Subseasonal-to-Seasonal Arctic Sea Ice Forecast Skill Improvement from Sea Ice Concentration Assimilation

YONG-FEI ZHANG,[a,b] MITCHELL BUSHUK,[a,c] MICHAEL WINTON,[a] BILL HURLIN,[a] THOMAS DELWORTH,[a] MATTHEW HARRISON,[a] LIWEI JIA,[a,c] FEIYU LU,[b] ANTHONY ROSATI,[a] AND XIAOSONG YANG[a]

[a] *National Oceanic and Atmospheric Administration/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey*
[b] *Atmospheric and Oceanic Sciences Program, Princeton University, Princeton, New Jersey*
[c] *University Corporation for Atmospheric Research, Boulder, Colorado*

(Manuscript received 16 July 2021, in final form 18 March 2022)

ABSTRACT: The current GFDL seasonal prediction system, the Seamless System for Prediction and Earth System Research (SPEAR), has shown skillful prediction of Arctic sea ice extent with atmosphere and ocean constrained by observations. In this study we present improvements in subseasonal and seasonal predictions of Arctic sea ice by directly assimilating sea ice observations. The sea ice initial conditions from a data assimilation (DA) system that assimilates satellite sea ice concentration (SIC) observations are used to produce a set of reforecast experiments (IceDA) starting from the first day of each month from 1992 to 2017. Our evaluation of daily sea ice extent prediction skill concludes that the SPEAR system generally outperforms the anomaly persistence forecast at lead times beyond 1 month. We primarily focus our analysis on daily gridcell-level sea ice fields. SIC DA improves prediction skill of SIC forecasts prominently in the June-, July-, August-, and September-initialized reforecasts. We evaluate two additional user-oriented metrics: the ice-free probability (IFP) and ice-free date (IFD). IFP is the probability of a grid cell experiencing ice-free conditions in a given year, and IFD is the first date on which a grid cell is ice free. A combined analysis of IFP and IFD demonstrates that the SPEAR model can make skillful predictions of local ice melt as early as May, with modest improvements from SIC DA.

KEYWORDS: Arctic; Sea ice; Data assimilation; Seasonal forecasting

1. Introduction

The rapid decline in Arctic summer sea ice threatens local communities and ecosystems and also creates economic opportunities for marine fishing, shipping and resource extraction. Subseasonal-to-seasonal forecasts of Arctic sea ice hence can aid mitigation of impacts and provide valuable planning information. The current generation of dynamical climate prediction systems have been shown to skillfully predict Arctic sea ice cover (e.g., Sigmond et al. 2013; Wang et al. 2013; Chevallier et al. 2013; Msadek et al. 2014; Bushuk et al. 2017; Dirkson et al. 2019; Bushuk et al. 2019). These skill evaluations have been mostly focused on monthly area-integrated sea ice variables, such as sea ice extent (SIE) or sea ice area (SIA), rather than local and daily fluctuations.

Zampieri et al. (2018) was the first study to assess predictions of Arctic sea ice on subseasonal time scales using daily data from a subseasonal-to-seasonal (S2S) prediction database (Vitart et al. 2017). They found that coupled forecast systems exhibit a large range of skill in predicting the sea ice edge position and that systems with direct assimilation of sea ice observations show better skill. Wayand et al. (2019) assessed the prediction skill for gridcell-level SIE of several dynamical models using daily data, also finding a wide range of skill across models and finding that most models beat the persistence forecast for lead times beyond a few weeks. Liu et al. (2018) analyzed NCEP's Climate Forecast System, version 2, gridcell-level sea ice concentration (SIC) using weekly data and found high skill in the marginal ice zones and that Arctic-averaged SIC skill exceeds the anomaly persistence forecast for lead times beyond a few weeks.

The Sea Ice Prediction Network has started collecting submissions of full fields of predicted SIC, sea ice probability [ice-free probability (IFP)], and ice-free date (IFD; Bhatt et al. 2020). Their report highlighted the difficulty of forecasting the timing of melt, given the large uncertainty among the participating models. Sigmond et al. (2016) evaluated ice retreat and advance date predictions in the Canadian Seasonal to Interannual Prediction System and showed that their system provides additional value beyond the anomaly persistence forecast for all initialization months. However, their system also has mean model biases that lead to reduced forecast skill. Dirkson et al. (2021) applied ensemble calibration methods to the same dynamical forecasting system and demonstrated a boost in skill for both IFP and IFD.

The sources of Arctic sea ice predictability at subseasonal-to-seasonal time scales mainly come from the persistence and reemergence of SIC, sea ice thickness (SIT), and upper-ocean heat content anomalies (Guemas et al. 2016; Blanchard-Wrigglesworth et al. 2011). Capitalizing on these sources of predictability requires accurate sea ice and ocean initial conditions for forecasts, which is one of the major challenges in Arctic sea ice prediction (Liu et al. 2019). Efforts have been made in the past two decades at adopting various data assimilation methods to constrain sea ice initial conditions with

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-21-0548.s1.

Corresponding author: Yong-Fei Zhang, yongfeiz@princeton.edu

DOI: 10.1175/JCLI-D-21-0548.1

direct sea ice observations. Most of the work has been focused on assimilating SIC observations (e.g., Lindsay and Zhang 2006; Lisæter et al. 2003; Stark et al. 2008; Kimmritz et al. 2018; Zhang et al. 2021), with fewer studies assimilating SIT observations (Chen et al. 2017; Fritzner et al. 2019; Xie et al. 2016; Blockley and Peterson 2018).

No comprehensive evaluation of the benefits of sea ice data assimilation on sea ice prediction from subseasonal-to-seasonal time scales has been made, however, to the authors' knowledge. Van Woert et al. (2004) looked at May-initialized forecasts at the hourly time scale. Caya et al. (2010) conducted short-term (48 h) forecast in an extreme winter. Massonnet et al. (2015) conducted winter-initialized hindcast experiments to explore the potential benefits of improved sea ice initial conditions at seasonal time scale. Kimmritz et al. (2019) conducted reforecast experiments throughout the year every other month, finding that regional SIE skill (particularly in summer) is improved up to four months as a result of SIC DA. Their DA frequency was low (once per month) and hence did not explore the full advantage of SIC DA. They focused their evaluation on monthly regional SIE.

Our preceding work (Zhang et al. 2021) assimilated satellite-retrieved SIC observations in GFDL's new seasonal prediction system, the Seamless System for Prediction and Earth System Research (SPEAR) and suggested that improved initial conditions of SIC and SIE would enhance the prediction skill of Arctic sea ice at lead times of 0–2 months. In this study, we test this hypothesis and investigate how much improvement SIC DA has on predicting different sea ice variables in the Arctic, including the area-integrated SIE, gridcell-level SIC, IFP, and IFD. We describe the GFDL prediction system, experimental design of the reforecasts, observations used to evaluate model results, evaluation metrics, and reference forecasts in section 2. An Arctic sea ice prediction skill assessment from the reforecast experiments is presented in section 3. We summarize and discuss our findings in section 4.

2. Methods

a. The SPEAR system

The recently developed GFDL seasonal prediction system, SPEAR, is used in this study. The SPEAR forecast system employs the newest component models from GFDL (Delworth et al. 2020): the AM4 atmosphere model, MOM6 ocean model, LM4 land model, and SIS2 sea ice model. It has two versions: SPEAR_LO and SPEAR_MED, which share the same nominal 1° ocean and sea ice resolution. SPEAR_LO uses 1° resolution in the atmosphere/land components and SPEAR_MED uses 0.5° resolution in these components. The two versions show very similar prediction skill for Arctic sea ice cover (Bushuk et al. 2022); thus, SPEAR_LO is used in this study for computational considerations.

The SPEAR seasonal prediction system uses initial conditions from two separate assimilation experiments. The ocean initial conditions are from an ocean DA run performed with SPEAR_LO that assimilates observations of sea surface temperature (SST), and subsurface temperature and salinity data from various sources (Lu et al. 2020). The atmosphere, land, and sea ice initial conditions are from an atmosphere and SST nudging run of SPEAR_LO in which the wind, temperature, and humidity are nudged to the NOAA/NCEP Climate Forecast System Reanalysis (CFSR), and SST is nudged toward the Optimum Interpolation Sea Surface Temperature (OISST). The SSTs in sea ice covered grid cells (SIC $>$ 30% in OISST) are nudged toward the seawater freezing point calculated based on the model-predicted local salinity to avoid using the low-quality SST observations under sea ice.

No sea ice observations are directly assimilated in these two SPEAR_LO assimilation experiments. The assimilation or nudging of SST observations can constrain sea ice conditions well if SIC is 100% or 0%. In the case where a grid cell is partially covered by sea ice, the SST represents an average of the underice SST and open-ocean SST and does not uniquely determine the SIC; thus, the SIC DA will provide additional information.

b. Two sets of reforecast experiments

Two sets of 15-member reforecast experiments spanning the years from 1992 to 2017, one without (nIceDA) and one with (IceDA) SIC assimilation, are initialized from the first day of each month and run for a 1-yr ensemble forecast. Their experimental configurations are listed in Table 1. The two reforecast suites use identical atmosphere, ocean, and land initial conditions, which are the standard initial conditions used in the SPEAR seasonal predictions as described in section 2a. The only difference between the two reforecast sets are in their sea ice initial conditions: the experiment nIceDA is initialized from the atmosphere and SST-nudged SPEAR_LO run (SPEAR_Nudged hereinafter), and the experiment IceDA uses sea ice initial conditions from a sea ice DA system described next (Zhang et al. 2021; SPEAR_IceDA hereinafter).

The major difference between the two sets of sea ice initial conditions is that SPEAR_IceDA directly assimilates SIC observations and SPEAR_Nudged does not. Satellite-retrieved SIC observations are assimilated every 5 days into SPEAR_IceDA using an ice-ocean model setup forced by the Japanese 55-year Reanalysis (JRA55-do; Tsujino et al. 2018) to generate a sea ice reanalysis product from 1982 to 2017. The SIC DA is done within the framework of the linked SIS2 and the Data Assimilation Research Testbed (DART), with the ensemble adjustment Kalman filter (EAKF). The observation error is set to 10%. The sea ice concentration of each ice-thickness category is the state variable that is updated by EAKF, while the mean thickness of each category remains unchanged. A Gaspari–Cohn half-width of 0.03 radians (~190 km) is applied for horizontal localization. More details can be found in Zhang et al. (2021).

SST in SPEAR_IceDA is nudged toward OISST in the same way as in SPEAR_Nudged but the SIC observations used to determine sea ice edges are different in the two initialization products. SPEAR_Nudged and SPEAR_IceDA also use different atmospheric reanalysis datasets to constrain the atmosphere. These might cause some differences in the sea ice initial conditions. However, as shown in Zhang et al. (2021), the case driven by the CFSR forcing shows very

TABLE 1. The experimental configurations of the reforecast experiments. The configurations are the same except for the sea ice initial conditions.

	nIceDA	IceDA
Ocean model	MOM6; 1.0° (0.5° in the Arctic), 75 vertical levels	MOM6; 1.0° (0.5° in the Arctic), 75 vertical levels
Sea ice model	SIS2; 1.0° (0.5° in the Arctic), 5 category ITD	SIS2; 1.0° (0.5° in the Arctic), 5 category ITD
Atmosphere model	AM4; 1.0°, 33 vertical levels	AM4; 1.0°, 33 vertical levels
Land model	LM4; 1.0°	LM4; 1.0°
Ocean IC	SPEAR_LO Ocean DA	SPEAR_LO Ocean DA
Sea ice IC	SPEAR_Nudged: atmosphere nudged to CFSR, SST nudged to OISST, and no SIC DA	SPEAR_IceDA: ice-ocean forced by JRA-55do, SST nudged to OISST, and SIC DA
Atmosphere IC	SPEAR_Nudged	SPEAR_Nudged
Land IC	SPEAR_Nudged	SPEAR_Nudged
Reforecast period	1992–2017	1992–2017
Ensemble size	15	15
Initialization dates	First of each month	First of each month
Prediction length	One year	One year

similar skill of regional SIE to (or even slightly higher in some subregions than) the case driven by the JRA-55do, used in initializing SPEAR_IceDA. Their gridcell-level SIC skills are also very similar as evaluated by RMSE (Fig. S1 in the online supplemental material) and ACC (supplemental Fig. S2). Hence, any improvements we identify in IceDA can be attributed to the improved sea ice initial condition from SIC DA.

Both the SPEAR_LO ocean DA experiment and SPEAR_IceDA have 30 ensemble members. Since each ensemble member is equally possible to represent our best knowledge of the truth, we do not pick ensemble members specifically. The first 15 ensemble members are used to initialize the reforecast experiments. The reduction in ensemble size from 30 to 15 is for computational considerations.

c. Observations

We use two observational products of SIC retrieved from the same satellite measurements from the Scanning Multichannel Microwave Radiometer (SMMR) on the satellite *Nimbus-7* and the Special Sensor Microwave Imager (SSM/I) sensors on the Defense Meteorological Satellite Program (DMSP) satellites. We use the daily SIC product produced by the National Snow and Ice Data Center (NSIDC) using the "NASA Team" algorithm, version 1.1 (Cavalieri et al. 1996), which is also the SIC observations assimilated in SPEAR_IceDA. We also use the daily NSIDC SIC product derived using the "Bootstrap" method (Comiso 2017) to evaluate the observational uncertainty that stems from satellite SIC retrieval algorithms.

d. Evaluation skill metrics

The root-mean-square error (RMSE) and detrended or full anomaly correlation coefficient (ACC) are used in this study to evaluate the forecast skill of different sea ice variables. No bias correction is applied to any of the forecast variables. The RMSE averaged over time is defined as

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2},$$

where n is the number of years, i is the year index, x is the model variable, and y is the observation. The ACC is defined as

$$\mathrm{ACC} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}},$$

where n is the number of years, i is the year index, x is the model variable, and y is the observation. For the detrended ACC $\overline{x}$ and $\overline{y}$ are linear trend fits to the data, and for the full ACC they are temporal-mean values.

We detrend the model time series of SIC and SIE for detrended ACC calculations. Instead of removing a linear trend of the whole time series, we remove the trend that only uses past years' data. This avoids the dependence of the anomaly on future data. The time series are detrended as follows, assuming there is no trend in the first 3 years:

$$\begin{aligned} y_i' &= y_i - \sum_{j=1}^{i-1} y_i \quad \text{if} \quad i \le 3 \\ y_i' &= y_i - (a_{i-1}t_i + b_{i-1}) \quad \text{if} \quad i > 3 \end{aligned},$$

where y_i and y_i' are the original and detrended values at year i, respectively; j is the year index; a_{i-1} and b_{i-1} are respectively the linear trend and intercept calculated based on data in the past years; and t_i is the time at year i.

e. Evaluated sea ice variables and the definitions of IFP and IFD

The sea ice variables evaluated in the study are SIC, SIE, IFP, and IFD. The SIE is the most commonly evaluated variable in sea ice forecast studies. Its skill shows how models represent the areally integrated sea ice coverage. The commonly used 15% SIC threshold has been adopted to define a grid cell as ice covered (i.e., SIE = 1). In addition

to SIE, we also evaluate gridcell-level SIC, which provides a local assessment of skill and is also sensitive to SIC anomalies within the ice pack.

We define a grid cell as ice free if its SIC is below 15% for 10 consecutive days, in a given year. The 10-day threshold is selected to match the criteria used by Sigmond et al. (2016). We also find that qualitative resulting forecast skill does not change when we switch to a 3-day threshold. The IFD is defined as the first day when a grid cell encounters the ice-free condition. If a model ensemble member in a given grid cell meets the ice-free condition at the beginning of the forecast or has not met the ice-free condition by the end of the summer, it does not have a valid IFD. However, this information should still contribute to the final IFD to be evaluated against observations. Hence, we define the model-predicted IFD as the median of the IFD of the 15 ensemble members and assign an IFD of negative infinity or positive infinity to those ensemble members that begin ice free or never go ice free, respectively. The ensemble median statistic allows for this information to be incorporated more naturally than the ensemble mean. The model-forecast IFP is defined as the fraction of ensemble members that have SIC $< 15\%$ for 10 consecutive days at some point in the year. For observations, the IFP of a grid cell is 1 if it meets the ice-free condition and is 0 if it does not.

f. Reference forecasts

We use different reference forecasts to evaluate the sea ice metrics. We claim that the model can make skillful predictions of a variable when its skill is better than the reference forecast of that variable at 95% significance level. For SIE and SIC, we construct anomaly persistence forecasts that use the observed anomaly one day prior to the initialization date as predictors. The anomaly is calculated by removing the linear trend of the past years as described in section 2d. The SIE persistence forecast is calculated for each subregion and the SIC persistence forecast each grid cell.

The reference forecasts for IFP and IFD are a time-evolving climatology based on the previous 10 years of sea ice observation data. We define an IFP reference forecast for each forecasting year and each grid cell as the fraction of years that have a valid IFD in the past 10 years. The IFD reference forecast is defined as the median IFD over the past 10 years. For example, the IFP in a grid cell for year 2017 is the number of years that have ever experienced ice-free conditions from 2007 to 2016 divided by 10. Similarly, the IFD for 2017 is the median of the IFDs in years from 2007 to 2016. We refer to these reference forecasts as Clim_10yr below. Considering that the observations have binary IFP, and the difficulty of incorporating nonvalid IFD information, we do not detrend the time series of IFP and IFD for the ACC calculation. However, since the reference forecasts are based on data only from the past 10 years, their time series contain trend information. Hence, the skill compared between the model forecasts and the reference forecasts has incorporated the skill from trends. Comparing the skill of model-predicted IFD and IFP with these reference forecasts allows us to assess whether the SPEAR system has additional skill beyond the long-term trend in these quantities.

3. Results

a. Observational uncertainty

We show the root-mean-square difference (RMSD) and ACC of SIC, IFP, and IFD between the two NSIDC SIC datasets retrieved from the algorithms of NASA Team and Bootstrap in Fig. 1. The largest RMSD of SIC occurs in September (Fig. 1a; figures in other seasons are not shown) and is relatively uniform across the Arctic, with magnitude ranging from 10% to 25%. The large summer uncertainty is likely due to surface melt and increased atmospheric moisture (Meier 2005). The two datasets are consistent in trend and interannual variability leading to their high correlation (Fig. 1b). Figure 1c shows the observed September SIC climatology in the Arctic.

IFP and IFD are much less commonly evaluated by the sea ice prediction community. The RMSD of the IFP observations shows a ring pattern corresponding to the zone of September sea ice variability (Fig. 1d). The two products agree well in the central Arctic region of perennial ice and an enclosing region of perennial open water. The difference shows up in between these two regions, which highlights the sea ice variability zone. The largest differences are found in the Canadian Archipelago. The ACC map shows similar patterns of agreement and difference between the two observational products (Fig. 1e).

We expect that the ice-free condition has large variations in the sea ice peripheral regions where SIC has large interannual variability. Figure 1f shows the observed IFP represented as the fraction of years in which a grid cell meets the ice-free condition in the years from 2008 to 2017. The regions with values between 0 and 0.5 are where less than 50% of years have met the ice-free condition. Spurious skill values may emerge because of small sample sizes in those regions; thus, we mask out grid cells where less than 50% of years have valid IFDs (i.e., meet the ice-free condition). This leaves a slightly smaller Arctic area to be evaluated as shown in Fig. 1i. The two observational datasets agree well in the evaluated area. The largest differences are in the Barents Sea where the RMSD can be 60 days (Fig. 1g) and the ACC as low as 0.7 (Fig. 1h).

b. Regional sea ice extent predictions

Our preceding SIC DA work (Zhang et al. 2021) shows that SPEAR_IceDA produces better interannual variability of SIE than SPEAR_Nudged in summer and speculates that using SPEAR_IceDA as sea ice initial conditions will improve summer predictions of SIE at short lead times. Figure 2 shows the detrended ACC of SIE averaged over initialization months of June, July, August, and September. The September-initialized forecast includes a forecast for October as well, but excluding the September month does not change the overall conclusions (figures not shown). Figure 2 confirms that IceDA shows higher ACC of detrended SIE than nIceDA in the first 60 days in the GIN Seas, Laptev, Chukchi, and Beaufort Seas, Canadian

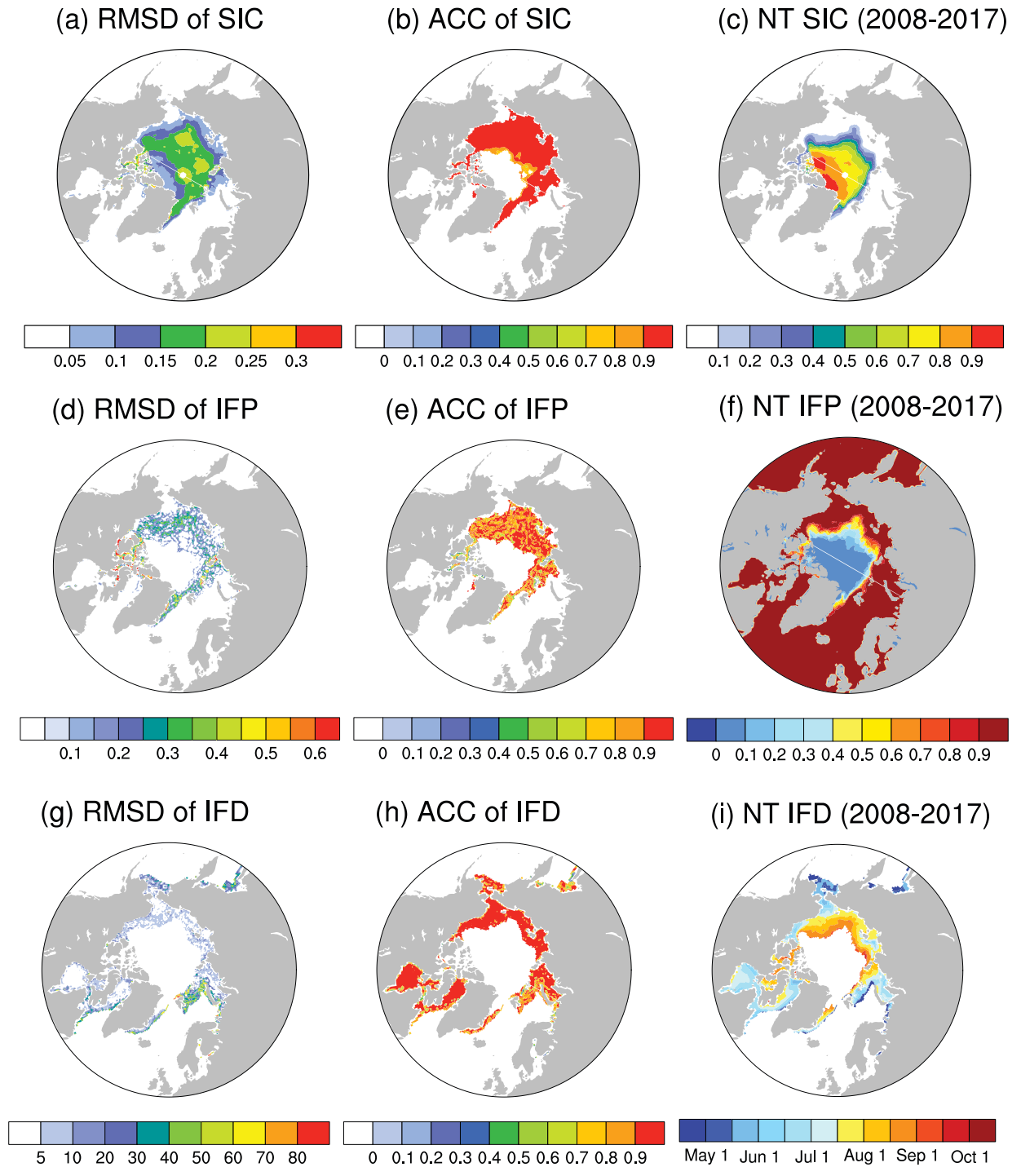

FIG. 1. The RMSD of September (a) sea ice concentration (SIC), (d) ice-free probability (IFP), and (g) ice-free date (IFD), and the ACC of September (b) SIC, (e) IFP, and (h) IFD between the NSIDC NASA Team (NT) and Bootstrap (BT) observations. Examples of the NT product are also shown for the 2008–17 (c) mean September SIC, (f) IFP, and (i) median IFD.

Archipelago, and the pan-Arctic, while IceDA converges with nIceDA after 30 days in the Barents Sea and shows lower ACC after 30 days in the East Siberian Sea. The full ACC of SIE is always slightly higher than the detrended ACC due to the model's ability to capture the decreasing trend of Arctic sea ice. Since the differences between the reforecasts shown in the full ACC are qualitatively similar to that in the detrended ACC, only the detrended ACC plots are presented.

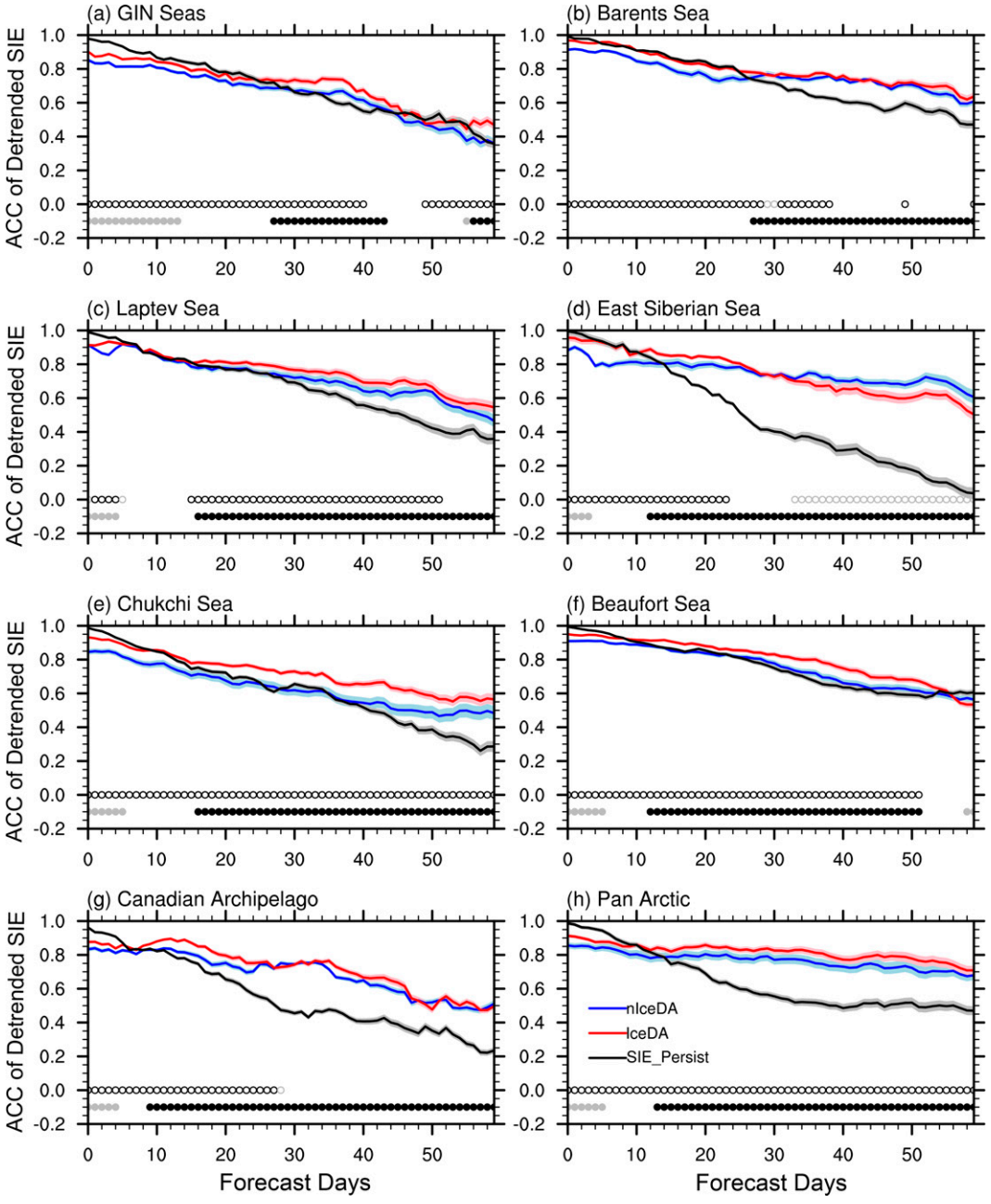

FIG. 2. ACC of detrended sea ice extent (SIE) for regions of summer sea ice variability for nIceDA (blue), IceDA (red), and anomaly persistence forecasts (black). The time series shown as a function of forecast days are averaged over initialization months of June, July, August, and September. The shading represents $\pm(2)^{1/2}$ standard deviations of the 14-day running mean ACC values calculated from the Bootstrap procedure. Open circles on the bottom indicate that the 14-day running mean detrended ACC values between nIceDA and Ice DA are significantly different at the 95% confidence level. Filled circles indicate that IceDA and the SIE persistence forecast are significantly different at the 95% confidence level. Black circles indicate that the skill of IceDA exceeds nIceDA or the persistence forecast, and gray circles indicate that the skill of IceDA is lower than nIceDA or the persistence forecast.

In comparing with the reference persistence forecast, it is seen that the skill of both SPEAR reforecasts is lower than persistence for short lead times of 0–10 days and exceeds the persistence forecast within 35 days in most regions of summer sea ice variability. The experiment IceDA begins to outperform the reference at lead times of 7 days in the Canadian Archipelago, 10 days in the East Siberian Sea, Chukchi, and Beaufort Seas, 15 days in the Laptev Sea, and 25 days in the

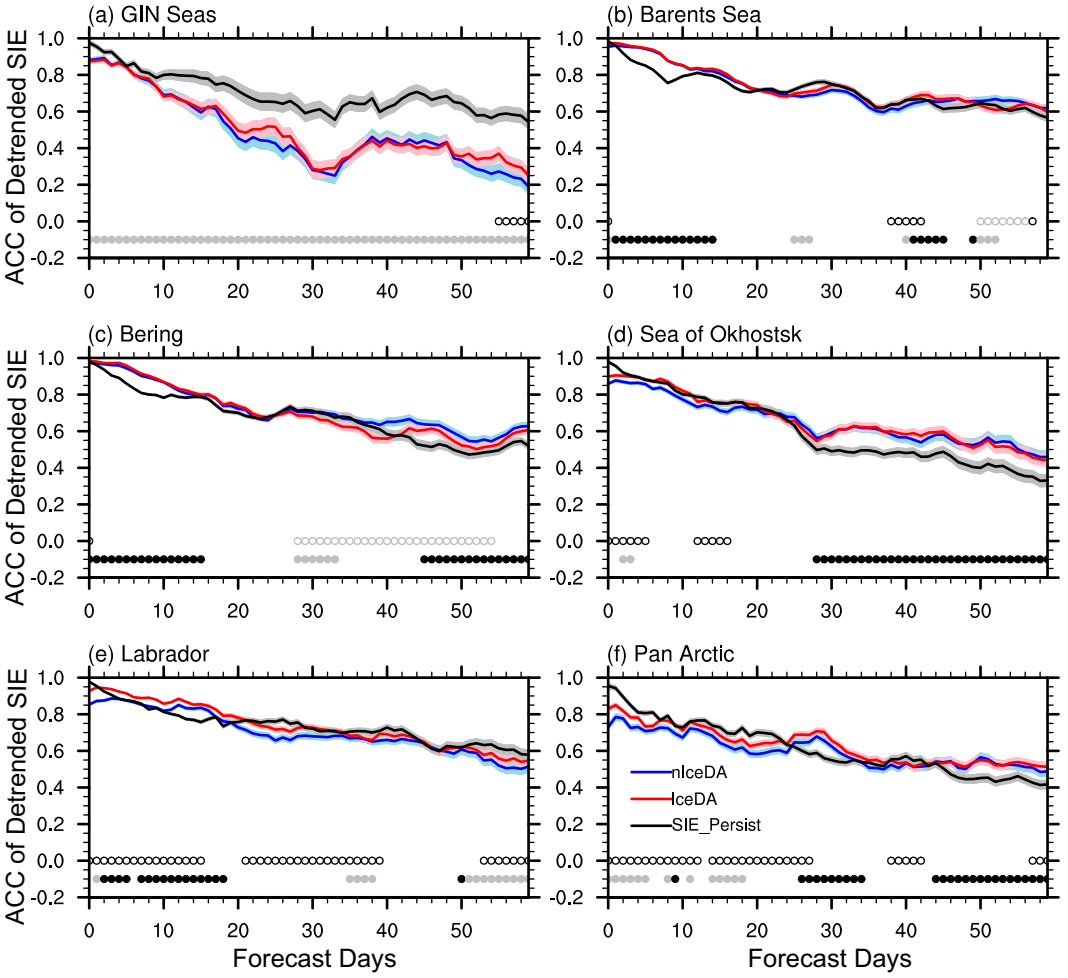

FIG. 3. As in Fig. 2, but for winter sea ice variability. The time series shown as a function of forecast days are averaged over initialization months of January, February, and March.

GIN and Barents Seas. The experiment nIceDA is slower to outperform the reference, taking approximately 10 days in the Canadian Archipelago, 15 days in the East Siberian Sea, 25 days in the Barents, Laptev, and Beaufort Seas, 30 days in the GIN Seas, and 35 days in the Chukchi Sea. The monthly mean skill (Fig. S3 in the online supplemental material) shows that the SPEAR reforecast generally exceeds the anomaly persistence forecast in the first month and again shows a clear skill improvement in the IceDA forecasts.

The winter forecasts (January-, February-, and March-initialized forecasts) show different results than the summer forecasts (Fig. 3). Although IceDA overall has higher detrended ACC than nIceDA, the influence of ice DA initialization is much smaller and less significant, and the two experiments converge quickly except in the Labrador Sea. Neither forecast system surpasses persistence until around 25 days in the Pan Arctic and both lose to the persistence forecast in general in the GIN and Barents Sea. The reforecasts perform well in the Bering Sea and Sea of Okhotsk. In the Labrador Sea, nIceDA loses to the persistence forecast overall, while IceDA exceeds the persistence forecast up to day 20 and is comparable to it afterward.

The ensemble mean of the 15 forecast members has been used to evaluate model results so far. We now look at the ensemble spreads of pan-Arctic SIE from the two reforecast experiments initialized from April to November in Fig. 4. The stabilized spread of SIE for forecast years 3–5 (Fig. S4 in the online supplemental material) from April-initialized reforecast runs is shown for comparison (solid black line). The interannual standard deviation of detrended SIE anomalies calculated from the NASA Team observations over the same time period (1992–2017) is also added for reference (dashed black line). Both sets of initial conditions, SPEAR_IceDA and SPEAR_Nudged, are tightly constrained, with SPEAR_IceDA slightly more constrained mainly due to SIC DA. The ensemble spreads of IceDA and nIceDA grow at almost the same rate as a function of forecast time, suggesting that the processes driving the loss of predictability are common between the two sets of reforecast experiments. The ensemble spread shown in the stabilized run represents the inherent error saturation level of pan-Arctic SIE in the SPEAR forecast system without initialization. The limit of predictability is reached when the spread of the initialized ensembles reaches

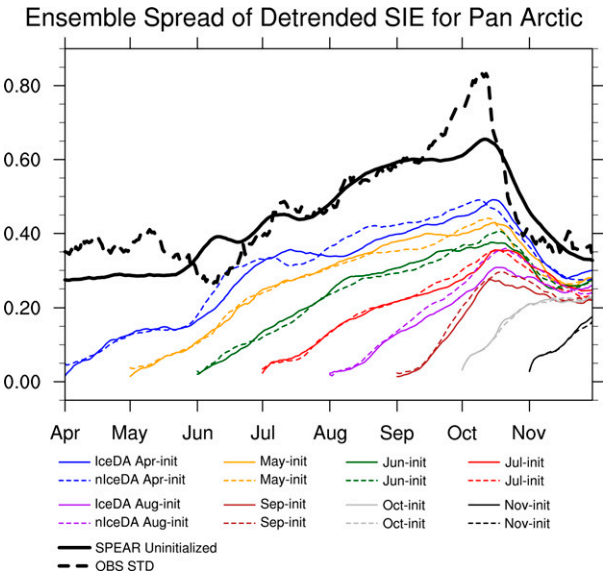

FIG. 4. Ensemble spread of detrended pan-Arctic SIE from the two reforecast experiments (colored lines and thin black lines), the estimated ensemble spread of the uninitialized SPEAR model (thick black line), and the interannual standard deviation of the detrended pan-Arctic SIE in the NASA Team observation from 1992 to 2017 (thick dashed black line).

this free-running model spread. The seasonal cycle of internal variability represented by the forecast system matches the observations well. They both increase as sea ice melts, peak in early October at the beginning of the growth season, and drop again as sea ice advances. The ensemble spreads in the reforecast experiments are much lower than the internal variability at short lead times because of the initialization and grow toward the model internal variability, which also peaks in early October (except for those initialized after October).

c. *Sea ice concentration predictions*

In this section we evaluate the forecast skill of SIC at the gridcell level. A comparison with the SIC anomaly persistence forecast indicates that both experiments show lower skill in the beginning of the forecast period and exceed the reference by the end of the first month for summer predictions (Fig. 5), which is similar to the regional SIE results. The gap between the reference and the experiments in the beginning, however, is larger for SIC than SIE. This indicates that SIE forecasts benefit from some cancellation of small-scale errors and that gridcell-level SIC is a more challenging quantity to predict. The SPEAR forecast experiments in general show lower ACC scores for SIC than SIE. IceDA (red lines in Fig. 5) has increased ACC significantly as compared with nIceDA (blue lines in Fig. 5), although the gap with the persistence forecast is still notable over the first 10 days, indicating that there is additional room-for-improvement in the sea ice initial conditions. The winter skill behaves similarly relative to the anomaly persistence forecast (Fig. S4 in the online supplemental material). The following evaluations will be focused on the differences between IceDA and nIceDA.

Figure 6 summarizes the area-integrated ACC of detrended SIC at lead times of 0–2 months for the two SPEAR reforecast experiments. Similar to the SIE results, IceDA shows superior performance to nIceDA at lead times of 0–1 months. We also see notable improvements in winter/spring in the GIN, Kara, Labrador Seas and Sea of Okhotsk. Their performance is very similar at the lead time of 2 months. Together with Fig. 5 and Fig. S5 in the online supplemental material, Fig. 6 shows that the improvement appears largest in the regions of summer sea ice variability and tapers with lead time.

Figure 7 summarizes the pan-Arctic averaged detrended SIC skill with initialization months from April to November. Since the largest improvements from SIC DA initialization is in summer, and summer Arctic sea ice attracts high interest because of its large variability, trend, and stakeholder interest, we focus on September-targeted forecasts initialized from April to September. The plots in individual regions are noisier, but in general show similar results (Fig. S6 in the online supplemental material). As also shown in Fig. 6, the two experiments have similar September-targeted forecast skill at lead time of 2 months. Notable increases in September skill from SIC DA are found in August- and September-initialized forecasts. This is consistent with our a priori knowledge that the anomaly persistence of ice area or extent contributes to the short-term forecast skill of September Arctic sea ice.

Figure 7 also shows that IceDA clearly has higher ACC values in the first two months of forecast across all the summer initialization months. The ACC difference between the two experiments is largest in the beginning and decreases with forecast time. The skill of the reforecast experiments increases as the lead time decreases. IceDA and nIceDA show no obvious differences in skill in the October- and November-initialized forecasts. This is expected since sea ice anomalies are not the major source of autumn sea ice skill, which is largely contributed by SST anomalies south of the September sea ice edge position. It is also noticeable that the ACC values in the initialization months from April to August start to increase by the end of September. The increase of skill after late September can be explained by the reemergence of skill from the melt season to the growth season due to the persistence of SST anomalies (Blanchard-Wrigglesworth et al. 2011; Guemas et al. 2016). IceDA also shows a better reemergence of skill than nIceDA, especially in August-initialized forecasts. Figure S7 in the online supplemental material shows that the August-initialized forecast of IceDA has a stronger negative correlation between the October SIE and the model-predicted SST of earlier months, confirming that the SST-based reemergence mechanism is stronger in IceDA. Although both initial conditions are strongly nudged toward the same SST observational product, this suggests that with SIC DA, IceDA predicts more accurate better SST anomalies than nIceDA, leading to better SIC prediction skill.

To illustrate the improvement at short lead times in IceDA from nIceDA, we show spatial plots of SIC RMSE averaged over forecast days 0–45 for IceDA (Fig. 8) and their difference with the nIceDA forecasts (Fig. 9). The errors are mainly concentrated in the marginal ice zones, while the

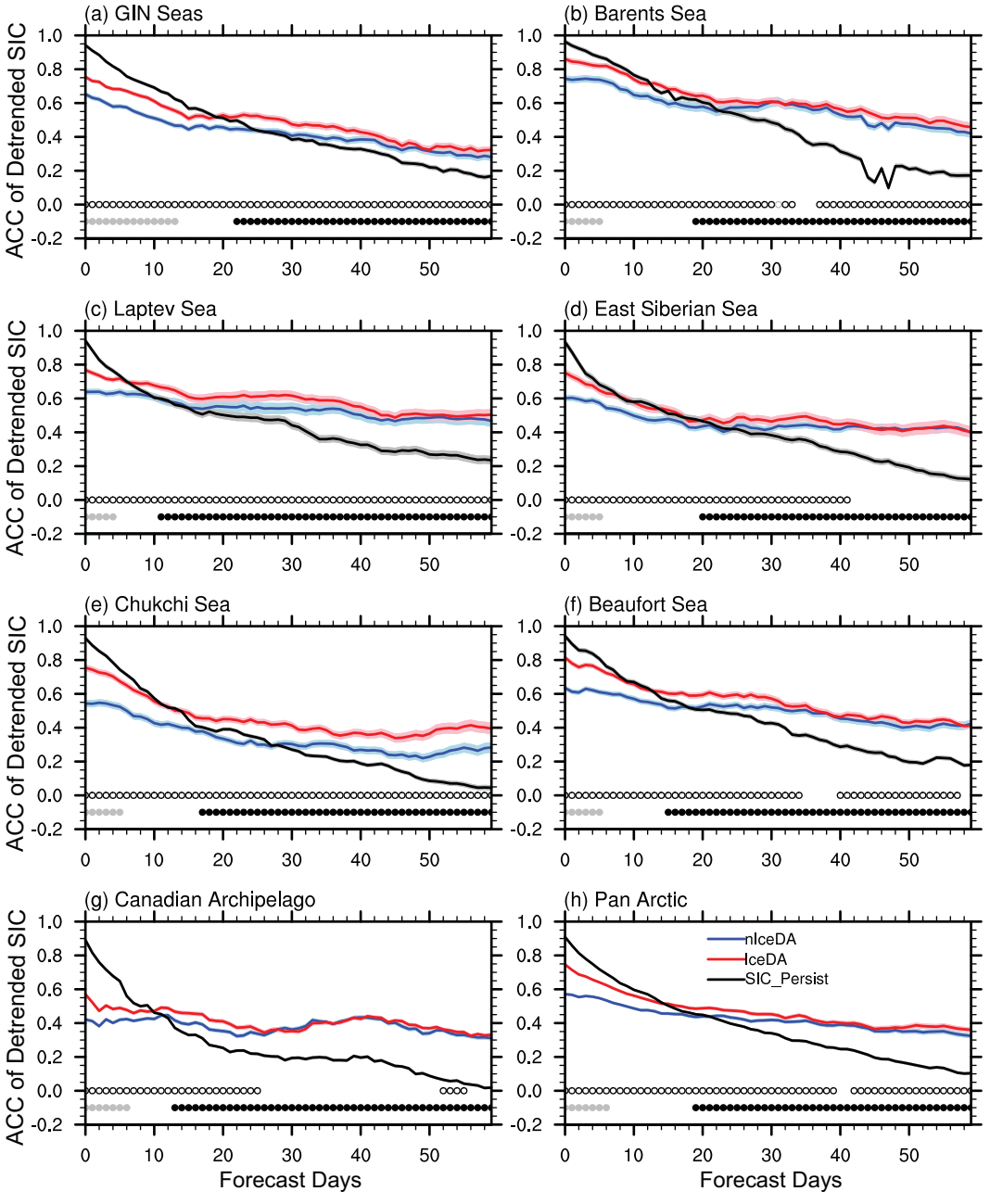

FIG. 5. As in Fig. 2, but for area-averaged ACC.

observational uncertainty is uniformly distributed across the interior Arctic. Unlike the observational uncertainty that is largest in summer, the magnitude of the forecast error in summer is smaller than (Fig. 8i) or similar to (Figs. 8g,h) that in other seasons. The magnitude of forecast error is comparable to the observational uncertainty in September, but larger in other months in the marginal ice zones. The error reduction is largest in summer, modest in spring, small in winter, and not discernible or even degraded in autumn. The small differences in autumn and winter likely result from the fact that the sea ice edge position is largely controlled by SST during the ice growth season, which is used as a constraint in both the IceDA and nIceDA reforecasts.

The experiment IceDA shows high ACC of detrended SIC in the ice variability zone (Fig. 10). While Fig. 6 shows that IceDA has higher area-aggregated detrended ACC of SIC in most regions than nIceDA at lead times of 0 and 1 months, Fig. 11 further demonstrates that the improvement is quite homogeneous across the whole Arctic. The skill improvement of IceDA over nIceDA shown in the detrended ACC (Fig. 11) is quite similar to the RMSE results; improvements are largest in summer, moderate in spring, and marginal in

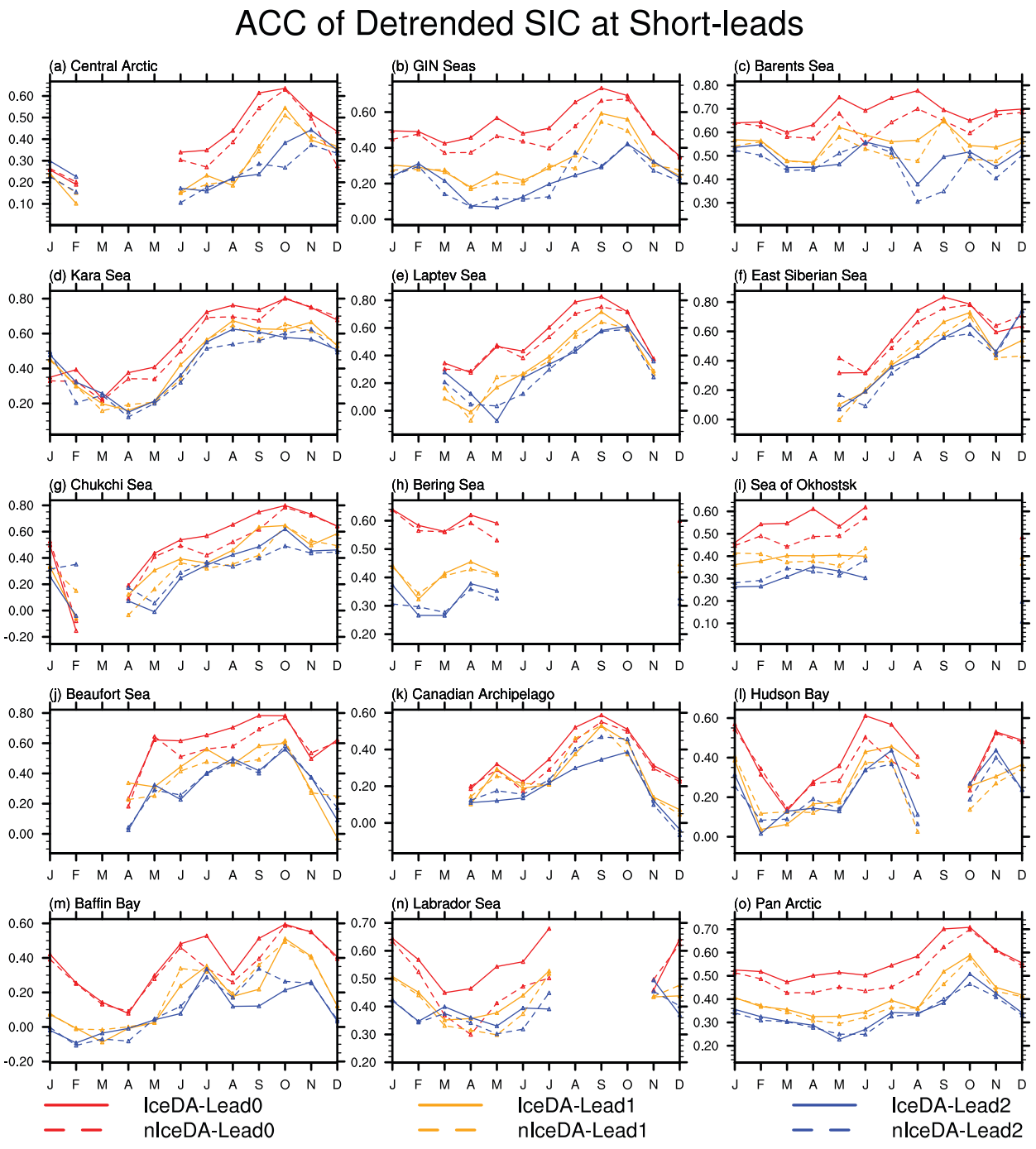

FIG. 6. Area-averaged ACC of detrended SIC for each target month from lead 0 (red), lead 1 (orange), and lead 2 (blue). ACC is calculated each day of the year using data from 1993 to 2017 and then averaged for each month. Only grid cells that have > 10% SIC interannual variability are taken into average for each day.

other seasons. The reduction of RMSE generally matches well with the increase of ACC, for example, from May to July in Hudson Bay and the Barents Sea and from June to September in the Laptev, Chukchi, and Beaufort Seas. Slight degradation of skill is found in the Baffin Bay in May and June (Figs. 9e,f, and 11e,f). The error reduction is persistent from January to August in the GIN Seas (Figs. 9a–h), while the increase of ACC is patchier in the GIN Seas (Figs. 11a–h).

d. Ice-free probability prediction

We showed in section 3c that IceDA generally outperforms nIceDA in predictions of gridcell-level SIC and that these

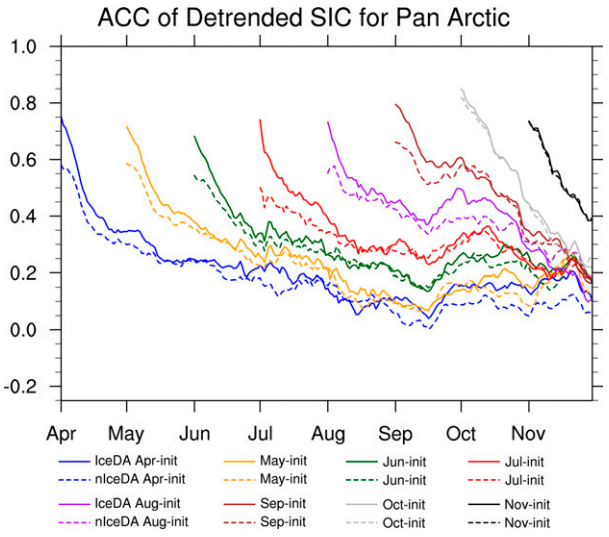

FIG. 7. Pan-Arctic averaged ACC of detrended SIC as a function of forecast month. ACC is calculated everyday of the year using data from 1993 to 2017. Only grid cells that have > 10% SIC interannual variability are taken into average for each day. In the area-averaging step, values averaged over less than 10% area of a certain region are filtered out.

improvements are mainly in spring and summer. In this section we demonstrate that IceDA can improve probabilistic forecasts of ice-free conditions as well. We assess the IFP skill of the reforecast experiments initialized from May to September. IceDA has a similar spatial pattern of IFP RMSE to the Clim_10yr reference forecast and its RMSE decreases as the initialization month progresses into summer (Fig. 12a). The error magnitude of IFP for IceDA remains similar to Clim_10yr until the initialization month of August, which is also shown in the difference map (Fig. 12b). In the August- and September-initialized forecasts, IceDA exceeds the reference forecast in most of the Arctic, with the largest RMSE difference in the East Siberian, Laptev, Chukchi, and Kara Seas. The magnitude of the model's IFP RMSE is generally larger than the observational uncertainty in IFP (Fig. 1), indicating additional room for improvement in these predictions. Their spatial patterns are very similar. The forecast errors and observational uncertainties are concentrated in the seasonal ice zones, which may be contributed by the errors in predicting and uncertainties in observing melting processes. In comparing IceDA with nIceDA, it is seen that the improvement also emerges in the initialization months of August and September, in which reductions of RMSE up to 0.15 are seen in the Beaufort, Chukchi, East Siberian, Laptev, and Kara Seas, and part of the central Arctic (Fig. 12c).

The experiment IceDA shows higher ACC of IFP than Clim_10yr even in the May-initialized forecasts (Fig. 13a), which suggests that IceDA has the potential to make skillful forecast of IFP as early as in May. Clim_10yr has advantages in the Chukchi and Beaufort Seas as compared with the May-, June-, and July-initialized forecast (Fig. 13b), likely related to SPEAR's model bias of overly thick and extensive ice in these regions (Bushuk et al. 2022). IceDA outperforms nIceDA in general, especially in these regions of model bias. The largest improvement from IceDA is seen in the initialization months of August and September, which is consistent with the largest reduction in RMSE in these months. The difference values of RMSE above 0.02 and ACC above 0.05 are in general significant at the 95% confidence level (Figs. S8 and S9 in the online supplemental material).

e. Ice-free date prediction

In this section, we further examine the skill of the SPEAR system in forecasting the timing of the transition from ice cover to open water. The IFD RMSE for the Clim_10yr reference forecast is similar to the May-initialized forecast for IceDA (Fig. 14a). The largest error (up to ~80 days) is seen in the Labrador, GIN, Barents, and Bering Seas. The forecast skill of IceDA gradually increases with shorter lead time. Dominant advantages of July- and August-initialized IceDA over Clim_10yr are found in almost all the peripheral ice regions (Fig. 14b). The data sample for September-initialized forecast is very small since most of the regions either have already met the ice-free condition before 1 September or are located in the perennial ice zone. The difference between IceDA and nIceDA are in general marginal (Fig. 14c). The May-, June-, and July-initialized forecasts of IceDA shows decreased RMSE (up to 10 days) in Hudson Bay and Baffin Bay, and the July- and August-initialized forecasts of IceDA in the Beaufort Sea. Slight increases of RMSE are seen in the June-, July-, and August-initialized forecasts of IceDA in the East Siberian Sea.

Relative to the observational uncertainty (Fig. 1g), the forecast errors of IFD have larger magnitudes. However, their spatial patterns are very similar, with the largest values showing up in the Atlantic sector including the Barents, Kara, and Labrador Seas. Particularly large IFD errors occur near Svalbard and Novaya Zemlya, which are also regions of large observational uncertainty potentially associated with land contamination. This indicates the need to take better account of observational uncertainty when evaluating model performance.

The SPEAR model shows much higher IFD ACC than Clim_10yr in general. The model has high ACC values in the Labrador, Beaufort, Chukchi, Kara, Laptev, and Barents Seas (Fig. 15a), which are also the regions IceDA shows advantages over Clim_10yr (Fig. 15b). The experiment IceDA shows a slight but persistent increase of ACC from nIceDA (Fig. 15c). The large improvement in Hudson Bay and Baffin Bay seen in the RMSE does not appear in the ACC map, which suggests that the error in those regions is dominated by model bias. The difference values of RMSE above 2 days and ACC above 0.1 generally pass the 95% significance test (Figs. S10 and S11 in the online supplemental material).

4. Conclusions and discussion

We conducted two sets of SPEAR reforecast experiments to evaluate the benefits of improved sea ice initial conditions on subseasonal-to-seasonal forecasts of the Arctic sea ice.

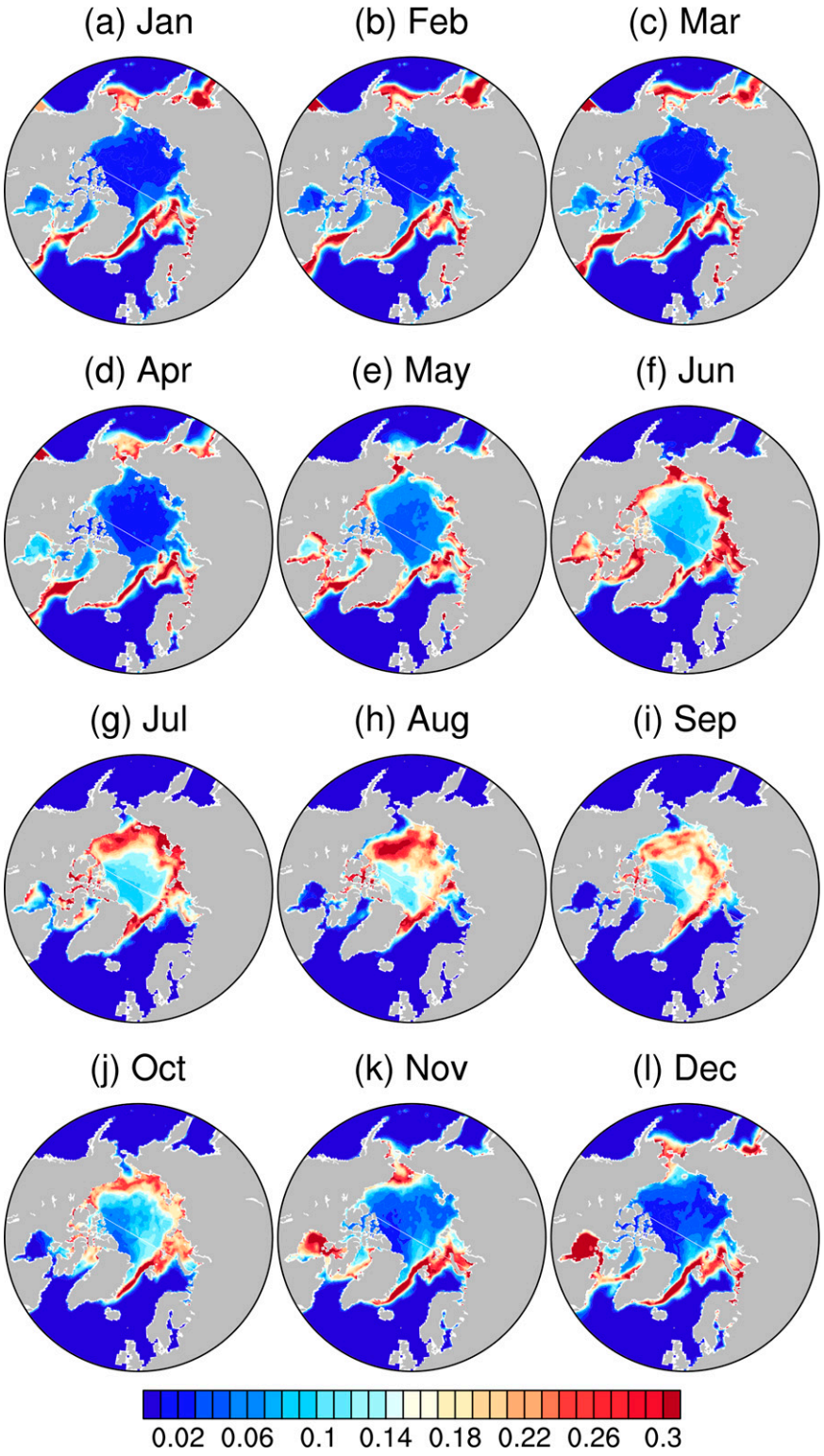

FIG. 8. SIC RMSE averaged over forecast days 0–45 for each initialization month in the experiment IceDA.

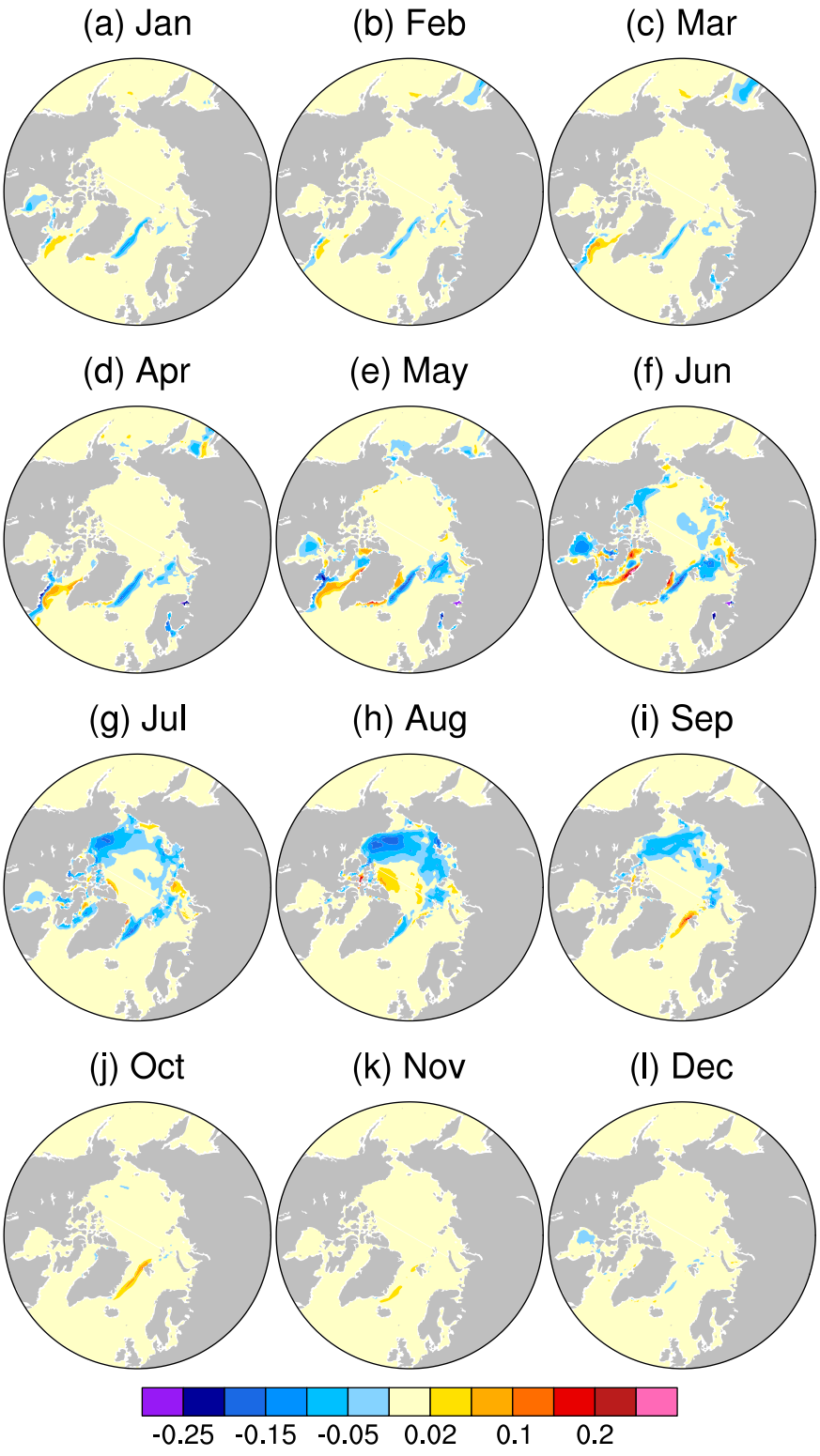

FIG. 9. RMSE differences of SIC averaged over forecast days 0–45 between the experiments IceDA and nIceDA (IceDA − nIceDA) for each initialization month.

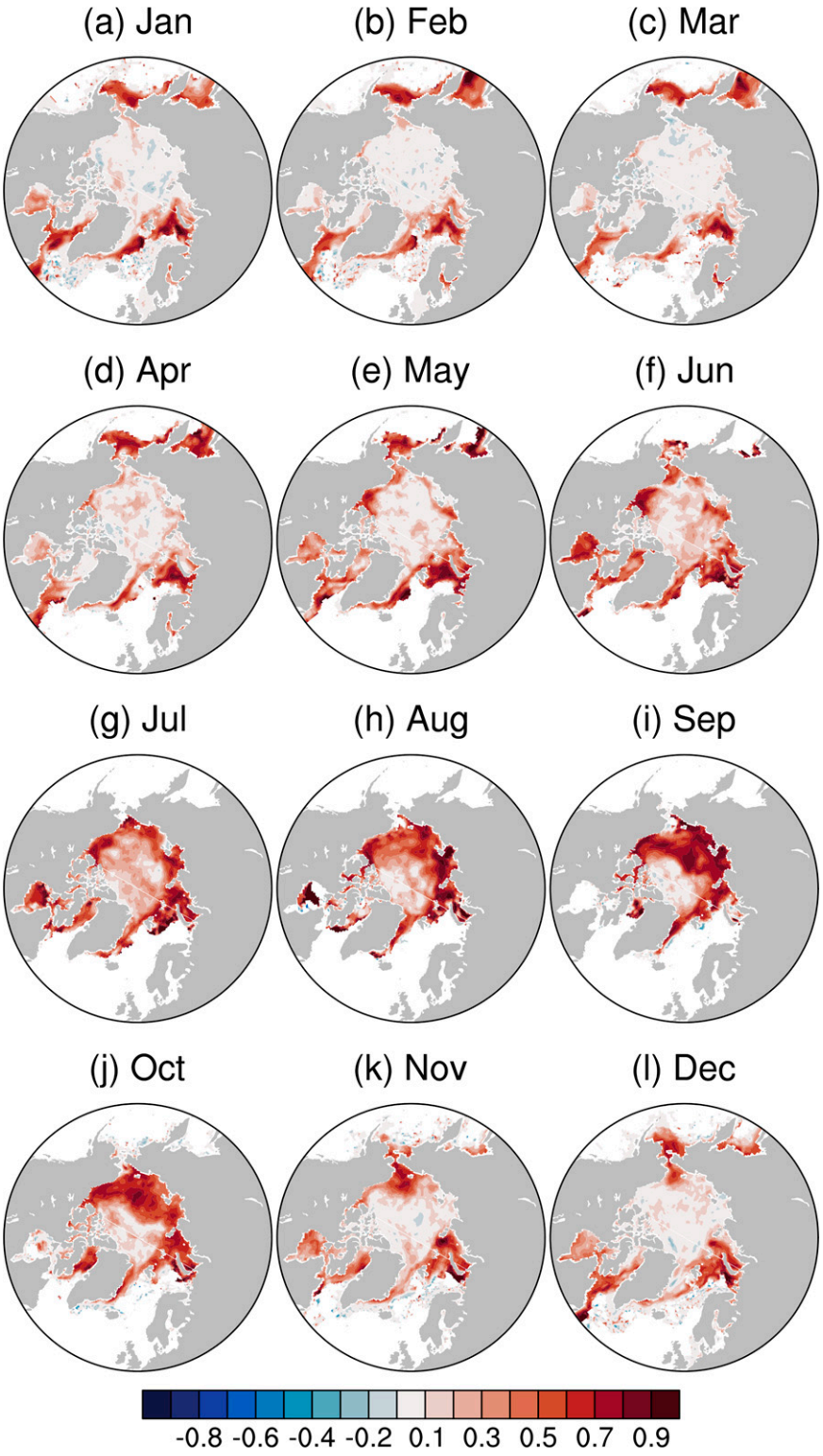

FIG. 10. Detrended SIC ACC averaged over forecast days 0–45 for each initialization month in the experiment IceDA.

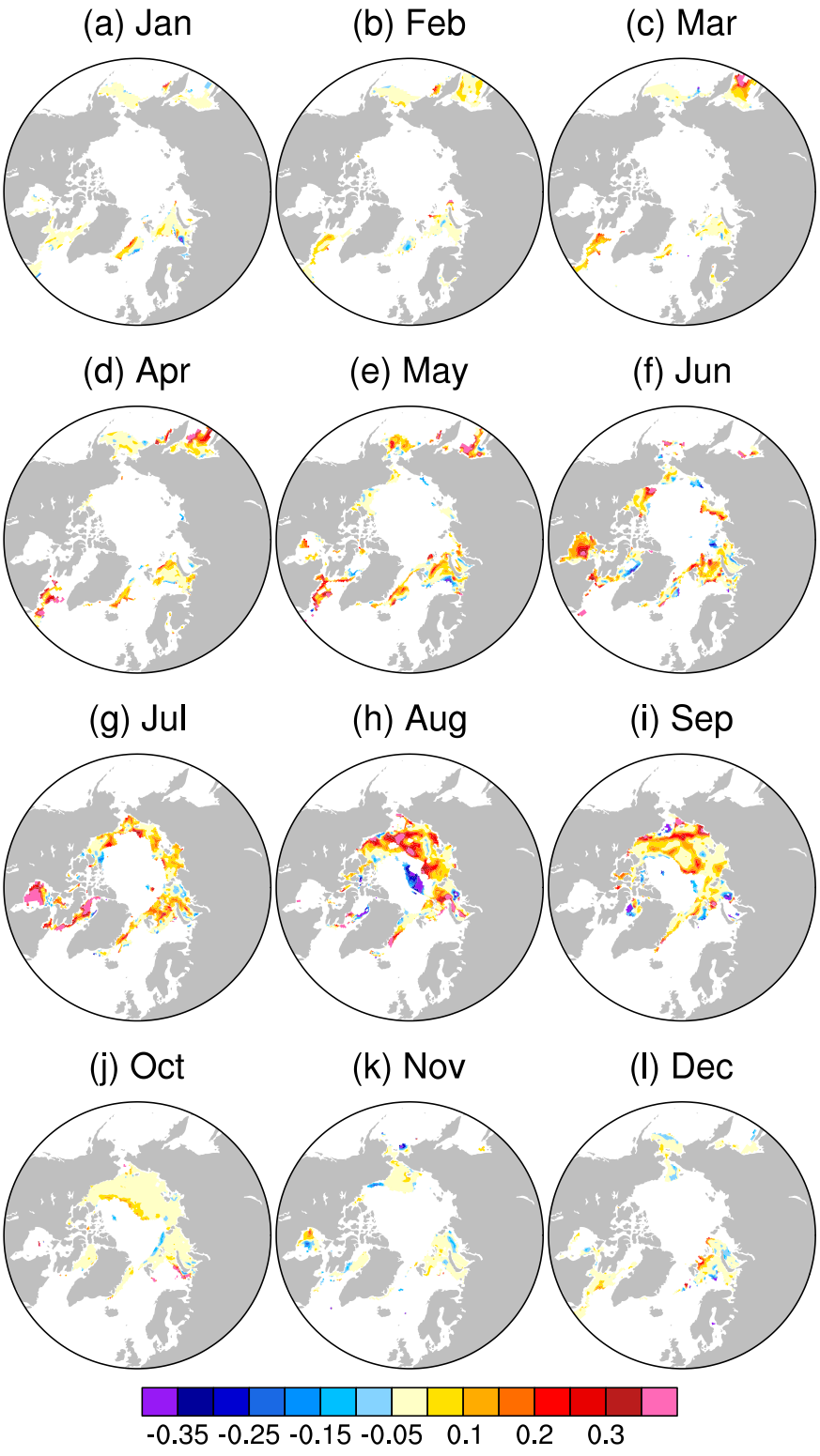

FIG. 11. ACC differences of detrended SIC averaged over forecast days 0–45 between the two experiments (IceDA − nIceDA) for each initialization month.

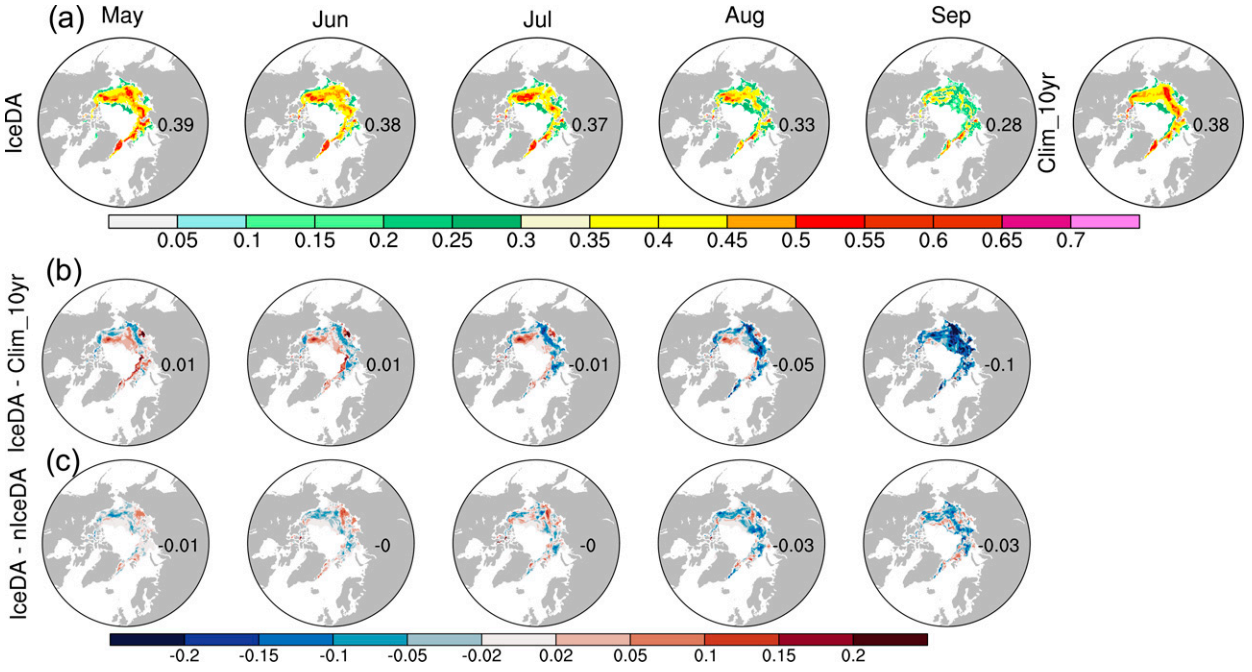

FIG. 12. (a) RMSE of IFP for IceDA initialized from May, June, July, August, and September, and the reference forecast Clim_10yr, (b) the difference of RMSE of IFP between IceDA and Clim_10yr, and (c) the difference of RMSE of IFP between IceDA and nIceDA. Numbers in each plot indicate the pan-Arctic averaged RMSE of IFP.

The experiment nIceDA is initialized from SPEAR_Nudged that does not directly assimilate any sea ice observations, and the experiment IceDA is initialized from SPEAR_IceDA that assimilates daily SIC observations. The sea ice variables evaluated are sea ice extent, sea ice concentration, ice-free probability, and ice-free date.

In both winter and summer forecasts of regional Arctic SIE, IceDA generally shows better skill than nIceDA in the

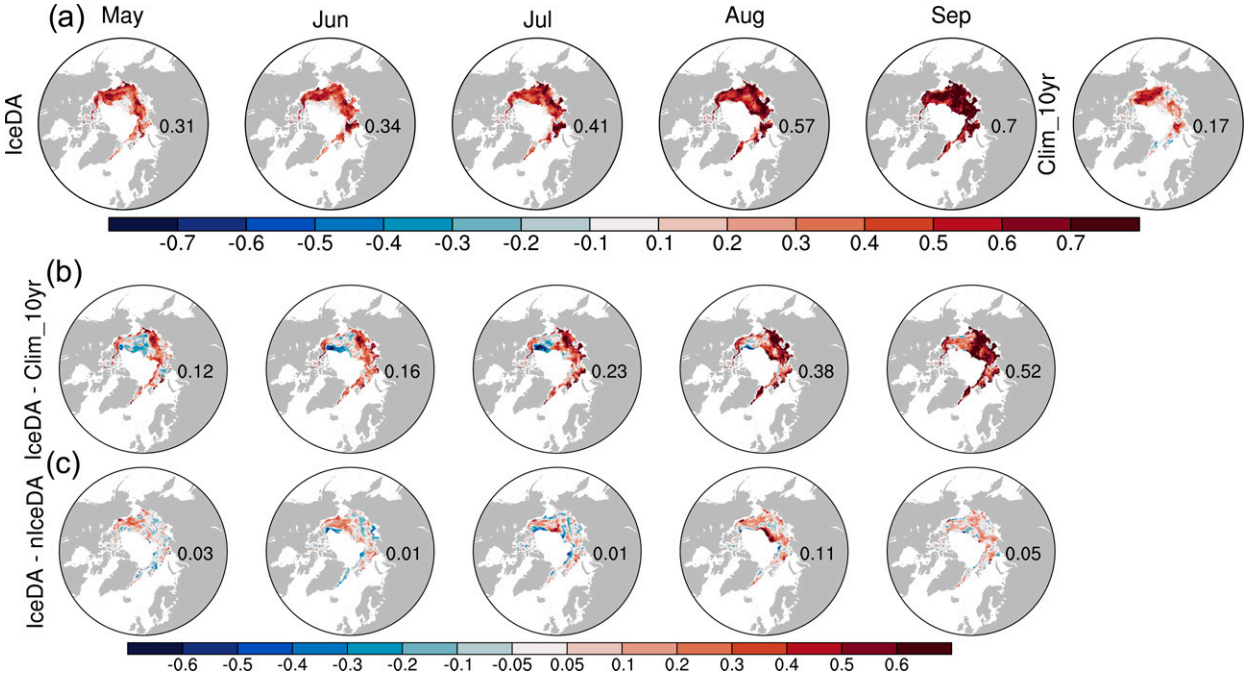

FIG. 13. As in Fig. 12, but for ACC.

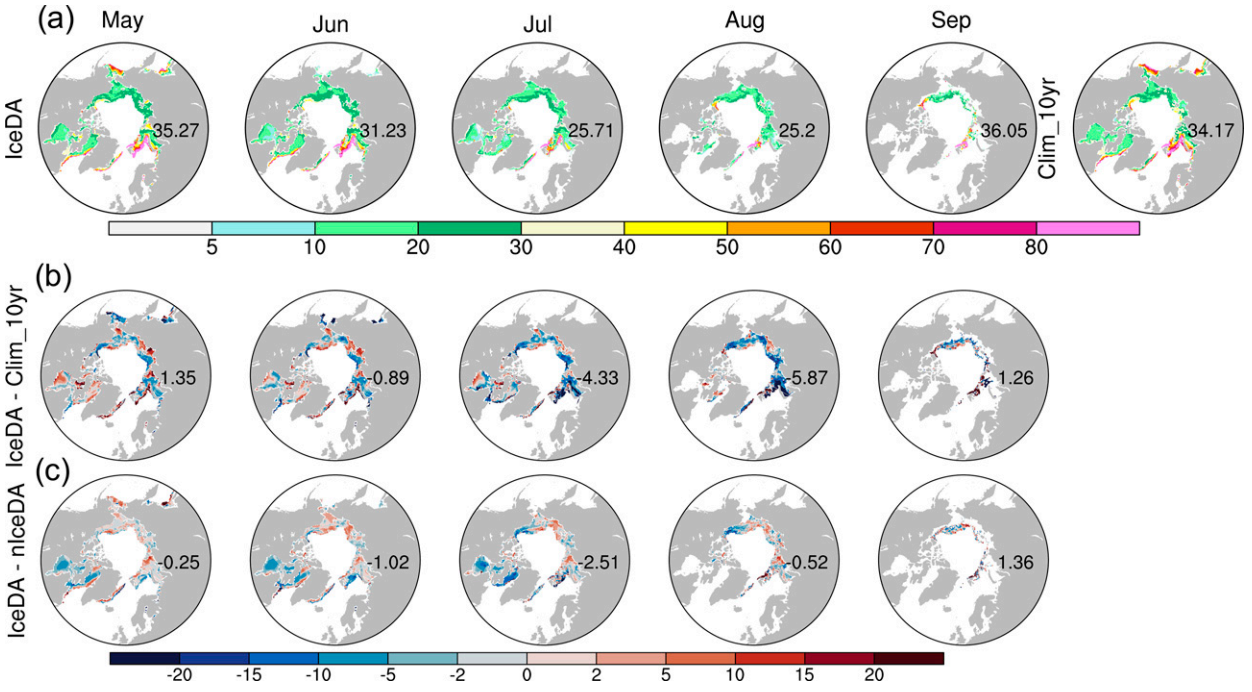

FIG. 14. As in Fig. 12, but for IFD.

first two months of the forecast. The improvement is largest in summer and limited in winter. A comparison with the SIE anomaly persistence forecast shows that the SPEAR system has better skill in summer forecasts at lead times of 0–1 months but does not perform as well in winter forecasts. The conclusions from our daily analysis are mostly consistent with that from the monthly analysis in our companion paper (Bushuk et al. 2022).

The gridcell-level SIC results are similar to the SIE results in that IceDA outperforms nIceDA in most regions and its advantages taper with lead time. The difference between the two experiments is largest in summer, moderate in spring and

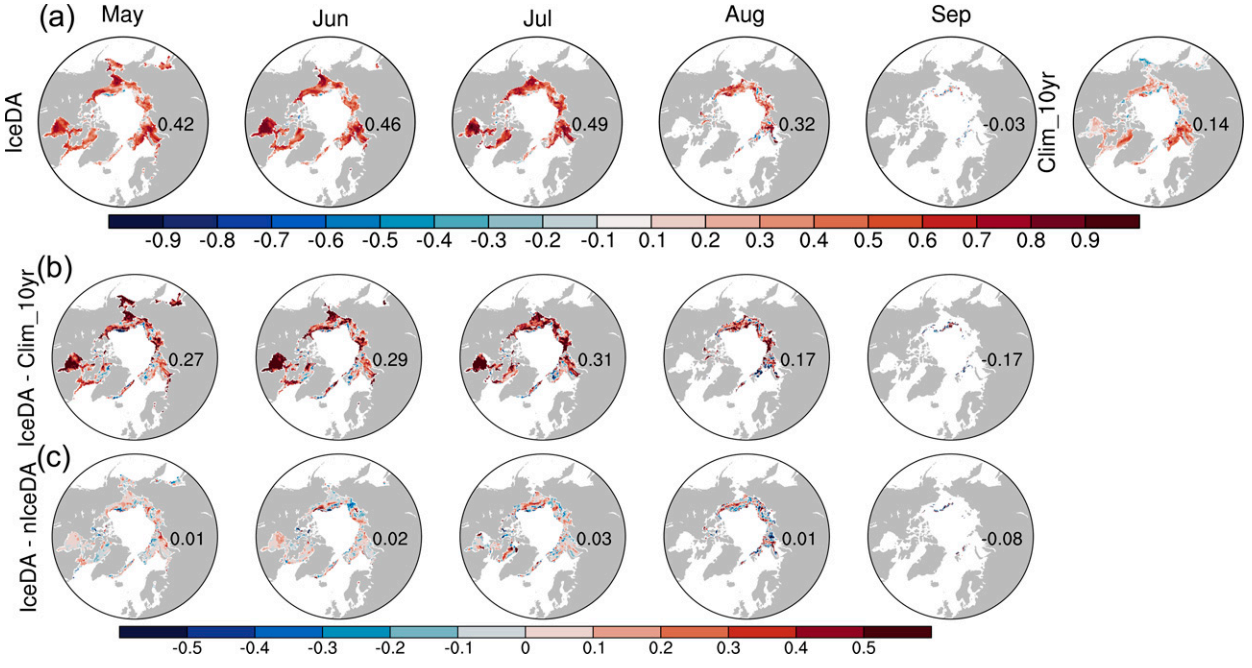

FIG. 15. As in Fig. 13, but for IFD.

winter, and not discernible in autumn. This can be explained by different predictability sources in different seasons. Whereas summer sea ice predictability mainly comes from sea ice area and volume anomalies, winter sea ice predictability is influenced most by ocean conditions, which are very similar in the two reforecast experiments. The SPEAR forecast skill also exceeds the SIC anomaly persistence forecast skill within the first month of the forecast period. The inferiority of the SPEAR reforecasts to the anomaly persistence forecasts in the first 0–10 days indicates that our sea ice initial conditions are not perfect and there is still room for improvement. The initial skill gap of SIC relative to persistence is larger than that of SIE. This suggests that the gridcell-level SIC is a more challenging variable to predict than regional SIE. The spatial patterns of the differences in RMSE and ACC between the two experiments show that IceDA improves SIC forecasts mainly in the marginal ice zones. The reduction in RMSE and increase in ACC agree well in general except the GIN Seas where RMSE is reduced persistently from January to August yet the ACC increase is not as obvious. This is likely due to a large model bias in the GIN Seas that hinders the RMSE forecast skill.

Our finding that IceDA outperforms nIceDA mostly in summer-initialized short-term forecasts suggests that by improving SIE/SIC initializations in IceDA, we are able to boost the short-term predictions of summer Arctic sea ice. The two initializations have applied the same SST restoring technique and hence have similar SST anomalies. The autumn-initialized short-term forecasts, therefore, are quite similar between the two reforecasts, which is consistent with previous findings that upper-ocean heat content anomaly is the major predictability source of autumn Arctic sea ice.

Since assimilation of SST usually incorporates information about sea ice edges, it indirectly improves sea ice extent anomalies, and will therefore provide an important predictability source for Arctic sea ice. However, we show in our preceding paper on SIC DA (Zhang et al. 2021) that direct assimilation of SIC will further improve SIE and more importantly, improve SIC. We further demonstrate in this study that both SIE and SIC forecasts are improved in IceDA. Furthermore, SIC DA also improves SST anomaly persistence in summer, which plays an important role in forecasting autumn sea ice, although this improvement is only obvious in August-initialized forecast and small in other initialization months.

Another important predictability source is SIT and the assimilation of SIT has been actively explored by the Arctic sea ice community to improve summer predictions at lead times longer than several months. The large uncertainty associated with the satellite observations and their availability only in winter and the limited spatial and temporal coverage of in situ observations make it challenging to generate a reliable year-round SIT reanalysis (Liu et al. 2019). Our next immediate step is to investigate how to utilize the currently available ice thickness satellite retrievals using ensemble Kalman filters with the purpose of improving reanalysis SIT interannual variability and eventually improving longer-term predictions of Arctic sea ice.

We also evaluated IFP and IFD to assess the skill of forecasting ice-free probability and the timing of the seasonal transition to open water. The SPEAR system can make skillful predictions of both variables at lead times as early as May, which is demonstrated by comparing their ACC with the reference forecast. However, the difference in RMSE between the SPEAR experiments and the reference forecast is not clear until the initialization time of August for IFP and July for IFD, and the magnitude of the difference is relatively small in comparison with the absolute RMSE. The enhanced skill represented by ACC relative to RMSE indicates that the model is capturing the interannual variability of the sea ice–free conditions well but the model may contain considerable IFD biases that need further investigation.

The largest improvement in IFP skill from SIC DA is in August- and September-initialized forecasts. SIC DA does not have much influence in the IFP variability zone in the earlier forecasts, and hence, it does not show large impact on the IFP skill until August. The improvement in IFD from IceDA is limited in general, showing smaller improvements than those found for IFP.

In summary, we showed skillful subseasonal-to-seasonal forecasts for Arctic sea ice cover in the newly developed prediction system, SPEAR. Our work also represents an important step toward comprehensively studying on the impact of SIC DA on Arctic sea ice forecast skill. The prediction skill for regional SIE and gridcell-level SIC exceed the reference anomaly persistence forecasts within the first month of the forecast period. This indicates that SPEAR is among the skillful dynamical prediction systems previously evaluated. By constraining the sea ice initial conditions with SIC DA, improvements in SIC skill are prominent in summer-initialized forecasts, moderate in spring, and small in winter and autumn. This agrees well with Kimmritz et al. (2019) that summer-initialized forecasts particularly benefit from SIC DA. By evaluating two user-oriented metrics, IFP and IFD, we demonstrated that the SPEAR system also makes skillful predictions of local ice melt conditions as early as May. SIC DA shows improvements in the forecast skill for IFP, especially for summer-initialized forecasts, and shows a smaller impact on IFD predictions. Overall, this study demonstrates that SIC DA improves both regional and local-scale Arctic sea ice predictions, and therefore represents an essential ingredient for future sea ice prediction systems.

Acknowledgments. Yong-Fei Zhang received Award NA18OAR4320123 under the Cooperative Institute for Modeling the Earth System (CIMES) at Princeton University, and the National Oceanic and Atmospheric Administration, U.S. Department of Commerce. We thank Maike Sonnewald and Alistair Adcroft for helpful comments on earlier versions of the paper. The University Center for Atmospheric Research is sponsored by the National Science Foundation.

Data availability statement. The NASA Team and Bootstrap sea ice concentration products from the National Snow and Ice

Data Center are available online (https://nsidc.org/data/nsidc-0051 and https://nsidc.org/data/nsidc-0079). The SPEAR reforecast data are available at https://figshare.com/articles/dataset/IceDA_reforecast_SIC/19684290.

REFERENCES

Bhatt, U., and Coauthors, 2020: Sea ice outlook: 2020 August report. ARCUS, https://www.arcus.org/sipn/sea-ice-outlook/2020/august.

Blanchard-Wrigglesworth, E., K. C. Armour, C. M. Bitz, and E. DeWeaver, 2011: Persistence and inherent predictability of Arctic sea ice in a GCM ensemble and observations. *J. Climate*, **24**, 231–250, https://doi.org/10.1175/2010JCLI3775.1.

Blockley, E. W., and K. A. Peterson, 2018: Improving met office seasonal predictions of Arctic sea ice using assimilation of CryoSat-2 thickness. *Cryosphere*, **12**, 3419–3438, https://doi.org/10.5194/tc-12-3419-2018.

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

——, ——, ——, ——, X. Yang, A. Rosati, and R. Gudgel, 2019: Regional Arctic sea–ice prediction: Potential versus operational seasonal forecast skill. *Climate Dyn.*, **52**, 2721–2743, https://doi.org/10.1007/s00382-018-4288-y.

——, and Coauthors, 2022: Mechanisms of regional Arctic sea ice predictability in two dynamical seasonal forecast systems. *J. Climate*, **35**, 4207–4231, https://doi.org/10.1175/JCLI-D-21-0544.1.

Cavalieri, D. J., and Coauthors, 1996: Sea ice concentrations from *Nimbus-7* SMMR and DMSP SSM/I-SSMIS passive microwave data, version 1. NSIDC, accessed 14 September 2020, https://doi.org/10.5067/8GQ8LZQVL0VL.

Caya, A., M. Buehner, and T. Carrieres, 2010: Analysis and forecasting of sea ice conditions with three-dimensional variational data assimilation and a coupled ice–ocean model. *J. Atmos. Oceanic Technol.*, **27**, 353–369, https://doi.org/10.1175/2009JTECHO701.1.

Chen, Z., J. Liu, M. Song, Q. Yang, and S. Xu, 2017: Impacts of assimilating satellite sea ice concentration and thickness on Arctic sea ice prediction in the NCEP Climate Forecast System. *J. Climate*, **30**, 8429–8446, https://doi.org/10.1175/JCLI-D-17-0093.1.

Chevallier, M., D. S. Mélia, A. Voldoire, M. Déqué, and G. Garric, 2013: Seasonal forecasts of the pan-Arctic sea ice extent using a GCM-based seasonal prediction system. *J. Climate*, **26**, 6092–6104, https://doi.org/10.1175/JCLI-D-12-00612.1.

Comiso, J. C., 2017: Bootstrap sea ice concentrations from *Nimbus-7* SMMR and DMSP SSM/I-SSMIS, version 3. NSIDC, accessed 19 October 2021, https://doi.org/10.5067/7Q8HCCWS4I0R.

Delworth, T. L., and Coauthors, 2020: SPEAR: The next generation GFDL modeling system for seasonal to multidecadal prediction and projection. *J. Adv. Model. Earth Syst.*, **12**, e2019MS001895, https://doi.org/10.1029/2019MS001895.

Dirkson, A., B. Denis, and W. Merryfield, 2019: A multimodel approach for improving seasonal probabilistic forecasts of regional Arctic sea ice. *Geophys. Res. Lett.*, **46**, 10 844–10 853, https://doi.org/10.1029/2019GL083831.

——, ——, M. Sigmond, and W. J. Merryfield, 2021: Development and calibration of seasonal probabilistic forecasts of ice-free dates and freeze-up dates. *Wea. Forecasting*, **36**, 301–324, https://doi.org/10.1175/WAF-D-20-0066.1.

Fritzner, S., R. Graversen, K. H. Christensen, P. Rostosky, and K. Wang, 2019: Impact of assimilating sea ice concentration, sea ice thickness and snow depth in a coupled ocean–sea ice modelling system. *Cryosphere*, **13**, 491–509, https://doi.org/10.5194/tc-13-491-2019.

Guemas, V., and Coauthors, 2016: A review on Arctic sea-ice predictability and prediction on seasonal to decadal time-scales. *Quart. J. Roy. Meteor. Soc.*, **142**, 546–561, https://doi.org/10.1002/qj.2401.

Kimmritz, M., F. Counillon, C. Bitz, F. Massonnet, I. Bethke, and Y. Gao, 2018: Optimising assimilation of sea ice concentration in an Earth system model with a multicategory sea ice model. *Tellus*, **70A**, 1–23, https://doi.org/10.1080/16000870.2018.1435945.

——, ——, L. H. Smedsrud, I. Bethke, N. Keenlyside, F. Ogawa, and Y. Wang, 2019: Impact of ocean and sea ice initialisation on seasonal prediction skill in the Arctic. *J. Adv. Model. Earth Syst.*, **11**, 4147–4166, https://doi.org/10.1029/2019MS001825.

Lindsay, R., and J. Zhang, 2006: Assimilation of ice concentration in an ice–ocean model. *J. Atmos. Oceanic Technol.*, **23**, 742–749, https://doi.org/10.1175/JTECH1871.1.

Lisæter, K. A., J. Rosanova, and G. Evensen, 2003: Assimilation of ice concentration in a coupled ice–ocean model, using the ensemble Kalman filter. *Ocean Dyn.*, **53**, 368–388, https://doi.org/10.1007/s10236-003-0049-4.

Liu, J., and Coauthors, 2019: Towards reliable Arctic sea ice prediction using multivariate data assimilation. *Sci. Bull.*, **64**, 63–72, https://doi.org/10.1016/j.scib.2018.11.018.

Liu, Y., W. Wang, and A. Kumar, 2018: Multiweek prediction skill assessment of Arctic sea ice variability in the CFSv2. *Wea. Forecasting*, **33**, 1453–1476, https://doi.org/10.1175/WAF-D-18-0046.1.

Lu, F., and Coauthors, 2020: GFDL's SPEAR seasonal prediction system: Initialization and ocean tendency adjustment (OTA) for coupled model predictions. *J. Adv. Model. Earth Syst.*, **12**, e2020MS002149, https://doi.org/10.1029/2020MS002149.

Massonnet, F., T. Fichefet, and H. Goosse, 2015: Prospects for improved seasonal Arctic sea ice predictions from multivariate data assimilation. *Ocean Modell.*, **88**, 16–25, https://doi.org/10.1016/j.ocemod.2014.12.013.

Meier, W. N., 2005: Comparison of passive microwave ice concentration algorithm retrievals with AVHRR imagery in Arctic peripheral seas. *IEEE Trans. Geosci. Remote Sens.*, **43**, 1324–1337, https://doi.org/10.1109/TGRS.2005.846151.

Msadek, R., G. A. Vecchi, M. Winton, and R. Gudgel, 2014: Importance of initial conditions in seasonal predictions of Arctic sea ice extent. *Geophys. Res. Lett.*, **41**, 5208–5215, https://doi.org/10.1002/2014GL060799.

Sigmond, M., J. Fyfe, G. Flato, V. Kharin, and W. Merryfield, 2013: Seasonal forecast skill of Arctic sea ice area in a dynamical forecast system. *Geophys. Res. Lett.*, **40**, 529–534, https://doi.org/10.1002/grl.50129.

——, M. Reader, G. Flato, W. Merryfield, and A. Tivy, 2016: Skillful seasonal forecasts of Arctic sea ice retreat and advance dates in a dynamical forecast system. *Geophys. Res. Lett.*, **43**, 12–457, https://doi.org/10.1002/2016GL071396.

Stark, J. D., J. Ridley, M. Martin, and A. Hines, 2008: Sea ice concentration and motion assimilation in a sea ice–ocean model. *J. Geophys. Res.*, **113**, C05S91, https://doi.org/10.1029/2007JC004224.

Tsujino, H., and Coauthors, 2018: JRA-55 based surface dataset for driving ocean–sea-ice models (JRA55-do). *Ocean Modell.*, **130**, 79–139, https://doi.org/10.1016/j.ocemod.2018.07.002.

Van Woert, M. L., C.-Z. Zou, W. N. Meier, P. D. Hovey, R. H. Preller, and P. G. Posey, 2004: Forecast verification of the Polar Ice Prediction System (PIPS) sea ice concentration fields. *J. Atmos. Oceanic Technol.*, **21**, 944–957, https://doi.org/10.1175/1520-0426(2004)021<0944:FVOTPI>2.0.CO;2.

Vitart, F., and Coauthors, 2017: The Subseasonal to Seasonal (S2S) Prediction project database. *Bull. Amer. Meteor. Soc.*, **98**, 163–173, https://doi.org/10.1175/BAMS-D-16-0017.1.

Wang, W., M. Chen, and A. Kumar, 2013: Seasonal prediction of Arctic sea ice extent from a coupled dynamical forecast system. *Mon. Wea. Rev.*, **141**, 1375–1394, https://doi.org/10.1175/MWR-D-12-00057.1.

Wayand, N., C. Bitz, and E. Blanchard-Wrigglesworth, 2019: A year-round subseasonal-to-seasonal sea ice prediction portal. *Geophys. Res. Lett.*, **46**, 3298–3307, https://doi.org/10.1029/2018GL081565.

Xie, J., F. Counillon, L. Bertino, X. Tian-Kunze, and L. Kaleschke, 2016: Benefits of assimilating thin sea ice thickness from SMOS into the TOPAZ system. *Cryosphere*, **10**, 2745–2761, https://doi.org/10.5194/tc-10-2745-2016.

Zampieri, L., H. F. Goessling, and T. Jung, 2018: Bright prospects for Arctic sea ice prediction on subseasonal time scales. *Geophys. Res. Lett.*, **45**, 9731–9738, https://doi.org/10.1029/2018GL079394.

Zhang, Y.-F., M. Bushuk, M. Winton, B. Hurlin, X. Yang, T. Delworth, and L. Jia, 2021: Assimilation of satellite-retrieved sea ice concentration and prospects for September predictions of Arctic sea ice. *J. Climate*, **34**, 2107–2126, https://doi.org/10.1175/JCLI-D-20-0469.1.