# Mechanisms of Regional Arctic Sea Ice Predictability in Two Dynamical Seasonal Forecast Systems

Mitchell Bushuk,[a,b] Yongfei Zhang,[c] Michael Winton,[a] Bill Hurlin,[a] Thomas Delworth,[a] Feiyu Lu,[c] Liwei Jia,[a,b] Liping Zhang,[a,b] William Cooke,[a] Matthew Harrison,[a] Nathaniel C. Johnson,[a] Sarah Kapnick,[a] Colleen McHugh,[a,d] Hiroyuki Murakami,[a,b] Anthony Rosati,[a,b] Kai-Chih Tseng,[c] Andrew T. Wittenberg,[a] Xiaosong Yang,[a] and Fanrong Zeng[a]

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

ABSTRACT: Research over the past decade has demonstrated that dynamical forecast systems can skillfully predict pan-Arctic sea ice extent (SIE) on the seasonal time scale; however, there have been fewer assessments of prediction skill on user-relevant spatial scales. In this work, we evaluate regional Arctic SIE predictions made with the Forecast-Oriented Low Ocean Resolution (FLOR) and Seamless System for Prediction and Earth System Research (SPEAR_MED) dynamical seasonal forecast systems developed at the NOAA/Geophysical Fluid Dynamics Laboratory. Compared to FLOR, we find that the recently developed SPEAR_MED system displays improved skill in predicting regional detrended SIE anomalies, partially owing to improvements in sea ice concentration (SIC) and thickness (SIT) initial conditions. In both systems, winter SIE is skillfully predicted up to 11 months in advance, whereas summer minimum SIE predictions are limited by the Arctic spring predictability barrier, with typical skill horizons of roughly 4 months. We construct a parsimonious set of simple statistical prediction models to investigate the mechanisms of sea ice predictability in these systems. Three distinct predictability regimes are identified: a summer regime dominated by SIE and SIT anomaly persistence; a winter regime dominated by SIE and upper-ocean heat content (uOHC) anomaly persistence; and a combined regime in the Chukchi Sea, characterized by a trade-off between uOHC-based and SIT-based predictability that occurs as the sea ice edge position evolves seasonally. The combination of regional SIE, SIT, and uOHC predictors is able to reproduce the SIE skill of the dynamical models in nearly all regions, suggesting that these statistical predictors provide a stringent skill benchmark for assessing seasonal sea ice prediction systems.

KEYWORDS: Arctic; Sea ice; Climate variability; Seasonal forecasting; Climate models; Data assimilation

## 1. Introduction

The decline of Arctic sea ice extent (SIE) and the associated increase in open-water season length (Stroeve and Notz 2018) has motivated a need for seasonal Arctic sea ice predictions (Jung et al. 2016). Skillfully predicting Arctic sea ice could benefit communities and stakeholders in the region and also provide a source of prediction skill for other components of the Earth system (Jung et al. 2014; Meehl et al. 2021). Sea ice forecasting capability has developed over the past decade due to interacting advancements in polar observations, coupled climate modeling, data assimilation techniques, and sea ice predictability research (Guemas et al. 2016b). Despite these notable advances, for many forecasting applications, sea ice predictions have yet to achieve the prediction skill levels or user-oriented metrics required for forecasts to be useful to stakeholders (Caron et al. 2020).

The basic picture of seasonal-to-interannual Arctic sea ice predictability involves a competition between slowly evolving and predictable variations in the ocean and sea ice components and comparatively unpredictable fluctuations that occur in the atmosphere (e.g., Tietsche et al. 2016). Both the atmosphere and ocean interact strongly with sea ice, each exerting thermodynamic and dynamic influences, but the relative importance of these factors varies with season, lead time, and region (Tietsche et al. 2014).

The direction of ice-edge motion dictates the dominant sources of SIE predictability for each season. In fall and winter months, as the ocean loses heat to the atmosphere and cools, new sea ice forms and the ice edge advances southward. The rate of ice formation is strongly modulated by the ocean conditions south of the sea ice edge and, therefore, upper-ocean heat content (uOHC) (Blanchard-Wrigglesworth et al. 2011a; Schlichtholz 2011; Day et al. 2014b; Bushuk and Giannakis 2015; Sigmond et al. 2016; Bushuk et al. 2019b; Kimmritz et al. 2019) and ocean heat transport (OHT) convergence (Årthun et al. 2012; Onarheim et al. 2015; Serreze et al. 2016; Lenetsky

---

---

et al. 2021) both provide crucial sources of predictability for fall and winter SIE. Conversely, in spring and summer months, sea ice begins to melt and the ice edge retreats northward. The energy required to melt sea ice is primarily determined by the sea ice thickness (SIT), and thus the amount of northward retreat is strongly modulated by the SIT conditions located north of the sea ice edge. Many previous studies have shown that SIT provides a key source of predictability for summer SIE conditions (Holland et al. 2011; Blanchard-Wrigglesworth et al. 2011b; Chevallier and Salas y Mélia 2012; Krumpen et al. 2013; Day et al. 2014a; Collow et al. 2015; Massonnet et al. 2015; Guemas et al. 2016a; Williams et al. 2016; Blanchard-Wrigglesworth et al. 2017; Bushuk et al. 2017b; Dirkson et al. 2017; Blockley and Peterson 2018; Holland et al. 2019; Bonan et al. 2019; Brunette et al. 2019; Babb et al. 2019; Bushuk et al. 2020; Babb et al. 2020).

Dynamical prediction systems based on global climate models (GCMs) seek to leverage these sources of predictability by initializing the sea ice, ocean, atmosphere, and land components using observational constraints and by accurately simulating the time-evolving state of the coupled system. A number of recent studies have established that these dynamical prediction systems can skillfully predict detrended pan-Arctic SIE anomalies at lead times ranging from 1 to 11 months for winter SIE and from 1 to 6 months for summer SIE (e.g., Wang et al. 2013; Chevallier et al. 2013; Sigmond et al. 2013; Merryfield et al. 2013; Msadek et al. 2014; Peterson et al. 2015; Blanchard-Wrigglesworth et al. 2015; Collow et al. 2015; Guemas et al. 2016a; Dirkson et al. 2017; Harnos et al. 2019; Batté et al. 2020). Pan-Arctic SIE predictions have limited utility for forecast users, who typically require local- or regional-scale information, and, therefore, recent efforts have focused on assessing regional sea ice prediction skill (Sigmond et al. 2016; Krikken et al. 2016; Bushuk et al. 2017a; Dirkson et al. 2019; Batté et al. 2020). These studies have shown that detrended regional winter sea ice can be skillfully predicted up to 11 months in advance, whereas regional summer sea ice predictions are skillful up to 4 months in advance because they are limited by the Arctic sea ice spring predictability barrier (Bonan et al. 2019; Bushuk et al. 2020). In parallel to this work on dynamical sea ice predictions, a number of skillful prediction methods based on statistical methodologies have been developed (Drobot et al. 2006; Drobot 2007; Lindsay et al. 2008; Schröder et al. 2014; Kapsch et al. 2014; Liu et al. 2015; Yuan et al. 2016; Williams et al. 2016; Petty et al. 2017; Kondrashov et al. 2018; Brunette et al. 2019; Walsh et al. 2019; Gregory et al. 2020; Andersson et al. 2021). These studies have primarily focused on summer SIE and generally show similar skill horizons to those found in dynamical models, but often have higher skill values (e.g., Yuan et al. 2016; Gregory et al. 2020; Andersson et al. 2021).

"Perfect model" ensemble experiments performed with GCMs have been used to estimate the upper limits of Arctic sea ice predictability and have shown that there is a significant gap between the current skill of initialized dynamical prediction systems and the skill achievable by perfect model forecasts (Koenigk and Mikolajewicz 2009; Holland et al. 2011;

Blanchard-Wrigglesworth et al. 2011b; Tietsche et al. 2014; Germe et al. 2014; Day et al. 2014b, 2016; Bushuk et al. 2019a; Holland et al. 2019). These perfect model experiments suggest that there is substantial room for improvement in Arctic sea ice predictions; however, perfect model predictability estimates are subject to model errors and thus may be biased relative to nature's true predictability limits (Kumar et al. 2014). Indeed, Arctic SIE anomalies in GCMs are generally more persistent than observations and this overpersistence likely results in an overestimation of predictability in perfect model experiments (Blanchard-Wrigglesworth and Bushuk 2019; Giesse et al. 2021). This suggests that the true upper limits of sea ice predictability likely lie somewhere between previously documented perfect model skill estimates and the current skill of operational systems.

In this study, we take a different tack to exploring predictability, analyzing the sources of prediction skill within dynamical seasonal prediction systems. We first assess the regional detrended SIE prediction skill in two seasonal prediction systems developed at the Geophysical Fluid Dynamics Laboratory. After establishing the regional SIE prediction skill of each system, we next seek a parsimonious description of the sources of prediction skill. In particular, we construct a series of simple linear regression models that use the forecast initial conditions (ICs) as predictors for regional Arctic SIE. The skill of these simple statistical models can then be compared to the dynamical models' skill in order to assess the key sources of prediction skill in each system. Compared to the perfect model framework, this approach focuses on predictions of observed regional SIE, and thus does not suffer from the problem of potentially overestimating predictability. However, this approach may miss sources of skill that are present in nature and may also be limited by the chosen statistical prediction models. Therefore, this approach of quantifying predictability provides a *lower bound* on predictability that complements the *upper bound* provided by perfect model experiments. These lower bounds, if sufficiently skillful, can help to better constrain the true limits of regional Arctic sea ice predictability. The simple statistical models also provide a direct quantification of the prediction skill attributable to particular sources of predictability. It is important to note that this quantification of skill is based on an earlier time period of 1992–2020, and therefore it is not necessarily indicative of future skill due to the nonstationarity of Arctic sea ice predictability (e.g., Holland and Stroeve 2011; Cheng et al. 2016; Holland et al. 2019). Additionally, statistical models can overestimate future skill due to overfitting, a factor we attempt to mitigate by considering single-predictor linear regression models.

The outline of this paper is as follows. In section 2, we introduce the dynamical seasonal prediction systems, describe suites of retrospective seasonal prediction experiments, and outline our methods for prediction skill assessment and model evaluation. In section 3, we evaluate the suitability of the dynamical models and their corresponding ICs for Arctic sea ice predictions by considering their simulated mean state, trends, and interannual variability. In section 4, we provide an assessment of regional Arctic SIE prediction skill in each

TABLE 1. Summary of GFDL seasonal prediction systems and retrospective forecasts considered in this study.

| System property | FLOR | SPEAR_MED |
|---|---|---|
| Ocean model | MOM5; 1.0°, 50 vertical levels | MOM6; 1.0°, 75 vertical levels |
| Sea ice model | SIS1; 1.0°, 5-category ITD | SIS2; 1.0°, 5-category ITD |
| Atmosphere model | AM2.5; 0.5°, 32 vertical levels | AM4; 0.5°, 33 vertical levels |
| Land model | LM3; 0.5° | LM4; 0.5° |
| | | |
| Ocean data | Satellite SST, Argo, XBT, MRB, CTD, MBT; daily | Satellite SST, Argo, XBT, MRB; daily |
| Atmosphere data | 3D temperature from NCEP-2; 6-hourly | 3D temperature, winds, humidity from CFSR; 6-hourly |
| Sea ice data | None | Satellite SIC used to adjust under-ice SST; daily |
| Land data | None | None |
| Ocean ICs | ECDA | SPEAR ODA |
| Sea ice ICs | ECDA | SPEAR_MED nudged run |
| Atmosphere ICs | AMIP run | SPEAR_MED nudged run |
| Land ICs | AMIP run | SPEAR_MED nudged run |
| | | |
| Reforecast period | 1992–2020 | 1992–2020 |
| Ensemble size | 12 | 15 |
| Initialization dates | First of each month | First of each month |
| Prediction length | 1 year | 1 year |

system and discuss the skill differences between the models. In section 5 and 6, we consider the sources of predictability for summer and winter regional Arctic SIE, respectively. In section 7, we analyze predictions in the Chukchi Seas, highlighting a combined predictability regime present in this region. We conclude and discuss the results in section 8.

## 2. Methods

### a. FLOR seasonal prediction system

In this study, we assess seasonal sea ice predictions made with two dynamical prediction systems developed at the Geophysical Fluid Dynamics Laboratory. We first describe the Forecast-Oriented Low Ocean Resolution (FLOR) system in this subsection and describe the recently developed Seamless System for Prediction and Earth System Research (SPEAR) in the following subsection (see Table 1 for an overview of the key features of each prediction system).

Previous studies have shown that the FLOR prediction system skillfully predicts (see definition of skill metrics in section 2e ahead) detrended regional SIE in both the Arctic (Bushuk et al. 2017a) and Antarctic (Bushuk et al. 2021). FLOR is a fully coupled global atmosphere–ocean–land–sea ice model, with nominal horizontal resolutions of 0.5° in the atmosphere–land components and a coarser resolution of 1.0° in the ocean and sea ice components (Vecchi et al. 2014). FLOR's sea ice component is based upon the sea ice simulator, version 1 (SIS1; Delworth et al. 2006), which employs a modified Semtner thermodynamic scheme with two ice layers and one snow layer (Winton 2000), a subgrid-scale ice-thickness distribution (ITD) with five thickness categories (Thorndike et al. 1975; Bitz et al. 2001), a formulation of ice dynamics based on the elastic–viscous–plastic (EVP) rheology (Hunke and Dukowicz 1997), and a surface-temperature-dependent albedo parameterization (see section 3.6.2 of Hunke et al. 2015). FLOR's ocean, atmosphere, and

land components are based on an updated version of Ocean Model, version 2.1 (OM2.1; Gnanadesikan et al. 2006; Delworth et al. 2012); Atmosphere Model, version 2.5 (AM2.5; Anderson et al. 2004; Delworth et al. 2012); and Land Model, version 3 (LM3; Milly et al. 2014), respectively.

The FLOR seasonal predictions are initialized using ocean and sea ice ICs from the GFDL Ensemble Coupled Data Assimilation system (ECDA; Zhang et al. 2007), which is based upon the CM2.1 model (Delworth et al. 2006). ECDA is a weakly coupled assimilation system spanning 1961–2020 that provides "full field" ICs using an ensemble adjustment Kalman filter (EAKF; Anderson 2001) approach. ECDA assimilates observations of ocean temperature and salinity (T/S) and atmospheric temperature data from the NCEP–DOE Atmospheric Model Intercomparison Project (AMIP-II) reanalysis (Kanamitsu et al. 2002) (see Table 1). The ocean observations comprise sea surface temperature (SST) data from the Hadley Centre's Sea Ice and SST dataset (HadISST1; Rayner et al. 2003) (prior to 2011) and NOAA's daily Optimum Interpolation SST dataset (OISST; Reynolds et al. 2007) (post 2011), and T/S profiles from the World Ocean Database (WOD; Levitus et al. 2013), the Global Temperature and Salinity Profile Programme (GTSPP; Sun et al. 2010), and the Argo program (Roemmich et al. 2004). The T/S profiles include profiling floats (PFL), expendable bathythermograph (XBT), mooring (MRB), ocean station (OSD), mechanical bathythermograph (MBT), and conductivity–temperature–depth (CTD) data [see Levitus et al. (2013) for descriptions of the various data types]. ECDA does not directly assimilate sea ice data, but the sea ice state variables are constrained via the assimilation of oceanic and atmospheric data and the associated heat fluxes and interfacial stresses that are passed to the sea ice model (Bushuk et al. 2019b). FLOR's atmosphere and land ICs come from a suite of AMIP-style atmosphere–land simulations forced by observed interannually varying SST and sea ice conditions from OISST.

### b. SPEAR seasonal prediction system

SPEAR is a recently developed seasonal prediction system, based on the coupled model components of the CM4 model (Held et al. 2019), but employing two factors to optimize for seasonal-to-decadal prediction research: 1) a lower resolution in the ice–ocean components to achieve the computational efficiency required for ensemble prediction, and 2) a higher-resolution option (SPEAR_MED) in the atmosphere and land components geared toward regional climate impacts. The SPEAR coupled model has two versions, SPEAR_LO and SPEAR_MED, which share the same 1° nominal ocean–sea ice resolution but have differing atmosphere–land resolutions of 1° and 0.5°, respectively (Delworth et al. 2020). SPEAR_MED is GFDL's quasi-operational seasonal prediction system, which submits near-real-time experimental predictions each month to the North American Multi-Model Ensemble (NMME; Kirtman et al. 2014). In this manuscript, we focus our analysis on the SPEAR_MED NMME system since we find that SPEAR_LO has very similar prediction skill for Arctic regional SIE. Note that while the systems have similar skill for Arctic sea ice, SPEAR_MED has better performance for other climate variables, such as tropical cyclone intensity and Antarctic sea ice (Bushuk et al. 2021).

SPEAR's sea ice model is based upon the sea ice simulator version 2 (SIS2; Adcroft et al. 2019). Similar to SIS1, the SIS2 model uses a five-category ITD and computes internal ice stresses using an EVP rheology. The key differences in the SIS2 model include improved shortwave radiation physics based on the Delta-Eddington scheme of Briegleb and Light (2007), an updated thermodynamic solver similar to Bitz and Lipscomb (1999), improved vertical ice temperature resolution with four ice layers and one snow layer, updated numerical algorithms to reduce the occurrence of coupled ice–ocean numerical instabilities, and ice dynamics that are solved using a C-grid stencil (Bouillon et al. 2009). SPEAR's ocean, atmosphere, and land models are based on the Modular Ocean Model, version 6 (MOM6; Adcroft et al. 2019); Atmosphere Model, version 4; and Land Model, version 4 (Zhao et al. 2018a,b), respectively.

SPEAR_MED is initialized using two separate assimilation experiments spanning 1990–2020 (see Table 1). The sea ice, atmosphere, and land ICs come from a nudged atmosphere and SST ensemble run of the coupled SPEAR_MED model. In this run, the atmospheric fields are nudged toward 3D temperature, wind, and humidity data from the NOAA/NCEP Climate Forecast System Reanalysis (CFSR; Saha et al. 2010) and the SST field is nudged toward daily OISST data (Reynolds et al. 2007). This run also incorporates a constraint from OISST SIC data, which are used to modify the SST values under sea ice. Specifically, if a gridpoint is ice covered in the observations (defined as SIC $\geq$ 30%), then the SST value is replaced by the freezing point of seawater given by $T_f = -0.054 \times$ SSS, where $T_f$ is the freezing point and SSS is the model-predicted sea surface salinity. This SST modification was found to significantly improve the fidelity of simulated sea ice in the nudged experiment compared to observations. The nudging is performed using a 6-h $e$-folding time scale for atmospheric temperatures and winds, a 1-day time scale for humidity, and a 4 m day$^{-1}$ piston velocity for SST (corresponding to a 12.5-day $e$-folding time scale for a 50-m mixed layer). The nudged experiment is run as a 5-member ensemble, which is initialized in 1990 using ensemble members 1–5 of the SPEAR_MED Large Ensemble, described in the following subsection. These five members are repeated three times to create a 15-member ensemble of ICs.

The SPEAR_MED ocean ICs come from an ocean data assimilation (ODA) system based on the SPEAR_LO coupled model and an EAKF approach (Lu et al. 2020). The ODA system assimilates daily OISST observations and ocean $T/S$ profiles from Argo (PFL), XBT, and MRB. The ODA system does not include CTD, OSD, drifting buoy (DRB), glider (GLD), and instrumented marine mammal (APB) data types, each of which have some coverage in the Arctic region. The SST values under sea ice are modified using the same freezing-point correction as used in the nudged run. The use of a common SST dataset between the ODA and nudged experiments allows for the ICs from these individual DA experiments to be combined to initialize the coupled model.

The SPEAR_MED system also uses an ocean-tendency adjustment (OTA) procedure to reduce the ocean biases of both the ODA run and the free-running model. This procedure applies the climatological DA increments from a previous ODA run spanning 2003–18 as 3D ocean $T/S$ tendency terms. This OTA technique improves both DA performance as well as coupled model prediction skill for the El Niño–Southern Oscillation (Lu et al. 2020).

### c. Retrospective seasonal predictions and large ensemble experiments

We perform suites of retrospective seasonal prediction experiments with each prediction system spanning the years 1992–2020 (see Table 1). These ensemble seasonal predictions are initialized on the first of each month and integrated for 1 year. The FLOR and SPEAR_MED systems are run with 12 and 15 ensemble members, respectively.

We also use large ensembles (LEs) of historical simulations performed with each model to assess the mean state biases of FLOR and SPEAR_MED (Bushuk et al. 2020; Delworth et al. 2020). For both models, the LEs use 30 members and span the period 1921–2100. The FLOR LE uses historical radiative forcings from 1921 to 2005 and representative concentration pathway 8.5 (RCP8.5; Meinshausen et al. 2011) from 2006 to 2100, and the SPEAR_MED LE uses historical radiative forcings from 1921 to 2014 and shared socioeconomic pathways 5–8.5 (SSP5–8.5; Riahi et al. 2017) from 2015 to 2100.

### d. Observational datasets

We assess the model simulations and predictions using monthly averaged passive microwave satellite SIC observations from the National Snow and Ice Data Center (NSIDC) processed using the NASA Team retrieval algorithm (dataset ID: NSIDC-0051; Cavalieri et al. 1996). The observed

monthly SIC data is regridded to the model grid to assess model biases and to compute prediction skill metrics. We evaluate SIT using the Alfred Wegener Institute monthly *CryoSat-2* SIT product, version 2.3, which derives SIT estimates from satellite radar altimeter measurements of the snow–ice interface and provides winter data for October–April of years 2010/11 through 2020/21 (Ricker et al. 2014). We also use SIT reanalysis data from the Pan-Arctic Ice Ocean Modeling and Assimilation System (PIOMAS; Zhang and Rothrock 2003). Sea ice drift is assessed using the low-resolution sea ice drift product of the EUMETSAT Ocean and Sea Ice Satellite Application Facility (OSISAF; Lavergne et al. 2010), which provides daily velocity data spanning 2006–21. We average the daily data into monthly means for our analysis.

### e. Prediction skill metrics, significance testing, and reference forecasts

In this manuscript, we focus on monthly mean and regional-scale quantities, with an associated companion manuscript focusing on daily and gridpoint-scale evaluations (Zhang et al. 2022). We consider prediction skill for regional SIE, defined as the regional areal sum of all grid points with SIC $\geq 15\%$. Throughout the manuscript, we use the term "target month" to refer to the month that is being predicted, and "lead time" to refer to the number of months prior to the target month that the forecast was initialized. For example, a 1 January initialized forecast of the January-mean SIE is referred to as a lead-0 prediction.

We let $o$ be an observed time series of regional SIE in a given target month, and let $p_{ij}(\tau)$ be the model-predicted value for year $i$, ensemble member $j$, and lead time $\tau$. We assess prediction skill using the ensemble-mean prediction $p_i(\tau)$ given by

$$p_i(\tau) = \frac{1}{K} \sum_{j=1}^{K} p_{ij}(\tau), \qquad (1)$$

where $K$ is the number of ensemble members. We consider prediction skill based on both the anomaly correlation coefficient (ACC) and the detrended ACC. We primarily focus our analysis on detrended ACC, since forced Arctic SIE trends provide the dominant source of prediction skill for nondetrended anomalies (Sigmond et al. 2013). The ACC is defined as

$$\text{ACC}(\tau) = \frac{\sum_{i=1}^{N} \left[p_i(\tau) - \overline{p(\tau)}\right](o_i - \overline{o})}{\sqrt{\sum_{i=1}^{N} \left[p_i(\tau) - \overline{p(\tau)}\right]^2} \sqrt{\sum_{i=1}^{N} (o_i - \overline{o})^2}}, \qquad (2)$$

where $N$ is the number of years in the observed and predicted time series and the overbar denotes a temporal mean.

The detrended ACC uses anomalies relative to a linear trend prediction, thereby removing skill associated with secular trends and focusing on interannual anomalies. The detrended ACC is defined as

$$\text{ACC}_{\text{detrend}}(\tau) = \frac{\sum_{i=1}^{N} \left[p_i(\tau) - p_i^L(\tau)\right]\left(o_i - o_i^L\right)}{\sqrt{\sum_{i=1}^{N} \left[p_i(\tau) - p_i^L(\tau)\right]^2} \sqrt{\sum_{i=1}^{N} \left(o_i - o_i^L\right)^2}}, \qquad (3)$$

where $o_i^L$ and $p_i^L(\tau)$ are linear trend predictions for year $i$ computed using all available past observed and predicted data, respectively. This approach is chosen to avoid using future data to compute the detrended anomaly in a given year. We assume a linear trend of zero for the first 3 years.

We assess the statistical significance of the computed ACC values using a bootstrapped resampling approach. This procedure involves producing an empirical distribution of ACC statistics based on resampling the prediction ensemble with replacement (Efron 1982). We compute a 95% confidence interval for each target month, region, and lead time based on a bootstrapped distribution of 1000 realizations. The ACC value is considered statistically significant at the 95% confidence level if the lower limit of this confidence interval is greater than zero. We also use the bootstrapped confidence intervals to assess the statistical significance of skill differences between the models. To do this, we transform the ACC values and confidence intervals to $z$ space using the Fisher $z$ transformation, compute standard errors for FLOR and SPEAR_MED (using the appropriate side of the confidence interval, depending on which model has the higher ACC value), and add the standard errors in quadrature to obtain a standard error for the correlation difference. If the correlation difference exceeds 1.96 standard errors, then the difference is considered statistically significant at the 95% confidence level.

We compare the models' prediction skill to the skill of a reference forecast based on anomaly persistence. The anomaly persistence forecast uses the observed anomaly at the forecast initialization time. These anomalies can be computed either relative to the linear trend (detrended anomalies) or relative to the climatology (nondetrended anomalies).

## 3. Simulated Arctic sea ice mean state, trends, and interannual variability

We first evaluate the fidelity of simulated Arctic sea ice in the FLOR and SPEAR_MED models. We assess the simulated mean state, trends, and interannual variability in both the free-running LEs of historical simulations and the data assimilation runs used to produce sea ice ICs.

### a. Arctic sea ice mean state

Figure 1 shows simulated climatologies of pan-Arctic SIE and sea ice volume (SIV) and the spatial patterns of SIC biases in the FLOR and SPEAR_MED models. We find that the SPEAR_MED ICs accurately capture the observed SIE seasonal cycle and also generally reflect the climatological SIC spatial pattern. The SPEAR_MED LE simulates pan-Arctic SIE that is biased high throughout the year, due to overly extensive winter sea ice in the Greenland–Iceland–Norwegian (GIN), Barents, and Bering Seas and positive summer SIC biases in the Beaufort, Chukchi, and East Siberian
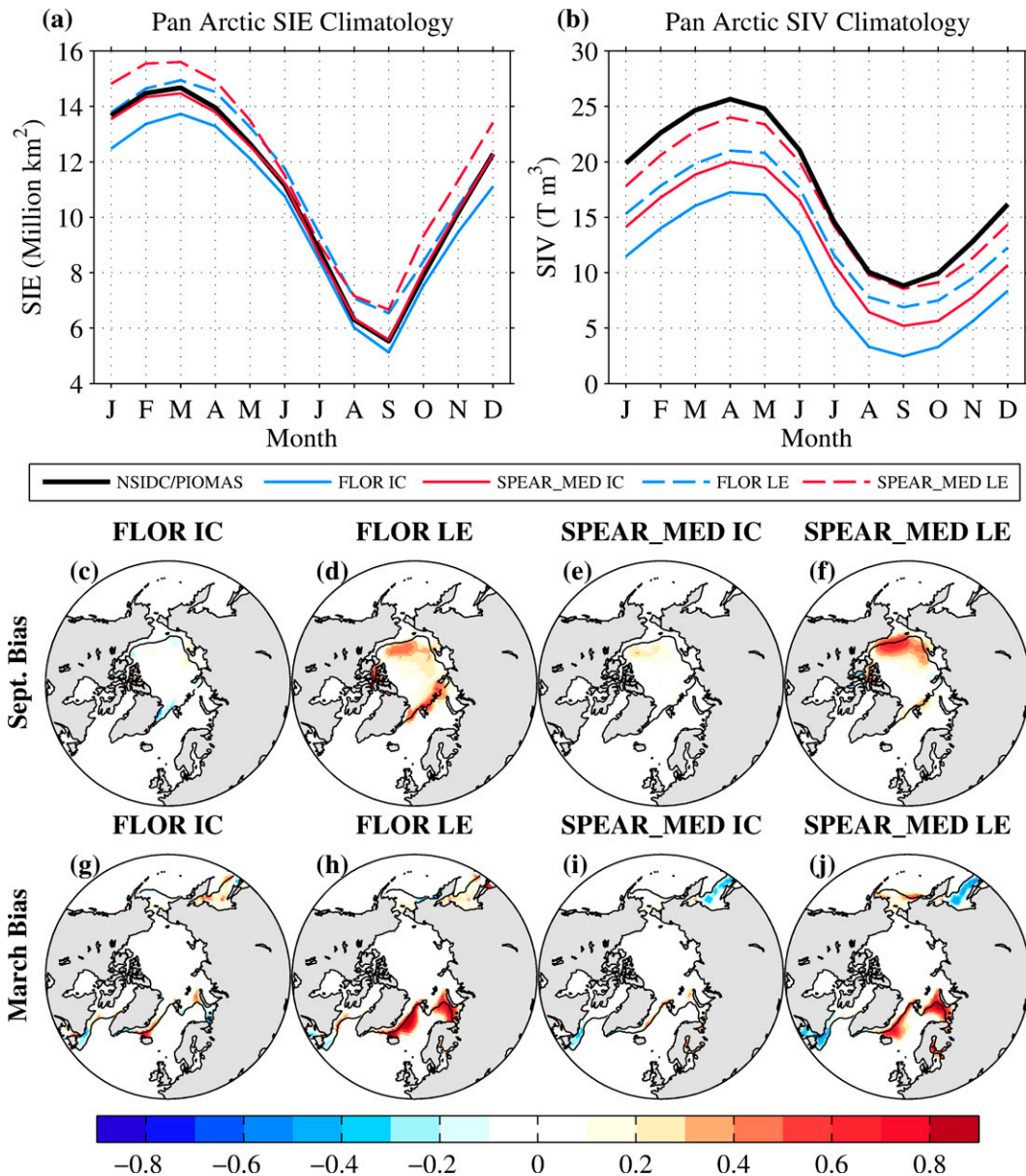
FIG. 1. Pan-Arctic (a) SIE and (b) SIV climatologies over 1990–2020 in assimilation runs used for ICs (solid lines) and LE historical simulations (dashed lines) performed with FLOR (blue) and SPEAR_MED (red). Black lines show the SIE climatology from NSIDC observations and the SIV climatology from PIOMAS reanalysis, respectively. Sea ice concentration biases (model minus observations; 1990–2020) in (c)–(f) September and (g)–(j) March in the IC and LE runs. The black contours show the observed climatological sea ice edge position.

Seas. SPEAR_MED LE also has negative winter SIC biases in the Sea of Okhotsk and the Labrador Sea. These SIC biases are reduced substantially in the SPEAR_MED IC run, but some modest biases remain with a spatial pattern that mirrors the LE biases. Similar to SPEAR_MED, FLOR LE's positive winter bias is dominated by the GIN and Barents Seas, whereas its summer SIC bias is more zonally symmetric. These SIC biases are substantially reduced in the FLOR IC run. Note that the FLOR ICs have pan-Arctic winter SIE values that are lower than observed but have roughly offsetting

positive and negative winter SIC biases. This discrepancy is primarily due to differences in the land–sea mask, which are particularly relevant in winter months when the ice edge interacts with many coastlines.

The pan-Arctic SIV climatology simulated by the SPEAR_MED LE agrees reasonably well with PIOMAS reanalysis and improves upon the FLOR LE, which is biased thin relative to PIOMAS throughout the year (Fig. 1b). PIOMAS is a sea ice reanalysis product, which does not assimilate SIT, but compares reasonably well with available satellite, aircraft,
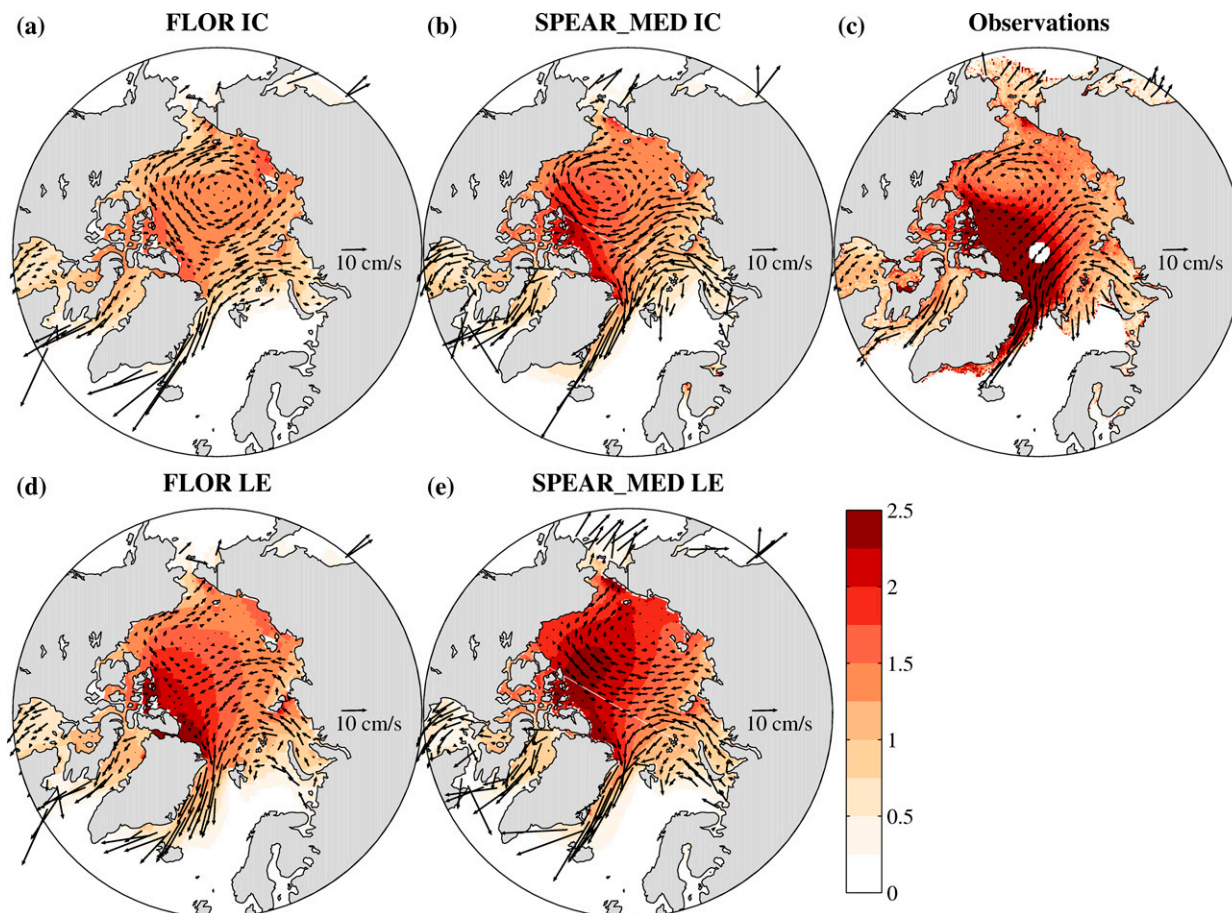
FIG. 2. Winter sea ice thickness and drift climatologies (January–March; 2010–20) in the data assimilation runs used for (a),(b) ICs and (d),(e) LE historical simulations. (c) The observed thickness is from AWI's *CryoSat-2* product, and the observed drift is from OSISAF.

and in situ SIT data (Schweiger et al. 2011; Wang et al. 2016). Both the FLOR and SPEAR_MED ICs are notably thinner than their LE counterparts, implying that winds and temperatures constrained by reanalysis data tend to produce thinner ice than the atmospheric conditions simulated by the free-running models. In Fig. 2, we plot winter climatologies of SIT and ice drift. We find that while the SPEAR_MED LE agrees with PIOMAS in terms of pan-Arctic SIV, it has large errors in its spatial SIT pattern, characterized by too much thick ice in the Beaufort and Chukchi Seas and a lack of thick ice directly north of Greenland. These SIT pattern errors resemble those in the GFDL-CM4 model (Held et al. 2019) and are associated with SPEAR_MED LE's biases in sea ice drift, which promote advection of ice into the Beaufort Sea. The SPEAR_MED Beaufort and Chukchi thickness bias is spatially coincident with the model's positive summer SIC bias, suggesting that the model is unable to completely melt this overly thick ice (cf. Figs. 1f–2e). The FLOR LE has an SIT spatial pattern that more closely resembles observations; however, its mean SIT is biased thin, resulting in the negative SIV bias shown in Fig. 1b. We find that both models have too much Fram Strait SIA and SIV

export (not shown), which may be related to their positive winter SIC biases in the GIN Seas. The IC runs have atmospheric reanalysis constraints, which lead to substantial improvements in simulated ice drift, including a well-defined Beaufort Gyre and a transpolar drift stream. It is notable that even with these drift improvements, the SPEAR_MED ICs retain errors in their SIT spatial pattern that resemble a muted version of the SPEAR_MED LE bias. Both IC runs are biased thin, resulting in the negative SIV biases shown in Fig. 1b.

*b. Trends and interannual variability*

In Fig. 3, we plot time series of September and March pan-Arctic SIE in the assimilation experiments used to produce ICs and the historical LE experiments. Both the FLOR and SPEAR_MED ICs capture the observed SIE trends and interannual variability with some skill (see Figs. 3a,b); however, the SPEAR_MED March trend is less negative than observed. In Fig. 4, we plot regional correlations between detrended SIE ICs and NSIDC observations for each month of the year. We find that in essentially all regions and all months of the year, the SPEAR_MED regional SIE ICs have improved interannual variability compared with the FLOR ICs. The reason for
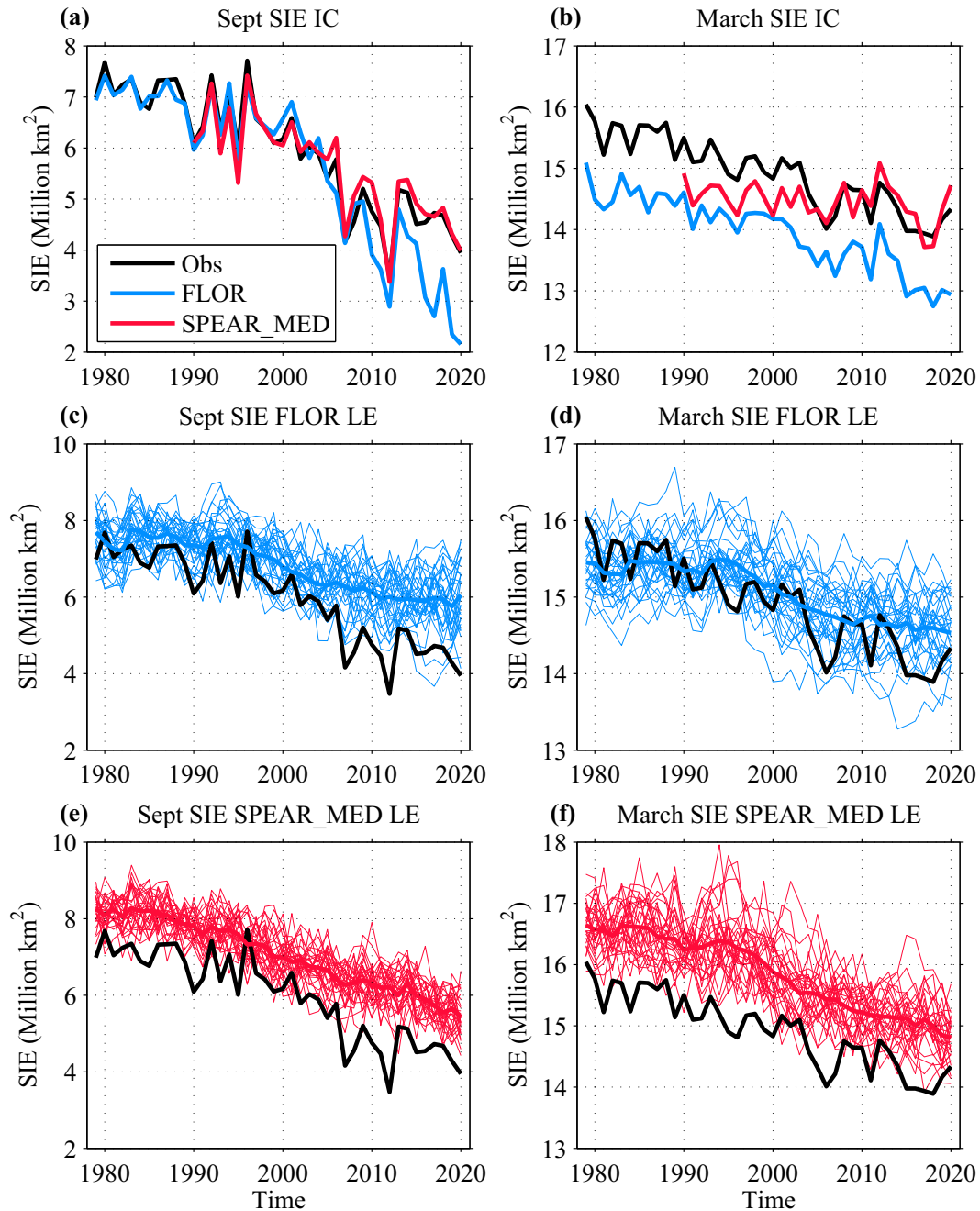
FIG. 3. Pan-Arctic SIE time series in FLOR (blue), SPEAR_MED (red), and NSIDC observations (black). (a),(b) Simulated SIE in the assimilation runs used for ICs. Simulated SIE in large ensembles of historical simulations performed with (c),(d) FLOR and (e),(f) SPEAR_MED.

this improvement is that SPEAR_MED uses observed SIC data to modify its under-ice SSTs, which results in SSTs that provide a strong constraint on SIE. These correlations can be further improved by direct assimilation of SIC data (Zhang et al. 2021), but the SPEAR_MED system does not include sea ice data assimilation. The FLOR system also incorporates an SST constraint, but this constraint is weaker due to the spatial subsampling of Arctic SST in the ECDA system and the

lack of SIC data used in this system. We also find that SPEAR_MED has improved regional SIV ICs compared to FLOR, showing higher detrended correlation with PIOMAS in essentially all Arctic regions (see supplementary Fig. S1).

Both the FLOR and SPEAR_MED LEs capture the observed SIE declines in summer and winter seasons with good fidelity, despite their biases in pan-Arctic SIE (see Figs. 3c–f). The observed Arctic regional SIE trends generally
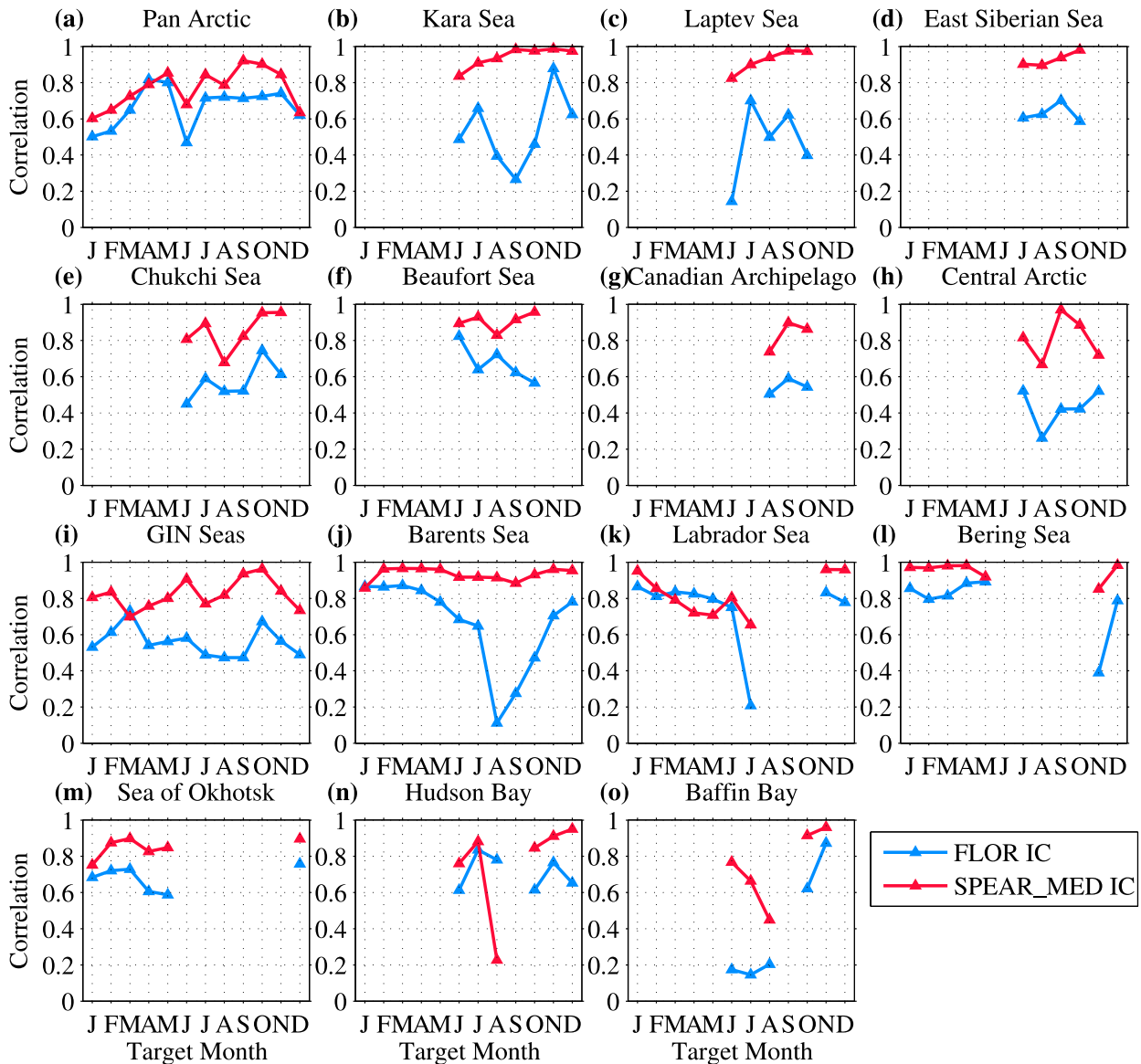
FIG. 4. Regional SIE detrended correlations between FLOR (blue) and SPEAR_MED (red) ICs and NSIDC observations. The correlations are computed over years 1990–2020. For region definitions, see Figs. 6 and 7. Correlation values are plotted for months with SIE standard deviation greater than 0.03 million km$^2$.

fall within the ensemble distribution of simulated trends (not shown), suggesting that the simulated Arctic response to climate forcings in these models is reasonable. The overall performance of the FLOR and SPEAR_MED models, in terms of their mean state, interannual variability, and trends, suggests that they are "fit for purpose" for use as Arctic seasonal sea ice prediction systems, which we explore next.

## 4. Regional sea ice prediction skill

### a. Overview of prediction skill characteristics

In this section, we assess the regional SIE prediction skill of FLOR and SPEAR_MED, focusing on common features and

skill differences between the two systems. We first consider the pan-Arctic September SIE prediction skill, a prediction of great interest to the stakeholder community and the primary focus of the Sea Ice Outlook (Meier et al. 2021). Figure 5 shows that both systems skillfully predict September pan-Arctic SIE at lead times of 0–3 months, capturing both the long-term trend as well as some of the interannual fluctuations. The predictions in Fig. 5 have been bias corrected by a linear regression adjustment, obtained by regressing the observations onto the set of retrospective forecasts. This correction adjusts the mean value and amplitude of the predictions but does not change the ACC values. SPEAR_MED has higher September SIE skill than FLOR for both full and
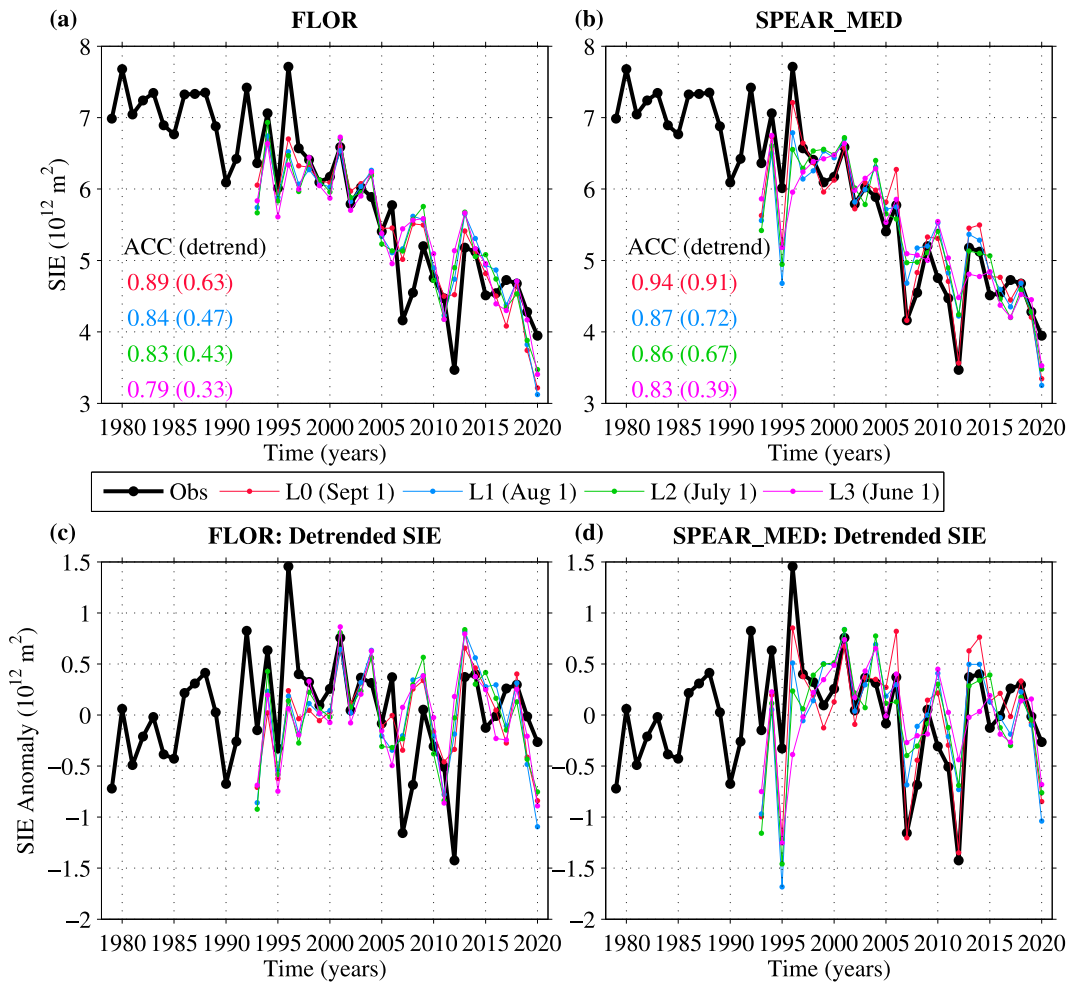
FIG. 5. Predictions of September pan-Arctic SIE in (a) FLOR and (b) SPEAR_MED at lead times of 0, 1, 2, and 3 months. (c),(d) Detrended SIE predictions. ACC values are indicated in colored text and detrended ACC values are indicated in parentheses. These predictions have been bias corrected by a linear regression adjustment.

detrended anomalies. The Sea Ice Outlook requests predictions initialized on 1 June, 1 July, and 1 August, and SPEAR_MED has detrended correlation values at these initialization times of 0.39, 0.67, and 0.72, respectively, showing that skillful detrended predictions of September SIE are possible at these lead times.

In Figs. 6 and 7, we take a more comprehensive view of the skill of these systems, plotting detrended ACC values for all Arctic regions, target months, and lead times of 0–11 months. Note that ACC values are only plotted for target months with SIE standard deviation greater than 0.03 million $km^2$. The nondetrended ACC values are shown in supplementary Figs. S2 and S3. We find that both models exhibit statistically significant seasonal prediction skill for detrended regional SIE anomalies with skill horizons that depend strongly on region, target month, and model. Both FLOR and SPEAR_MED have skill that generally exceeds the skill of the persistence forecast (indicated by square markers); however, there are months and lead times when the models' skill is significant but

lower than persistence (indicated by dot markers). The right columns of Figs. 6 and 7 plot skill differences between the models and show that SPEAR_MED has higher skill than FLOR in nearly all regions of the Arctic (statistically significant differences are indicated by upward triangles), with the exception of the GIN Seas where FLOR's skill is superior to SPEAR_MED's (statistically significant differences are indicated by downward triangles). A generic feature of the regional skill differences is that SPEAR_MED is consistently more skillful than FLOR at lead times of 0–1 month, owing to SPEAR_MED's improved SIE ICs (see Fig. 4). For some target months and regions, FLOR's lead-0 skill is lower than its skill at longer lead times, which is suggestive of errors in SIE ICs. This behavior is not found in SPEAR_MED, consistent with its improved SIE ICs. The sources of seasonal predictability in these systems are explored further in section 5–7.

We also find that the SPEAR_LO system, which employs lower-resolution atmosphere–land components than SPEAR_MED, has very similar skill to SPEAR_MED
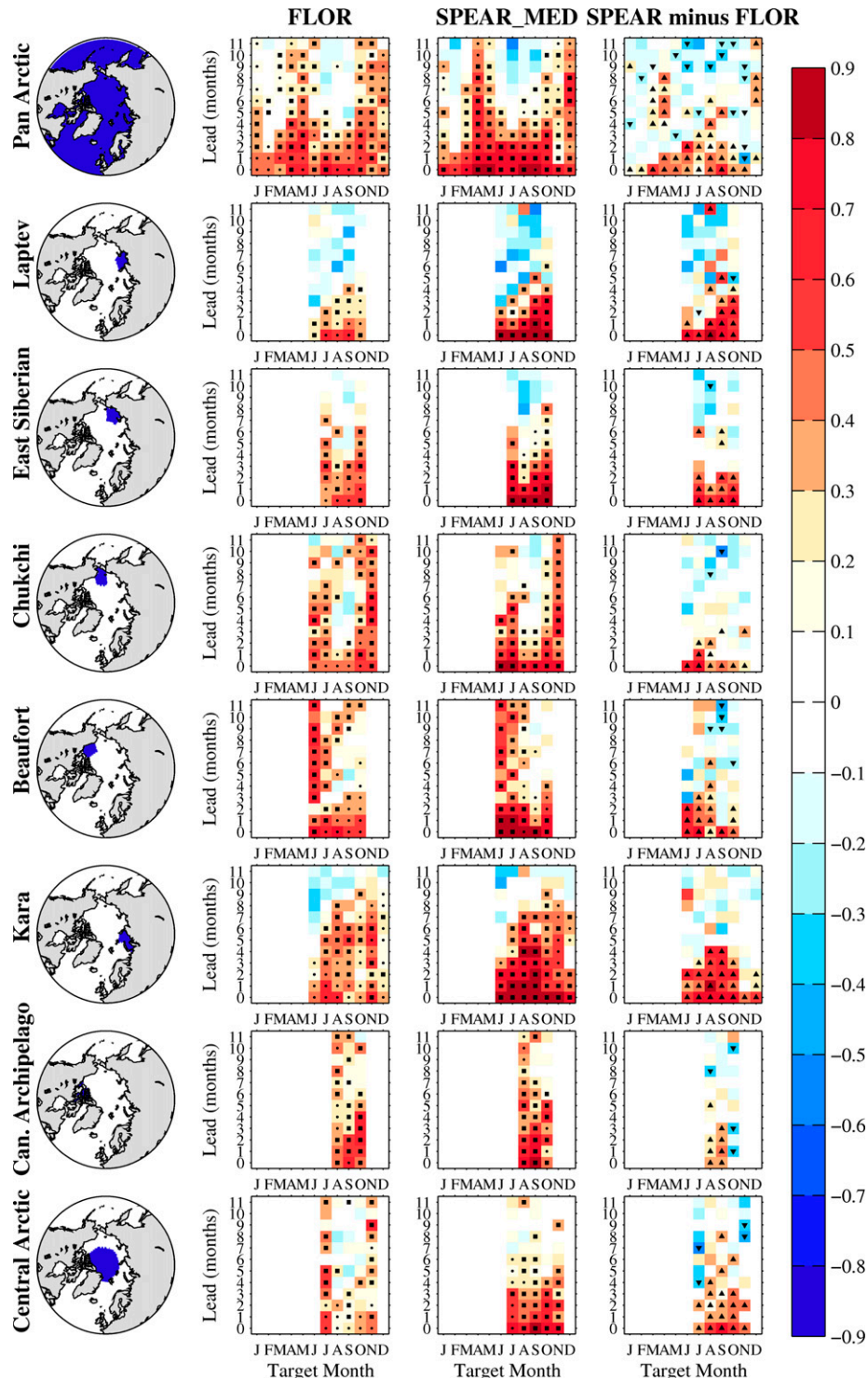
FIG. 6. Regional SIE prediction skill (detrended ACC) in FLOR and SPEAR_MED for regions of summer ice variability. Square and dot markers indicate months in which the ACC values are statistically significant at the 95% confidence level. Squares indicate months where the model's skill beats the persistence forecast, and dots indicate months where the model's skill is significant but lower than persistence. ACC differences (SPEAR_MED minus FLOR) computed using the Fisher $z$ transformation are plotted in the third column. Upward (downward) triangles indicate months where the SPEAR_MED ACC values are statistically significantly greater than (less than) the FLOR ACC values at the 95% confidence level. ACC values are masked for target months with SIE standard deviation less than 0.03 million km$^2$.
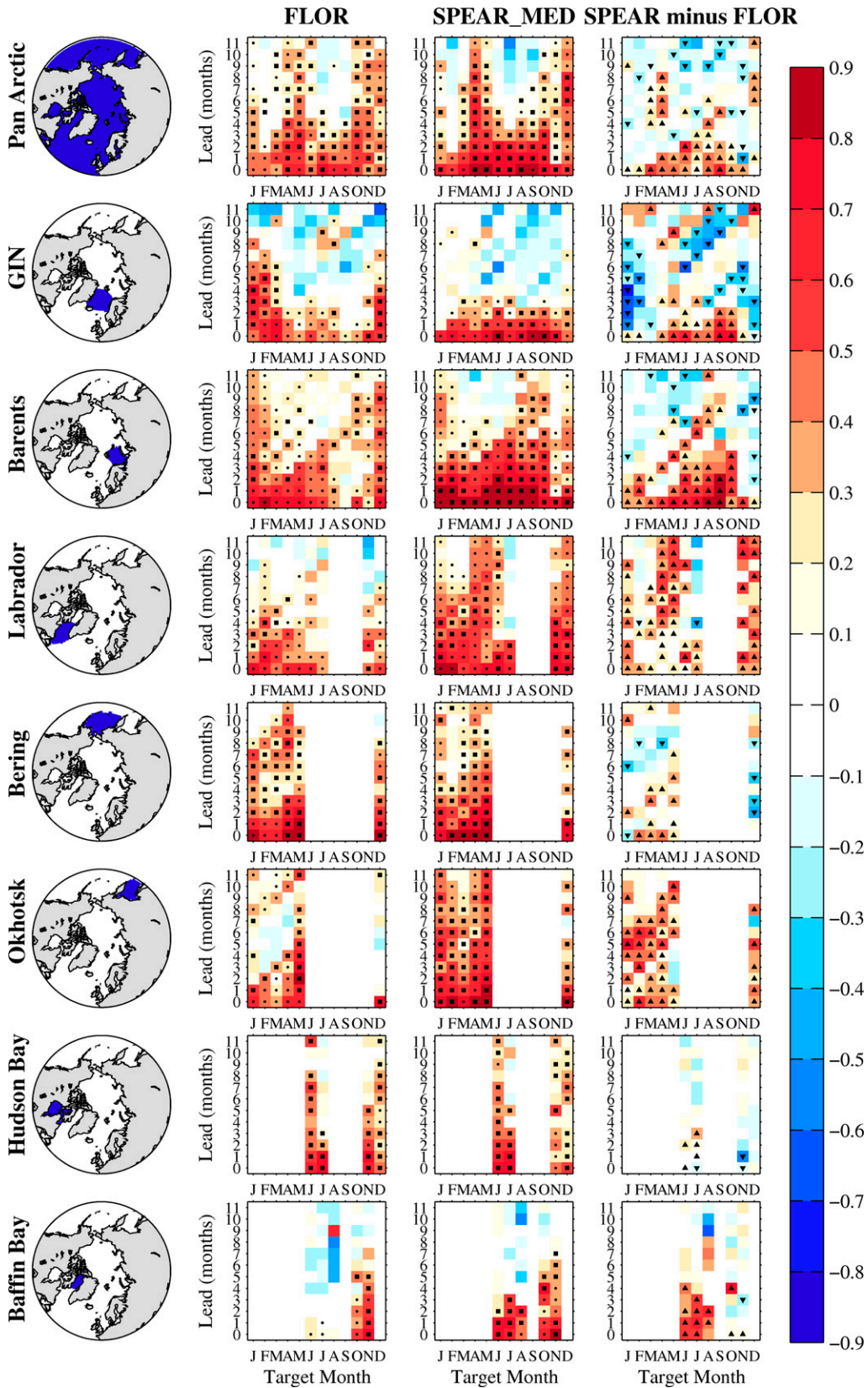
FIG. 7. As in Fig. 6, but for regions of winter sea ice variability.

(see Figs. S4 and S5). This motivates our focus on SPEAR_MED in this study, and also suggests that the computationally efficient SPEAR_LO model can be equally used for Arctic sea ice prediction studies, which is the approach we take in the companion paper of Zhang et al. (2022). Next, we discuss the regional skill characteristics of FLOR and SPEAR_MED.

## b. Summer SIE prediction skill

Figure 6 shows detrended prediction skill values in the regions of summer Arctic sea ice variability. The regional ACC skill structures are generally similar between the two models, but SPEAR_MED consistently has superior skill to FLOR. We find that the Laptev and East Siberian Seas show skillful predictions of summer SIE 2–6 months in advance, with both regions showing a diagonal ACC structure consistent with a spring predictability barrier (Bonan et al. 2019). This barrier corresponds approximately to an initialization date of 1 June: predictions initialized on or after 1 June skillfully predict summer SIE, whereas skill drops off rapidly for predictions initialized prior to this date. Both the Chukchi and Beaufort Seas show distinctive skill structures that are quite consistent between the models. The Chukchi Sea shows a diagonal correlation structure suggestive of a spring barrier for target months August–October, but these months are flanked by longer-lead prediction skill for June, July, and November target months. This skill pattern is suggestive of a seasonally dependent combination of predictability regimes in the Chukchi Sea, which we will return to in section 7. The Beaufort Sea shows skill at 0–3-month lead times for summer SIE and also has longer-lead prediction skill, however, the skill is not continuous across lead times and displays gaps corresponding to April and March initialized predictions. The Kara Sea stands out as a region with high summer prediction skill up to lead times of 8 months in advance, which appears to not be limited by the thickness-based spring predictability barrier. We also find that both systems have skill in the Canadian Arctic Archipelago and that SPEAR_MED is more skillful in the central Arctic domain. We will return to the sources of summer sea ice predictability in these systems in section 5.

## c. Winter SIE prediction skill

Figure 7 shows detrended prediction skill values in the regions of winter Arctic sea ice variability. Compared to the summer regions, we find larger differences in skill structures between FLOR and SPEAR_MED in these regions. The GIN Seas stand out as the only region where FLOR unequivocally outperforms SPEAR_MED. FLOR has skill at 6–8-month lead times for December–March, whereas SPEAR_MED does not have skill beyond 3 months in advance. In the Barents, Labrador, Bering, and Okhotsk Seas, we find skill up to 11 months in advance for predictions of winter and spring sea ice. It is notable that in each of these regions there are lead times and target months for which the dynamical model predictions lose to the persistence forecast, particularly for long-lead predictions of winter and spring SIE. This suggests that both models have

room for improvement in their winter and spring SIE prediction skill, potentially achievable by improving SIE ICs and the model representation of physical processes responsible for persisting SIE anomalies. In the Barents Sea, SPEAR_MED has significantly higher skill than FLOR for target months of May–October and slightly lower skill for November–March, although these winter skill differences are not statistically significant. The Labrador Sea skill is higher in SPEAR_MED than FLOR, however, the SPEAR_MED skill only exceeds persistence for forecasts initialized in fall and summer months. Note that the FLOR Labrador Sea skill reported here is lower than that shown in Bushuk et al. (2017a). This is due to the different reforecast periods considered in these studies (1992–2020 and 1981–2015, respectively), and the associated inclusion of the decadal-scale Labrador sea ice anomalies that occurred between the mid-1980s and the late 1990s.

Bushuk et al. (2017a) highlighted differences between winter prediction skill in the North Atlantic and the North Pacific, with North Atlantic regions having higher skill than their North Pacific counterparts in the FLOR system. Here we find that SPEAR_MED's skill does not clearly reflect this Atlantic–Pacific dichotomy, as this system has long-lead skill in both the Bering Sea and the Sea of Okhotsk. We also find that FLOR's North Pacific skill is somewhat higher over the reforecast evaluation period considered in this study, potentially related to increased interannual-to-decadal Pacific sea ice variability over this period. Both of these results raise questions as to whether there is an inherent Atlantic–Pacific difference in sea ice predictability. The systems have similar skill in the Bering Sea, whereas SPEAR_MED has higher skill in the Sea of Okhotsk owing to improved skill for predictions initialized over the summer and fall months. We also find that both systems have skill in predicting melt season and growth season SIE anomalies in Hudson Bay and Baffin Bay, with SPEAR_MED having better skill for Baffin Bay melt season predictions.

## d. Pan-Arctic SIE prediction skill

Pan-Arctic SIE prediction skill represents a complex combination of these regional skill contributions. Both systems show skillful predictions of summer pan-Arctic SIE up to 3–4 months in advance, with the SPEAR_MED ACC values being consistently higher than FLOR's at lead times of 0–3 months. Target months of October–January and March–May show pan-Arctic SIE skill at 7–11-month lead times in both systems. Interestingly, both systems show relatively low skill for target month February despite the systems having skill in this month on the regional scale.

## 5. Sources of summer Arctic sea ice predictability

We next explore the mechanisms of summer sea ice predictability in the SPEAR_MED and FLOR prediction systems. A body of earlier work has shown that the persistence of SIE and SIV anomalies provide key sources of predictability for summer SIE (see references listed in the Introduction). Inspired by this, we seek a parsimonious description of summer SIE prediction skill, considering

linear predictors based on regional SIE and SIV. Specifically, we construct linear regression models for observed regional SIE using the regional SIE and SIV ICs from each prediction system. For the pan-Arctic domain, the SIV predictor excludes the central Arctic region, as we found that removing this region improved SIV-based prediction skill. This is due to the fact that thickness anomalies in the central Arctic domain do not strongly correlate with September SIE in these systems.

In Fig. 8, we compare the skill of these simple statistical models to the dynamical model skill of FLOR and SPEAR_MED. We find that, in most regions of summer SIE variability, the dynamical model skill (blue curves) can be effectively reconstructed via a combination of regional SIE (black curves) and regional SIV (red curves) predictors. A multilinear regression model based on regional SIE and SIV has skill that closely resembles the maximum skill of the single-variable predictors at each lead time (see Fig. S6). Regional SIE ICs provide the key source of predictability at short lead times (typically 0–1-month lead times), whereas regional SIV ICs provide the key source of predictability at longer lead times of 2–3 months. This analysis reveals the two primary reasons for the summer SIE skill improvements in SPEAR_MED relative to FLOR. First, SPEAR_MED has improved SIE ICs, which improve its regional SIE prediction skill at lead times of 0–1 months. Second, SPEAR_MED has improved SIV ICs, which can explain the model's improved skill at lead times of 2–3 months.

We also experimented with other statistical predictors for summer regional SIE, including surface air temperature, sea level pressure, upper-ocean salt content, and the North Atlantic Oscillation (NAO; Hurrell 1995) index, generally finding that these alternative predictors were unable to provide additional skill beyond that of the SIE and SIV predictors. While the combination of SIE and SIV statistical models can reproduce the dynamical models' skill in most regions, there are some notable exceptions to this. In the Beaufort Sea, we find that the statistical models clearly outperform the dynamical models, particularly in the SPEAR_MED system. This suggests that the models are not capitalizing on sources of sea ice predictability that are present in their ICs. The poor performance of SPEAR_MED in the Beaufort Sea is potentially related to the model's positive SIC bias (Fig. 1f) and biases in SIT and drift (Fig. 2e) in this region. The SPEAR_MED Kara Sea predictions are the one region where the statistical models are unable to fully capture the dynamical model skill, particularly at lead times of 2–4 months (skill differences are statistically significant at leads 2 and 4 months, but not at lead 3 months), suggesting that an additional source of summer SIE predictability is relevant in this region. Future work is required to further investigate this topic.

## 6. Sources of winter Arctic sea ice predictability

### a. Statistical prediction models for winter sea ice

We next consider the sources of predictability for regional winter Arctic SIE. As in the previous section, we draw upon earlier work that has highlighted uOHC and SIE persistence as key sources of winter SIE predictability (see references listed in the introduction), and consider statistical models based on these predictors. We construct linear regression models for regional SIE based on the SPEAR_MED and FLOR uOHC and SIE ICs. The uOHC predictors are based on ocean temperature ICs regionally averaged over the upper 200 m, except for the Bering Sea where we use upper-50-m ocean temperatures due to the shallow bathymetry of the Bering Sea shelf region. The uOHC prediction skill results are robust to the choice of depth range, and produce similar values for lower depth limits ranging from 100 to 300 m. The uOHC prediction skill is generally lower if only subsurface temperatures (excluding the upper 50 m) or only SSTs are used, indicating that both surface and subsurface temperatures provide predictive value.

Figure 9 compares the winter SIE prediction skill of FLOR and SPEAR_MED to the uOHC and SIE statistical predictions. We find that in most regions of winter sea ice variability, the combination of SIE and uOHC predictors are able to reproduce, or even exceed, the skill of the dynamical models. The key source of prediction skill at short lead times (0–2 months) is SIE persistence, whereas uOHC becomes the dominant source of skill at longer lead times (3–11 months). We also find that the SIE predictor displays a clear reemergence of skill for predictions initialized the previous spring. The SIE-based skill at lead times of 7–11 months is generally similar to the uOHC-based skill, consistent with earlier work that has shown that melt-to-growth season SIE reemergence is due to the persistence of uOHC anomalies (Blanchard-Wrigglesworth et al. 2011a). Analogous to the summer skill results, we find that a multilinear regression model based on regional SIE and uOHC has skill that resembles the maximum skill of the single-variable predictors (see Fig. S7). We also investigated other statistical predictors, including SIV, surface air temperature, sea level pressure, upper-ocean salt content, and the NAO index, finding that these predictors generally did not provide additional skill to the SIE and uOHC predictors.

In the North Atlantic regions, the combination of statistical predictors generally exceeds the skill of the dynamical models, implying that there is clear room for improvement in the dynamical predictions. The one exception to this is the FLOR GIN Sea predictions, where dynamical model skill exceeds the FLOR statistical predictions, but shows comparable skill to the SPEAR_MED GIN uOHC predictor. The poor winter skill of SPEAR_MED in the GIN Seas is associated with a region of relatively shallow mixed layers near the Greenland Sea ice edge in the SPEAR_MED ocean ICs. This region of shallower mixing tends to form ice more quickly in winter months, thereby systematically producing positive SIC forecast errors (not shown). The Sea of Okhotsk is the only winter region in which the SPEAR_MED prediction skill is clearly higher than the statistical predictors, suggesting that other sources of predictability are relevant in this region.

### b. The importance of large-scale ocean variability

We have shown that the combination of initialized regional SIE and regional uOHC can explain most of the winter sea
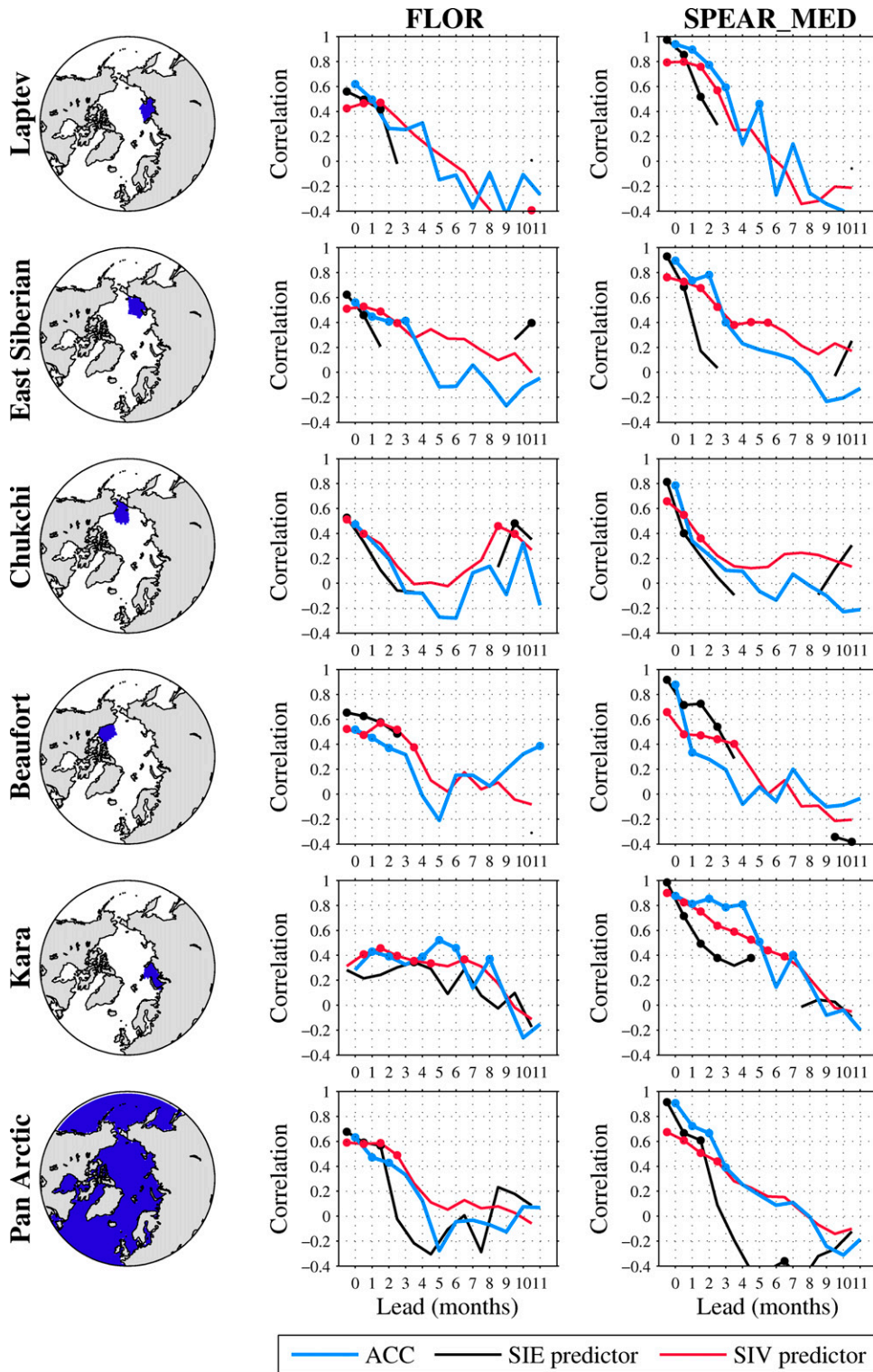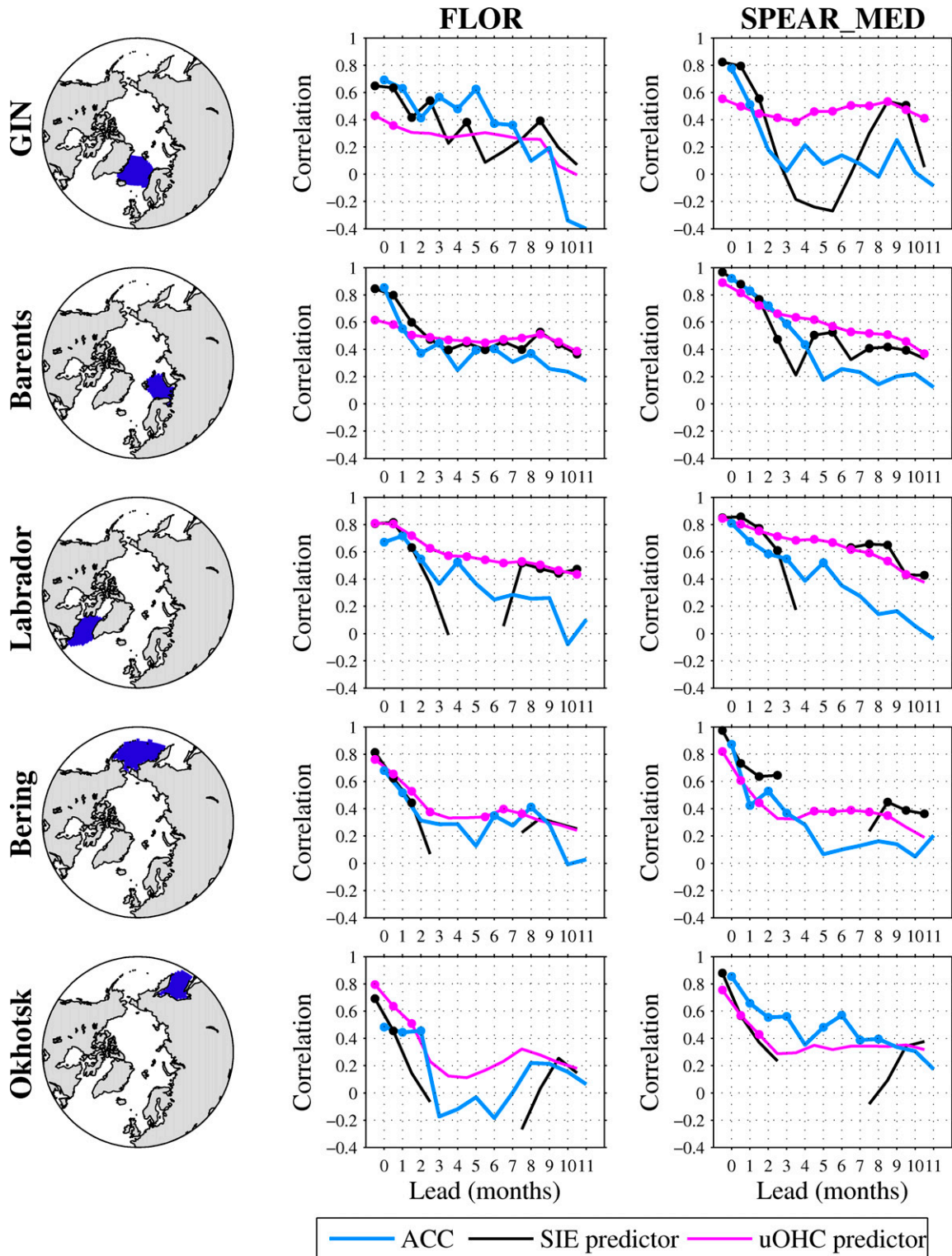
FIG. 8. Sources of regional September SIE prediction skill. Blue lines show the detrended ACC skill in (center) FLOR and (right) SPEAR_MED. Black and red lines show the skill of linear regression forecasts based on regional SIE and regional SIV, respectively. The SIE predictor is plotted for all initial months that have an SIE standard deviation greater than 0.03 million km$^2$. Dots indicate correlation values that are significant at the 95% confidence level based on a $t$ test. Note that the statistical predictions are shifted by 0.5-month lead time since these are computed using monthly mean quantities, whereas the dynamical predictions are initialized on the first of each month.

FIG. 9. Sources of regional February SIE prediction skill. Blue lines show the detrended ACC skill in (center) FLOR and (right) SPEAR_MED. Black and magenta lines show the skill of linear regression forecasts based on regional SIE and regional uOHC, respectively. The SIE predictor is plotted for all initial months that have an SIE standard deviation greater than 0.03 million km$^2$. Dots indicate correlation values that are significant at the 95% confidence level based on a $t$ test. Note that the statistical predictions are shifted by 0.5-month lead time since these are computed using monthly mean quantities, whereas the dynamical predictions are initialized on the first of each month.
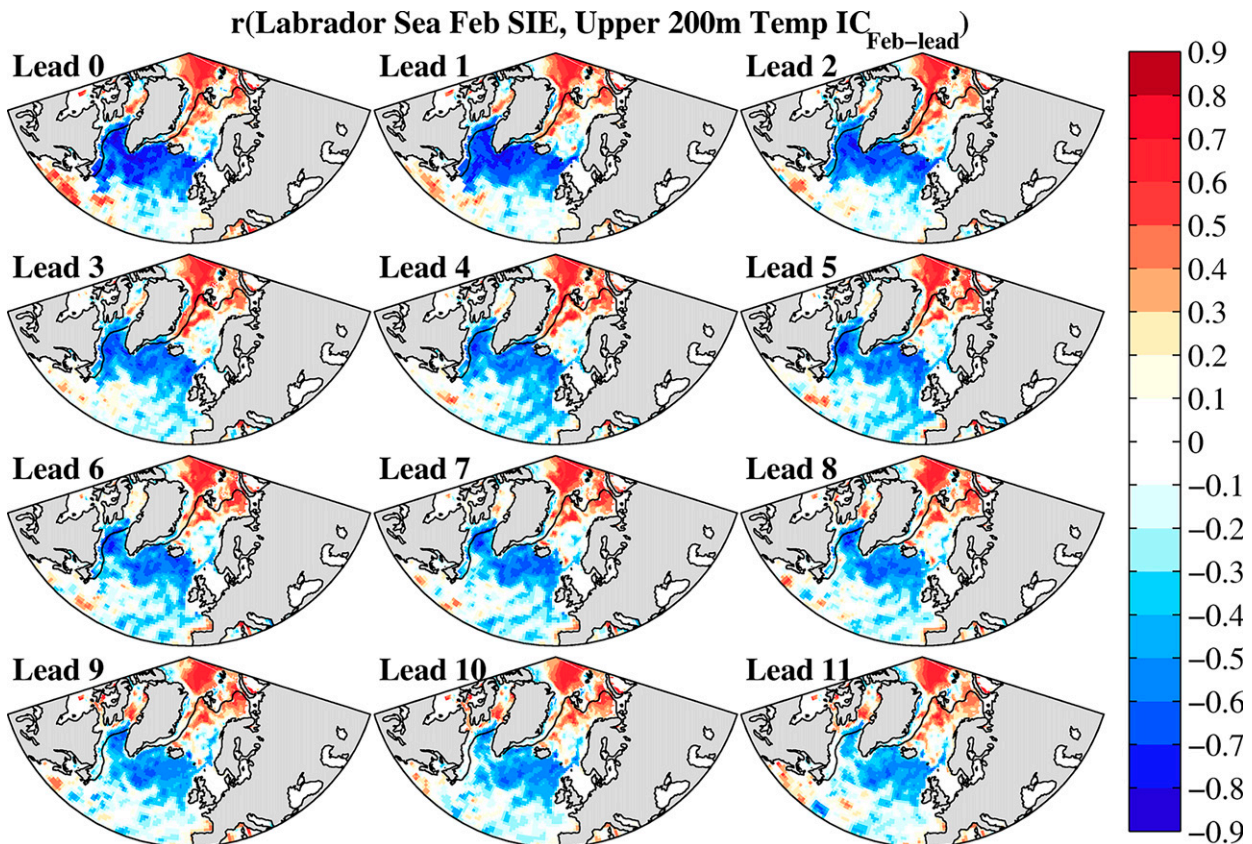
FIG. 10. Correlations between observed February Labrador SIE and upper-200-m ocean temperatures used for SPEAR_MED ICs in earlier months (leads of 0–11 months). The black contours show the observed February sea ice edge.

ice prediction skill in the FLOR and SPEAR_MED models. However, this does not preclude a role for larger-scale ocean variability and dynamics in winter sea ice predictability. To illustrate this point, we consider lagged correlations between observed regional SIE and the earlier upper-200-m ocean temperatures used for ICs in SPEAR_MED.

In Figs. 10 and 11, we plot upper-ocean temperature correlation values with February SIE at different lead times in the Labrador Sea and the Sea of Okhotsk, respectively. As expected physically, we observe negative SIE–temperature correlations local to the region at short lead times. Interestingly, these negative correlations extend spatially throughout the subpolar gyre regions of the North Atlantic and North Pacific sectors, respectively. These broad spatial correlation patterns are present at all lead times, with values that decay for lead times of 0–2 months and are generally maintained at longer lead times of 3–11 months.

The Labrador SIE–temperature correlation pattern closely resembles the SST loading pattern of the NAO, whereas the Sea of Okhotsk correlation pattern closely resembles the SST loading pattern of the North Pacific Gyre Oscillation (NPGO; Di Lorenzo et al. 2008). This suggests that these large-scale modes of climate variability can be leveraged as a potential source of winter sea ice predictability. This predictability can either come from 1) direct

initialization of large-scale uOHC anomalies and the subsequent persistence and advection of these anomalies; or 2) successful dynamical prediction of the mode of variability itself. We find that the NAO index is a skillful predictor of regional winter SIE in the Labrador and Barents Seas, however, the skill is generally lower than that of SIE and uOHC (see Fig. S7). This suggests that the direct initialization approach is critical for winter SIE predictions, and that skillful prediction of the NAO or NPGO index in seasonal prediction systems may extend sea ice prediction skill to longer lead times.

The NAO index prediction skill is relatively modest in both FLOR and SPEAR_MED, with little skill beyond 1- month lead times (not shown), suggesting that most of the SIE skill in these systems is attributable to ocean and sea ice initialization. However, recent work with other dynamical prediction systems has shown that the NAO index can be skillfully predicted months, or even 1 year, in advance, implying a potential for associated long-lead sea ice predictability (Dunstone et al. 2016). These skillful NAO predictions are hampered by the so-called signal-to-noise paradox, which necessitates the use of large prediction ensembles (more than 30 members) to extract the predictable part of the NAO signal (Scaife and Smith 2018), which may be a computational challenge for modern seasonal prediction systems.

### r(Sea of Okhotsk Feb SIE, Upper 200m Temp IC$_{\text{Feb–lead}}$)
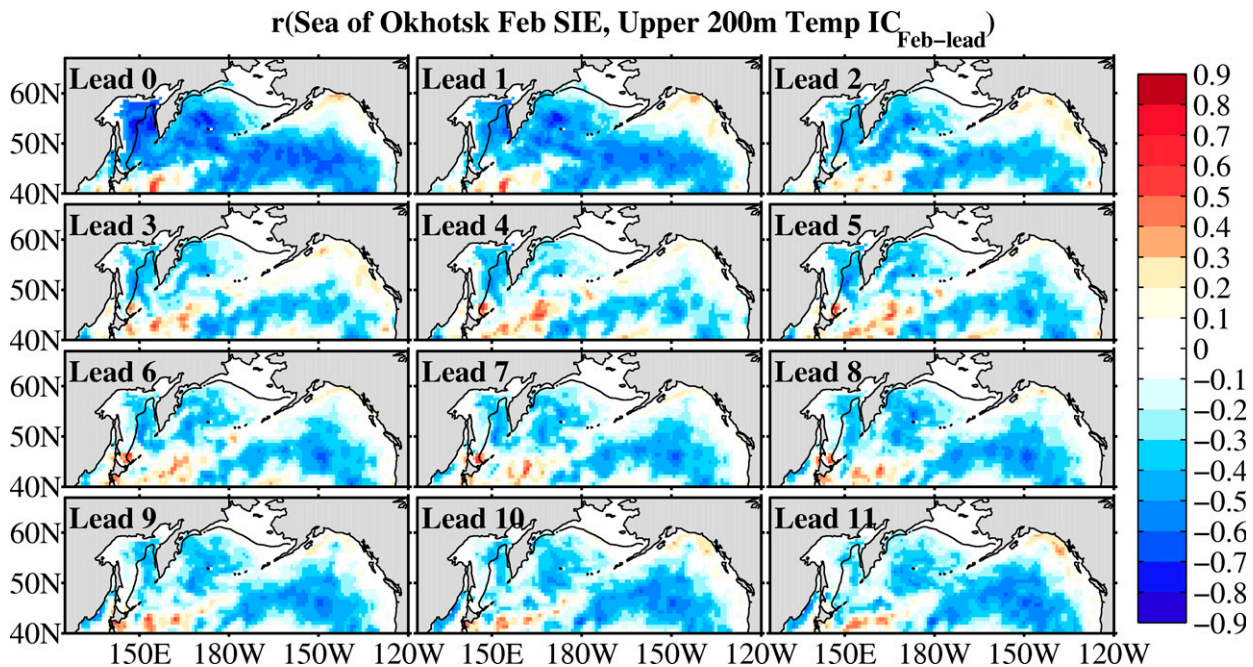


FIG. 11. Correlations between observed February Okhotsk SIE and upper-200-m ocean temperatures used for SPEAR_MED ICs in earlier months (leads of 0–11 months). The black contours show the observed February sea ice edge.

## 7. Combined predictability regimes in the Chukchi Sea

We now consider the prediction skill in the Chukchi Sea. In both SPEAR_MED and FLOR, the Chukchi skill values are relatively modest for target months of August–October but are flanked by longer-lead prediction skill for target months of June, July, and November (see Fig. 6). This skill pattern suggests that a combination of predictability regimes is present in the Chukchi Sea, which depend on the target month. We investigate this seasonally dependent predictability using similar statistical models to section 5 and 6, based on regional SIE, regional SIV, and uOHC. The uOHC predictor is based on upper-50-m ocean temperatures averaged over the Chukchi and Bering Seas. We include the Bering Sea domain in order to capture the waters that flow northward through the Bering Strait into the Chukchi Sea (Woodgate 2018).

The skill of these statistical predictors is compared to the dynamical models' skill in Fig. 12. In both systems, two distinct predictability regimes are evident. The first regime—occurring in June, July, and November—involves uOHC providing the dominant source of predictability and SIE persistence providing short lead (lead 0) predictability. The second regime—occurring in August, September, and October—mirrors the summer sea ice predictability findings of section 5, with SIE persistence providing predictability at short leads (0–1 month) and SIV persistence providing predictability at longer leads. Again, we find that the combination of SIE, SIV, and uOHC statistical predictors is able to match, or in some cases exceed, the skill of the dynamical models.

The ocean-based predictability regime has much higher skill than the thickness-based regime, retaining significant skill values up to lead times of 6–7 months for June and July predictions and up to 11 months for November predictions. It is also notable that the skill of the uOHC predictor drops off sharply for target months of August–October, before reemerging to a much higher value in November. This finding supports the recent work of Lenetsky et al. (2021), who found that observed Bering Strait OHT (averaged over April and May) skillfully predicted Chukchi SIA in June, July, and November, but did not have skill over the intervening SIA minimum months. Lenetsky et al. (2021) additionally found that the prediction skill from Bering Strait OHT was attributable to ocean temperature anomalies rather than volume transport anomalies, consistent with our finding here that uOHC is a skillful predictor.

These findings raise the natural question: Why does ocean-based predictability in the Chukchi Sea drop off abruptly for target months of August–October? We investigate this in Fig. 13, where we plot the observed sea ice edge positions in different months of the year and the climatological (annual mean) ocean surface speed from the SPEAR_MED ocean ICs. The SPEAR_MED ODA system resolves a northward inflow through Bering Strait, which roughly splits into two branches—one flowing north and east along the northern Alaskan coastline and one flowing north and west toward Wrangel Island. The Alaskan branch carries on eastward along the coastline and is located south of the westward surface flow of the Beaufort gyre. The Wrangel Island branch transports waters northward to roughly Wrangel Island, where the current loses strength. The spatial structure of this inflow relative to the sea ice edge position suggests a reason for the loss of ocean-based predictability in August–October.
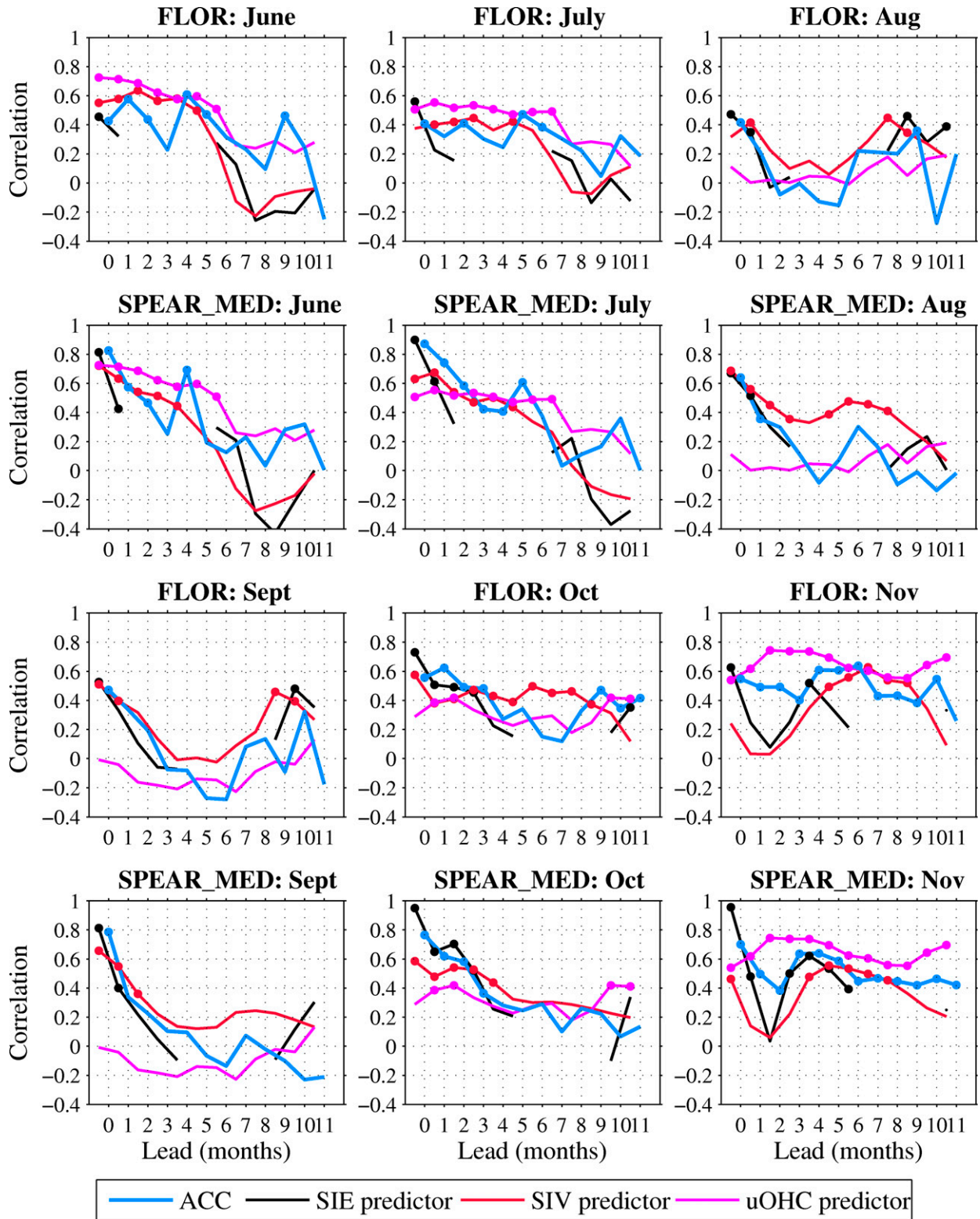
FIG. 12. Sources of Chukchi Sea SIE prediction skill for target months of June–November. Blue lines show the detrended ACC skill in FLOR and SPEAR_MED. Black, red, and magenta lines show the skill of linear regression forecasts based on regional SIE, regional SIV and uOHC, respectively. The uOHC predictor is based on a regional mean over the Chukchi and Bering Seas. The SIE predictor is plotted for all initial months that have an SIE standard deviation greater than 0.03 million km$^2$. Dots indicate correlation values that are significant at the 95% confidence level based on a $t$ test. Note that the statistical predictions are shifted by 0.5-month lead time since these are computed using monthly mean quantities, whereas the dynamical predictions are initialized on the first of each month.
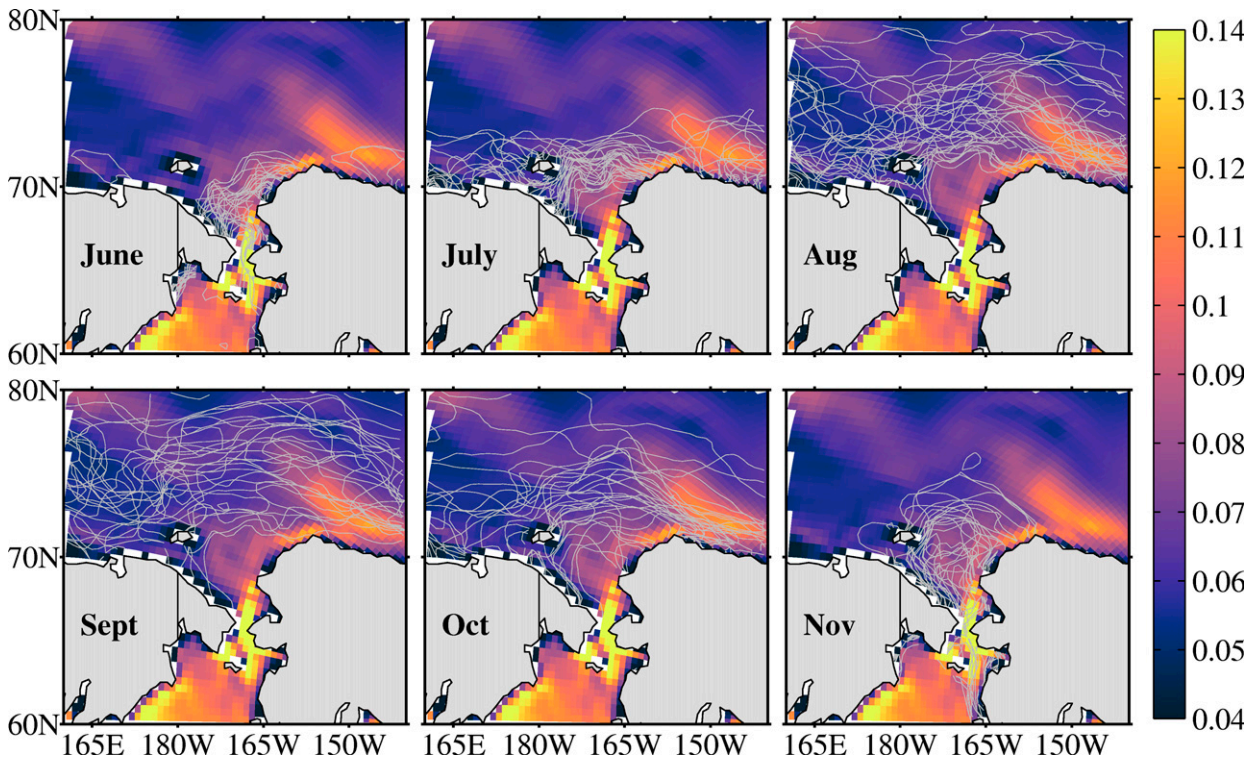
FIG. 13. Interaction of the Chukchi Sea ice edge with surface ocean currents from June through November. The panels show monthly observed sea ice edges plotted as gray contours for each year from 1992 to 2020. The climatological (annual mean) ocean surface speed (m s$^{-1}$) from the SPEAR_MED ocean ICs is plotted in color.

In particular, in June and July, the sea ice edge interacts directly with inflowing waters from the Bering Strait, implying a potential control of upstream uOHC anomalies on the sea ice edge position. By August, however, the ice edge has retreated sufficiently far northward that there is no longer direct interaction with these inflowing waters, suggesting a limited ability for the ocean inflow to impact the ice edge position. The ice edge remains geographically isolated from the inflow waters in September and October, before returning to the inflow region in November, where it interacts with the uOHC anomaly that was present the previous June and July. Note that June, July, and November SIE can each be skillfully predicted using the uOHC anomaly present the previous December (lead times of 5, 6, and 10 months, respectively; see Fig. 12), suggesting that these SIE anomalies arise from a common and persistent uOHC anomaly that is initially generated in early winter.

## 8. Discussion and conclusions

In this study, we have analyzed the regional prediction skill and predictability of Arctic sea ice extent (SIE) using retrospective seasonal forecasts performed with the FLOR and SPEAR_MED dynamical prediction systems. We found that both systems skillfully predict detrended regional Arctic SIE on the seasonal time scale, with skill that is typically higher than that of a persistence forecast. The recently developed

SPEAR_MED system generally has improved skill relative to FLOR, both for pan-Arctic and regional predictions. This improved skill in SPEAR_MED is largely attributable to improved representation of interannual variability and trends in the sea ice concentration (SIC) and sea ice thickness (SIT) fields used for initial conditions (ICs). Compared to FLOR, SPEAR_MED may also benefit from improved initialization of its atmospheric state and differences in model physics, but these aspects have not been explicitly explored in this work.

While SPEAR_MED generally has higher skill values, the correlation patterns of regional SIE skill are similar between the two systems. This consistency in skill patterns across the two independent prediction systems builds our confidence in the robustness of particular skill features. We found that predictions initialized 1 June and later skillfully predicted summer SIE, and predictions initialized 1 May or earlier generally had lower skill values, consistent with the SIT-based spring sea ice predictability barrier (Bushuk et al. 2020). Some notable exceptions to this barrier-type skill behavior were found, such as long-lead prediction skill for June, July, and November SIE in the Chukchi Sea and long-lead prediction skill for July–November SIE in the Kara Sea. We hypothesize that this longer lead summer prediction skill arises due to ocean-based predictability, which extends skill beyond the typical limits expected for SIT-based predictability.

Both FLOR and SPEAR_MED were found to skillfully predict regional winter SIE up to 11 months in advance, but

unlike summer predictions, the systems do not consistently beat the persistence forecast for winter target months. We found that the dichotomy between North Atlantic and North Pacific winter SIE prediction skill identified earlier in FLOR (Bushuk et al. 2017a) is not apparent in SPEAR_MED and is also reduced in this study's analysis of FLOR due to the different time period considered for skill evaluation. This finding suggests that there may not be an inherent Atlantic–Pacific predictability difference, but similar analyses with other prediction systems are required to further investigate this topic.

We explored the physical mechanisms underlying regional sea ice predictability by attempting to reconstruct the dynamical prediction skill of the systems using a parsimonious set of simple statistical models. These statistical models were based on the regional SIE, regional sea ice volume (SIV), and regional upper-ocean heat content (uOHC) ICs used for the forecasts. We found that the skill of the dynamical models can generally be reproduced by a combination of these predictors, suggesting that these three factors constitute the crucial sources of regional sea ice predictability in these prediction systems. We also investigated other statistical predictors, including surface air temperature, sea level pressure, upper-ocean salt content, and the NAO index, generally finding that these alternative predictors were unable to provide additional skill beyond that of SIE, SIV, and uOHC. The key sources of summer sea ice predictability were identified as SIE persistence at short lead times (0–1 month) and SIV persistence at longer lead times (2–3 months). We found that the key sources of winter sea ice predictability are SIE persistence at short lead times (0–2 months) and uOHC persistence at longer lead times (3–11 months). We also found that winter sea ice variations display long-lead covariability with uOHC patterns resembling large-scale modes of climate variability. This analysis suggested that the NAO and NPGO modes provide a source of winter sea ice predictability, and that skillfully predicting these modes may improve sea ice prediction skill.

The Chukchi Sea was found to have seasonally dependent predictability regimes, which are controlled by SIE and uOHC in June, July, and November, and SIE and SIV in August–October. This seasonality in the efficacy of uOHC as a predictor of Chukchi SIE is consistent with the results of Lenetsky et al. (2021). We found that this trade-off in predictability sources was due to the seasonal evolution of the ice edge position and its associated interaction with inflowing waters from the Bering Strait.

Many earlier works have argued for the role of SIE, SIV, and uOHC as sources of seasonal sea ice predictability. To our knowledge, this study provides the first quantitative assessment of the skill attributable to each of these factors. We found that linear regression models based on regional SIE, SIV, and uOHC provide a stringent skill benchmark, which can effectively capture, if not exceed, the skill of the dynamical models. Therefore, we advocate that these regional SIE, SIV, and uOHC regression models should be used as a standard benchmark test for evaluating the skill of regional sea ice predictions. Dynamical prediction skill that exceeds these benchmarks, such as the Sea of Okhotsk and the Kara Sea in this study, will then motivate further research into

additional mechanisms of sea ice predictability. These benchmarks also serve as useful metrics for assessing the fidelity of simulated interannual variability in reanalysis-forced ice–ocean simulations.

## 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.

——, and Coauthors, 2004: The new GFDL global atmosphere and land model AM2–LM2: Evaluation with prescribed SST simulations. *J. Climate*, **17**, 4641–4673, https://doi.org/10.1175/JCLI-3223.1.

Andersson, T. R., and Coauthors, 2021: Seasonal Arctic sea ice forecasting with probabilistic deep learning. *Nat. Commun.*, **12**, 5124, https://doi.org/10.1038/s41467-021-25257-4.

Årthun, M., T. Eldevik, L. H. Smedsrud, Ø. Skagseth, and R. Ingvaldsen, 2012: Quantifying the influence of Atlantic heat on Barents Sea ice variability and retreat. *J. Climate*, **25**, 4736–4743, https://doi.org/10.1175/JCLI-D-11-00466.1.

Babb, D., J. Landy, D. Barber, and R. Galley, 2019: Winter sea ice export from the Beaufort Sea as a preconditioning mechanism for enhanced summer melt: A case study of 2016. *J. Geophys. Res. Oceans*, **124**, 6575–6600, https://doi.org/10.1029/2019JC015053.

——, ——, J. Lukovich, C. Haas, S. Hendricks, D. Barber, and R. Galley, 2020: The 2017 reversal of the Beaufort Gyre: Can dynamic thickening of a seasonal ice cover during a reversal limit summer ice melt in the Beaufort Sea? *J. Geophys. Res. Oceans*, **125**, e2020JC016796, https://doi.org/10.1029/2020JC016796.

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.

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., and M. Bushuk, 2019: Robustness of Arctic sea-ice predictability in GCMs. *Climate Dyn.*, **52**, 5555–5566, https://doi.org/10.1007/s00382-018-4461-3.

——, K. C. Armour, C. M. Bitz, and E. DeWeaver, 2011a: 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.

——, C. Bitz, and M. Holland, 2011b: Influence of initial conditions and climate forcing on predicting Arctic sea ice. *Geophys. Res. Lett.*, **38**, L18503, https://doi.org/10.1029/2011GL048807.

——, R. Cullather, W. Wang, J. Zhang, and C. Bitz, 2015: Model forecast skill and sensitivity to initial conditions in the seasonal Sea Ice Outlook. *Geophys. Res. Lett.*, **42**, 8042–8048, https://doi.org/10.1002/2015GL065860.

——, and Coauthors, 2017: Multi-model seasonal forecast of Arctic sea-ice: Forecast uncertainty at pan-Arctic and regional scales. *Climate Dyn.*, **49**, 1399–1410, https://doi.org/10.1007/s00382-016-3388-9.

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., 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.

Bouillon, S., M. A. M. Maqueda, V. Legat, and T. Fichefet, 2009: An elastic–viscous–plastic sea ice model formulated on Arakawa B and C grids. *Ocean Modell.*, **27**, 174–184, https://doi.org/10.1016/j.ocemod.2009.01.004.

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. Rep. NCAR/TN-472+STR, 100 pp., https://doi.org/10.5065/D6B27S71.

Brunette, C., B. Tremblay, and R. Newton, 2019: Winter coastal divergence as a predictor for the minimum sea ice extent in the Laptev Sea. *J. Climate*, **32**, 1063–1080, https://doi.org/10.1175/JCLI-D-18-0169.1.

Bushuk, M., and D. Giannakis, 2015: Sea-ice reemergence in a model hierarchy. *Geophys. Res. Lett.*, **42**, 5337–5345, https://doi.org/10.1002/2015GL063972.

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

——, ——, ——, ——, ——, ——, and ——, 2017b: Summer enhancement of Arctic sea ice volume anomalies in the September-ice zone. *J. Climate*, **30**, 2341–2362, https://doi.org/10.1175/JCLI-D-16-0470.1.

——, ——, ——, ——, X. Yang, A. Rosati, and R. Gudgel, 2019a: 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.

——, 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.

——, M. Winton, D. B. Bonan, E. Blanchard-Wrigglesworth, and T. Delworth, 2020: A mechanism for the Arctic sea ice spring predictability barrier. *Geophys. Res. Lett.*, **47**, e2020GL088335, https://doi.org/10.1029/2020GL088335.

——, and Coauthors, 2021: Seasonal prediction and predictability of regional Antarctic sea ice. *J. Climate*, **34**, 6207–6233, https://doi.org/10.1175/JCLI-D-20-0965.1.

Caron, L.-P., F. Massonnet, P. J. Klotzbach, T. J. Philp, and J. Stroeve, 2020: Making seasonal outlooks of Arctic sea ice and Atlantic hurricanes valuable—Not just skillful. *Bull. Amer. Meteor. Soc.*, **101**, E36–E42, https://doi.org/10.1175/BAMS-D-18-0314.1.

Cavalieri, D. J., C. L. Parkinson, P. Gloersen, and H. J. Zwally, 1996: Sea ice concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS passive microwave data, version 1. NASA DAAC at the National Snow and Ice Data Center, accessed 1 June 2021, https://doi.org/10.5067/8GQ8LZQVL0VL.

Cheng, W., E. Blanchard-Wrigglesworth, C. M. Bitz, C. Ladd, and P. J. Stabeno, 2016: Diagnostic sea ice predictability in the pan-Arctic and U.S. Arctic regional seas. *Geophys. Res. Lett.*, **43**, 11 688–11 696, https://doi.org/10.1002/2016GL070735.

Chevallier, M., and D. Salas y Mélia, 2012: The role of sea ice thickness distribution in the Arctic sea ice potential predictability: A diagnostic approach with a coupled GCM. *J. Climate*, **25**, 3025–3038, https://doi.org/10.1175/JCLI-D-11-00209.1.

——, ——, 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.

Collow, T. W., W. Wang, A. Kumar, and J. Zhang, 2015: Improving Arctic sea ice prediction using PIOMAS initial sea ice thickness in a coupled ocean–atmosphere model. *Mon. Wea. Rev.*, **143**, 4618–4630, https://doi.org/10.1175/MWR-D-15-0097.1.

Day, J. J., E. Hawkins, and S. Tietsche, 2014a: Will Arctic sea ice thickness initialization improve seasonal forecast skill? *Geophys. Res. Lett.*, **41**, 7566–7575, https://doi.org/10.1002/2014GL061694.

——, S. Tietsche, and E. Hawkins, 2014b: Pan-Arctic and regional sea ice predictability: Initialization month dependence. *J. Climate*, **27**, 4371–4390, https://doi.org/10.1175/JCLI-D-13-00614.1.

——, H. F. Goessling, W. J. Hurlin, and S. P. Keeley, 2016: The Arctic Predictability and Prediction on Seasonal-to-Interannual Timescales (APPOSITE) data set version 1. *Geosci. Model Dev.*, **9**, 2255–2270, https://doi.org/10.5194/gmd-9-2255-2016.

Delworth, T. L., and Coauthors, 2006: GFDL's CM2 global coupled climate models. Part I: Formulation and simulation

characteristics. *J. Climate*, **19**, 643–674, https://doi.org/10.1175/JCLI3629.1.

——, and Coauthors, 2012: Simulated climate and climate change in the GFDL CM2.5 high-resolution coupled climate model. *J. Climate*, **25**, 2755–2781, https://doi.org/10.1175/JCLI-D-11-00316.1.

——, 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.

Di Lorenzo, E., and Coauthors, 2008: North Pacific Gyre Oscillation links ocean climate and ecosystem change. *Geophys. Res. Lett.*, **35**, L08607, https://doi.org/10.1029/2007GL032838.

Dirkson, A., W. J. Merryfield, and A. Monahan, 2017: Impacts of sea ice thickness initialization on seasonal Arctic sea ice predictions. *J. Climate*, **30**, 1001–1017, https://doi.org/10.1175/JCLI-D-16-0437.1.

——, 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.

Drobot, S. D., 2007: Using remote sensing data to develop seasonal outlooks for Arctic regional sea-ice minimum extent. *Remote Sens. Environ.*, **111**, 136–147, https://doi.org/10.1016/j.rse.2007.03.024.

——, J. A. Maslanik, and C. Fowler, 2006: A long-range forecast of Arctic summer sea-ice minimum extent. *Geophys. Res. Lett.*, **33**, L10501, https://doi.org/10.1029/2006GL026216.

Dunstone, N., D. Smith, A. Scaife, L. Hermanson, R. Eade, N. Robinson, M. Andrews, and J. Knight, 2016: Skilful predictions of the winter North Atlantic Oscillation one year ahead. *Nat. Geosci.*, **9**, 809–814, https://doi.org/10.1038/ngeo2824.

Efron, B., 1982: *The Jackknife, the Bootstrap, and Other Resampling Plans.* CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 38, SIAM, 99 pp.

Germe, A., M. Chevallier, D. S. y Mélia, E. Sanchez-Gomez, and C. Cassou, 2014: Interannual predictability of Arctic sea ice in a global climate model: Regional contrasts and temporal evolution. *Climate Dyn.*, **43**, 2519–2538, https://doi.org/10.1007/s00382-014-2071-2.

Giesse, C., D. Notz, and J. Baehr, 2021: On the origin of discrepancies between observed and simulated memory of Arctic sea ice. *Geophys. Res. Lett.*, **48**, e2020GL091784, https://doi.org/10.1029/2020GL091784.

Gnanadesikan, A., and Coauthors, 2006: GFDL's CM2 global coupled climate models. Part II: The baseline ocean simulation. *J. Climate*, **19**, 675–697, https://doi.org/10.1175/JCLI3630.1.

Gregory, W., M. Tsamados, J. Stroeve, and P. Sollich, 2020: Regional September sea ice forecasting with complex networks and Gaussian processes. *Wea. Forecasting*, **35**, 793–806, https://doi.org/10.1175/WAF-D-19-0107.1.

Guemas, V., M. Chevallier, M. Déqué, O. Bellprat, and F. Doblas-Reyes, 2016a: Impact of sea ice initialisation on sea ice and atmosphere prediction skill on seasonal timescales. *Geophys. Res. Lett.*, **43**, 3889–3896, https://doi.org/10.1002/2015GL066626.

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

Harnos, K., M. L'Heureux, Q. Ding, and Q. Zhang, 2019: Skill of seasonal Arctic sea ice extent predictions using the North American Multimodel Ensemble. *J. Climate*, **32**, 623–638, https://doi.org/10.1175/JCLI-D-17-0766.1.

Held, I., and Coauthors, 2019: Structure and performance of GFDL's CM4.0 climate model. *J. Adv. Model. Earth Syst.*, **11**, 3691–3727, https://doi.org/10.1029/2019MS001829.

Holland, M. M., and J. Stroeve, 2011: Changing seasonal sea ice predictor relationships in a changing arctic climate. *Geophys. Res. Lett.*, **38**, L18501, https://doi.org/10.1029/2011GL049303.

——, D. A. Bailey, and S. Vavrus, 2011: Inherent sea ice predictability in the rapidly changing Arctic environment of the Community Climate System Model, version 3. *Climate Dyn.*, **36**, 1239–1253, https://doi.org/10.1007/s00382-010-0792-4.

——, L. Landrum, D. Bailey, and S. Vavrus, 2019: Changing seasonal predictability of Arctic summer sea ice area in a warming climate. *J. Climate*, **32**, 4963–4979, https://doi.org/10.1175/JCLI-D-19-0034.1.

Hunke, E. C., and J. 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.

——, W. H. Lipscomb, A. K. Turner, N. Jeffery, and S. Elliot, 2015: CICE: The Los Alamos Sea Ice Model documentation and software user's manual version 5.1. Doc. LA-CC-06-012, 116 pp.

Hurrell, J. W., 1995: Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science*, **269**, 676–679, https://doi.org/10.1126/science.269.5224.676.

Jung, T., M. A. Kasper, T. Semmler, and S. Serrar, 2014: Arctic influence on subseasonal midlatitude prediction. *Geophys. Res. Lett.*, **41**, 3676–3680, https://doi.org/10.1002/2014GL059961.

——, and Coauthors, 2016: Advancing polar prediction capabilities on daily to seasonal time scales. *Bull. Amer. Meteor. Soc.*, **97**, 1631–1647, https://doi.org/10.1175/BAMS-D-14-00246.1.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. Hnilo, M. Fiorino, and G. Potter, 2002: NCEP–DOE AMIP-II Reanalysis (R-2). *Bull. Amer. Meteor. Soc.*, **83**, 1631–1644, https://doi.org/10.1175/BAMS-83-11-1631.

Kapsch, M.-L., R. G. Graversen, T. Economou, and M. Tjernström, 2014: The importance of spring atmospheric conditions for predictions of the Arctic summer sea ice extent. *Geophys. Res. Lett.*, **41**, 5288–5296, https://doi.org/10.1002/2014GL060826.

Kimmritz, M., F. Counillon, 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.

Kirtman, B. P., and Coauthors, 2014: The North American multimodel ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bull. Amer. Meteor. Soc.*, **95**, 585–601, https://doi.org/10.1175/BAMS-D-12-00050.1.

Koenigk, T., and U. Mikolajewicz, 2009: Seasonal to interannual climate predictability in mid and high northern latitudes in a global coupled model. *Climate Dyn.*, **32**, 783–798, https://doi.org/10.1007/s00382-008-0419-1.

Kondrashov, D., M. D. Chekroun, and M. Ghil, 2018: Data-adaptive harmonic decomposition and prediction of Arctic sea ice extent. *Dyn. Stat. Climate Syst.*, **3**, dzy001, https://doi.org/10.1093/climsys/dzy001.

Krikken, F., M. Schmeits, W. Vlot, V. Guemas, and W. Hazeleger, 2016: Skill improvement of dynamical seasonal Arctic sea ice forecasts. *Geophys. Res. Lett.*, **43**, 5124–5132, https://doi.org/10.1002/2016GL068462.

Krumpen, T., M. Janout, K. Hodges, R. Gerdes, F. Girard-Ardhuin, J. Hölemann, and S. Willmes, 2013: Variability and trends in Laptev Sea ice outflow between 1992–2011. *Cryosphere*, **7**, 349–363, https://doi.org/10.5194/tc-7-349-2013.

Kumar, A., P. Peng, and M. Chen, 2014: Is there a relationship between potential and actual skill? *Mon. Wea. Rev.*, **142**, 2220–2227, https://doi.org/10.1175/MWR-D-13-00287.1.

Lavergne, T., S. Eastwood, Z. Teffah, H. Schyberg, and L.-A. Breivik, 2010: Sea ice motion from low-resolution satellite sensors: An alternative method and its validation in the Arctic. *J. Geophys. Res.*, **115**, 2009JC005958, https://doi.org/10.1029/2009JC005958.

Lenetsky, J. E., B. Tremblay, C. Brunette, and G. Meneghello, 2021: Subseasonal predictability of Arctic Ocean sea ice conditions: Bering Strait and Ekman-driven ocean heat transport. *J. Climate*, **34**, 4449–4462, https://doi.org/10.1175/JCLI-D-20-0544.1.

Levitus, S., and Coauthors, 2013: The World Ocean Database. *Data Sci. J.*, **12**, WDS229–WDS234.

Lindsay, R., J. Zhang, A. Schweiger, and M. Steele, 2008: Seasonal predictions of ice extent in the Arctic Ocean. *J. Geophys. Res.*, **113**, C02023, https://doi.org/10.1029/2007JC004259.

Liu, J., M. Song, R. M. Horton, and Y. Hu, 2015: Revisiting the potential of melt pond fraction as a predictor for the seasonal Arctic sea ice extent minimum. *Environ. Res. Lett.*, **10**, 054017, https://doi.org/10.1088/1748-9326/10/5/054017.

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.

Meehl, G. A., and Coauthors, 2021: Initialized Earth system prediction from subseasonal to decadal timescales. *Nat. Rev. Earth Environ.*, **2**, 340–357, https://doi.org/10.1038/s43017-021-00155-x.

Meier, W., and Coauthors, 2021: 2020 Sea ice outlook post-season report. Sea Ice Prediction Network, https://www.arcus.org/sipn/sea-ice-outlook/2020/post-season.

Meinshausen, M., and Coauthors, 2011: The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change*, **109**, 213–241, https://doi.org/10.1007/s10584-011-0156-z.

Merryfield, W., W.-S. Lee, W. Wang, M. Chen, and A. Kumar, 2013: Multi-system seasonal predictions of Arctic sea ice. *Geophys. Res. Lett.*, **40**, 1551–1556, https://doi.org/10.1002/grl.50317.

Milly, P. C., and Coauthors, 2014: An enhanced model of land water and energy for global hydrologic and Earth-system studies. *J. Hydrometeor.*, **15**, 1739–1761, https://doi.org/10.1175/JHM-D-13-0162.1.

Msadek, R., G. 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.

Onarheim, I. H., T. Eldevik, M. Årthun, R. B. Ingvaldsen, and L. H. Smedsrud, 2015: Skillful prediction of Barents Sea ice cover. *Geophys. Res. Lett.*, **42**, 5364–5371, https://doi.org/10.1002/2015GL064359.

Peterson, K. A., A. Arribas, H. Hewitt, A. Keen, D. Lea, and A. McLaren, 2015: Assessing the forecast skill of Arctic sea ice extent in the GloSea4 seasonal prediction system. *Climate Dyn.*, **44**, 147–162, https://doi.org/10.1007/s00382-014-2190-9.

Petty, A. A., D. Schröder, J. Stroeve, T. Markus, J. Miller, N. Kurtz, D. Feltham, and D. Flocco, 2017: Skillful spring forecasts of September Arctic sea ice extent using passive microwave sea ice observations. *Earth's Future*, **5**, 254–263, https://doi.org/10.1002/2016EF000495.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, https://doi.org/10.1029/2002JD002670.

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.

Riahi, K., and Coauthors, 2017: The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environ. Change*, **42**, 153–168, https://doi.org/10.1016/j.gloenvcha.2016.05.009.

Ricker, R., S. Hendricks, V. Helm, H. Skourup, and M. Davidson, 2014: Sensitivity of CryoSat-2 Arctic sea-ice freeboard and thickness on radar-waveform interpretation. *Cryosphere*, **8**, 1607–1622, https://doi.org/10.5194/tc-8-1607-2014.

Roemmich, D., S. Riser, R. Davis, and Y. Desaubies, 2004: Autonomous profiling floats: Workhorse for broad-scale ocean observations. *Mar. Technol. Soc. J.*, **38**, 21–29, https://doi.org/10.4031/002533204787522802.

Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1058, https://doi.org/10.1175/2010BAMS3001.1.

Scaife, A. A., and D. Smith, 2018: A signal-to-noise paradox in climate science. *npj Climate Atmos. Sci.*, **1**, 28, https://doi.org/10.1038/s41612-018-0038-4.

Schlichtholz, P., 2011: Influence of oceanic heat variability on sea ice anomalies in the Nordic Seas. *Geophys. Res. Lett.*, **38**, L05705, https://doi.org/10.1029/2010GL045894.

Schröder, D., D. L. Feltham, D. Flocco, and M. Tsamados, 2014: September Arctic sea-ice minimum predicted by spring melt-pond fraction. *Nat. Climate Change*, **4**, 353–357, https://doi.org/10.1038/nclimate2203.

Schweiger, A., R. Lindsay, J. Zhang, M. Steele, H. Stern, and R. Kwok, 2011: Uncertainty in modeled Arctic sea ice volume. *J. Geophys. Res.*, **116**, C00D06, https://doi.org/10.1029/2011JC007084.

Serreze, M. C., A. D. Crawford, J. C. Stroeve, A. P. Barrett, and R. A. Woodgate, 2016: Variability, trends, and predictability of seasonal sea ice retreat and advance in the Chukchi Sea. *J. Geophys. Res. Oceans*, **121**, 7308–7325, https://doi.org/10.1002/2016JC011977.

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**, 12 457–12 465, https://doi.org/10.1002/2016GL071396.

Stroeve, J., and D. Notz, 2018: Changing state of Arctic sea ice across all seasons. *Environ. Res. Lett.*, **13**, 103001, https://doi.org/10.1088/1748-9326/aade56.

Sun, C., and Coauthors, 2010: The data management system for the global temperature and salinity profile programme. *Proc. OceanObs'09: Sustained Ocean Observations and Information for Society*, Venice, Italy, ESA,WPP-306, https://doi.org/10.5270/OceanObs09.cwp.86.

Thorndike, A. S., D. Rothrock, G. Maykut, and R. Colony, 1975: The thickness distribution of sea ice. *J. Geophys. Res.*, **80**, 4501–4513, https://doi.org/10.1029/JC080i033p04501.

Tietsche, S., and Coauthors, 2014: Seasonal to interannual Arctic sea ice predictability in current global climate models. *Geophys. Res. Lett.*, **41**, 1035–1043, https://doi.org/10.1002/2013GL058755.

——, E. Hawkins, and J. J. Day, 2016: Atmospheric and oceanic contributions to irreducible forecast uncertainty of Arctic surface climate. *J. Climate*, **29**, 331–346, https://doi.org/10.1175/JCLI-D-15-0421.1.

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.

Walsh, J. E., J. S. Stewart, and F. Fetterer, 2019: Benchmark seasonal prediction skill estimates based on regional indices. *Cryosphere*, **13**, 1073–1088, https://doi.org/10.5194/tc-13-1073-2019.

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.

Wang, X., J. Key, R. Kwok, and J. Zhang, 2016: Comparison of Arctic sea ice thickness from satellites, aircraft, and PIOMAS data. *Remote Sens.*, **8**, 713, https://doi.org/10.3390/rs8090713.

Williams, J., B. Tremblay, R. Newton, and R. Allard, 2016: Dynamic preconditioning of the minimum September sea ice extent. *J. Climate*, **29**, 5879–5891, https://doi.org/10.1175/JCLI-D-15-0515.1.

Winton, M., 2000: A reformulated three-layer sea ice model. *J. Atmos. Oceanic Technol.*, **17**, 525–531, https://doi.org/10.1175/1520-0426(2000)017<0525:ARTLSI>2.0.CO;2.

Woodgate, R. A., 2018: Increases in the Pacific inflow to the Arctic from 1990 to 2015, and insights into seasonal trends and driving mechanisms from year-round Bering Strait mooring data. *Prog. Oceanogr.*, **160**, 124–154, https://doi.org/10.1016/j.pocean.2017.12.007.

Yuan, X., D. Chen, C. Li, L. Wang, and W. Wang, 2016: Arctic sea ice seasonal prediction by a linear Markov model. *J. Climate*, **29**, 8151–8173, https://doi.org/10.1175/JCLI-D-15-0858.1.

Zhang, J., and D. Rothrock, 2003: Modeling global sea ice with a thickness and enthalpy distribution model in generalized curvilinear coordinates. *Mon. Wea. Rev.*, **131**, 845–861, https://doi.org/10.1175/1520-0493(2003)131<0845:MGSIWA>2.0.CO;2.

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

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.

——, and Coauthors, 2022: Subseasonal-to-seasonal sea ice forecast skill improvement from sea ice concentration assimilation. *J. Climate*, **35**, 4233–4252, https://doi.org/10.1175/JCLI-D-21-0548.1.

Zhao, M., and Coauthors, 2018a: The GFDL global atmosphere and land model AM4.0/LM4.0: 1. Simulation characteristics with prescribed SSTs. *J. Adv. Model. Earth Syst.*, **10**, 691–734, https://doi.org/10.1002/2017MS001208.

——, and Coauthors, 2018b: The GFDL global atmosphere and land model AM4.0/LM4.0: 2. Model description, sensitivity studies, and tuning strategies. *J. Adv. Model. Earth Syst.*, **10**, 735–769, https://doi.org/10.1002/2017MS001209.