CrossMark

Subseasonal forecast of Arctic sea ice concentration via statistical approaches

Lei Wang[1,2] · Xiaojun Yuan[2] · Cuihua Li[2]

Abstract

Subseasonal forecast of Arctic sea ice has received less attention than the seasonal counterpart, as prediction skill of dynamical models generally exhibits a significant drop in the extended range (> 2 weeks). The predictability of pan-Arctic sea ice concentration is evaluated by statistical models using weekly time series for the first time. Two statistical models, the vector auto-regressive model and the vector Markov model, are evaluated for predicting the 1979–2014 weekly Arctic sea ice concentration (SIC) anomalies at the subseasonal time scale, using combined information from the sea ice, atmosphere and ocean. The vector auto-regressive model is slightly inferior to the vector Markov model for the subseasonal forecast of Arctic SIC, as the latter captures more effectively the subseasonal transition of the underlying dynamics. The cross-validated forecast skill of the vector Markov model is found to be superior to both the anomaly persistence and damped anomaly persistence at lead times > 3 weeks. Surface air and ocean temperatures can be included to further improve the forecast skill for lead times > 4 weeks. The long-term trends in SIC due to global warming and its polar amplification contribute significantly to the subseasonal sea ice predictability in summer and fall. The vector Markov model shows much higher skill than the NCEP CFSv2 model for lead times of 3–6 weeks, as evaluated for the period of 1999–2010.

1 Introduction

The trans-Arctic shipping routes are projected to be more and more navigable as the Arctic sea ice shrinks due to global warming and Arctic amplification (i.e., the Arctic has shown much more warming than lower latitudes in the past few decades, e.g., Smith and Stephenson 2013). The longer opening season in the Arctic will also stimulate more social activities such as transpolar shipping, tourism, and natural resources (e.g., O'Garra 2017; Petrick et al. 2017). All these economic activities call for better understanding of the Arctic sea ice predictability, especially over subseasonal time scales.

Subseasonal variability of Arctic sea ice concentration (SIC) is usually predicted using two types of models: dynamical (or numerical) and statistical models. For day to day SIC variability, dynamical forecasts are usually skillful in the short to medium range (1–10 days; e.g., Smith et al. 2016), whereas statistical forecasts are often only skillful for longer lead times (e.g., Wang et al. 2016b). However, Wang et al. (2016b) only examined the summer predictability of daily SIC variability, and further attempts revealed poor skill in other seasons with similar settings (not shown). Also, daily forecasts beyond 2 weeks seem to lack adequate skill and necessity in general. Therefore, it might be better to study weekly-mean, instead of daily, SIC variability in the subseasonal range, as week-3 and 4 forecasts receive increasing attention (e.g., Robertson et al. 2015). Based on the local correlation with SIC, oceanic and atmospheric information will be used in addition to SIC itself to improve the forecast skill, following Yuan et al. (2016).

Weekly Arctic SIC time series show large variability over marginal seas, particularly in the Barents Sea (Fig. 1a; note that the total variability includes contributions from long-term trends and variability over all smaller time scales), and a visible reduction in variability can be seen after long-term linear trends are removed (Fig. 1b). Although relatively small in total variability, sea ice in the Chukchi Plateau regions has the largest contribution from long-term trends, up to one-third to the total variability (red regions

✉ Lei Wang
  wanglei_ias@fudan.edu.cn

1 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
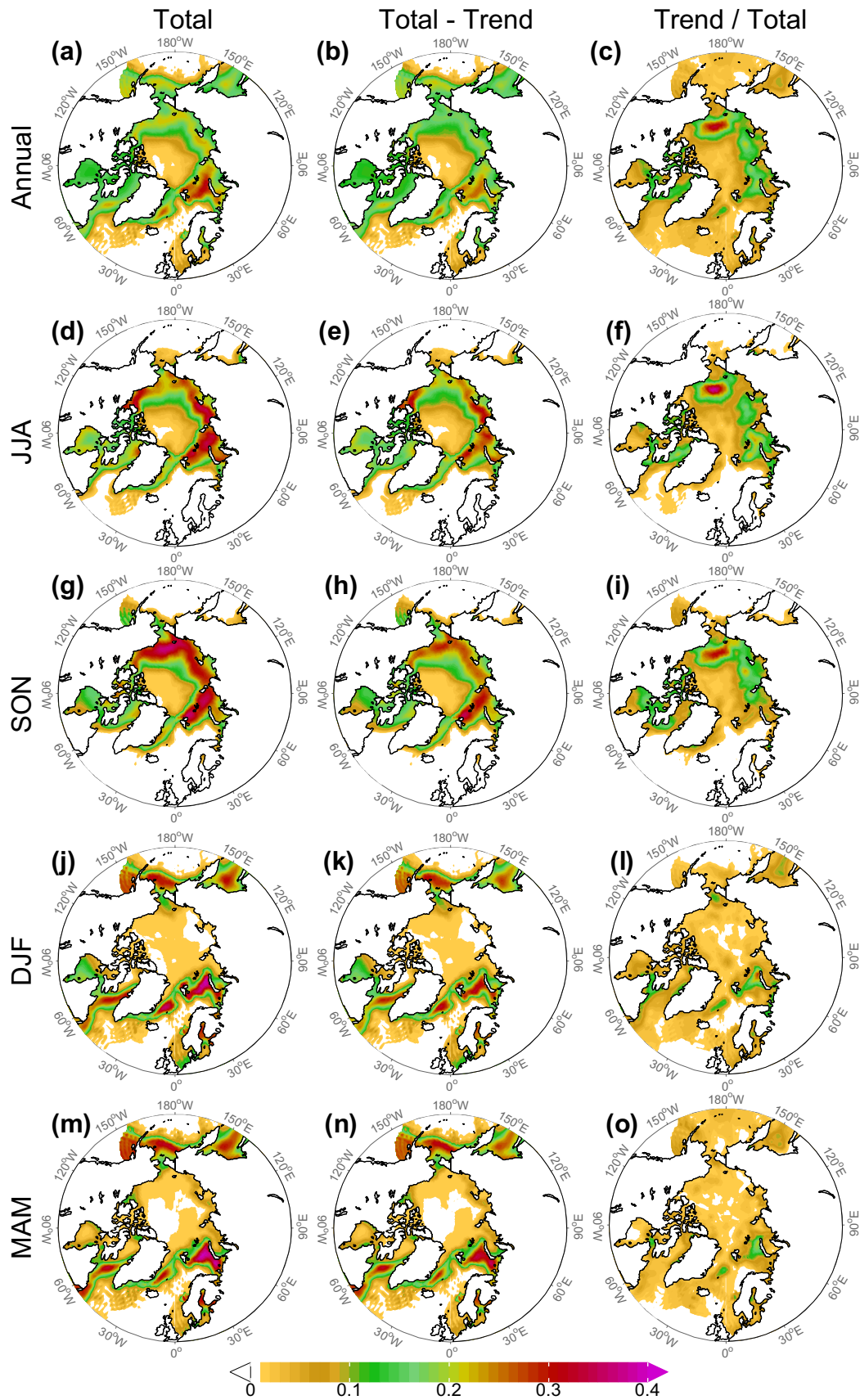
◀**Fig. 1** Weekly SIC variability for all seasons, JJA, SON, DJF, and MAM [top to bottom with (left) and without (middle) trends, and the portion of the trends to the total variability (right)]. The color scales indicate the standard deviation of SIC anomalies (in fraction)

in Fig. 1c). The SIC variability varies significantly in seasons following the retreat and advance of ice edge (Fig. 1d, g, j, m). During the retreat seasons (summer and fall), the SIC variability is highest in the marginal seas including the Beaufort, Chukchi, East Siberian, Laptev, Kara, and Barents Seas (Fig. 1d, g). On the other hand, during the advance seasons (winter and spring), the SIC variability is largest over the Barents, Greenland, Labrador, and Bering Seas, as well as the Sea of Okhotsk (Fig. 1j, m). The large contribution of long-term trends in the Chukchi Plateau regions seen in Fig. 1c mainly occurs in summer and fall (Fig. 1f, i). The Barents Sea appears to be the only area that has large variability in all seasons, consistent with Yang et al. (2016).

Observational data and statistical models used in this study are described in Sect. 2, which also includes the cross-validation methods used to evaluate model skill. The skill of the take-1-year-out cross-validated and retrospective forecasts is presented in Sects. 3 and 4, respectively. The statistical model skill is also compared with that of a dynamical model in Sect. 5, followed by the summary and discussion in Sect. 6.

2 Data and method

The Arctic SIC comes from the 1979–2014 satellite data provided by the National Snow and Ice Data Center (NSIDC). This SIC data set has been generated using the Bootstrap Algorithm (Comiso 2010). The daily SIC data are combined into 4 weeks per month, i.e., days 1–7 as the first week, days 8–14 as the second, days 15–21 as the third, and the rest of the month as the fourth. This definition of weeks makes it easier to calculate anomaly and organize training data than using the traditional weeks. The SIC is resampled from the original 25×25 km to the 100×100 km grids (with a sample size of 2033 grids) to reduce the spatial dimension of the data.

Atmospheric and oceanic data from ERA-Interim reanalysis (Dee et al. 2011) for the same period (1979–2014) are used to improve the forecast skill on top of SIC. The variables used here include sea surface temperature (SST), surface air temperature (SAT), 300 hPa geopotential height (Z300) and winds (UV300), following Yuan et al. (2016). The surface temperatures represent the thermal coupling between the sea ice and its upper and lower boundaries, i.e., air and water, while the upper tropospheric variables are used to represent the impact of large-scale and low-frequency atmospheric motions on the sea ice (surface winds

are often too noisy and vibrant and thus not included). These data are spatially sampled on a $2° \times 2°$ grid within 40–90°N, where the sample size is 4680 (180 longitudinal grids × 26 latitudinal grids) for atmospheric data and 2762 for SST (number of grids over ocean poleward of 40°N).

The temporal evolution of the coupled ice–ocean–atmosphere system is extracted from a multivariate empirical orthogonal function (MEOF) space. The oceanic and atmospheric variables are added incrementally on top of SIC in the order of + SST, + SAT, + Z300, + UV300, similar to Yuan et al. (2016). Such an order is determined by adding these variables separately to SIC and evaluating the corresponding skill (now shown). Each variable is standardized by its total variability (a single number for both spatial and temporal dimensions) before added into the matrix for MEOF decomposition. For 3–6-variable cases (starting from + SAT), the SIC is given a weight of 2 to emphasize the sea ice variability and all the other variables have a weight of 1, according to sensitivity tests (not shown). Too little weight for the SIC will reduce the sea ice representation in the leading MEOF modes and thus more modes need to be included in the statistical model in order to represent enough variance for the SIC, whereas too much weight for the SIC might underestimate the role of sea ice in the ice–air–sea coupling. The weekly climatology is taken out from all variables before calculating MEOFs, where trends are retained in the anomalies of all variables to avoid artificial spikes in the SIC time series. The trends are mostly represented by the first principal component (PC) of the MEOF analysis. Linear weekly trends are removed from PCs before they are fed into the statistical models, as the statistical models only represent stationary processes. In the evaluation, the forecast is made with the residual (detrended) PC time series and the trends are added back to the predicted PCs before the PCs are converted to the SIC forecast.

Two statistical models are tested for the Arctic sea ice subseasonal forecast, the vector autoregressive (VAR) model and the vector Markov model. The VAR model has been used to predict daily variability of summer Arctic SIC and NH atmospheric circulation (Wang et al. 2016a, b), while the scaler Markov model has been used for the seasonal forecast of Antarctic and Arctic SIC (Chen and Yuan 2004; Yuan et al. 2016). The vector Markov model used here can be written as $\mathbf{x}_i = \mathbf{A} \cdot \mathbf{x}_{i-1} + \mathbf{e}_i$, where $\mathbf{x}_i$ represents the state vector $\mathbf{x}$ at the i-th time step and is a linear function of its states at the previous time step with the coefficient matrix $\mathbf{A}$ plus white noises $\mathbf{e}$. The main difference from the scaler Markov model used by Chen and Yuan (2004) and Yuan et al. (2016) is that they apply the Markov model to individual PCs, i.e. each PC is only a function of itself, whereas in the vector Markov model each PC is a function of all leading PCs retained. The scaler Markov model better suits short time series, which is the main reason for which it has been used

for the seasonal forecast of sea ice by Chen and Yuan (2004) and Yuan et al. (2016). In contrast, the vector Markov model requires much longer records to train but may achieve better skill than the scaler version, as to be shown in this study.

The VAR model and the vector Markov model differ in their assumptions on the transition matrix **A**, i.e., the underlying dynamics. To make a forecast at the lead time of N time steps, the vector Markov model is iteratively implemented for N times and a different transition matrix **A** is used at each step. On the other hand, the VAR model directly makes the lead-N forecast in one implementation: $\mathbf{x}_i = \mathbf{A} \cdot \mathbf{x}_{i-N} + \mathbf{e}_i$, with the main assumption that the underlying dynamics are governed by the same processes. The VAR model could avoid accumulating errors as the vector Markov model might do in the iterations, but might insufficiently represent the subseasonal transition of the underlying dynamics for long lead times. The VAR model seems more skillful in predicting the daily fluctuations than the vector Markov model (Wang et al. 2016a, b) but has not been tested for the weekly (or longer time scales) variability of sea ice.

Two different cross-validation methods are adopted to examine the subseasonal predictability of Arctic SIC. One is the commonly used take-1-year-out cross-validation. The leading PCs of a single week are first selected as the target to predict, i.e., **x**. The PCs of the corresponding week plus the preceding 3 and subsequent 3 weeks in all the other years are used to retrieve the coefficient matrix **A**, which is used to make predictions of the target week (i.e., out-of-sample forecast). The extra 6 weeks are included to increase the training sample size (i.e., the degree of freedom) for a more robust retrieval of **A**. The above routine is repeated for all weeks (of all years) to obtain the predicted PCs, which are converted back into SIC anomalies at each spatial grid point. The predicted SIC is then verified against the observed data. Note that the MEOF is not recalculated for each training period because of the large trends included in SIC. The other method is the much stricter retrospective forecast, which uses a fixed period of data to train the model. The MEOF is calculated using only 20 years of data (1979–1998) and then projected on the SIC anomalies to obtain the PCs. The PCs from the training period (1979–1998) are used to estimate the coefficient matrix **A**. and the PCs from the verification period (1999–2014) only serve as initial conditions for predictions. In this way, the validation is done using exclusively future (new) data and is thus much stricter than the take-1-year-out cross-validation. Moreover, the trends in SIC are not linear in time and the verification periods have larger trends and variability than the training period, making it more challenging to predict (Holland and Stroeve 2011).

Two metrics are used to evaluate the forecast skill: the anomaly correlation coefficient (ACC) and the root-mean-square error (RMSE). The observed and predicted anomalies of SIC are both calculated relative to the observed climatology. This is a strict criterion because of the truncation in the MEOF. Since only the leading MEOF modes are retained, the predicted SIC would have a slightly different climatology from the observed, even if the PCs were perfectly predicted. Such differences in climatology will contribute to the total error in SIC. The Fisher's Z transformation is used to estimate the statistical significance of the difference of ACCs, as the ACC is not an additive quantity (see Wilks 1995 for more details). The correlation coefficient r is first transformed to z, which is approximately Gaussian, by $z = \text{arctanh}\ (r)$. Then one may use the following quantity, $z_d = |z_1 - z_2|/\text{sqrt}\ [1/(n_1 - 3) + 1/(n_2 - 3)]$, to test whether the difference between r_1 and r_2 is statistically significant. n_1 and n_2 represent the length of time series for r_1 and r_2, respectively. The critical value of z_d is 1.96 for the confidence interval of 95%, or the confidence level of 5%, by a two-tailed t test. $n_1 = n_2 = 1728$ (number of weeks within 1979–2014) for all seasons and 432 for individual seasons at each grid point. To test the statistical significance in differences between two spatially averaged correlation coefficients, the sample size of spatial grids (denoted as m) needs to be taken into account and the critical value of z_d reduces to 1.96/sqrt (m).

The SIC forecast skills, ACC and RMSE, are compared with the climatology and the (damped) anomaly persistence at each grid point. The climatology forecast predicts zero anomaly and the future state of SIC follows the climatological annual cycle. On the other hand, the anomaly persistence assumes the anomaly constant in time, using the current anomaly plus the climatology at the target time as an estimate of the future state. The damped anomaly persistence assumes the anomaly dissipative in time following the local auto-correlation (r), i.e., the amplitude of the anomaly becomes r. r usually vanishes at long lead time and thus damped anomaly persistence gradually approaches the climatology. Note that the (damped) persistence has zero error (ACC = 1) at zero lead time and thus not shown in the plots.

The SIC forecast skill is also compared between our statistical models and the National Centers for Environmental Prediction (NCEP) Climate Forecast System, version 2 (CFSv2; Saha et al. 2014) reforecasts, contribution to the Subseasonal to Seasonal (S2S) Prediction Project, a recent World Meteorological Organization effort (Vitart et al. 2017). The sea ice in the NCEP model is simulated by the GFDL Sea Ice Simulator (Griffies et al. 2004), as part of the ocean component, the Geophysical Fluid Dynamics Laboratory (GFDL) Modular Ocean Model Version 4 (MOM4p0). MOM4p0 has a horizontal resolution of 0.5° north of 30°N and 40 vertical levels. The NCEP SIC output is available on a 1.5° × 1.5° grid and the statistical forecast skill is interpolated onto this grid for a direct comparison. The NCEP reforecasts are available daily from 1999 to 2010 and have four ensemble members. The daily SIC reforecasts are

converted to weekly format using the same calendar weeks as above for observations. The ensemble mean skill is compared with the statistical forecast skill over lead times of 1–6 weeks as the NCEP forecast length is 45 days.

3 Take-1-year-out cross-validated forecast skill

The forecast skill depends on the number of MEOF modes retained in the statistical models. The more modes are included, the more variance the forecasts represent. The forecast skill usually increases with the number of modes included. However, higher modes tend to be less persistent than the leading PCs and become unpredictable (i.e., become noises) to degrade the forecasts. Therefore, there exists an optimal number of modes to be included in each statistical models for any given lead time. In Fig. 2 we show the pan-Arctic (SIC variability > 0.05) mean forecast skill (ACC and RMSE) as a function of the number

of modes, averaged over lead times from 5 to 12 weeks. This range of the lead time is chosen because the persistence of SIC is pretty high over short leads (i.e., there is much less need for prediction) and the skill drops to too low and is no longer useful for lead times longer than 12 weeks. The vector Markov model shows highest ACC for 30 modes and the MEOF of SIC + SST + SAT and lowest RMSE for 34 modes (green curves in Fig. 2a, b). Since the RMSE is roughly the same between 30 and 40 modes, the optimal number of modes is chosen as 30 for the vector Markov model. The VAR model shows highest ACC and lowest RMSE for 23 modes of SIC + SST + SAT (green curves in Fig. 2c, d). Although SIC + SST shows slightly higher ACC for 20 modes (blue curve in Fig. 2c), the corresponding RMSE is much larger (blue curve in Fig. 2d). Therefore, we choose 23 as the optimal modes for the VAR model. It is also worth noting that SIC + SST + SAT + Z300 can reach the similar level of skill with more modes included (orange curves in Fig. 2). Given the large spatial (m = 2033 grid points) and
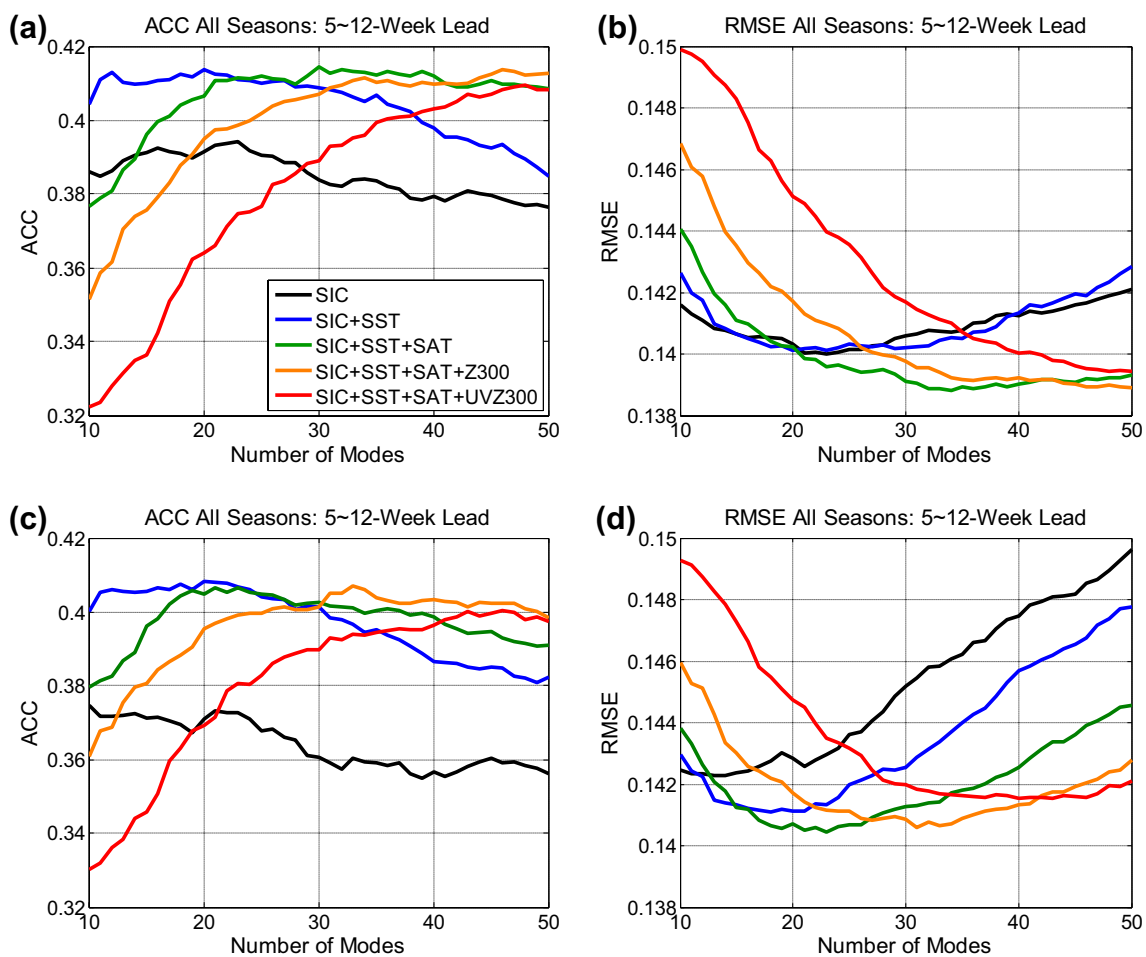
Fig. 2 Pan-Arctic mean ACC (left) and RMSE (right) of the Markov (upper) and VAR (lower) models, evaluated with all seasons within 1979–2014 by take-1-year-out cross-validation

temporal (n = 1728 weeks for all seasons) sample sizes, a 0.002 difference in the pan-Arctic mean ACC is considered statistically significant at the 95% confidence interval.

The vector Markov and VAR models show very similar forecast skill for all lead times. The vector Markov model shows slight higher ACC and lower RMSE over marginal seas (Fig. 3) where variability is relatively large (Fig. 1). The difference between the two models is much smaller for short leads (week-3–4; Fig. 3a, b) than that for long leads (week-5–12; Fig. 3c, d). The ACC difference over most grid points is less than the critical value of 0.06 for the 95% confidence interval for all seasons. Given the small difference between these two models, only results from the vector Markov model are presented hereafter.

The forecast skill shows significant seasonal dependence (Fig. 4). The pan-Arctic mean ACC is higher in summer and fall than in winter and spring by about 0.1 (blue curves in Fig. 4). The (damped) anomaly persistence

Fig. 4 1979–2014 pan-Arctic mean ACC for target time in **a** JJA, **b** SON, **c** DJF, and **d** MAM for the Markov model with 30 leading PCs. **e–h** Same as **a–d** but for RMSE

shows similar seasonal differences as anomaly persistence (red and black curves in Fig. 4). The vector Markov model starts to be superior to the (damped) anomaly persistence from week-3 in summer and spring (Fig. 4a, d) but from week-2 in fall and winter (Fig. 4b, c). The skill in summer and fall is above 0.5 at the 4-week lead, higher than in winter and spring. On the other hand, the RMSE follows the seasonal variations of SIC variability (compare Fig. 4e–h with Fig. 1). The RMSE is largest in SON, but the model skill improvements over the (damped) anomaly persistence and climatology are also larger than all the other seasons (Fig. 4f). The damped anomaly persistence is most skillful among the three skill references. The vector Markov model significantly reduces RMSE from the damped persistence

Fig. 3 Maps of **a** ACC and **b** RMSE differences between the Markov and VAR models (Markov–VAR) averaged for lead times of week-3–4. **c, d** Same as **a, b** but for lead times of week-5–12
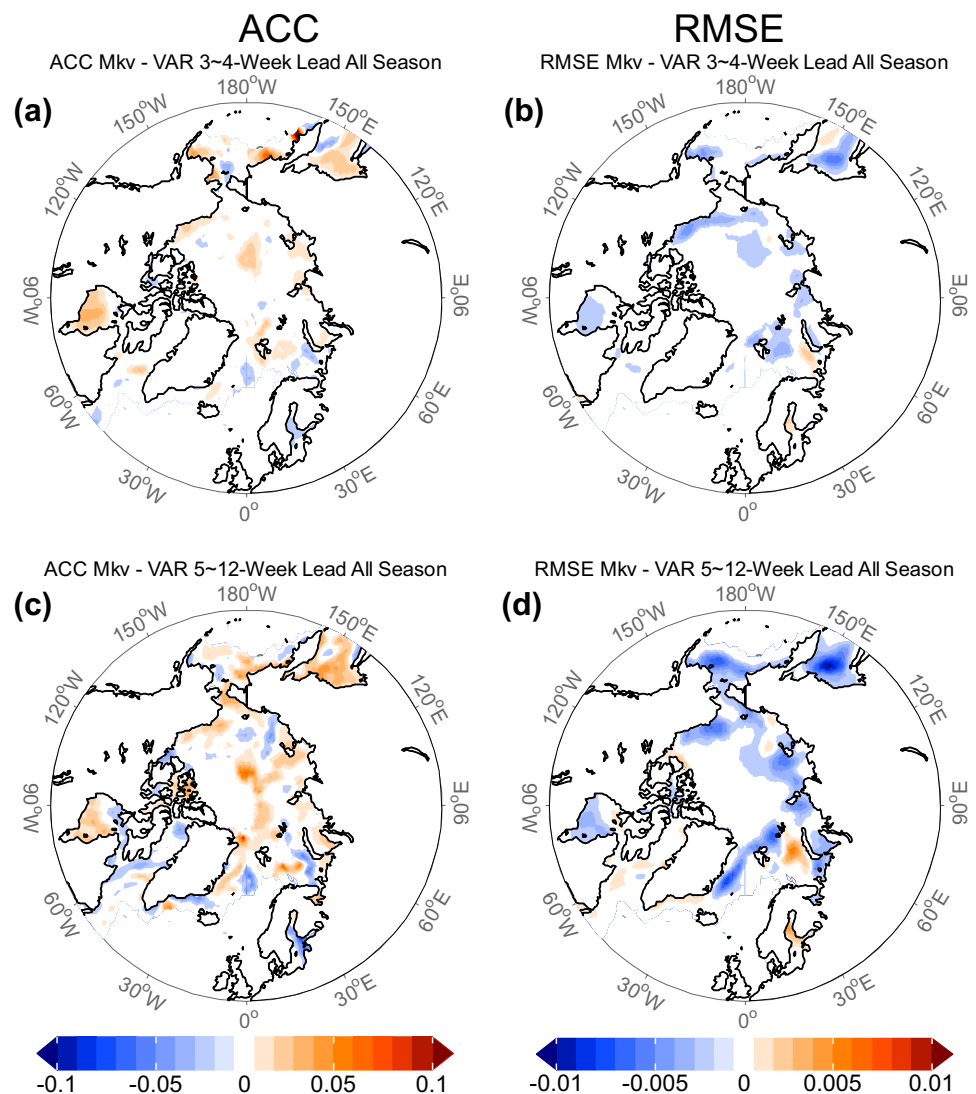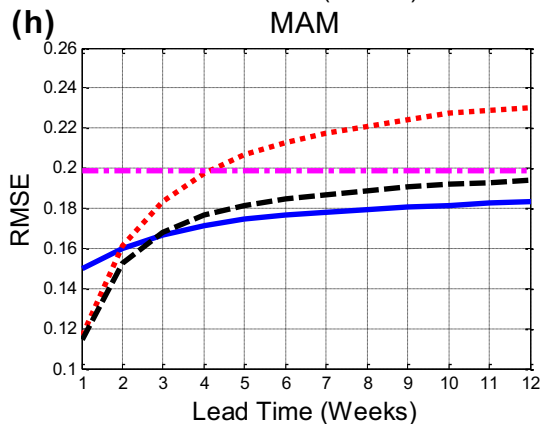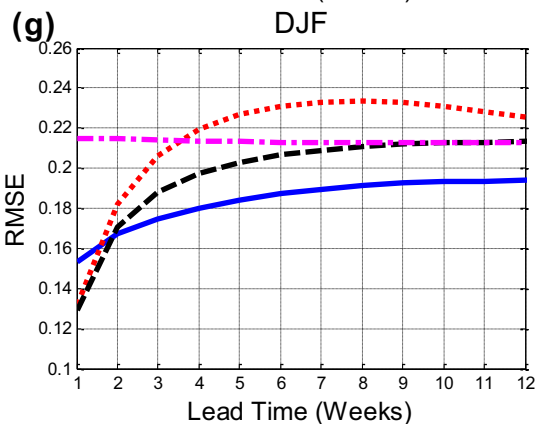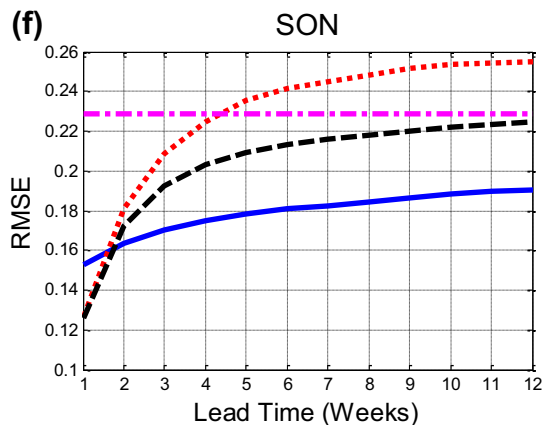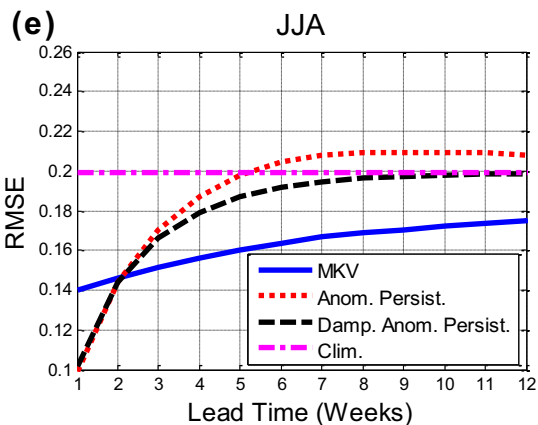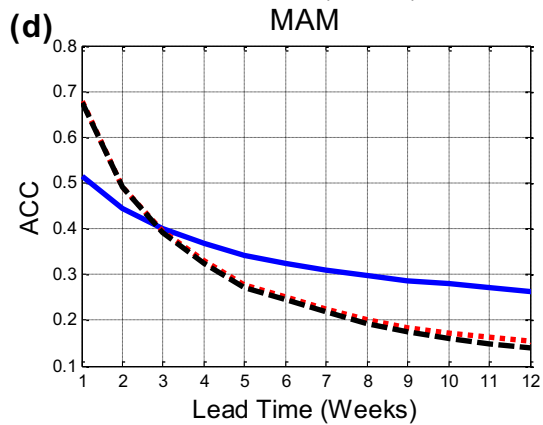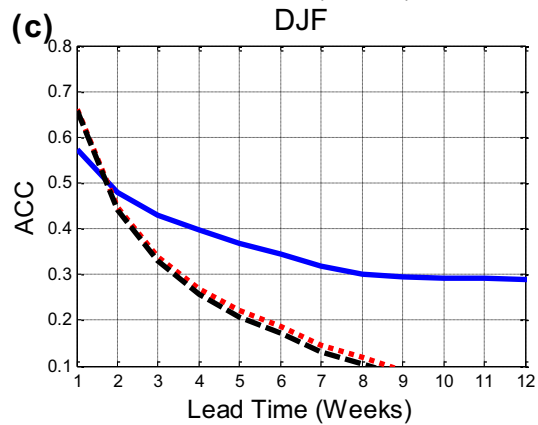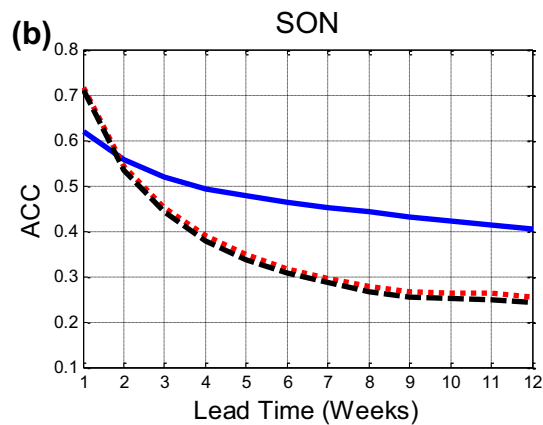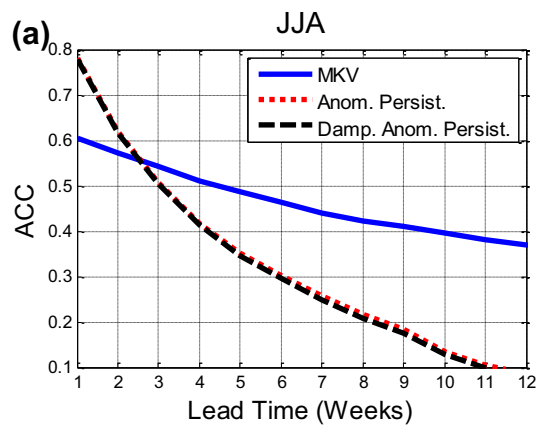
at 3-week and longer lead times for all seasons. The RMSE reduction is larger in summer and fall than in winter and spring. In general, the vector Markov model is skillful for the week-3–4 range in all seasons. Recently, the weather forecast community has focused on improving the forecast skill at the week-3–4 lead time (e.g., Vitart et al. 2017). The vector Markov model delivers promising skill for sea ice prediction at this prediction range.

The advantage of the vector Markov model skill over the damped anomaly persistence exhibits pronounced spatial variation (Fig. 5 for week-3–4), following the seasonal shifts of ice edges. In summer, the modeled ACC shows the most pronounced advantage over the Laptev Sea, East Siberian Sea, and Beaufort Sea (Fig. 5a), where correlation increases by 0.1–0.2 (any correlation difference > 0.1 is generally significant at the 95% confidence interval for individual seasons). On the other hand, the model's RMSE has an advantage over much broader regions (Fig. 5b). The advantageous regions moved more towards continents (lower latitudes) in fall and winter (Fig. 5c–f). The vector Markov model skill is almost identical to the damped anomaly persistence in the spring (Fig. 5g, h). The advantageous regions are similar with much larger amplitudes for longer lead times (Fig. 6 for week-5–2). The ACC improvements are most significant (up to 0.5) at the Beaufort and Laptev Seas in summer (Fig. 6a), whereas the RMSE decreases most in the Kara Sea in summer and at the Beaufort Sea in summer and fall (Fig. 6d). The spring skill shows some visible improvements over the damped anomaly persistence at the coastal regions of the Hudson Bay, Baffin Bay, Sea of Okhotsk, and Chukchi Sea (Fig. 6g, h). In summary, the vector Markov model starts to be superior to the damped anomaly persistence in week-3–4, and the advantage becomes much more significant in week-5–12 for summer, fall, and winter. The advantageous regions follow the seasonal evolution of ice edges.

Since the model skill shows considerable spatial variations, it is natural to examine the spatial distribution of the skill boost by adding SST and SAT. The model with SIC + SST + SAT does not improve but slight degrade the week-3–4 skill over most of the Arctic except for the Beaufort Sea in summer, the Hudson Bay in fall, the Sea of Okhotsk in winter (left column of Fig. 7). The skill boost by SST + SAT is more visible for longer lead times (right column of Fig. 7). The pan-Arctic mean RMSE of the SIC-only model is about 0.02 larger than that of the SIC + SST + SAT model for the week-5–12 lead (Fig. 2b). However, the local differences can be more than 0.02, about 1/5 of the skill improvements over the damped anomaly persistence (compare right columns of Figs. 6 and 7; note the different color scales). The skill boost by SST + SAT is most significant over the Beaufort and East Siberian Seas in the melting seasons (summer and fall; Fig. 7b, d). The Chukchi Sea

and Laptev Sea RMSEs are slightly degraded in summer (Fig. 7b), as well as the Laptev Sea and Kara Sea in fall (Fig. 7d). The winter skill is improved mainly over the Barents Sea, Hudson Bay, and Baffin Bay (Fig. 7f). SST + SAT has a minor impact on the spring skill (Fig. 7h).

The pronounced decreasing trend in the SIC since the 1990s has been shown to explain a significant portion of the seasonal forecast skill, especially for long leads (e.g., Yuan et al. 2016). It also applies to the subseasonal SIC predictability. After removing the linear weekly trends at each grid point for both observations and forecasts, the detrended skill is much lower than the full skill and the reduction in skill increases with lead time (Fig. 8a–d). The most pronounced reduction occurs in summer and is almost up to 0.2 in the pan-Arctic mean ACC (Fig. 8a). On the other hand, the model's RMSE seems not sensitive to the trends, whereas the damped anomaly persistence is significantly reduced without trends (Fig. 8e–h).

4 Retrospective forecast skill

The robustness of the vector Markov model skill can be tested in the retrospective forecast experiments. Different from the take-1-year-out cross-validation experiments shown in Sect. 3, here we choose 1979–1998 as the training period and the 1999–2014 data are used for forecast verification. That is, no future data enters the training processes, including EOF decomposition, model construction, and retrieval. This procedure is a much stricter validation method than the take-1-year-out cross-validation, and thus the forecast skill is usually lower than that of the latter. Furthermore, one particular challenge for the Arctic sea ice is that the long-term trend and inter-annual variability are both much larger in the verification period than in the training period, due to accelerated ice melting in response to climate change. Therefore, the optimal number of variables increases to four, with 300 hPa geopotential height added, for lead times of week-5–12 (orange curves in Fig. 9c, d). For the short lead of week-3–4, SIC only gives the best forecast skill (black curves in Fig. 9a, b), although just slightly better than other combinations. The optimal number of modes for lead times of week-5–12 is 23 for four variables, which corresponds to the highest ACC and lowest RMSE (Fig. 9c, d).

The retrospective forecast skill with 23 modes of four variables also shows significant seasonal dependence (Fig. 10). The pan-Arctic mean ACC is still fairly skillful in summer compared with the (damped) anomaly persistence for lead times longer than 5 weeks (Fig. 10a) but is only slightly higher or comparable in the other seasons (Fig. 10b–d). Note that the fall skill is much higher than the counterpart in the take-1-year-out cross-validation experiments (compare Fig. 10b with Fig. 4b), as the damped anomaly persistence is

Fig. 5 Maps of JJA, SON, DJF, and MAM (top to bottom) ACC (left) and RMSE (right) differences between the Markov model and the damped anomaly persistence averaged for lead times of week-3–4. Positive values in ACC and negative values in RMSE indicate that the Markov model is superior to the damped anomaly persistence
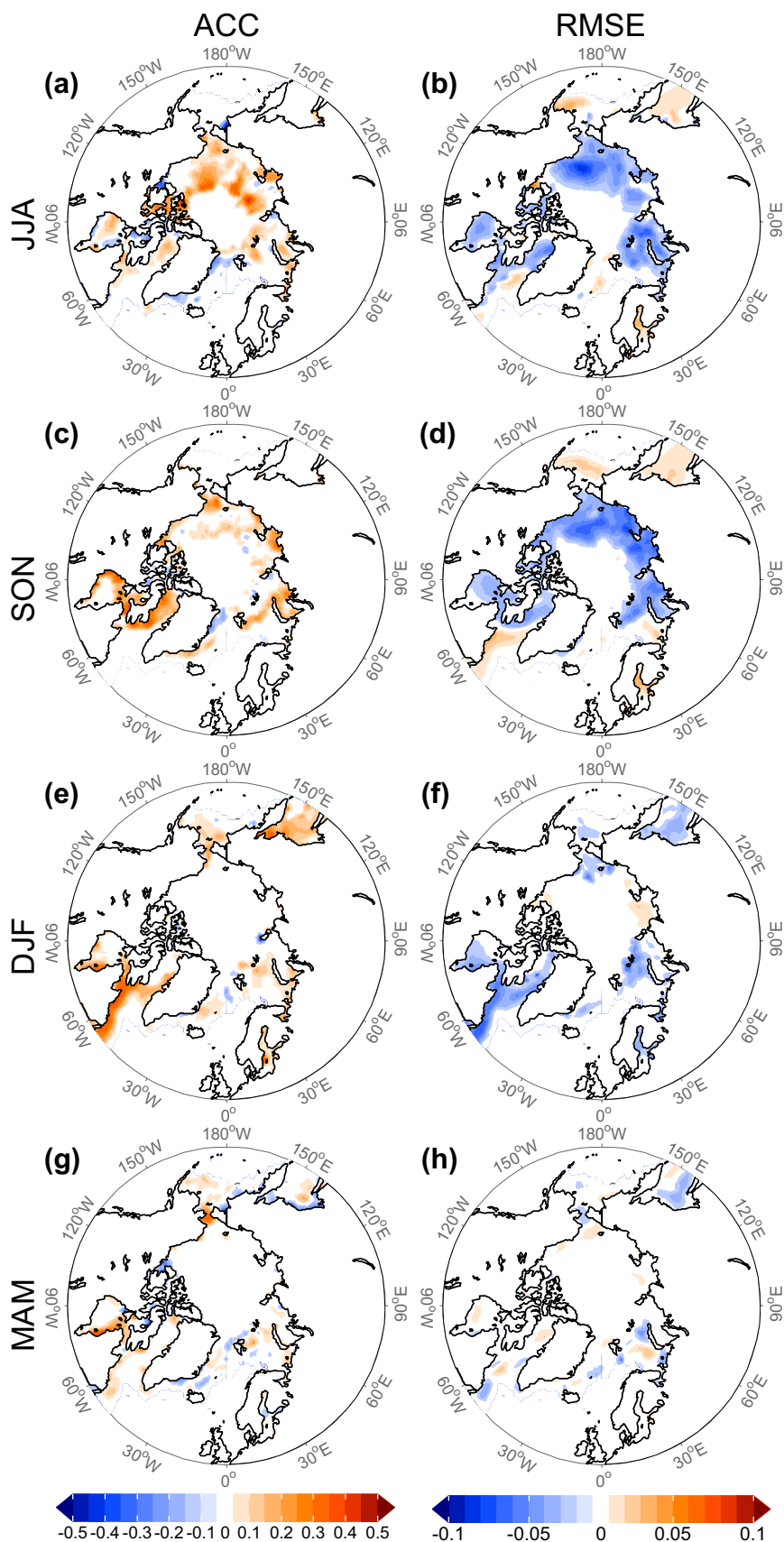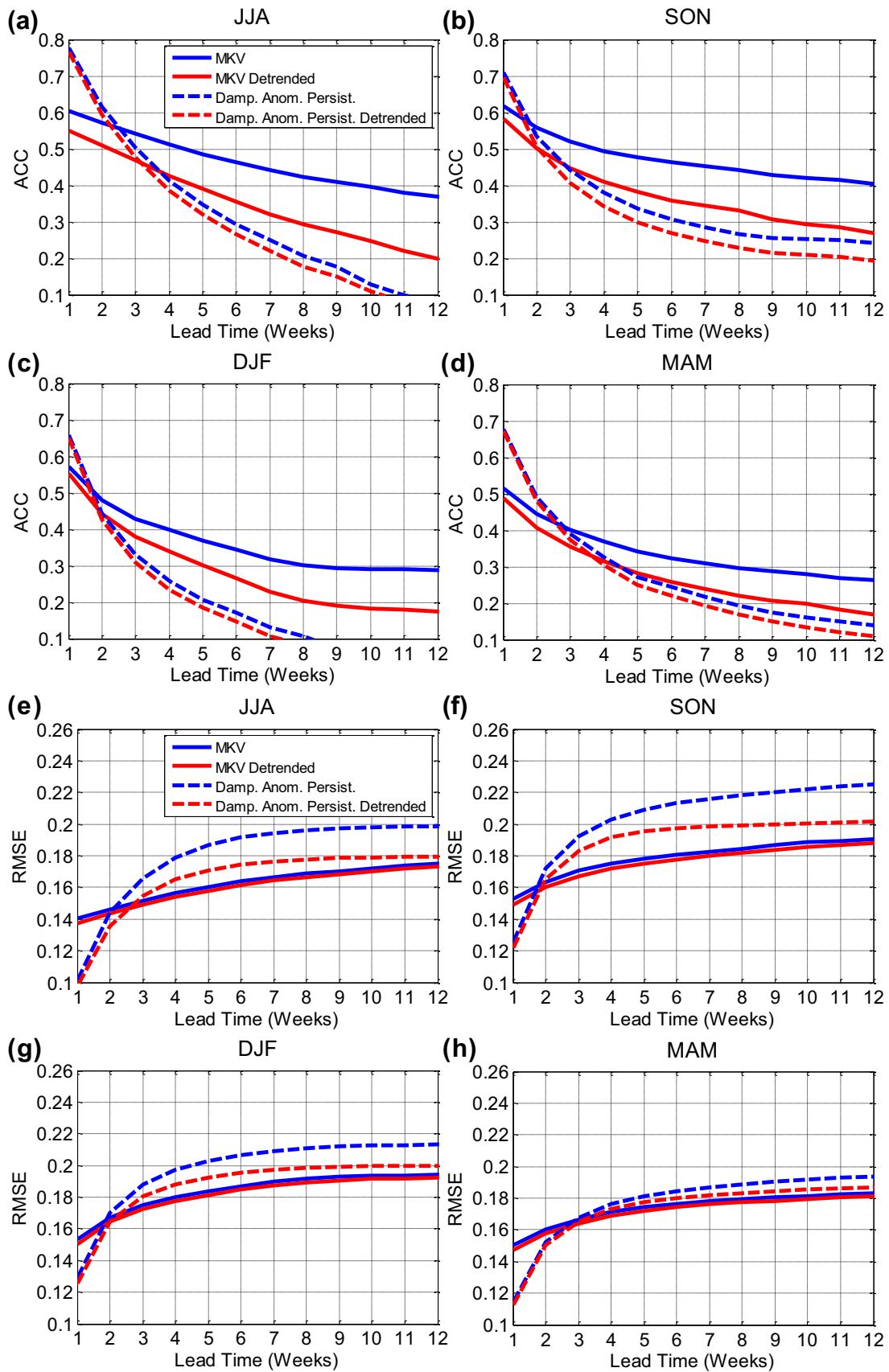
Fig. 6 Same as Fig. 5 but averaged for lead times of week-5–12

ACC
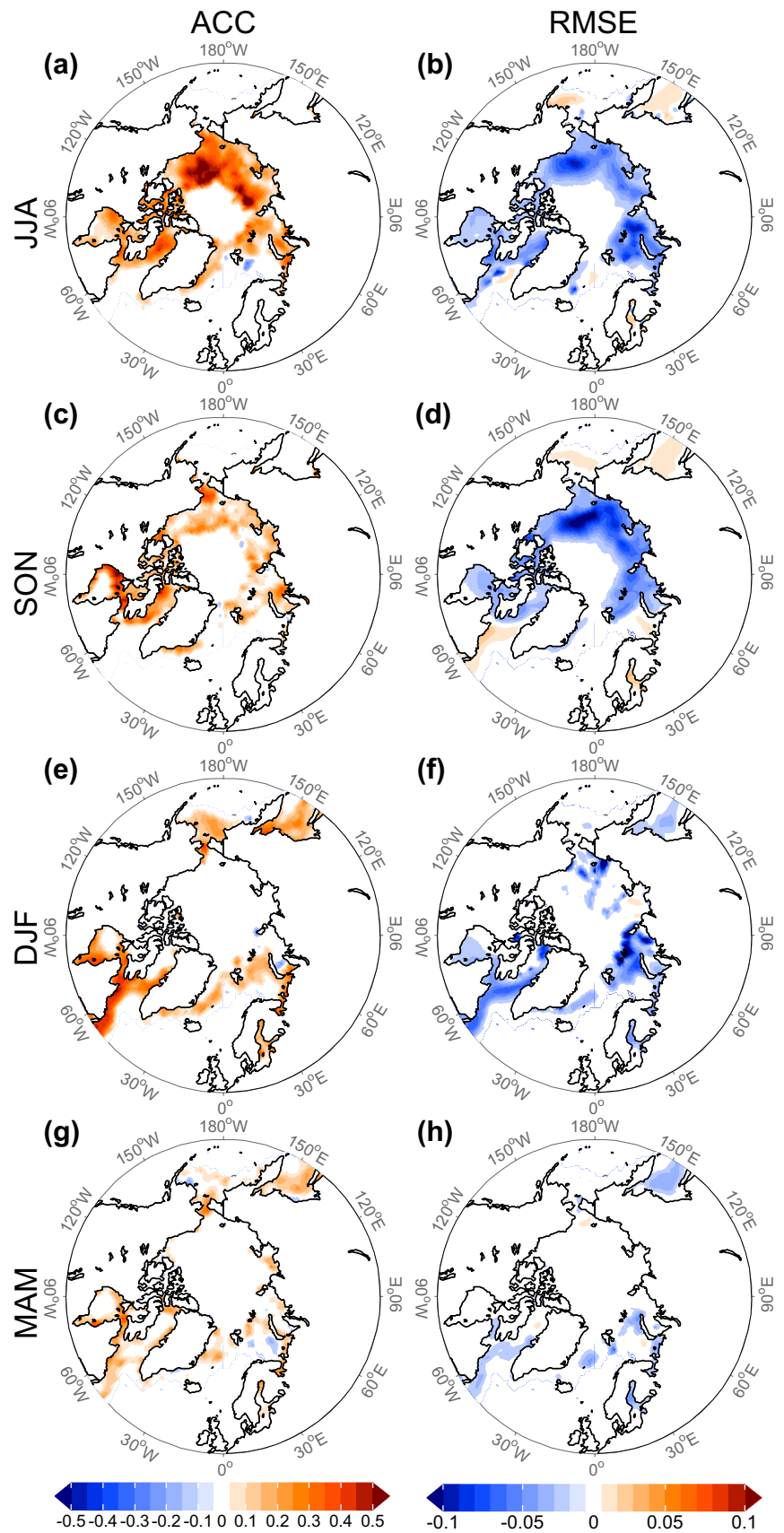
RMSE

Fig. 7 RMSE differences between the Markov model with SIC + SST + SAT and the model with SIC only in JJA, SON, DJF, and MAM (top to bottom), averaged for lead times of week-3–4 (left) and week-5–12 (right)
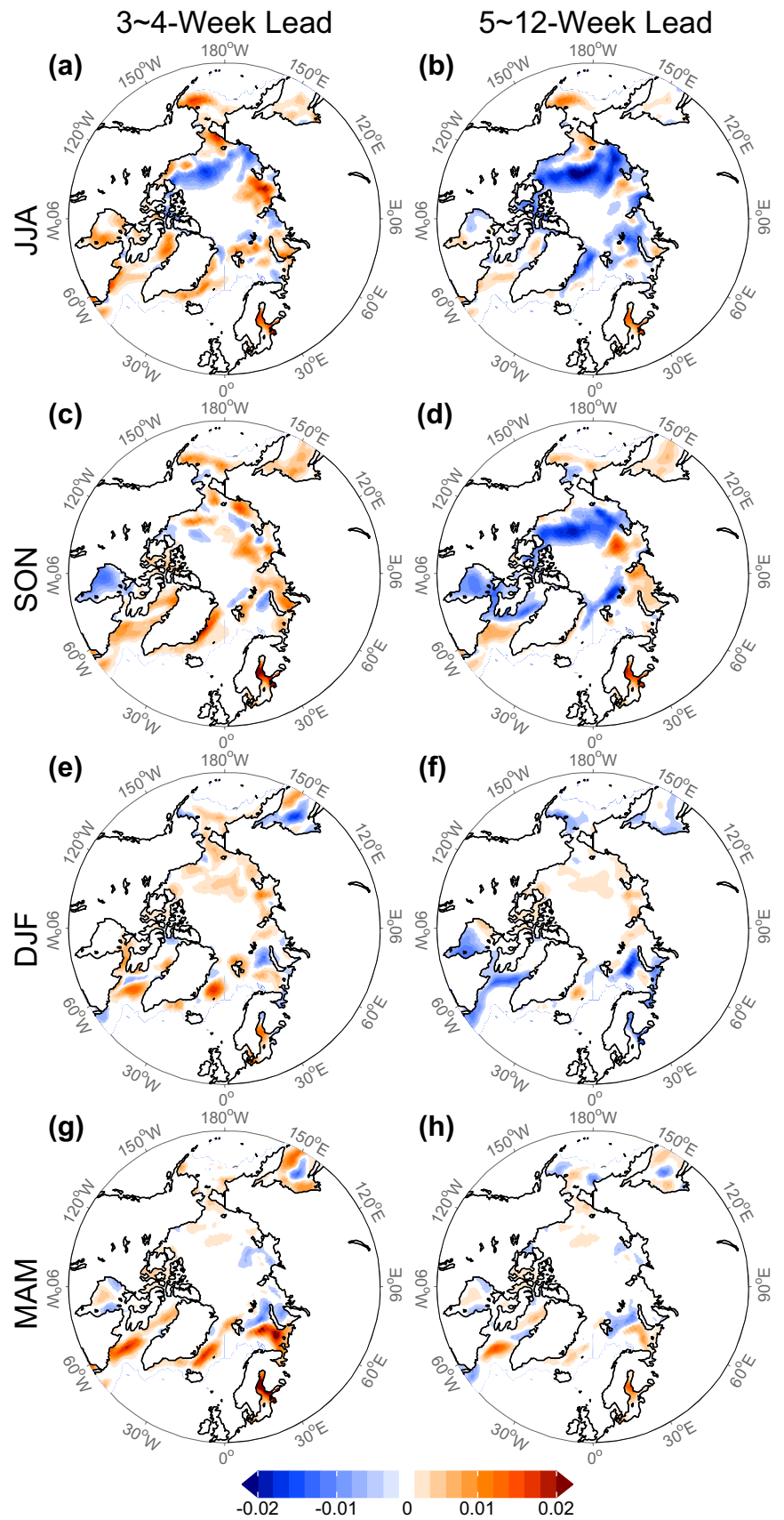
◀**Fig. 8** The pan-Arctic mean ACC for the Markov model with (blue) and without trends (red) in **a** JJA, **b** SON, **c** DJF, and **d** MAM. The panels in **e**–**h** are the same as **a**–**d** but for RMSE

much greater in the later period. In contrast, the pan-Arctic mean RMSE of the vector Markov model is still quite skillful compared with the climatology and (damped) anomaly persistence in most seasons (Fig. 10e–h), with advantage reduces only slightly compared with the take-1-year-out cross-validation experiments (compare Fig. 10e–h with Fig. 4e–h). From the RMSE perspective, the much stricter retrospective forecast is still able to make a skillful forecast for long leads (> 4 weeks).

The ACC advantage of the retrospective forecast skill over the damped anomaly persistence is fairly limited in some regions of the central Arctic and part of the Laptev Sea for week-3–4 (red regions in Fig. 11a) and expands to a much broader area in longer leads (red regions in Fig. 11c for week-5–12). On the other hand, the advantageous region for RMSE covers most of the high variability regions (i.e.,

marginal seas), even starting from week-3–4 (blue regions in Fig. 11b, d). When evaluated with ACC, the damped anomaly persistence shows better skill in a fair amount of regions (blue regions in Fig. 11a, c). However, most of these regions have comparable RMSEs by the damped anomaly persistence and the vector Markov model, as shown by in white and light blue at corresponding regions in Fig. 11b, d (such as the Greenland Sea, Bering Sea, and the Canadian Arctic Archipelago). Therefore, the pan-Arctic mean of ACC might over-emphasize these regions and underestimate the vector Markov model's skill. The pan-Arctic mean of RMSE is thus a better representation of the forecast skill.

5 Comparison with a dynamical model

The 1999–2010 NCEP ensemble mean reforecasts skill is compared with that of both the take-1-year-out and retrospective cross-validated forecasts by the vector Markov model. The pan-Arctic mean ACC for the NCEP model is
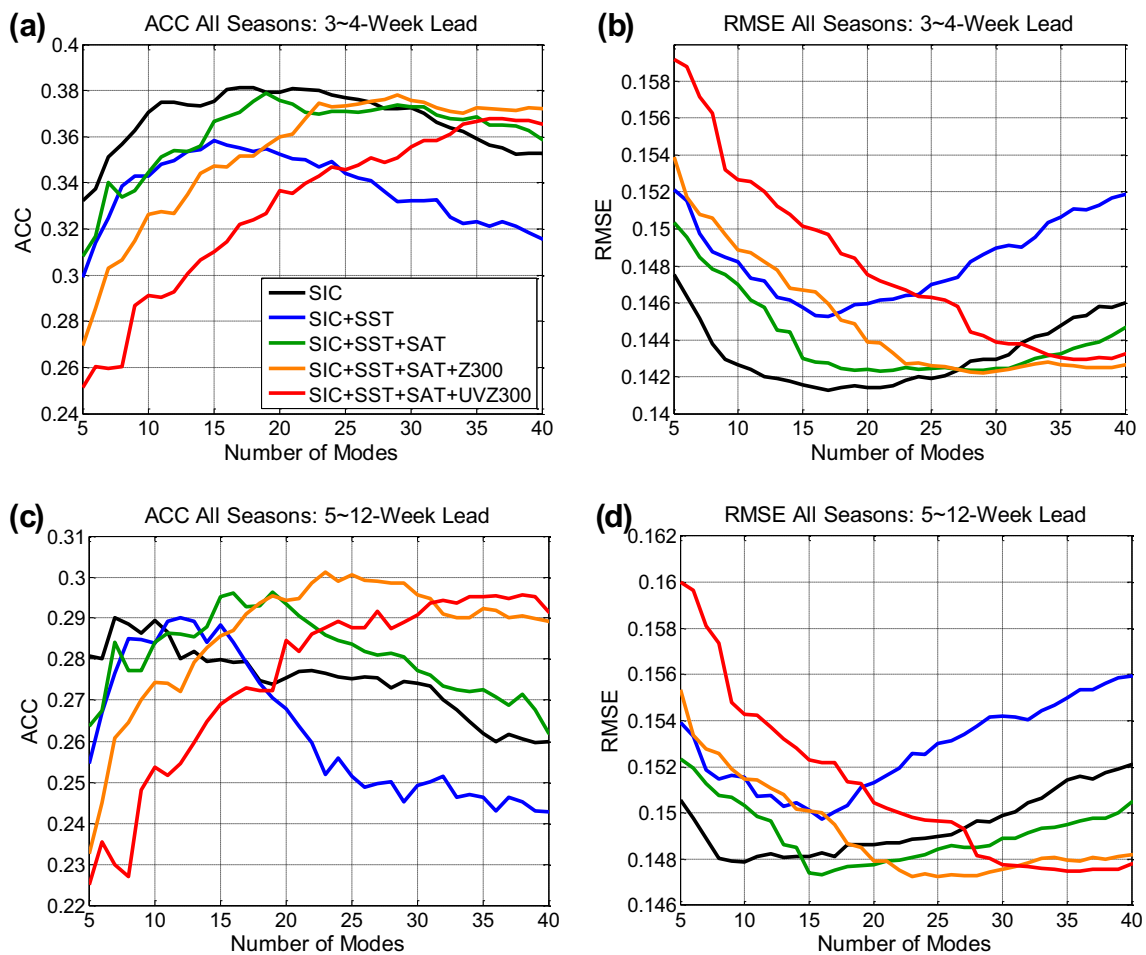
Fig. 9 Pan-Arctic mean ACC (left) and RMSE (right) for lead times of week-3–4 (upper) and week-5–12 (lower), evaluated with all seasons within 1999–2014 using a 20-year training period of 1979–1998
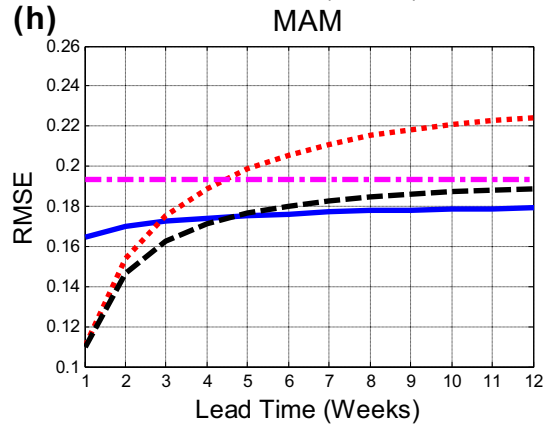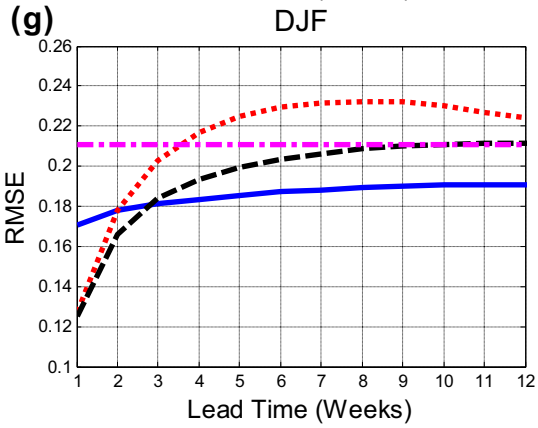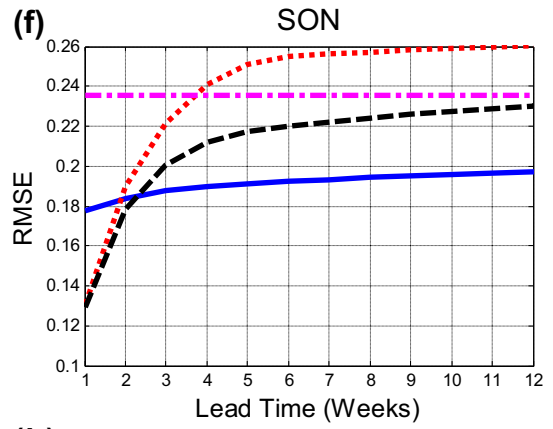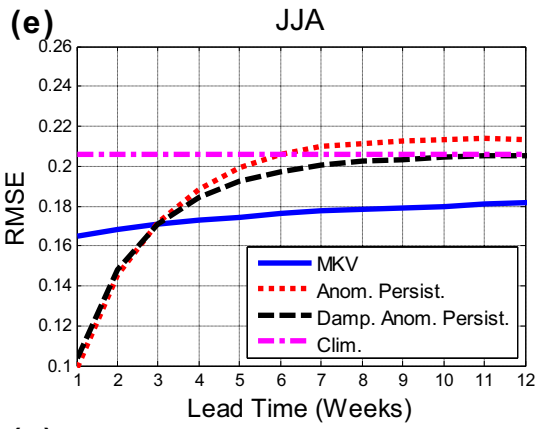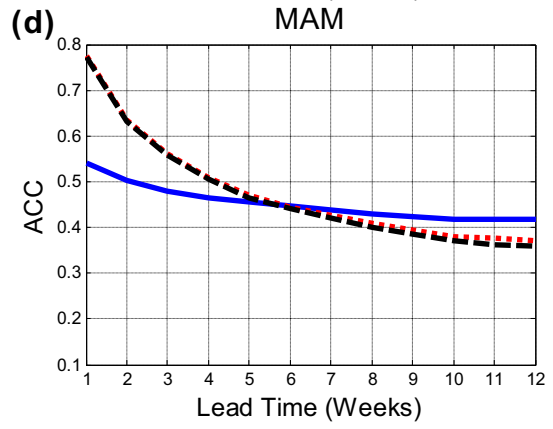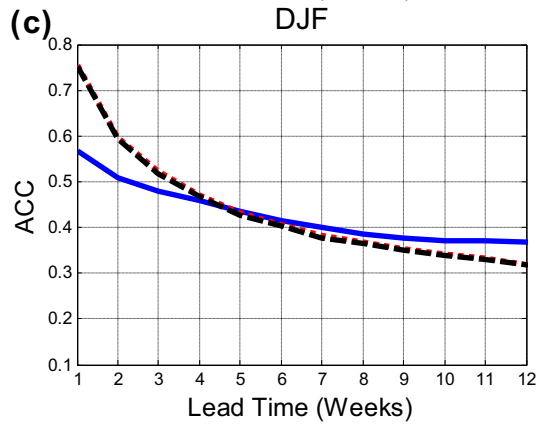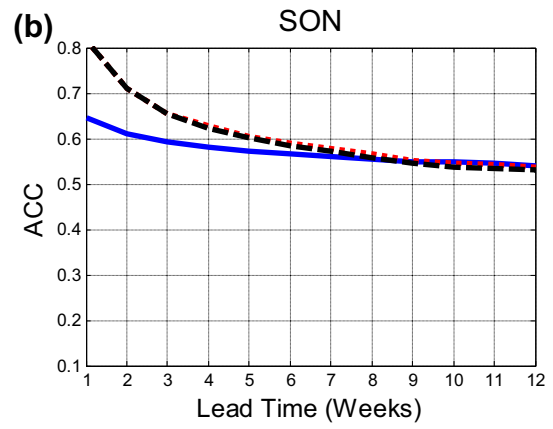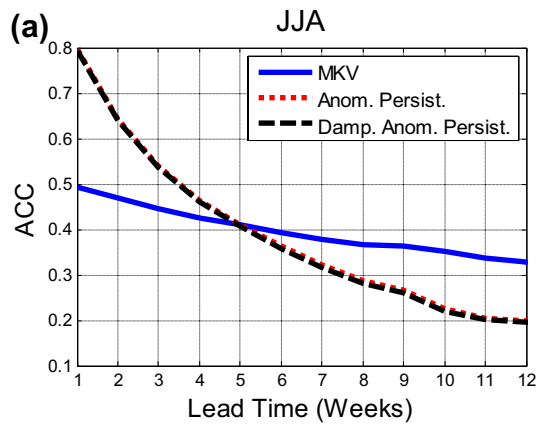
◄**Fig. 10** The pan-Arctic mean ACC for the Markov model with 23 leading PCs of the 4-variable MEOF case in **a** JJA, **b** SON, **c** DJF, and **d** MAM, evaluated during 1999–2014. The panels of **e–h** are the same as **a–d** but for RMSE

poorer than the (damped) anomaly persistence for all seasons and lead times (red vs. black curves in Fig. 12). In contrast, the vector Markov model shows comparable ACC with the (damped) anomaly persistence starting from the 3-week lead time and better skill for 5–6-week lead times. Note that, for this shorter period, the critical value of seasonal correlation differences is about 0.01 for pan-Arctic mean and is about 0.16 at each grid point at the 95% confidence interval. In particular, the take-1-year-out ACC (blue curves in Fig. 12) is better than the (damped) anomaly persistence from week-4 in spring and from week-3 in the other three seasons. The retrospective ACC (green curves in Fig. 12) is lower than the take-1-year-out skill but is close to the (damped) anomaly persistence for week-3 and 4 and higher the latter for longer lead times in summer, fall, and winter. The NCEP skill is generally lowest among all compared
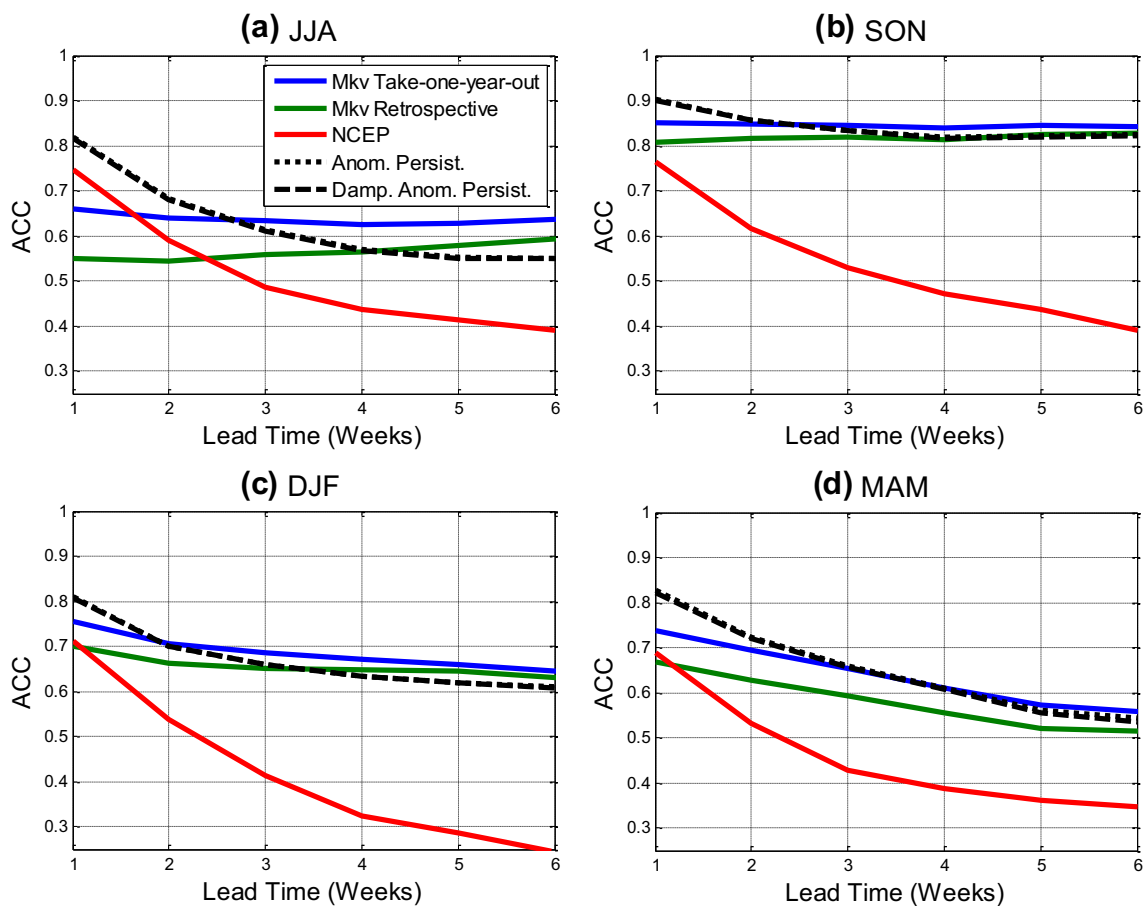
ACCs except for week-1 and 2 in summer (Fig. 12a) and shows increasing disadvantage with lead time. The skill for individual members of the NCEP model is slightly poorer than the ensemble mean (not shown), indicating that either a much larger ensemble size might be necessary to obtain satisfying forecast skill or the coupling between the atmosphere, ocean, and sea ice might not be well represented. The summer Markov model skill shows noticeable increases with lead time (blue and green curves in Fig. 12a) because the starting time is fixed to accommodate the NCEP reforecasts. This is different from previous pan-Arctic mean ACC plots where the target time is fixed. For long lead times, the target time in Fig. 12a extends to fall when the persistence is higher than that in summer. Note that the RMSE is not compared here because the NCEP model has different climatology than the observed.

The summer week-6 ACC of the take-1-year-out forecast by the vector Markov model is fairly high over most of the Chukchi, Kara, and Barents Seas, as well as in the Hudson and Baffin Bays (Fig. 13a), while the NCEP reforecasts are only skillful in a small portion of these

Fig. 11 a ACC and **b** RMSE differences between the Markov model and the damped anomaly persistence during 1999–2014, averaged for lead times of week-3–4. **c, d** Same as **a, b** but for lead times of week-5–12
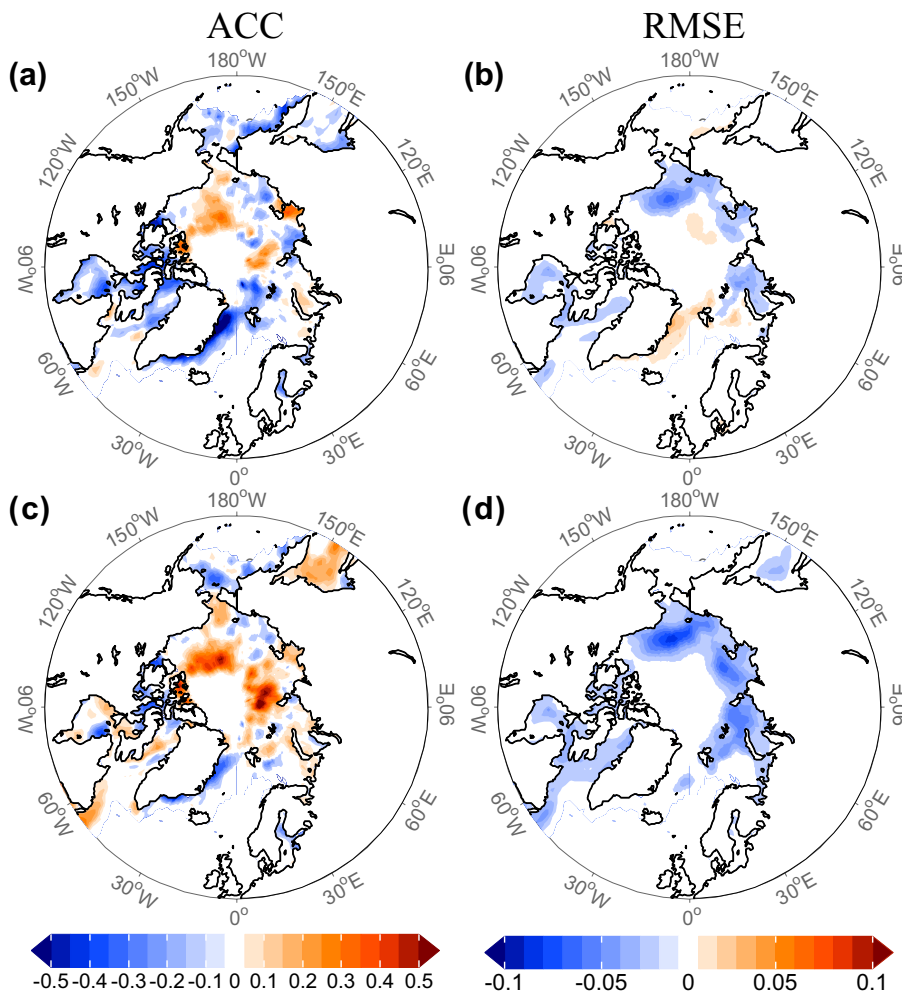
regions (Fig. 13b). The red regions in Fig. 13c represent the advantage in summer ACC of the vector Markov model over the NCEP model (any correlation difference > 0.16 is significant at the 95% confidence interval at each grid point). The week-6 forecast skill in fall is highest among all seasons for both the vector Markov and NCEP models, whereas the former shows even greater advantage than the latter compared with summer (Fig. 13f). The winter and spring predictability is much lower than the other seasons, due to much lower SIC variability. However, the vector Markov model is still fairly skillful over part of the Chukchi Sea, Kara Sea, and Hudson Bay in winter (Fig. 13g), as well as over part of the Bering Sea, Sea of Okhotsk, Barents Sea, Greenland Sea, and Baffin Bay in spring (Fig. 13j). The NCEP model is, on the other hand, much less skillful over most regions (Fig. 13h, k; see also the differences in Fig. 13i, l).

Fig. 13 1999–2010 JJA, SON, DJF, and MAM (top to bottom) ACC ▶ of the Markov model with 3 variables and 30 leading PCs using take-1-year-out cross-validation (left), NCEP S2S reforecast (middle), and their differences (Mkv-NCEP right) at the 6-week lead. Skill is only shown at grid points where weekly SIC variability is at least 0.05

6 Summary and discussion

With increasing demands from social activities in the Arctic, the seasonal forecast of Arctic sea ice has been a popular topic in recent years, and the scientific community has made significant progress on it. In contrast, the sub-seasonal forecast of Arctic sea ice is still fairly challenging for most dynamical models, as most dynamical models show insufficient prediction skill beyond a few weeks. This study evaluated two statistical models for the subseasonal predictability of 1979–2014 weekly Arctic sea ice concentration, with integrated information from the sea ice,


Fig. 12 The pan-Arctic mean 1999–2010 ACC for the take-1-year-out and retrospective cross-validation by the Markov model and the NCEP S2S reforecast with starting time in **a** JJA, **b** SON, **c** DJF, and **d** MAM. The take-one-year-out experiment uses 3 variables and 30 leading PCs, while the retrospective experiment uses 4 variables and 23 leading PCs, respectively. The ACC is averaged over grid points where weekly sea ice variability is at least 0.05 in the corresponding season of 1999–2010
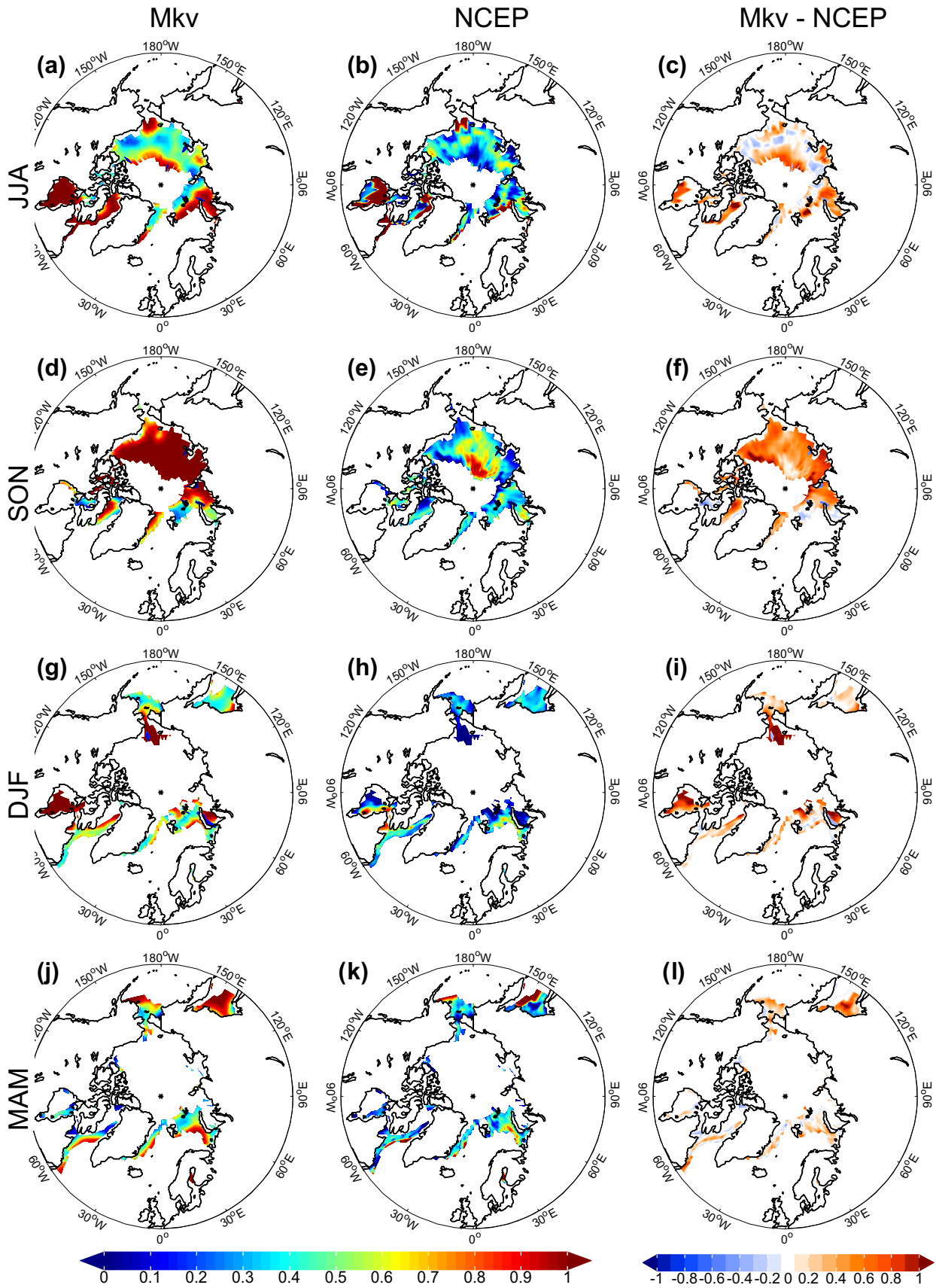
ocean and atmosphere. Although previously shown to be more skillful in predicting summer sea ice daily variability (Wang et al. 2016b), the vector auto-regressive model is slightly inferior to the vector Markov model for the sub-seasonal forecast of weekly Arctic sea ice concentration.

The cross-validated forecast skill of the vector Markov model is found to be superior to both the anomaly persistence and damped anomaly persistence at a lead time of 3 weeks or longer, as revealed by the take-1-year-out experiments for the period of 1979–2014. The regions where the vector Markov model is more skillful than the damped anomaly persistence are usually close to the ice edge and shift with seasons. The advantageous regions are mainly located in the Laptev Sea, East Siberian Sea, and the Beaufort Sea in summer and move slightly southward (towards continents) in fall and winter. The spring skill of the vector Markov model is fairly close to the damped anomaly persistence. Surface air and ocean temperatures can help further improve the long-range forecast skill for lead times longer than 5 weeks. The vector Markov model skill is most significantly enhanced along the Beaufort and East Siberian Seas in summer and fall. The winter and spring skill remains relatively unchanged.

The global warming and the polar amplification have caused pronounced long- term trends in the Arctic sea ice, which contribute significantly to the subseasonal sea ice predictability in summer and fall. The pan-Arctic mean anomaly correlation is reduced by up to 0.2 in summer and up to 0.1 in the other seasons when linear trends are removed. Since linear trends are quite predictable for the subseasonal time scale, the existence of significant long-term trends in sea ice is not considered a prediction barrier but more an opportunity.

The above conclusions are fairly robust in the more challenging retrospective forecast experiments using only 20 years of data (1979–1998) to train the vector Markov model. Although the ACC shows much less advantage over the damped anomaly persistence, the overall ACC is even higher in the period of 1999–2014 as SIC anomalies become more persistent than the early period. In addition, the model's RMSE for 1999–2014 is slightly higher in summer and fall but even lower in winter and spring, compared with that in the early period. The advantage in the model's RMSE over the damped anomaly persistence also shows only minor decreases compared with the take-1-year-out cross-validation skill.

The vector Markov model performance is also directly compared with the NCEP CFSv2 reforecast skill for the period of 1999–2010 for lead times from 1 to 6 weeks. The statistical model exhibits much better skill than the dynamical model in most cases except for the 1-week lead in summer. The NCEP model forecasts quickly become unsatisfying after 1 or 2 weeks, indicating that the chaotic forecast errors from the atmospheric component contaminate the sea ice component shortly and much room remains for the NCEP model to improve its representation of the coupling between these subcomponents of the Earth system.

Although the vector Markov model has shown good skill in predicting the subseasonal variability of Arctic sea ice, there is still room to improve the accuracy and robustness of the predictions. First, the robustness across different ice concentration datasets, such as the NASA team SIC (Cavalieri et al. 1996), should be examined. Second, our study focused on the extended range (i.e., > 2 weeks) but the short range skill can be largely improved by including more principal components and/or more variables in the vector Markov model. The surface winds and sea level pressure might be more useful in the near-term sea ice forecast to help decision-making processes during emergencies such as the RV Akademik Shokalskiy rescue operation (e.g., Zhai et al. 2015). Last but not least, the vector Markov model can also be applied for regional sea ice predictions, where the MEOF modes built for a local domain might better represent the physical processes that are more relevant to the local sea ice changes. Such regional models thus might provide better forecast ability than the pan-Arctic model introduced in this study.

References

Cavalieri DJ, Parkinson CL, Gloersen P, Zwally HJ (1996) Sea ice concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS passive microwave data, version 1. In: NASA National Snow and Ice Data Center Distributed Active Archive Center. Boulder, Colorado. https://doi.org/10.5067/8GQ8LZQVL0VL **(updated yearly)**

Chen D, Yuan X (2004) A Markov model for seasonal forecast of Antarctic sea ice. J Clim 17(16):3156–3168

Comiso JC (2010) Polar Oceans from space. Springer, New York

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q J R Meteorol Soc 137(656):553–597

Griffies SM, Harrison MJ, Pacanowski RC, Rosati A (2004) Technical guide to MOM4. GFDL Ocean Gr Tech Rep 5:337. http://www.gfdl.noaa.gov/fms. Accessed 15 Mar 2011

Holland MM, Stroeve J (2011) Changing seasonal sea ice predictor relationships in a changing Arctic climate. Geophys Res Lett 38(18):1–6

O'Garra T (2017) Economic value of ecosystem services, minerals and oil in a melting Arctic: a preliminary assessment. Ecosyst Serv 24:180–186

Petrick S, Riemann-Campe K, Hoog S, Growitsch C, Schwind H, Gerdes R, Rehdanz K (2017) Climate change, future Arctic Sea ice, and the competitiveness of European Arctic offshore oil and gas production on world markets. Ambio 46(3):410–422

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

Saha S, Moorthi S, Wu X, Wang J, Nadiga S, Tripp P, Behringer D, Hou YT, Chuang HY, Iredell M, Ek M (2014) The NCEP climate forecast system version 2. J Clim 27(6):2185–2208

Smith LC, Stephenson SR (2013) New trans-Arctic shipping routes navigable by midcentury. Proc Natl Acad Sci 110(13):E1191-E1195

Smith GC, Roy F, Reszka M, Surcel Colan D, He Z, Deacu D, Belanger JM, Skachko S, Liu Y, Dupont F, Lemieux JF (2016) Sea ice forecast verification in the Canadian global ice ocean prediction system. Q J R Meteorol Soc 142(695):659–671

Vitart F, Ardilouze C, Bonet A, Brookshaw A, Chen M, Codorean C, Déqué M, Ferranti L, Fucile E, Fuentes M, Hendon H (2017) The subseasonal to seasonal (S2S) prediction project database. Bull Am Meteor Soc 98(1):163–173

Wang L, Ting M, Chapman D, Lee DE, Henderson N, Yuan X (2016a) Prediction of northern summer low-frequency circulation using a high-order vector auto-regressive model. Clim Dyn 46(3–4):693–709

Wang L, Yuan X, Ting M, Li C (2016b) Predicting summer Arctic sea ice concentration intraseasonal variability using a vector autoregressive model. J Clim 29(4):1529–1543

Wilks DS (1995) Statistical methods in the atmospheric sciences: an introduction. Academic Press, San Diego

Yang XY, Yuan X, Ting M (2016) Dynamical link between Arctic sea ice and the Arctic Oscillation. J Clim. https://doi.org/10.1175/JCLI-D-15-0669.1

Yuan X, Chen D, Li C, Wang L, Wang W (2016) Arctic sea ice seasonal prediction by a linear Markov model. J Clim 29(22):8151–8173

Zhai M, Li X, Hui F, Cheng X, Heil P, Zhao T, Jiang T, Cheng C, Ci T, Liu Y, Chi Z (2015) Sea-ice conditions in the Adélie Depression, Antarctica, during besetment of the icebreaker RV Xuelong. Ann Glaciol 56(69):160–166