Assimilation of Satellite-Retrieved Sea Ice Concentration and Prospects for September Predictions of Arctic Sea Ice

YONG-FEI ZHANG,[a,b] MITCHELL BUSHUK,[a] MICHAEL WINTON,[a] BILL HURLIN,[a] XIAOSONG YANG,[a] TOM DELWORTH,[a] AND LIWEI JIA[a,c]

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

(Manuscript received 22 June 2020, in final form 9 December 2020)

ABSTRACT: The current GFDL seasonal prediction system achieved retrospective sea ice extent (SIE) skill without direct sea ice data assimilation. Here we develop sea ice data assimilation, shown to be a key source of skill for seasonal sea ice predictions, in GFDL's next-generation prediction system, the Seamless System for Prediction and Earth System Research (SPEAR). Satellite sea ice concentration (SIC) observations are assimilated into the GFDL Sea Ice Simulator version 2 (SIS2) using the ensemble adjustment Kalman filter (EAKF). Sea ice physics is perturbed to form an ensemble of ice–ocean members with atmospheric forcing from the JRA-55 reanalysis. Assimilation is performed every 5 days from 1982 to 2017 and the evaluation is conducted at pan-Arctic and regional scales over the same period. To mitigate an assimilation overshoot problem and improve the analysis, sea surface temperatures (SSTs) are restored to the daily Optimum Interpolation Sea Surface Temperature version 2 (OISSTv2). The combination of SIC assimilation and SST restoring reduces analysis errors to the observational error level (~10%) from up to 3 times larger than this (~30%) in the free-running model. Sensitivity experiments show that the choice of assimilation localization half-width (190 km) is near optimal and that SIC analysis errors can be further reduced slightly either by reducing the observational error or by increasing the assimilation frequency from every 5 days to daily. A lagged-correlation analysis suggests substantial prediction skill improvements from SIC initialization at lead times of less than 2 months.

KEYWORDS: Arctic; Sea ice; Data assimilation

1. Introduction

Arctic sea ice has undergone rapid changes in recent decades, which imposes threats on the wildlife and local people whose habitats largely rely on sea ice. Meanwhile, it brings economic opportunities including marine fishing, more direct shipping routes through the Arctic, and petroleum extraction. Predicting Arctic sea ice, especially in the summertime, has great implications for environmental protection, human activity regulations, and stakeholder decision making.

Recent work has shown that coupled Earth system models are capable of making skillful seasonal predictions of Arctic sea ice (e.g., W. Wang et al. 2013; Chevallier et al. 2013; Sigmond et al. 2013; Msadek et al. 2014; Smith et al. 2015; Bushuk et al. 2017; Batté et al. 2020). Evaluations of several state-of-the-art modeling systems' Arctic sea ice prediction skill revealed that the lack of knowledge of sea ice initial conditions is one of the major limitations of current forecasts (Dirkson et al. 2019). Four out of the six evaluated climate models in Dirkson et al. (2019) use sea ice observation–constrained initializations (Table 1), and the data assimilation method applied in these systems ranges from simple nudging to three-dimensional variational approaches. Despite the growing spatial and temporal coverage of satellite observations of Arctic sea ice, large initial errors remain in the current major subseasonal to seasonal prediction systems (Zampieri et al. 2018). Additionally, perfect model studies, which estimate the upper limit of predictability of a particular model, have demonstrated a large skill gap between current operational prediction systems and their potential prediction skill (Bushuk et al. 2019a). These findings imply an opportunity to improve sea ice data assimilation (DA) techniques to make better use of the existing observations and advance prediction capability.

Previous sea ice DA studies either designed perfect model observing system simulation experiments (OSSEs) to test various DA techniques or assimilated real observations in a variety of ice–ocean models. The DA methods actively applied in the sea ice community include nudging (e.g., K. Wang et al. 2013; Lindsay and Zhang 2006), optimal interpolation (e.g., Stark et al. 2008), three-dimensional variational approach (3D-VAR) (e.g., Caya et al. 2010; Toyoda et al. 2015), and ensemble Kalman filters (EnKFs) (e.g., Massonnet et al. 2015; Kimmritz et al. 2018; Zhang et al. 2018; Fritzner et al. 2019). These methods involve different levels of complexity and computational cost. All have proven to reduce the difference between the model states and observations efficiently. They differ in how relative weights are given to the model and observations,

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-20-0469.s1.

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

DOI: 10.1175/JCLI-D-20-0469.1

TABLE 1. List of models discussed in section 1. Modified based on Table 1 in Dirkson et al. (2019).

Center	Model	Sea ice component/ physics	Resolution (°)	Observations assimilated in sea ice initialization	DA method	DA frequency
ECMWF	SEAS5 (Johnson et al. 2019)	LIM2, implicit ice categories, EVP	0.25	SIC	3DVAR	Every 5 days
CMC	CanCM3 (Merryfield et al. 2013)	SIC/SIT, cavitating fluid	2.8	SIC	Nudging	Daily
CMC	CanCM4 (Merryfield et al. 2013)	SIC/SIT, cavitating fluid	2.8	SIC	Nudging	Daily
GFDL	FLOR (Vecchi et al. 2014)	SIS1, five ice categories, EVP	1	None; constrained by ocean and atmosphere DA	—	—
Météo France	SYSTEM 6 (Dorel et al. 2017)	GELATO v6, five ice categories, EVP	1	NONE	—	Daily
U.K. Met Office	GloSea5 (MacLachlan et al. 2015)	CICE, five ice categories, EVP	0.25	SIC	3D VAR	Every 2 days

whether and how the covariance matrices of the model states are calculated, and how other nonobserved model states are affected. 3D-VAR and EnKFs consider spatially varying error covariance matrix of the model states, and the latter additionally considers the temporal dimension (Liu et al. 2019). A variant of EnKF, the ensemble adjustment Kalman filter (Anderson 2001), is applied in this study.

All studies find that the assimilation of sea ice concentration (SIC) leads to significant reductions of errors in SIC or sea ice extent (SIE) (e.g., Lisæter et al. 2003; Stark et al. 2008; Mathiot et al. 2012; Sakov et al. 2012; Kimmritz et al. 2018; Zhang et al. 2018), while they have different conclusions as to whether SIC DA improves sea ice thickness (SIT). Most studies agree that SIC DA has limited influence on SIT (e.g., Lisæter et al. (2003); Sakov et al. 2012; Kimmritz et al. 2018; Zhang et al. 2018), whereas Stark et al. (2008) find some impact of SIC DA on the thinnest ice category, and Mathiot et al. (2012) and Tietsche et al. (2013) show that SIC DA is also able to significantly improve grid cell-averaged SIT. Both OSSEs and observation-based DA studies show that SIT DA can significantly improve the modeled SIT (e.g., Yang et al. 2014; Xie et al. 2016; Chen et al. 2017; Mu et al. 2018; Zhang et al. 2018; Fritzner et al. 2019). The assimilation of SIT observations is challenging given the large errors and short temporal coverage of satellite retrievals (Zygmuntowska et al. 2014) and the lack of spatial representativeness of observations from ground, submarine, or airborne measurements (Lindsay and Schweiger 2015). Blockley and Peterson (2018) directly assimilated satellite-retrieved winter SIT data into an ice–ocean model and showed improvements in the summer pan-Arctic SIE and ice edge predictions, indicating the potential benefits of this challenging task.

Among the dynamical models that contribute to the September Arctic sea ice forecasts collected by the Sea Ice Outlook (SIO), more than half of them assimilated sea ice observations into their initial conditions as of 2019 (Bhatt et al. 2020). The Geophysical Fluid Dynamics Laboratory (GFDL) forecast model is among the other half that have not constrained their initializations with sea ice observations. This paper aims to build a sea ice DA framework within a GFDL sea ice/ocean model, find the proper DA methods to maximize the utility of SIC observations, explore the benefits of improved sea ice initial conditions for summertime Arctic sea ice predictions at short lead times, and lay foundations for DA of other observation types in the future.

In this study, we assimilate satellite SIC observations within the GFDL sea ice–ocean model. We develop an SIC DA control run, and test various DA configurations to explore how sensitive our DA results are to the choices we make. We describe the model setup and data assimilation system, observations assimilated, experimental design, and evaluation methods in section 2. The DA results are discussed in section 3, followed by the summary and conclusions in section 4.

2. Data and methodology

a. The sea ice–ocean model and data assimilation system

The sea ice simulator version 2 (SIS2; Adcroft et al. 2019) developed at the Geophysical Fluid Dynamics Laboratory is used in this study. SIS2 employs an elastic–viscous–plastic rheology to calculate internal ice forces (Hunke and Dukowicz 1997) and an energy-conserving thermodynamic scheme with four ice layers and one snow layer, similar to Bitz and Lipscomb (1999). Multiple ice thickness categories are simulated, between which sea ice moves due to dynamic or thermodynamic changes (Bitz et al 2001). The default setting of five sea ice thickness categories is used, with category boundaries of 0.1, 0.3, 0.7, and 1.1 m. SIS2 is coupled with GFDL's Modular Ocean Model version 6 (MOM6) (Adcroft et al. 2019) and is forced by the Japanese 55-year Reanalysis (JRA55-do) (Tsujino et al. 2018) that spans from 1958 to 2017.

We create an interface between SIS2 and the Data Assimilation Research Testbed (DART) to conduct data assimilation (DA) experiments. DART is a software developed at the National Center for Atmospheric Sciences (NCAR) (Anderson et al. 2009). DART provides a variety of ensemble-based data assimilation algorithms and has been linked with all the components of the Community Earth System Model (CESM) for atmosphere (Raeder et al. 2012), ocean (Danabasoglu et al. 2012), land (Zhang et al. 2014), and sea ice (Zhang et al. 2018) DA.

TABLE 2. List of experiments with different configurations.

Experiment name	SST restoring	Assimilation of SIC	Localization half-width (radians/ approximated in km)	Observation error	DA frequency	Experiment period
FREE	N	N	—	—	—	1982–2017
noDA-SSTrest	Y	N	—	—	—	1982–2017
DA-noSSTrest	N	Y	0.03/190	10%	Every 5 days	1982–2017
DActr	Y	Y	0.03/190	10%	Every 5 days	1982–2017
DALocS	Y	Y	0.01/64	10%	Every 5 days	1982–91
DALocL	Y	Y	0.05/320	10%	Every 5 days	1982–91
DAErrS	Y	Y	0.03/190	5%	Every 5 days	1982–91
DAErrL	Y	Y	0.03/190	15%	Every 5 days	1982–91
DAFreqH	Y	Y	0.03/190	10%	Every day	1982–91
DAFreqL	Y	Y	0.03/190	10%	Every month	1982–91

The SIS2/MOM6 model writes out restart files when observations are to be assimilated and sends the selected model state variables from the restart files to DART. In this study, the state variable is the SIC of each category (SICN). The thickness of each category remains unchanged as suggested by previous work (Kimmritz et al. 2018; Zhang et al. 2018). DART then calculates the aggregate SIC in the observation operator and combines it with observations through the filter to generate the analysis states. Certain postprocessing is necessary before the analysis states replace the model states in the restart files to avoid unphysical values generated from the filter. When the analysis aggregate SIC is greater than one, SICN is scaled proportionally to make the aggregate SIC equal one. When the analysis SIC goes negative, SICN is set to zero. A five-layer sea ice profile is created when sea ice is added by the filter in places where sea ice was absent. The added SICN has the middle-point thickness in its category and is snow-free (the midpoint thickness given in category 5 is 1.3 m). The added sea ice has a bulk salinity of 5 ppt, and its temperature is calculated based on the liquidus relation as a function of brine salinity from Assur (1958). The updated restart fields serve as the initial conditions for the next DA cycle.

We use the ensemble adjustment Kalman filter (EAKF) (Anderson 2001), a variant of the ensemble Kalman filter (EnKF) (Evensen 1994), to assimilate observations in this study. The EAKF does not perturb the observations as the traditional EnKF does and shows outperformance when the ensemble size is small (Anderson 2001). The EAKF requires an ensemble of sea ice states to estimate the error statistics of the model state. The ensemble size we choose is 30 in this study. The ensemble of SIS2/MOM6 simulations is created by perturbing a set of sea ice parameters, including the ice strength parameter P^* from Hibler (1979) and the albedo parameters of snow R_snw, ice R_ice, and pond R_pnd from Briegleb and Light (2007). The sea ice strength parameter P^* is chosen randomly from a uniform distribution spanning from 20 000 to 50 000 N m^{-1}, which covers the default value of 27 500 N m^{-1}. The albedo parameters are each chosen independently from a uniform distribution spanning from −1.6 to 1.6 standard deviations, while the default value is 0. The resulting ensemble spread of sea ice concentration is discussed in section 3a.

b. Observations

The SIC products from *Nimbus-7* SMMR and DMSP SSM/I-SSMIS passive microwave data downloaded from the National Snow and Ice Data Center (NSIDC) website are used in the study. The product retrieved using the NASA Team (NT) algorithm (https://nsidc.org/data/nsidc-0051) is assimilated and the product retrieved using the bootstrap (BT) method (https://nsidc.org/data/nsidc-0079) is used for additional evaluation.

The daily Optimum Interpolation Sea Surface Temperature version 2 (OISSTv2) (Reynolds et al. 2007; Banzon et al. 2016) product is used to constrain SST. We apply the SIC data that comes with OISST to mask out SST values at grid cells where SIC is larger than 30%.

c. Experimental design

We list the configurations of all the experiments in Table 2. The FREE experiment does not assimilate any observations but has the same perturbed parameters as the DA experiments, with each ensemble member having a unique set of the four parameters. To reduce thermohaline drift, the surface salinity is restored globally to a climatology based on observations at the piston velocity of 1/6 m day^{-1} (Griffies et al. 2009) corresponding to a time scale of 300 days for a 50-m length scale (Table 2 in Danabasoglu et al. 2014). Salinity restoring is applied to all the experiments, both in the open ocean and under sea ice.

The experiment DA-noSSTrest assimilates the NSIDC SIC observations every 5 days but has no SST restoring. The observational error we choose is 10%, which is within the range of the values used in previous SIC DA studies (e.g., Lisæter et al. 2003; Kimmritz et al. 2018) and agrees reasonably well with error estimates of SIC observational products (e.g., Meier 2005). The horizontal localization half-width distance is 0.03 radians (about 190 km).

To ameliorate an overshooting problem identified in DA-noSSTrest (the problem is further discussed in later sections), we restore the SST to the daily OISST value with a piston velocity of four meters per day. We apply a sea ice mask to the observed SST data so that SST is only restored to observations where SIC is below 30%. The sea ice mask is derived from the SIC data in the OISST product and is applied to avoid using unreliable SST observations in the Arctic region in the version

of OISST product we use (Lu et al. 2020). The threshold value of 30% is chosen based on sensitivity tests (30% performs the best among the threshold values of 10%, 25%, and 30%). In the regions where SIC is above 30%, we restore SST to the freezing point of seawater calculated based on the modeled salinity. We conduct an experiment (noDA-SSTrest) with SST restoring but no SIC DA to evaluate the sole effects of SST restoring on the sea ice field.

The DActr is our DA control experiment that has the same DA configurations as DA-noSSTrest and does SST restoring with the same piston velocity as noDA-SSTrest. We also conducted a set of DA sensitivity experiments stemming from DActr to investigate the influence of the localization half-width, observation error, and DA frequency on the DA results. The experiments DALocS and DALocL use a smaller and a larger localization half-width, 0.01 radians (about 64 km) and 0.05 radians (about 320 km), respectively. The observation errors in the experiments DAErrS and DAErrL are 5% and 15%, respectively. The experiments DAFreqH and DAFreqL assimilate observations every day and every month, respectively.

A single-member SIS2/MOM6 simulation forced by JRA55-do is run from 1958 to 1979 to provide a "spun-up" ice-ocean state to be used as initial conditions for the ensemble simulations. A separate SIS2/MOM6 experiment that has 30 ensemble members with the same set of perturbed parameters is then run from 1979 to 1981, from which the ensemble of initial conditions is obtained for the DA experiments. The experiments FREE, noDA-SSTrest, DA-noSSTrest, and DActr are each run for 36 years from 1982 to 2017, and the set of DA sensitivity experiments run for 10 years from 1982 to 1991.

d. *Two additional experiments from GFDL's seasonal prediction systems*

We compare our DA experiments with two additional experiments from the current and next-generation versions of GFDL's seasonal prediction system. The ensemble coupled data assimilation (ECDA) system (Zhang et al. 2007) is based on GFDL's fully coupled climate model CM2.1 (Delworth et al. 2006) with atmosphere constrained by the NCEP–NCAR reanalysis and ocean constrained by temperature and salinity profiles and satellite SST observations (Zhang et al. 2007; Chang et al. 2013; Xue et al. 2017). The sea ice state in this system is constrained via fluxes associated with assimilation in the atmospheric and oceanic components (Bushuk et al. 2019b). The Seamless System for Prediction and Earth System Research (SPEAR) is a new prediction system developed at GFDL with updated versions of the atmosphere, land, ocean, and sea ice component models (Delworth et al. 2020). The configuration of the sea ice and ocean models in SPEAR is the same as that of this study except that our study perturbs four sea ice parameters as described in section 2a. The initial conditions for sea ice in this system come from an experiment that nudges atmosphere and SST toward the Climate Forecast System Reanalysis and daily OISST (SPEAR-Nudged hereafter; Lu et al. 2020).

Both ECDA and SPEAR-Nudged constrain the atmosphere by reanalysis and ocean by observations. They are most comparable to the experiment noDA-SSTrest in terms of their configuration. Comparing our DA experiments with ECDA and SPEAR-Nudged provides insights on the relative improvements from more advanced model components and the additional information from sea ice DA.

e. *Evaluation methods*

The root-mean-square error (RMSE) and mean absolute error (MAE) are used to evaluate the general performance of all the experiments. They are defined as follows:

$$\mathrm{RMSE}_j = \sqrt{\frac{\sum_{i=1}^{N}(x_{i,j}-y_{i,j})^2}{N}};\ \mathrm{RMSE}_{\mathrm{region}} = \frac{\sum_{j=1}^{M}\alpha_j \mathrm{RMSE}_j}{\sum_{j=1}^{M}\alpha_j};$$

$$\mathrm{MAE}_{\mathrm{region},i} = \frac{\sum_{j=1}^{M}\alpha_j|x_j-y_j|}{\sum_{j=1}^{M}\alpha_j},$$

where RMSE_j is the RMSE of SIC at the jth grid cell, $\mathrm{RMSE}_{\mathrm{region}}$ is the mean RMSE of SIC for a specified subregion, $\mathrm{MAE}_{\mathrm{region},i}$ is the MAE of SIC for a subregion in the ith month, x is the ensemble mean of the monthly SIC for each grid cell from the model experiments, y is the observed SIC, i is the index in time, j is the index in space, N is the number of years, M is the number of grid cells in a specified subregion, and α_j is the areal weight of the jth grid cell. We do not evaluate individual ensemble members so x always refers to the ensemble mean.

Model bias is defined as the 30-member ensemble mean of the model outputs minus the observations. The detrended Pearson correlation is defined as follows:

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

where R is the detrended correlation between model x and observation y, i is the index in time, $\overline{x}$ and $\overline{y}$ are linear trend fits to the data, and N is the sample size.

3. Results

a. *The ensemble spread of sea ice concentration*

For data assimilation with ice–ocean models or ice-only models, perturbing only initial conditions is not sufficient to maintain a reasonably large ensemble spread. Different approaches have been used to generate ensemble members and maintain ensemble spread in previous studies. The majority force the ice models with an ensemble of atmospheric fields, by adding random fields to a single atmospheric forcing (e.g., Lisæter et al. 2003; Mathiot et al. 2012; Massonnet et al. 2015), employing an ensemble atmospheric forcing coming from an ensemble-based reanalysis product (e.g., Zhang et al. 2018; Yang et al. 2015), or adding perturbations from an ensemble

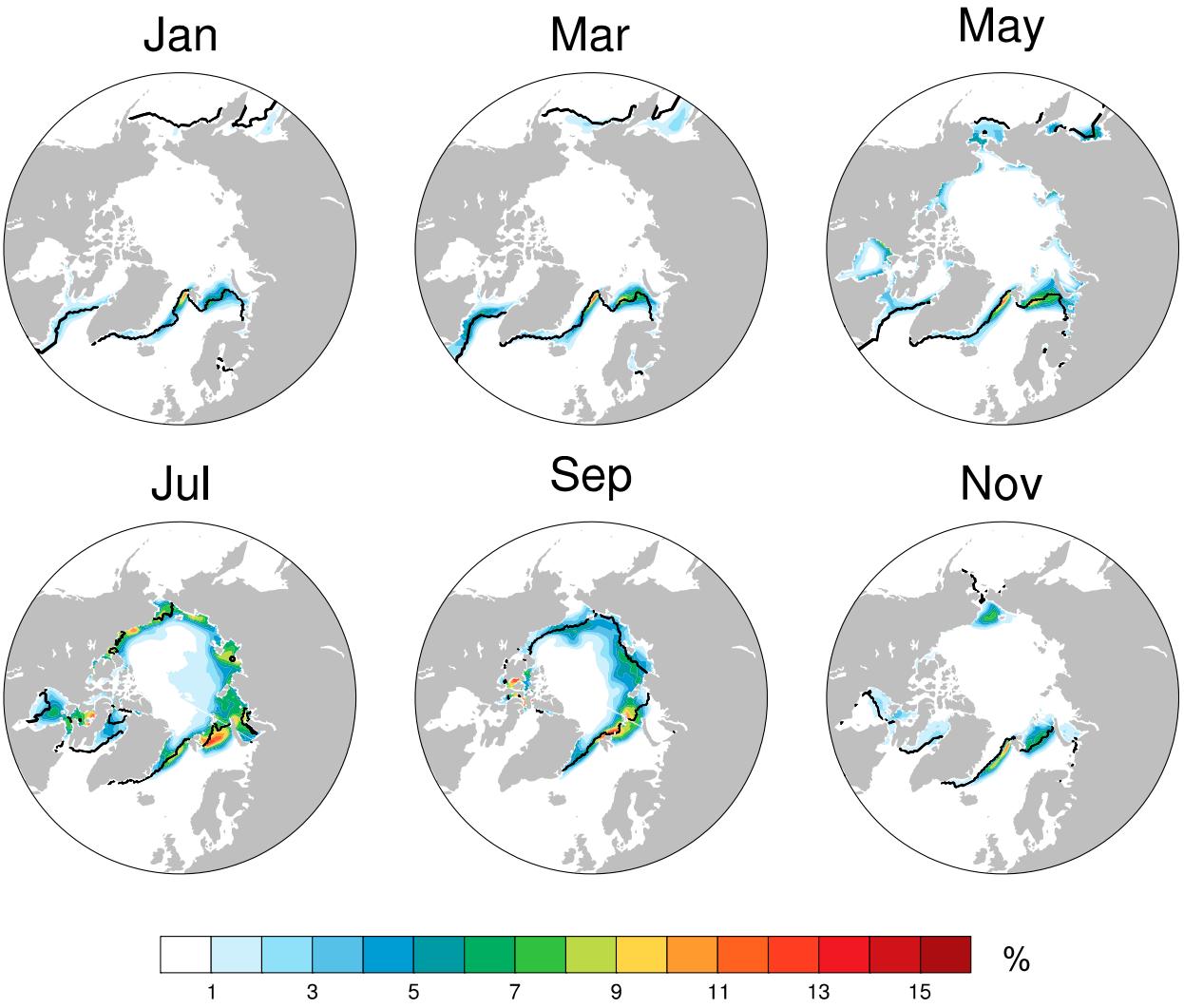

FIG. 1. The 36-yr mean ensemble spread for every other month from the experiment FREE with black contours representing the climatological sea ice edge from the NSIDC NT observations from 1982 to 2017.

reanalysis system to a deterministic forcing (e.g., Shlyaeva et al. 2016). Very few studies consider uncertainty from model physics. Zhang et al. (2018) perturbed an albedo parameter and a sea ice strength parameter in addition to forcing the ice model with an ensemble atmospheric reanalysis and found that perturbing model physics increases ensemble spread throughout the year but has little effect in winter along the sea ice edges. Shlyaeva et al. (2016) found that perturbing an ice strength parameter, an ice-albedo parameter, and ice–atmosphere and ice–ocean drag coefficients increases the ensemble spread of SIC in the 0%–10% and 90%–100% ice concentration areas. Since the parameters are static, it also helps sustain the ensemble spread of the model states over time.

We choose to perturb an ice strength parameter and three albedo parameters and force the ice–ocean model with a single atmospheric forcing (JRA-55-do). To evaluate the general performance of our perturbation method, we show the 36-yr-average ensemble standard deviation of SIC every other month in Fig. 1. The spatial pattern of the ensemble spread looks very similar to those of other studies: the spread is the largest in the ice marginal zones and small in the internal ice pack throughout the year. The ensemble spread is generally sufficiently large to encompass the observed sea ice edge, suggesting that the filter should be well behaved when assimilating observations. A notable exception to this is the Pacific sector in winter, where the ensemble spread is quite limited. Perturbing more parameters, for example, the drag coefficients, may introduce more spread in winter. But examining the effects of different approaches on the SIC ensemble spread is outside the scope of this study, which focuses primarily on summer Arctic sea ice.

b. Evaluation of the sea ice DA experiments

Figure 2 shows the seasonal cycle of $\mathrm{RMSE}_{\mathrm{region}}$ of SIC (calculated against the NSIDC NT observations) for seven selected subregions and the pan-Arctic from all the experiments. RMSEs result from a combination of (i) climatological

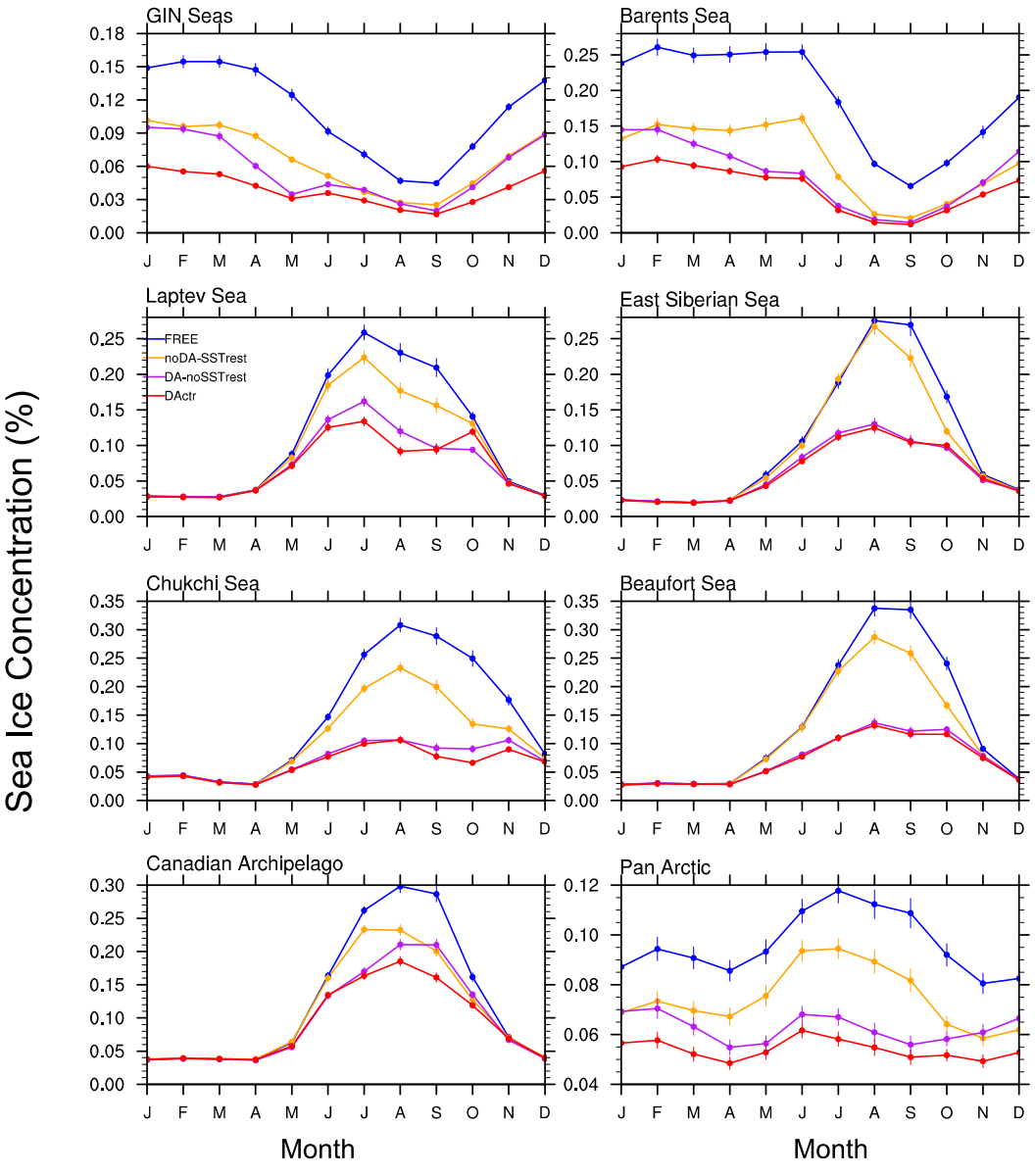

FIG. 2. The 36-yr mean seasonal cycle of $RMSE_{region}$ of SIC. $RMSE_{region}$ of SIC is calculated against the NSIDC NT observations for each month across the selected the pan-Arctic regions for the experiments FREE (blue), noDA-SSTrest (orange), DA-noSSTrest (purple), and DActr (red). Each error bar represents two standard deviations of $RMSE_{region}$ calculated from a bootstrap procedure.

biases, (ii) errors in interannual variability, and (iii) errors in the trend. We begin by assessing the full RMSE in Figs. 2–4, and assess the contributions of these three factors in subsequent figures. The FREE experiment consistently has the largest $RMSE_{region}$ in all the regions throughout the year. The errors are mostly concentrated in the seasonal ice zones so $RMSE_{region}$ is relatively low in the winter months in all the regions except the Greenland, Iceland, and Norwegian (GIN) Seas and Barents Sea (these regions are nearly ice-free in summer). Figure 3 displays the spatial pattern of $RMSE_j$ from our four experiments, together with ECDA and SPEAR-Nudged for reference. Figure 3 confirms that model errors are concentrated on the ice marginal regions. Large errors appear in the GIN Seas and the Barents Sea in March and June. In September, FREE has a hotspot in the Beaufort Sea, and ECDA has the largest errors around the Laptev and Kara Seas. The pan-Arctic averaged $RMSE_j$ values of noDA-SSTrest and SPEAR-nudged are similar and distinctly less than ECDA (Fig. 3). This indicates that, even without ocean data assimilation, the newer generation modeling system shows some improvement over the older ECDA system.

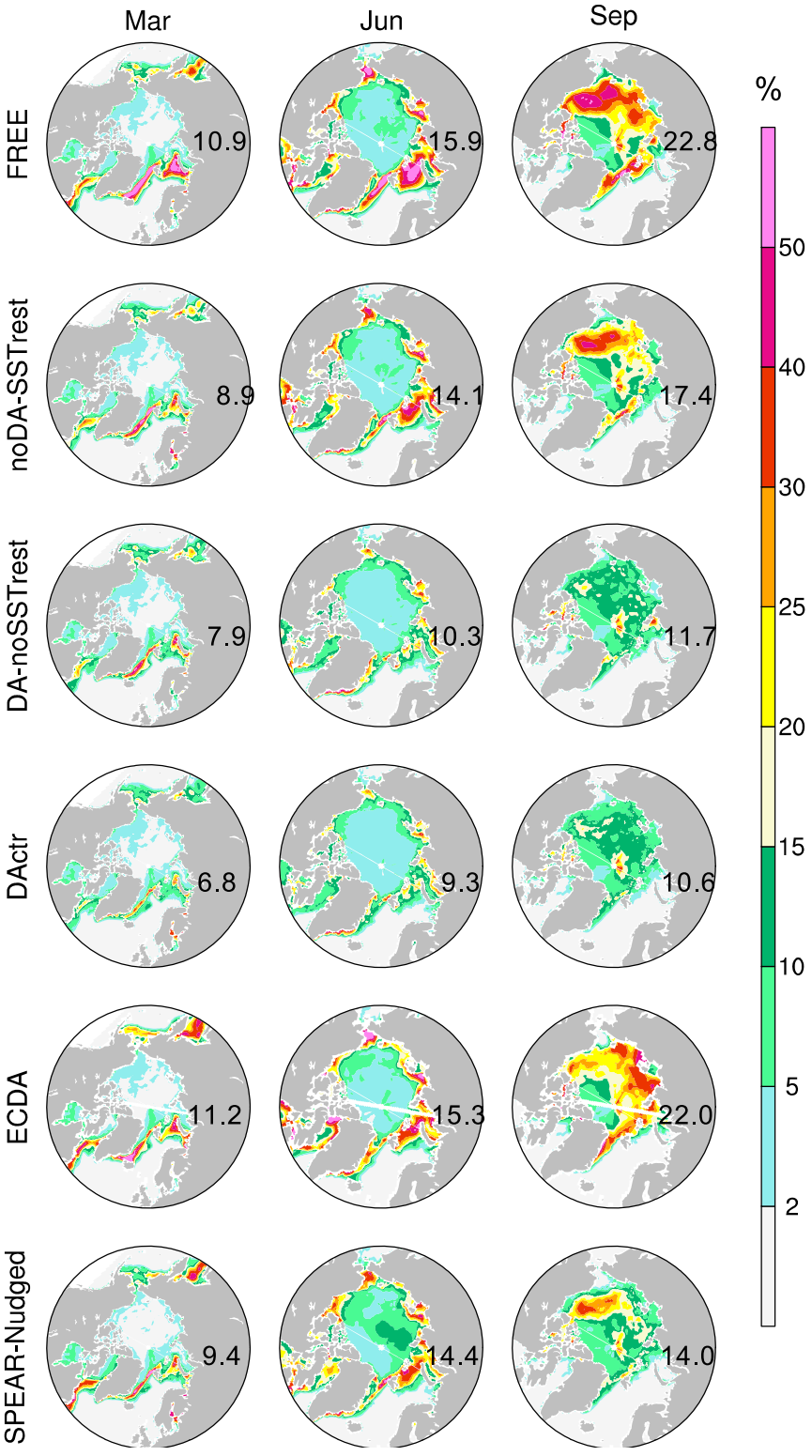

FIG. 3. The 36-yr mean March, June, and September $RMSE_{js}$ of SIC for our four experiments (FREE, noDA-SSTrest, DA-noSSTrest, and DActr), ECDA, and SPEAR-Nudged. Numbers on each plot indicate the $RMSE_{pan\text{-}Arctic}$. RMSEs are calculated against the NSIDC NT observations.

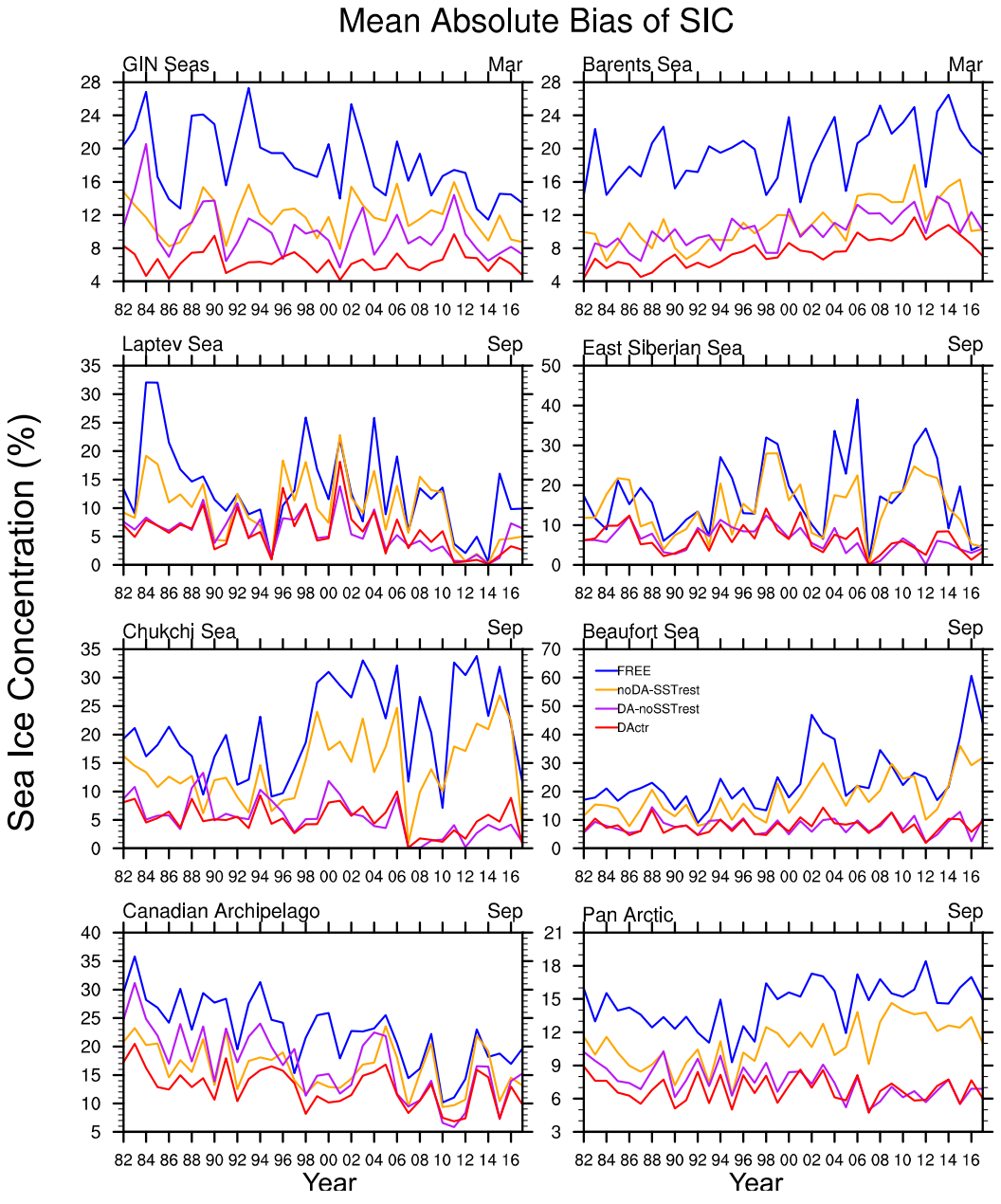

FIG. 4. March $MAE_{region,i}$ of SIC for the GIN Seas and the Barents Sea, and September $RMSE_t$ for the rest of the Arctic regions for the experiments FREE (blue), noDA-SSTrest (orange), DA-noSSTrest (purple), and DActr (red). $MAE_{region,i}$ is calculated against the NSIDC NT observations.

SST restoring reduces the $RMSE_{region}$ from FREE in all the regions (experiment noDA-SSTrest in Fig. 2), especially in the GIN Sea, Barents Sea, Beaufort Sea, Chukchi Sea, and East Siberian Sea (Fig. 3). Relative to the FREE experiment, the pan-Arctic averaged $RMSE_j$ decreases by ~18%, 11%, and 24% in March, June, and September, respectively (Fig. 3). The experiment SPEAR-Nudged has a very similar spatial pattern of SIC $RMSE_j$ to noDA-SSTrest. SPEAR-Nudged has slightly smaller errors in the Atlantic sector and larger errors in the Pacific sector in March and June. SPEAR-Nudged outperforms noDA-SSTrest in September, especially in the Beaufort and Chukchi Seas (Fig. 3).

The experiment DA-noSSTrest shows a more homogeneous error reduction across the whole Arctic, bringing the $RMSE_{region}$ below 15% in most areas except the GIN Sea and the Barents Sea in March (Fig. 3). Relative to the FREE experiment, the pan-Arctic averaged $RMSE_j$ is reduced by ~28%, 34%, and 49% in the three selected months. Combining SST restoring and SIC DA, the experiment DActr performs the best among all the experiments. It further reduces $RMSE_{region}$ to a similar

magnitude to the observation error (10%) in all the regions. Compared with DA-noSSTrest, the most obvious improvement is seen in the GIN Seas, Barents Sea, and the Canadian Archipelago (Figs. 2 and 3). The pan-Arctic averaged $RMSE_j$ is reduced by ~38%, 42%, and 54% in the three months as compared with those of FREE.

We look at the interannual variations of $MAE_{region,i}$ of SIC in Fig. 4, focusing on March in the GIN Seas and Barents Sea, and on September in the other regions. In general, the performances of the DA experiments have similar features to those in Fig. 2. FREE has the largest errors throughout the experiment period, noDA-SSTrest reduces error from FREE in all the regions, and DA-noSSTrest reduces more biases. In September, DActr does not show an obvious improvement over DA-noSSTrest except in the Canadian Archipelago and the first 20 years of the experiment in the pan-Arctic. In March, it is clear that SST restoring provides additional error reduction to SIC DA in the GIN Seas and the Barents Sea.

The RMSE and MAE values provide an overall SIC error, whereas biases show the systematic differences between the model and observations. Figure 5 displays the seasonal cycle of regional and pan-Arctic sea ice extent (SIE) for the DA experiments. Both NSIDC SIC products are plotted, which offers a brief view of the uncertainty in observations. The SIE of a region is defined as the total area of grid cells in that region where ice coverage exceeds a certain threshold. We use the same threshold as the NSIDC sea ice index products, which is 15%. The pan-Arctic SIE from FREE has an overall positive bias throughout the year. The positive biases show up in the GIN Seas and the Barents Sea in winter and become prominent in the Beaufort Sea and the Chukchi Sea in summer (Fig. 5). The spatial pattern of SIC bias shown in Fig. 6 confirms that FREE in general overestimates sea ice cover. SST restoring (the experiment noDA-SSTrest) reduces the positive bias in SIE in all the regions (Figs. 5 and 6).

The March and June positive biases in SIC in the Barents Sea and September positive biases in SIC in the Beaufort Sea and the Chukchi Sea are almost removed by SIC DA (the experiment DA-noSSTrest in Fig. 6). However, the negative biases in FREE still show up in DA-noSSTrest; for example, the red spots in the Laptev Sea, East Siberian Sea, and Kara Sea in FREE can also be found in DA-noSSTrest. While the negative errors and positive errors in SIC partly cancel out in FREE, the asymmetric correction on the biases from SIC DA could lead to degradation in SIE. This partly explains why DA-noSSTrest shows worse SIE than FREE in several months in those regions (Fig. 5), but shows overall reductions in RMSE of SIC (Fig. 2). This suggests that SIC DA is able to efficiently remove sea ice but is less effective at adding sea ice. The addition of sea ice requires both SST restoring and SIC DA, as evidenced by the improved performance of DActr compared to the DA-noSSTrest experiment.

We also identify the problem that SIC DA sometimes overcorrects the positive biases and flips the sign to negative, which can be found in the GIN Seas in June and in the East Siberian Sea in September (Fig. 6). This overshooting problem also contributes to the degradations of SIE in those regions and months (Fig. 5). This happens when DA abruptly removes sea ice and triggers the ice-albedo feedback. The reduction of sea ice decreases the surface albedo, which increases the amount of solar radiation that is absorbed by the ocean. This warms up SST, enhances bottom melting, and reduces more sea ice (see Fig. S1 in the online supplemental material). However, if we restore the SST to the observations (OISST), the ice-albedo feedback is damped and the overshooting problem is ameliorated as shown in Figs. 5 and 6.

The time series of March and September SIE in Fig. 7 show the SIE trends and interannual variations captured by the different experiments and observations. The two observations show quite consistent trends. We show linear trends of the March and September SIE in Fig. 8. FREE generally underestimates the decreasing trends. The experiment DA-noSSTrest has better trends of September SIE in the East Siberian Sea, Chukchi Sea, Beaufort Sea, and pan-Arctic region. By adding SST restoring, the experiment DActr does not perform as well as DA-noSSTrest in terms of trends. Figure 7 shows that SST restoring is mainly influential in the first 20 years if we compare noDA-SSTrest to FREE and DActr to DA-noSSTrest.

To figure out why SST restoring shows different performance in the first and second halves of our experiment period, we look at the spatial maps of the SIC bias in selected months in 1990 (Fig. S2) and 2010 (Fig. S3), which show quite different patterns (similar spatial patterns are found in other years within each period). SIC bias in the FREE experiment in 1990 is generally positive in the Beaufort, Chukchi, and Barents Seas, and negative in the East Siberian, Laptev, and Kara Seas. The negative biases are largely reduced in noDA-SSTrest. DA-noSSTrest, however, does not correct the negative biases or even increases the magnitude of the negative biases, while it reduces the positive biases effectively. DActr combines the advantages of the two and outperforms all the other experiments. By contrast, the SIC bias in FREE is mostly positive in the year 2010. While SIC DA alone is very effective at removing sea ice, there is not much residual error for SST restoring to correct, which partly explains why DActr does not show significant improvements from DA-noSSTrest, especially in summer. Another cause is the asymmetric correction from SST restoring due to the fact that we apply a sea ice mask to the observed SST data so that SST is only restored to observations where SIC is less than 30%. When a negative SIC bias exists (SST has a warm bias accordingly), restoring to the colder SST helps form new sea ice; however, when a positive SIC bias exists in regions where the observed SIC is greater than 30% (SST might have a cold bias; this mostly happens in summer), SST is restored to a salinity-based freezing point, which does not promote sea ice melting. In another situation, when a positive SIC bias exists in regions where the observed SIC is less than 30% (this mostly happens in winter), SST is restored to the OISST Daily data and hence will enhance sea ice melting. In summary, SST restoring can correct negative SIC biases effectively year-round, which complements the SIC DA. Its effectiveness on positive SIC biases is more complicated and depends on whether sea ice is present in the biased regions, which varies with space and time.

Figure 7 shows that compared to the satellite observations, the experiments can capture the interannual variations of September SIE in the Arctic. We further look at the detrended

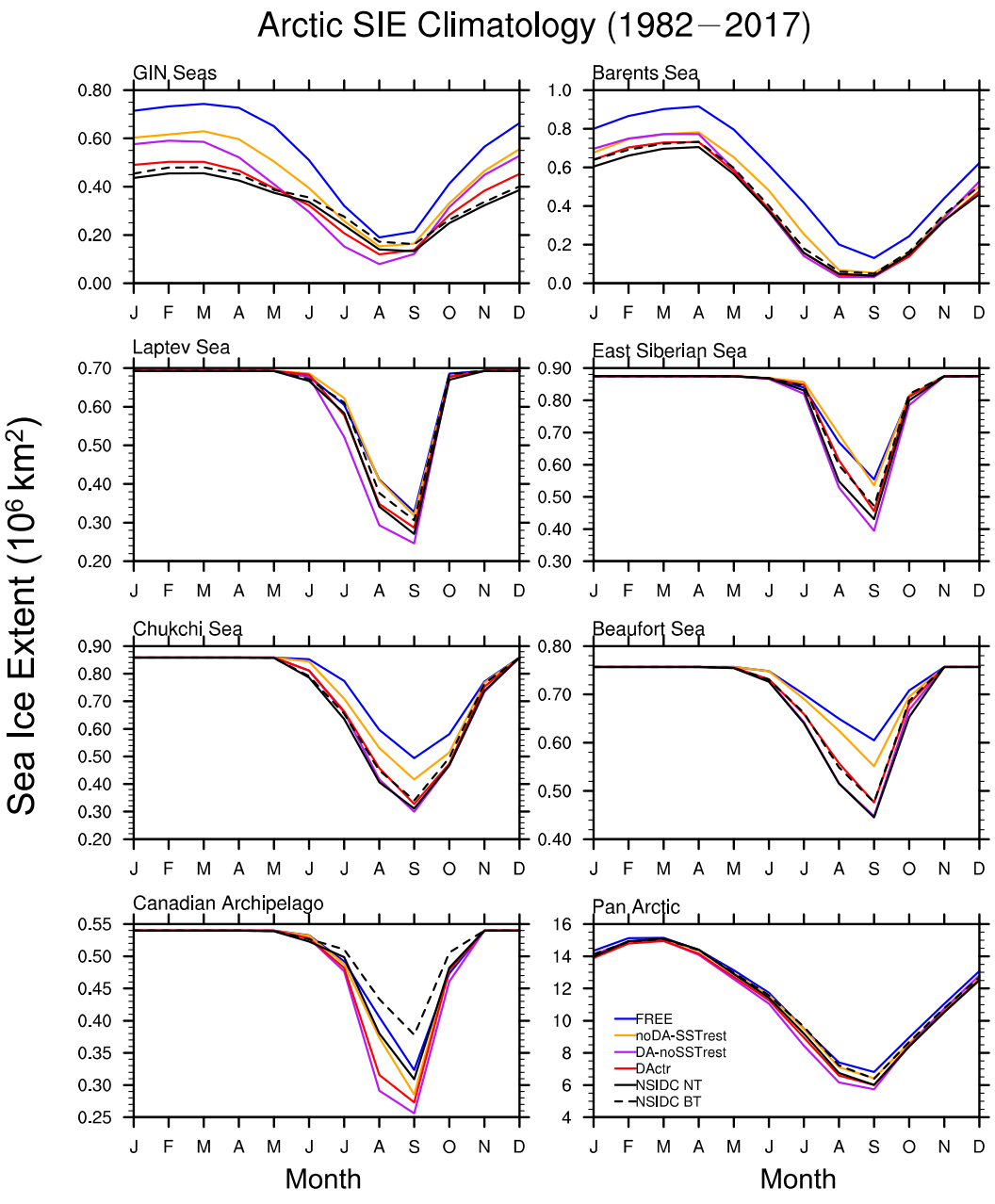

FIG. 5. Seasonal cycle of regional and pan-Arctic sea ice extent (SIE) for the experiments FREE (blue), noDA-SSTrest (orange), DA-noSSTrest (purple), and DActr (red), the NSIDC NT observations (black solid), and the NSIDC BT observations (black dashed).

correlations between the modeled SIE and the observed SIE for different months (Fig. 9). Since SPEAR-Nudged provides sea ice initial conditions for the new seasonal prediction system developed at GFDL, a comparison between SPEAR-Nudged and DActr will provide insights into the potential improvements brought by SIC DA. As expected, FREE generally has the lowest correlation. SPEAR-Nudged and noDA-SSTrest have similar correlation values in general. DA-noSSTrest improves the correlations and DActr consistently performs the best. The improvement seen from noDA-SSTrest (or SPEAR-Nudged) to DActr suggests that using initial conditions from DActr has the potential to substantially improve the short-term prediction skill of Arctic sea ice.

c. *The sensitivity of DA results to different DA configurations*

The DA configurations we test in this study include the localization half-width, observation error, and DA frequency. Previous studies generally set up DA configurations based on assumptions and rarely explore the sensitivity of sea ice DA

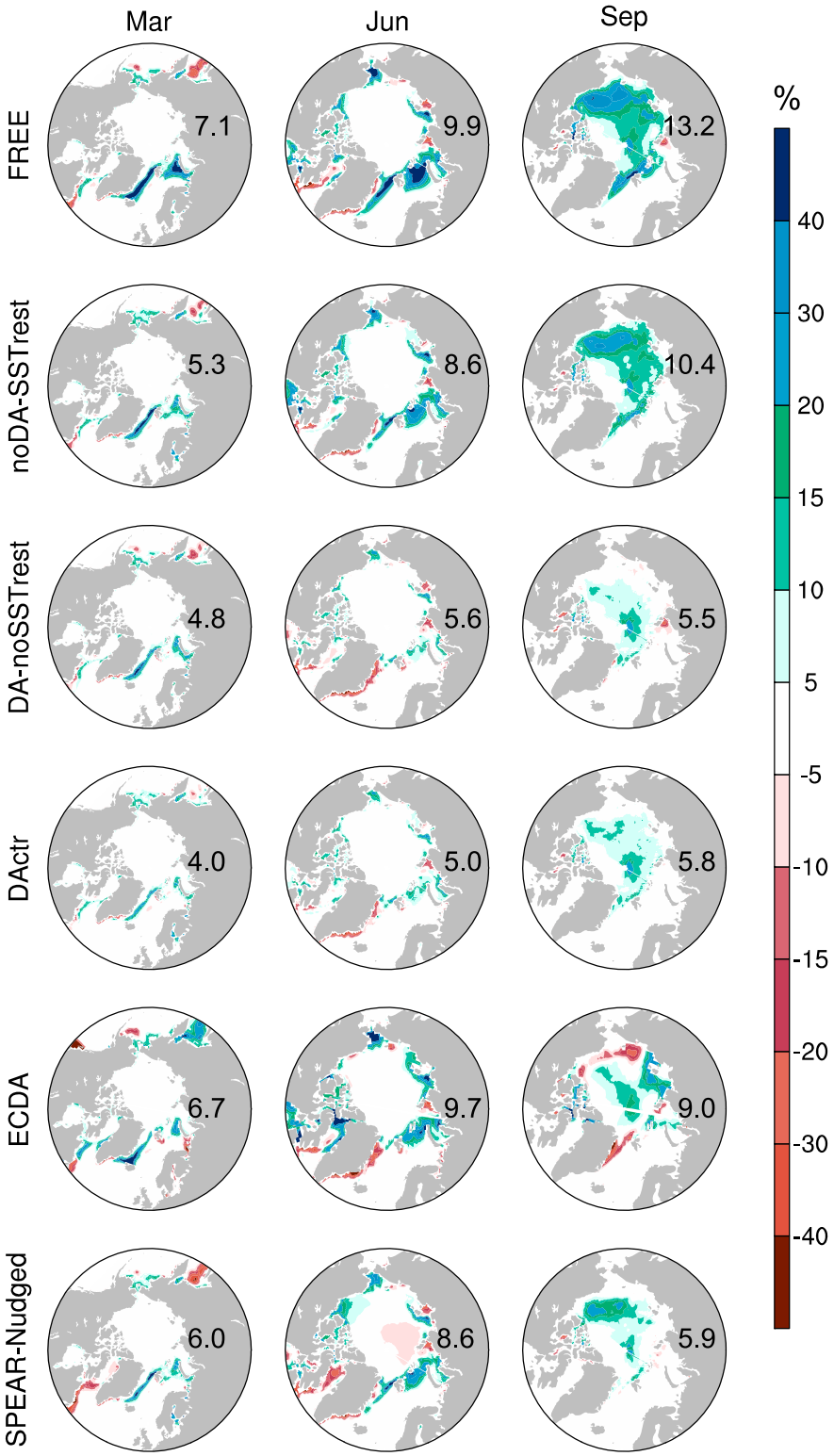

FIG. 6. Mean March, June, and September biases of sea ice concentration (SIC) for four of our experiments, ECDA, and SPEAR-Nudged. Bias is calculated as model minus the NSIDC NT observations. Numbers on the map are the area-weighted average of absolute mean biases over the grid cells where SIC either from observation or models is higher than 1%.

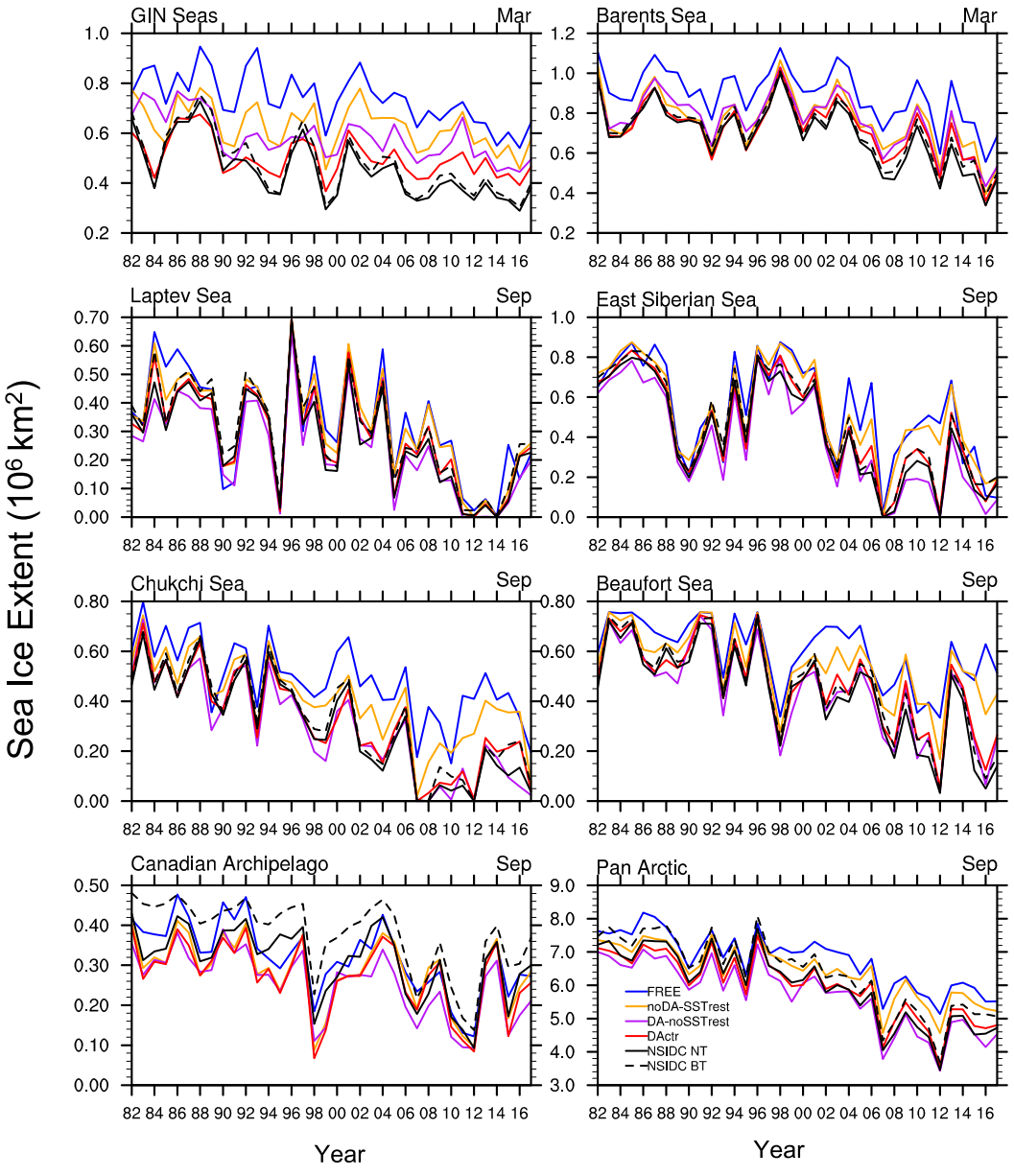

FIG. 7. Time series of regional and pan-Arctic September sea ice extent (SIE) from 1982 to 2017 for FREE (blue), noDA-SSTrest (orange), DA-noSSTrest (purple), DActr (red), the NSIDC NT product (black solid), and the NSIDC BT product (black dashed).

performance to those configurations. We aim to analyze the relative importance of each choice and provide guidance for future sea ice DA systems. The 10-yr mean RMSE of selected months for each set of experiments is shown in Fig. 10. The DA performance is most sensitive to DA frequency across all the regions. Although its sensitivity depends on seasons, increasing the DA frequency has a clear advantage. The result suggests that assimilating SIC observations every month is not favorable (however, it still shows significant improvements over FREE), and that daily assimilation provides the best results. We assimilate observations every 5 days in this study as a compromise of the computational cost and DA performance. Higher-frequency DA is suggested in future online sea ice DA systems that are embedded in the sea ice-ocean models and do not require file input and output in the DA step. This has been done efficiently in some coupled sea ice–ocean model frameworks (e.g., Mu et al. 2020; Nerger et al. 2020).

Reducing the observation error from 10% to 5% also improves our DA performance, while increasing the error to 15% always produces larger SIC RMSE. The current satellite retrievals of SIC observations have uncertainties larger than 5%, and can be as high as 20% over thin ice (Ivanova et al. 2015).

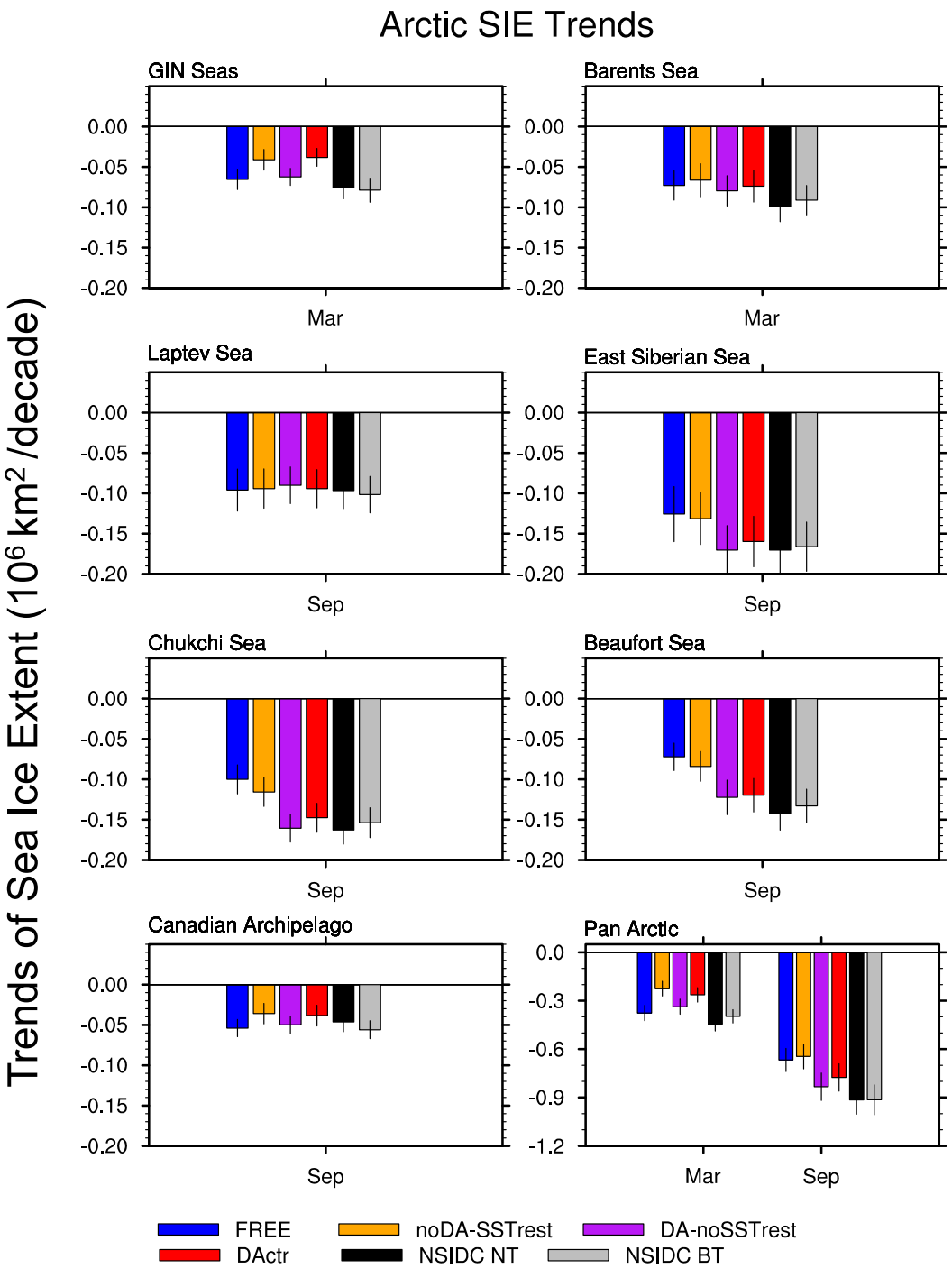

FIG. 8. Linear trends of SIE in the Arctic regions for March and September for FREE (blue), noDA-SSTrest (orange), DA-noSSTrest (purple), DActr (red), the NSIDC NT product (black), and the NSIDC BT product (gray). Each error bar represents one standard deviation of the linear regression coefficient (the trend). The experiments are conducted from 1982 to 2017.

The DAErrS experiment result indicates that improving the quality of observations provides the potential for a better SIC DA result as such improvement would provide a justification for using a smaller observational error.

The DA performance is least sensitive to the localization half-width. The strictest localization (DALocS) has a half-width of ~64 km and hence the influence of a given observation will be spread over a few model grid cells within the radius of

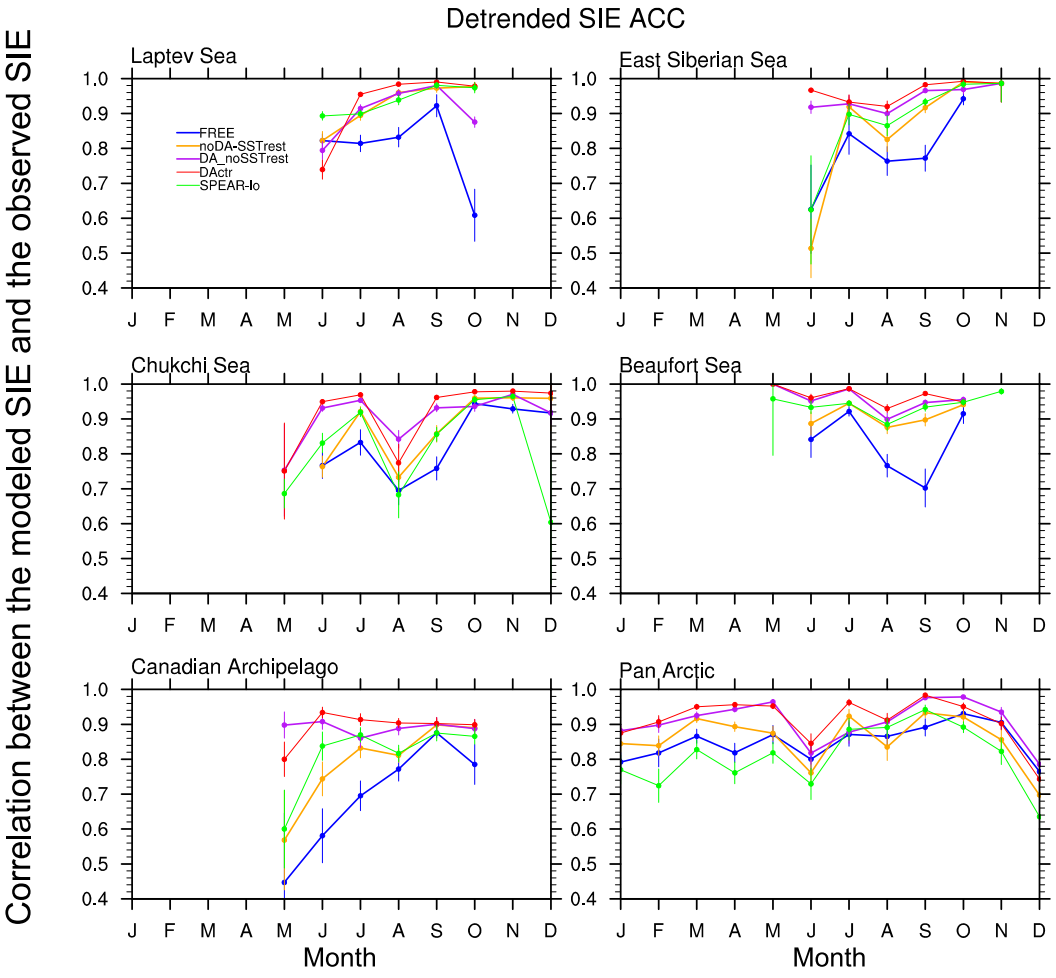

FIG. 9. Detrended correlations between the observed SIE and the modeled SIE for the experiments FREE (blue), noDA-SSTrest (orange), DA-noSSTrest (purple), DActr (red), and SPEAR-Nudged (green) for the Arctic regions. Months when the standard deviations of the observed SIE are larger than 0.01 million km^2 are shown. The NSIDC NT observations are used in the calculation. Each error bar represents two standard deviations of correlation coefficient calculated from a bootstrap procedure.

~128 km, given that the 1° grid has spacing of 40–50 km in the Arctic. By increasing the localization half-width, we let observations influence more neighboring grid cells. DALocS produces slightly yet persistently higher RMSE than DActr. Further relaxing localization by increasing the half-width from 0.03 radians (DActr) to 0.05 radians (DALocL) does not show an obvious reduction or even slight increase in RMSE, which is presumably due to the high density of the satellite observations. The sensitivity of DA to localization depends on the observation density, model resolution, and the correlation length scale of the observation. Hence the choice of a localization half-width of 0.03 radians that fits our model configuration may not be proper for other DA systems.

d. Impacts of SIC DA on SIT and sea surface salinity

The thickness of each category remains unchanged during DA, but the grid cell averaged SIT is updated by changing the ice concentration of each category. We compare the grid cell averaged SIT of our experiments with Pan-Arctic Ice Ocean Modeling and Assimilation System (PIOMAS) SIT in Fig. S4. The experiment FREE generally has thinner sea ice than PIOMAS except in the GIN Seas and Canadian Archipelago. Meanwhile, FREE has general positive SIC/SIE biases (Figs. 5 and 6), even in regions where SIT is thinner than PIOMAS SIT. As a result, the SIT of the DA experiments is further thinned relative to PIOMAS via the removal of ice during SIC DA. This occurs in most regions, except the GIN Seas and Canadian Archipelago where the removal of ice improves the FREE run's thick bias.

Whether SIC DA improves or degrades SIT is still debatable. First, it depends on how SIT is updated within the SIC assimilation procedure (Tietsche et al. 2013). The updating method that leads to significant improvement in SIT in Tietsche et al. (2013), however, creates a SIT drift in Kimmritz et al. (2018) and Zhang et al. (2018). It also depends on the consistency between the SIC/SIE bias and SIT bias in the model. The

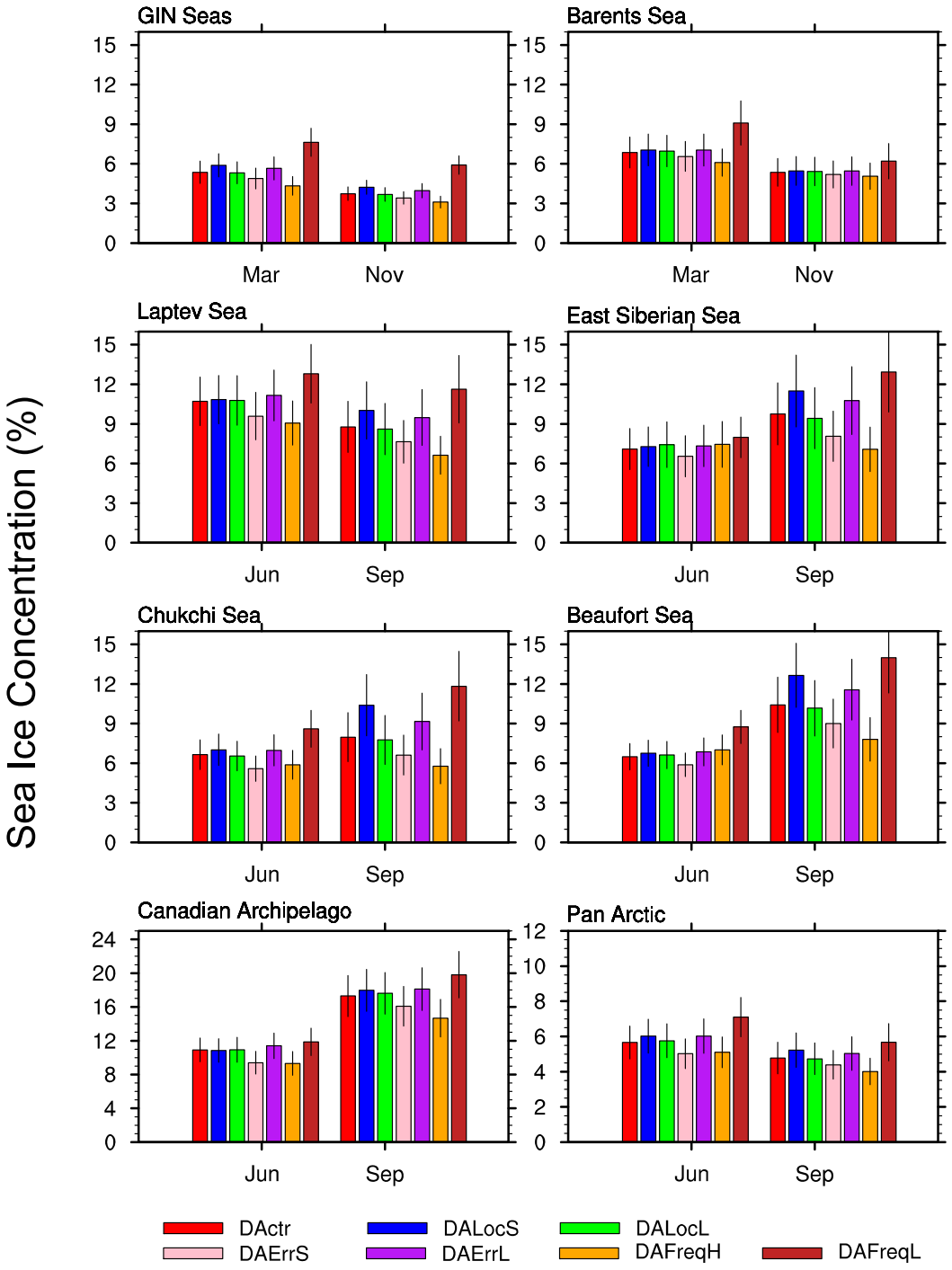

FIG. 10. $RMSE_{region}$ of sea ice concentration (SIC) in selected months for the Arctic regions for the DA experiments: DActr (red), DALocS (blue), DALocL (green), DAErrS (pink), DAErrL (purple), DAFreqH (orange), and DAFreqL (brown). $RMSE_{region}$ is calculated against the NSIDC NT observations from 1982 to 1991. Each error bar represents two standard deviations of RMSEregion calculated from a bootstrap procedure.

studies that achieved significant improvements in SIT tend to have consistent positive SIE/SIC and SIT biases (Mathiot et al. 2012) or negative SIE/SIC and SIT biases (Tietsche et al. 2013) in their models compared to the studies finding mild or no improvements in SIT (e.g., Kimmritz et al 2018; Zhang et al. 2018).

Our study chooses the SIT updating method that preserves the mean sea ice thickness of each category, which is

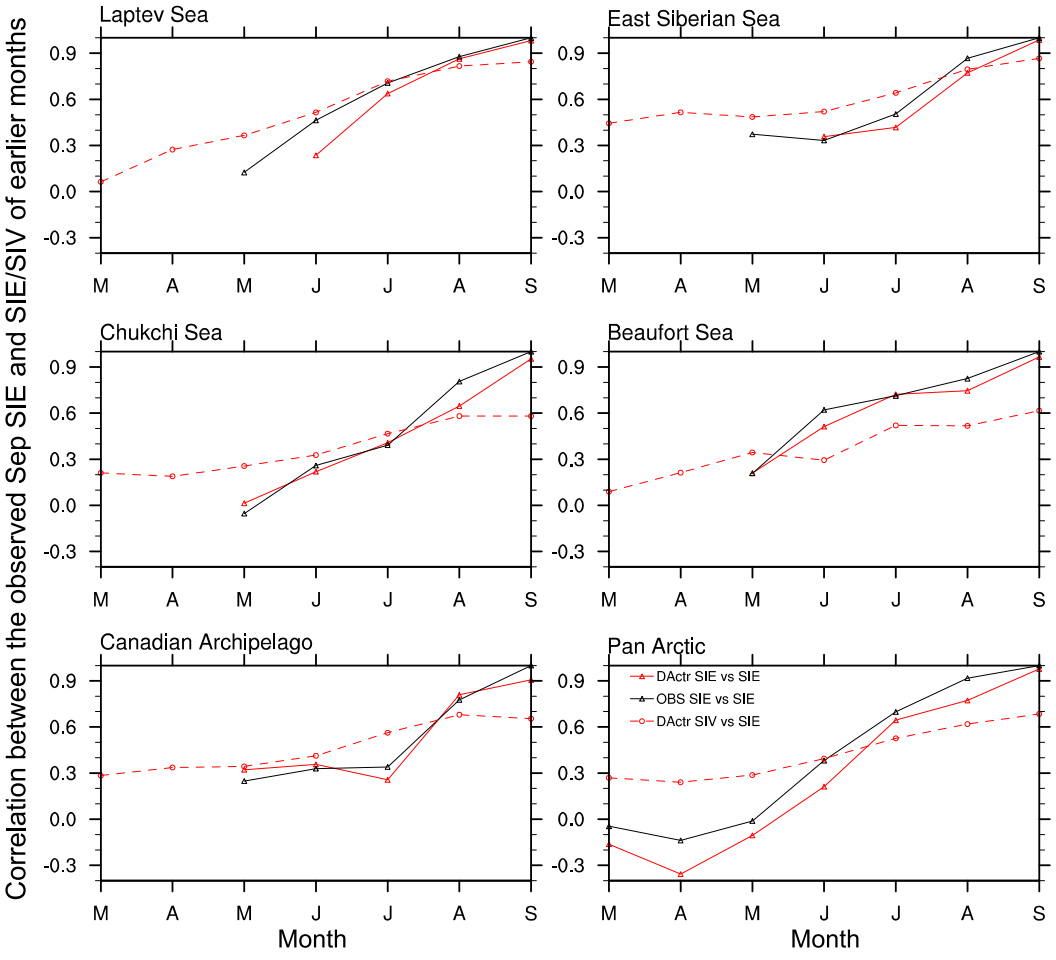

FIG. 11. Detrended correlations between the observed September SIE and the modeled SIE in earlier months (red solid), the modeled SIV in earlier months (red dashed), and the observed SIE in earlier months (black solid). The NSIDC NT observations are used in the calculation.

demonstrated not to cause SIT drift and has limited influence on SIT (Tietsche et al. 2013; Kimmritz et al. 2018; Zhang et al. 2018). The model in this study has opposite signs of biases in SIC/SIE and SIT. As a result of generally removing SIC during DA, the updated SIT is further thinned relative to the FREE experiment. Consequently, the DA sensitivity experiments that have better performance of SIC/SIE (DAFreqH and DAErrS) also have thinner biases than DActr (figures not shown).

We do not consider the conservation of freshwater or salt in the SIC DA. When sea ice is added or removed, the freshwater and salt are not modified accordingly, which is a common choice made in DA studies. As a result, the sea surface salinity (SSS) is generally saltier in the DA experiments than in FREE (Fig. S5), which is due to the general removal of sea ice by DA. The differences of SSS between the experiments are relatively small compared to their differences with the restored climatology (Fig. S4b). This shows that the impact of SIC DA on SSS is modest relative to the biases of the model. The central Arctic SSS in DActr is much fresher than the Atlantic influx water (~35 psu) and thus it is unlikely that SIC DA will cause spurious convection.

e. *Prospects for improved summertime predictions of Arctic sea ice*

Previous studies suggest that the anomaly persistence of SIE is the major predictor of summertime Arctic sea ice at the short lead times, while the anomaly persistence of SIT is the key for longer-term predictions (Blanchard-Wrigglesworth et al. 2011; Sigmond et al. 2013; Bonan et al. 2019). The lagged correlations between observed September SIE and SIE/SIT of earlier months from DActr (Fig. 11) are good indicators of the model's forecast skills at different lead times (Bonan et al. 2019). We also show the autocorrelation of SIE in the observation that indicates the persistence of SIE in reality. The correlation between the observed September SIE and SIE of earlier months from the model is very close to that of the observation. This confirms that our SIC DA has done well in capturing the anomaly persistence of SIE. It decays relatively quickly with increased lead time. The correlation between the

observed September SIE and the modeled SIV of earlier months is flatter. It has smaller values than that of SIE in the zero- to one-month lead times in most regions and has larger values in lead times longer than one month except over the Beaufort Sea and Pan-Arctic. This confirms that SIE is a better predictor for near-term September Arctic sea ice predictions but the prediction skill from the memory of SIE decreases with time more quickly than that from SIV. Beyond a few months, SIV is more important for prediction skill than SIE. Thus, our success in improving SIE provides bright prospects for improving summertime Arctic sea ice predictions at short lead times, and future work on SIT assimilation is needed to improve skill at lead times of two months and beyond. Dedicated dynamical forecast experiments that explicitly evaluate the impact of these improved SIC initial conditions on forecast skill are required to properly demonstrate this. This is planned as future work.

4. Conclusions and discussion

This paper documents the first effort in assimilating sea ice observations into the GFDL sea ice–ocean model. The GFDL Sea Ice Simulator version 2 (SIS2) is coupled with the Modular Ocean Model version 6 (MOM6) and driven by the Japanese 55-year Reanalysis (JRA-55-do) atmospheric forcing. SIS2 is linked with the Data Assimilation Research Testbed (DART) to conduct data assimilation (DA) experiments. The satellite-derived sea ice concentration (SIC) observations from the National Snow and Ice Data Center (NSIDC) NASA Team (NT) product are assimilated through an ensemble Kalman filter (EnKF) in this study. Although variants of EnKF have been applied in sea ice DA studies, they are not used extensively in the current modeling prediction systems.

Previous studies have demonstrated the capability of the GFDL prediction system in predicting summertime Arctic sea ice (Bushuk et al. 2017). One of the major hurdles in near-term prediction is the lack of knowledge of the sea ice initial conditions. With the SIC DA via the ensemble adjustment Kalman filter (EAKF), we show that the climatology, interannual variability, and trends of SIC and SIE are largely improved, which offers a more accurate sea ice initial condition for not only the pan-Arctic but also the critical subregions. The lag correlations between the September SIE and SIE/SIV of earlier months indicate that the persistence of SIE anomalies is essential to the summertime predictions of Arctic sea ice at lead times of zero to one month in all the regions. Hence the improved initial condition of SIE shows bright prospects for September predictions of Arctic sea ice at short lead times.

The assimilation of SIC is found to effectively reduce the positive errors of SIC in the model and capture interannual variations and trends with high skill. An overshooting problem is identified in some regions where sea ice is removed abruptly by DA and the ice-albedo feedback is triggered. This feedback amplifies the initial reduction in SIC and results in negative biases. To dampen the overshooting issue, we restore the sea surface temperature (SST) to the daily Optimum Interpolation Sea Surface Temperature (OISST), which shuts off the albedo feedback and largely improves the DA results. SST restoring can effectively correct negative biases, which offers complementary improvement to the SIC DA. However, when positive bias prevails, SST restoring is only effective in winter and does not show additional improvements in summer because (i) most of the positive biases are removed by the SIC DA and (ii) SST is restored to the local freezing point, which does not promote sea ice melting.

Analysis of the DA sensitivity experiments reveals that a variation in DA frequency shows the largest perturbation to the DA performance. The results suggest that daily SIC DA is favorable, while monthly SIC DA degrades the DA performance substantially. The experiment set with different observation errors shows that the SIC observations of higher quality have the potential to improve the SIC DA results. By relaxing the localization, we get slightly better results, but the improvement saturates as the localization half-width approaches the correlation length scale of the observation type.

The evaluation of the impact of SIC DA on sea ice thickness (SIT) shows that whether SIT is improved or degraded depends on the updating method and also the signs of biases in SIC and SIT. Our model has a positive bias in SIC and a negative bias in SIT. By removing SIC, we reduce SIT as well, which leads to a further thinned sea ice pack. Based on previous studies and our study, it is still debatable whether SIT can be improved by SIC DA. Studies with like-signed biases in SIC and SIT report SIT improvements, whereas this study with opposite-signed biases reports an SIT degradation. This implies the importance of assimilating SIT observations directly and we plan to conduct SIT DA in the future work. The influence of SIC DA on the sea surface salinity (SSS) is modest relative to the biases of the model.

In summary, our results agree with previous sea ice DA studies that applying an EnKF to assimilate SIC could largely improve the pan-Arctic sea ice extent. In addition, we evaluate the DA performance at regional scales and demonstrate that the trends, interannual variability, and climatology of the critical subregions are improved as well. During the sea ice DA practice, we identified an overshooting problem that may be common in the ocean–sea ice weakly coupled DA systems, which could cause degradations in sea ice DA performance. We suggest that applying SST restoring or DA in combination will ameliorate the overshooting issue. Our study also offers insights into how DA options may influence the performance, which has been rarely examined by previous studies. DA frequency causes the largest perturbation and daily DA is recommended as computational resources allow. The choice of a smaller observation error leads to decreased RMSE, indicating the potential benefits of higher-quality observations. A localization half-width around 190 km is ideal for conducting SIC DA with our ocean–sea ice model at 1° resolution. We suggest that future SIC DA design studies consider localization tuning based on their assimilated observation type and model configurations.

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. This work was supported by

ExxonMobil through its e-filliates membership in the Andlinger Center for Energy and Environment. We thank Matthew Harrison and Feiyu Lu for helpful discussions, Niki Zadeh and William Cooke for instructions on setting up models, and Rusty Benson, Seth Underwood, and Jeffrey Durachta for help on job scripting. We thank Feiyu Lu and Hyung-Gyu Lim for helpful discussions on earlier versions of the manuscript. We also thank the DART group at NCAR for discussions on the linkage between SIS2 and DART.

REFERENCES

Adcroft, A., and Coauthors, 2019: The GFDL global ocean and sea ice model OM4.0: Model description and simulation features. *J. Adv. Model. Earth Syst.*, **11**, 3167–3211, https://doi.org/10.1029/2019MS001726.

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903, https://doi.org/10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2.

——, T. Hoar, K. Raeder, H. Liu, N. Collins, R. Torn, and A. Arellano, 2009: The Data Assimilation Research Testbed: A community facility. *Bull. Amer. Meteor. Soc.*, **90**, 1283–1296, https://doi.org/10.1175/2009BAMS2618.1.

Assur, A., 1958: Composition of sea ice and its tensile strength. *Arctic Sea Ice*, U.S. National Academy of Sciences, 106–138.

Banzon, V., T. M. Smith, T. M. Chin, C. Liu, and W. Hankins, 2016: A long-term record of blended satellite and in situ sea-surface temperature for climate monitoring, modeling and environmental studies. *Earth Syst. Sci. Data*, **8**, 165–176, https://doi.org/10.5194/essd-8-165-2016.

Batté, L., I. Välisuo, M. Chevallier, J. C. A. Navarro, P. Ortega, and D. Smith, 2020: Summer predictions of Arctic sea ice edge in multi-model seasonal re-forecasts. *Climate Dyn.*, **54**, 5013–5029, https://doi.org/10.1007/s00382-020-05273-8.

Bhatt, U. S., and Coauthors, 2020: 2019 Sea ice outlook full post-season report. https://www.arcus.org/sipn/sea-ice-outlook/2019/post-season.

Bitz, C. M., and W. H. Lipscomb, 1999: An energy-conserving thermodynamic model of sea ice. *J. Geophys. Res.*, **104**, 15 669–15 677, https://doi.org/10.1029/1999JC900100.

——, M. M. Holland, A. J. Weaver, and M. Eby, 2001: Simulating the ice-thickness distribution in a coupled climate model. *J. Geophys. Res.*, **106**, 2441–2463, https://doi.org/10.1029/1999JC000113.

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.

Bonan, D. B., M. Bushuk, and M. Winton, 2019: A spring barrier for regional predictions of summer Arctic sea ice. *Geophys. Res. Lett.*, **46**, 5937–5947, https://doi.org/10.1029/2019GL082947.

Briegleb, B. P., and B. Light 2007: A delta-Eddington multiple scattering parameterization for solar radiation in the sea ice component of the Community Climate System Model. NCAR Tech. Note NCAR/TN-472+STR, 100 pp., https://doi.org/10.5065/D6B27S71.

Bushuk, M., R. Msadek, M. Winton, G. 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, 2019a: Regional Arctic sea–ice prediction: Potential versus operational seasonal forecast. *Climate Dyn.*, **52**, 2721–2743, https://doi.org/10.1007/s00382-018-4288-y.

——, X. Yang, M. Winton, R. Msadek, M. Harrison, A. Rosati, and R. Gudgel, 2019b: The value of sustained ocean observations for sea ice predictions in the Barents Sea. *J. Climate*, **32**, 7017–7035, https://doi.org/10.1175/JCLI-D-19-0179.1.

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.

Chang, Y.-S., S. Zhang, A. Rosati, T. L. Delworth, and W. F. Stern, 2013: An assessment of oceanic variability for 1960–2010 from the GFDL ensemble coupled data assimilation. *Climate Dyn.*, **40**, 775–803, https://doi.org/10.1007/s00382-012-1412-2.

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. Salas y 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.

Danabasoglu, G., and Coauthors, 2012: The CCSM4 ocean component. *J. Climate*, **25**, 1361–1389, https://doi.org/10.1175/JCLI-D-11-00091.1.

——, and Coauthors, 2014: North Atlantic simulations in Coordinated Ocean-ice Reference Experiments phase II (CORE-II). Part I: Mean states. *Ocean Modell.*, **73**, 76–107, https://doi.org/10.1016/j.ocemod.2013.10.005.

Delworth, T. L., and Coauthors, 2006: GFDL's CM2 global coupled climate model. Part I: Formulation and simulation characteristics. *J. Climate*, **19**, 643–674, https://doi.org/10.1175/JCLI3629.1.

——, 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. J. 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.

Dorel, L., C. Ardilouze, M. Déqué, L. Batté, and J. F. Guérémy, 2017: Documentation of the Météo-France pre-operational seasonal forecasting system. Météo-France Tech. Rep. 401, 36 pp., https://www.umr-cnrm.fr/IMG/pdf/system6-technical.pdf.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10 143–10 162, https://doi.org/10.1029/94JC00572.

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 modeling system. *Cryosphere*, **13**, 491–509, https://doi.org/10.5194/tc-13-491-2019.

Griffies, S. M., and Coauthors, 2009: Coordinated Ocean-ice Reference Experiments (CORes). *Ocean Modell.*, **26** (1-2), 1–46, https://doi.org/10.1016/j.ocemod.2008.08.007.

Hibler, W. D., 1979: A dynamic thermodynamic sea ice model. *J. Phys. Oceanogr.*, **9**, 815–846, https://doi.org/10.1175/1520-0485(1979)009<0815:ADTSIM>2.0.CO;2.

Hunke, E. C., and J. K. Dukowicz, 1997: An elastic–viscous–plastic model for sea ice dynamics. *J. Phys. Oceanogr.*, **27**, 1849–1867, https://doi.org/10.1175/1520-0485(1997)027<1849:AEVPMF>2.0.CO;2.

Ivanova, I., and Coauthors, 2015: Inter-comparison and evaluation of sea ice algorithms: Towards further identification of challenges and optional approach using passive microwave observations. *Cryosphere*, **9**, 1797–1817, https://doi.org/10.5194/tc-9-1797-2015.

Johnson, S. J., and Coauthors, 2019: SEAS5: The new ECMWF seasonal forecast system. *Geosci. Model Dev.*, **12**, 1087–1117, https://doi.org/10.5194/gmd-12-1087-2019.

Kimmritz, M., F. Counillon, C. M. 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), 1–23, https://doi.org/10.1080/16000870.2018.1435945.

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.

——, and A. Schweiger, 2015: Arctic sea ice thickness loss determined using subsurface, aircraft, and satellite observations. *Cryosphere*, **9**, 269–283, https://doi.org/10.5194/tc-9-269-2015.

Lisæter, K., 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.

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

MacLachlan, C., and Coauthors, 2015: Global seasonal forecast system version 5 (GloSea5): A high-resolution seasonal forecast system. *Quart. J. Roy. Meteor. Soc.*, **141**, 1072–1084, https://doi.org/10.1002/qj.2396.

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.

Mathiot, P., C. K. Beatty, T. Fichefet, H. Goosse, F. Massonnet, and M. Vancoppenolle, 2012: Better constraints on the sea-ice state using global sea-ice data assimilation. *Geosci. Model Dev.*, **5**, 1501–1515, https://doi.org/10.5194/gmd-5-1501-2012.

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

Merryfield, W. J., and Coauthors, 2013: The Canadian seasonal to interannual prediction system. Part I: Models and initialization. *Mon. Wea. Rev.*, **141**, 2910–2945, https://doi.org/10.1175/MWR-D-12-00216.1.

Msadek, R., G. A. Vecchi, M. Winton, and R. G. 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.

Mu, L., Q. Yang, M. Losch, S. N. Losa, R. Ricker, L. Nerger, and X. Liang, 2018: Improving sea ice thickness estimates by assimilating CryoSat-2 and SMOS sea ice thickness data simultaneously. *Quart. J. Roy. Meteor. Soc.*, **144**, 529–538, https://doi.org/10.1002/QJ.3225.

——, and Coauthors, 2020: Toward a data assimilation system for seamless sea ice prediction based on the AWI climate model. *J. Adv. Model. Earth Syst.*, **12**, e2019MS001937, https://doi.org/10.1029/2019MS001937.

Nerger, L., Q. Tang, and L. Mu, 2020: Efficient ensemble data assimilation for coupled models with parallel data assimilation framework: Example of AWI-CM (AWI-CM-PDAF 1.0). *Geosci. Model Dev.*, **13**, 4305–4321, https://doi.org/10.5194/gmd-13-4305-2020.

Raeder, K., J. L. Anderson, N. Collins, T. J. Hoar, J. E. Kay, P. H. Lauritzen, and R. Pincus, 2012: DART/CAM: An ensemble data assimilation system for CESM atmospheric models. *J. Climate*, **25**, 6304–6317, https://doi.org/10.1175/JCLI-D-11-00395.1.

Reynolds, R. W., T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, and M. G. Schlax, 2007: Daily high-resolution-blended analyses for sea surface temperature. *J. Climate*, **20**, 5473–5496, https://doi.org/10.1175/2007JCLI1824.1.

Sakov, P., F. Counillon, L. Bertino, K. A. Lisaeter, P. R. Oke, and A. Korablev, 2012: TOPAZ4: An ocean–sea ice data assimilation system for the North Atlantic and Arctic. *Ocean Sci. Discuss.*, **9**, 1519–1575, https://doi.org/10.5194/osd-9-1519-2012.

Shlyaeva, A., M. Buehner, A. Caya, J.-F. Lemieux, G. C. Smith, F. Roy, F. Dupont, and T. Carrieres, 2016: Towards ensemble data assimilation for the Environment Canada Regional Ice Prediction System. *Quart. J. Roy. Meteor. Soc.*, **142**, 1090–1099, https://doi.org/10.1002/qj.2712.

Sigmond, M., J. C. Fyfe, G. M. Flato, V. V. Kharin, and W. J. 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.

Smith, G. C., and Coauthors, 2015: Sea ice forecast verification in the Canadian Global Ice Ocean Prediction System. *Quart. J. Roy. Meteor. Soc.*, **142**, 659–671, https://doi.org/10.1002/qj.2555.

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.

Tietsche, S., D. Notz, J. H. Jungclaus, and J. Marotzke, 2013: Assimilation of sea-ice concentration in a global climate model—Physical and statistical aspects. *Ocean Sci.*, **9**, 19–36, https://doi.org/10.5194/os-9-19-2013.

Toyoda, T., Y. Fujii, T. Yasuda, N. Usui, K. Ogawa, T. Kuragano, H. Tsujino, and M. Kamachi, 2015: *J. Oceanogr.*, **72**, 235–262, https://doi.org/10.1007/s10872-015-0326-0.

Tsujino, H. and Coauthors, 2018: JRA-55 based surface dataset for driving ocean–sea-ice models (JRA55-do). *Ocean Modell.*, **30**, 79–139, https://doi.org/10.1016/j.ocemod.2018.07.002.

Vecchi, G. A., and Coauthors, 2014: On the seasonal forecasting of regional tropical cyclone activity. *J. Climate*, **27**, 7994–8016, https://doi.org/10.1175/JCLI-D-14-00158.1.

Wang, K., J. Debernard, A. K. Sperrevik, P. E. Isachsen, and T. Lavergne, 2013: A combined optimal interpolation and nudging scheme to assimilate OSISAF sea-ice concentration into ROMS. *Ann. Glaciol.*, **54**, 8–12, https://doi.org/10.3189/2013AoG62A138.

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

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

Xue, Y., C. Wen, X. Yang, D. Behringer, A. Kumar, G. Vecchi, A. Rosati, and R. Gudgel, 2017: Evaluation of tropical Pacific observing systems using NCEP and GFDL ocean data assimilation systems. *Climate Dyn.*, **49**, 843–868, https://doi.org/10.1007/s00382-015-2743-6.

Yang, Q., S. N. Losa, M. Losch, X. Tian-Kunze, L. Nerger, J. Liu, L. Kaleschke, and Z. Zhang, 2014: Assimilating SMOS sea ice thickness into a coupled ice-ocean model using a local SEIK filter. *J. Geophys. Res. Oceans*, **119**, 6680–6692, https://doi.org/10.1002/2014JC009963.

——, ——, ——, T. Jung, and L. Nerger, 2015: The role of atmospheric uncertainty in Arctic summer sea ice data assimilation and prediction. *Quart. J. Roy. Meteor. Soc.*, **141**, 2314–2323, https://doi.org/10.1002/qj.2523.

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, S., M. J. Harrison, A. Rosati, and A. Wittenberg, 2007: System design and evaluation of coupled ensemble data assimilation for global oceanic studies. *Mon. Wea. Rev.*, **135**, 3541–3564, https://doi.org/10.1175/MWR3466.1.

Zhang, Y.-F., T. J. Hoar, Z.-L. Yang, J. L. Anderson, A. M. Toure, and M. Rodell, 2014: Assimilation of MODIS snow cover through the data assimilation research testbed and the Community Land Model version 4. *J. Geophys. Res. Atmos.*, **119**, 7091–7103, https://doi.org/10.1002/2013JD021329.

——, C. M. Bitz, J. L. Anderson, N. Collins, J. Hendricks, T. Hoar, and K. Raeder, 2018: Insights on sea ice data assimilation from perfect model observing system simulation experiments. *J. Climate*, **31**, 5911–5926, https://doi.org/10.1175/JCLI-D-17-0904.1.

Zygmuntowska, M., P. Rampal, N. Ivanova, and L. H. Smedsrud, 2014: Uncertainties in Arctic sea ice thickness and volume: New estimates and implications for trends. *Cryosphere*, **8**, 705–720, https://doi.org/10.5194/tc-8-705-2014.