Seasonal Forecast Skill and Potential Predictability of Arctic Sea Ice in Two Versions of a Dynamical Forecast System

by

Joseph Martin
BEng, Royal Military College of Canada, 2013
MBA, Management Centre Innsbruck, 2019
MGM, Royal Roads University, 2020

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

in the School of Earth and Ocean Sciences

Seasonal Forecast Skill and Potential Predictability of Arctic Sea Ice in Two Versions of a Dynamical Forecast System

by

Joseph Martin
BEng, Royal Military College of Canada, 2013
MBA, Management Centre Innsbruck, 2019
MGM, Royal Roads University, 2020

Supervisory Committee

Dr. Michael Sigmond, Supervisor
(School of Earth and Ocean Sciences)

Dr. Adam H. Monahan, Co-Supervisor
(School of Earth and Ocean Sciences)

Supervisory Committee

Dr. Michael Sigmond, Supervisor
(School of Earth and Ocean Sciences)

Dr. Adam H. Monahan, Co-Supervisor
(School of Earth and Ocean Sciences)

ABSTRACT

As the decline in Arctic sea ice extent makes this region more accessible, the need is increasing for effective seasonal sea ice forecasting to facilitate operational planning. Recently, coupled global climate models (CGCMs) have been used to address the need for effective sea ice forecasting on seasonal time scales. This thesis assesses the operational utility of the Canadian Seasonal to Interannual Prediction System (CanSIPS) for seasonal sea ice forecasting. This assessment consists of two separate studies. The first uses hindcasting to analyze the skill of two versions of CanSIPS, as well as an intermediate version, on the pan-Arctic as well as regional scales. This approach allows for an overall assessment of the system's skill in addition to providing insight with regards to the features in each version which improved that skill. This study finds that the use of a new initialization procedure for sea ice concentration and thickness improved forecast skill on the pan-Arctic scale as well as in the Central Arctic, Barents Sea, Laptev Sea, and Sea of Okhotsk. This study also shows that the substitution of one of the constituent models in the system improved forecast skill on the pan-Arctic scale as well as in the GIN, Barents, Kara, East Siberian, Chukchi, Bering, and Beaufort Seas. Overall, the new version of CanSIPS was found to be generally more skillful than previous versions. The second study conducts a potential predictability experiment on CanCM4, the constituent CGCM common to all versions of CanSIPS considered in this study. This study follows the methodology introduced by Bushuk et al. (2018) which allows for a more complete assessment of the dependency of potential predictability on initialization month than previous studies and for comparisons to be made between potential predictability and operational skill. This

analysis is again done on both the pan-Arctic and regional scale. The findings of this experiment show that CanCM4 has relatively low potential predictability relative to other models and explains results previously presented in a multi-model study by Day et al. (2016). Further, the characteristics of CanCM4's potential predictability share similarities with other models including greater predictability at longer lead times for winter target months than summer target months, greater predictability in the Atlantic sector than the Pacific sector, and the presence of the spring predictability barrier on the pan-Arctic scale as well as in several regions. The comparison of operational skill to potential predictability provides a general overview of the "skill gap" which may be closed with improvements in initialization procedures and model physics. This comparison does, however, come with some caveats due to differences in the statistical characteristics of the perfect model and the climate system it represents. Together, the operational skill assessment of different versions of CanSIPS and the potential predictability experiment conducted on one of its constituent models, CanCM4, demonstrate that while room for improvement exists, the recent development of this forecast system has clearly increased its operational utility as a seasonal sea ice forecasting tool.

Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my family. My wife Leta for being the partner and mom I needed her to be while completing this research and my daughter Corrie for always reminding me of the important things in life. Your love and support were the constants in two years filled with variables and I am forever grateful for having you in my life. I'd also like to thank my parents and siblings for their constant encouragement and especially my sister Sarah who is far and away the best scientist in the family.

I would further like to thank my supervisors, Dr. Michael Sigmond and Dr. Adam Monahan, for their support throughout this process. I could not have navigated this field in which I had little prior experience without your expertise, support, and patience. While I wrote this thesis myself, every word is the result of the knowledge and advice you provided to me and I am deeply appreciative of your understanding especially as I went through several life changes (and a global pandemic) during this program. I would also like to thank the third member of my committee, Dr. Bill Merryfield, and my external examiner, Dr. Marika Holland, for their contributions to my success and support in getting this thesis across the finish line. I would also like to express my gratitude to the Royal Canadian Navy for this opportunity and also thank my supervisor, Ulrich Suesser, and predecessor, Josée Belcourt, for their support during my program and in the transition into the Formation Oceanographer position at Maritime Forces Pacific.

This program, like any endeavour, would not have been near as enjoyable without the fantastic people I got to work with every day. It was a privilege to share a workplace with my colleagues in the UVic Climate Lab and broader School of Earth and Ocean Sciences. Specifically I would like to thank Liz Ramsey, my cubicle neighbour and partner-in-crime; Pat Duke for all the much-needed football nights; and the regulars at Friday beers. I hope our time together will outlive our time in grad school.

Finally, I would like to raise my hands with respect to the Lekwungen-speaking peoples on whose lands I have been privileged to live and grow — including during this research. I would also like to acknowledge the many Indigenous peoples of the Arctic, the original knowledge keepers of the land I studied in this thesis, who will continue to bear a far-outsized burden as a result of our changing climate.

DEDICATION

While I had the privilege of completing this research through the sponsorship of the Royal Canadian Navy, six sailors and airmen — my sister- and brothers-in-arms — were taken by the sea on April 29th, 2020 when their Cyclone helicopter crashed off the coast of Greece during a deployment with Standing NATO Maritime Group 2.

This work is dedicated to the crew and passengers of Stalker 22:

Captain Brenden Ian Macdonald
Pilot

Captain Kevin Hagen
Co-Pilot

Captain Maxime Miron-Morin
Tactical Coordinator

Master Corporal Matthew Cousins
Airborne Electronic Sensors Operator

Sub-Lieutenant Matthew Pyke
Naval Warfare Officer

Sub-Lieutenant Abigail Cowbrough
Marine Systems Engineer

"... and they can never lonely be for when they lived, they chose the sea."

Rest in Peace shipmates, Secure Flying Stations.

Chapter 1

Introduction

The past decades have seen a historically unprecedented decline in the extent of sea ice in the Arctic Ocean. The resulting increase in accessibility to the region has led to widespread interest in high latitude operations from numerous stakeholders in both the private and public sectors (Eicken, 2013). Given the prohibitive cost of ice breaking vessels, many of these new Arctic stakeholders will be increasingly dependent on sea ice forecasting to understand the seasonal navigability of Arctic waters and thus conduct timely and effective operational planning. Further, anthropogenic climate change has changed the nature of the environments which the Indigenous peoples of the North have forecast using traditional knowledge since time immemorial (Lee et al., 2020). The need for effective seasonal forecasting of sea ice for both original, modern, and prospective stakeholders will therefore only increase with the rapidly changing Arctic.

Efforts to produce skillful forecasts on seasonal timescales have been underway for over half a century. Early efforts utilized an anomaly persistence forecast (Namias, 1964). This approach is still used by some for seasonal sea ice prediction (Walsh et al., 2019) and is particularly effective for predicting phenomena whose time series feature substantial autocorrelation. Other seasonal forecasting strategies have relied on more advanced statistical methods (Wang et al., 2012; Kämäräinen et al., 2019). With the capabilities of modern computing, however, forecasting systems that use climate models implemented on supercomputers are increasingly used for seasonal sea ice forecasting. This thesis will assess the forecast skill with regards to seasonal sea ice extent of one such system.

1.1 The Canadian Seasonal to Interannual Prediction System (CanSIPS)

The Canadian Seasonal to Interannual Prediction System (CanSIPS) is a forecast system developed at Environment and Climate Change Canada. This system leverages the benefits of a multi-model forecasting system by combining two Coupled Global Climate Models (CGCMs) in order to provide dynamical seasonal and interannual forecasts. Of specific importance to the present study, CanSIPS was one of the first dynamical seasonal forecast systems with an interactive sea ice component which allowed its use as a seasonal sea ice forecasting tool. Two versions of CanSIPS have been developed and released, version 1 (Merryfield et al., 2013a) and version 2 (Lin et al., 2020), with version 2.1 set to be released in November 2021 (Lin and Muncaster, 2021; Merryfield, 2021).

There have been several previous analyses of the sea ice forecasting skill of various versions of CanSIPS. Sigmond et al. (2013) presented the first operational skill assessment of CanSIPS version 1 which found that the dynamical system's forecast skill was greater than that of an anomaly persistence forecast. This study also highlighted the important role of the long-term trend and the system's need for sea ice thickness initialization. Merryfield et al. (2013b) analyzed the detrended forecast skill of CanSIPSv1 and NCEP CFSv2 individually and combined (as a multi-model ensemble) and found that despite the generally higher skill of CFSv2, the multi-model system of the two models combined was more skillful in forecasting pan-Arctic sea ice area. Sigmond et al. (2016) was the first regionally-focused assessment of CanSIPS and considered prediction of advance and retreat dates of sea ice using CanSIPSv1. This study found on average, CanSIPSv1 is able to skillfully predict advance dates with a lead time of 5 months and retreat dates with a lead time of 3 months. Dirkson et al. (2017) developed three separate initialization procedures for CanCM3, one of the constituent models of CanSIPSv1, to provide synthetic "real-time" observations of sea ice thickness. One of these three methods, Statistical Model version 3 (SMv3), was integrated into CanSIPSv2. Dirkson et al. (2019) developed probabilistic sea ice concentration forecasts and Dirkson et al. (2021) developed probabilistic ice-free date and freeze-up date forecasts both using CanSIPSv1 with the SMv3 initialization procedure using different post-processing calibration methods. This thesis builds on this previous work to provide an assessment and analysis of the sea ice extent forecast skill of CanSIPS.

1.2 Objectives

This thesis will assess the operational utility of CanSIPS as a seasonal sea ice forecasting tool by addressing the following objectives. First, an operational skill assessment will be conducted on the three versions of CanSIPS. This analysis will provide a skill assessment of the current version on the pan-Arctic scale as well as in various Arctic regions and, through analyses of previous versions of CanSIPS, identify sources of skill improvement in the new versions. Second, a potential predictability experiment using one of the constituent models of CanSIPS, CanCM4, will be presented. This experiment aims to estimate the upper limit of predictability of seasonal sea ice as represented in CanCM4. Further analysis will illustrate features of predictability both on the pan-Arctic scale and in various Arctic regions. Finally, comparison of potential predictability to operational skill of CanCM4 can be used to assess any "skill gaps" between the operational model and a theoretical perfect model which could be addressed through improved model physics or initialization.

These proposed objectives will provide multiple novel contributions to the literature. This analysis will be the first comprehensive analysis of the seasonal sea ice extent forecast skill of CanSIPSv2. It will also be the first CanSIPS sea ice extent forecast skill assessment to compare skill between Arctic regions and to compare skill between different versions of CanSIPS. The potential predictability experiment represents the second use of the potential predictability methodology developed by Bushuk et al. (2018). It will further be the first perfect model experiment of a constituent CanSIPS model which analyzes the predictability of forecasts from a variety of initialization months and which compares the operational skill of the model to the potential predictability. These findings will both quantify the skill of CanSIPSv2 as an operational sea ice forecasting tool and inform the development of future versions of this forecasting system.

1.3 Terms and Abbreviations

The following terms and abbreviations are commonly used in the study of sea ice and will be used repeatedly in this thesis. They are included here with definitions for reference:

Pan-Arctic refers to all sea ice in the Northern Hemisphere. It should be noted that this deviates from some definitions of the boundaries of the Arctic and

especially the Arctic Ocean particularly around the lower latitudes of the GIN, Barents, Labrador, and Bering Seas as well as with the inclusion of the Sea of Okhotsk.

SIC refers to "Sea Ice Concentration" which is the fraction of a defined area covered by sea ice.

SIA refers to "Sea Ice Area" which is the area of the sea covered by sea ice. SIA is calculated from gridded SIC data by taking the sum of the products of each grid cell's area and its SIC.

SIE refers to "Sea Ice Extent" which is the area of all grid cells with an SIC greater than 0.15.

SIT refers to "Sea Ice Thickness" and is a vertical measure of sea ice including freeboard (ice above the waterline) and draught (ice below the waterline).

SIV refers to "Sea Ice Volume" which is the three-dimensional measure of sea ice and is thus pointwise the product of SIA and SIT.

1.4 Outline

This thesis is organized into four chapters (including the current chapter) with Chapters 2 and 3 representing two analyses of the forecast skill of CanSIPS with regards to sea ice as follows:

Chapter 2 analyzes the operational skill of different versions of CanSIPS in predicting SIE both on the pan-Arctic and regional scales.

Chapter 3 presents a potential predictability experiment conducted on one of the constituent models of CanSIPS, CanCM4, in order to quantify the upper bound of sea ice predictability on both the pan-Arctic and regional scales and compare the model's operational skill against this boundary.

Chapter 4 will provide a summary of these assessments of operational skill and potential predictability, consider avenues of further inquiry, and discuss the overall conclusions of this work.

Chapter 2

Assessment of the Seasonal Forecast Skill of Arctic Sea Ice in CanSIPS Versions 1 and 2

This chapter presents an assessment of the forecast skill of the Canadian Seasonal to Interannual Prediction System versions 1 and 2 in predicting Arctic sea ice extent on both the pan-Arctic and regional scales. Forecast skill is primarily quantified as the anomaly correlation coefficients of detrended time series for each target month of the year with initialization lead times of up to eleven months. The forecast skill is also compared to the skill of a much simpler anomaly persistence forecasting method. By comparison of the skills of CanSIPSv1 and CanSIPSv2 to that of an intermediate version of CanSIPS, CanSIPSv1b, we can attribute differences between CanSIPSv1 and CanSIPSv2 to two main sources. First, the new procedure for initializing sea ice thickness developed by Dirkson et al. (2017) 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. Second, the change in model combination from CanSIPSv1 to CanSIPSv2 (exchanging CanCM3 for GEM-NEMO) improves forecast skill in the Bering, Kara, Chukchi, Beaufort, East Siberian, Barents, and GIN Seas. In the more constrained waters of Hudson and Baffin Bay, as well as the Labrador Sea, there is limited and unsystematic improvement in forecasts of CanSIPSv2 as compared to CanSIPSv1. Overall, there is a net increase of forecast skill 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 which, in previous studies, have been associated with the spring predictability barrier (Day et al., 2014b).

2.1 Introduction

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 several seasonal sea ice forecasting systems (Bunzel et al., 2016; Krikken et al., 2016; Dirkson et al., 2017; Bushuk et al., 2017). The first hindcast analysis of the skill of CanSIPS in forecasting Arctic SIE was done by Sigmond et al. (2013) using CanSIPSv1. Other hindcast analyses of sea ice in CanSIPS with different objectives include Sigmond et al. (2016), Dirkson et al. (2017), and Dirkson et al. (2021). One known issue with CanSIPSv1 was the use of a fixed monthly SIT climatology for initialization which lacked both the decreasing trend in SIT and any interannual variability. Since Sigmond et al. (2016), a new version of CanSIPS, CanSIPSv2 (Lin et al., 2020), has been developed as well as an intermediate version, CanSIPSv1b (Dirkson et al., 2021), which was used during the transition between operational systems. This chapter will present an assessment of the forecast skill of each these three versions as well as notable changes in forecast skill which emerge in the transition from CanSIPSv1 to CanSIPSv1b and again from CanSIPSv1b to CanSIPSv2. This assessment will be done by analysing 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.

Certain features of dynamical sea ice forecasting systems which appear to be common among multiple systems have appeared in the past decade. These include the presence of a "melting season" or "spring" predictability barrier affecting forecasts initialized in the spring, a growing consensus regarding the importance of SIT initial conditions, and a move towards analyzing forecast skill at a regional scale in an attempt to work toward greater utility of these systems for end users. The following subsections will provide more detailed discussion of these topics.

2.1.1 The Spring Predictability Barrier

The term "spring predictability barrier" refers to a substantial decline in forecast skill of sea ice conditions when predicted at certain lead times, often associated with a sharp drop in skill between June- and May-initialized forecasts. 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 CGCMs by contrasting the superior skill in predicting September SIE of forecasts initialized in July with those initialized in May. This same barrier was also found in analyses of other models (in some cases retrospectively) including CanSIPSv1 (Sigmond et al., 2013), NCEP CFSv2 (Wang et al., 2013), and GFDL-FLOR (Bushuk et al., 2017).

The spring predictability barrier has also been identified in a perfect model experiment involving the GFDL-FLOR model (Bushuk et al., 2018) and a correlation analysis of SIV and 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 itself and thus may not be entirely eliminated by improvements in initialization procedures or model physics (Bushuk et al., 2018). Despite the inherent nature of this barrier, however, the difficulty in observing SIT, as compared to SIA, causes the SIT dimension to often be a source of error in the initialization of SIV in sea ice models. The findings of Bonan et al. (2019) indicate that SIV is an important predictor of summer sea ice conditions and thus can help to partially overcome the decline in skill associated with the spring predictability barrier.

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. Synoptic-scale atmospheric processes, which are mostly unpredictable on seasonal timescales, export variable amounts of sea ice into warmer waters increasing the unpredictability of SIC and SIT prior to the beginning of the melt season. Once melting begins, the seasonal export-driven anomaly is effectively determined and persists as well as amplifies through the melt season, partially due to positive ice-albedo feedbacks, resulting in a corresponding anomaly in September SIA. This change in predictability that occurs across the melt season as a result of this persisting anomaly is the spring predictability barrier. In this study,

Bushuk et al. (2020) further underscored the potential for increased predictability with improved initialization of SIT after the onset of melting.

2.1.2 The Importance of Sea Ice Thickness Initialization

Day et al. (2014a) identified the requirement of accurate SIT initialization as of particular importance in predicting September SIE. The perfect model forecast skill of a set of runs initialized with near-perfect initial conditions presented lower errors than those initialized with only a SIT climatology (such as is used for CanSIPSv1). The long-term decline in SIT is among many cryospheric indicators of climate change (Johannessen et al., 1999). Holland et al. (2019) found that SIT made 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 anomalies 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).

Despite substantial advances in remote sensing, accurate large-scale measurements of SIT sufficient to initialize CGCMs do not yet exist (Mu et al., 2018). Further, even with increased observational coverage of SIT, the lack of historical datasets prevent initialization of the hindcasts required for model skill assessment and calibration. Currently, the most commonly used substitute for SIT observations is the Pan-Arctic Ice Ocean Modeling and Assimilation System (PIOMAS) first described by Zhang and Rothrock (2003) which has proven effective when compared to various in situ observations obtained in the Atlantic sector of the Antarctic Ocean through drilling (Wadhams et al., 1987) and data from a 1993 submarine track in the Beaufort Sea and Central Arctic (Zhang and Rothrock, 2003). 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. (2017) developed a variety of statistical models to provide real-time SIT initialization fields by using PIOMAS SIC and SIT values up to one year prior to the initialization date to predict the required SIT. This study's Statistical Model version 3 (SMv3) was subsequently used to provide SIT initialization fields to CanSIPSv1b and CanSIPSv2, as described in Section 2.2.1.

The statistical models outlined in Dirkson et al. (2017) are estimated from a 15-year training period using known predictors $\mathbf{x}_m$ (either SIC from PIOMAS or sea level pressure from the ERA-Interim reanalysis (Dee, 2011)) and predictand $\mathbf{y}_m$ (SIT from PIOMAS) with the subscript m indexing month. The predictors are chosen due to their availability in real-time unlike SIT. These parameters are then applied to the real time predictor ($\mathbf{x}_m(t_e)$) to obtain a real time estimate of the predictand ($\mathbf{y}_m(t_e)$) for that month.

SMv3 was found to be the most skillful of the three statistical SIT models in reconstructing PIOMAS, especially for boreal summer predictions and in the Marginal Ice Zone (Dirkson et al., 2017). This statistical model adds the local detrended SIC anomaly ($\mathbf{x}_m^{\mathrm{SIC}}(t_e) - \hat{\mathbf{x}}_m^{\mathrm{SIC}}(t_e)$) multiplied by a proportionality constant α to a linear extrapolation of the local SIT $\hat{\mathbf{y}}_m(t_e)$:

$$\tilde{\mathbf{y}}_m(t_e) = \hat{\mathbf{y}}_m(t_e) + \alpha[\mathbf{x}_m^{\mathrm{SIC}}(t_e) - \hat{\mathbf{x}}_m^{\mathrm{SIC}}(t_e)] \tag{2.1}$$

This statistical model aimed to improve the forecast skill of various dynamical models by better representing the SIT field during initialization. Dirkson et al. (2017) showed that on the pan-Arctic scale, SMv3 used 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 its application to CanSIPSv1 and CanSIPSv2 (as outlined in Section 2.2.1), SMv3 is applied in the creation of initial conditions of each of the systems' two underlying models. The present analysis will expand the initial assessment in Dirkson et al. (2017), which assessed the skill of CanCM3 forecasts initialized in March, May, June, and September out to six months, by considering the forecast skill of CanSIPS, a multi-model system (versions of which include CanCM3), out to 11 months for every initialization month.

2.1.3 Regional Analyses

Few assessments of sea ice forecast skill on the regional scale for the entire Arctic have been reported in the literature. Previous regional analyses include Day et al. (2014b) which considered the predictability of sea ice in various regions and Sigmond et al. (2016) 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 five months and three months respectively). The assessment of EC-Earth v2.3 by

Krikken et al. (2016) represents the first analysis of regional sea ice forecast skill for all regions in the Arctic in the literature followed by the regional analysis of GFDL-FLOR by Bushuk et al. (2017) — the broad features of which will be compared to CanSIPS in Section 2.3.

Although there are few comprehensive assessments of operational SIE forecasts considering all Arctic regions, there is a significant body of research focusing on predicting sea ice in individual regions. In one such study, 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. 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. Additionally, 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 two years after their first occurrence. Finally, Bushuk et al. (2017) noted regional differences in the importance of ocean temperature initialization for winter regional skill and SIT initialization for summer regional skill. In particular, Bushuk et al. (2017) found similarities in target month/lead time structures of SIE forecast skill and the magnitude of the (negative) correlation between regional SIE and oceanic temperatures, especially in the Barents and Labrador Seas. Further, Bushuk et al. (2017) found a consistency between SIE forecast skill in the East Siberian, Chukchi, and Laptev Seas and the strength of SIE-SIT correlation in these regions. Regionally-focused analyses of drivers of sea ice conditions such as these can provide context to forecast skill analyses.

2.2 Experimental Design

This study presents an analysis of the seasonal SIE forecast skill of three different versions of CanSIPS. This section will describe 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. The

Model	CanCM3	CanCM4	GEM-NEMO
Atmosphere	31 levels, 2.8° × 2.8°	35 levels, 2.8° × 2.8°	79 levels, approx. 155 km
Ocean	40 levels, 1.4° × 0.94°	40 levels, 1.4° × 0.94°	50 levels, 1° × 1°
Sea Ice	cavitating fluid	cavitating fluid	CICE

Table 2.1: 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.

specifications of the components of the systems' constituent models are summarized in Table 2.1 while a summary of the initialization products used for each model can be found in Table 2.2.

2.2.1 The Models

The model physics and initialization of CanSIPSv1 are described in Merryfield et al. (2013a). This system combines ten ensembles each of the Third (Scinocca et al., 2008) and Fourth (Arora et al., 2011) generations of the Canadian Coupled Model (CanCM3 and CanCM4). The atmospheric components of these models are, respectively, the Third and Fourth Generation Atmosphere General Circulation Model and the ocean component of both models is the Fourth Generation Ocean General Circulation Model. Sea ice is represented as a cavitating fluid as described in Flato and Hibler (1992).

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. (2013a). The SIC in the CanSIPSv1 assimilation runs is relaxed towards the 1980-2012 Hadley Centre Sea Ice and Sea Surface Temperature (HadISST) version 1.1 SIC fields (Rayner, 2003) and the SIT is relaxed towards a seasonally-varying climatological mean due to a lack of time-varying and geographically-distributed SIT field observations. SIC after 2013 is nudged towards the Canadian Meteorological Centre (CMC) analysis used to initialize real-time predictions (Dirkson et al., 2021) and is derived from the assimilation of several datasets (Buehner et al., 2012, 2013, 2014).

In CanSIPSv1b, the CanSIPSv1 SIT initialization was changed using the SMv3 procedure outlined in Dirkson et al. (2017) which extrapolates trends from the local SIC and applies them to the local SIT as described in Section 2.1.2. Further, the SIC field in CanSIPSv1b is initialized using Had2CIS data (Lin et al., 2020) which is

comprised of observations of SIC from the HadISST version 2.2 dataset (Titchner and Rayner, 2014) combined with available digitized charts of the Canadian Ice Service (Tivy et al., 2011). Note that in this analysis, the CanSIPSv1 system with this initialization procedure is referred to as "CanSIPSv1b" in contrast to Dirkson et al. (2021) where it is referred to as "Mod-CanSIPS".

The model physics and initialization of the system's current operational version, CanSIPSv2, are described in Lin et al. (2020). 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), which has an evenly spaced 256 $\times$ 128 grid, coupled with the NEMO ocean model (Gurvan et al., 2019). The resolutions of these models are specified in Table 2.1. Sea ice is represented in GEM-NEMO using the Community of Ice Code (CICE) model (Hunke et al., 2010) which has five thickness categories. The initial SIC and SIT conditions are nudged to Had2CIS and ORAP5 (Zuo et al., 2017) conditions, respectively, from 1980-2010. From 2011-2018, both the initial SIC and SIT conditions are from the Canadian Centre for Meteorological and Environmental Prediction Global Ice Ocean Prediction System (CCMEP GIOPS) analysis (Smith et al., 2016). Both GEM and NEMO run on finer resolutions than CanCM3 and CanCM4.

2.2.2 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 in Section 2.2.1, was for this analysis interpolated to a $1^\circ \times 1^\circ$ grid to match the regional mask as described in Section 2.2.4 and then used to calculate both pan-Arctic and regional mean SIE. The SIC trends in HadISST version 2.2 were found to be more accurate than those underestimated in HadISST version 1.1 (Sigmond et al., 2013) and thus provide a better assessment of SIC (Lin et al., 2020).

2.2.3 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 (τ) as the Anomaly Correlation Coefficient (ACC)

System	Model	SIC initialization	SIT initialization	Reference
CanSIPSv1	CanCM3	HadISSTv1.1 (1980-2012) CMC analysis (2013-2018)	fixed monthly climatology	Merryfield et al. (2013a)
	CanCM4	HadISSTv1.1 (1980-2012) CMC analysis (2013-2018)	fixed monthly climatology	
CanSIPSv1b	CanCM3	Had2CIS (1980-2012) CMC analysis (2013-2018)	SMv3 (Dirkson et al., 2017)	Dirkson et al. (2021)
	CanCM4	Had2CIS (1980-2012) CMC analysis (2013-2018)	SMv3	
CanSIPSv2	CanCM4	Had2CIS (1980-2012) CMC analysis (2013-2018)	SMv3	Lin et al. (2020)
	GEM-NEMO	Had2CIS (1980-2010) CCMEP GIOPS analysis (2011-2018)	ORAP5 (1980-2010) CCMEP GIOPS analysis (2011-2018)	

Table 2.2: Products used to initialize SIC and SIT fields in each of the consituent models of the CanSIPS versions.

between the time series of the ensemble means of the modelled predictions (p) and the time series of the target month from the Had2CIS observations dataset (o) with overbars denoting time averaging and angled brackets denoting ensemble averaging:

$$ACC(\tau) = \frac{\Sigma_{j=1}^{K}\left(\langle p_j(\tau)\rangle - \overline{p(\tau)}\right)(o_j - \overline{o})}{\sqrt{\Sigma_{j=1}^{K}\left(\langle p_j(\tau)\rangle - \overline{p(\tau)}\right)^2}\sqrt{\Sigma_{j=1}^{K}(o_j - \overline{o})^2}} \tag{2.2}$$

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. As this trend is also present in the initial conditions of CanSIPS, the forecast skill of the system as represented by the ACC may be dominated by this trend and not be strongly affected by how well the forecast system captures the interannual variability of SIE (Sigmond et al., 2013). The present study will consider the forecast skill of the operational systems independent of trend and thus both the prediction and observation time series were subjected to linear detrending before computing the ACC. While many previous studies implement detrending using the entire time series (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 can not be used to remove the trend). In order 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. A further benefit of this approach is that it allows for the trend to evolve over time with the increasing rate of sea ice decline. The first three years have only the past mean removed — the linear trend is assumed to be zero. This method was first presented in Bushuk et al. (2017) and fully described in

Bushuk et al. (2018). Some minor differences were found between this method and traditional detrending; a representative example is provided as Figure A.1.

All statistical significance tests provided in this study are computed using a bootstrapping methodology. This procedure creates a distribution by randomly sampling with replacement the forecast and model outputs 1000 times and calculating the ACC for each random pair. 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. The null hypotheses in this study are, for the skill assessments of each version, that there is zero or smaller correlation between the forecasts and observations if the ACC is positive or that there is zero or greater correlation between the forecasts and observations if the ACC is negative. Further, in considering the skill differences between versions, the null hypothesis is that the two versions have equal correlation between forecasts and observations.

An anomaly persistence forecast is a long-used seasonal prediction methodology (Namias, 1964). It is created by applying the anomaly of a variable from the present period's climatological mean to the climatological mean of the period being forecasted. Due to the trivial effort required for such a forecast with modern computers, 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 Figure 2.1, this method can have both high and significant skill (relative to a null hypothesis of zero ACC) in seasonal sea ice prediction due to the substantial autocorrelation present in these time series. As with CGCM predictions, much of this skill is derived from the downward trend in sea ice. The right plot of Figure 2.1 demonstrates the drop in skill of a persistence forecast with trend removed and also how initialization months with persistent anomalies can be seen by following the cells diagonally up and to the right. For example, in the case of a February-initialized forecast, the persistent anomaly can be seen beginning at February lead 0 and continuing to March lead 1, April lead 2, etc.

2.2.4 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

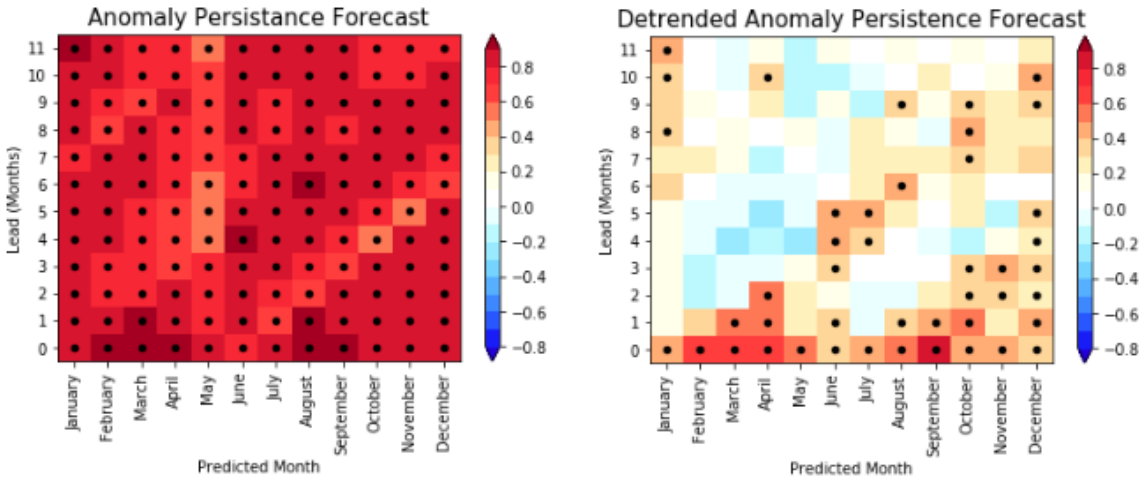

Figure 2.1: Pan-Arctic anomaly persistence forecasts using Had2CIS observations including the trend (left) and detrended (right). Dots represent forecasts whose skill is significant at the 95% confidence level (relative to a null hypothesis of zero ACC).

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 defined in this analysis (Figure 2.2) are the same as those defined in Day et al. (2014b) and also used in a number of other studies (eg. Bushuk et al. (2018, 2020); Bonan et al. (2019)). These regions are defined on a $1^\circ \times 1^\circ$ grid. Some interpolations of the regional mask and model outputs were conducted to create a common grid for the mask, model outputs, and observations used in the analysis.

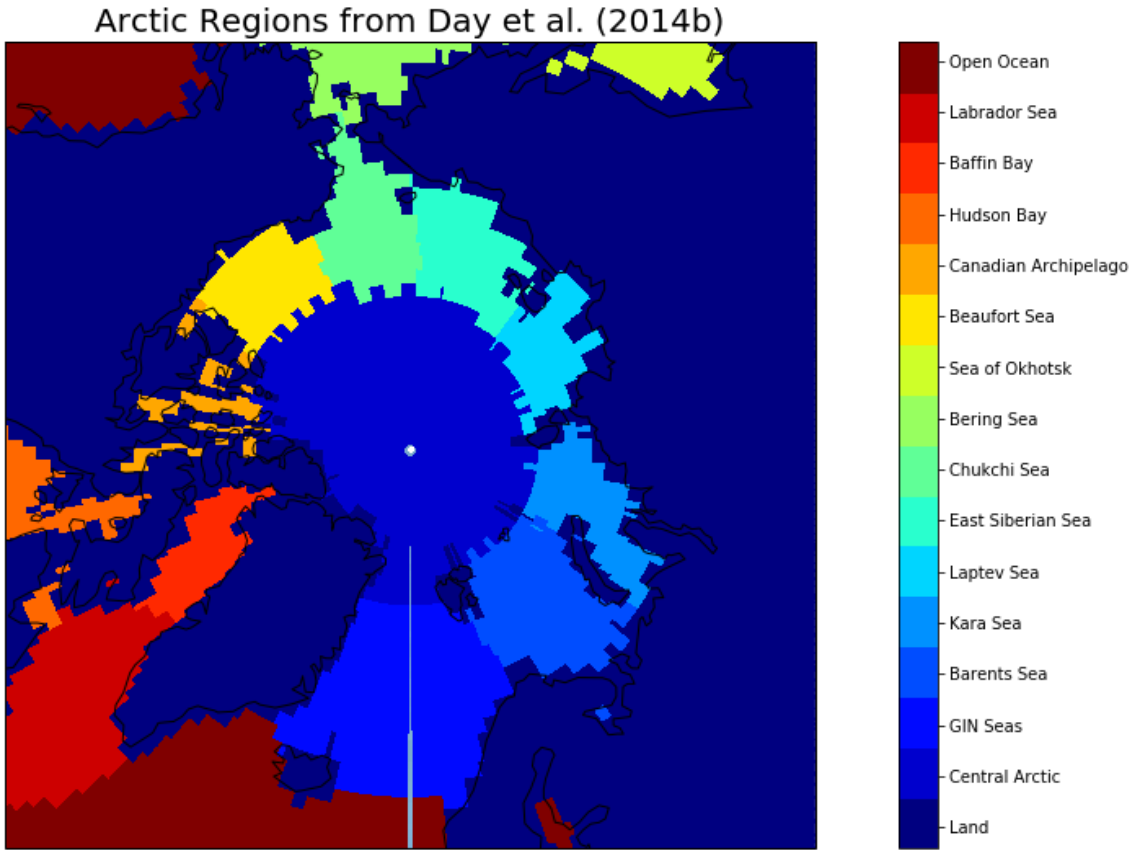

Figure 2.2: Arctic regions considered in this study.

2.3 Results

This section presents the results of the analysis of SIE forecast skills of CanSIPSv1, CanSIPSv1b, and CanSIPSv2 on the pan-Arctic and regional scales. First, a summary of the pan-Arctic and regional forecast skill of CanSIPSv1 will be presented in order to provide a baseline of the dynamical system's forecast skill. The improvements seen in the system's forecast skill at the pan-Arctic and regional scales as a result of the improved initialization procedure first implemented in CanSIPSv1b and the new model combination used for CanSIPSv2 are then discussed.

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 September 1st while lead 1 represents a September forecast that was initialized on August 1st. Further, the skill of the forecast initialized in a certain month can be followed diagonally such as for September-initialized forecasts which would 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 times series of the forecast and observations from Had2CIS as described in Section 2.2. Forecast skill values determined to be statistically significant at the 95% confidence level through bootstrapping are denoted with markers. Triangles indicate statistically significant forecast skill (relative to a null hypothesis of zero ACC) which is also higher than the corresponding anomaly persistence forecast, while dots represent significant skill which is not more skillful than persistence (Section 2.2.3). All results are presented for detrended forecasts. Analyses of the ACCs between forecasts and observations for target months where regions are nearly open waters or almost completely covered by sea ice are excluded in part to focus this study on prediction of sea ice variability. Forecast skill for a target month is therefore not calculated if the observed interannual standard deviation of the region's SIE for that month is less than 0.8% of the area of the region. It should be noted that a forecast of completely covered or completely ice-free waters during these months would indeed be a highly skillful forecast, although the ACC would be undefined if this occurs for all years.

A complete set of plots showing the skill of each of the three CanSIPS versions for each region is presented in the Appendix along with plots illustrating the differences in skill between versions (CanSIPSv1b-CanSIPSv1, CanSIPSv2-CanSIPSv1b, and CanSIPSv2-CanSIPSv1) (Figures A.2-A.14). 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 sufficent resolution to properly represent this region.

2.3.1 CanSIPSv1 prediction skill

The detrended skill of CanSIPSv1 on the pan-Arctic and regional scales is presented in Figure 2.3. Without exception, the forecasts with trend (not shown) are more skillful than the detrended forecasts. Pan-Arctic SIE was previously considered by Sigmond et al. (2013) for a shorter hindcast period (1979-2010 in contrast to the 1980-2018 in the present analysis) and was obtained using a different detrending technique (see Section 2.2.3). Three notable features can be seen in several regions: two lobes of elevated forecast skill for target months that are roughly coincident with the sea ice melt and growth seasons, a spring predictability barrier visible as a pronounced drop in skill from forecasts initialized in June to those initialized in earlier months, and notably high skill in the seas connected to the North Atlantic, with significant skill for lead times as long as 11 months in the GIN, Barents, and Labrador Seas.

In the pan-Arctic forecast of CanSIPSv1 (Figure 2.3), we see the presence of two seasonal lobes of skill in the target months of April and October, which were also noted in Bushuk et al. (2017), corresponding approximately with the beginning of the ice melt and growth seasons, respectively. These lobes are also visible in various regions although for the majority of regions only one lobe falls within a period of partial ice coverage. The first skill lobes for predictions of target months in early spring is most prominent at the beginning of the retreat in the Sea of Okhotsk in April and can also be seen to a lesser extent in the Bering Sea in April. It is not discernible in the Barents and Labrador Seas due to the presence of substantial skill for forecasts of January through to May and it is notably interrupted in the GIN Seas by low skill of October-, November-, and December-initialized forecasts. The second lobe, associated with the beginning of the season's sea ice growth, is visible in all regions which regularly have sea ice coverage in October. In most regions, the October skill is limited to approximately four months because of the spring predictability barrier, with skill generally much smaller for May- (lead 5) than June-initialized (lead 4) forecasts.

The spring predictability barrier described by Day et al. (2014b) is evident in most regions with summer sea ice coverage with Hudson and Baffin Bay being the

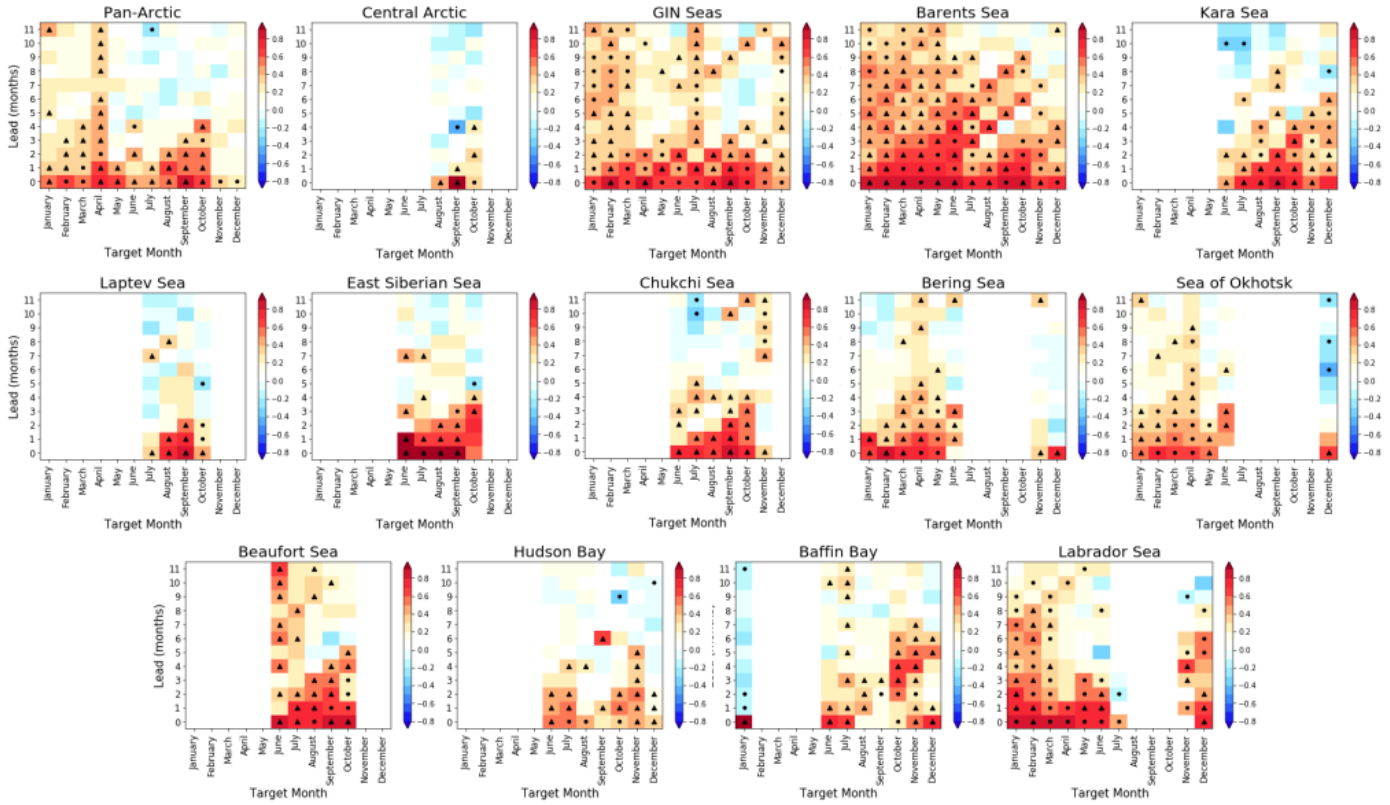

Figure 2.3: CanSIPSv1 forecast skill of detrended sea ice extent shown as functions of target month and lead time for 13 different regions, for a period spanning 1980 to 2018. Markers indicate statistical significance of the skill at the 95% confidence level, such that a triangle (dot) indicates a statistically significant forecast with greater (less) skill than an anomaly persistence forecast. The observations used to assess this skill are from the Had2CIS dataset and forecast skill is masked out for target months if the observed interannual standard deviation of the region's SIE for that month is less than 0.8% of the area of the region.

two major exceptions. It can often be seen by noting the substantially lower skill of May-initialized forecasts (beginning at May lead 0, then diagonally to June lead 1, July lead 2, etc.) compared to the June-initialized forecasts (beginning at June lead 0, then diagonally to July lead 1, August lead 2, etc.) as seen in the pan-Arctic subplot in Figure 2.3. On the pan-Arctic scale, significant skill greater than persistence can be seen out to October from forecasts initialized in June (four months lead time), but not for any late summer/early autumn forecasts initialized earlier. On the regional scale, this same pattern is seen with the May-initialized forecast providing no significant autumn skill (September, October, and November) in the GIN, Barents, Kara, Laptev, East Siberian, and Chukchi Seas. With the notable exception of the

April-initialized forecast in the Barents Sea, the June-initialized forecast is generally the earliest significantly skillful forecast of any region's autumn SIE. In the Beaufort Sea, the barrier occurs prior to the May-initialized forecast. While Baffin Bay and Hudson Bay show less evidence of the barrier (possibly due to being almost ice-free at the peak of melt season in autumn) the fact that the last skillful November forecasts are from the May and June initializations respectively suggests the barrier is present in these regions as well.

The CanSIPSv1 forecast has the lowest level of skill in the Central Arctic with only five forecasts having significant skill higher than the persistence forecast. Of the seasonal ice regions, the Laptev Sea (north of Siberia) represents the system's lowest skill, with regional skill increasing away toward the west (from the Kara and Barents Seas) and the east (through the East Siberian, Chukchi, and Beaufort Seas). The seas most closely connected to the North Atlantic — the GIN, Barents, and Labrador Seas — all show especially high skill in the winter months (particularly for the target months of February and March where all three regional analyses have skillful forecasts at lead 9 or greater). This elevated predictability may be a result of effective representation of regionally-important processes such as the Atlantic meridional overturning circulation (Krikken et al., 2016). This finding of high skill near the North Atlantic is also consistent with Bushuk et al. (2017) who found similar skill in these regions in the winter, attributing the elevated skill to the SIE predictability provided by ocean temperature anomalies.

There is some evidence of skill reemergence across the September minimum in more constrained 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 can be seen in the June-initialized forecasts in Hudson Bay (June lead 0 skill reemerges as November lead 5), Baffin Bay (June lead 0 skill reemerges as October lead 4), and the Labrador Sea (June lead 0 skill reemerges as December lead 6). These regional reemergence features were also 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 phenomena has also been seen in many other models (Day et al., 2014b; Guemas et al., 2016). 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. There is little evidence of the spring predictability barrier in these constrained water regions with the Labrador Sea having near ice-free water in October, May- and June-initialized forecasts having little skill in October in Hudson Bay, and moderate-to-high skill for the same target month in Baffin Bay.

2.3.2 Improvements in pan-Arctic SIE Forecast Skill

The forecast skill of pan-Arctic SIE of each version of CanSIPS (with trend) is shown in Figure 2.4, along with plots of differences in forecast skill between systems. CanSIPSv2, shows a marked improvement in overall forecast skill (Figure 2.4f) with statistically significant improvements seen in target months at lead times that correspond with the spring predictability barrier. Further, both CanSIPSv1b and CanSIPSv2 outperform persistence more often than CanSIPSv1 — especially in the winter months. In considering the step-wise evolution of CanSIPS, these improvements are shown to be due to the better SIT initial conditions provided by the statistical method of Dirkson et al. (2017) as well as the improved SIC initial conditions in Had2CIS observations as compared to HadISST version 1.1 (due to the better representation of trends). CanSIPSv2 shows little significant improvement from CanSIPSv1b on the pan-Arctic spatial scale (as shown in Figure 2.4e).

An analysis of detrended forecast skill (Figure 2.5), further underscores the improvements in skill of forecasts on the pan-Arctic scale that result from improving the SIT initialization. While an overall decrease in skill is clear once the trend is removed, relative to skill estimates including the trend, significant improvements for target months and lead times previously associated with the spring predictability barrier are evident (Figure 2.5d). A notable difference in CanSIPSv1b relative to CanSIPSv1 is improved skill of August and September forecasts, representing an extension of the maximum skillful lead times from 3 and 4 months to lead times of 5 and 7 months respectively. In particular, the improvement seen in May-initialized forecasts from CanCM3 using SMv3 in Dirkson et al. (2017) can also be seen in CanSIPSv1b (Figure 2.5b) at the pan-Arctic scale.

Several aspects of the forecast skill of CanSIPSv2 show improvement over CanSIPSv1. The first such aspect is the more frequent statistical significance of Can-

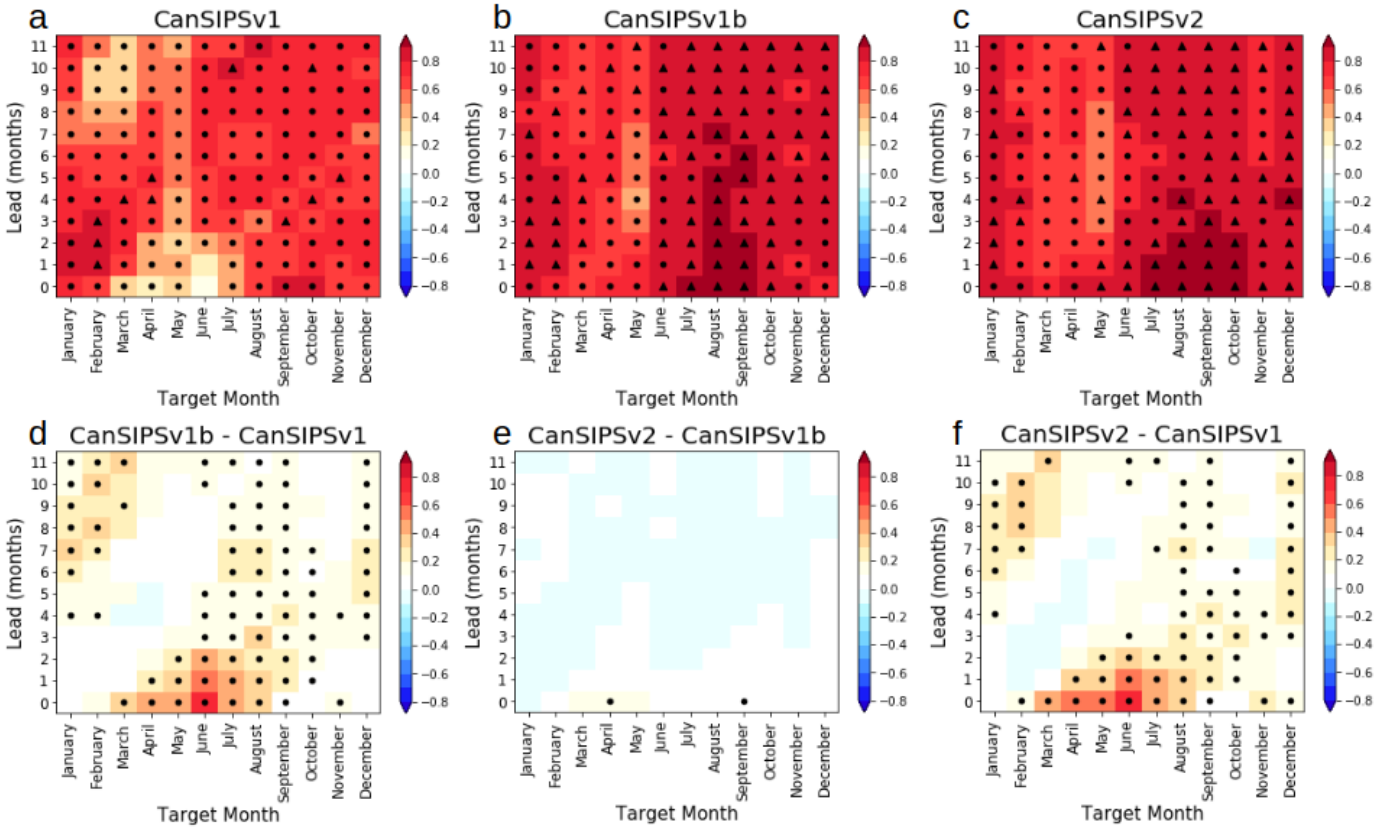

Figure 2.4: Comparison of the pan-Arctic forecast skills of sea ice extent for three versions of the CanSIPS model as a function of target month and lead time for a period spanning 1980 to 2018. Markers indicate statistical significance of the skill (or difference in skill) at the 95% confidence level, such that a triangle (dot) indicates a statistically significant forecast with greater (less) skill than an anomaly persistence forecast. The observations used to assess this skill are from the Had2CIS dataset.

SIPSv2 relative to an anomaly persistence forecast. Forecasts without detrending made by CanSIPSv1 at lead times up to eleven months were only significant and of higher skill than an anomaly persistence forecast for 8% of target month/lead pairs. This fraction increased to 52% with CanSIPSv2 — driven largely by improvements in forecasts for the months of July to December. Forecasts with detrending made by CanSIPSv1 were significant and of higher skill than an anomaly persistence forecast for 26% of target month/lead pairs which increased to 45% due mostly to improvements in the winter target months. 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 four months of a May-initialized forecast. In contrast, CanSIPSv2 provides significantly skillful forecasts through to September (Figure 2.5c). For April-

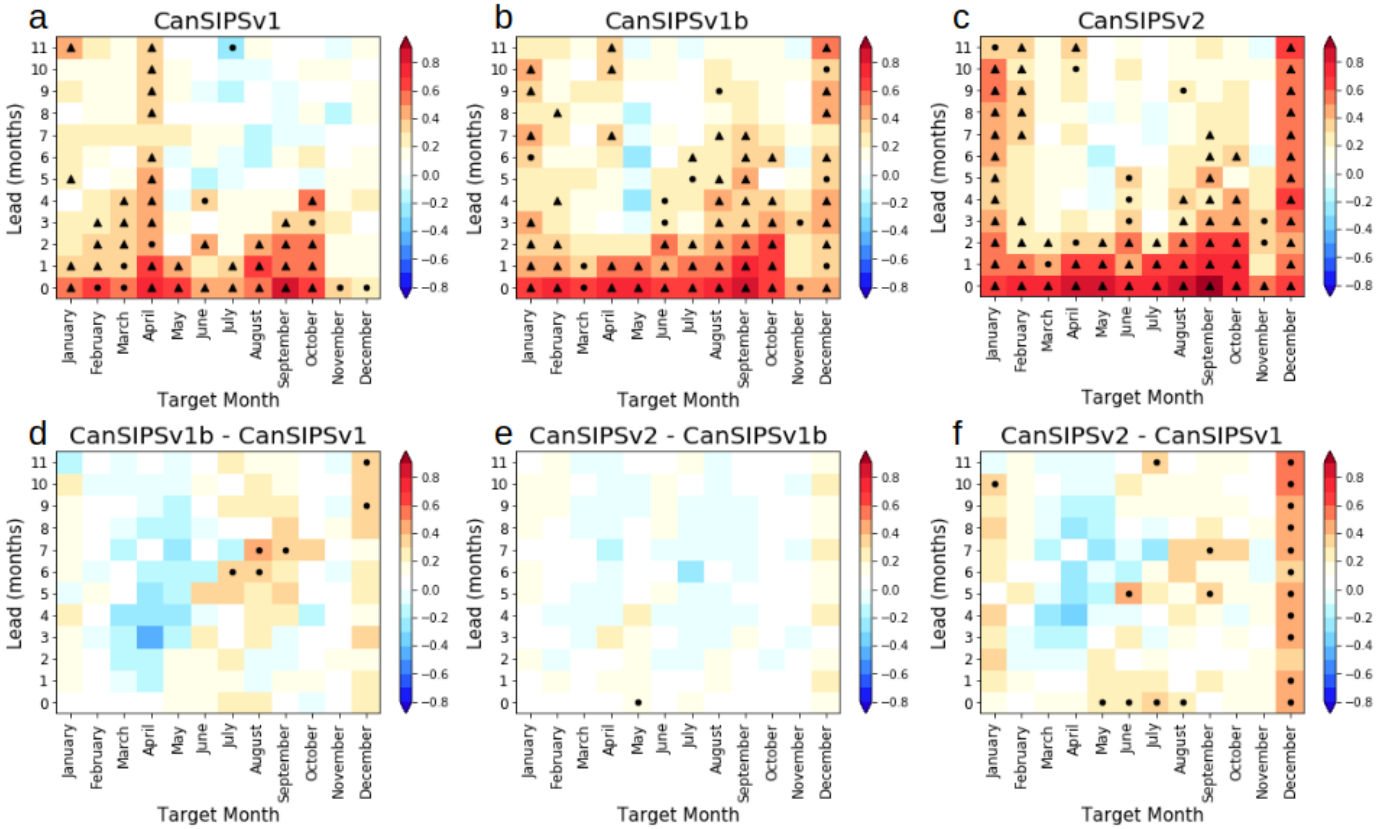

Figure 2.5: As for Figure 2.4 except for skill after detrending.

initialized forecasts, CanSIPSv2 provides significantly skillful forecasts through to October (with the exception of July) as shown in Figure 2.5c. Also noteworthy are the skills for forecasts of the target months December and January out to a lead time of eleven months. When considered together with a large decline of the skill of forecasts with a target month of April, it appears that the lobe of skill near the melt season onset in CanSIPSv1 (Section 2.3.1) migrates into the middle of winter in CanSIPSv2.

On the pan-Arctic scale, these gains in forecast skill appear to be almost entirely associated with the improved initialization first implemented in CanSIPSv1b, as highlighted in Figure 2.5e, although the totality of improvements present in CanSIPSv2 do contribute to significant skill improvements in the target months of January and especially December (Figure 2.5f). When compared to CanSIPSv1, whose skill for September to December forecasts disappears for May and earlier initializations, the success of the initialization method detailed in Dirkson et al. (2017) and implemented in CanSIPSv1b further underscores the importance of SIT for predictions through the

melting season (Chevallier and Salas-Mélia, 2012; Sigmond et al., 2013; Day et al., 2014a). It is also notable that the improvements to April- and May-initialized forecasts appear to come at the slight expense of spring forecasts initialized during or immediately before the Arctic sea ice growth season (such as a marginal though not significant decrease in skill for March and April for all lead times greater than one as seen in Figures 2.5d and 2.5f).

2.3.3 Improvements in Regional SIE Forecast Skill

Of greater interest to operational forecasting is that forecasts are skillful on regional scales. In the North Atlantic, forecast skill is generally higher overall and the improvements to the system are readily apparent. Conversely, in the regions nearer to the Pacific, forecasts are less skillful and in many cases some changes to the systems, particularly the new initialization procedure from Dirkson et al. (2017), have a modest or negative effect on skill. The more constrained waters of Hudson and Baffin Bay, as well as the Labrador Sea, have their own skill patterns with particularly pronounced skill reemergence but with little notable change in forecast skill between versions. Finally, the forecast skill in the Central Arctic and Sea of Okhotsk see unique but significant improvements from CanSIPSv1 to CanSIPSv2 due to the distinct geography of these regions.

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 (CanSIPSv1b-CanSIPSv1, CanSIPSv2-CanSIPSv1b, and CanSIPSv2-CanSIPSv1) in the Appendix (Figures A2-A14).

CanSIPSv1 already provided highly skillful forecasts in the North Atlantic (Figure 2.3). Previous studies have suggested that high skill in this region 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). With regards to the spring predictability barrier, we see substantial improvement in the Barents Sea (Figure 2.6, second row) between CanSIPSv1 and CanSIPSv2 as the number of May-initialized forecasts with significant skill increases from lead 2 to lead 10. In contrast, almost no change is found for the GIN Seas (Figure 2.6, first row) where the number of May-initialized forecasts with significant skill increases only from lead 1 to lead 3. This result is consistent with

the effects of improving SIT initialization with SMv3 in CanCM3 to the previous initialization method by (Dirkson et al., 2017), but it is clear from the low skill of the September lead 4 and October lead 5 forecasts for CanSIPSv1b in the Barents

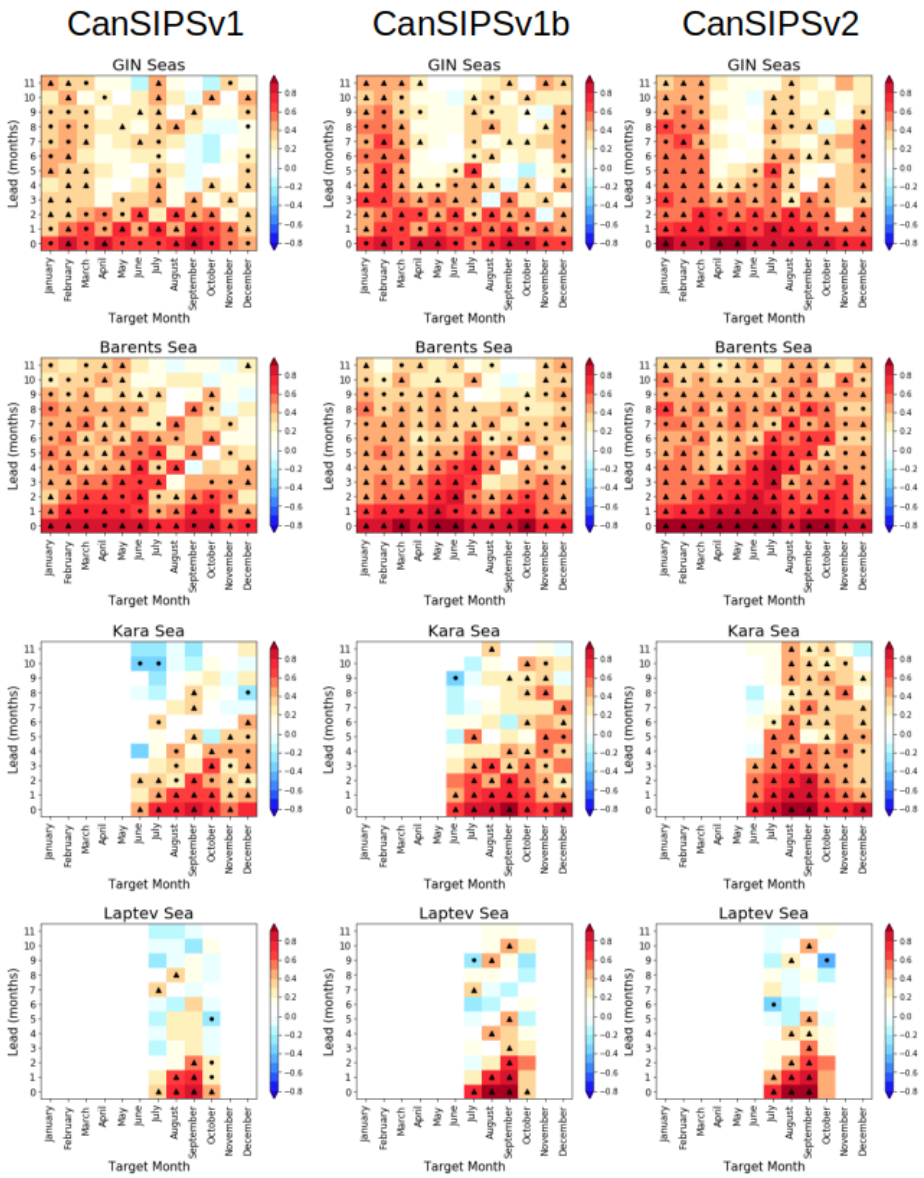

Figure 2.6: Forecast skill of detrended SIE shown as a function of target month and lead time shown for CanSIPSv1, CanSIPSv1b, and CanSIPSv2 over a period spanning 1980 to 2018 for the GIN, Barents, Kara, and Laptev Seas. Symbols are as in Figure 2.4 and the observations used to assess this skill are from the Had2CIS dataset. Forecast skill is not calculated for target months if the observed interannual standard deviation of the region's SIE for that month is less than 0.8% of the area of the region.

Sea that the change in constituent models is also important in addressing the spring predictability barrier in this region. It is noteworthy that the April-initialized forecast for the Barents Seas shows significant skill until October which is inconsistent with the spring predictability barrier. 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 (Figure 2.6, last row) shows a smaller but still significant improvement in late summer forecasts at short lead times when the new initialization method is applied suggesting that proper SIT initialization is important in this region. This result is consistent with the finding of Bushuk et al. (2017) that the forecast skill of the GFDL-FLOR model in the Laptev Sea matches closely with the correlation between observed SIE and the forecast's SIT initial conditions. The Kara Sea (Figure 2.6, third row) sees particular improvement as result of both the improved initialization (similar to the improvement seen in Dirkson et al. (2017) with regards to CanCM3) and the change of constituent models in the target months of August, September, and October which all have significant skill at all lead times up to eleven months in CanSIPSv2.

The Arctic seas nearest to the North Pacific demonstrate little improvement from the new initialization procedure, in contrast to the findings of Bushuk et al. (2017) that summer SIE prediction skill in these regions results from accurate SIT initialization (Figure 2.7). In fact, the improved initialization is associated with a slight decrease in skill in the East Siberian, Chukchi, Bering, and Beaufort Seas and these regions represent the majority of the decline seen in the study. It should first be noted that initial conditions provided by Dirkson et al. (2017) are improved but imperfect. Specifically in regions nearer to the Pacific, the initial conditions provided by SMv3 are, particularly in the winter, less accurate when compared against PIOMAS. In addition to potential errors in the initialization itself, these regions 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). 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. These circumstances potentially contribute to the fact that the spring predictability barrier is inherent in the climate system of these regions, as will be further demonstrated in Chapter 3. 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 and Chukchi Seas.

The regions of more constrained waters connected to the North Atlantic (Figure 2.8) show less skill in all forecast system versions than the more open GIN and Barents Seas (Figure 2.6). In fact, there is little improvement seen in Hudson and Baffin Bay as well as the Labrador from CanSIPSv1 to CanSIPSv2 aside from the slightly more skillful forecasts of January, February, and March at lead times of seven to eleven

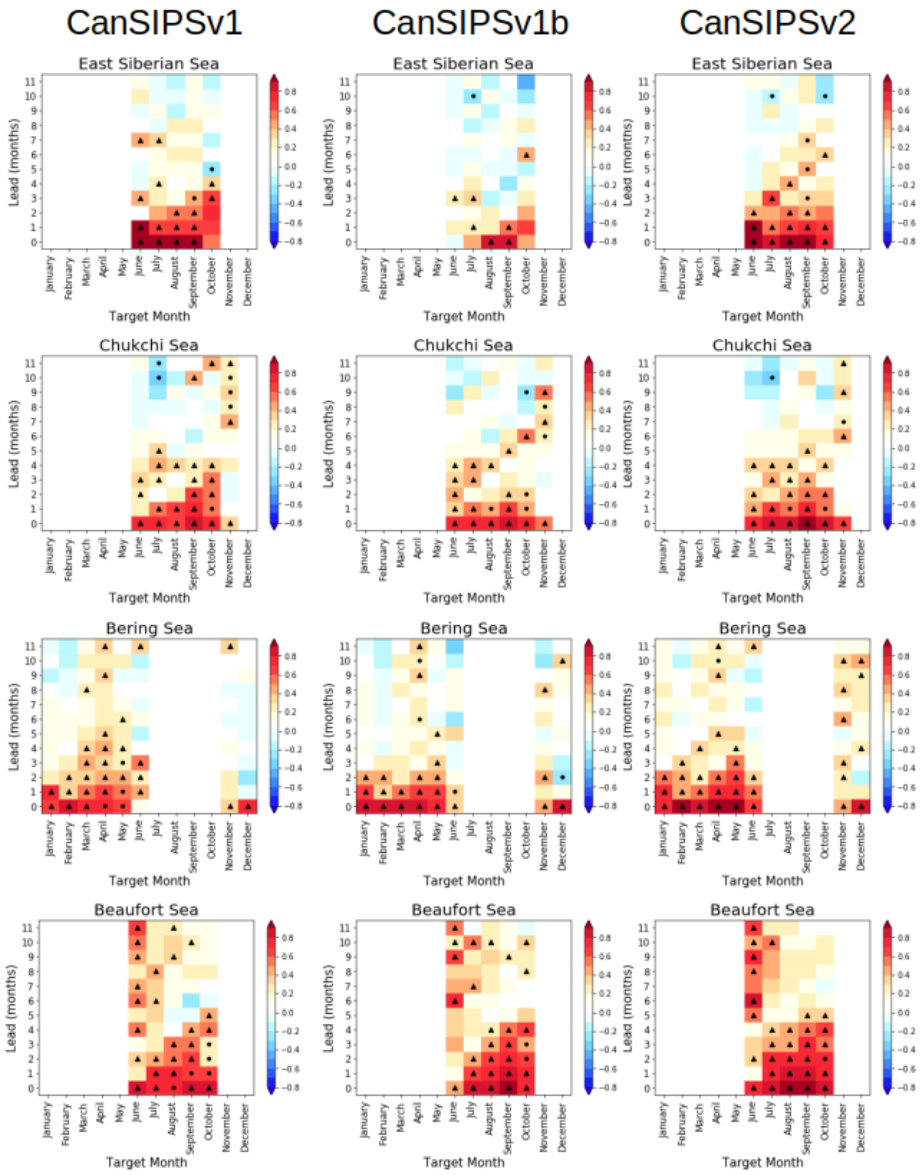

Figure 2.7: As in Figure 2.6 for the East Siberian, Chukchi, Bering, and Beaufort Seas.

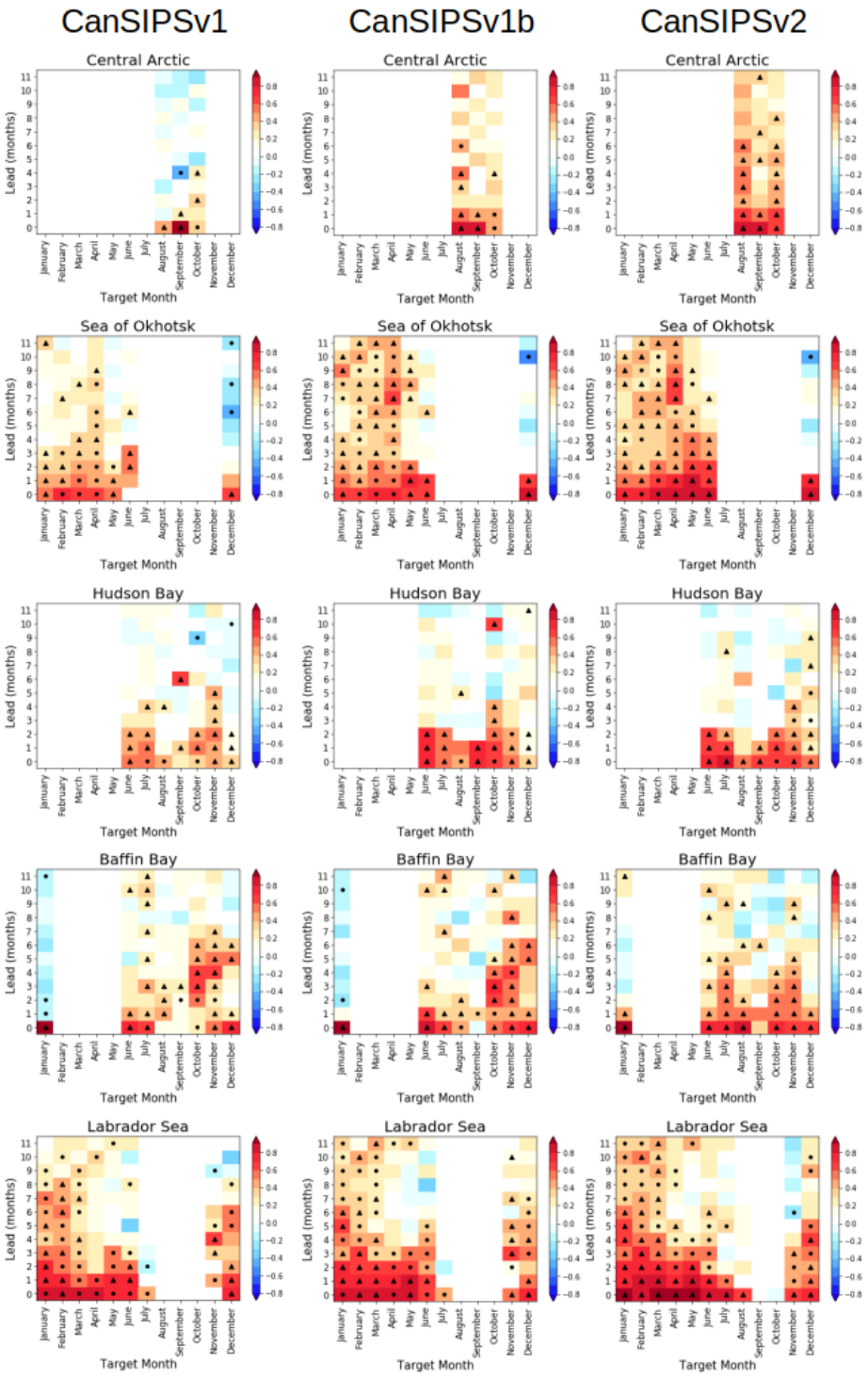

Figure 2.8: As in Figure 2.6 for the Central Arctic, Sea of Okhotsk, Hudson Bay, Baffin Bay, and the Labrador Sea.

months. Skill reemergence is still a feature in the skill of CanSIPSv1b and CanSIPS v2 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) reemerge in the November (lead 4) forecasts in Hudson and Baffin Bay while July-initialized forecast skill reemerges in December (lead 5) in the Labrador Sea. This reemergence is not substantially different from in CanSIPSv1.

The forecast skill in the Central Arctic (Figure 2.8, first row) appears to improve almost entirely due to the new initialization procedure with little improvement from the substitution of GEM-NEMO. Forecasts in this region improve from not having any significant skill at lead times greater than two 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). This significant improvement from CanSIPSv1 to CanSIPSv1b in the Central Arctic was similarly seen by Dirkson et al. (2017) when assessing the SIC forecast skill of CanCM3 in the region before and after using the new initialization procedure (SMv3).

Finally, substantial improvement in the winter and spring target months is seen in the Sea of Okhotsk (Figure 2.8, 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.

2.4 Conclusion

This chapter presented an assessment of the forecast skill of three CanSIPS forecast systems with regards to SIE with particular attention to the effect of improvements made in CanSIPSv1b and CanSIPSv2. It is clear from the results that substantial improvements in forecast skill were made through both the implementation of a new SIC and SIT initialization procedure as well as through changing one of the system's constituent models. The importance of proper initialization of SIT is underscored by the improvements CanSIPSv1b makes in the target months of August, September, and October with the lead times of forecasts with significant skill increasing from lead 2, 3, and 4 (June-initialized forecasts) to lead 4 (April-initialized forecast), lead 7 (February initialized forecast), and lead 7 (March-initialized forecast) on the pan-Arctic scale. Regionally, similar improvements resulting from improved SIT initialization in these target months are seen in the Central Arctic (significant skill in October seen for lead

5 vice lead 2) and the Laptev Sea (significant skill in September seen for lead 5 vice lead 2). Substantial improvement from the new initialization is also seen in the Sea of Okhotsk with little significant skill in CanSIPSv1 for forecasts of the target months of January through May initialized before November improving to show significant skill for all forecasts initialized from July onward in CanSIPSv2. The substitution of CanCM3 by the GEM-NEMO model in CanSIPSv2 appears to provide the majority of skill improvements in the seas of 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. Finally, more constrained waters such as Hudson and Baffin Bays as well as the Labrador Sea showed little improvement in the new versions with the majority of their forecast skill still derived from reemergence across the autumn SIE minimum — potentially the result of sea ice anomalies being captured in sea surface temperatures after melting. In summary, this analysis found CanSIPSv2 to be significantly more skillful than its predecessors on both the pan-Arctic and regional scales due to effective representation of the initial condition of sea ice thickness and a more skillful CGCM combination.

Chapter 3

Potential Predictability of Arctic Sea Ice in the Fourth Generation Canadian Coupled Model (CanCM4)

This chapter presents an assessment of the potential predictability of Arctic sea ice as estimated by CanCM4 — the model common to both versions of CanSIPS. The methodology of Bushuk et al. (2018) is used for this experiment except for a modification to the selection method of initial conditions to avoid systematic biases seen in their results. Three-year perfect model experiments are conducted, consisting of twelve ensemble runs each for six initialization months in six initialization years branched off a control run. Predictability is then calculated on the pan-Arctic and regional scales by designating individual ensemble members in turn as "the truth" and predicting these synthetic observations using the ensemble mean of the remaining ensemble members with the same initialization month and year. The perfect model skill, or predictability, is then analyzed and compared to the operational model's skill in order to assess the operational "skill gap" which may be closed through improved model physics or initialization. The skill of the CanCM4 perfect and operational models is then compared to the results presented for GFDL-FLOR in Bushuk et al. (2018). We find that CanCM4's estimate of potential predictability is lower than those of other GCMs, as had been suggested in Day et al. (2016). The potential predictability in CanCM4 is also lower than GFDL-FLOR (Bushuk et al., 2018) on

the pan-Arctic scale and in every region except the Labrador Sea. The results also show that predictability characteristics found in previous perfect model experiments are also found in CanCM4 such as greater predictability at longer lead times for winter target months than summer target months (Koenigk et al., 2008; Holland et al., 2010; Germe et al., 2014), greater predictability in the regions nearer to the Atlantic than the Pacific (Germe et al., 2014; Bushuk et al., 2018; Cruz-García et al., 2019), and the presence of a spring predictability barrier on the pan-Arctic scale and in several regions (Day et al., 2014b; Bushuk et al., 2018). It should be noted that for some forecast month/lead time pairs the operational model shows higher skill than the predictability. This phenomenon could be either the result of sampling errors or an underestimation of the potential predictability. This assessment provides both an estimate of the upper limit of predictability for the CanCM4 model on the pan-Arctic and regional scales as well as a quantification of the potential skill improvements that could be achieved with regards to the current operational model.

3.1 Introduction

Studies considering the potential predictability of sea ice conditions using dynamical models have been conducted in concert with past assessments of the forecast skill of operational prediction models discussed in Chapter 2. Many of these studies are perfect model experiments in which ensembles initialized with a minute difference in the initial conditions are compared against each other in turn by designating one ensemble member the "truth" or "synthetic observations" against which the other ensemble members are compared. The perfect model is assumed to represent the climate system and the skill obtained using this experiment is assumed to represent the upper limit of skill of a model which perfectly represents the climate system and is initialized with near-perfect initial conditions. This perfect model skill will be hereafter referred to as predictability. As with any modelled quantity, however, this upper limit may still be biased relative to the true predictability of sea ice.

In one of the earliest such experiments conducted using the Max Planck Institute's coupled model, Koenigk and Mikolajewicz (2008) found that SIT was highly predictable due to persistence, especially in the Central Arctic, while SIC was less predictable. Koenigk and Mikolajewicz (2008) further noted that predictability was greater for forecasts initialized in the summer than those in the winter. Holland et al. (2010) found similarly high SIT predictability on seasonal and interannual timescales

in a perfect model experiment using the Community Climate System Model. Holland et al. (2010) also found a thickness state-dependence of predictability such that SIC was more predictable in regimes of thicker sea ice. This state-dependence was also visible as a decrease in predictability in the spring followed by an increase during the summer for their January-initialized experiment. This instance of predictability reemergence suggests that preconditioning of winter sea ice is a source of some forecast skill for prediction of summer sea ice conditions. Germe et al. (2014) analyzed the results of a perfect model experiment run for the CNRM-CM5.1 model on decadal time scales. In line with Holland et al. (2010), this study found that SIE for winter target months was predictable for longer than SIE for summer target months and that SIT predictability extends out to two years. In considering regional SIE predictability, Germe et al. (2014) found higher predictability in the North Atlantic as compared to the Pacific, with the GIN Seas in particular having the highest predictability of any region in the study.

Perfect model studies have more recently included multiple models to provide a more complete assessment of the upper limit of predictability in the climate system. The first such study was conducted by Tietsche et al. (2014) in which the predictability of each of four models' July-initialized forecasts decreased to an ACC of 0.7 (high predictability) over the initial five months and then varied with an overall decrease to between 0.3 and 0.5 after 12 months and between 0.2 and 0.5 after 36 months. Tietsche et al. (2014) compared these results to the detrended forecast skill of the CanSIPSv1 results from Sigmond et al. (2013) to suggest that, with improvements, operational models could achieve higher levels of predictability. In a complementary study which was part of a contemporary project known as Arctic Predictability and Prediction On Seasonal to Interannual Timescales (APPOSITE), Day et al. (2016) attempted to quantify the maximum predictability of Arctic sea ice using the four climate models from Tietsche et al. (2014) as well as three additional models including CanCM4. The experimental design for the CanCM4 simulations was slightly different from those of the other models with ensemble members branching off from a transient climate run vice control runs with forcings fixed at one year as was done for the other models (either 1990, 2000, or 2005 depending on the model). Furthermore, the ensemble members for CanCM4 were run for only one year in Day et al. (2016) while the ensemble members of the other models were run for three. Day et al. (2016) found that, for the models considered, winter SIE was predictable at longer timescales than summer SIE and that SIV was better predicted than SIE. The second result is

consistent with other studies' findings regarding SIT as a source of predictability (Holland et al., 2010; Germe et al., 2014). SIE and SIV predictability in the single-year ensembles of CanCM4 was found to be lower than all other models and Day et al. (2016) hypothesized that this may be due to the difference in the setup of the control run. Cruz-García et al. (2019) further assessed the predictability of the APPOSITE models other than CanCM4 (which was excluded due to the shorter time period of its ensemble members) on a regional scale finding that the Baffin Bay and GIN Seas showed robust predictability across all models, that this predictability and that of other Atlantic sectors is primarily driven by persistence of thermal anomalies in the ocean, and that pan-Arctic SIE and SIV predictability is the result of the persistence of SIT anomalies in the Central Arctic. This last finding further supported those of Blanchard-Wrigglesworth and Bushuk (2018) using the APPOSITE simulations that predictability is linked to the time scales of sea ice anomaly persistence specific to each GCM.

Other potential predictability studies have considered the patterns and sources of predictability. Blanchard-Wrigglesworth et al. (2011) conducted a perfect model experiment with the Community Climate System Model version 4 to assess the dependence of predictability on initial conditions and climate forcing at different timescales. Their results showed that for CCSM4, initial conditions substantially affect predictability on time scales of 1-2 years while past 3 years, predictability was dominated by climate forcings. As discussed in Section 2.1.1, Day et al. (2014b) used the perfect model experiment conducted with the APPOSITE models to demonstrate the existence of a spring predictability barrier both on the pan-Arctic scale and in several regions. Finally, Goessling et al. (2016) used the APPOSITE experiment to develop and demonstrate a metric called the integrated ice-edge error to gain a better understanding of the geographic context of the ice edge than is provided by the overall SIE. Goessling et al. (2016) found that the ice edge is substantially less predictable than SIE especially in the target month of September.

Bushuk et al. (2018) conducted a detailed examination of the potential predictability of SIE as represented in GFDL-FLOR by considering forecasts initialized in six different months and quantifying predictability on the pan-Arctic as well as regional scales. Most previous studies have assessed predictability exclusively for a January or July forecast. Bushuk et al. (2018) was also the first perfect model study to directly compare operational skill of the same model in order to quantify the "room for improvement" in model skill. Due to a methodology of picking initialization years

from the control integration based on magnitude of SIV anomaly, a high predictability bias was introduced in Bushuk et al. (2018), which will be further discussed in Section 3.2.1. Kumar et al. (2014) emphasized that using a comparison of two models to infer a skill gap requires that the two models be statistically similar, which may not be the case in a potential predictability analysis. Bushuk et al. (2018) use a comparison of autocorrelation structures between the perfect model and observations to ascertain the validity of the assessment of potential improvement. The study described in this chapter is modelled on the methodology of Bushuk et al. (2018) with one change made to avoid that study's aforementioned high predictability bias.

3.2 Experimental Design

The overarching goal in designing the methodology presented here was to replicate with CanCM4 the perfect model experiment conducted on GFDL-FLOR by Bushuk et al. (2018) while avoiding the high predictability bias seen in that study. This experiment allows comparisons both of the perfect and operational models in each system, as well as between the two forecasting systems. The dynamical model used in this experiment, CanCM4, was described in detail in Section 2.2.1. This section will describe the control integration and the perfect model experiment which was started from it as well as the metrics used to assess both the predictability and operational forecast skill.

3.2.1 The Control Integration and Perfect Model Experiment

The perfect model simulations branched off from various points in a 450-year control integration of CanCM4 that was run repeatedly with external forcings (including greenhouse gases, anthropogenic aerosols, tropospheric and stratospheric ozone, and land use) observed in the year 1990. To help ensure a well-equilibrated control run and avoid skill biases associated with model spin up, the initialization years considered for the experiment were selected from the final 150 years of the control run.

Six start years were selected from the control run, based on SIV anomalies. Bushuk et al. (2018) chose two initialization years from each of low, typical, and high SIV anomaly years in order to assess the initial state dependence of predictability. The low and high years were chosen from years where all months had an SIV anomaly

greater than +/- 1.2σ respectively and the typical years were chosen from years where all months had an SIV anomaly less than +/- 0.25σ where σ is the standard deviation of the control run. Systematic high predictability biases were a consequence of having large anomalies in two-thirds of the start years unlike in the climatological distribution. To better represent the climatological distribution of sea ice anomalies in the initial conditions of the present perfect model experiments, the last 150 years of the control run of this experiment were divided into sextiles based on SIV anomaly and a single year was randomly selected from each (Figure 3.1). This approach imposed a closer probabilistic similarity of the ensemble members to the climatological distribution of SIV than simply randomly selecting any six years which could by chance select initialization years that would introduce a similar bias to that seen in Bushuk et al. (2018). Further, to ensure independence of the selected initial states, start years were required to be a minimum of 20 years apart.

For each start year, three-year forecasts with 12 ensemble members were initialized from the first day of every odd month of the year (January, March, May, July, September, and November). The ensemble members were generated by introducing a small perturbation to the control simulation's wind fields on the initialization date. The resulting 432 simulations (twelve ensemble members initialized at the beginning of six different months for six different years) were used to assess potential predictability. For each target month and lead time pair, each ensemble member is in turn designated "the truth" and the mean of the 11 remaining ensemble members is used to predict these synthetic observations.

3.2.2 Predictability and Operational Forecast Skill Metrics

The Anomaly Correlation Coefficient (ACC) is again used to quantify skill in this study. Predictability is calculated by predicting in turn each ensemble forecast x_{ij} for each of $M = 12$ ensemble members (j) initialized in each of $N = 6$ years (i), with the ensemble mean of the remaining ensemble forecasts $\langle \mathbf{x}_{\hat{i}j}(\tau) \rangle$. In both cases anomalies are defined relative to the climatological mean of the last 150 years of the control run, $\overline{\mu(\tau)}$, such that:

$$ACC(\tau) = \frac{\Sigma_{j=1}^{M} \Sigma_{i=1}^{N} \left(\langle \mathbf{x}_{\hat{i}j}(\tau) \rangle - \overline{\mu(\tau)} \right) \left(x_{ij} - \overline{\mu(\tau)} \right)}{\sqrt{\Sigma_{j=1}^{M} \Sigma_{i=1}^{N} \left(\langle \mathbf{x}_{\hat{i}j}(\tau) \rangle - \overline{\mu(\tau)} \right)^2} \sqrt{\left(x_{ij} - \overline{\mu(\tau)} \right)^2}} \tag{3.1}$$

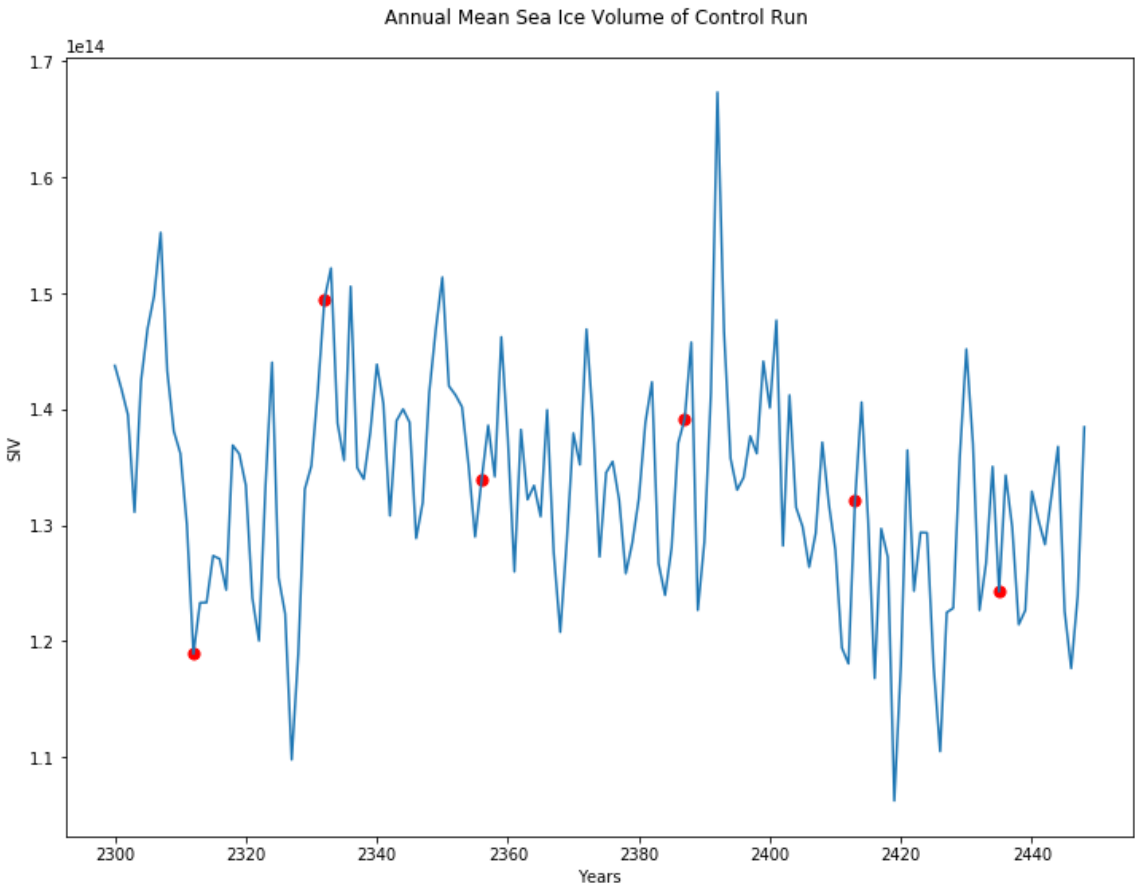

Figure 3.1: Time series of annual mean pan-Arctic SIV over the last 150 years of the control run. The six start years (2312, 2435, 2413, 2356, 2387, 2332) were selected based on SIV anomaly (red dots).

As the experiment was only conducted for six initialization months, the predictability was linearly interpolated for all remaining target month and lead time pairs, as was done by Bushuk et al. (2018). As in Chapter 2, statistical significance was assigned at the 95% confidence level estimated using bootstrapping. The bootstrapping method, which was repeated 1000 times to build the distribution, resampled with replacement the truth and the ensemble mean of the remaining ensemble members and calculated the ACC between them as in Equation 3.1. The same method used in Chapter 2 was also used to assess the statistical significance of the skill differences between predictability and operational forecast skill.

The operational forecast skill of the model is calculated using the same observations as described in 2.2.1 and the same metric of detrended forecast skill as described in 2.2.3. It should be noted, however, that the operational skill in this chapter is solely

based on CanCM4 with the Dirkson et al. (2017) procedure, one of the two models used in the CanSIPSv1b and CanSIPSv2 forecasts systems analyzed in Chapter 2. Regional analyses are conducted using the same regions as in Chapter 2 and depicted in Figure 2.2.

3.2.3 Statistical Similarity of the Perfect Model and Observations

Kumar et al. (2014) noted that model biases may preclude a comparison of predictability from perfect model simulations to operational skill if the signal-to-noise ratio of the perfect model is substantially different from that of real world observations. To address this potential limitation, Bushuk et al. (2018) compared the autocorrelation structures for given target months and lead times of the perfect model and relevant SIE observations as a proxy of the signal-to-noise ratio suggested by Kumar et al. (2014). Autocorrelations were also calculated in the present analysis using Had2CIS data (Titchner and Rayner, 2014; Tivy et al., 2011) to provide some guidance regarding the validity of interpreting the difference between the predictability and operational forecast skill as a measure of the "skill gap". Assessments are made through a qualitative comparison of autocorrelation structure to determine if the predictability was overestimated, underestimated, or correctly estimated as compared to the operational model skill. It should be emphasized that the methodology used in Bushuk et al. (2018) and the present study where operational sea ice models are compared to potential predictability is still relatively new in the literature. It should also be underscored that comparison of autocorrelation structures is a convenient proxy for comparing the signal-to-noise ratios of the perfect model and climate system but cannot be taken as a flawless quantitative measure. While the similarity of the autocorrelation structures may support this comparison for a given region, the criterion is not perfect. It should further be noted that in Bushuk et al. (2018) and other studies such as of sea surface temperature (Newman and Sardeshmukh, 2017) and the North Atlantic Oscillation (Weisheimer et al., 2019), the actual skill of the operational model has been greater in some instances than the potential predictability would suggest is possible. This reaffirms that the potential predictability measure itself may not necessarily be a precise reflection of the upper boundary of the climate system's predictability. These caveats do not invalidate the perfect predictability methodology, but should be taken into account in its interpretation.

3.3 Results

The results of the perfect model experiment are presented in this section along with a comparison to the skill of the operational model. Predictability and operational forecast skill plots are presented as in Chapter 2 with the target month of a forecast on the horizontal axis and forecast lead time on the vertical axis (e.g. September lead 3 is the September forecast initialized on June 1st). Also as in the previous chapter, the Canadian Arctic Archipelago is excluded from the regional analysis as it is not adequately represented in CanCM4 due to the low resolution of the model. Finally, as this study is again interested in the model's skill with regards to predicting variability, neither predictability nor operational forecast skill are calculated where the annual standard deviation of a region's observed SIE or the perfect model's SIE is less than 0.8% of the region's area for a target month. It should be noted that in the excluded target months, the model could make a perfectly skillful forecast of complete ice coverage or completely open water; skillful forecasts of this kind are not captured in this analysis.

3.3.1 Control Run Biases

As was done by Bushuk et al. (2018), the sea ice biases of the model were analyzed by comparing annual cycles of monthly mean SIE of the first year of each of the 72 ensemble members initialized in January against NSIDC observations from 1979-2016 (Fetterer et al., 2017) and of the monthly mean SIV against PIOMAS (Zhang and Rothrock, 2003). Similar to the biases of GFDL-FLOR found by Bushuk et al. (2018), the control run climatology agrees with the observations with regards to SIE but the SIV climatology has a notable thin bias (Figure 3.2). As discussed in Section 2.1.2, SIT is an important source of predictability, especially for summer target months, and thus this thin bias may lead to an underestimation of the model's potential predictability in those months. Such a bias should be reflected in the autocorrelation analysis described in Section 3.2.3.

3.3.2 Pan-Arctic SIE Predictability

The predictability of CanCM4 on the pan-Arctic scale (Figure 3.3) has two peaks of high predictability (ACC $>$ 0.7) extending from the January lead 0 and July lead 0 initialization months diagonally out to May lead 4 and October lead 3, respec-

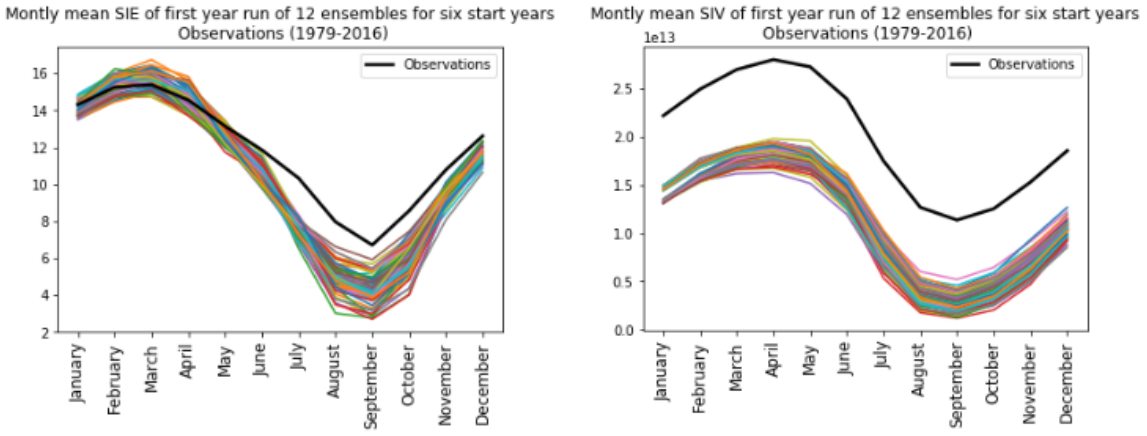

Figure 3.2: Seasonal cycles of SIE (left) and SIV (right) of January-initialized ensemble members (colours) compared against NSIDC observations and PIOMAS model data respectively (black).

tively. Annual predictability reemergence is seen as the predictability of both the January- and July-initialized forecasts increases at lead 12 and remains significant for four subsequent months. Predictability from forecasts initialized in other months decreases much more quickly, particularly those initialized in May and November. This more rapid decline indicates that there is a seasonality in initialization month predictability. It should be considered that previous studies have generally focused on the initialization months of this pattern which produce the highest predictability forecast (January and July) and that Bushuk et al. (2018) did not find similar peaks in the predictability of GFDL-FLOR.

The predictability of CanCM4 in the present study is substantially lower than found for GFDL-FLOR by Bushuk et al. (2018). This result is particularly notable given that a comparison of the autocorrelation structures of the perfect model and the observations suggests that the predictability in CanCM4 should be overestimated (Figure A.17). Day et al. (2016) suggested that the lower potential predictability found for CanCM4 compared to other models was the result of a difference of control run setup particular to that study. The predictability of July-initialized forecasts here demonstrates, however, that CanCM4 is inherently less predictable than other models and that the lower CanCM4 predictability reported in Day et al. (2016) is not exclusively due to the different control run setup. In assessing predictability out to 36 months, this study further shows that CanCM4 has especially low predictability past 18 months as compared to the other APPOSITE models.

Two lobes of predictability form around the target months associated with the seasonal advance (March and April) and retreat (October and November) of Arctic sea ice. In concurrence with previous studies (Holland et al., 2010; Germe et al., 2014; Day et al., 2016), predictability of winter target months is found to exist at longer lead times than summer target months. Significant predictability of March and April forecasts is seen almost continuously from lead 0 out to 18 months lead time and then sporadically thereafter with statistically significant forecasts for March lead 34 and April lead 35. By contrast, fewer than half of October and November forecasts are significant for lead 6 out to lead 18 with only one significant forecast (November lead 34) seen after lead 20. It should be noted that given a confidence interval of 95%,

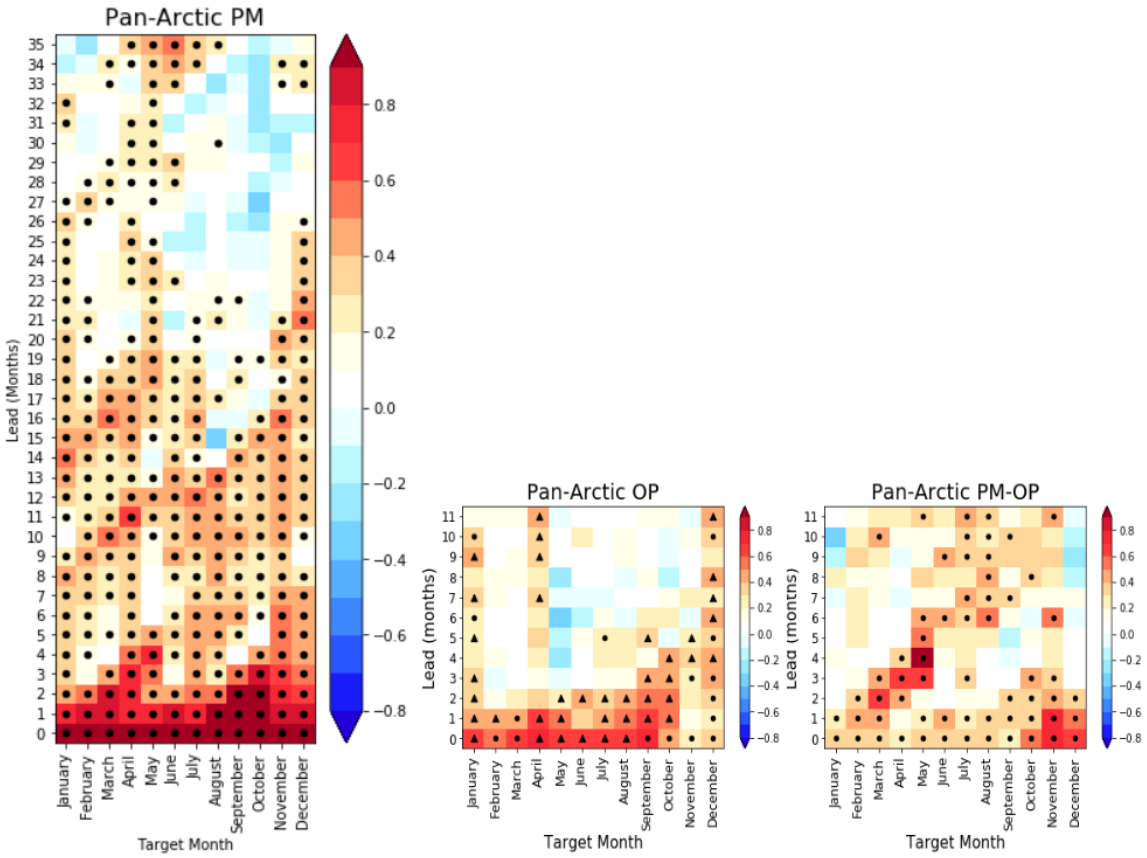

Figure 3.3: Pan-Arctic predictability (left), operational model skill (centre), and the skill difference (right). Markers (dots and triangles) in the left and middle plots indicate predictability or skill significant at the 95% confidence level. Triangles in the operational model skill indicate the forecast has higher skill than an anomaly persistence forecast (Section 2.2.3). Dots in the right plot indicate a statistically significant difference in skill between the predictability and operational forecast skill.

one in twenty forecasts are expected to appear significant by chance.

Finally, the spring predictability barrier (Day et al., 2014b) is clearly evident on the pan-Arctic scale with an abrupt decrease in predictability of September and October when initialized in May versus June. Given the assumption of perfect initial conditions in this experiment, this result supports the interpretation that this barrier is inherent to the climate system, as has been suggested in other studies (Day et al., 2014b; Bushuk et al., 2018).

3.3.3 Regional SIE Predictability

The regional predictability is presented in Figures 3.4 to 3.6. In the majority of regions, predictability decreases quickly with only sporadic significant predictability for lead times longer than six months for all target months. Notable exceptions are the GIN and Barents Seas, Bering Sea, Sea of Okhotsk, and Labrador Sea. Some regions also exhibit isolated predictability for specific forecasts (such as lead 20-23 of the November-initialized forecast in the Laptev Sea), though these may be chance occurrences given the sample size. Additionally, predictability reemergence is evident in some regions such as annual reemergence of the June- and July-initialized forecasts in the Beaufort Sea and near annual reemergence of the October-, November-, and December-initialized forecasts in the Chukchi Sea.

The predictability in the GIN Seas and, to a lesser extent the Barents Sea, is exceptionally high especially in the winter months. The longest lead times for which there are high predictability (ACC $>$ 0.7) are January lead 12 in the GIN Seas and February lead 7 for the Barents Sea. The majority of forecasts in the GIN and Barents Seas for the target months of February and March in the first year have high predictability. By contrast, the GIN Seas have the lowest summer and fall predictability of all regions with partial ice coverage; only sporadic significant predictability is seen for lead times greater than 2 months. Given that the GIN and Barents Seas see partial ice coverage all year, the seasonality of this predictability can be most easily compared to the patterns seen in previous studies on the pan-Arctic. As on the pan-Arctic scale, significant predictability appears for forecasts with longer lead times for winter target months than summer target months (Holland et al., 2010; Germe et al., 2014; Day et al., 2016). Robust predictability in the GIN Seas was also found in CNRM-CM5.1 (Germe et al., 2014) and the other APPOSITE models (Cruz-García et al., 2019), but was less apparent in the Barents Sea in those

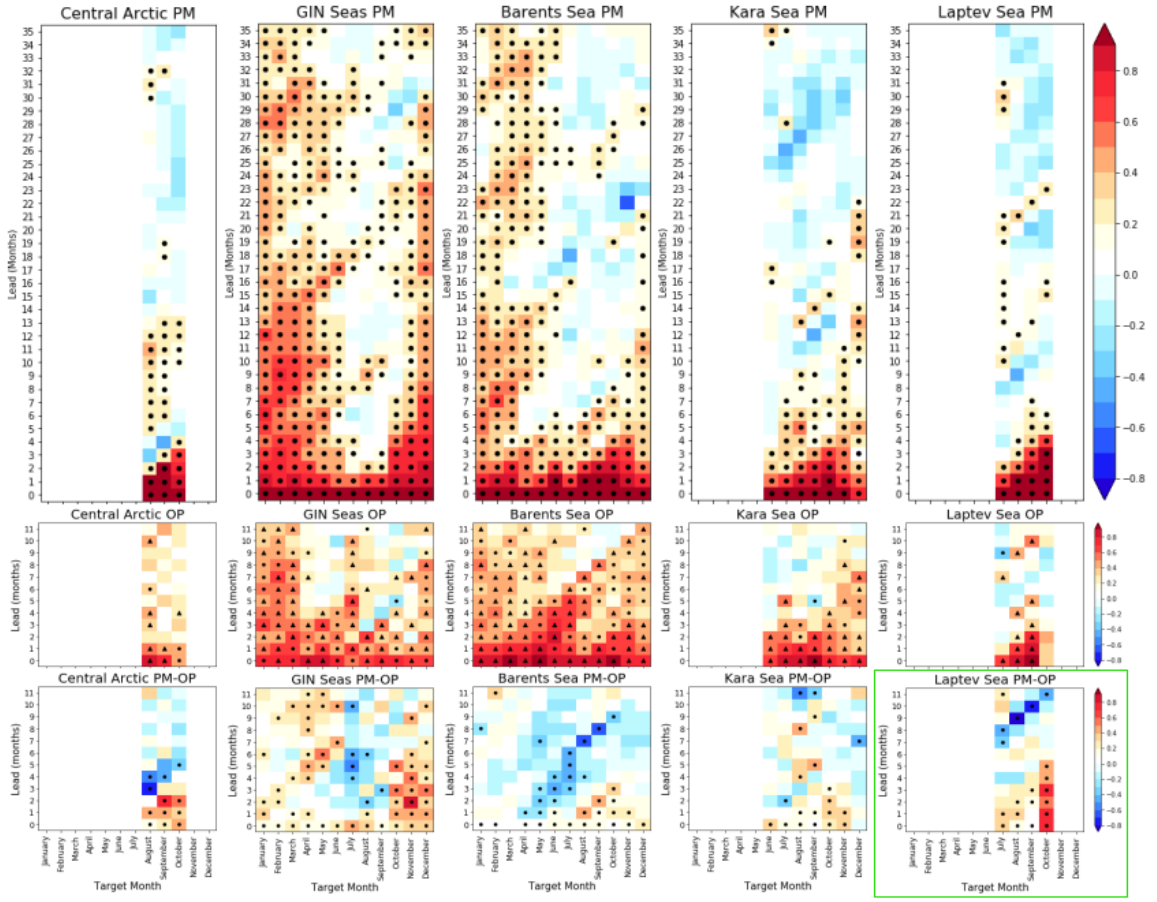

Figure 3.4: Predictability (top), operational model skill (middle), and the skill difference (bottom) for the Central Arctic, GIN Seas, Barents Sea, Kara Sea, and Laptev Sea. Markers (dots and triangles) in the top and middle plots indicate predictability or skill significant at the 95% confidence level. Triangles in the operational model skill indicate the forecast has higher skill than an anomaly persistence forecast (Section 2.2.3). Dots in the bottom plots indicate a statistically significant difference in skill between the predictability and operational forecast skill. Neither predictability nor skill are calculated for target months where the annual standard deviation of observed SIE in the region from 1980-2018, or of the control run SIE, is less than 0.8% of the region's area. Skill difference plots are outlined in green where the autocorrelation structures (Figures A.15-A.17) of the model and observations are sufficiently similar to allow for an assessment of the "skill gap" (Section 3.2.3).

studies.

Although the predictability in the Labrador Sea differs between the present study and previous studies, it is notable that in all studies the region exhibits a pattern of predictability distinct from other regions of the same study. In the present study,

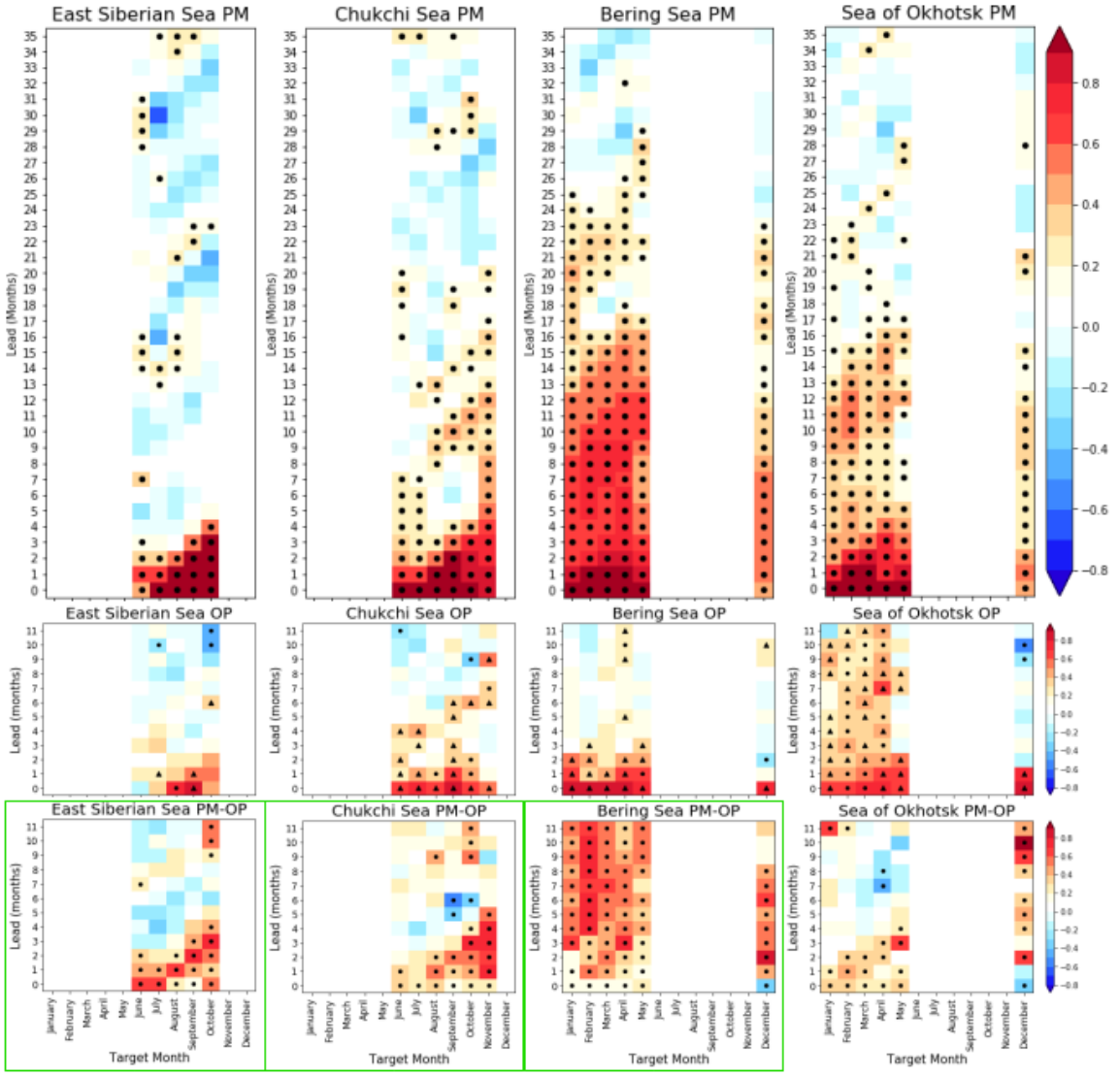

Figure 3.5: As in Figure 3.4 for the East Siberian Sea, Chukchi Sea, Bering Sea, and Sea of Okhotsk.

forecasts for the Labrador Sea have the highest predictability compared to any other regions. All but one forecast of January through May out to lead 16 show significant predictability. Further, almost continuous skillful forecasts are seen for March, April, and May out to lead 35 with notably high predictability (ACC > 0.7) for April lead 17-19, lead 31 and lead 35 as well as May lead 28. This result is generally different from those of the July-initialized APPOSITE perfect model experiments which saw overall low predictability other than high predictability peaks in January and April which were attributed to sub-polar gyre persistence in the first year and advection of

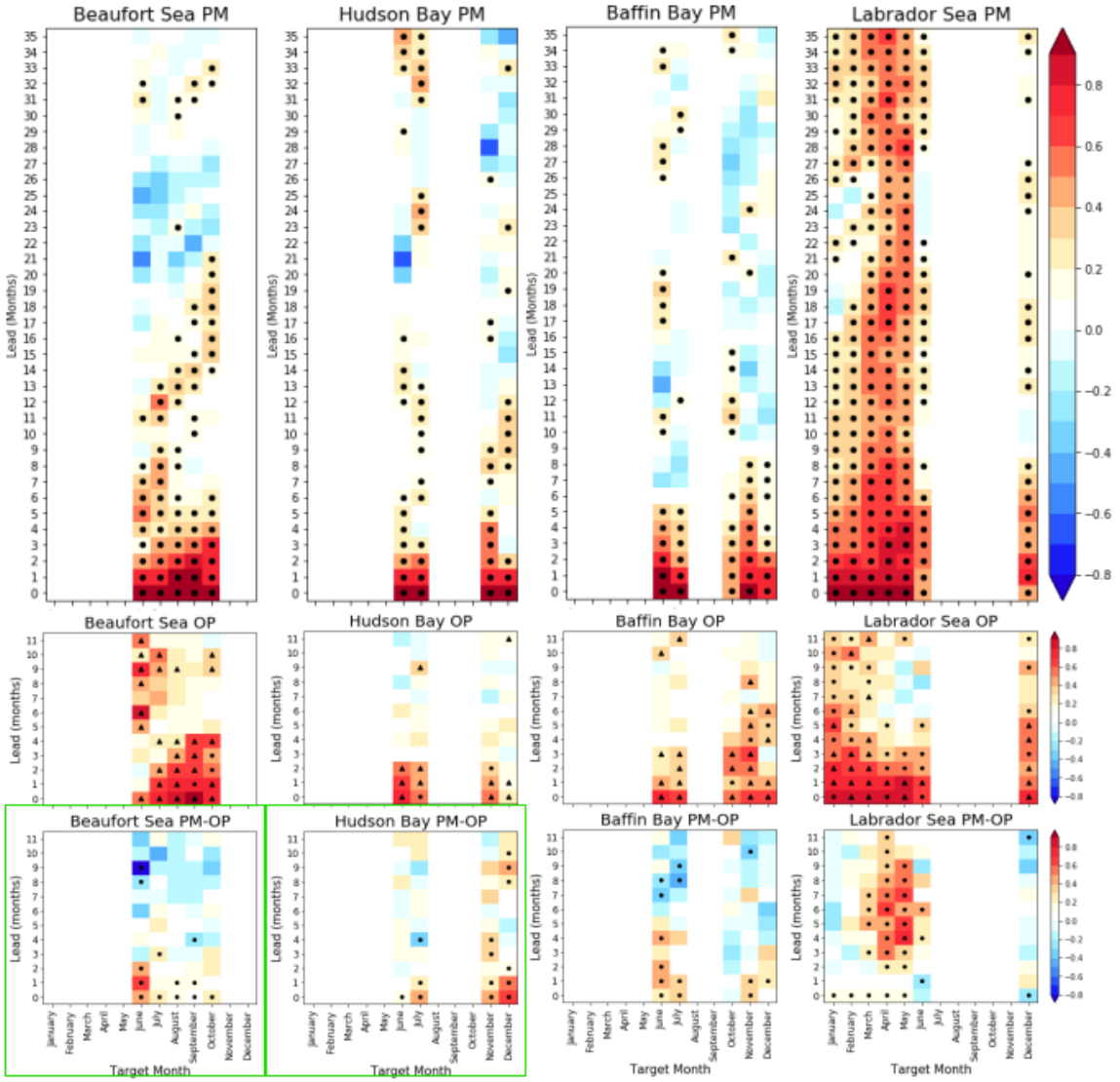

Figure 3.6: As in Figure 3.4 for the Beaufort Sea, Hudson Bay, Baffin Bay, and Labrador Sea.

ocean temperatures in the second year (Cruz-García et al., 2019). When considering July-initialized forecasts in the present study, however, the annual April target month predictability peak seen in the APPOSITE models is reproduced as can be seen at April lead 10 (with predictability increase from March lead 9 and decrease to May lead 11 in year 1), April lead 22 (with predictability increase from March lead 21 and decrease to May lead 23 in year 2); and April lead 34 (with predictability increase from March lead 33 and decrease to May lead 35 in year 3).

As was found on the pan-Arctic scale, the spring predictability barrier is also

present in the seas nearer to the Pacific as well as the Central Arctic. In the Central Arctic, Kara, and East Siberian Seas, the barrier can be seen as a substantial increase in predictability from the May- to June-initialized forecasts. This sudden predictability increase is seen between the April- and May-initialized forecasts in the Laptev and Chukchi Seas and March- and April-initialized forecasts in the Beaufort Sea. As previously discussed, this result indicates that the spring predictability barrier is inherent in the climate system and explains the limited operational forecast skill improvement seen with the improved SIT initialization procedure (Dirkson et al., 2017) in some regions in Chapter 2 — particularly the East Siberian and Chukchi Seas. Indeed, even for June- and July-initialized forecasts of August, September, and October, which are relatively high predictability forecasts in most regions, there is low potential predictability in the East Siberian and Chukchi Seas which is consistent with the relatively low skill of the operational model for these target months.

3.3.4 Skill Gap Between Perfect and Operational Models

As outlined in Section 3.2.3, for predictability to be used as a benchmark for the operational model, the perfect model must be statistically similar to the observations. In the present study, as in Bushuk et al. (2018), an assessment of autocorrelation structure was used to make this determination. This assessment is presented in the Appendix (Figures A.15-A.17). On the pan-Arctic scale, the assessment of autocorrelation suggests that CanCM4 predictability overestimates that of the real climate system. In the Central Arctic, Sea of Okhotsk, Baffin Bay, and the Kara and Labrador Seas, the perfect model should underestimate the predictability. The GIN and Barents Seas are notable in that predictability should be overestimated by the perfect model for lead times greater than lead 11 and underestimated for lead time of 11 months or less (those which are compared against operational forecast skill). Relative similarity between the autocorrelation structures of the perfect model and the observations were found in the remaining regions for which the ACC difference or "skill gap" plots are outlined in green.

In Hudson Bay (a region for which the autocorrelation structure of the model corresponds well with the observations), the operational model's skill is close to the upper limit of predictability as estimated from the perfect model. The Laptev, East Siberian, and Chukchi Seas, other regions with reasonable autocorrelation agreement between the model and observations, each show moderate skill gaps. The October

forecast skill of the operational model in the Laptev Sea in particular shows evidence of room for improvement. The Bering Sea has the largest skill gap with strong predictability (ACC > 0.7) seen for the majority of forecasts out to lead 11 not captured in the operational model's forecast skill. Finally, while there was strong agreement between the autocorrelation structure of the Beaufort Sea's perfect model and observations, the substantial number of forecasts for which the operational model has significantly higher skill than the predictability suggests that the perfect model's skill may not be entirely representative of the region's potential predictability. This result could partially be a consequence of sampling error as will be discussed in Section 3.3.5 and regardless further demonstrates that the comparison of autocorrelation structures may not be a perfect representation of the similarity of signal-to-noise ratios.

Despite statistical dissimilarity of the autocorrelation structure of the perfect model and observations, the results for the Labrador Sea still reveals useful information. Given that the predictability in the Labrador Sea is likely underestimated, the "room for improvement", especially for fall- and winter-initialized forecasts of April and May, may be even more substantial than indicated.

3.3.5 Comparison to GFDL-FLOR Model

As part of the assessment of the predictability of SIE using the CanCM4 model, this study also compared those findings against the previous potential predictability experiment conducted by Bushuk et al. (2018). It is important to first note that the methodology for picking start years used in Bushuk et al. (2018) resulted in a high predictability bias and corrective post-processing was used to reduce this bias in that study. The primary change in the methodology for the present study was made to avoid this systematic bias. As shown in Figure 3.7, when predictability is calculated for each of the six initialization years individually, the estimated predictability generally increases as the magnitude of the initial anomaly increases. This result corroborates the bias seen in Bushuk et al. (2018) which has been reduced in the present study through a selection of initialization years for representative of the climatological SIV distribution.

Overall, the predictability of GFDL-FLOR reported in Bushuk et al. (2018) is higher than that of CanCM4. This predictability is found on the pan-Arctic scale as well as in all regions except the Labrador Sea where, as previously discussed, the predictability of CanCM4 extends throughout the winter months out to lead

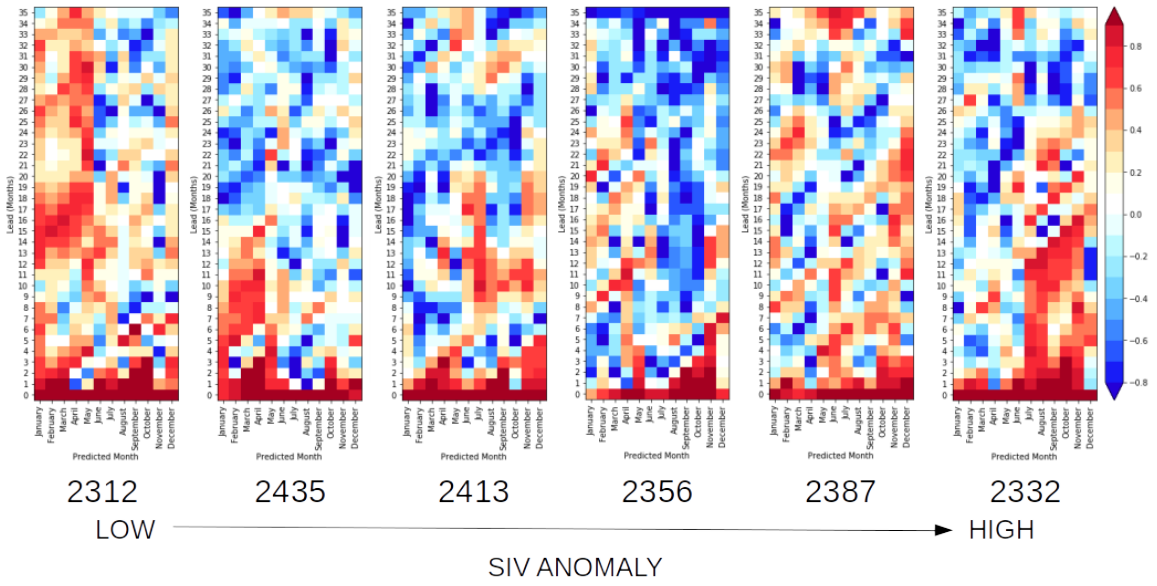

Figure 3.7: Predictability for each individual initialization year from lowest initial SIV anomaly (2312, left) to highest initial SIV anomaly (2332, right). It should be noted that the magnitude of the anomalies, the larger of which caused the high predictability bias in Bushuk et al. (2018), increases from the middle two plots outwards.

35 whereas GFDL-FLOR sees sparse predictability after lead 12. The skill of the operational models is similar in almost all contexts with the exception of the Central Arctic, GIN Seas, and Sea of Okhotsk where CanCM4 has greater skill and the East Siberian Sea where GFDL-FLOR has greater operational forecast skill. The findings with regards to predictability indicate that the SIE is clearly more predictable in GFDL-FLOR. It is worth noting that given the relative similarity of operational forecast skill, this high predictability also means that GFDL-FLOR generally has a larger skill gap. In comparing such skill gaps, however, it should be considered that the only regions where the autocorrelation structures of the perfect model and detrended observations were similar in both studies were: the Laptev Sea, East Siberian Sea, Bering Sea, Beaufort Sea, and Hudson Bay.

Despite the consistently higher predictability of GFDL-FLOR and some differences in operational forecast skill, similar predictability characteristics can be seen between the two models. Both CanCM4 and GFDL-FLOR show higher predictability in the Atlantic sector than in the Pacific sector, in line with the findings of Cruz-García et al. (2019) regarding the other APPOSITE models. Additionally, the predictability of several regions in GFDL-FLOR shows clear evidence of a spring predictability bar-

rier as was also found in the APPOSITE models (Day et al., 2014b). These results suggest that while the magnitude of predictability is clearly different between these two models, they may each have similar sources of predictability.

Bushuk et al. (2018) noted that in rare instances, the operational model had higher skill than predictability for a given target month and lead time and interpreted these inconsistencies as resulting from sampling errors. Such instances were much more common in this study and in some cases the differences were quite large (for example August lead 9 and September lead 10 in the Laptev Sea) or quite common in a specific region (for example in the Beaufort Sea). While such sampling errors may simply be more prolific in this study, or more evident due to the high predictability bias in Bushuk et al. (2018), future comparisons of predictability and operational forecast skill should monitor for such discrepancies and consider their implications for the validity of this methodology.

3.4 Conclusion

The potential predictability of Arctic sea ice in CanCM4 was assessed in this study based on a perfect model experiment. The methodology of Bushuk et al. (2018) was largely followed with the exception of the process for selecting initialization years which was altered to reduce the high predictability bias seen in the GFDL-FLOR study. On the pan-Arctic scale, CanCM4 presented lower predictability than seen for other models. This finding shows that the low predictability of CanCM4 in the APPOSITE study (Day et al., 2016) is not the result of methodological differences, but correctly represents the relatively low predictability of Arctic sea ice in this model compared to other models. On the regional scale, the predictability of CanCM4 was below that of GFDL-FLOR (Bushuk et al., 2018) in every region except the Labrador Sea although the forecast skill of the operational models was broadly similar. In regions for which the skill gap can be meaningfully assessed (based on statistical similarity of the perfect model and the observations), there is evidence of large room for improvement in the Bering Sea and a moderate skill gap in the Laptev, East Siberian, and Chukchi Seas. The operational model appears to be performing near the upper bound of predictability in Hudson's Bay and the Kara Sea.

Several predictability characteristics seen in previous models were also evident in the predictability of CanCM4. Generally, two lobes of predictability could be seen in the winter and fall target months where there is partial ice coverage at that time

with peaks occurring in February or March and October or November respectively. As was found in previous studies (Koenigk et al., 2008; Holland et al., 2010; Germe et al., 2014), the predictions of winter target months showed higher predictability and maintained that predictability at longer lead times. As was found for GFDL-FLOR (Bushuk et al., 2018) and the other APPOSITE models (Cruz-García et al., 2019), the regions in the vicinity of the North Atlantic, especially the GIN and Barents Seas, are substantially more predictable than the regions nearer to the Pacific in CanCM4. As was seen in GFDL-FLOR, the spring predictability barrier is evident on the pan-Arctic scale and in several regions in CanCM4 suggesting that this feature is inherent in the climate system and explains, given the perfect model assumption, why initialization improvements made by Dirkson et al. (2017) did not completely remove this phenomenon (Chapter 2). Finally, instances in which the operational model showed higher skill than predictability were more frequent or larger in magnitude than those seen in Bushuk et al. (2018). While this result may be in part due to sampling error in this study, an investigation of the occurrence of such negative skill gaps in other models would provide an interesting direction of research with regards to the use of perfect model experiments to assess predictability of sea ice.

Chapter 4

Conclusions

Motivated by the increased need for operational planning tools in the Arctic, this thesis assessed the skill of CanSIPS as a seasonal sea ice forecasting system. In Chapter 2, hindcasting skill was assessed for CanSIPSv1 (Merryfield et al., 2013a) and CanSIPSv2 (Lin et al., 2020) as well as an intermediate version, CanSIPSv1b, in order to assess the operational forecast skill of the current system and also to ascertain features of the new versions which result in an improvement of skill. This analysis was conducted both on the pan-Arctic scale as well as regionally. In Chapter 3, a perfect model experiment was conducted in order to assess the pan-Arctic and regional potential predictability of CanCM4, the constituent model common to all three versions of CanSIPS. The experimental design was based on the experiment conducted by Bushuk et al. (2018) on the GFDL-FLOR model. With the assumption that the perfect model skill or predictability provides an upper limit of forecast skill, a "skill gap" can be quantified by comparing the operational forecast skill of CanCM4 against the potential predictability.

From the analyses of operational skill and potential predictability, several conclusions can be made regarding the seasonal sea ice forecasting skill of CanSIPS. First, it is clear that CanSIPSv2 has overall greater forecast skill than CanSIPSv1 with most regions exhibiting a predominance of significant increases over decreases in forecast skill. The source of this improvement in skill for detrended forecasts (whether it comes from the improved SIC and SIT initialization developed by Dirkson et al. (2017) or from the substitution of GEM-NEMO for CanCM3) varies by region. Second, the potential predictability of CanCM4 is relatively low compared to other models such as those included in the APPOSITE study (Day et al., 2016) and GFDL-FLOR (Bushuk et al., 2018). Given the relative similarity in operational skill of CanCM4 and GFDL-

FLOR, the corollary to this finding is that the "skill gap" between the CanCM4 operational model and its potential predictability is much smaller than in GFDL-FLOR. This apparent lower predictability must be interpreted with the consideration of the high predictability bias in Bushuk et al. (2018) and the use of only January- and July-initialized forecasts in previous potential predictability studies. The results of the present study suggest that there is a substantial inherent lack of predictability in seasonal sea ice forecasting, especially in thinner regimes (ie. the summer months) which will only become more frequent with climate change. Indeed, as both improvements on several operational forecast systems and climate change continue, many may be approaching the upper limits of SIE predictability. Finally, it is clear that despite an overall lower potential predictability, CanCM4 exhibits the same skill characteristics as many other dynamical seasonal sea ice forecasting models. These include: higher potential predictability for winter target months than summer target months (Koenigk et al., 2008; Holland et al., 2010; Germe et al., 2014; Bushuk et al., 2018), higher potential predictability and operational skill in the Atlantic than Pacific sector (Sigmond et al., 2016; Bushuk et al., 2017, 2018; Cruz-García et al., 2019), and the presence of a spring predictability barrier in the potential predictability on the pan-Arctic scale and in most regions, suggesting that this feature is inherent to the climate system.

The assessments conducted in the present study provide a foundation for several avenues of further inquiry. Although thirteen regions of the Arctic were assessed in both the analysis of operational skill and of potential predictability, the scope of this analysis did not allow for an in-depth consideration of the processes unique to each region which affect both of these aspects of predictability. The physical mechanisms which result in the differences in seasonal and regional predictability could be further identified through more focused case studies. Such research could also better explain why modifications such as the improved SIT initialization procedure or change in constituent models have greater effects in some regions than others.

The perfect predictability study conducted in Chapter 3 represents only the second use of the methodology introduced by Bushuk et al. (2018). Unique features of this methodology include the use of six initialization months in each year in order to examine the seasonality of predictability and the comparison of potential predictability as represented in the subject model to the operational forecast skill of the same model. As a result of the nascent status of this methodology in the literature, it is difficult to fully assess the implications of unexpected results. First, the use of autocorrela-

tion structures of the perfect model and detrended observations for the assessment of the similarity of signal-to-noise ratios recommended by Kumar et al. (2014), cautions against directly comparing the operational skill to potential predictability for half of the regions studied (including the pan-Arctic scale). This assessment should be taken as an important caveat, but should not completely discount the comparison of operational skill to potential predictability. This is particularly true in regions where predictability is underestimated by the perfect predictability analysis but there is a large skill gap (such as in the Labrador Sea). It should also be noted that the autocorrelation structure comparison is not a perfect representation of the similarity of signal-to-noise ratios of the perfect model and climate system. Finally, the presence of numerous target month and lead time pairs for which the operational model had higher skill than the predictability could be simply the result of sampling errors. It is worth noting that potential predictability experiments studying Indo-Pacific sea surface temperatures (Newman and Sardeshmukh, 2017) and the North Atlantic Oscillation (Weisheimer et al., 2019) have seen similar occurrences of actual skill exceeding potential predictability. Given the occurrence of such features in both studies to date which have used the methodology these instances of a negative "skill gap" warrant further analysis that will be possible when this methodology is used on other CGCMs.

These results show that CanSIPSv2 is as skillful as other current operational models at forecasting seasonal sea ice conditions. Further, the system's forecast skill varies both geographically and seasonally in similar ways to comparable operational modelling systems. The analysis here does suggest, however, that one of CanSIPS' constituent models represents the climate system's sea ice as less predictable than in other CGCM-based systems. The importance of proper SIC and SIT initialization as well as the benefits of an effective combination of models in a multi-model system, as shown in Chapter 2, should be considered in future efforts to improve seasonal sea ice forecasting. Those efforts should also be informed by an understanding of the inherent unpredictability of Arctic sea ice especially in thinner regimes as depicted by the potential predictability experiment described in Chapter 3. From CanSIPSv1 to CanSIPSv2, significant improvements in seasonal sea ice forecast skill have been achieved. Further developments currently underway will likely bring this operational model closer to the maximum achievable skill of a dynamical model.

Appendix A

Additional Information

A.1 Differences in Detrending Methods

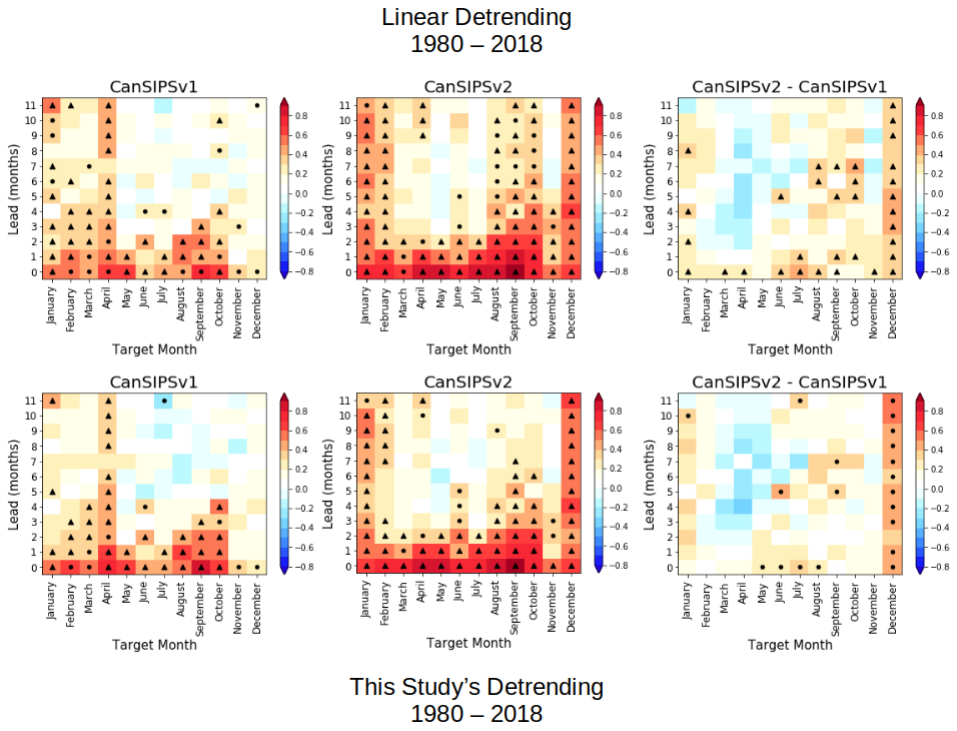

Figure A.1: Difference in traditional linear detrending and the detrending conducted in this study shown for the assessments of pan-Arctic SIE forecast skill

A.2 Regional Operational Skill with Difference Plots

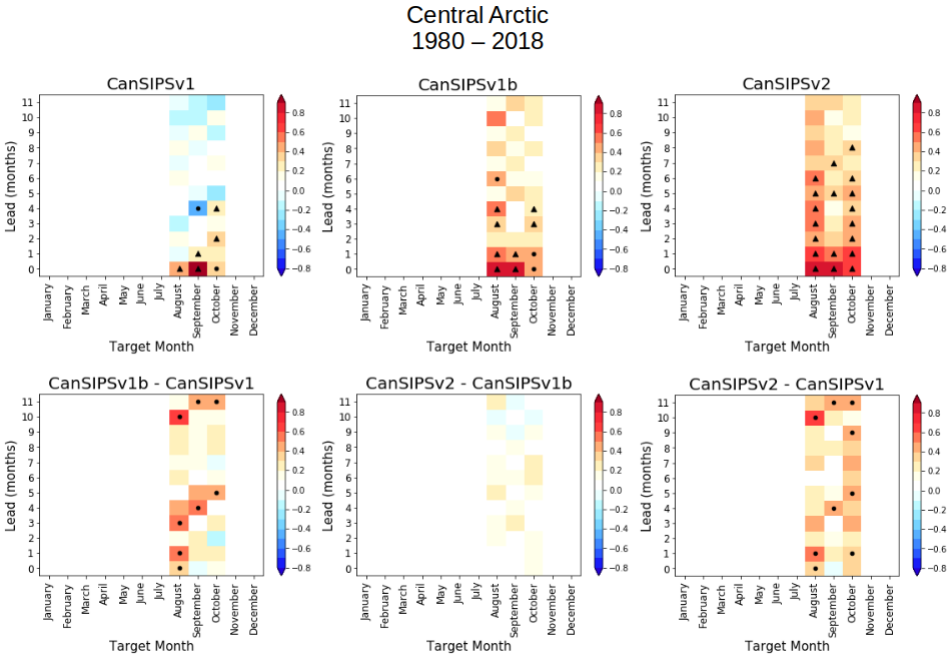

Figure A.2: Operational skill of CanSIPS in the Central Arctic.

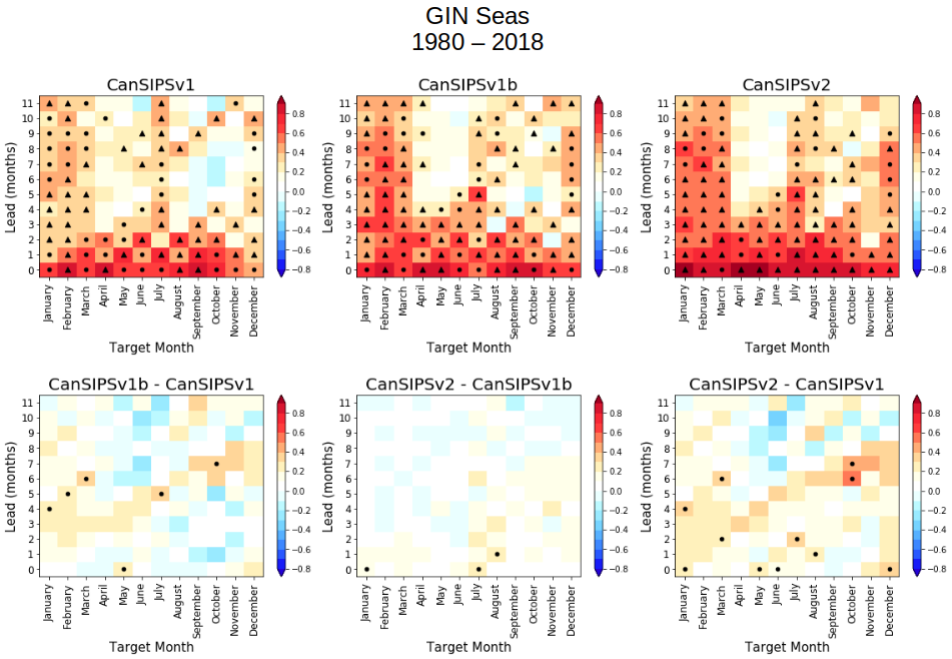

Figure A.3: Operational skill of CanSIPS in the GIN Seas.

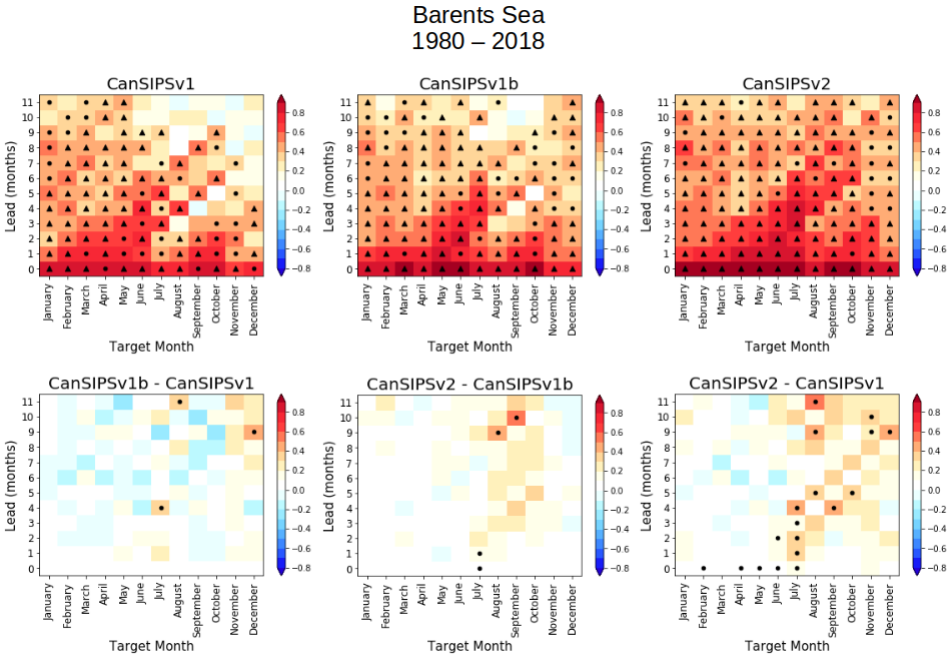

Figure A.4: Operational skill of CanSIPS in the Barents Sea.

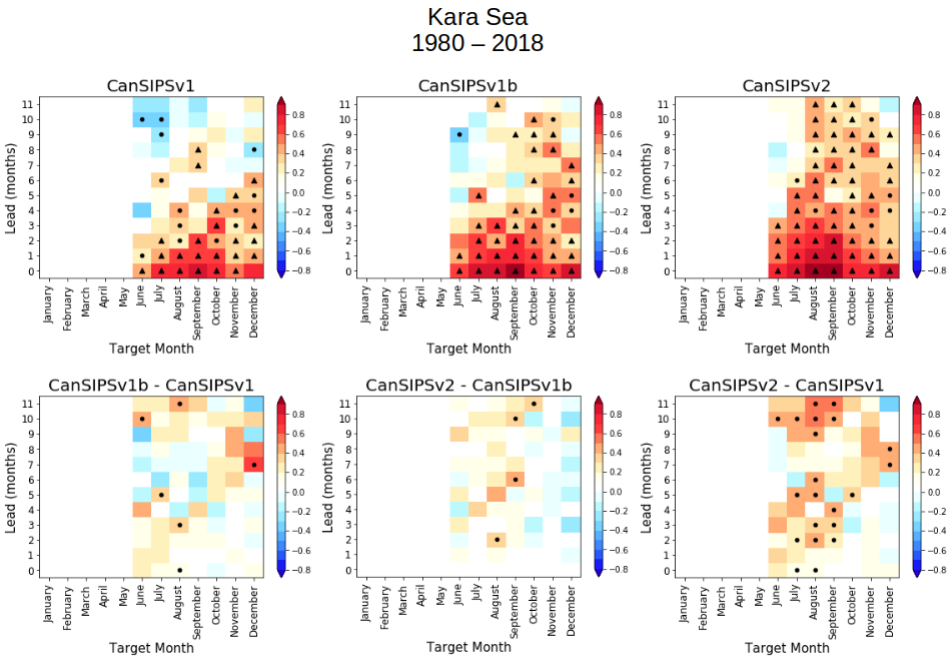

Figure A.5: Operational skill of CanSIPS in the Kara Sea.

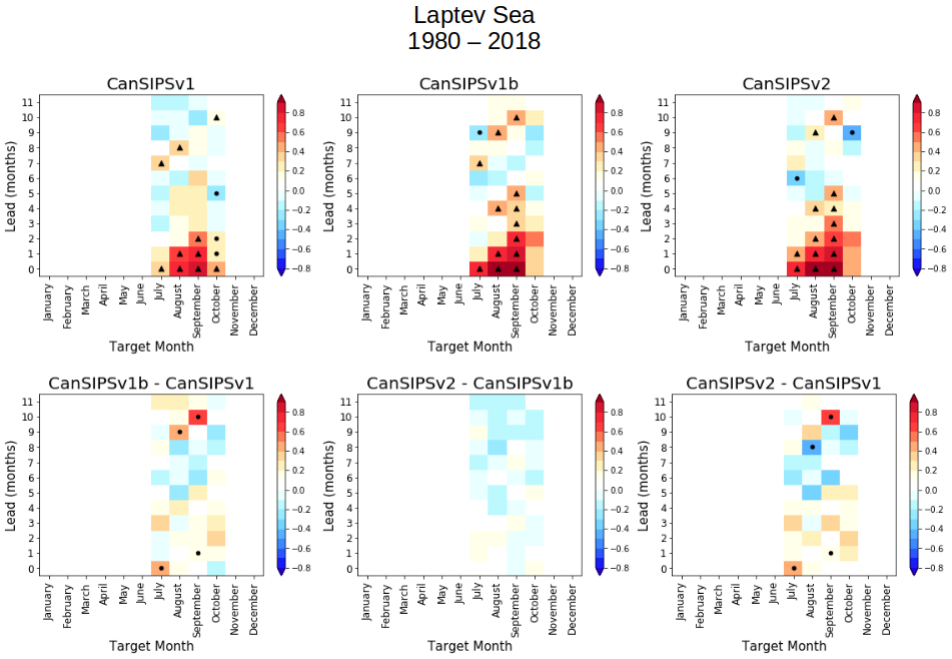

Figure A.6: Operational skill of CanSIPS in the Laptev Sea.

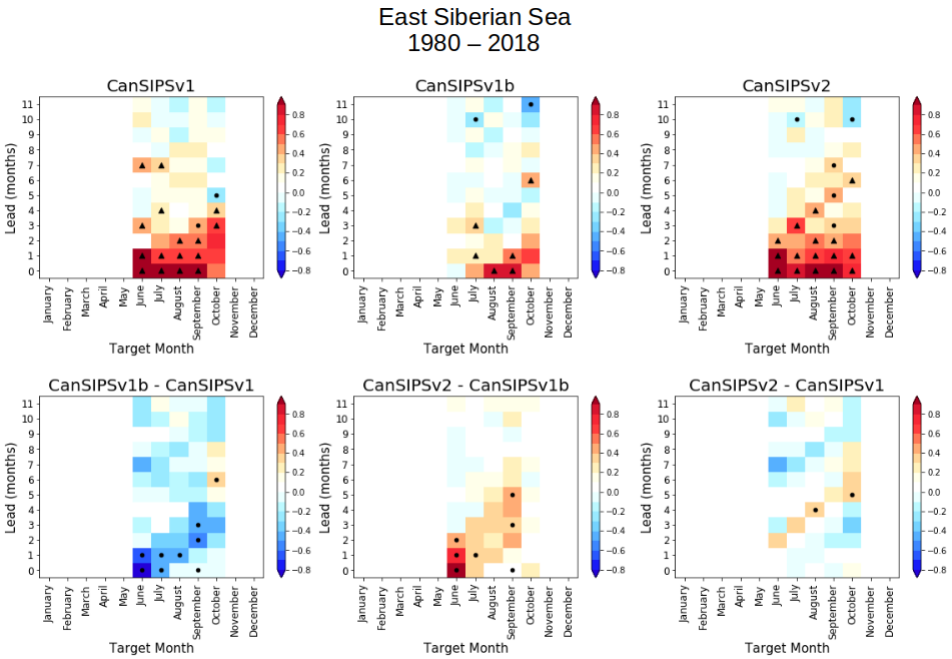

Figure A.7: Operational skill of CanSIPS in the East Siberian Sea.

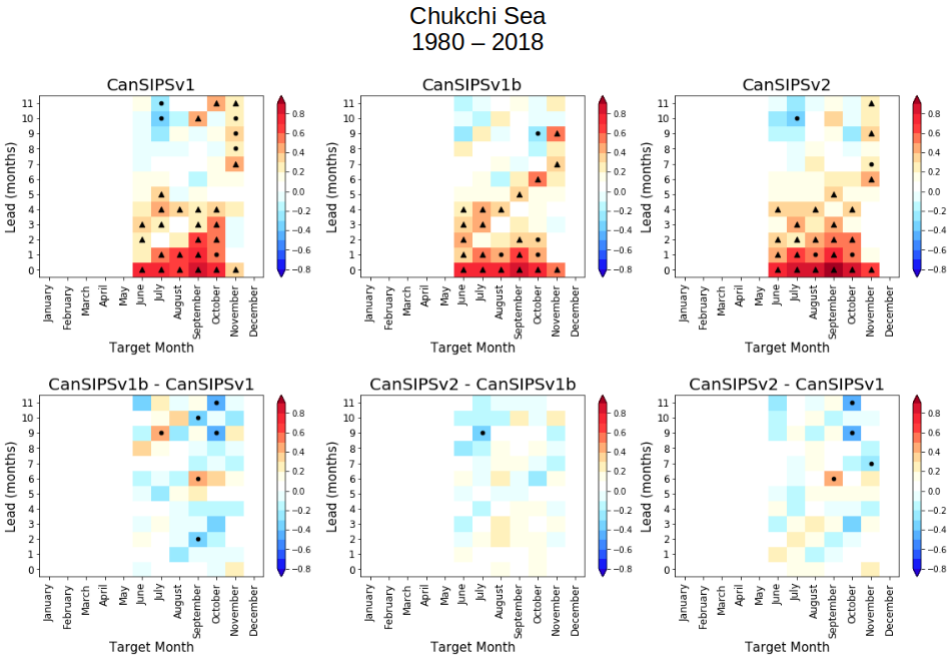

Figure A.8: Operational skill of CanSIPS in the Chukchi Sea.

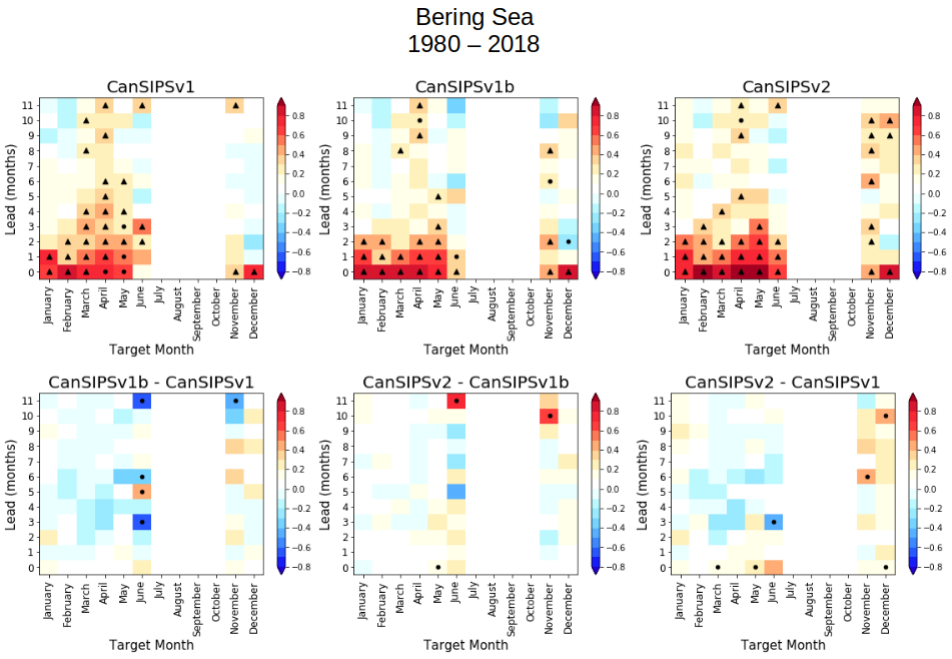

Figure A.9: Operational skill of CanSIPS in the Bering Sea.

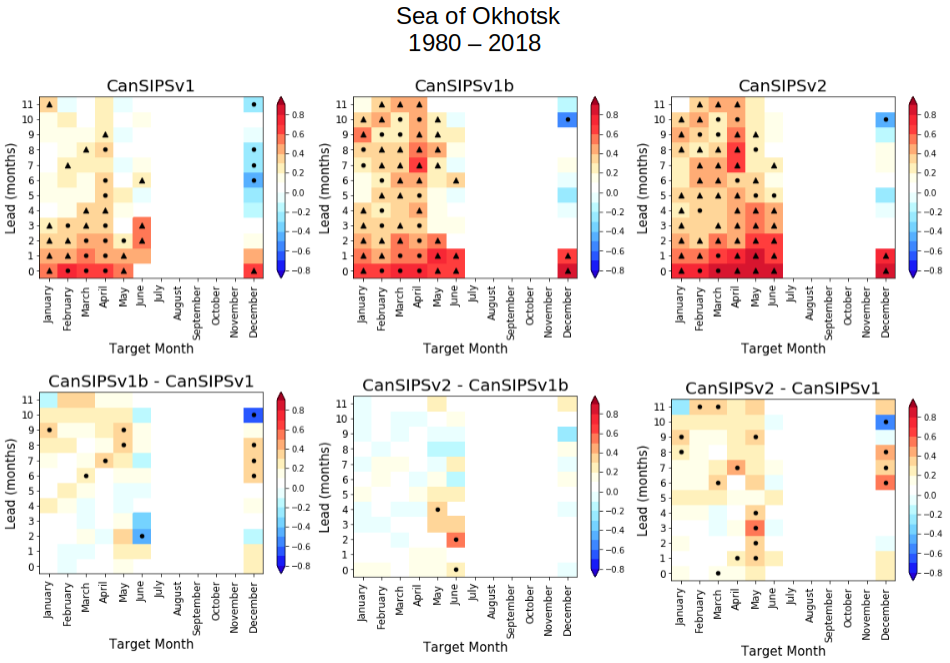

Figure A.10: Operational skill of CanSIPS in the Sea of Okhotsk.

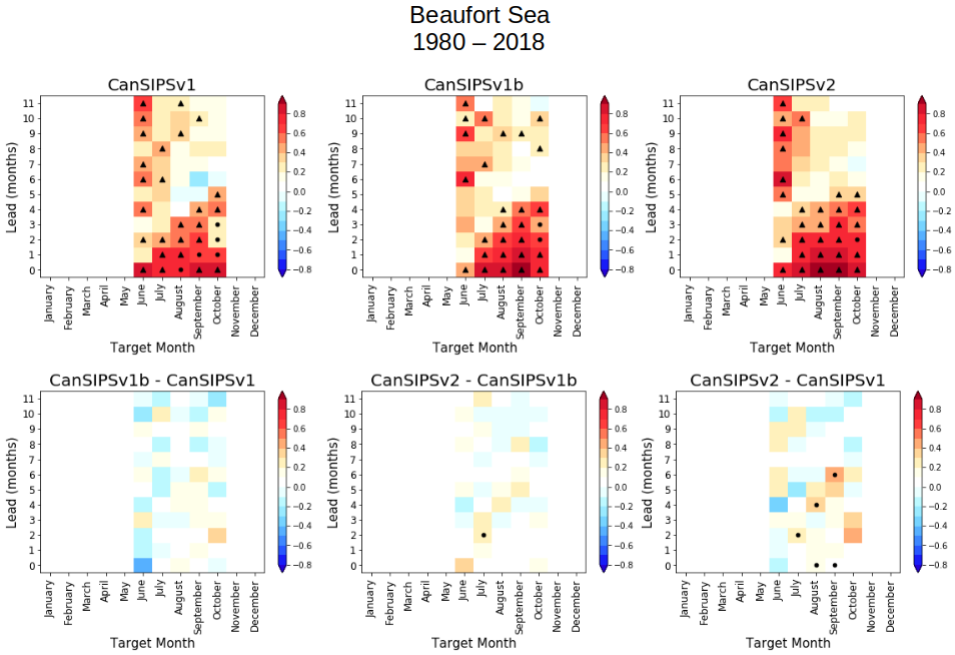

Figure A.11: Operational skill of CanSIPS in the Beaufort Sea.

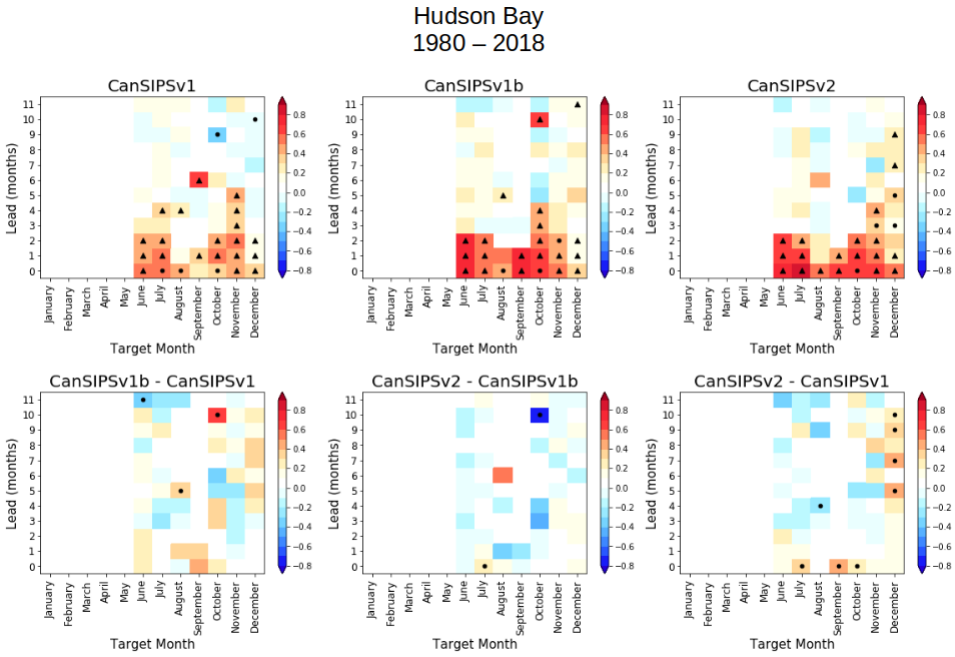

Figure A.12: Operational skill of CanSIPS in Hudson Bay.

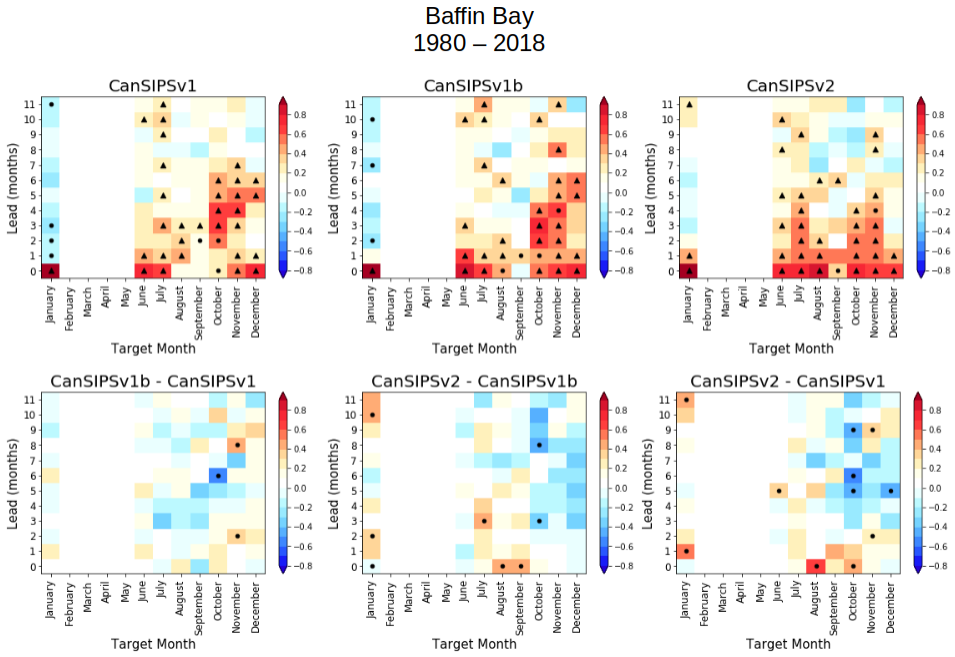

Figure A.13: Operational skill of CanSIPS in Baffin Bay.

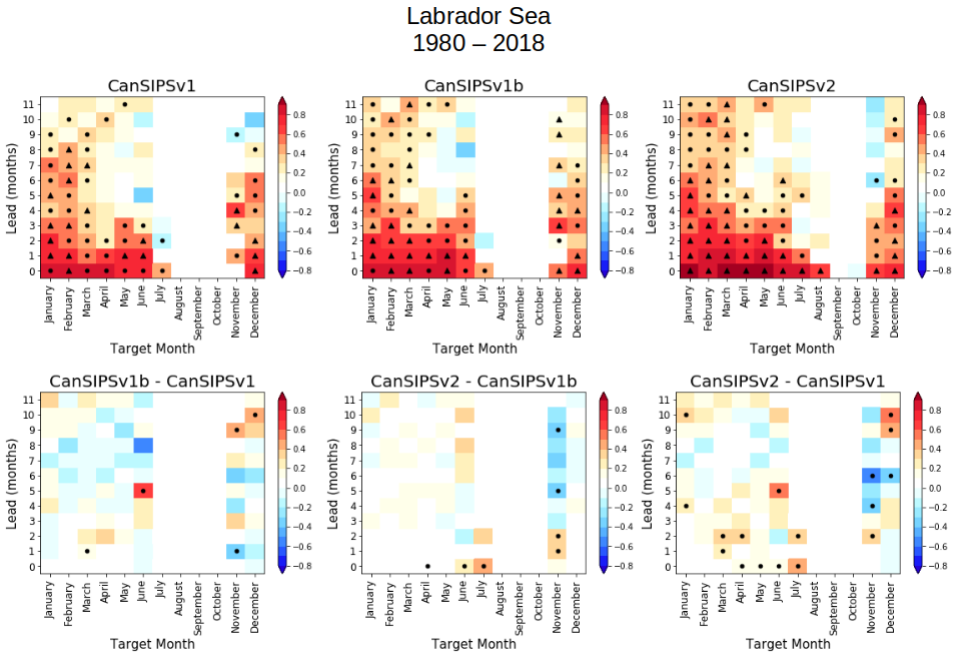

Figure A.14: Operational skill of CanSIPS in the Labrador Sea.

A.3 Regional Autocorrelation Comparisons

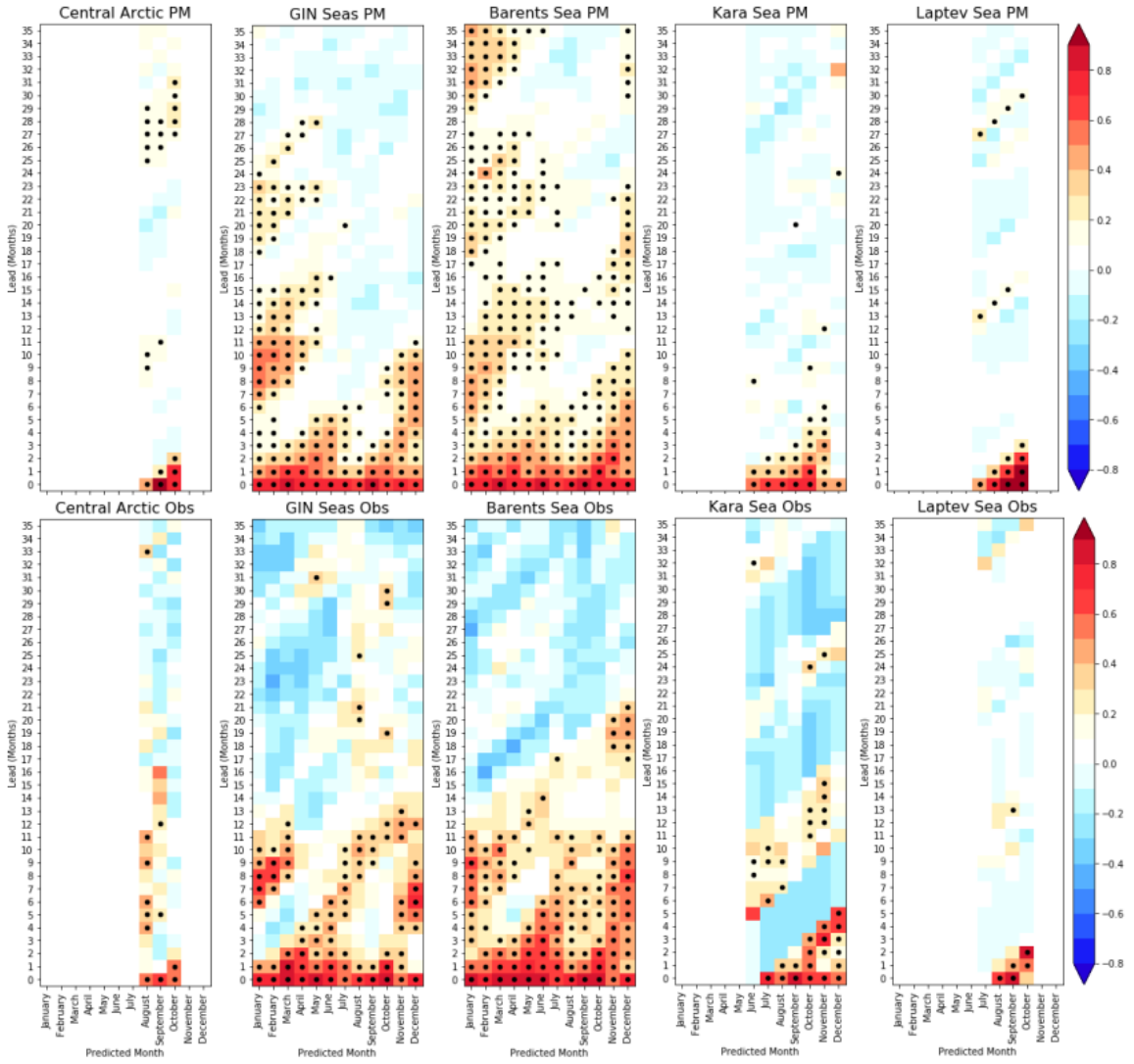

Figure A.15: Comparisons of autocorrelation of the perfect model and detrended observations for the Central Arctic, GIN Seas, Barents Sea, Kara Sea, and Laptev Sea. Dots indicate autocorrelation significant at the 95% confidence level. Autocorrelation is not calculated for target months where the annual standard deviation of observed SIE in the region from 1980-2018, or of the control run SIE, is less than 0.8% of the region's area.

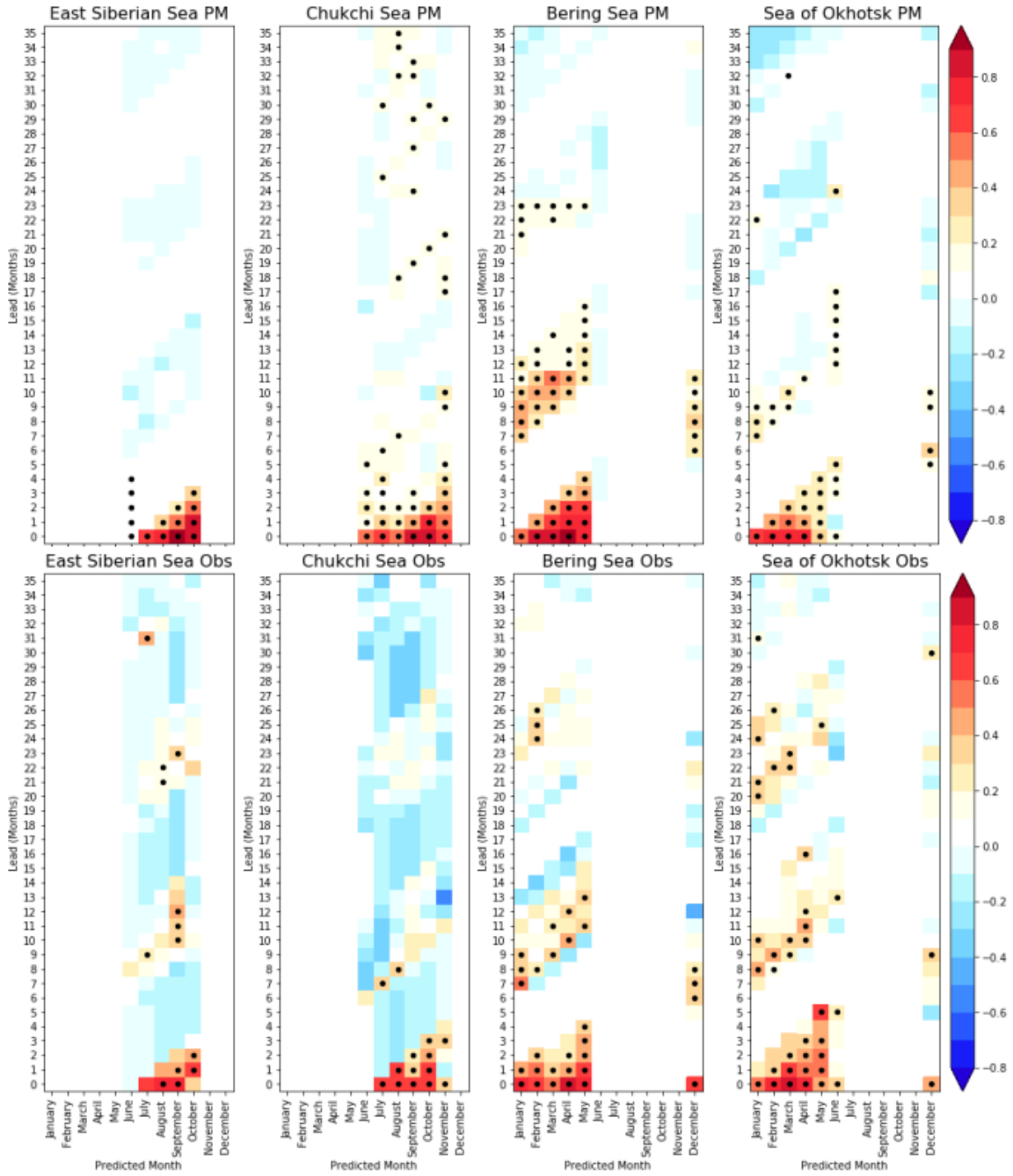

Figure A.16: As in Figure A.15 for the East Siberian, Chukchi, and Bering Seas as well as the Sea of Okhotsk.

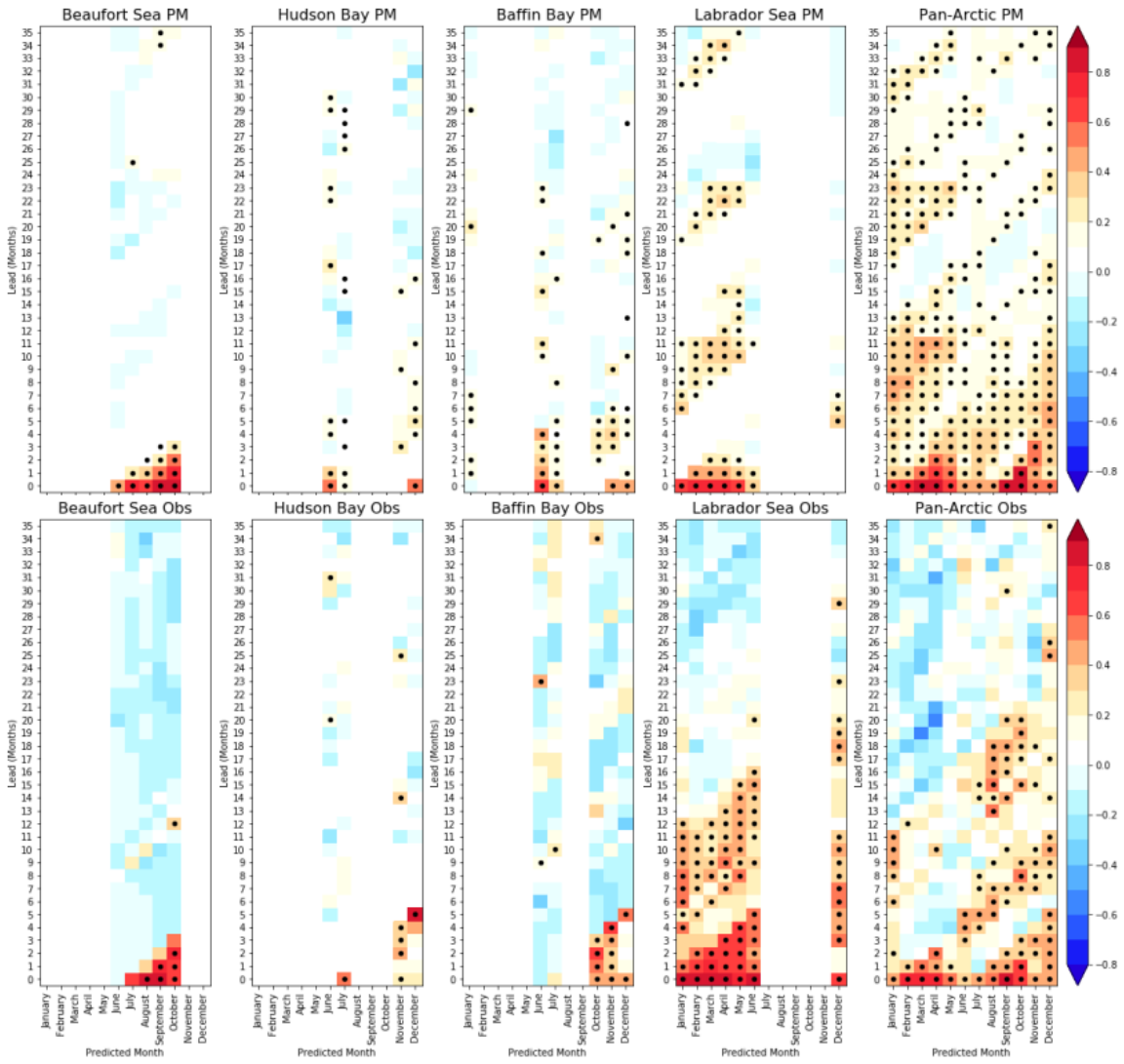

Figure A.17: As in Figure A.15 for the Beaufort Sea, Hudson Bay, Baffin Bay, Labrador Sea, and on the pan-Arctic scale.

Bibliography

Arora, V., Scinocca, J., Boer, G., Christian, J., Denman, K., Flato, G., Kharin, V., Lee, W., and Merryfield, W. (2011). Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases. *Geophysical Research Letters*, 38(5).

Babb, D. G., Landy, J. C., Lukovich, J. V., Haas, C., Hendricks, S., Barber, D. G., and Galley, R. J. (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? *Journal of Geophysical Research: Oceans*, 125(12):e2020JC016796. e2020JC016796 2020JC016796.

Blanchard-Wrigglesworth, E., Bitz, C. M., and Holland, M. M. (2011). Influence of initial conditions and climate forcing on predicting Arctic sea ice. *Geophysical Research Letters*, 38(18):n/a–n/a.

Blanchard-Wrigglesworth, E. and Bushuk, M. (2018). Robustness of Arctic sea-ice predictability in GCMs. *Climate Dynamics*, 52(9-10):5555–5566.

Bonan, D. B., Bushuk, M., and Winton, M. (2019). A spring barrier for regional predictions of summer Arctic sea ice. *Geophysical Research Letters*, 46(11):5937–5947.

Buehner, M., Caya, A., Carrieres, T., and Pogson, L. (2014). Assimilation of SSMIS and ASCAT data and the replacement of highly uncertain estimates in the Environment Canada regional ice prediction system. *Quarterly Journal of the Royal Meteorological Society*, 142(695):562–573.

Buehner, M., Caya, A., Carrieres, T., Pogson, L., and Lajoie, M. (2013). Overview of ice sea data assimilation activities at Environment Canada.

Buehner, M., Caya, A., Pogson, L., Carrieres, T., and Pestieau, P. (2012). A new Environment Canada regional ice analysis system. *Atmosphere-Ocean*, 51(1):18–34.

Bunzel, F., Notz, D., Baehr, J., Müller, W. A., and Fröhlich, K. (2016). Seasonal climate forecasts significantly affected by observational uncertainty of Arctic sea ice concentration. *Geophysical Research Letters*, 43(2):852–859.

Bushuk, M., Msadek, R., Winton, M., Vecchi, G., Yang, X., Rosati, A., and Gudgel, R. (2018). Regional Arctic sea–ice prediction: potential versus operational seasonal forecast skill. *Climate Dynamics*, 52(5-6):2721–2743.

Bushuk, M., Msadek, R., Winton, M., Vecchi, G. A., Gudgel, R., Rosati, A., and Yang, X. (2017). Skillful regional prediction of Arctic sea ice on seasonal timescales. *Geophysical Research Letters*, 44(10):4953–4964.

Bushuk, M., Winton, M., Bonan, D. B., Blanchard-Wrigglesworth, E., and Delworth, T. L. (2020). A mechanism for the Arctic sea ice spring predictability barrier. *Geophysical Research Letters*, 47(13).

Chevallier, M. and Salas-Mélia, D. (2012). The role of sea ice thickness distribution in the Arctic sea ice potential predictability: A diagnostic approach with a coupled gcm. *Journal of Climate*, 25(8):3025–3038.

Cruz-García, R., Guemas, V., Chevallier, M., and Massonnet, F. (2019). An assessment of regional sea ice predictability in the Arctic Ocean. *Climate Dynamics*, 53(1-2):427–440.

Day, J., Hawkins, E., and Tietsche, S. (2014a). Will Arctic sea ice thickness initialization improve seasonal forecast skill? *Geophysical Research Letters*, 41(21):7566–7575.

Day, J. J., Tietsche, S., Collins, M., Goessling, H. F., Guemas, V., Guillory, A., Hurlin, W. J., Ishii, M., Keeley, S. P. E., Matei, D., Msadek, R., Sigmond, M., Tatebe, H., and Hawkins, E. (2016). The Arctic Predictability and Prediction on Seasonal-to-Interannual TimEscales (APPOSITE) data set version 1. *Geoscientific Model Development*, 9(6):2255–2270.

Day, J. J., Tietsche, S., and Hawkins, E. (2014b). Pan-Arctic and regional sea ice predictability: Initialization month dependence. *Journal of Climate*, 27(12):4371–4390.

Dee, D. P. (2011). The era-interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, 137:553–597.

Dirkson, A., Denis, B., Sigmond, M., and Merryfield, W. J. (2021). Development and calibration of seasonal probabilistic forecasts of ice-free dates and freeze-up dates. *Weather and Forecasting*, 36(1):301–324.

Dirkson, A., Merryfield, W. J., and Monahan, A. (2017). Impacts of sea ice thickness initialization on seasonal Arctic sea ice predictions. *Journal of Climate*, 30(3):1001–1017.

Dirkson, A., Merryfield, W. J., and Monahan, A. H. (2019). Calibrated probabilistic forecasts of arctic sea ice concentration. *Journal of Climate*, 32(4):1251 – 1271.

Eicken, H. (2013). Arctic sea ice needs better forecasts. *Nature*, 497(7450):431–433.

Fetterer, F., Knowles, K., Meier, W., Savoie, M., and Windnagel, A. (2017). Sea ice index, version 3.

Flato, G. M. and Hibler, W. D. (1992). Modeling pack ice as a cavitating fluid. *Journal of Physical Oceanography*, 22(6):626–651.

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

Girard, C., Plante, A., Desgagné, M., McTaggart-Cowan, R., Côté, J., Charron, M., Gravel, S., Lee, V., Patoine, A., Qaddouri, A., Roch, M., Spacek, L., Tanguay, M., Vaillancourt, P. A., and Zadra, A. (2014). Staggered vertical discretization of the canadian environmental multiscale (GEM) model using a coordinate of the log-hydrostatic-pressure type. *Monthly Weather Review*, 142(3):1183–1196.

Goessling, H. F., Tietsche, S., Day, J. J., Hawkins, E., and Jung, T. (2016). Predictability of the Arctic sea ice edge. *Geophysical Research Letters*, 43(4):1642–1650.

Guemas, V., Blanchard-Wrigglesworth, E., Chevallier, M., Day, J. J., Déqué, M., Doblas-Reyes, F. J., Fučkar, N. S., Germe, A., Hawkins, E., Keeley, S., et al. (2016). A review on Arctic sea-ice predictability and prediction on seasonal to decadal timescales. *Quarterly Journal of the Royal Meteorological Society*, 142(695):546–561.

Gurvan, M., Bourdallé-Badie, R., Chanut, J., Clementi, E., Coward, A., Ethé, C., Iovino, D., Lea, D., Lévy, C., Lovato, T., Martin, N., Masson, S., Mocavero, S., Rousset, C., Storkey, D., Vancoppenolle, M., Müeller, S., Nurser, G., Bell, M., and Samson, G. (2019). NEMO ocean engine.

Holland, M. M., Bailey, D. A., and Vavrus, S. (2010). Inherent sea ice predictability in the rapidly changing Arctic environment of the Community Climate System Model, version 3. *Climate Dynamics*, 36(7-8):1239–1253.

Holland, M. M., Landrum, L., Bailey, D., and Vavrus, S. (15 Aug. 2019). Changing seasonal predictability of Arctic summer sea ice area in a warming climate. *Journal of Climate*, 32(16):4963 – 4979.

Hunke, E. C., Lipscomb, W. H., Turner, A., Jeffery, N., and Elliott, S. (2010). CICE: the Los Alamos sea ice model documentation and software user's manual version 4.1. *Los Alamos National Laboratory, Los Alamos, NM*, pages 1–115.

Johannessen, O. M., Shalina, E. V., and Miles, M. W. (1999). Satellite evidence for an Arctic sea ice cover in transformation. *Science*, 286(5446):1937–1939.

Kämäräinen, M., Uotila, P., Karpechko, A. Y., Hyvärinen, O., Lehtonen, I., and Räisänen, J. (2019). Statistical learning methods as a basis for skillful seasonal temperature forecasts in europe. *Journal of Climate*, 32(17):5363–5379.

Koenigk, T. and Mikolajewicz, U. (2008). Seasonal to interannual climate predictability in mid and high northern latitudes in a global coupled model. *Climate Dynamics*, 32(6):783–798.

Koenigk, T., Mikolajewicz, U., Jungclaus, J. H., and Kroll, A. (2008). Sea ice in the barents sea: seasonal to interannual variability and climate feedbacks in a global coupled model. *Climate Dynamics*, 32(7-8):1119–1138.

Krikken, F., Schmeits, M., Vlot, W., Guemas, V., and Hazeleger, W. (2016). Skill improvement of dynamical seasonal Arctic sea ice forecasts. *Geophysical Research Letters*, 43(10):5124–5132.

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

Lee, O., Guy, L. S., Johnson, F., Metcalf, V., Eicken, H., Nayokpuk, C., Waghiyi, A., Schreck, M.-B., Irrigoo, C., Plumb, E., et al. (2020). Spring sea ice forecasts and bering sea indigenous marine mammal harvests. *Earth and Space Science Open Archive ESSOAr*.

Lin, H., Merryfield, W. J., Muncaster, R., Smith, G. C., Markovic, M., Dupont, F., Roy, F., Lemieux, J.-F., Dirkson, A., Kharin, V. V., et al. (2020). The Canadian Seasonal to Interannual Prediction System version 2 (CanSIPSv2). *Weather and Forecasting*, 35(4):1317–1343.

Lin, H. and Muncaster, R. (2021). Improving seasonal predictions with gem5 based coupled model. 55th Canadian Meteorological and Oceanographic Society (CMOS) Congress.

Merryfield, W. (2021). Application of CanESM5 to seasonal and decadal forecasting. 55th Canadian Meteorological and Oceanographic Society (CMOS) Congress.

Merryfield, W. J., Lee, W.-S., Boer, G. J., Kharin, V. V., Scinocca, J. F., Flato, G. M., Ajayamohan, R., Fyfe, J. C., Tang, Y., and Polavarapu, S. (2013a). The Canadian Seasonal to Interannual Prediction System. Part I: Models and initialization. *Monthly weather review*, 141(8):2910–2945.

Merryfield, W. J., Lee, W.-S., Wang, W., Chen, M., and Kumar, A. (2013b). Multi-system seasonal predictions of arctic sea ice. *Geophysical Research Letters*, 40(8):1551–1556.

Mu, L., Losch, M., Yang, Q., Ricker, R., Losa, S. N., and Nerger, L. (2018). Arctic-wide sea ice thickness estimates from combining satellite remote sensing data and a dynamic ice-ocean model with data assimilation during the cryosat-2 period. *Journal of Geophysical Research: Oceans*, 123(11):7763–7780.

Namias, J. (1964). A 5-year experiment in the preparation of seasonal outlooks. *Monthly weather review*, 92(10):449–464.

Newman, M. and Sardeshmukh, P. D. (2017). Are we near the predictability limit of tropical Indo-Pacific sea surface temperatures? *Geophysical Research Letters*, 44(16):8520–8529.

Petrich, C., Eicken, H., Zhang, J., Krieger, J., Fukamachi, Y., and Ohshima, K. I. (2012). Coastal landfast sea ice decay and breakup in northern Alaska: Key processes and seasonal prediction. *Journal of Geophysical Research: Oceans*, 117(C2).

Rayner, N. A. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, 108(D14).

Scinocca, J., McFarlane, N., Lazare, M., Li, J., and Plummer, D. (2008). The CCCma third generation agcm and its extension into the middle atmosphere.

Serreze, M. C., Crawford, A. D., Stroeve, J. C., Barrett, A. P., and Woodgate, R. A. (2016). Variability, trends, and predictability of seasonal sea ice retreat and advance in the Chukchi Sea. *Journal of Geophysical Research: Oceans*, 121(10):7308–7325.

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

Sigmond, M., Reader, M., Flato, G., Merryfield, W., and Tivy, A. (2016). Skillful seasonal forecasts of Arctic sea ice retreat and advance dates in a dynamical forecast system. *Geophysical Research Letters*, 43(24):12–457.

Smith, G. C., Roy, F., Reszka, M., Surcel Colan, D., He, Z., Deacu, D., Belanger, J.-M., Skachko, S., Liu, Y., Dupont, F., Lemieux, J.-F., Beaudoin, C., Tranchant, B., Drévillon, M., Garric, G., Testut, C.-E., Lellouche, J.-M., Pellerin, P., Ritchie, H., Lu, Y., Davidson, F., Buehner, M., Caya, A., and Lajoie, M. (2016). Sea ice forecast verification in the Canadian Global Ice Ocean Prediction System. *Quarterly Journal of the Royal Meteorological Society*, 142(695):659–671.

Tietsche, S., Day, J. J., Guemas, V., Hurlin, W. J., Keeley, S. P. E., Matei, D., Msadek, R., Collins, M., and Hawkins, E. (2014). Seasonal to interannual Arctic sea ice predictability in current global climate models. *Geophysical Research Letters*, 41(3):1035–1043.

Titchner, H. A. and Rayner, N. A. (2014). The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. sea ice concentrations. *Journal of Geophysical Research: Atmospheres*, 119(6):2864–2889.

Tivy, A., Howell, S. E. L., Alt, B., McCourt, S., Chagnon, R., Crocker, G., Carrieres, T., and Yackel, J. J. (2011). Trends and variability in summer sea ice cover in the Canadian Arctic based on the Canadian ice service digital archive, 1960–2008 and 1968–2008. *Journal of Geophysical Research: Oceans*, 116(C3).

Wadhams, P., Lange, M. A., and Ackley, S. F. (1987). The ice thickness distribution across the Atlantic sector of the Antarctic Ocean in midwinter. *Journal of Geophysical Research: Oceans*, 92(C13):14535–14552.

Walsh, J. E., Stewart, J. S., and Fetterer, F. (2019). Benchmark seasonal prediction skill estimates based on regional indices. *The Cryosphere*, 13(4):1073–1088.

Wang, Q. J., Schepen, A., and Robertson, D. E. (2012). Merging seasonal rainfall forecasts from multiple statistical models through bayesian model averaging. *Journal of Climate*, 25(16):5524–5537.

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

Weisheimer, A., Decremer, D., MacLeod, D., O'Reilly, C., Stockdale, T. N., Johnson, S., and Palmer, T. N. (2019). How confident are predictability estimates of the winter North Atlantic Oscillation? *Quarterly Journal of the Royal Meteorological Society*, 145:140–159.

Zhang, J. and Rothrock, D. (2003). Modeling global sea ice with a thickness and enthalpy distribution model in generalized curvilinear coordinates. *Monthly Weather Review*, 131(5):845–861.

Zuo, H., Balmaseda, M. A., and Mogensen, K. (2017). The new eddy-permitting ORAP5 ocean reanalysis: Description, evaluation and uncertainties in climate signals. *Climate Dynamics*, 49:791–811.