hPa) geopotential height (Fig. 3), except in spring over the North American Monsoon regions which are more influenced by tropical air surges (e.g., Hales 1972; Adams and Comrie 1997). It is worth noting that the NCEP model achieves comparable skill to the ECMWF model with a three times coarser resolution.

The high skills in the surface weather and atmospheric circulation are shown to be qualitatively attributable to ENSO, NAO, and PNA teleconnections in winter over the
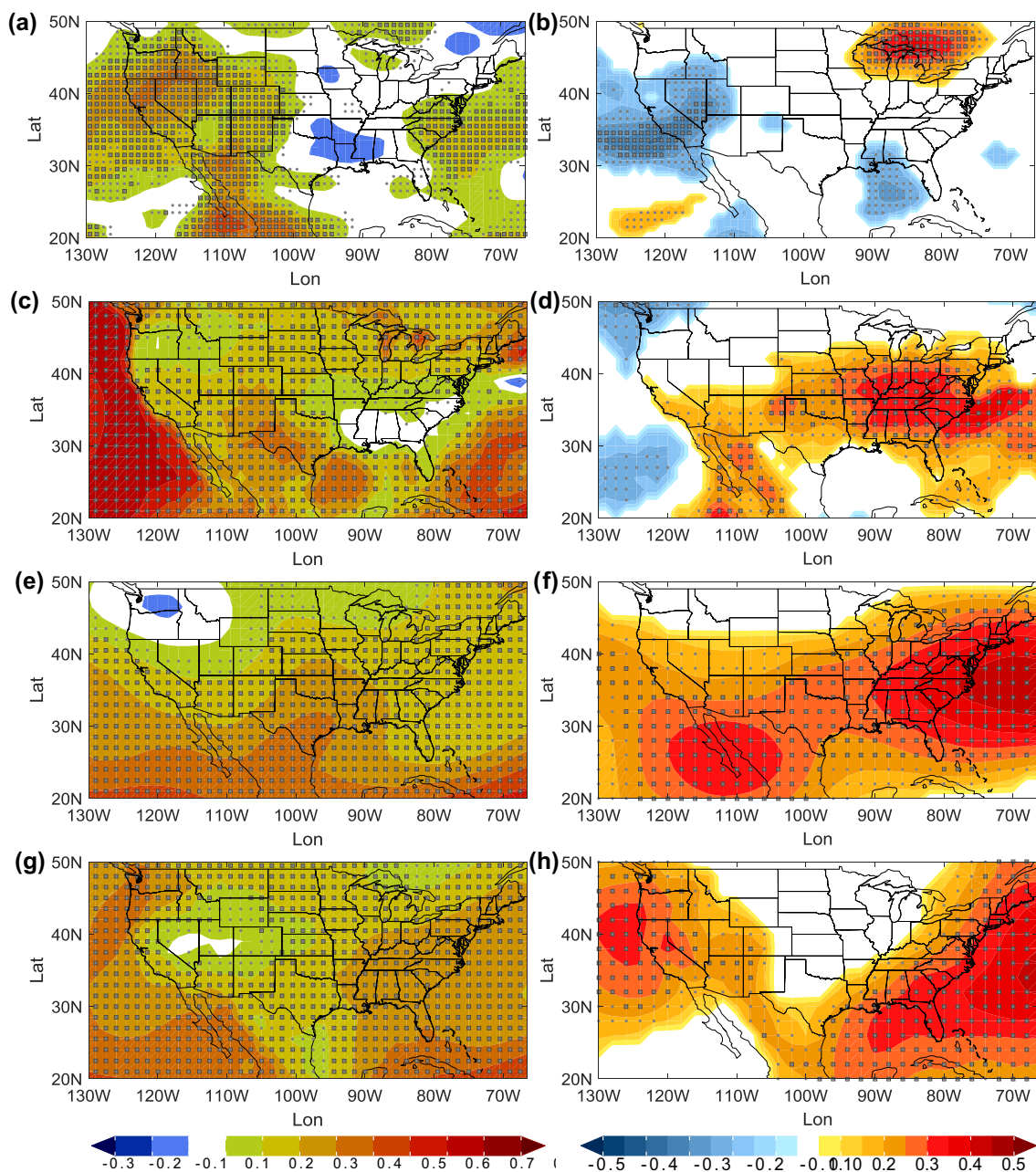
Fig. 7 Maps of **a** MAM CORA skill for ECMWF week 3–4 total precipitation, and **b** MAM correlation maps between dekadal GPCP total precipitation and observed AO index. Same as **a, b** but for **c, d** 2 m temperature, **e, f** 500 hPa geopotential height, and **g, h** sea level pressure

southern and eastern US (Figs. 4, 5, 6), and to the AO in spring over the western US (Fig. 7) which appears to be dominated by the interannual component of variability (Fig. 8). The two ensemble prediction systems both show good week 3–4 forecast skill in predicting the teleconnection modes themselves (Fig. 9), especially in winter. This skill is found to be dominated by subseasonal timescale variations, except for the winter AO and NAO. By inference, this indicates an important seasonal time scale contribution to week 3–4 forecast skill in the Atlantic sector in winter.

Nevertheless, the relatively short periods (12 years and 20 years) considered here, are not sufficient to sample the interannual and decadal variations of these teleconnection modes (e.g. Weishemeir et al. 2017; O'Reilly et al. 2017). Consequently, caution is required in extrapolating the seasonal skill estimated for these periods to other periods or to the future.

These results underscore the overlap of intraseasonal and interannual sources of predictability in the week 3–4 time range, leading to the prospect of enhanced predictability
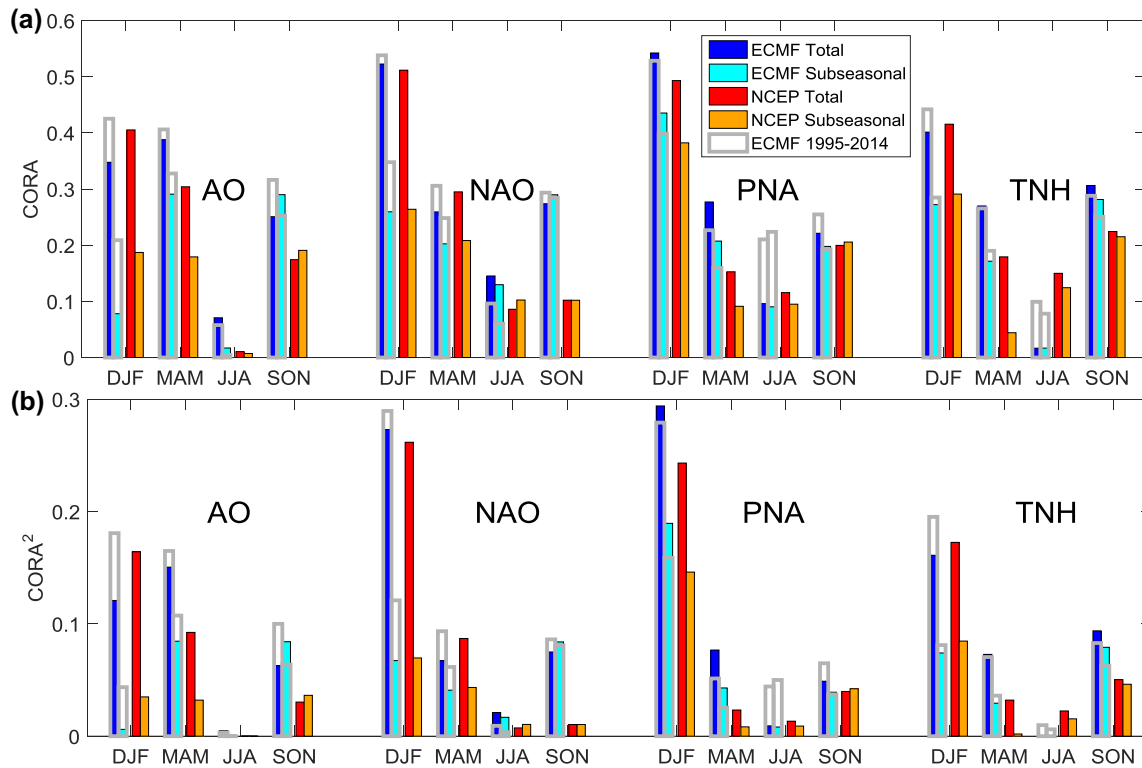
Table 1 CORA for 1999–2010 (1995–2014) (a) AO index, (b) NAO index, (c) PNA index, (d) TNH index

Lead (week)	ECMWF					NCEP				
	1	2	3	4	3&4	1	2	3	4	3&4
(a) AO index										
DJF	**0.93 (0.96)**	**0.68 (0.71)**	**0.34 (0.40)**	0.25 (**0.33**)	**0.35 (0.43)**	**0.90**	**0.61**	**0.41**	0.28	**0.41**
MAM	**0.94 (0.96)**	**0.59 (0.65)**	**0.40 (0.39)**	0.27 (**0.30**)	**0.39 (0.41)**	**0.89**	**0.50**	**0.35**	0.12	**0.30**
JJA	**0.91 (0.93)**	**0.43 (0.48)**	0.12 (0.12)	0.02 (−0.01)	0.07 (0.06)	**0.86**	**0.40**	0.13	−0.08	0.01
SON	**0.92 (0.95)**	**0.55 (0.55)**	0.27 (**0.30**)	0.16 (0.23)	0.25 (**0.32**)	**0.87**	**0.41**	0.14	0.17	0.17
(b) NAO index										
DJF	**0.94 (0.96)**	**0.71 (0.74)**	**0.52 (0.53)**	**0.43 (0.44)**	**0.52 (0.54)**	**0.93**	**0.67**	**0.51**	**0.37**	**0.51**
MAM	**0.94 (0.95)**	**0.55 (0.61)**	**0.31 (0.33)**	0.14 (0.20)	0.26 (**0.31**)	**0.91**	**0.45**	**0.35**	0.05	0.29
JJA	**0.91 (0.93)**	**0.47 (0.51)**	0.20 (0.17)	0.09 (0.03)	0.15 (0.10)	**0.87**	**0.38**	0.11	0.06	0.09
SON	**0.93 (0.95)**	**0.54 (0.56)**	**0.31 (0.32)**	0.14 (0.15)	0.27 (0.29)	**0.86**	**0.38**	0.14	0.04	0.10
(c) PNA index										
DJF	**0.98 (0.97)**	**0.86 (0.83)**	**0.60 (0.57)**	**0.39 (0.41)**	**0.54 (0.53)**	**0.96**	**0.77**	**0.55**	**0.36**	**0.49**
MAM	**0.96 (0.96)**	**0.66 (0.66)**	**0.33 (0.31)**	0.16 (0.08)	0.28 (0.23)	**0.92**	**0.53**	0.21	0.07	0.15
JJA	**0.96 (0.95)**	**0.55 (0.55)**	0.05 (0.18)	0.12 (0.18)	0.10 (0.21)	**0.90**	**0.45**	0.19	0.07	0.12
SON	**0.96 (0.95)**	**0.64 (0.63)**	0.26 (**0.31**)	0.12 (0.14)	0.22 (0.25)	**0.90**	**0.49**	0.22	0.15	0.20
(d) TNH index										
DJF	**0.96 (0.96)**	**0.74 (0.74)**	**0.47 (0.46)**	0.23 (0.29)	**0.40 (0.44)**	**0.92**	**0.66**	**0.45**	0.26	**0.42**
MAM	**0.96 (0.95)**	**0.61 (0.61)**	**0.32 (0.28)**	0.13 (0.13)	0.27 (0.27)	**0.93**	**0.56**	0.29	−0.01	0.18
JJA	**0.92 (0.92)**	**0.50 (0.47)**	0.08 (0.13)	0.01 (0.04)	0.02 (0.10)	**0.84**	**0.30**	0.16	0.08	0.15
SON	**0.94 (0.94)**	**0.62 (0.62)**	**0.34 (0.32)**	0.21 (0.20)	**0.31 (0.29)**	**0.88**	**0.54**	0.27	0.08	0.22

Bold numbers (CORA > 0.3) are statistically significant at the 99% confidence interval by a one-tailed t-test, while CORA > 0.2 corresponds to the 95% interval (hereafter for all tables)

Fig. 8 **a** 110–125°W, 35–45°N mean week 3–4 CORA skill and its subseasonal component for 1999–2010 ECMWF and NCEP total precipitation. **b** Same as **a** but for the corresponding squared CORA, which is proportional to the variances explained by each component. The 1995–2014 ECMWF counterparts are also included for comparison. The horizontal line in **a** represents the 95% confidence interval by a one-tailed *t*-test
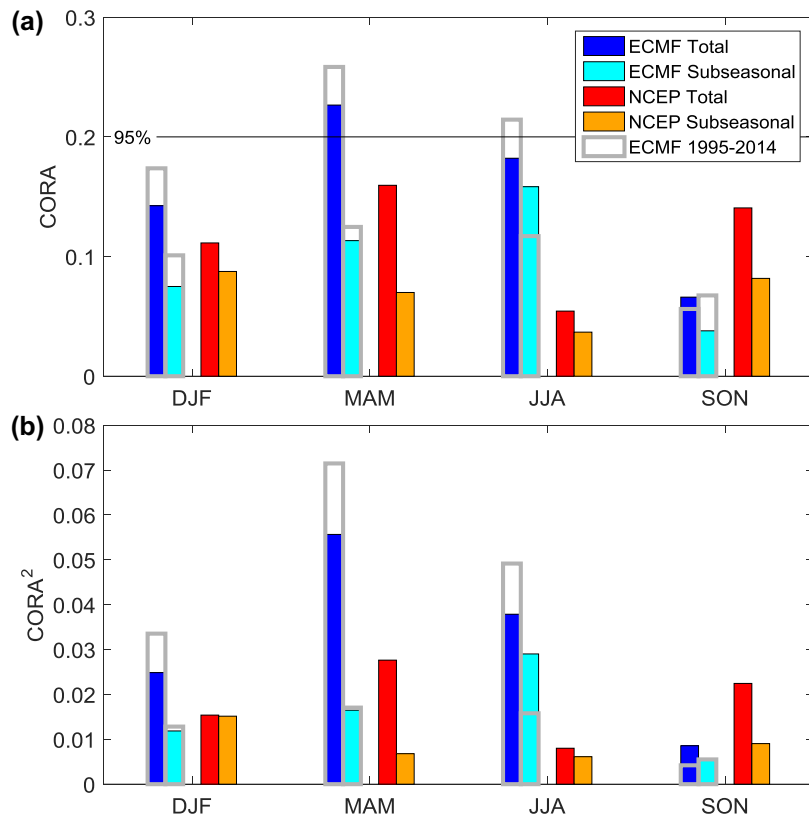
Fig. 9 **a** Week 3–4 CORA skill and the subseasonal components for ECMWF and NCEP AO, NAO, and PNA indices. **b** Same as **a** but for the corresponding squared CORA. Any CORA greater than 0.3 (0.2) is statistically significant at the 99% (95%) confidence interval by a one-tailed *t*-test

Table 2 Subseasonal CORA for 1999–2010 (1995–2014) (a) AO index, (b) NAO index, (c) PNA index, (d) TNH index

Lead (week)	ECMWF					NCEP				
	1	2	3	4	3&4	1	2	3	4	3&4
(a) AO index										
DJF	**0.91 (0.95)**	**0.57 (0.63)**	0.12 (0.22)	0.02 (0.14)	0.08 (0.21)	**0.86**	**0.48**	0.23	0.10	0.19
MAM	**0.93 (0.95)**	**0.51 (0.57)**	**0.31 (0.31)**	0.20 (0.25)	0.29 (**0.33**)	**0.87**	**0.40**	0.25	0.02	0.18
JJA	**0.89 (0.92)**	**0.38 (0.41)**	0.05 (0.05)	0.01 (−0.03)	0.02 (0.00)	**0.84**	**0.36**	0.11	−0.04	0.01
SON	**0.92 (0.94)**	**0.55 (0.50)**	0.29 (0.24)	0.18 (0.17)	0.29 (0.25)	**0.86**	**0.41**	0.15	0.18	0.19
(b) NAO index										
DJF	**0.91 (0.95)**	**0.58 (0.63)**	**0.30 (0.35)**	0.20 (0.28)	0.26 (**0.35**)	**0.89**	**0.52**	**0.30**	0.14	0.26
MAM	**0.93 (0.95)**	**0.50 (0.57)**	0.27 (0.38)	0.11 (0.15)	0.20 (0.25)	**0.90**	**0.40**	**0.30**	−0.05	0.21
JJA	**0.91 (0.93)**	**0.47 (0.47)**	0.19 (0.13)	0.08 (0.02)	0.13 (0.06)	**0.86**	**0.38**	0.13	0.06	0.10
SON	**0.93 (0.94)**	**0.54 (0.55)**	**0.32 (0.32)**	0.14 (0.13)	0.29 (0.28)	**0.85**	**0.34**	0.12	0.05	0.10
(c) PNA index										
DJF	**0.98 (0.97)**	**0.83 (0.81)**	**0.54 (0.50)**	0.26 (0.27)	**0.44 (0.40)**	**0.95**	**0.75**	**0.49**	0.21	**0.38**
MAM	**0.96 (0.95)**	**0.62 (0.61)**	0.28 (0.25)	0.12 (0.06)	0.21 (0.16)	**0.90**	**0.47**	0.15	0.05	0.09
JJA	**0.95 (0.95)**	**0.52 (0.54)**	0.06 (0.20)	0.11 (0.17)	0.09 (0.22)	**0.89**	**0.41**	0.18	0.06	0.09
SON	**0.96 (0.95)**	**0.64 (0.61)**	0.25 (0.27)	0.10 (0.10)	0.20 (0.19)	**0.90**	**0.49**	0.23	0.14	0.21
(d) TNH index										
DJF	**0.96 (0.95)**	**0.72 (0.71)**	**0.40 (0.37)**	0.11 (0.12)	0.27 (0.28)	**0.91**	**0.62**	**0.37**	0.15	0.29
MAM	**0.94 (0.94)**	**0.54 (0.54)**	0.23 (0.21)	0.07 (0.08)	0.17 (0.19)	**0.90**	**0.49**	0.23	−0.15	0.04
JJA	**0.91 (0.92)**	**0.47 (0.45)**	0.05 (0.11)	0.04 (0.03)	0.02 (0.08)	**0.83**	0.27	0.12	0.09	0.12
SON	**0.93 (0.94)**	**0.61 (0.60)**	**0.34 (0.30)**	0.18 (0.17)	0.28 (0.25)	**0.87**	**0.52**	0.26	0.08	0.22

Fig. 10 Composite of monthly-mean CORP on the negative (Nino3.4 index < − 1; 28 sample months), neutral (between − 0.5 and 0.5; 45 sample months), and positive (> 1; 24 sample months) ENSO phases for ECMWF (blue) and NCEP (red) models over the US domain (70–130°W, 20–50°N) for 1999–2010 reforecasts of (**a**) total precipitation, **b** 2 m temperature, **c** 500 hPa geopotential height, and **d** sea level pressure. The ECMWF 1995–2014 results are marked with grey edges. The errorbars represent the standard error (spread) of the composites
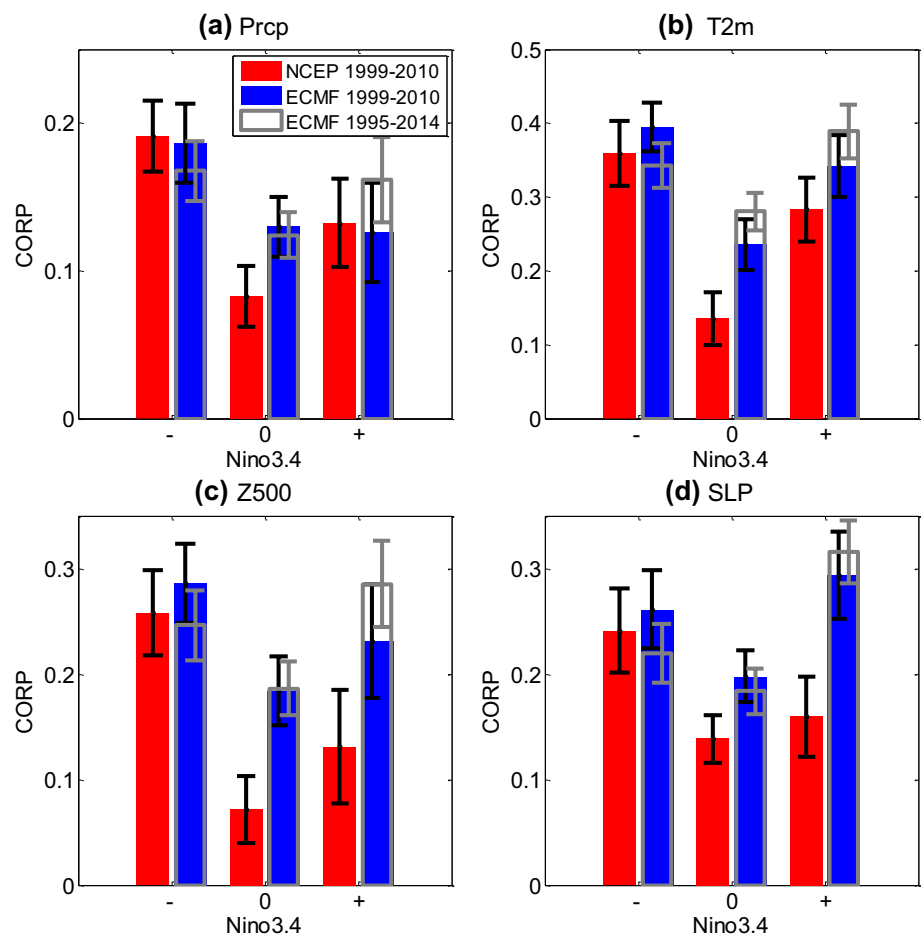

during certain episodes when multiple teleconnection modes are strong and interfere constructively. For example, high skill in temperature in the 2009/2010 winter is attributable to large excursions in both Atlantic and Pacific teleconnection indices, while the high temperature skill in 1997/1998 and 2006/2007 was mostly attributable to the Pacific and Atlantic teleconnections respectively (Fig. 11). Wang and Chen (2010) illustrated the role of AO in driving the cold temperature in the northern extratropics in 2009/2010 winter. On the other hand, Lin (2014) found significant influences of the PNA on the North American surface temperature in winter.

While week 3–4 predictability in precipitation and temperature over the US is not large, the results in this paper demonstrate that it is robust between two different ensemble prediction systems and that its spatial structure and seasonality can be interpreted physically in terms of well-known teleconnection patterns. The highly seasonal nature of midlatitude teleconnections, which have largest amplitudes in winter (e.g., Hurrell et al. 2003), helps explain the seasonality of precipitation and temperature skill which also peaks in winter. An interesting exception is found in spring, where precipitation (but not temperature) has a predictable component seen in both models in the northwestern US. This

appears to be associated with interannual variability of the AO and a possible physical interpretation is that the storm track movement driven by the AO variability may influence the spring onset in the western US (McAfee and Russell 2008). A more recent study revealed impacts of the Arctic stratospheric ozone variability on the April precipitation in this region through a series of feedbacks that involve the winter AO variability (Ma et al. 2018), which may explain the seasonal component of the skill. Nevertheless, there exist shifts in teleconnection modes and their correlation with precipitation over decadal or longer time scales (e.g., Cole and Cook 1998; Cook et al. 2011). A further exception is found in summer over the southwest monsoon region, seen most clearly in the ECMWF model, but also visible in the NCEP model.

Returning to the question of the sources of week 3–4 predictability beyond ENSO, the PNA is notable for the large subseasonal fraction of its skill, and how this clearly contrasts with the NAO where the interannual fraction dominates. Recent studies have found that interannual NAO predictability is fairly high in winter, which arises primarily from the variability of Arctic sea ice, Atlantic ocean heat content, and stratospheric circulation that might further
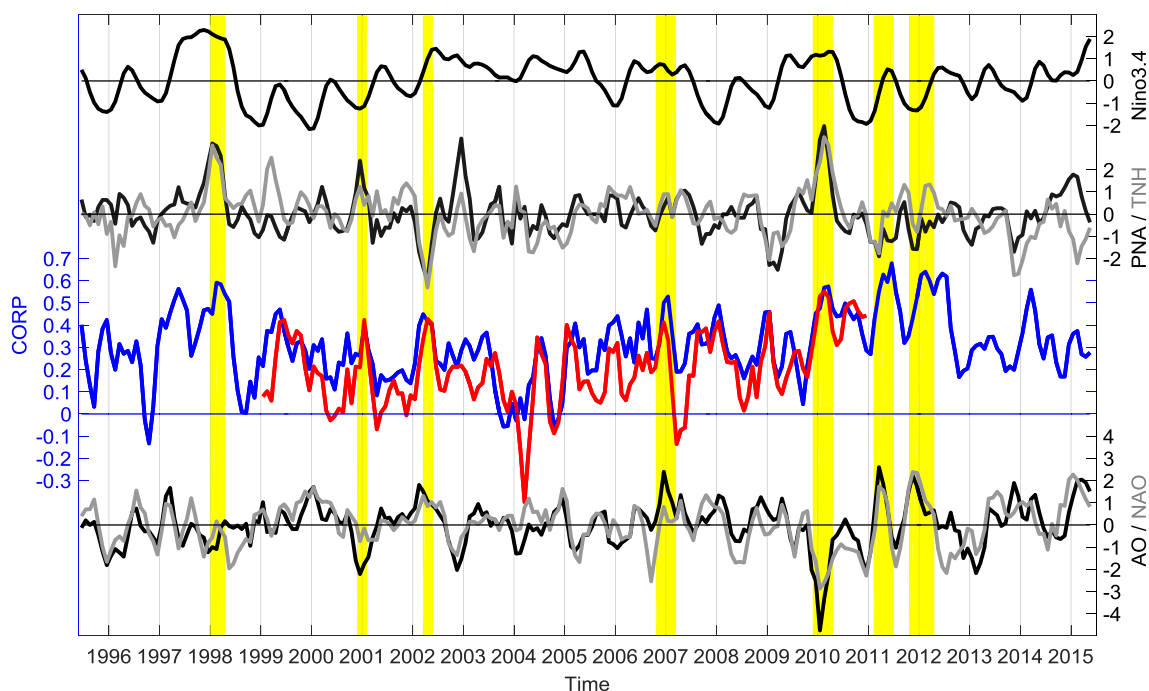
Fig. 11 Seasonal-mean ECMWF (blue) and NCEP (red) T2m Week 3–4 CORP of the US domain (70–130°W, 20–50°N), compared with the standardized Nino3.4/PNA/TNH/AO/NAO indices. Each point in the CORP time series represents the average of all (26 or 27) starts within the 3 months centered at each calendar month

link to ENSO and QBO (e.g., Scaife et al. 2014; Wang et al. 2017). Therefore, the subseasonal predictability might be masked by the interannual component, even when a strong MJO-NAO teleconnection exists (e.g., Cassou 2008). Another possibility is the model deficits that degrade the models' subseasonal skill, such as the too slow MJO propagation in the NCEP model (Wang et al. 2014) and the dry bias in the lower tropospheric moisture that affects the MJO-NAO teleconnection in the EMCWF model (Kim 2017).

References

Adams DK, Comrie AC (1997) The North American monsoon. Bull Am Meteorol Soc 78(10):2197–2213

Baldwin MP, Dunkerton TJ (2001) Stratospheric harbingers of anomalous weather regimes. Science 294(5542):581–584

Barnston A, Livezey RE (1987) Classification, seasonality, and persistence of low-frequency circulation patterns. Mon Weather Rev 115:1083–1126

Bauer P, Thorpe A, Brunet G (2015) The quiet revolution of numerical weather prediction. Nature 525:47–55

Becker E, Van Den Dool H (2016) Probabilistic seasonal forecasts in the North American multimodel ensemble: a baseline skill assessment. J Clim 29(8):3015–3026

Black J, et al. (2017) The predictors and forecast skill of northern hemisphere teleconnection patterns for lead times of 3–4 weeks. Mon Weather Rev 148:2855–2877

Brunet G, et al. (2010) Collaboration of the weather and climate communities to advance sub-seasonal to seasonal prediction. Bull Am Meteorol Soc 91:1397–1406

Cai W, Van Rensch P, Cowan T, Sullivan A (2010) Asymmetry in ENSO teleconnection with regional rainfall, its multidecadal variability, and impact. J Clim 23(18):4944–4955

Cassou C (2008) Intraseasonal interaction between the Madden–Julian oscillation and the North Atlantic oscillation. Nature 455(7212):523–527

Chen M, Wang W, Kumar A (2013) Lagged ensembles, forecast configuration, and seasonal predictions. Mon Weather Rev 141(10):3477–3497

Cohen J, Entekhabi D (1999) Eurasian snow cover variability and Northern Hemisphere climate predictability. Geophys Res Lett 26:345–348

Cole JE, Cook ER (1998) The changing relationship between ENSO variability and moisture balance in the continental United States. Geophys Res Lett 25:4529–4532

Cook BI, Seager R, Miller RL (2011) On the causes and dynamics of the early twentieth century North American pluvial. J Climate 24:5043–5060. https://doi.org/10.1175/2011JCLI4201.1

Dee DP, et al. (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q J R Meteorol Soc 137(656):553–597

DelSole T, Trenary L, Tippett MK, Pegion K (2017) Predictability of week-3–4 average temperature and precipitation over the contiguous United States. J Clim 30(10):3499–3512

Deser C, Timlin MS (1997) Atmosphere–ocean interaction on weekly timescales in the North Atlantic and Pacific. J Clim 10(3):393–408

Deser C, Alexander MA, Timlin MS (2003) Understanding the persistence of sea surface temperature anomalies in midlatitudes. J Clim 16(1):57–72

Dunstone N, Co-authors (2016) Skilful predictions of the winter North Atlantic Oscillation one year ahead. Nat Geosci 9(11):809–814

García-Serrano J, Frankignoul C, Gastineau G, De La Càmara A (2015) On the predictability of the winter Euro-Atlantic climate: lagged influence of autumn Arctic sea ice. J Clim 28(13):5195–5216

Hales JE Jr (1972) Surges of maritime tropical air northward over the Gulf of California. Mon Weather Rev 100(4):298–306

Higgins RW, Yao Y, Wang XL (1997) Influence of the North American monsoon system on the US summer precipitation regime. J Clim 10(10):2600–2622

Huffman GJ, Bolvin DT (2012) Version 1.2 GPCP one- degree daily precipitation data set documentation. WDC-A, NCDC, p 27. ftp://meso.gsfc.nasa.gov/ pub/1dd-v1.2/1DD_v1.2_doc.pdf. Accessed 19 Jan 2017

Hurrell JW, Kushnir Y, Ottersen G, Visbeck M (2003) An Overview of the North Atlantic Oscillation. In: Hurrell JW, Kushnir Y, Ottersen G, Visbeck M (eds) The North Atlantic Oscillation), climatic significance and environmental impact. American Geophysical Union, Washington

Johansson A (2007) Prediction skill of the NAO and PNA from daily to seasonal time-scales. J Clim 20(10):1957–1975

Johnson NC, Collins DC, Feldstein SB, L'Heureux ML, Riddle EE (2014) Skillful wintertime North American temperature forecasts out to 4 weeks based on the state of ENSO and the MJO. Wea Forecasting 29:23–38

Kim HM (2017) The impact of the mean moisture bias on the key physics of MJO propagation in the ECMWF reforecast. J Geophys Res Atmos 122:7772–7784

Kirtman BP, Co-authors (2014) The North American multimodel ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. Bull Am Meteorol Soc 95:585–601

Koster RD, et al. (2010) Contribution of land surface initialization to subseasonal forecast skill: first results from a multi-model experiment. Geophys Res Lett 37(2):L02402

Kumar A (2009) Finite samples and uncertainty estimates for skill measures for seasonal prediction. Mon Weather Rev 137(8):2622–2631

Li S, Robertson AW (2015) Evaluation of submonthly precipitation forecast skill from global ensemble prediction systems. Mon Weather Rev 143(7):2871–2889

Lin H (2014) Subseasonal variability of North American wintertime surface air temperature. Clim Dyn 45(5–6):1–19

Ma X, Xie F, Li J, Tian W, Ding R, Sun C, Zhang J (2018) Effects of Arctic stratospheric ozone changes on spring precipitation in the northwestern United States. Atmos Chem Phys Discuss. https://doi.org/10.5194/acp-2018-414, in review

McAfee SA, Russell JL (2008) Northern annular mode impact on spring climate in the western United States. Geophys Res Lett 35:L17701

Moron V, Gouirand I (2003) Seasonal modulation of the El Niño–Southern oscillation relationship with sea level pressure anomalies over the North Atlantic in october–march 1873–1996. Int J Climatol 23:143–155

Murphy JM (1988) The impact of ensemble forecasts on predictability. Q J R Meteorol Soc 114(480):463–493

O'Reilly CH, Heatley J, MacLeod D, Weisheimer A, Palmer TN, Schaller N, Woollings T (2017) Variability in seasonal forecast skill of Northern Hemisphere winters over the 20th century. Geophys Res Lett 44:5729–5738

Robertson AW, Ghil M (1999) Large-scale weather regimes and local climate over the western United States. J Clim 12(6):1796–1813

Robertson AW, Kumar A, Peña M, Vitart F (2015) Improving and promoting subseasonal to seasonal prediction. Bull Am Meteorol Soc 96(3):ES49–ES53

Rodney M, Lin H, Derome J (2013) Subseasonal prediction of wintertime North American surface air temperature during strong MJO events. Mon Weather Rev 141(8):2897–2909

Saha S, et al. (2010) The NCEP climate forecast system reanalysis. Bull Am Meteorol Soc 91:1015–1057

Scaife AA, et al. (2014) Skillful long-range prediction of European and North American winters. Geophys Res Lett 41(7):2514–2519

Scaife AA, et al. (2016) Seasonal winter forecasts and the stratosphere. Atmos Sci Lett 17(1):51–56

Stan C, Straus DM, Frederiksen JS, Lin H, Maloney ED, Schumacher C (2017), Review of tropical-extratropical teleconnections on intraseasonal time scales. Rev Geophys. https://doi.org/10.1002/2016RG000538

Straus DM, Shukla J (2002) Does ENSO force the PNA? J Clim 15(17):2340–2358

Vigaud N, Robertson AW, Tippett MK (2017) Multimodel ensembling of subseasonal precipitation forecasts over North America. Mon Weather Rev 145:3913–3928

Vitart F (2014) Evolution of ECMWF sub-seasonal forecast skill scores. Q J R Meteorol Soc 140:1889–1899

Vitart F, et al. (2008) The new VAREPS-monthly forecasting system: a first step towards seamless prediction. Q J R Meteorol Soc 134:1789–1799

Vitart F, Robertson AW, Anderson DL (2012) Subseasonal to Seasonal Prediction Project: Bridging the gap between weather and climate. Bull World Meteorol Organ 61(2):23

Vitart F, et al. (2017) The subseasonal to seasonal (S2S) prediction project database. Bull Am Meteorol Soc 98(1):163–173

Wallace JM, Gutzler DS (1981) Teleconnections in the geopotential height field during the Northern Hemisphere winter. Mon Weather Rev 109(4):784–812

Wang L, Chen W (2010) Downward Arctic oscillation signal associated with moderate weak stratospheric polar vortex and the cold December 2009. Geophys Res Lett 37:L09707

Wang W, Hung MP, Weaver SJ, Kumar A, Fu X (2014) MJO prediction in the NCEP climate forecast system version 2. Clim Dyn 42(9–10):2509–2520

Wang L, Ting M, Kushner PJ (2017) A robust empirical seasonal prediction of winter NAO and surface climate. Sci Rep 7(1):279

Weisheimer A, Schaller N, O'Reilly C, MacLeod DA, Palmer T (2017) Atmospheric seasonal forecasts of the twentieth century: multidecadal variability in predictive skill of the winter North Atlantic oscillation (NAO) and their potential value for extreme event attribution. Q J R Meteorol Soc 143(703):917–926

Wheeler MC, Zhu H, Sobel AH, Hudson D, Vitart F (2017) Seamless precipitation prediction skill comparison between two global models. Q J R Meteorol Soc 143(702):374–383

Younas W, Tang YM (2013) PNA predictability at various time scales. J Clim 26(22):9090–9114

Zhou S, L'Heureux M, Weaver S, Kumar A (2012) A composite study of the MJO influence on the surface air temperature and precipitation over the Continental United States. Clim Dyn 38:1459–1471