ABSTRACT: This study assesses the forecast skill of the Canadian Seasonal to Interannual Prediction System (CanSIPS), version 2, in predicting Arctic sea ice extent on both the pan-Arctic and regional scales. In addition, the forecast skill is compared to that of CanSIPS, version 1. Overall, there is a net increase of forecast skill when considering detrended data due to the changes made in the development of CanSIPSv2. The most notable improvements are for forecasts of late summer and autumn target months that have been initialized in the months of April and May that, in previous studies, have been associated with the spring predictability barrier. By comparison of the skills of CanSIPSv1 and CanSIPSv2 to that of an intermediate version of CanSIPS, CanSIPSv1b, we can attribute skill differences between CanSIPSv1 and CanSIPSv2 to two main sources. First, an improved initialization procedure for sea ice initial conditions markedly improves forecast skill on the pan-Arctic scale as well as regionally in the central Arctic, Laptev Sea, Sea of Okhotsk, and Barents Sea. This conclusion is further supported by analysis of the predictive skill of the sea ice volume initialization field. Second, the change in model combination from CanSIPSv1 to CanSIPSv2 (exchanging the constituent CanCM3 model for GEM-NEMO) improves forecast skill in the Bering, Kara, Chukchi, Beaufort, East Siberian, Barents, and the Greenland–Iceland–Norwegian (GIN) Seas. In Hudson and Baffin Bay, as well as the Labrador Sea, there is limited and unsystematic improvement in forecasts of CanSIPSv2 as compared to CanSIPSv1.

KEYWORDS: Arctic; Sea ice; Seasonal forecasting; Dynamical system model; Model evaluation/performance; Model initialization

1. Introduction

We assess the forecast skill of several versions of the Canadian Seasonal to Interannual Prediction System (CanSIPS), a dynamical seasonal forecasting system based on two coupled climate models. The forecast skill of prediction systems is generally assessed through the comparison of forecasts initialized by, and compared to, past observations—a method known as hindcasting. The analysis of hindcasts has been used to assess the skill of various seasonal sea ice forecasting systems (e.g., Bunzel et al. 2016; Krikken et al. 2016; Dirkson et al. 2017; Bushuk et al. 2017). The metric forecasted and observed in these studies is sea ice extent (SIE), which is the total area of all grid cells with ice concentration greater than 15% in a given geographic area. Originally, SIE was primarily used as a metric due to difficulties in discerning grid cells with sea ice concentration less than 15% with remote sensing methods during the summer due to the presence of melt ponds (Comiso and Zwally 1982). While there are noted issues with the use of SIE to assess model skill due to its grid dependence and nonlinear nature (Notz 2014; Notz and SIMIP Community 2020), it is used here due to its continued relevance in the field including by the most recent Sea Ice Outlook (Blanchard-Wrigglesworth et al. 2023), a comparatively similar study (Bushuk et al. 2017), and the reference papers for CanSIPSv1 (Merryfield et al. 2013) and CanSIPSv2 (Lin et al. 2020).

Several common aspects have emerged in studies of dynamical sea ice forecasting systems over the past decade. These include the presence of a predictability barrier affecting forecasts of autumn SIE initialized in the spring, a growing consensus regarding the importance of sea ice thickness (SIT) initial conditions, and a move toward analyzing forecast skill at a regional scale in order to establish greater utility of these systems for end users.

a. The spring predictability barrier

The term “spring predictability barrier” (sometimes “melting season predictability barrier”) refers to a substantial decline in forecast skill of sea ice conditions when predicted at certain lead times. The barrier is often seen as a sharp drop in skill for June-initialized forecasts relative to those initialized in May. This phenomenon was first described as a “melt season predictability barrier” by Day et al. (2014b), in which the authors identified its presence in five coupled climate models. By contrasting the superior skill in predicting September SIE of forecasts initialized in July with those initialized in May, such behavior was also found in the skill of operational seasonal forecasts systems including CanSIPSv1 (Sigmond et al. 2013), NCEP CFSv2 (Wang et al. 2013), and GFDL-FLOR (Bushuk et al. 2017). A pronounced but slightly later skill decline was found for forecasts initialized in the months of June and July in the Met Office’s GloSea4 (Peterson et al. 2015) and the months...
of May and June in GloSea5 (Peterson 2015) when compared to April-initialized SIE forecasts. While still falling under the melting season, the phenomena found in GloSea may be better described as a “summer predictability barrier.” The spring predictability barrier has also been identified in a perfect model experiment involving the GFDL-FLOR model. A perfect model experiment simulates the skill of a system with “perfect” model physics and initial conditions by using the system to predict its own ensemble members (Bushuk et al. 2019). The barrier was also identified in a correlation analysis of sea ice volume (SIV) and sea ice area (SIA) in CMIP5 models (Bonan et al. 2019). The presence of the spring predictability barrier in perfect model experiments suggests that it may be inherent in the climate system and thus may not be entirely eliminated by improvements in initialization procedures or model physics (Bushuk et al. 2019).

To identify the physical mechanism of the spring predictability barrier, Bushuk et al. (2020) decomposed the regional sea ice mass budget into growth, melt, and export terms in two models, CESM1 and GFDL-FLOR. The study concluded that the lack of predictability of sea ice concentrations prior to the spring melt is caused by the variability of sea ice export between regions, which is primarily driven by synoptic-scale atmospheric processes.

b. The importance of sea ice thickness initialization for summer and autumn forecasts

Day et al. (2014a) identified the importance of accurate SIT initialization, in particular for predicting September SIE. Both the interannual variation and long-term decline of SIT as a result of anthropogenic warming (Meredith et al. 2019) have an effect on sea ice evolution. The perfect model forecast skill of a set of runs initialized with near-perfect SIT initial conditions presented higher skill than those initialized with only a SIT climatology. Holland et al. (2019) found that SIT makes a substantial contribution to summer sea ice predictability in the Arctic, noting that the decline of SIT in a warming climate reduced the growth rate of SIT-related forecast errors thereby increasing predictability. These gains are counterbalanced by the fact that a warming climate was found to increase the impact of early summer SIT errors on September sea ice conditions. These two factors together result in an optimal thickness for predictability, found to occur around the year 2010 in the Community Earth System Model Large Ensemble (CESM-LE).

As a result of these effects combined with the difficulty in observing SIT, as compared to SIA, SIT initialization errors substantially affect forecast skill in sea ice models. The findings of Bonan et al. (2019), for example, indicate that a proper SIT initialization in dynamical forecasting systems can help to partially overcome the decline in skill associated with the spring predictability barrier. Bushuk et al. (2020) further underscored the potential for increased predictability with improved initialization of SIT after the onset of melting. Bushuk et al. (2022) assessed the mechanisms of regional predictability for GFDL’s FLOR and SPEAR_MED models. Linear regression forecasts of SIE and SIV as well as SIE and upper ocean heat content were compared against the skill of both dynamical models. The SIV predictor, as a representation of SIT persistence, was found to have higher skill than both dynamical models in many instances.

Despite substantial advances in remote sensing, current observational coverage of SIT is limited, constraining the accuracy of initial SIT conditions for dynamical model-based forecast systems (Mu et al. 2018). There has, however, been recent success in using SIT initialization fields obtained from CryoSat-2. Blockley and Peterson (2018) used all CryoSat-2 SIT observations between 1 and 7 m to initialize GloSea seasonal forecasts noting that these winter SIT initial conditions substantially reduced errors in summer SIE. Further, Allard et al. (2020) found a 43% reduction in error for the U.S. Navy Earth System Prediction Capability’s forecast of the September 2018 minimum SIE when using a CryoSat-2 reanalysis (Allard et al. 2018). Despite such increased observational coverage of SIT, the lack of historical datasets prevents accurate initialization of many hindcasts.

Currently, the most commonly used substitute for SIT observations is the Pan-Arctic Ice Ocean Modeling and Assimilation System (PIOMAS) dataset first described by Zhang and Rothrock (2003), which was first evaluated through comparison to observed thicknesses from a 1993 submarine track in the Beaufort Sea and central Arctic (Zhang and Rothrock 2003). PIOMAS “observations” have become broadly accepted as indicated, for example, by their use as a benchmark to evaluate year-round SIT observations derived from CryoSat-2 data (Landy et al. 2022). Further, Collow et al. (2015) found that the skill in predicting sea ice of 9-month hindcasts produced by Climate Forecast System, version 2 was improved when the model was initialized with PIOMAS SITs. Despite these successes and open access to regular updates of SIT fields, availability of PIOMAS information is not sufficiently timely to be used for the initialization of real-time forecasts. In light of this unmet need, Dirkson et al. (2015, 2017) explored statistical models to provide real-time SIT initialization fields by using PIOMAS SIC and SIT values up to 1 year prior to the initialization date to predict the required SIT for initialization. One of these, Statistical Model version 3 (SMv3) from Dirkson et al. (2017) was subsequently used to provide SIT initialization fields for CanSIPSv1b and CanSIPSv2.

SMv3 was found to be the most skillful of three statistical SIT models in reconstructing fields from PIOMAS, especially for boreal summer predictions and in the Marginal Ice Zone (Dirkson et al. 2017). It was also the most straightforward to implement. SMv3 estimates SIT using a 15-yr training period with known predictors $s_m$ (SIC from PIOMAS) and predictand $\tilde{y}_m$ (SIT from PIOMAS) with the subscript $m$ indexing month. The SIT initialization field value is computed for the target year for initialization ($t_e$) as the local detrended SIC anomaly $s_m^{SIC}(t_e) - s_m^{SIC}(t)$, where $s_m^{SIC}(t)$ represents an extrapolation of the local linear SIC trend, multiplied by a proportionality constant $\alpha$ and then added to a linear extrapolation of the local SIT $\hat{y}_m(t_e)$:
\[ \tilde{y}_m(t_e) = y_m(t_e) + \alpha[x_m^{SIC}(t_e) - \tilde{x}_m^{SIC}(t_e)]. \] (1)

This statistical model was designed to provide real-time SIT fields to initialize CanSIPS models, improving upon the climatological SIT initialization in the CanSIPSv1 models. Dirkson et al. (2017) showed that on the pan-Arctic scale, SMv3 applied in CanCM3 primarily improves March- and May-initialized forecasts. Regionally, the largest increases in skill of the May-initialized forecasts were seen in the central Arctic and extended to the Kara Sea. In CanSIPSv1b, SMv3 is used for the generation of SIT initial conditions for the system’s constituent models, one of which (CanCM4) is also employed in CanSIPSv2. The present analysis will expand upon the initial assessment in Dirkson et al. (2017), which assessed the skill of forecasts by the single model CanCM3 initialized in only the months of March, May, June, and September out to 6 months, by considering the forecast skill of three versions of the CanSIPS multimodel system (two of which include CanCM3), out to 11 months lead time for every initialization month.

c. Regional analyses

Few assessments of sea ice forecast skill in separate regions across the entire Arctic exist in the literature. Sigmond et al. (2016) is one such study which demonstrated the ability of CanSIPSv1 to predict local sea ice advance and retreat dates in various regions (with significant skill generally seen for lead times of 5 months and 3 months, respectively). Krikken et al. (2016) analyzed the regional forecast skill of EC-Earth v2.3 with regards to SIA for all regions in the Arctic, whereas Bushuk et al. (2017) similarly assessed the regional forecast skill of GFDL-FLOR with regards to SIE.

In addition to these comprehensive assessments of seasonal SIE forecasts considering all Arctic regions, there is a significant body of research focusing on identifying mechanisms responsible for predictability of sea ice in individual regions. In such studies, Babb et al. (2020) studied the variety of physical processes in the Beaufort Sea, with particular attention to 2017 when despite high SIV in the winter, substantial melting led to a near ice-free September. Their conclusion, after a set of case studies, was that the region’s sea ice cover was most heavily affected by synoptically driven convergence or divergence of sea ice in a given year, which is generally not predictable on seasonal time scales.

Similar studies considered the predictability of sea ice in the Chukchi Sea with Petrich et al. (2012) considering the sources of predictability for thermal and mechanical breakup of landfast ice and Serreze et al. (2016) identifying ocean heat inflow from the Bering Strait as the single most important predictor of sea ice advance and retreat in the region. Bushuk et al. (2022) outlined a combined regime of predictability in the Chukchi Sea based on both ocean heat transport and atmospheric processes.

Koenigk et al. (2008) found in an analysis of the Barents Sea that regional sea ice anomalies are primarily driven by sea ice import from the central Arctic. These anomalies were found to normally persist for 2 years after their first occurrence. Finally, Bushuk et al. (2017) identified particularly high forecast skill of winter SIE in the North Atlantic sector and through correlation analysis attributed this to upper ocean initialization in the preceding summer and its impact on sea ice during the growth season. Bushuk et al. (2017) also identified high skill for summer SIE in the East Siberian, Laptev, Beaufort, and Chukchi Seas (at lead times of 1–4 months) which through correlation analysis was attributed to SIT initialization in the melt season. In the first three of these regions, a strong spring predictability barrier was identified.

d. Previous analyses of CanSIPS

The first hindcast analysis of the skill of CanSIPS in forecasting Arctic sea ice was that of Sigmond et al. (2013) based on CanSIPSv1. Subsequent hindcast analyses of sea ice in CanSIPS include Sigmond et al. (2016), Dirkson et al. (2017, 2021). One known issue with CanSIPSv1 is the relatively simple SIT initialization, with SIT in the assimilation runs (which provide initial conditions) being nudged toward a fixed, monthly SIT climatology. As SIT has proven to be an important source of sea ice prediction skill, this simplified SIT initialization procedure was viewed as potentially limiting forecast skill. A second issue that potentially limited sea ice prediction skill in CanSIPSv1 is the fact that sea ice concentration in the assimilation runs was nudged toward a dataset that was known to underestimate long-term trends, which are a main source of sea ice prediction skill. Since Sigmond et al. (2016), a new version of CanSIPS, CanSIPSv2 (Lin et al. 2020), has been developed. The main two differences with CanSIPSv1 are 1) improved sea ice thickness and concentration initialization and 2) a change in model combination. The main goal of the present analysis is to document the capability of CanSIPSv2 to forecast Arctic sea ice on seasonal time scales, both on a pan-Arctic scale and for different Arctic regions. A secondary goal is to document improvements compared to CanSIPSv1. By considering an intermediate version that only includes improvements in the SIT and SIC initialization (CanSIPSv1b; Dirkson et al. 2021), the improvements between CanSIPSv1 and CanSIPSv2 are attributed to either the improved sea ice initialization or the change of the model combination. This assessment will be done by analyzing each system’s skill in forecasting SIE in the Arctic from hindcasts initialized at the beginning of every month from January 1980 to December 2018 on both the pan-Arctic and regional scales.

2. Experimental design

This study presents an analysis of the seasonal pan-Arctic and regional SIE forecast skill of three different versions of CanSIPS. This section describes the models used in these systems, the observations used to initialize these models and to assess the hindcasts, the skill metrics used in this assessment, and the approach to determine regional skill. A summary of the initialization products used for each forecast system can be found in Table 1 while the specifications of the systems’ constituent models are summarized in Table 2.
TABLE 1. Overview of constituent models of versions of CanSIPS analyzed in this study and the products used to initialize their SIC and SIT fields.

<table>
<thead>
<tr>
<th>System</th>
<th>Model</th>
<th>SIC initialization</th>
<th>SIT initialization</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEM-NEMO</td>
<td>SMv3</td>
<td>CCMEP GIOPS analysis (2011–18)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. The forecast systems

CanSIPSv1 came into operational use at the Canadian Meteorological Centre (CMC) in December 2011 (Merryfield et al. 2013). This system combines 10 ensemble members from each of the third and fourth generations of the Canadian Coupled Model (CanCM3 and CanCM4) (Merryfield et al. 2013). The atmospheric components of these models are, respectively, the third and fourth generation Canadian Atmosphere General Circulation Model (CanAM3 and CanAM4) and the ocean component of both models is the fourth generation Canadian Ocean General Circulation Model (CanOM4). Sea ice is represented as a cavitating fluid as described in Flato and Hibler (1992) and land surface processes are represented identically in both models (Merryfield et al. 2013).

The resolutions of these models are described in Table 2. The models’ atmospheric components are spectral models considering the atmosphere from the surface to 1 hPa whose physical tendencies are computed on to a linear transform grid. CanAM4 includes upgraded physical parameterizations with regards to radiative transfer, cloud microphysics, shallow convection, and bulk aerosols. Both models also include anthropogenic influences on radiative forcing. The ocean model found in both coupled models, CanOM4, increased resolution from 29 vertical levels in CanOM3 to 40 vertical levels resulting in vertical spacings from 10 m near the surface to greater than 300 m at abyssal depths. Processes represented include subsurface heating by shortwave radiation, diapycnal mixing, and tidal mixing as well as eddy-induced and along-isopycnal diffusion. Coupling of the atmosphere and ocean models occurs once per day. The model physics and initialization of CanSIPSv1 are further described in Merryfield et al. (2013).

The atmosphere and ocean components of each CanSIPSv1 model are initialized using the output of assimilation runs, one conducted for each ensemble member, as detailed in Merryfield et al. (2013). SIC values in the CanSIPSv1 assimilation runs for 1980–2012 are relaxed with a 3-day time scale toward daily observation values obtained from a linear interpolation of the monthly Hadley Centre Sea Ice and Sea Surface Temperature (HadISST) version 1.1 SIC fields (Rayner et al. 2003). SIT is relaxed toward a seasonally varying climatological mean derived from a climate model simulation (Merryfield et al. 2013). From 2013 onward, SIC is nudged toward the Canadian Meteorological Centre (CMC) operational model known as the Global Deterministic Prediction System (GDPS). It is used to initialize real-time predictions (Dirkson et al. 2021) which are derived from the assimilation of several datasets (Buehner et al. 2012, 2013, 2016).

In CanSIPSv1b, SIT was initialized using the SMv3 procedure outlined in Dirkson et al. (2017) and described in the Introduction. This system should be considered as a by-product of CanSIPS development and has never been an operational system. Further, the SIC field in CanSIPSv1b is initialized using Had2CIS data (Lin et al. 2020) for the years 1980–2012 which is comprised of observations of SIC from the HadISST fields.

TABLE 2. Specifications (including vertical and horizontal resolutions) of the atmosphere, ocean, and sea ice components of all the constituent models of the various versions of CanSIPS.

<table>
<thead>
<tr>
<th>Model</th>
<th>CanCM3</th>
<th>CanCM4</th>
<th>GEM-NEMO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atmosphere</td>
<td>31 levels, 2.8° × 2.8°</td>
<td>35 levels, 2.8° × 2.8°</td>
<td>79 levels, 1.4° × 1.4°</td>
</tr>
<tr>
<td>Ocean</td>
<td>40 levels, 1.4° × 0.94°</td>
<td>40 levels, 1.4° × 0.94°</td>
<td>50 levels, 1° × 1°</td>
</tr>
<tr>
<td>Sea ice</td>
<td>Cavitating fluid</td>
<td>Cavitating fluid</td>
<td>CICE 4.0</td>
</tr>
</tbody>
</table>
version 2.2 dataset (Titchener and Rayner 2014) combined with digitized charts from the Canadian Ice Service (Tivy et al. 2011). The SIC trends in HadISST version 2.2 are more accurate than those in HadISST version 1.1 (Sigmond et al. 2013) used in initializing CanSIPSv1. Both these changes are expected to enhance forecast skill. From 2013 to 2018, SIC is initialized using GDPS. Note that in Dirkson et al. (2021) CanSIPSv1b is referred to as “Mod-CanSIPS.”

The model physics and initialization of CanSIPSv2 are described in Lin et al. (2020). The operational implementation of CanSIPSv2 replacing CanSIPSv1 began in July 2019 (Canadian Centre for Climate Modelling and Analysis 2019). Like CanSIPSv1b, CanSIPSv2 utilizes CanCM4 with the SMv3 initialization method [referred to by Lin et al. (2020) as CanCM4i], but CanCM3 is replaced with the GEM-NEMO model. GEM-NEMO uses the GEM atmospheric model (Girard et al. 2014) coupled with the NEMO ocean model (Girvan et al. 2019). The resolutions of these models are specified in Table 2. The GEM model is version 4.8-LTS.13 and the NEMO version is NEMO 3.6 ORCA 1 and they are coupled once an hour through a Globally Organized System for Simulation Information Passing (GOSSIP) coupler (Canadian Centre for Climate Modelling and Analysis 2019). Sea ice is represented in GEM-NEMO using version 4.0 of the CICE model ( Hunke and Lipscomb 2010), which has five thickness categories. The initial SIC and SIT conditions are from Had2CIS and ORAP5 (Zuo et al. 2017), respectively, from 1980 to 2010. Had2CIS SIC was used for initial SIC in lieu of ORAP5 SIC as it was found to be more consistent with the deterministic forecasts of the Global Deterministic Prediction System (W. Merryfield 2023, personal communication). From 2011 to 2018, both the initial SIC and SIT conditions are from the Canadian Centre for Meteorological and Environmental Prediction Global Ice Ocean Prediction System (CCMIP GIOPS) analysis (Smith et al. 2016). Unlike for CanCM4, initial conditions are prescribed directly from observational datasets without assimilation runs. Both GEM, NEMO, and CICE employ finer resolutions for the atmosphere, ocean, and sea ice components than CanCM3 and CanCM4.

Each of the three constituent models (CanCM3, CanCM4, and GEM-NEMO) used a different method to perturb their initial conditions in order to create 10 ensemble members. The initial conditions for CanCM3 and CanCM4 were obtained from assimilation runs that were initialized in, respectively, 1948 and 1958. The 10 initial states for the CanCM3 assimilation runs were obtained from two 50-yr control runs with 1990 radiative forcing that were initialized from PHC/ WOA climatology (Steele et al. 2001), while the 10 initial states for the CanCM4 assimilation runs were obtained from the end of a 350-yr run that was initialized from PHC/WAO conditions and used a procedure to bring the deep ocean into approximate thermal equilibrium with the 1960s climate forcing (Merryfield et al. 2013). By contrast, GEM-NEMO did not use assimilation runs, but instead used reanalysis fields of the atmosphere with random homogenous and isotropic perturbations added which are transformed into wind, temperature, and surface pressure to create 10 ensemble members (Lin et al. 2016, 2020). Lin et al. (2020) noted that there is a potential for initialization shock in GEM-NEMO due to inconsistencies and imbalance resulting from the separate initialization of the atmosphere, ocean and land components.

b. Observations

The Had2CIS dataset is used throughout the present analysis as the observations against which the operational forecast system is compared. This dataset provides SIC which, in addition to being used to initialize CanSIPSv1b and CanSIPSv2 as described above, was interpolated to a $1^\circ \times 1^\circ$ grid to match the regional mask as described below and then used to calculate both pan-Arctic and regional mean SIE.

c. Forecast skill metrics

The forecast skill of each system is assessed using a hindcast of $K = 39$ sample years from 1980 to 2018. Forecast skill in this analysis is defined for the forecast of a given target month and lead time ($\tau$) as the anomaly correlation coefficient (ACC) between the time series of the ensemble means of the modeled predictions ($p$) and the time series of the target month from the Had2CIS observations dataset ($o$) with over-bars denoting time averaging over the set of hindcast years and angle brackets denoting ensemble averaging such that

$$\text{ACC}(\tau) = \frac{\sum_{j=1}^{K} [(\bar{p}(\tau)) - \bar{p}(\tau)](\bar{o}_j - \bar{o})}{\sqrt{\sum_{j=1}^{K} [(\bar{p}(\tau)) - \bar{p}(\tau)]^2} \sqrt{\sum_{j=1}^{K} (\bar{o}_j - \bar{o})^2}. \quad (2)$$

The long-term historical decline in Arctic sea ice as a result of climate change can be seen in a consistent decline in mean SIE for each individual month throughout the observed period. This long-term negative trend is relatively predictable and is the main source of skill when forecasting Arctic sea ice (e.g., Sigmond et al. 2013). It is more challenging for forecast systems to predict interannual deviations from the long-term trend (Stroeve et al. 2014). While we will consider some results regarding forecast skill that include the long-term trend, our main focus is on the ability of CanSIPS to predict deviations from the long-term trend (hereafter referred to as simply interannual variability). Forecast skill for interannual variations is quantified by the ACC after first linearly detrending the time series of the model forecasts and observations. While many previous studies implement detrending using the entire time series (e.g., Sigmond et al. 2013; Guemas et al. 2016; Dirkson et al. 2017), this detrending method is inconsistent with the constraints of an operational forecast where future information is not known and thus cannot be used to determine the trend.

To better frame this analysis in the context of operational forecasting, the detrending conducted in this analysis for any given year removes only the trend from 1980 to the year previous to the given forecast, following Bushuk et al. (2017, 2019). The first 3 years have only the past mean removed—the linear trend for these years is taken to be zero. Only some minor
differences were found between this method and detrending using the entire time series; a representative example is provided in appendix A (Fig. A1). It should be noted that the implication that the Bushuk method linearity and global linearity are similar is not what would be expected from first principles as the global trend of sea ice extent is certainly not linear.

Analyses of ACC for target months where regions are nearly open water or almost completely covered by sea ice for the entire hindcast period are excluded in part to focus this study on prediction of sea ice variability. More practically, target months for which completely open waters are forecasted or observed will have an undefined ACC. Forecast skill for a target month is therefore not calculated if the modeled or observed interannual standard deviation of the region’s SIE for that month is less than 0.8% of the area of the region. This value was chosen as one which excluded substantially the same target months as Bushuk et al. (2017) in order to facilitate a comparative analysis. It is acknowledged that this limitation of the analysis and the primary metric will exclude both high skill forecasts that correctly predict completely ice covered or open waters and low skill forecasts that do not.

All statistical significance tests conducted in this study are computed using a bootstrapping methodology. This procedure creates a distribution of ACCs based on randomly selecting with replacement 39 observation-model forecast pairs from the original time series, calculating the ACC, and repeating this 1000 times. This distribution is then used to determine the p value, or probability that the null hypothesis can be rejected, with statistical significance being defined as the 95% confidence level (a p value of less than 0.05). In considering differences in ACC, the distribution of the differences of the 1000 ACCs is used for the hypothesis test. Two null hypotheses are considered in this study. For the skill assessments, the null hypothesis, which will be referred to as N1, is that there is zero correlation or correlation of the opposite sign of the calculated ACC between the forecasts and observations. For considering the skill differences between versions, the null hypothesis, which will be referred to as N2, is that the two versions have equal ACC (the absolute value of 1000 differences between the calculated ACC of each model has a p value of less than 0.05).

d. Regional analysis

While pan-Arctic analyses can provide insight into broad features of forecast skill, a more detailed picture can be provided by regional analyses. Further, the changing nature of the Arctic climate continues to drive the need for increased regionalization of seasonal sea ice forecasting for both current stakeholders such as government and Indigenous peoples, as well as future stakeholders who may be drawn by the region’s increased accessibility (Eicken 2013). While the regional scale presented here is not sufficiently fine for maritime navigation, regional predictions can inform better operational planning for activities in the Arctic Ocean.

The regions considered in this analysis (Fig. 1) are the same as those defined in Day et al. (2014b) and also used in a number of other studies (e.g., Bushuk et al. 2019, 2020; Bonan et al. 2019). These regions are defined on a 1° × 1° grid.

3. Results

This section presents the SIE forecast skill of CanSIPSv2 and then analyzes improvements in skill compared to CanSIPSv1. By comparison to the forecast skill in CanSIPSv1b, these improvements can be attributed either to the improved sea ice initialization (between CanSIPSv1 and CanSIPSv1b) or change in the model combination (from CanSIPSv1b to CanSIPSv2). In these comparisons, the skill of a forecast of SIE is shown as a function of target month and lead time. For the target month of September, for example, a forecast of lead 0 represents a forecast of September mean SIE that was
initialized on 1 September while lead 1 represents a September forecast that was initialized on 1 August. Further, the skill of the forecast initialized in a certain month can be followed diagonally to longer lead times such as in Fig. 2, for September-initialized forecasts, which begin at September lead 0 and then continue to October lead 1, November lead 2, out to August lead 11.

Skill is quantified as the ACC between the time series of the forecast and observations from Had2CIS as described in section 2. Forecast skill values determined to be statistically significant at the 95% confidence level through bootstrapping are denoted with markers. Dots and triangles indicate significant positive correlation whereas “x” markers indicate significant negative correlation. Triangles indicate a statistically significant forecast skill (relative to the null hypothesis N1) which is also higher than the corresponding anomaly persistence forecast (see below), while dots represent significant skill which is not more skillful than persistence.

It should be noted that in a case of anticorrelation for one version changing to no correlation for the next version, this will appear as an “improvement” in skill on the difference plot. These scenarios are relatively rare, however, with the only significant “improvements” in fact being reductions of significant anticorrelation occurring for: the July forecast initialized 11 months prior on the pan-Arctic scale (Figs. 3a,b), the September forecast initialized 4 months prior in the central Arctic (Fig. 9 top row, second column), the June forecast initialized 10 months prior in the Kara Sea (Fig. 7 third row, second and third columns), and the December forecasts initialized from 6 to 8 months prior in the Sea of Okhotsk (Fig. 9 second row, third column).

Results for the anomaly persistence forecast (Fig. 2) and on the pan-Arctic scale (Fig. 3) are initially presented for time series including trends to illustrate the substantial skill provided by the downward trend in Arctic SIE. As the difference between full and detrended forecasts skill is consistent across all regions, all other results are presented for detrended forecasts only.

Anomaly persistence forecasts represent one of the simplest statistically based seasonal prediction methodologies (Namias 1964). This forecast is created by applying the anomaly of a variable from the present calendar month’s climatological mean to the climatological mean of the calendar month being forecasted. Due to the trivial computational effort required for such a forecast, the anomaly persistence forecast is often used as a baseline from which to assess the merits of far more computationally expensive forecast systems such as CanSIPS (e.g., Sigmond et al. 2013). As shown in the left-hand panel of Fig. 2, anomaly persistence forecasts can have high and significant skill (relative to the null hypothesis N1) in seasonal sea ice prediction due to the substantial autocorrelation present in these time series. Much of this skill is derived from the downward trend in sea ice based on comparison with the right-hand panel of Fig. 2, which shows skill of an anomaly persistence forecast with linear trends removed.

A complete set of plots showing the skill of each of the three CanSIPS versions for each region is presented in appendix B (Figs. B1–B13) along with plots illustrating the differences in skill between versions (CanSIPsv1b-CanSIPsv1, CanSIPsv2-CanSIPsv1b, and CanSIPsv2-CanSIPsv1). While the Canadian Arctic Archipelago region is included in Day et al. (2014a) and other studies, it is not included here as CanCM3 and CanCM4 do not operate at sufficient resolution to properly represent this region.

a. Improvements in pan-Arctic forecast skill for total and detrended anomalies

Forecasts of pan-Arctic SIE predict the total extent of sea ice in the Northern Hemisphere. As found by Sigmond et al. (2013), a large amount of the forecast system’s skill on this scale is rooted in the downward trend of SIE in the Arctic.
Figures 3 and 4 show the forecast skill of all three versions of CanSIPS (top rows) with and independent of the trend, respectively, as well as the differences between versions. Markers in the plots of the top row indicate statistical significance of the skill at the 95% confidence level (relative to the null hypothesis N1), such that a triangle (dot) indicates a statistically significant positive ACC score with greater (less) skill than an anomaly persistence forecast and an “x” indicates a statistically significant negative ACC score. Dots in the plots of the bottom row indicate a statistically significant difference between the two subject forecast systems (relative to the null hypothesis N2). The observations used to assess skill are from the Had2C18 dataset.

Fig. 3. Comparison of the pan-Arctic forecast skills of sea ice extent (quantified as the anomaly correlation coefficient) without detrending for three versions of the CanSIPS model as a function of target month and lead time. (top) Absolute skill of each version of system and (bottom) differences between versions. Markers in the plots of the top row indicate statistical significance of the skill at the 95% confidence level (relative to the null hypothesis N1), such that a triangle (dot) indicates a statistically significant positive ACC score with greater (less) skill than an anomaly persistence forecast and an “x” indicates a statistically significant negative ACC score. Dots in the plots of the bottom row indicate a statistically significant difference between the two subject forecast systems (relative to the null hypothesis N2). The observations used to assess skill are from the Had2C18 dataset.

Regarding the forecast skill of pan-Arctic SIE with trend, we find that CanSIPSv2 shows statistically significant improvements compared to CanSIPSv1 for 43% of the lead and initial month pairs (Fig. 3f). The fraction of forecasts whose skill exceeds that of a persistence forecast increases from 9% in CanSIPSv1 (Fig. 3a), to 51% in CanSIPSv2 (Fig. 3c), driven largely by improvements in forecasts for the months of July to December. Comparison to CanSIPSv1b reveals that most of the forecast skill is due to the improved sea ice initialization procedures. Note that the downward trends in initial fields of both SIC and SIT in CanSIPSv2 are larger (and more realistic) than in CanSIPSv1, which are both expected to contribute to the larger skill with trend.

Regarding the forecast skill of pan-Arctic SIE without trend, we find that CanSIPSv2 shows statistically significant improvements compared to CanSIPSv1 in 14% of the lead and initial month pairs. The fraction of forecasts whose skill for detrended forecasts exceeds that of a detrended persistence forecast increases from 26% in CanSIPSv1 (Fig. 4a) to 45% in CanSIPSv2 (Fig. 4c). As for skill with the trend, we find that most of the skill improvement after removing the trend is due to the improved sea ice initialization. In particular we find the following improvements:

- Improved skill of August and September forecasts, representing an extension of the maximum skillful lead times from 3 to 4 months in CanSIPSv1 (Fig. 4a) to lead times of 5 and 7 months, respectively, in CanSIPSv2 (Fig. 4c).
- In particular, the improvement seen in May-initialized forecasts from CanCM3 using SMv3 in Dirkson et al. (2017) can also be seen in CanSIPSv1b (Fig. 4b) at the pan-Arctic scale. As in the models analyzed by Day et al. (2014b), the spring predictability barrier reveals itself in CanSIPSv1 as a rapid decline in skill over the first 4 months of a May-initialized forecast. In contrast, CanSIPSv2 provides significantly skillful forecasts through to September (Fig. 4c).
- For April-initialized forecasts, CanSIPSv2 provides significantly skillful forecasts through to October (with the exception of July).
- Improved skill in CanSIPSv1b and CanSIPSv2 for the target months of September–November which CanSIPSv1 could not provide significantly skillful forecasts for when initialized earlier than May.
Improved skill for forecasts of the target months of December and January out to a lead time of 11 months.

These improvements when considered together with a large decline of the skill of forecasts with a target month of April result in the more pronounced area of higher skill target month and lead time pairs near the melt season onset in CanSIPSv1 migrating into the middle of winter in CanSIPSv2. There is also a notable though not significant decrease in skill for March and April for all lead times greater than one (as seen in Figs. 4d,f). A precise cause of this decrease in skillful forecasts at and immediately following the growth season maximum is unclear and worthy of future investigation.

b. CanSIPSv2 regional forecast skill

Of greater interest to operational forecasting than pan-Arctic forecast skill is regional skill. The detrended skill of CanSIPSv2 on the pan-Arctic and regional scales is presented in Fig. 5. Regional skill is near its minimum near the Laptev and East Siberian Seas and then increases in skill toward the west (from the Kara and Barents Seas) and the east (through the Chukchi and Beaufort Seas). The seas most closely connected to the North Atlantic—the Greenland–Iceland–Norwegian (GIN) Seas, as well as the Barents and Labrador Seas—all show especially high skill in the winter months (particularly for the target months of January through March where all three regional analyses have skillful forecasts for all lead times up to and including 11 months.).

The majority of CanSIPSv2 regions exhibit some evidence of the spring predictability barrier. In the Kara, Laptev, and Beaufort Seas, the barrier is clearly visible as a pronounced drop in skill with forecasts initialized in May or June being clearly more skillful than those initialized in earlier months. In the GIN Seas, the barrier is seen with May being less skillful than April and indeed tied with October as the initialization month with the fewest forecasts of significant skill (eight). Other regions in which May-initialized forecasts have lower skill than forecasts initialized in April include: the Laptev, East Siberian, and Chukchi Seas. This feature is particularly noticeable when comparing forecasts initialized in May to those initialized in April, especially in the Barents and East Siberian Seas.

There is some evidence of skill reemergence across the September minimum in more confined basins, where the forecast skill for a given initialization month declines across a set of target months and then increases, or reemerges, at longer lead times. This phenomenon is illustrated by the May-initialized forecast in Hudson Bay where June lead 1 and July lead 2 skill reemerges after the melt season as increased skill for November lead 6 and significant skill for December lead 7. Skill reemergence can also be seen in the July-initialized forecast in Hudson Bay (Fig. 5, bottom row) as well as the June-, July-, and August-initialized forecasts in Baffin Bay. There is little evidence of the spring predictability barrier in these confined water regions where waters are virtually ice free in September.

c. Improvement in regional forecast skill

The detrended SIE forecast skill of CanSIPSv2 followed by the skill differences between versions (CanSIPSv2-CanSIPSv1, CanSIPSv1b-CanSIPSv1, and CanSIPSv2-CanSIPSv1b) in each
of the Arctic regions defined in Fig. 1 are shown below (see Figs. 7–9). A complete set of plots showing the skill of each of the three versions for each region is presented side-by-side along with plots illustrating the difference in skill between each version in appendix B (Figs. B1–B13).

Specific regional trends in the different versions of CanSIPS can be found in Fig. 6, which presents the percentage of forecasts with significant positive skill as shown in Figs. 3, 4, and B1–B13. The greater skill of all versions of CanSIPS in the Atlantic regions (GIN, Barents, and Kara Seas) is apparent as compared to the Pacific regions (East Siberian, Chukchi, Bering, and Beaufort Seas). Further evident is the improvement in subsequent versions of CanSIPS in the Atlantic regions and central Arctic. Finally, the same limited improvement (or indeed, deterioration, in the case of the East Siberian Sea) seen in the Pacific regions is also present in the more confined waters of Hudson Bay, Baffin Bay, and the Labrador Sea.

In the North Atlantic (the GIN and Barents Seas, Fig. 7), forecast skill is generally higher overall in all versions of CanSIPS and the improvements to the forecast system with each successive version are readily apparent. As shown in Fig. 6, the GIN, Kara, and Barents Seas are the three regions with the highest percentage of significant positive CanSIPSv2 forecasts. Conversely, in the regions nearer to the Pacific (Fig. 8), forecasts are less skillful and in many cases some changes to the systems have a modest or negative effect on skill. The more confined waters of Hudson and Baffin Bay, as well as the Labrador Sea (Fig. 9), have their own skill patterns with particularly pronounced skill reemergence but little notable change in forecast skill between versions (Fig. 6). Of these three regions, the Labrador Sea has the highest percentage of significant positive forecasts for all three versions of CanSIPS. Finally, the forecast skills in the central Arctic and Sea of Okhotsk (Fig. 9) see unique but significant improvements from CanSIPSv1 to CanSIPSv2 as would be expected given the distinct geography of these regions. In both regions, the most substantial increase in skill is from CanSIPSv1 to CanSIPSv1b (Fig. 6).

CanSIPSv1 forecasts for regions in the Atlantic sector of the Arctic are of substantially greater skill than for other Arctic regions (Figs. B2–B4). With regards to the spring predictability barrier, we see substantial improvement in the Barents Sea (Fig. 7, second row) between CanSIPSv1 and CanSIPSv2 as the May-initialized forecasts with significant skill extend from lead 2 to lead 10. In contrast, relatively little change is found for the GIN Seas (Fig. 7, first row) where the number of May-initialized forecasts with significant skill increases only from lead 1 to lead 3. The new model combination of CanSIPSv2 almost eliminates the spring predictability barrier in the Barents Sea with forecasts as early as the January-initialized forecast providing significant skill higher than persistence.
The Laptev Sea (Fig. 7, last row) shows a smaller but still significant improvement in late summer forecasts at short lead times from CanSIPSv1 to CanSIPSv1b. The Kara Sea (Fig. 7, third row) sees particular improvement from CanSIPSv1 to CanSIPSv1b and from CanSIPSv1b to CanSIPSv2 in the target months of August, September, and October, which all have significant skill at all lead times up to 11 months in CanSIPSv2.

The Arctic seas nearest to the North Pacific demonstrate little improvement when comparing CanSIPSv1b and CanSIPSv2 to CanSIPSv1 (Fig. 8). In fact, the improved sea ice initialization from CanSIPSv1 to CanSIPSv1b is associated with a slight decrease in skill in the East Siberian, Chukchi, Bering, and Beaufort Seas. These regions represent the majority of the skill decline seen in the study. This decline in skill with improved sea ice initialization may point to compensating errors in CanSIPSv1. Despite the modest increase or decline in skill from CanSIPSv1 to CanSIPSv1b, the change in constituent models from CanSIPSv1b to CanSIPSv2 resulted in substantial skill improvements in the East Siberian Sea.

The regions of more confined waters connected to the North Atlantic (Hudson Bay, Baffin Bay, and the Labrador Sea) (Fig. 9) show less skill in all forecast system versions than the more open GIN and Barents Seas (Fig. 7). In fact, there is little improvement seen in Hudson and Baffin Bay as well as the Labrador Sea from CanSIPSv1 to CanSIPSv2 aside from the slightly more skillful forecasts of January, February, and March at lead times of 7–11 months. Skill reemergence is still a feature in the skill of CanSIPSv1b and CanSIPSv2 in these regions. Specifically, in both systems, the skill of July-initialized forecasts before the late summer minimum (when the regions see nearly ice-free waters) reemerge in the November (lead 4) forecasts in Hudson and Baffin Bay (Figs. B11 and B12). This reemergence is not substantially different than in CanSIPSv1.

The forecast skill in the central Arctic (Fig. 9, first row) increases substantially due to the improved sea ice initialization between CanSIPSv1 and CanSIPSv1b with little improvement seen transitioning from CanSIPSv1b to CanSIPSv2. Forecasts in this region improve from not having any significant skill at lead times greater than 2 months to showing significant skill at the majority of lead times for forecasts of October (representing the month of the onset of sea ice growth in a usually ice-covered region).

Finally, substantial improvement in the winter and spring target months is seen in the Sea of Okhotsk (Fig. 9, second row) with the change from CanSIPSv1 to CanSIPSv1b. Whereas few forecasts for these target months initialized before November showed significant skill in CanSIPSv1, all forecasts for January–May initialized as early as July show significant skill in CanSIPSv1b and CanSIPSv2.

d. Comparison to the skill of GFDL-FLOR

The analysis of CanSIPSv2 conducted in the present study can be compared to that of GFDL-FLOR by Bushuk et al. (2017) given the near-identical methodologies particularly in
presenting forecast skill and in detrending of the relevant time series. Indeed of the few differences between the analyses, the most notable is that the analysis of Bushuk et al. (2017) considers the years 1981–2015 versus 1980–2018 for the present study. While CanSIPSv2 and GFDL-FLOR have similar forecast skill on the pan-Arctic scale, CanSIPSv2 has notably higher skill in a number of regions including: the GIN, Barents, and Kara Seas, as well as the central Arctic and Sea of Okhotsk. GFDL-FLOR has substantially higher skill for the target month of August in Hudson Bay and overall in the Labrador Sea. The forecast systems have reasonably similar forecast skill in the remaining regions.

Some notable skill features of CanSIPSv2 are present but less pronounced in GFDL-FLOR. There is a similar geographic distribution of skill in GFDL-FLOR as in CanSIPSv2, but given the lower skill of GFDL-FLOR in the Atlantic sector (GIN, Barents, and Kara Seas), the difference in skill between regions is not as substantial. The spring predictability barrier is present on the pan-Arctic scale in GFDL-FLOR but regionally is only clearly visible in the Laptev Sea (as in CanSIPSv2) and East Siberian Sea (unlike in CanSIPSv2). Skill reemergence in the more confined waters of Hudson Bay, Baffin Bay, and the Labrador Sea occurs comparably in both forecast systems.

e. Predictive skill of sea ice volume initial conditions

The increase in skill of SIE forecasts between CanSIPSv1 and CanSIPSv1b supports the importance of sea ice initialization given that the SIC and SIT initialization procedure is the sole difference between the systems. Recent methods have
been developed to further assess forecast skill that is associated with sea ice initialization. Bushuk et al. (2017) compared spatial grids of SIT and SIE throughout the Arctic with a focus on the East Siberian Sea to ascertain whether SIT is a source of summer predictability. Further, Bushuk et al. (2022) constructed a simple statistical linear regression model, in which the SIV initial conditions in the GFDL-FLOR and SPEAR_MED models were used as a predictor of the observed SIE. The skill of such statistical predictions can be interpreted as the skill that is associated with sea ice initialization.

As GEM-NEMO SIV fields were not archived and CanCM3 SIV fields are virtually identical to those in CanCM4 (not shown), this analysis will compare the skill of the SIV initial conditions of CanCM4 and CanCM4i as statistical predictors of the observed SIE. The skill of such statistical predictions can be interpreted as the skill that is associated with sea ice initialization.

As GEM-NEMO SIV fields were not archived and CanCM3 SIV fields are virtually identical to those in CanCM4 (not shown), this analysis will compare the skill of the SIV initial conditions of CanCM4 and CanCM4i as statistical predictors of the observed SIE. Figure 10 shows the correlation between the initial regional or pan-Arctic SIV and the observed SIE in the following months. For example, September lead 2 presents the correlation between July SIV initial conditions and observed September SIE. The results shown are for the pan-Arctic and three particular regions (central Arctic, Kara Sea, and East Siberian Sea). Also as in previous analyses, skill is not presented if the interannual standard deviation of the observed SIE is less than 0.8% of the area of the region. Plots for all regions can be found in appendix C (Figs. C1 and C2).

The predictive skill of CanCM4 initial SIV conditions is not negligible, despite the fact that the CanCM4 initial conditions are obtained from assimilation runs in which SIT is nudged toward a climatological mean. We find, that the initial SIV conditions in CanCM4 have sufficiently realistic interannual variability to have statistical predictive power for future SIE. This is likely due to the fact that SIV is not only determined by sea ice thickness, but also by sea ice area. Since the initial
sea ice area is initialized with observed interannual variations, SIV initial conditions contain some of the interannual variability that contributes to predictive skill of SIE. In addition, despite the nudging to climatological SIT in the assimilation runs, the initial SIT contains some interannual variations, which are obtained indirectly through the nudging of atmosphere, ocean and SIC fields in the assimilation run toward observational values. Such interannual variability in initial SIT further contributes to the interannual variations in initial SIV.

Figure 10 shows regions in which changes in skill between CanSIPSv1 and CanSIPSv1b match the difference in lagged correlation of the respective initial SIV fields to SIE. This analysis suggests that for these regions, the forecast skill increases (of detrended anomalies) can be mainly attributed to the change in initialization of the model. As described in
The improved initialization procedure developed by Dirksen et al. (2017) produced a more skillful forecast system on both the pan-Arctic scale and in several regions. This fact is demonstrated by the skill increase seen on the pan-Arctic scale as well as in the central Arctic, GIN Sea, Barents Sea, Kara Sea, and Sea of Okhotsk (Fig. 6). It should be noted, however, that forecast skill in the East Siberian Sea decreased in CanSIPSv1b and little or no difference was seen in the remaining regions (the Laptev, Chukchi, Bering, and Beaufort Seas as well as Hudson Bay and Baffin Bay). A potential cause of this could be substantial sea ice export in these regions reducing the amount of predictive skill provided by effective SIT initialization. The substitution of CanCM3 with the GEM-NEMO model in CanSIPSv2 appears to provide the majority of skill
improvements in the seas north of Siberia (the East Siberian, Kara, and Chukchi Seas), in combination with the new initialization procedure to improve on the already high skill seen in the GIN and Barents Seas in the North Atlantic.

Consideration should be given to the fact that initial conditions provided by Dirkson et al. (2017), while improved, remain imperfect. Specifically, in regions nearer to the Pacific, the initial conditions provided by SMv3 are less accurate when compared against PIOMAS, particularly during the winter (Dirkson et al. 2017). The analysis of the lagged correlation of SIV initial fields to SIE found that on the pan-Arctic scale and in the central Arctic and Kara Sea, SIV initialized with the new procedure was a far better predictor of regional SIE than the SIV initialized from assimilation runs that were nudged toward a SIT climatology and used HadISST vice with the new procedure was a far better predictor of regional scale and in the central Arctic and Kara Sea, SIV initialized with the new procedure was a far better predictor of regional SIE than the SIV initialized from assimilation runs that were nudged toward a SIT climatology and used HadISST vice Had2CIS. These findings further support the initialization procedure’s contribution to the system’s skill in these regions. The analysis similarly demonstrated that the SIV produced using the new initialization procedure was a poor predictor of SIE in the East Siberian Sea potentially contributing to the decrease in skill in this region between CanSIPSv1 and CanSIPSv1b. Finally, consideration should be given to the fact that in any dynamical model forecast system, skill may be rooted in compensating model biases (as was suggested in the case of CanSIPSv1 regarding the seas north of Greenland by Dirkson et al. 2017). One example could be a cold bias which results in incorrectly thin sea ice growing to its correct thickness. Further investigation of model biases is merited to help illuminate such cases. In a case where improved initialization reduces some model biases but not others, forecast skill could in fact decrease as a result.

It is further important to note that in addition to potential errors in the initial conditions themselves, the regions of the Pacific sector are also notable for the influence that unpredictable atmospheric circulation patterns have on the regions’ sea ice (Petrich et al. 2012; Serreze et al. 2016). Indeed such processes can dampen the predictability provided by other mechanisms such as ocean heat transport as Bushuk et al. (2022) noted in the Chukchi Sea. This limitation is further supported by the results of Babb et al. (2020) which show that a variety of local dynamic processes occurring throughout the melt season may dampen the correlation between winter and summer sea ice conditions in the Beaufort Sea. In the Atlantic, previous studies have suggested that high skill of seasonal sea ice forecasting systems can be partially attributed to the models’ ability to represent the thermohaline circulation from the broader Atlantic and Arctic Oceans (Guemas et al. 2016) due to the importance of ocean temperatures in predicting SIE, particularly in the winter (Sigmond et al. 2013; Bushuk et al. 2017, 2022).

Regional reemergence features have been previously noted for CanSIPSv1 in Sigmond et al. (2016) which suggested they were the result of the persistence of ocean temperatures acting as a memory for melt season sea ice anomalies, allowing those anomalies to reemerge during the growth season. This phenomenon has also been seen in other models (Day et al. 2014b; Guemas et al. 2016) and is also present in CanSIPSv2. The Labrador Sea as well as Hudson and Baffin Bays show the strongest signals of skill reemergence following the melting season of all regions considered. This feature is likely most prominently displayed in these regions due to less heat transport from other regions affecting the local sea surface temperatures as compared to regions exposed to the circulation within the larger Arctic Basin or to the Pacific and Atlantic Oceans through the Bering and Fram Straits, respectively.

In this paper, we have focused on quantifying the skill of ensemble mean deterministic forecasts. A limitation of this approach is that we did not use information that is contained in the ensemble spread, which can be used to provide arguably more user-relevant probabilistic forecasts. Due to the bounded nature of sea ice, calibration of probabilistic sea ice forecasts and quantifying the skill of such forecasts is not a straightforward exercise (see e.g., Dirkson et al. 2021). Therefore, while we recognize the importance of probabilistic forecasting, quantification of skill differences of probabilistic forecasts between CanSIPSv1 and CanSIPSv2 is left for future investigation. One factor likely affecting prediction skill is the use of the cavitation fluid rheology of Flato and Hibler (1992) in both versions of CanCM3 and CanCM4. Assessment of the consequences for ice predictive skill of using a more modern rheology is also an interesting direction of future study. Indeed in the next major version of CanSIPS, CanCM4 will likely be replaced with the Canadian Earth System version 5 (Swart et al. 2019), which includes a more modern sea ice component.

This analysis of the forecast skill found that CanSIPSv2 has higher skill than previous CanSIPS versions on the pan-Arctic scale and in nearly all regions of the Arctic. This improvement in skill can be partly attributed to the improved initialization procedure developed by Dirkson et al. (2017) as evidenced by the improvements in skill from CanSIPSv1 to CanSIPSv1b and by an analysis of the predictive skill of SIV in various regions. Other regions also saw an improvement in skill from the replacement of the constituent model CanCM3 with GEM-NEMO. Previously noted phenomena in other models such as the spring predictability barrier and skill reemergence in some regions remain in CanSIPSv2 although to a lesser extent in the case of the former. CanSIPSv2 more broadly demonstrates the benefits of the changes made in the development of the CanSIPS forecast system and specifically underscores the importance of properly initializing sea ice thickness in particular when developing skillful Arctic seasonal sea ice forecast models.

Acknowledgments. The authors would like to acknowledge the Royal Canadian Navy for its financial support of Joseph Martin during the graduate program that produced this research, as well as Bill Merryfield and Marika Holland for their review of the work and Woosung Lee for providing access to the relevant model data. They would further like to thank three anonymous peer reviewers for their thorough revision and helpful insights, which improved this manuscript.

Data availability statement. CanSIPS hindcast data and regional mask are available upon request.
APPENDIX A

Differences in Detrending Methods

Figure A1 shows the sensitivity of the forecast skill of Pan-Arctic SIE to the detrending method. Comparison between the top row (linear detrending using the entire time series) and the bottom row (linear detrending using the Bushuk et al. (2017) method) shows that differences of forecast skill are minimal.

Fig. A1. Pan-Arctic forecast skills of sea ice extent for versions 1 and 2 of the CanSIPS model as a function of target month and lead time based on (top) standard linear detrending and (bottom) the method described in Bushuk et al. (2017) used in the present study. Markers in the left and center panels indicate statistical significance of the skill at the 95% confidence level (relative to the null hypothesis N1), such that a triangle (dot) indicates a statistically significant forecast with greater (less) skill than an anomaly persistence forecast. Dots in the right panels indicate a statistically significant difference between the two subject forecast systems (relative to the null hypothesis N2). The observations used to assess this skill are from the Had2CIS dataset.
APPENDIX B

Forecast Skill and Differences between all Versions of CanSIPS

Figures B1–B13 show the forecast skill of, and difference in forecast skill between, different versions of CanSIPS for each region.

Fig. B1. As in Fig. 4, but for the central Arctic.
FIG. B2. As in Fig. 4, but for the Greenland–Iceland–Norwegian Seas.

FIG. B3. As in Fig. 4, but for the Barents Sea.
Fig. B4. As in Fig. 4, but for the Kara Sea.

Fig. B5. As in Fig. 4, but for the Laptev Sea.
Fig. B6. As in Fig. 4, but for the East Siberian Sea.

Fig. B7. As in Fig. 4, but for the Chukchi Sea.
FIG. B8. As in Fig. 4, but for the Bering Sea.

FIG. B9. As in Fig. 4, but for the Sea of Okhotsk.
FIG. B10. As in Fig. 4, but for the Beaufort Sea.

FIG. B11. As in Fig. 4, but for Hudson Bay.
FIG. B12. As in Fig. 4, but for Baffin Bay.

FIG. B13. As in Fig. 4, but for the Labrador Sea.
APPENDIX C

Correlation of Pan-Arctic and Regional SIV and SIE

Figures C1 and C2 display the lagged correlation (quantified as the ACC) between the mean of the SIV initialization fields of CanCM4 and CanCM4i and observed SIE for all regions considered in this study. Figure 10 shows only those regions where the changes in lagged correlation match the changes in skill between CanSIPSv1 and CanSIPSv1b.

Fig. C1. As in Fig. 10, but for the pan-Arctic, central Arctic, GIN Seas, Barents Sea, Kara Sea, and Laptev Sea.
REFERENCES


