Climate Dynamics
https://doi.org/10.1007/s00382-018-4461-3

Robustness of Arctic sea-ice predictability in GCMs

E. Blanchard-Wrigglesworth[1] · M. Bushuk[2,3]

Received: 20 October 2017 / Accepted: 19 September 2018

Abstract
General circulation models have been amply used to quantify Arctic sea-ice predictability. While models share some common aspects of predictability loss with increasing forecast lead time, there is significant model spread in the magnitude and timing of predictability loss. Here we show that inter-model differences in predictability are linked to inter-model differences in the persistence timescales of sea-ice anomalies that are unique to each model, with models that exhibit longer persistence having higher potential predictability. Given this result and previous work showing that in a single model control simulation the magnitude of persistence fluctuates between multi-annual periods of high and low persistence, we assess whether initial-value predictability is dependent on the persistence state of the initial conditions. We find that predictability is not clearly impacted by the persistence state of the initial conditions, suggesting that predictability may be robust within a constant climate mean state.

Keywords Sea ice · Predictability · Arctic · GCMs

1 Introduction

Over the last decade, a growing effort has taken place in the sea-ice research community to quantify and understand sea-ice predictability, in part spurred by recent Arctic sea-ice loss and a growing stakeholder interest in the region (Guemas et al. 2016; Jung et al. 2016). Operational environmental forecasting centers are beginning to include sea-ice variables as forecast output and assimilate sea-ice information into their initial conditions (e.g., Hebert et al. 2015), novel forecast techniques are being developed both using dynamical and statistical models (e.g., Wang et al. 2013; Merryfield et al. 2013; Sigmond et al. 2013; Msadek et al. 2014; Schröder et al. 2014; Peterson et al. 2015; Yuan et al. 2016; Bushuk et al. 2017; Petty et al. 2017), and since 2008 September Arctic sea-ice extent forecasts produced in the preceding months are regularly collected and disseminated by the Sea Ice Prediction Network (SIPN, http://www.arcus.org/sipn/sea-ice-outlook).

One of the foundations underpinning our understanding of the mechanisms and limits of sea-ice predictability has been the use of 'perfect-model' experiments (PMEs) run with fully-coupled general circulation models (GCMs) (e.g., Koenigk and Mikolajewicz 2009; Holland et al. 2010; Blanchard-Wrigglesworth et al. 2011b; Day et al. 2014, 2016; Tietsche et al. 2014; Germe et al. 2014). These experiments assess the inherent predictability present in a GCM since forecast initial conditions (ICs) are known perfectly and there are no model physics uncertainties, in the sense that the model is used to predict itself. The chaotic growth of infinitesimal errors that are added to the ICs is the only source of forecast error growth, and thus the predictability skill found in perfect-model studies is considered to be the upper limit of predictability for a specific GCM (e.g., Collins 2002).

Studies performing PMEs using different GCMs have found varying degrees of initial-value predictability of sea-ice area (SIA) and sea-ice volume (SIV). While Blanchard-Wrigglesworth et al. (2011b) and Tietsche et al. (2014) found that most pan-Arctic SIA and SIV predictability is lost after 1–2 years and 3–4 years respectively, Germe et al. (2014) found longer predictability that in some seasons and regions could last up to 8–9 years.

✉ E. Blanchard-Wrigglesworth
ed@atmos.washington.edu

1 Department of Atmospheric Sciences, University of Washington, Seattle, WA 98195-1640, USA

2 Geophysical Fluid Dynamics Laboratory, NOAA, Princeton, NJ, USA

3 University Corporation for Atmospheric Research, Boulder, CO, USA

Published online: 25 September 2018

Fewer studies have assessed how sea-ice predictability may be mean-state and model dependent. Holland et al. (2010) examined predictability in a greenhouse-gas forced GCM simulation and found that predictability of summer sea ice was lower in the 2010s compared to the 1970s, and attributed this change in predictability to a transition to a thinner ice mean state in the model. Similarly, Tietsche et al. (2013) found lower predictability of a mid twenty-first century sea-ice anomaly compared to a present-day anomaly. Tietsche et al. (2014) and Day et al. (2014) found that predictability skill differs across GCMs—while the patterns of loss of skill with increased forecast lead time are qualitatively similar, the magnitude of predictability skill for the same forecast lead time differs by a factor of two for forecast lead times of just one season.

Our goal in this work is to investigate inter-model differences in sea-ice predictability, and to determine whether the mechanisms that contribute to inter-model predictability differences may also result in intra-model differences in predictability. To do this, we consider two complementary sets of GCM experiments: (1) a multi-model suite of PMEs designed to investigate initial-value sea-ice predictability; and (2) a suite of PMEs performed with a single GCM whose ICs are chosen to sample different persistence states.

2 Inter-model predictability: insights from APPOSITE

A plausible metric that may offer insights into the range of predictability skill found across GCMs is the timescale of persistence of sea-ice anomalies. Day et al. (2014) found that seasonal fluctuations in the timescales of SIA persistence could help explain the different rates of perfect-model predictability loss for different initialization months in a single GCM, providing a link between persistence and predictability. Globally, inter-model differences in perfect model predictability of sea surface temperatures (SSTs) have been linked to inter-model differences in SST autocorrelation (Kumar et al. 2014).

We expand on these ideas by investigating the link between sea ice persistence and initial-value predictability in the Arctic Predictability and Prediction on Seasonal-to-Interannual TimEscales (APPOSITE) set of PMEs (Day et al. 2014; Tietsche et al. 2014; Day et al. 2016). APPOSITE quantifies initial-value sea-ice predictability in 6 different GCMs in a modern climate mean state (constant radiative forcing taken from 1990 to 2005 depending on the model). All model simulations include prognostic sea-ice components. The runs consist of century-long control simulations run under constant radiative forcing from which the PMEs are initialized after a spin-up phase of 100 years. Figure 1a shows the normalized root mean square error (NRMSE) of SIA for all APPOSITE PMEs. NRMSE is defined as:

$$NRMSE = \frac{\sqrt{\langle (x_{kj} - x_{ij}) \rangle^2_{i,j,k\neq i}}}{\sqrt{2\sigma^2}} \tag{1}$$

where $\langle \rangle_i$ denotes the expectation value, calculated by summing over the specified index with appropriate normalization, $x_{ij}(t)$ is either SIA or SIV at lead time t, index j indicates the ensemble, and indices i and k indicate the ensemble member. The σ in the denominator is the standard deviation in the control run calculated after detrending. An NRMSE of 1 indicates no predictability skill relative to a climatological forecast.

While the tendency in all models is for NRMSE to increase with longer lead times (i.e., for predictability to decrease), there is clear model spread in NRMSE as documented by Tietsche et al. (2014). Interestingly, there are also seasonal fluctuations in NRMSE that vary across models. Some models tend to have higher predictability in the winter (e.g., E6F and HadGEM1-2), others have higher predictability in the summer (e.g,. EC-Earth2.2) while others do not show any seasonal fluctuations (e.g., MPI-ESM).

Can persistence help explain inter-model spread in NRMSE across lead time and seasons? To address this we characterize persistence with two metrics: the year-to-year (Y2Y) autocorrelations of September and March SIA (hereafter referred to as Y2Y Sep and Y2Y Mar, respectively). We calculate the Y2Y metrics from the detrended September and March SIA timeseries of the APPOSITE control runs. In Fig. 1b, c we show scatter plots of NRMSE and the Y2Y metrics for forecast lead times of 1 through 16 months (see the supplementary material for timestep animations). Both persistence metrics are correlated with NRMSE at different times of the year. Y2Y Sep is correlated with NRMSE outside the winter maximum sea ice months (June through December), while Y2Y Mar is correlated with NMRSE during the winter maximum sea ice months (February through May) (Fig. 1d)—in other words, GCMs with high Y2Y Sep (Mar) persistence tend to have higher summer (winter) predictability. Interestingly, there is no correlation between the two Y2Y metrics across the APPOSITE models (Fig. 1e), which helps explain the different seasonal patterns of NMRSE predictability across models, as high Y2Y Sep values (and thus high summer predictability) do not imply high Y2Y Mar values, and vice versa. Exploring the relationship between Y2Y Sep and Y2Y Mar across all CMIP5 models shows that they are only weakly related (R^2 = 0.13, see Fig. 1e and Fig. S1 in the supplementary information), and that the APPOSITE models reasonably sample the CMIP5 state space of Y2Y autocorrelations. Interestingly, detrended observations (for which we use NSIDC's sea ice index, [10])

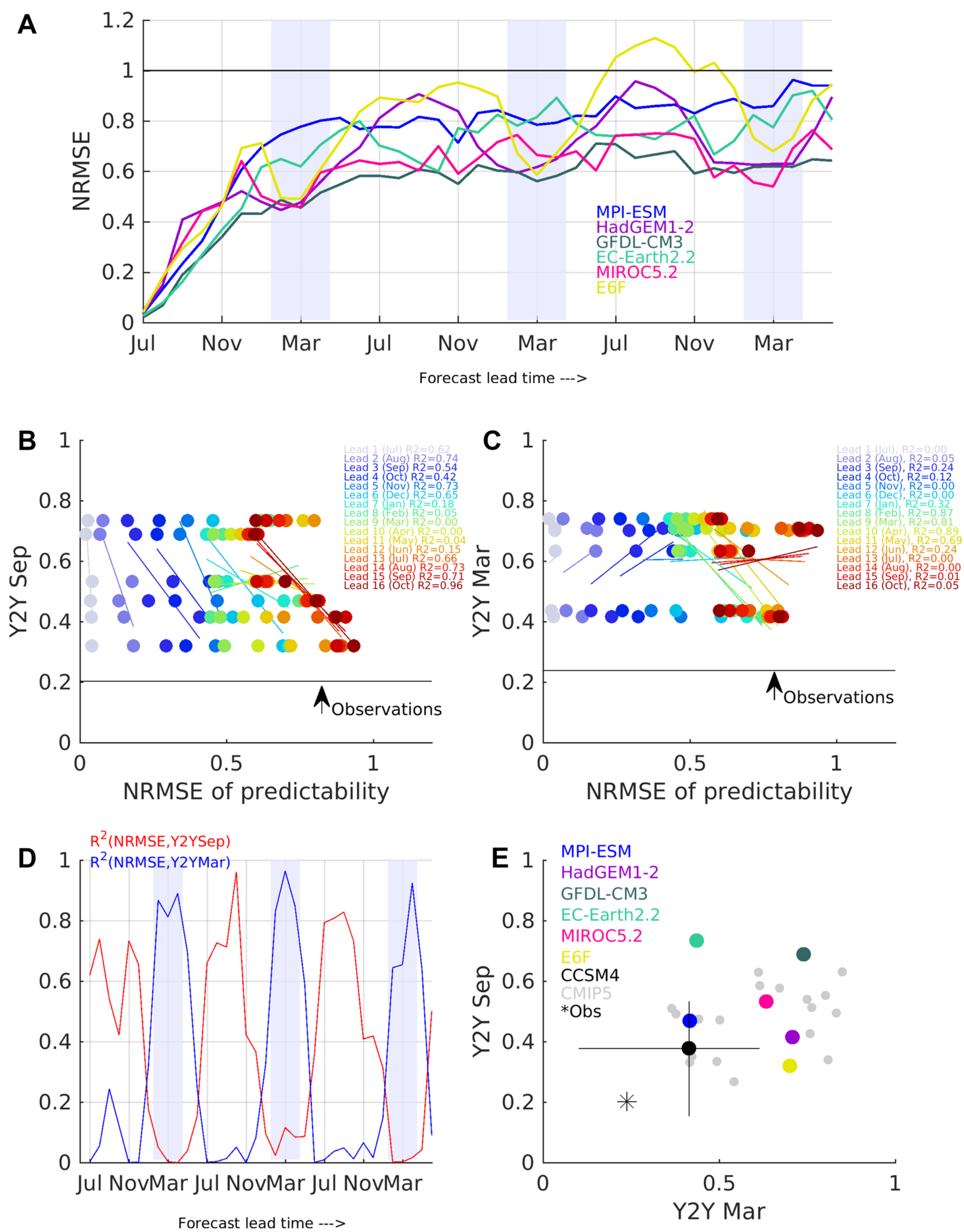

Fig. 1 **a** NRMSE of SIA in APPOSITE, **b**, **c** lead scatter plot of NRMSE and SIA Y2Y September and March autocorrelation respectively for forecast lead times of 1–16 months, **d** R^2 of the linear fits in **b** and **c**; and **e** scatter of Y2Y September and March autocorrelations, including values for the 1300-year long CCSM4 control used in the paper, CMIP5 models and observations. The CCSM4 cross-lines indicate the range of autocorrelations using centennial data. CMIP5 model correlations are calculated from detrended 1850–1955 data (see Fig S3 for further details)

show lower values in both Y2Y persistence metrics than the models. Using the anomaly correlation coefficient metric (ACC, see Eq. 4) to quantify predictability shows similar results (see Figure S2 in the supplementary information).

The two Y2Y metrics of persistence are linked to the two mechanisms of SIA anomaly re-emergence (Blanchard-Wrigglesworth et al. 2011a), which result from different physical processes at play in geographically distinct regions (Y2Y Sep driven by sea-ice thickness anomalies in the central Arctic, Y2Y March by upper-ocean heat anomalies in the sub-Arctic seas) which likely helps explain their uncoupled nature.

3 Intra-model persistence: variability and predictability

In recent work, Bushuk et al. (2015) and Bushuk and Giannakis (2015) have found that in a long control GCM simulation these Y2Y memory re-emergence events have significant temporal variability, characterized by fluctuating periods of strong and weak re-emergence that may last several years. Since inter-model variations in persistence/memory re-emergence can help explain inter-model variations in initial-value predictability, would intra-model variations in persistence lead to different predictability within a given model? Would predictability be significantly higher when the system is in a strong re-emergence regime? To explore this question, we perform two sets of PMEs with a GCM, initializing forecast ensembles from ICs that are taken at different times of both high and low memory re-emergence in a millennium-long control run. To further understand the model results we also build a simple auto-regressive statistical model of SIA that replicates the persistence of sea ice in the GCM and quantify its predictive skill.

3.1 Model and experiment design

We use model output from a 1300-yr equilibrated control integration of the Community Climate System Model version 4 (CCSM4, see Gent et al. 2011, see https://www.earthsystemgrid.org/dataset/ucar.cgd.ccsm4.joc.b40.1850.track1.1deg.006.html). The simulation (hereafter referred to as the control) is run under constant 'pre-industrial' radiative forcing with a CO2 concentration of 280ppm, employs fully coupled atmosphere, ocean, sea ice, and land components with a nominal 1° resolution, and has quasi-equilibrated northern hemisphere ice cover throughout the simulation (see Fig. 2a, b).

We begin by assessing the time-varying persistence of the control. We do this in two ways. First, we calculate the lead correlation r of SIA anomalies for all months (out to lead of 23 months) for consecutive 35 year periods as defined in Eq. 2. Second, we calculate the lead pattern correlation p of sea-ice

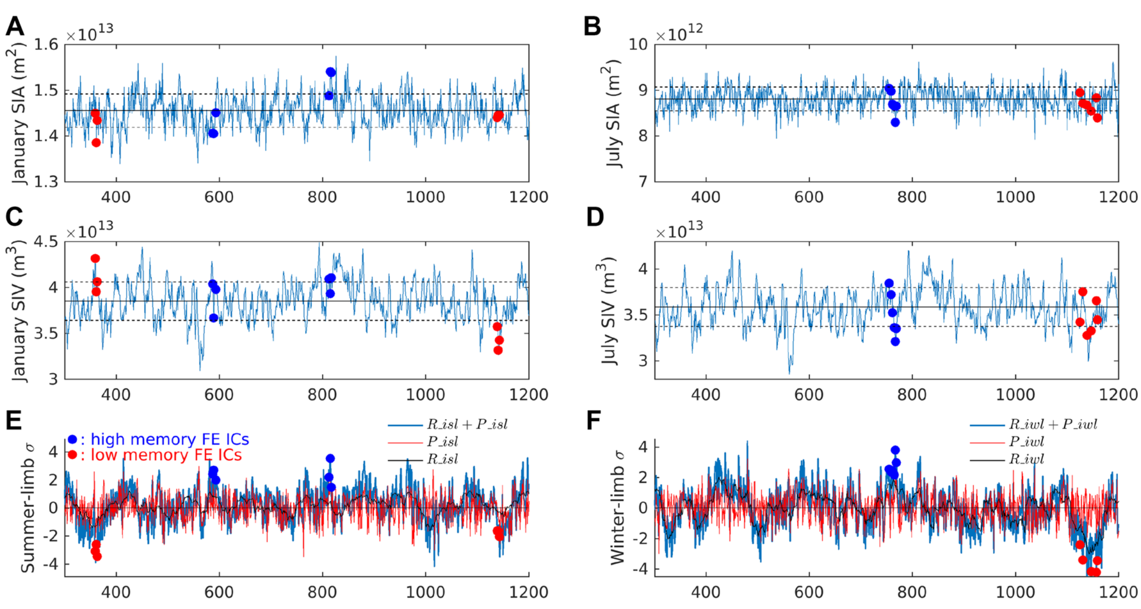

Fig. 2 January SIA (**a**), SIV (**c**) and July SIA (**b**), SIV (**d**) for years 300–1200 in the control, and indices of SIA persistence R_{isl} (black) and SIC pattern correlation P_{isl} (red) and the aggregate memory index $R_{isl} + P_{isl}$ (blue) for the summer-limb (**e**) and $R_{iwl}, P_{iwl}, R_{iwl} + P_{iwl}$ for winter-limb (**f**). The blue and red circles are the years from which the high memory and low memory forecast ensembles are initialized in each PME

concentration (SIC) anomalies for all months (also out to a lead of 23 months) and for all years as defined in Eq. 2.

$$
\begin{aligned}
r_{im}(\tau) &= \frac{cov(a_{i,m}, a_{i,m+\tau})}{\sigma_{a_{i,m}} \sigma_{a_{i,m+\tau}}}, \\
p_{im}(\tau) &= \frac{cov(c_{i,m}, c_{i,m+\tau})}{\sigma_{c_{i,m}} \sigma_{c_{i,m+\tau}}},
\end{aligned}
\tag{2}
$$

where i, m and τ are the year, initial month, and time lead in months respectively, $a_{i,m}$ is the 35-year time series of SIA anomalies centered around year i and month m and $c_{i,m}$ are the sea-ice concentration anomalies in year i and month m.

The time series of r and p offer complementary insights on decadal and annual sea-ice persistence variability respectively, as r is calculated using a multi-decadal time period (35 years) and p is calculated over a seasonal-to-annual period (from a predictor month to a target month at leads of 0 to 23 months). We use them together to pinpoint periods of high and low persistence. Since both metrics are of monthly resolution, to obtain yearly indexes of memory re-emergence R and P we average the lead correlation values along the 'summer-limb' and 'winter-limb' of memory re-emergence as follows

$$
\begin{aligned}
R_{isl} &= \langle r_{im}(\tau) \rangle, P_{isl} = \langle p_{im}(\tau) \rangle \\
R_{iwl} &= \langle r_{im}(\tau) \rangle, P_{iwl} = \langle p_{im}(\tau) \rangle
\end{aligned}
\tag{3}
$$

where $\tau = \{11, 9, 7, 5, 3, 1\}$ for $m = \{3, 4, 5, 6, 7, 8\}$ for the summer-limb (R_{isl}, P_{isl}) and $\tau = \{12, 10, 8, 6\}$ for $m = \{9, 10, 11, 12\}$ for the winter-limb (R_{iwl}, P_{iwl}—see Fig. 3 for a visual representation of these month-lead pairs). These month-lead pairs represent the strongest signal of memory re-emergence that takes place between melt (spring/summer) and posterior freeze-up (fall) months (the summer-limb) and between freeze-up (fall) and posterior melt months (spring/summer) months (the winter-limb, Blanchard-Wrigglesworth et al. 2011a and Bushuk et al. 2015) and are respectively related to the Y2Y Mar and Y2Y Sep metrics described above (note that Y2Y Mar and Y2Y Sep are equal to $r_m(\tau)$ for $\tau = 12$ and $m = 3$ and $m = 9$ respectively, and that performing the APPOSITE analysis shown in Fig. 1 with the limb metrics yields the same results, see Figure S3). Figure 2e, f shows the normalized yearly time-series of area persistence R and pattern correlation P in the control for both limbs of memory re-emergence. In accordance with Bushuk et al. (2015), SIA persistence and pattern correlation fluctuate in magnitude over both short and long timescales. Over the full control run, the non-normalized timeseries of $R + P$ exhibit almost identical variability for both limbs of memory re-emergence (not shown), indicating that the variability in persistence state is comparable for both limbs.

We create overall memory metrics (one for each limb) that capture both SIA persistence and pattern correlation by summing the normalized R and P timeseries. We next select periods

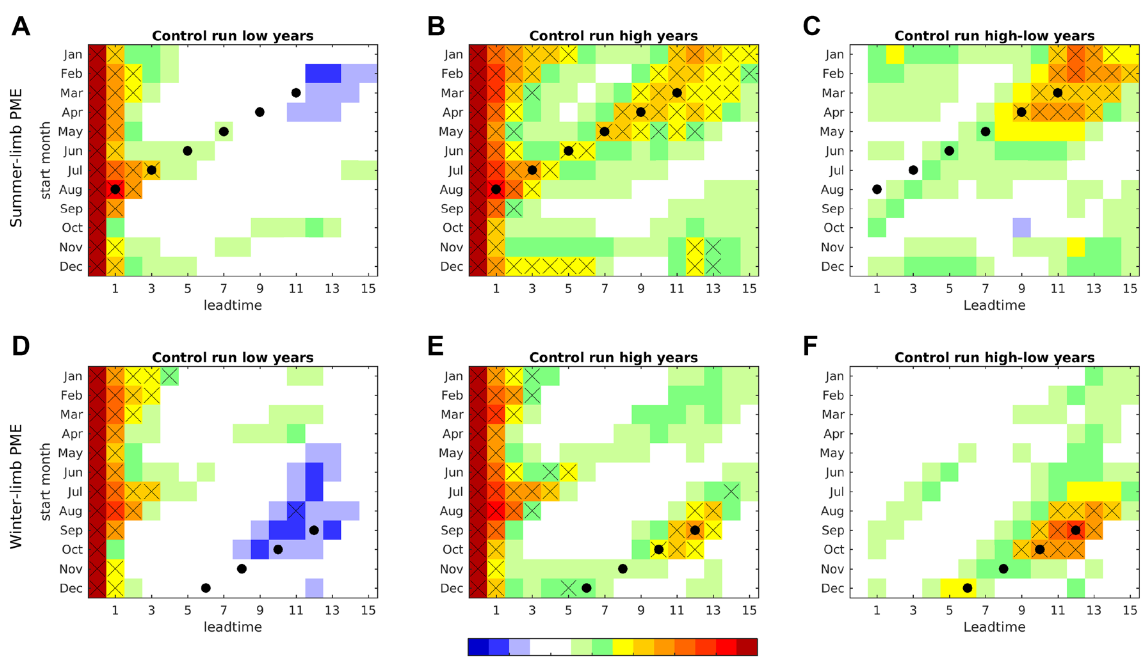

Fig. 3 Lead pattern correlation p in the control averaged over the years following the IC dates used in the high memory and low memory summer-limb PMEs (**a**, **b**), winter-limb PMEs (**d**, **e**) and the difference between the high and low memory in each PME (**c**, **f**). Filled black circles indicate the month pairs m and τ in Eq. 3 along which R_{isl}, P_{isl} (summer-limb PME, top row) and R_{iwl}, P_{iwl} (winter-limb PME, bottom row) are calculated. Xs indicate statistical significance at the 95% threshold, calculated with a t-test in **a**, **b**, **d**, **e** and a z-test using the Fisher transform in **c** and **f**

of high and low memory in the control run from which to initialize a summer-limb PME and a winter-limb PME. For each PME, we produce two sets of forecast ensembles (FEs), each set consisting of 6 individual ensembles initialized from control run years of high and low memory respectively. We initialize the high memory (low memory) summer-limb FEs on January 1 from years 587, 589, 593, 813, 815 and 817 (359, 361, 363, 1139, 1141 and 1143). The high memory (low memory) winter-limb FEs are initialized on July 1 from years 755, 759, 761, 765, 767 and 769 (1125, 1131, 1139, 1147, 1157, 1159, see Fig. 2). Each individual FE is 16-months long, and is composed of 10 members that differ only in their atmospheric ICs, which are taken from consecutive days centered around the start date. Table 1 summarizes the PMEs, and Fig. 3 shows the lead pattern correlation p for the control in the months following the ICs that make up the high and low memory PMEs (averaged over all IC years), illustrating the marked difference in persistence and memory re-emergence between the two sets of ICs. We assess statistical significance using a 95% confidence level. For NRMSE we use an F-test. For ACC and pattern correlation, we use a t-test to assess significance relative to zero, and a z-test using a Fisher's Z-transformation to assess significance of the difference between PME values. For relative entropy we solve for the mean and standard deviation of a population that is significantly different to the climatology at exactly the 95% level in an analogous manner to Blanchard-Wrigglesworth et al. (2011b).

3.2 Results: Pan-Arctic predictability

We begin by evaluating the growth of the ensemble spread by calculating the NRMSE of monthly Arctic SIA and SIV, shown in Fig. 4a, e (Fig. 4i, m) for the summer-limb (winter-limb) PME. As expected from earlier studies (e.g., Blanchard-Wrigglesworth et al. 2011b; Day et al. 2014), the NRMSE of SIA grows faster than that of SIV, indicating higher predictability of SIV over SIA. Importantly, there are no significant differences in NRMSE of SIA and SIV between the high and low memory PMEs, suggesting that predictability skill as quantified by NRMSE is not dependent on the background persistence in the control run from which the PMEs are initialized.

We next consider the anomaly correlation coefficient (ACC) of the PMEs. The ACC is defined as follows:

$$ACC = \frac{\langle (x_{ij} - \mu_j)(x_{kj} - \mu_j) \rangle_{i,j,k \neq i}}{\langle (x_{ij} - \mu_j) \rangle^2}, \quad (4)$$

where μ_j is the climatological mean at the time of the jth ensemble prediction. An ACC of 0 indicates no predictability. We show the ACC for SIA and SIV in Fig. 4b, f (Fig. 4i, n) for the summer-limb (winter-limb) PMEs. For the summer-limb PME, the high memory FE shows higher SIA ACC values than the low memory FE, indicating higher predictability as quantified by the ACC. Interestingly, the ACC for SIV shows the opposite, lower ACC values for the high memory FE relative to the low memory FE. We further explore this below in the discussion section. For the winter-limb PME, both high and low memory FEs show similar predictability.

While the NRMSE and ACC quantify different aspects of forecast skill (predictability of the spread versus predictability of the anomaly, respectively), a useful metric that can quantify both aspects simultaneously is relative entropy (Kleeman 2002). Relative entropy (RE), in its univariate form, is defined as follows:

$$RE = \frac{1}{2}\left[\ln\left(\frac{\sigma_c^2}{\sigma_e^2}\right) + \frac{\sigma_e^2}{\sigma_c^2} + \frac{(\mu_e - \mu_c)^2}{\sigma_c^2} - 1\right], \quad (5)$$

where σ_c and σ_e are standard deviations of the control and PME respectively, and μ_c and μ_e are the mean of the control and PME respectively. The first two and fourth terms in Eq. 5 are known as the *dispersion* component (predictability of the spread) and the third term as the *signal* component of RE (predictability of the mean). RE provides the information content (in bits) of the prediction ensemble relative to the climatological distribution. High RE values indicate high predictability and an RE value of zero indicates no predictability. We show the summer-limb PME RE for SIA for both FEs in Fig. 4c, d and for SIV in Fig. 4g , h. For SIA, the high memory FE shows higher RE values than the low memory FE which results from higher levels in the signal component. For SIV, the low memory FE shows higher RE values than the high memory FE which again results from higher levels in the signal component. These results agree with our findings above using NRMSE and ACC, as only the signal component, associated with deviations from the climatological mean, shows different levels of predictability across the summer-limb high and low memory PMEs. For the winter-limb PME (Fig. 4k, l, o, p), both high and low memory FEs show similar levels of RE and thus no

Table 1 Summary of PMEs

PME	High memory IC FE	Low memory IC FE	Size, start date, length
Summer-limb	587, 589, 593, 813, 815, 817	359, 361, 363, 1139, 1141, 1143	10 members, Jan 1, 16 months
Winter-limb	755, 759, 761, 765, 767, 769	1125, 1131, 1139, 1147, 1157, 1159	10 members, Jul 1, 16 months

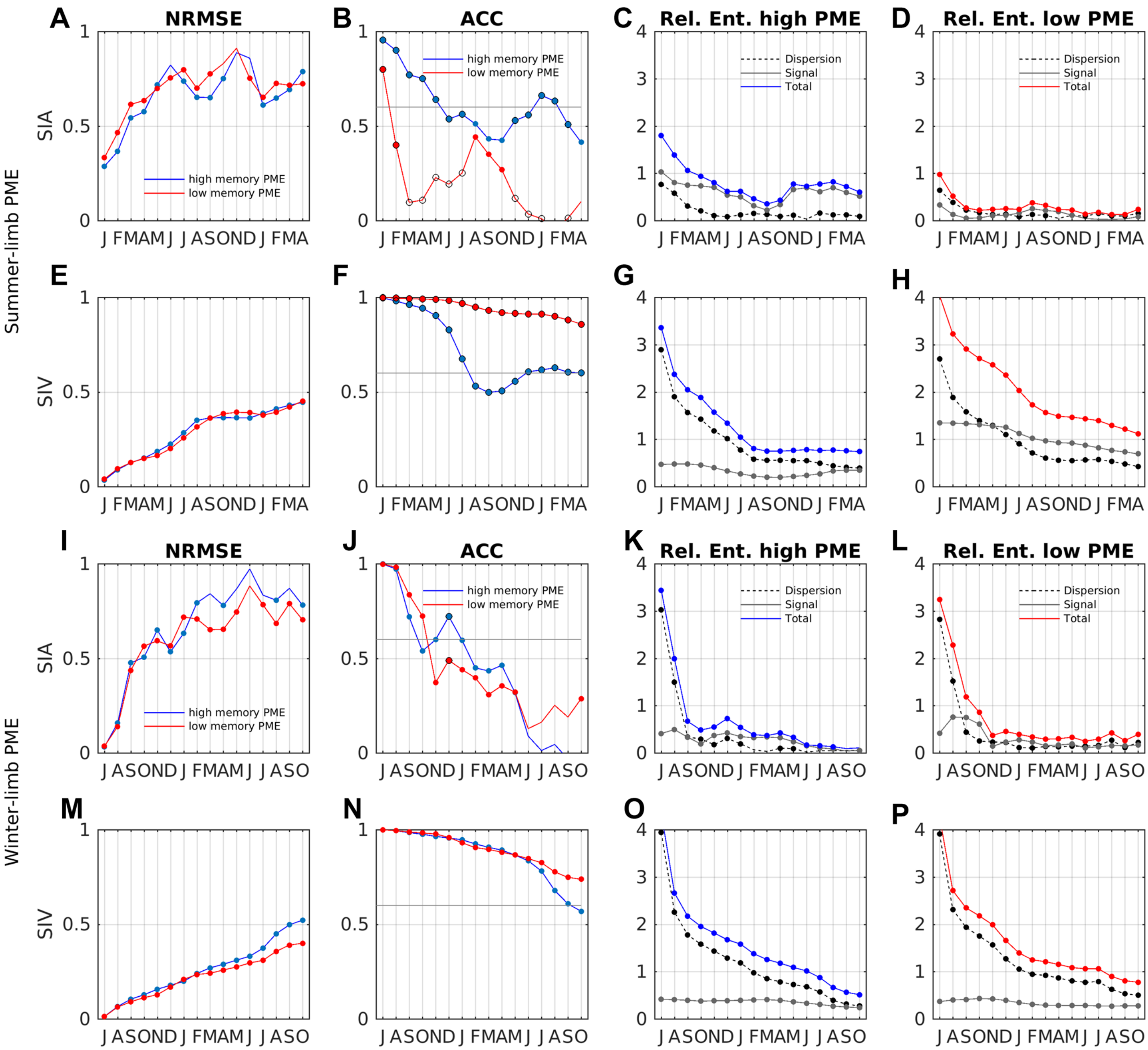

Fig. 4 Predictability metrics for the high and low memory PMEs of SIA and SIV: NRMSE (**a**, **e**, **i**, **m**), ACC (**b**, **f**, **j**, **n**), Relative entropy for the high and low memory PMEs of SIA (C, D, K, L) and SIV (**g**, **h**, **o**, **p**). Filled circles indicate values significantly different to 1 (for NRMSE) or 0 (for ACC and Rel. Ent.) at the 95% threshold. Black circles indicate PME ACC values that are significantly different to each other

differences in predictability, in line with the NRMSE and ACC results above.

We next show the lead pattern correlation p values from the PMEs in Fig. 5, calculated by averaging p across all ensemble members. In the summer-limb PME, there are only minor differences across both high and low memory FEs, with lead pattern correlation values only slightly higher for some months during the high memory FE. Compared to the control IC years, re-emergence in the high memory FE is not as strong as re-emergence in the high memory control IC years (Fig. 5b, e), and re-emergence in the low memory FE is slightly stronger than re-emergence in the low control IC years, when it is zero or even slightly negative (Fig. 5a, d). Contrasting the differences between both (Fig. 5c, f) it is clear that most of the pattern correlation signal seen at lags of 10–15 months from winter-to-winter in the control IC years is lost in the summer-limb PMEs, and therefore is not predictable. The lead pattern correlation p values for the winter-limb PME show even greater similarity between the high memory and low memory FEs (Fig. 5g–i), indicating no differences in predictability, and practically all the pattern correlation signal

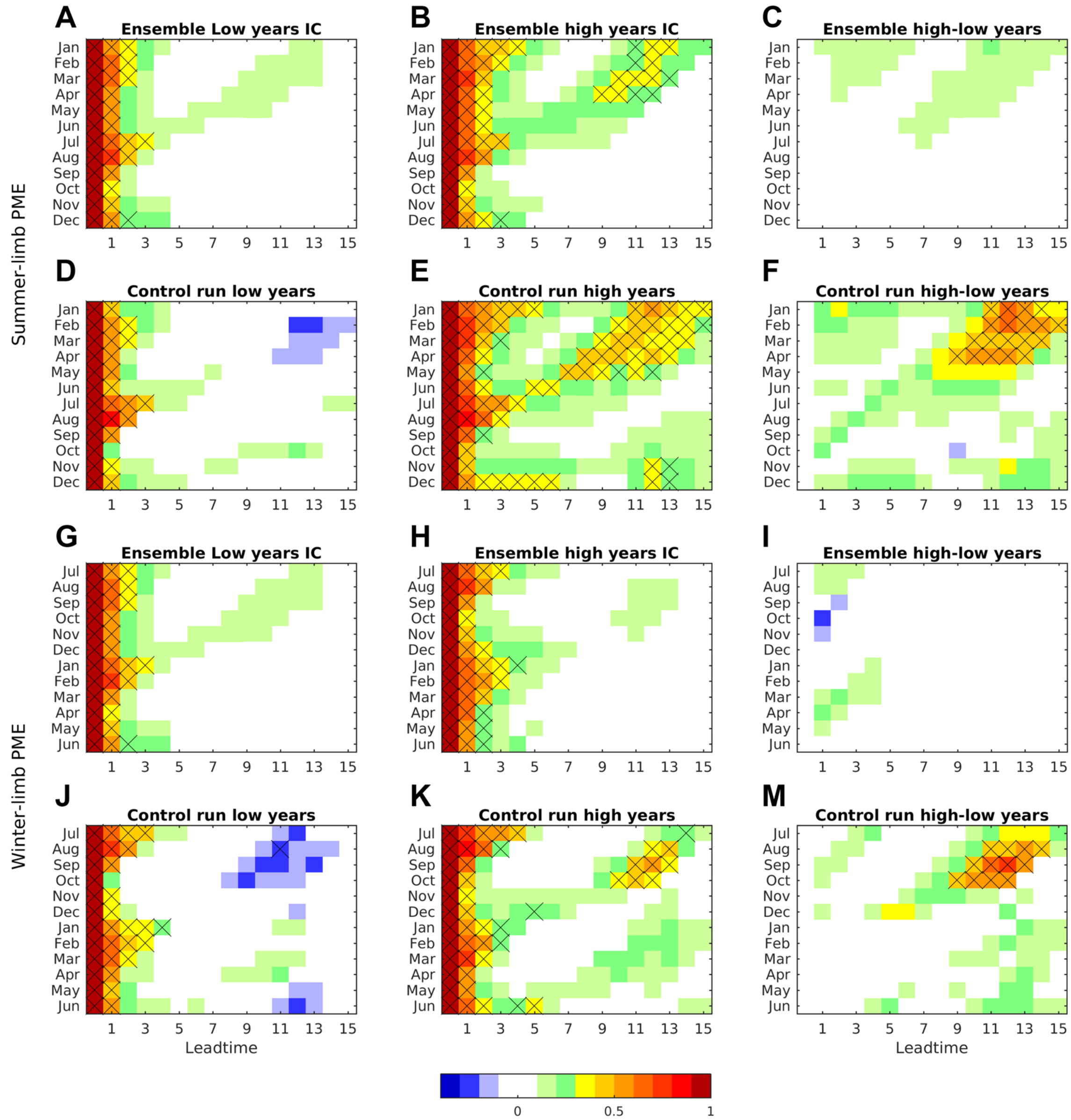

Fig. 5 Lead pattern correlation in the predictability ensemble low years (**a**), high years (**b**), difference (**c**) and in the control simulation years from which initial conditions for the predictability experiments are drawn [low (**d**), high (**e**) and difference (**f**)]. Xs indicate statistical significance at the 95% threshold

from summer-to-summer present in the control is lost in the FEs (contrast Fig. 5i, l).

3.3 Results: regional predictability

We now turn our attention to SIA predictability in five regions: the Barents/Kara, Greenland/Iceland/Norwegian (GIN) and Labrador Seas in the Atlantic, and the Okhotsk and Bering Seas in the Pacific. We define the geographic boundaries of these seas following (Bushuk et al. 2015). To capture both predictability of the spread and signal we just show the RE for the summer-limb PME in Fig. 6. Higher RE in the high memory FE is seen in the Atlantic seas, particularly in the Barents and GIN Seas. As with pan-Arctic SIA, the difference in RE originates from larger values in the signal component. By contrast, the Bering and Okhotsk Seas show no obvious differences in RE between the two FEs. In Fig. 7 we show the RE for the winter-limb PME, in this case focusing on the Atlantic seas and the Arctic basin seas which show summer sea ice variability. As is the case with the summer-limb PME, regional predictability is similar in both high and low memory FEs, particularly at long lead times when we might expect re-emergence. Small differences arise in the signal component of RE predictability in the first few

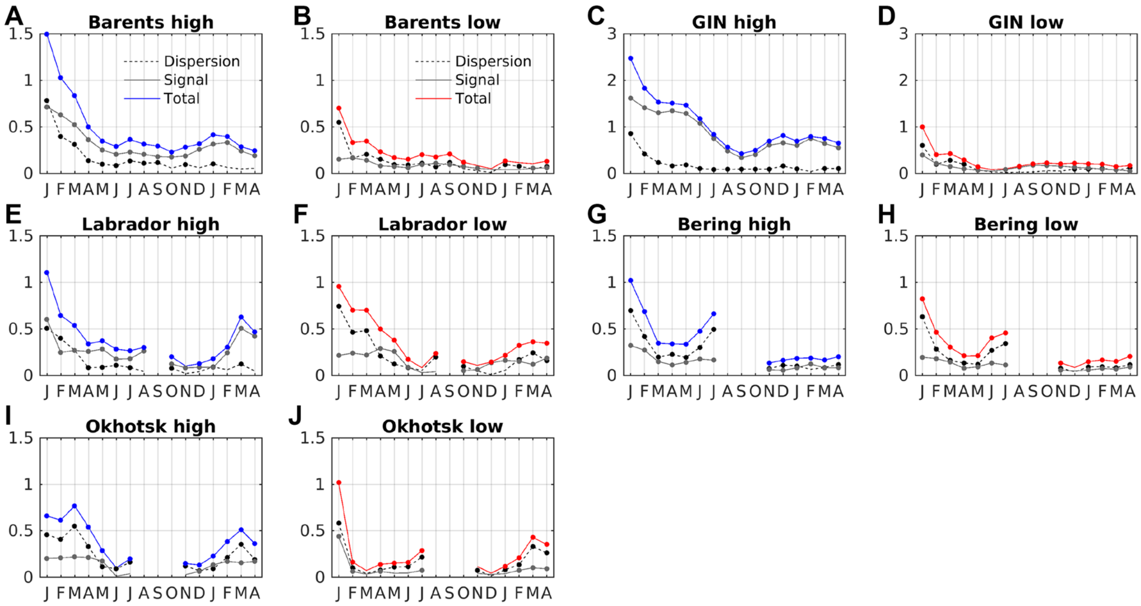

Fig. 6 Relative entropy in high and low memory summer-limb PMEs (denoted as high and low) in the Barents, GIN, Labrador, Bering and Okhotsk seas. Months in which the climatological SIA is close to zero are left blank (August through October in the Bering and Okhotsk seas, September in the Labrador sea)

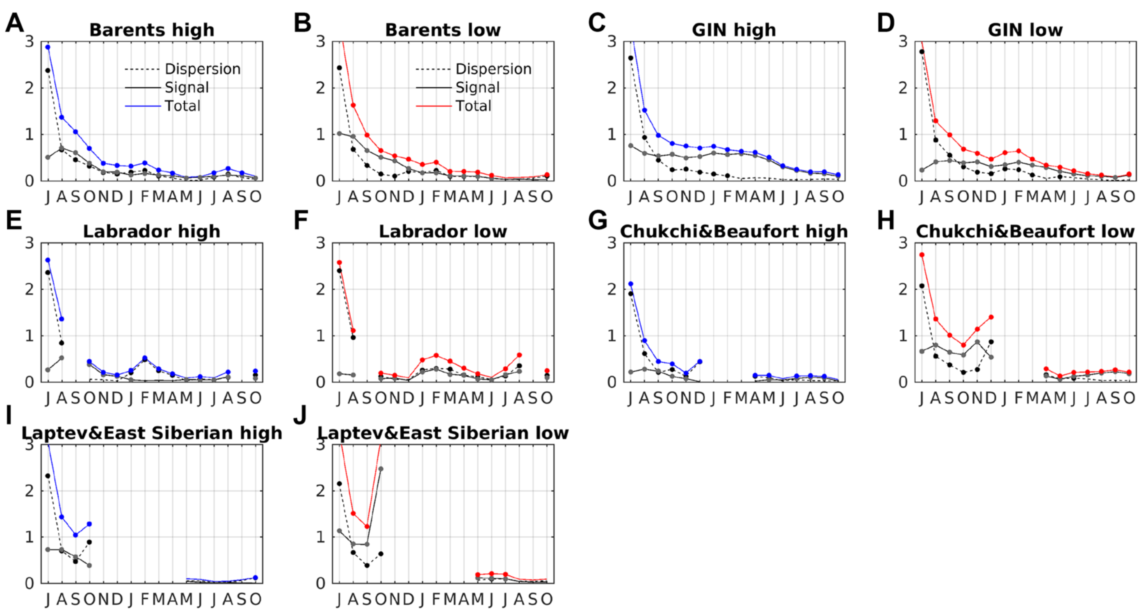

Fig. 7 As in Fi. 6, but for the winter-limb PME

months of the integration in the GIN, Chukchi & Beaufort, and Latpev & East Siberian Seas.

3.4 Insights from a simple statistical model

To further explore the *a priori* un-expected result that NMRSE predictability in the high and low memory FEs is not significantly different, we calculate the benchmark skill provided by a simple statistical forecast of synthetic SIA time series that reflect the high and low persistence regimes of the control.

To analyze the summer limb predictability, we first construct two sets of synthetic 16-month long time-series using an AR(1) model in which a synthetic SIA anomaly at time t is computed as follows

$$x_{dec} = \sigma_{dec} z, \quad x_t = \beta_t x_{dec} + e, \quad \beta_t = r_{dec,t} \frac{\sigma_t}{\sigma_{dec}} \tag{6}$$

where x_t is the SIA anomaly at time t, β_t is the auto-correlation coefficient between December and the month at time t weighted by the ratio of standard deviation of SIA anomalies in month t and December, and $e = \sigma_t z$, where σ_t is the standard deviation of SIA in month t and z is taken from a normal distribution with unit variance and zero mean (we select December since the summer-limb FEs are initialized on January 1). We compute two different sets of synthetic time series, one that takes the lead correlation $r_{dec,t}$ values in Eq. 6 from the highest and lowest $r_{im}(\tau)$ (for $m = 12$ and $\tau = 1{:}16$) values calculated from consecutive 100-year long time-series taken from the control.

Each set consists of 2000 independent iterations of Eq. 6 (i.e., 2000 16 month-long time series for each set). Next, we calculate a simple persistence forecast of all time series in each set as follows:

$$y_t = \beta_t x_{dec}, \tag{7}$$

where y_t is the area forecast at lag t months from December. We also perform an analogous procedure to analyze winter limb predictability, computing two sets of synthetic time series, corresponding to high and low winter limb autocorrelation regimes, respectively. Specifically, we compute lead correlations between June and month t, and define high and low memory regimes based on the highest and lowest $r_{im}(\tau)$ ($m = 6$ and $\tau = 1{:}16$) values calculated from consecutive 100 year long time series taken from the control.

We show the NRMSE of the simple persistence forecasts for each set in Fig. 8. The NRMSE of the high persistence set is consistently lower than the NMRSE of the low persistence set, which loses all predictability by the first spring. This result suggests that one would indeed expect significantly different predictability skill to arise from the different persistence/memory re-emergence regimes seen in the control, were these regimes predictable. We also show in Fig. 8 the NRMSE values from the summer-limb FEs above. The difference between the synthetic and the FE NRMSEs shows that there is enhanced predictability beyond that expected solely from persistence or memory re-emergence, echoing results in Day et al. (2014), and confirming that the PME predictability is independent of the IC state space with regards to persistence.

4 Discussion and conclusions

Inter-model spread in initial value sea-ice predictability is coupled to inter-model spread in sea-ice persistence across the APPOSITE set of models (Fig. 1). This link is modulated seasonally by the two different mechanisms of memory re-emergence: Y2Y March autocorrelation is coupled to predictability during the winter sea ice maximum months, whereas Y2Y September autocorrelation is coupled to predictability outside the winter sea ice maximum months. In the APPOSITE models, and to some extent in the CMIP5 models, these two autocorrelation metrics are

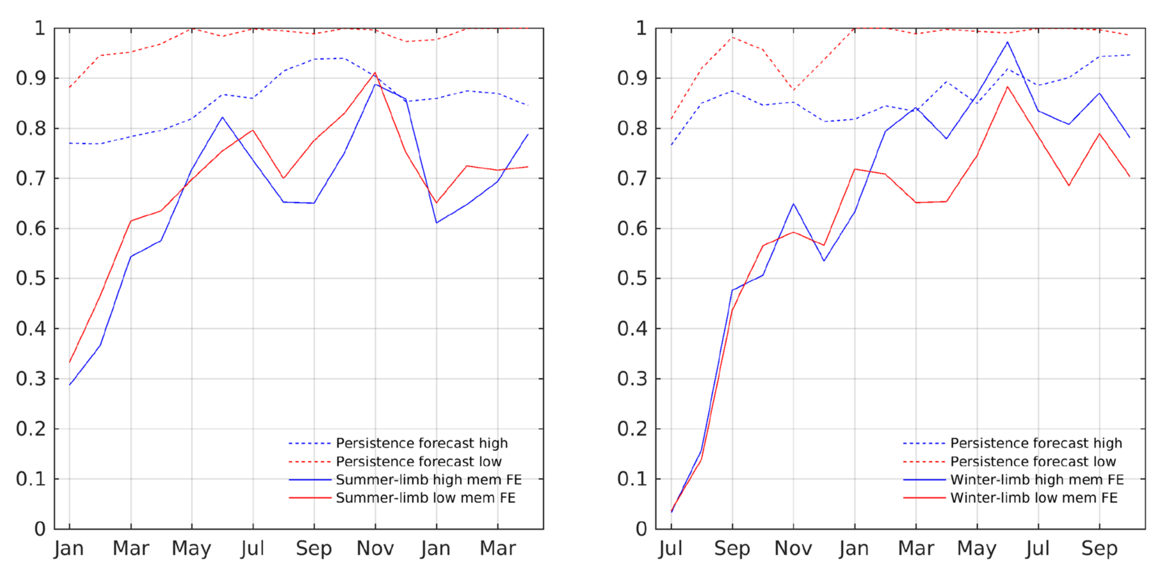

Fig. 8 **a** NRMSE of a lead persistence forecast of synthetic high and low persistence set (dashed), together with the summer-limb PME high and low memory FE NRMSE values (taken from Fig. 4a, b) as in **a**, but for the winter-limb PME

largely uncorrelated, helping explain why some models have either higher winter than summer predictability, vice versa, or no seasonal fluctuations in predictability. Since the Y2Y autocorrelations in the APPOSITE and CMIP5 models are higher than in detrended observations (a feature also seen across CMIP3, not shown), this poses the question whether predictability estimates from GCMs are biased high relative to the real world predictability, which could help explain a portion of the predictability gap between PMEs and real-world forecasts (Blanchard-Wrigglesworth et al. 2015). Identifying why GCMs show different persistence characteristics and relating the relevant model processes to observed processes should be of high priority.

These results should be treated with some caution given the relatively small number of models contributing to APPOSITE, and should encourage current efforts to expand the number of models contributing to APPOSITE (Day, personal communication). We note that the link between persistence and predictability is also present in Antarctica (see Figure S4), where interestingly only the winter-to-winter autocorrelation (Y2Y Sep in the austral calendar) is linked to winter predictability, possibly a result of the small summer sea-ice cover in Antarctica and the lack of summer memory re-emergence (note the much lower Y2Y Mar values in Antarctica in Fig. S4 compared to the Y2Y Sep values for the Arctic) and summer predictability in general, which was also found by Holland et al. (2013).

In contrast, intra-model temporal variations in persistence do not have a clear impact on predictability in CCSM4. We have used initial conditions taken from a long CCSM4 control run to initialize perfect-model forecasts at times when the control run was in high and low persistence states. Forecasts of the ensemble spread, as quantified by the NRMSE and dispersion component of RE, are similar in the high and low memory FEs for both the summer and winter limb PMEs (Fig. 4), and the spatial pattern correlations lose most of the re-emergence signal present in the ICs (Fig. 5). Given the range in centennial persistence of the Y2Y autocorrelation metrics in CCSM4 (Fig. 1e), if the inter-model relationship between persistence and predictability applied to the intra-model relationship, one would expect noticeable differences in intra-model predictability. This result suggests that perfect model NRMSE predictability in unforced (constant radiative forcing) GCM simulations is robust across initial conditions taken from different persistence regimes and complements (Day et al. 2016) who found the same result by comparing NRMSE predictability in high and low SIV initial conditions. By extension, differences in NRMSE predictability across different models are likely to be robust and represent the impact of different model physics on predictability.

On the other hand, predictability of the signal, as quantified by the ACC and signal component of RE, is dependent on the initial conditions. We have found higher ACC SIA predictability in the high memory FE relative to the low memory FE in the summer-limb PME but, seemingly counterintuitively, lower ACC SIV predictability in the high memory FE relative to the low memory FE in the same PME. We note that Day et al. (2016) found ACC predictability to be IC-dependent, with higher ACC when the ICs were anomalously high and low, reflecting a statistical artefact of the ACC metric. Considering the SIA and SIV anomalies in the summer-limb PMEs' ICs, it is likely that this feature of the ACC metric plays a role in our findings: the high memory FE summer-limb PME ICs contain more anomalous SIA initializations than the low memory PME (4 ICs vs 1 IC at least 1 standard deviation away from climatology, Fig. 2a), and the reverse is true for SIV (2 vs 4 anomalous ICs, see Fig. 2c). Our findings in the winter-limb PME, where both FEs have similar ACC values and the ICs sample similar average SIA and SIV conditions (Fig. 2b, d), support this interpretation of ACC predictability. These findings suggest that PM skill comparisons should be made using spread-based metrics (such as NRMSE), as these metrics are much less sensitive to start date sampling biases than the ACC metric.

To summarize, we have found that in CCSM4, persistence does not have a clear impact on predictability. Why is it that within one year the FEs lose most of the persistence and pattern correlations imprinted in the ICs? As discussed in Bushuk et al. (2015), one of the driving mechanisms of coherent patterns of SIC anomalies across separate, remote seas are planetary scale modes of atmospheric variability that induce synchronous sea-ice anomalies across remote regions. Considering the regional predictability results from individual seas in the summer-limb PME, whereby only the Atlantic seas show differences between high/low memory FEs, it is likely that (1) the hemispheric modes of atmospheric forcing are mostly not predictable at seasonal-to-yearly timescales, and (2) the upper ocean can only force annual re-emergence of sea-ice anomalies in the Atlantic seas from winter-to-winter. The key role of the atmosphere on the timescale of these seasonal-to-annual forecasts has recently been documented by Tietsche et al. (2016), who found that the atmosphere forced the majority of chaotic error growth in the first year of a forecast.

Acknowledgements We thank Cecilia Bitz and Dimitrios Giannakis for discussions on the work presented, Jonny Day for help with accessing the APPOSITE data, and Jen Kay, David Bailey and NCAR's Polar Climate Working Group for assisting with computing resources. EBW was supported by the Office of Naval Research grants N00014-13-1-0793 and N00014-17-1-2986, and MB was supported by NOAA's Climate Program Office, Climate Variability and Predictability Program (award GC15-504).

References

Blanchard-Wrigglesworth E, Armour K, Bitz CM, deWeaver E (2011) Persistence and inherent predictability of Arctic sea ice in a GCM ensemble and observations. J Clim. https://doi.org/10.1175/2010JCLI3775.1

Blanchard-Wrigglesworth E, Bitz CM, Holland MH (2011) Influence of initial conditions and climate forcing on predicting Arctic sea ice. Geophys Res Lett. https://doi.org/10.1029/2011GL048807

Blanchard-Wrigglesworth E, Cullather R, Wang W, Zhang J, Bitz C (2015) Model forecast skill and sensitivity to initial conditions in the seasonal sea ice outlook. Geophys Res Lett 42(19):8042–8048

Bushuk M, Giannakis D (2015) Sea-ice reemergence in a model hierarchy. Geophys Res Lett 42:5337–5345

Bushuk M, Giannakis D, Majda AJ (2015) Arctic sea-ice reemergence: the role of large-scale oceanic and atmospheric variability. J Clim 28:5477–5509

Bushuk M, Msadek R, Winton M, Vecchi G, Gudgel R, Rosati A, Yang X (2017) Skillful regional prediction of Arctic sea ice on seasonal timescales. Geophys Res Lett 44:4953–4964. https://doi.org/10.1002/2017GL073155

Collins M (2002) Climate predictability on interannual to decadal time scales: the initial value problem. Clim. Dyn. 19:671–692

Day J, Tietsche S, Hawkins E (2014) Pan-Arctic and regional sea ice predictability: initialisation month dependence. J Clim 27(12):4371–4390

Day JJ, Goessling HF, Hurlin WJ, Keeley SP (2016) The Arctic predictability and prediction on seasonal-to-interannual timescales (APPOSITE) data set version 1. Geosci Model Dev 9(6):2255

Fetterer F, Knowles K, Meier W, Savoie M (2002) Sea ice index. https://nsidc.org/data/docs/noaa/g02135_seaice_index/ (updated 2016)

Gent P, Danabasoglu G, Donner L, Holland M, Hunke E, Jayne S, Lawrence D, Neale R, Rasch P, Vertenstein M, Worley P, Yang ZL, Zhang M (2011) The community climate system model version 4. Jo Clim 24(19):4973–4991

Germe A, Chevallier M, y Mélia DS, Sanchez-Gomez E, Cassou C, (2014) Interannual predictability of Arctic sea ice in a global climate model: regional contrasts and temporal evolution. Clim Dyn 43(9–10):2519–2538

Guemas V, Blanchard-Wrigglesworth E, Chevallier M, Day JJ, Déqué M, Doblas-Reyes FJ, Fučkar NS, Germe A, Hawkins E, Keeley S, Koenigk T (2016) A review on Arctic sea-ice predictability and prediction on seasonal to decadal time-scales. Quart J R Meteorol Soc 142(695):546–561

Hebert DA, Allard RA, Metzger EJ, Posey PG, Preller RH, Wallcraft AJ, Phelps MW, Smedstad OM (2015) Short-term sea ice forecasting: an assessment of ice concentration and ice drift forecasts using the us navy's arctic cap nowcast/forecast system. J Geophys Res: Oceans 120(12):8327–8345

Holland MM, Bailey DA, Vavrus S (2010) Inherent sea ice predictability in the rapidly changing Arctic environment of the Community Climate System Model, version 3. Clim Dyn. https://doi.org/10.1007/s00382-010-0792-4

Holland MM, Blanchard-Wrigglesworth E, Kay J, Vavrus S (2013) Initial-value predictability of Antarctic sea ice in the Community Climate System Model 3. Geophys Res Lett 40(10):2121–2124

Jung T, Gordon ND, Bauer P, Bromwich DH, Chevallier M, Day JJ, Dawson J, Doblas-Reyes F, Fairall C, Goessling HF (2016) Advancing polar prediction capabilities on daily to seasonal time scales. Bull Am Meteorol Soc 97(9):1631–1647

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

Koenigk T, Mikolajewicz U (2009) Seasonal to interannual climate predictability in mid and high northern latitudes in a global coupled model. Clim Dyn. https://doi.org/10.1007/s00382-008-0419-1

Kumar A, Peng P, Chen M (2014) Is there a relationship between potential and actual skill? Mon Weather Rev 142(6):2220–2227

Merryfield W, Lee WS, Wang W, Chen M, Kumar A (2013) Multi-system seasonal predictions of Arctic sea ice. Geophys Res Lett 40(8):1551–1556

Msadek R, Vecchi G, Winton M, Gudgel R (2014) Importance of initial conditions in seasonal predictions of Arctic sea ice extent. Geophys Res Lett 41(14):5208–5215

Peterson KA, Arribas A, Hewitt HT, Keen AB, Lea DJ, McLaren AJ (2015) Assessing the forecast skill of Arctic sea ice extent in the GloSea4 seasonal prediction system. Clim Dyn 44(1–2):147–162

Petty AA, Schröder D, Stroeve J, Markus T, Miller J, Kurtz N, Feltham D, Flocco D (2017) Skillful spring forecasts of September Arctic sea ice extent using passive microwave sea ice observations. Earth's Future 5(2):254–263

Schröder D, Feltham DL, Flocco D, Tsamados M (2014) September Arctic sea-ice minimum predicted by spring melt-pond fraction. Nat Clim Change 4(5):353

Sigmond M, Fyfe J, Flato G, Kharin V, Merryfield W (2013) Seasonal forecast skill of Arctic sea ice area in a dynamical forecast system. Geophys Res Lett 40(3):529–534

Tietsche S, Day JJ, Guemas V, Hurlin WJ, Keeley SPE, Matei D, Msadek R, Collins M, Hawkins E (2014) Seasonal to interannual Arctic sea ice predictability in current global climate models. Geophys Res Lett 41(3):1035–1043

Tietsche S, Hawkins E, Day JJ (2016) Atmospheric and oceanic contributions to irreducible forecast uncertainty of Arctic surface climate. J Clim 29(1):331–346

Tietsche S, Notz D, Jungclaus JH, Marotzke J (2013) Predictability of large interannual Arctic sea-ice anomalies. Clim Dyn 41(9–10):2511–2526

Wang W, Chen M, Kumar A (2013) Seasonal prediction of Arctic sea ice extent from a coupled dynamical forecast system. Mon Weather Rev 141(4):1375–1394

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