Reconstructing Arctic Sea Ice in the Instrumental Era

M. Kathleen Brennan

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading Committee:

Gregory J. Hakim, Chair

Cecilia M. Bitz

Edward Blanchard-Wrigglesworth

Program Authorized to Offer Degree:
Atmospheric Sciences

University of Washington

Abstract

Reconstructing Arctic Sea Ice in the Instrumental Era

M. Kathleen Brennan

Chair of the Supervisory Committee:
Dr. Gregory J. Hakim
Atmospheric Sciences

Arctic sea ice has undergone rapid declines in recent decades. Given the short satellite record (1979–present), disentangling the relative role of natural variability and anthropogenic forcing on recent declines remains an important unresolved problem. In order to acquire a longer, reliable record of Arctic sea ice we employ data assimilation to combine temperature observations and climate model output. This technique results in fully gridded spatial fields of various climate variables throughout the Instrumental Era (1850-present). Specifically, the goal of this research is to reconstruct Arctic sea ice coverage and thickness on both annual and monthly timescales. We first reconstruct Arctic sea ice on annual timescales using an offline approach where each time step is independent from one another. This work reveals larger declines in total Arctic sea ice extent during the early 20th century (1900–1940) than previously estimated. Next, we build a Linear Inverse Model to forecast Arctic sea ice and other climate conditions on monthly timescales. We find that the Linear Inverse Model is able to skillfully forecast Arctic climate conditions during statistically stable time periods and is thus most useful when used as a model emulator. We then embed the Linear Inverse Model into a data assimilation scheme to produce monthly reconstructions of Arctic climate throughout the Instrumental Era.

TABLE OF CONTENTS

LIST OF FIGURES

GLOSSARY

CE: Common Era

CE VALUE: coefficient of efficiency

DA: data assimilation

ETCW: early 20th century warming

EOF: empirical orthogonal function

LIM: Linear Inverse Model

SIC: sea ice concentration

SIE: sea ice extent

SIT: sea ice thickness

SLP: sea-level pressure

TAS: 2m air temperature or surface air temperature

TOS: sea surface temperature

W17SIE: Walsh et al (2017) reconstruction

ZG: 500hPa geopotential heights

ACKNOWLEDGMENTS

I first would like to acknowledge that this work was done on the unceded ancestral lands of Duwamish People who have lived in this beautiful place for time immemorial.

I wish to express sincere appreciation to my advisor Dr. Greg Hakim who provided careful and helpful guidance throughout my entire graduate career. I would also like to thank Dr. Edward Blanchard-Wrigglesworth for providing guidance and understanding of the polar regions, which has helped me become a polar scientist. Dr. Cecilia Bitz also provided endless expertise and advice on how to apply data assimilation to sea ice variables, and also served as a role model to me. I would also like to thank Dr. Eric Steig for serving on my committee and Dr. Luanne Thompson for her serving as my Graduate School Representative. Both Dr. Steig and Dr. Thompson have also served as role models during different times of my graduate career and I thank them for their kindness and enthusiasm.

I'd also like to express gratitude to all the faculty in the Department of Atmospheric Sciences. I feel like I was able to build a relationship with most all of you which has greatly contributed to my graduate career. A special thanks to Dr. Abigail Swann, Dennis Hartman, David Battisti, Peter Rhines, Dale Durran, and Dargan Frierson who I had more regular contact with. Finally I would like to thank the office and administrative staff in the department who have helped keep me organized, paid, and resourced throughout my time at the University of Washington. Particularly Hettie Scofield, Jennifer Siembor, Erica Coleman and David Warren.

Next thanks to all the other graduate students and postdocs I interacted with during my time in the department. You all are an amazing bunch and I surely wouldn't have made it through without you. To members of my cohort for being there every step of the way,

particularly: Lucas Zeppetello, Robin Clancy Hamid Pahlavan and Yue Dong. To all you postdocs and older grads who mentored me in a big way: Luke Parsons, Dan Amrhein, Lettie Roach, Sara Sanchez, Kinya Toride, Nils Hutter, Maddie Smith, Isabel McCoy, and Marysa Lague. Members of the Hakim group who indulged me in countless discussions of data assimilation: Robert Tardif, Jessica Badgeley, Gemma O'Connor. Members of the Ice and climate group for teaching me about sea ice and python: Andrew Pauling, Molly Wieringa, Ellen Koukel, and Zac Espinosa. Officemates past and present for being there through the daily grind: Andre Perkins, Noah Brenowitz, Lindsey Taylor, and Anna Black.

Finally, I would like to thank my family and friends outside the university for their endless support. My parents, Bill and Jackie, for instilling a deep sense of curiosity in me from early age. My siblings, Bobby and Kiki, for serving as role models. My friends, Rosa, Sarah, Ori, Gabrielle, Toko, Jaya, Emily, Donna, Kate, Lian, Braydon, Ister, Carolyn, Erin, Katie, Kyle, Ty and so many more for the many many pep-talks, adventures and love that kept me going. Finally I have to thank my partner Oli for keeping me sane, making me take breaks, making me laugh, providing endless love and support, meals, and for believing in my ability as a scientist more than me sometimes. I promise I'm working on that. Thanks for making life brighter.

DEDICATION

to the next and future generations.

for Emmy, Eva, Mika, Quinn, Kai

"There are no passengers on Spaceship Earth. We are all crew."

-Marshall McLuhan

Chapter 1

INTRODUCTION

1.1 Dissertation Motivation

Arctic sea ice is one of the most rapidly changing components of the climate system, affecting surface albedo and modulating ocean–atmosphere interaction through surface fluxes. Large declines in sea ice can impact local ecosystems, human communities, and the global climate system (Meier et al., 2014). Documenting and understanding decadal–centennial variability in sea ice is limited by the availability of high-quality observations, which are only spatially and temporally complete during the satellite era (1978–present, Fetterer et al. (2017)). Furthermore, given the presence of strong radiative forcing during this period, it is difficult to partition the relative role of natural variability (e.g. Ding et al. (2017); England et al. (2019)) and radiative forcing (Notz and Marotzke, 2012) on the rapid Arctic sea ice decline observed in the satellite record. In order to estimate the natural variability of sea ice, a longer record is needed.

To date, researchers have relied on three main sources to study pre-satellite Arctic sea-ice conditions and variability: modeling, direct observations, and proxy records. Modeling past Arctic sea ice provides spatially and temporally complete information, yet models are typically not constrained to observations. The Coupled Model Intercomparison Project, phase 5 (CMIP5)–Paleoclimate Modelling Intercomparison Project phase 3 (PMIP3) represents a significant modeling effort for understanding climate conditions during the last millennium. Arctic sea-ice concentration is included in CMIP5–PMIP3 Last Millennium simulations spanning 850–1849 CE (Schmidt et al., 2011; Taylor et al., 2012) from eight modeling centers. Atwood et al. (2016) used the CMIP5–PMIP3 Last Millennium simulations to investigate the global cooling observed between 1200–1850 CE and found a small contribution of Arc-

tic sea-ice changes to this cooling via surface albedo feedbacks. However, the authors also point out some key differences between paleoclimate proxy records and model results, and note that large uncertainties in volcanic forcing and insolation over the last millennium prescribed in these model simulations could have a large role in cooling over this time period. Furthermore, Goosse et al. (2013) found that when simulating Arctic sea ice, the same model biases that exist in the modern era also exist when simulating Arctic sea ice of the past. In an attempt to circumvent some of these biases, Schweiger et al. (2019) forced Pan-Arctic Ice Ocean Modeling and Assimilation System (PIOMAS, Schweiger et al. (2011); Zhang and Rothrock (2003)), a sea-ice–ocean model, with atmospheric reanalysis data and observational data in order to reconstruct sea-ice volume from 1901–2010. This approach is sensitive to the atmospheric reanalysis used to force the model (Lindsay et al., 2014), and the initial conditions.

Walsh et al. (2017, 2019) provide a spatially complete reconstruction of Arctic sea-ice coverage from 1850–2013 CE at monthly resolution based on a synthesized and curated database of historical observations of sea ice. The Walsh et al. (2017, 2019) reconstructions of sea-ice extent are based on a priority ordering procedure of the observational datasets available, along with analog and linear interpolation methods to spatially infill regions and times when observations are not available. In particular, before around 1953, observations during winter are especially sparse, and results during that time mainly reflect the infill strategy (Brennan et al., 2020).

In order to extend the Arctic sea-ice record prior to the historical period, various sea-ice–sensitive proxies have been proposed and evaluated (de Vernal et al., 2013). For example, biogenic proxies such as IP_{25} (e.g., Belt et al., 2007) and sea salt in ice cores (e.g., Wolff et al., 2003) have been used to quantitatively reconstruct Arctic sea ice. However, biogenic and ice core sea-ice proxies show strong regional dependencies (Polyak et al., 2010) and as such have been used almost exclusively for regional reconstructions. Given the intricacies of each sea-ice proxy record, limited spatial coverage, coarse resolution, and dating uncertainties, they are mainly used to complement other proxy types when reconstructing hemispheric-

scale climate conditions (de Vernal et al., 2013). They do however, provide a source for out of sample comparison, which is particularly useful prior to the 20th century.

Models, direct observations, and proxy records all provide useful information about Arctic sea ice of the past, but discrepancies still exist between each of these sources. Data assimilation (DA) is a statistical technique that combines sparse observations and spatially complete model output to generate a better estimate of climate fields when compared to using models and observations individually. Using DA to reconstruct climate variability has undergone rapid development in recent years (Dirren and Hakim, 2005; Annan et al., 2005; Goosse et al., 2006; Ridgwell et al., 2007; Widmann et al., 2010; Bhend et al., 2012; Steiger et al., 2014; Hakim et al., 2016; Tardif et al., 2019), though few studies have used DA techniques to reconstruct sea ice. Klein et al. (2014) reconstruct Arctic sea ice over a 400-year period in the mid-Holocene with a particle-filter that uses a proxy based reconstruction (based on dinocyst assemblages) to constrain an ensemble of model runs at each time step. Klein et al. (2014) find that for simulations run with data assimilation show better data agreement than free running simulations exhibiting increased sea ice concentration in the Barents and Chukchi seas, which they attribute to differences in atmospheric circulation patterns. Singh et al. (2018) use an ensemble Kalman Filter DA approach to reconstruct various climate fields, including Arctic sea ice, in order to investigate Atlantic multidecadal variability over the last two millennia, and find that during a positive phase of the Atlantic Multidecadal Oscillation, sea ice thins throughout the Arctic and retreats over Greenland, Iceland, and Nordic seas.

Here we apply data assimilation to combine temperature observations during the Instrumental Era (1850–present) with climate models in order to better estimate the evolution of Arctic sea ice during this time period. We begin by reconstructing climate fields on annual timescales using an offline DA approach and then build on this work to obtain monthly reconstructions using online DA. For offline DA, each time step in our reconstruction is temporally independent from the others while in an online framework they are not. For an online framework, after observations are assimilated, the resulting climate state is forecasted

in time using a model emulator and used in the assimilation at the next time step.

1.2 Chapter Summaries

In Chapter 2, we first investigate the relationship between surface air temperature and Arctic sea ice extent and find that it is approximately linear. We then exploit this linear relationship using an offline Ensemble Kalman Filter to assimilate surface air temperature observations with Last Millennium climate model simulations to produce fully gridded annual reconstructions of Arctic sea-ice concentration from 1850–2018. The estimates of total Arctic sea ice extent derived from these reconstructions are skillful when compared to satellite observations with and squared correlation (R^2) of 0.89 between the two. After validating our reconstructions, we explore the variability they exhibit. We find a decline of $\sim$1.25 x10^6 km^2 of total Arctic sea ice extent during the early 20th-century warming (1910–1940). The 25-year trends during this period are $\sim$30-40% smaller than the satellite era (1979–2018) but almost twice as large as previous estimates. Additionally we find that variability of SIE on decadal timescales prior to satellite era is $\sim$40% greater than previously estimated.

Next, we aim to produce monthly reconstructions of Arctic sea using an online DA framework. Given it is too computationally expensive to run a full global climate model for each monthly time step over hundreds of years, in Chapter 3 we explore the use of model emulators to forecast a climate state in time. We build a Linear Inverse Model (LIM) trained on model output to predict Arctic sea ice concentration, thickness and other atmospheric and oceanic variables on monthly timescales. We begin by training a LIM on monthly last millennium simulation output to determine the amount of training and validation data necessary to generate skillful monthly forecasts. The LIM is able to make skillful forecasts when initialized with output from the same model used to produce training data for both in and out of sample time periods. The LIM is unable to make skillful predictions when initialized with model output from a different model than used for training, which indicates that the LIM inherits model bias. We find that LIM forecast skill is related to the variance in the verification data relative to the training data.

In Chapter 4 we embed the LIM developed in Chapter 3 into a data assimilation framework to produce online reconstructions of Arctic sea ice on monthly timescales in the Instrumental Era. At each time step, temperature observations are assimilated into a prior estimate of the climate state and then the LIM is then used to project the climate state to the next time step when more observations are available to assimilate. To test this new framework, we begin by performing perfect model experiment where sparse *pseudo* observations are drawn from a climate model simulation taken to be the target state. We find that in a perfect model framework, the online DA scheme is able to skillfully reconstruct hemispheric mean quantities for all six variables included in the framework with the highest skill for 2m air temperature, sea surface temperature and sea ice thickness. To be continued...

1.3 Skill Metric Definitions

Throughout all of the proceeding chapters we will use we will use squared correlation coefficient and coefficient of efficiency to validate our reconstructions and predictions. The correlation coefficient (R) measures relative phasing of two time series.

$$R = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(v_i - \bar{v})}{\sqrt{\sum_i^n (x_i - \bar{x})^2}\sqrt{\sum_i^n (v_i - \bar{v})^2}} \tag{1.1}$$

Here v is the verification data and x is the state being evaluated (e.g. reconstructed or forecasted value). The square of R describes the percentage of the variance in v that is explained by x. The coefficient of efficiency (CE value, (Nash and Sutcliffe, 1970)), like the correlation coefficient, measures the relative phasing of two datasets, but also takes into account bias and amplitude differences between the two timeseries.

$$CE = 1 - \frac{\sum_i^n (v_i - x_i)^2}{\sum_i^n (v_i - \bar{v})^2}. \tag{1.2}$$

This quantity has an upper bound of one and is unbounded in the negative direction. A CE value of one indicates the two datasets are identical, namely exhibiting the same relative phasing, amplitude of signal and no bias. A CE value of zero occurs when the sum of squared

errors is equal to the variance in the verification data, and negative CE values result from a bias between the two datasets, either in the mean or amplitude of the variability.

Chapter 2

ANNUAL OFFLINE ARCTIC SEA-ICE RECONSTRUCTIONS

A version of this work was published in Brennan et al. (2020).

2.1 Introduction

Arctic sea-ice extent (SIE) has declined drastically in recent decades, yet its evolution prior to the satellite era is highly uncertain. Studies using SIE observations find little variability prior to the 1970s; however, these reconstructions are based on limited data, especially prior to the 1950s. Here we introduce a novel method to reconstruct sea ice cover from 1850–present, using data assimilation (DA), numerical model data, and observations of surface air temperature (TAS).

The longest Arctic sea-ice extent (SIE) observation-based reconstruction combines various sea ice observation types, ranging from satellite data to shipping records, to extend Arctic sea ice records back to 1850 (Walsh et al., 2017). The Walsh et al. (2017) reconstruction of SIE (hereafter W17SIE) shows little variability before 1970, particularly on decadal to multi-decadal timescales. However, the fidelity of this dataset is limited by gaps in observation availability, particularly before 1953 and during winter months (Figure 2.1).

Walsh et al. (2017) uses sea ice observations from a ranked list of 12 different sources. When none of these observation types are available at a given time, temporal interpolation (for a single month of missing data) or analog based methods to fill in missing data (for periods with more than one month missing) are used. In Figure 2.1 we plot the percentage of longitude ocean grid cells with an observation available for each month separated into two seasons from 1850–2013. Before March of 1953 there is very little spatial coverage of sea ice observations in the winter months (September–March). Data coverage in the summer

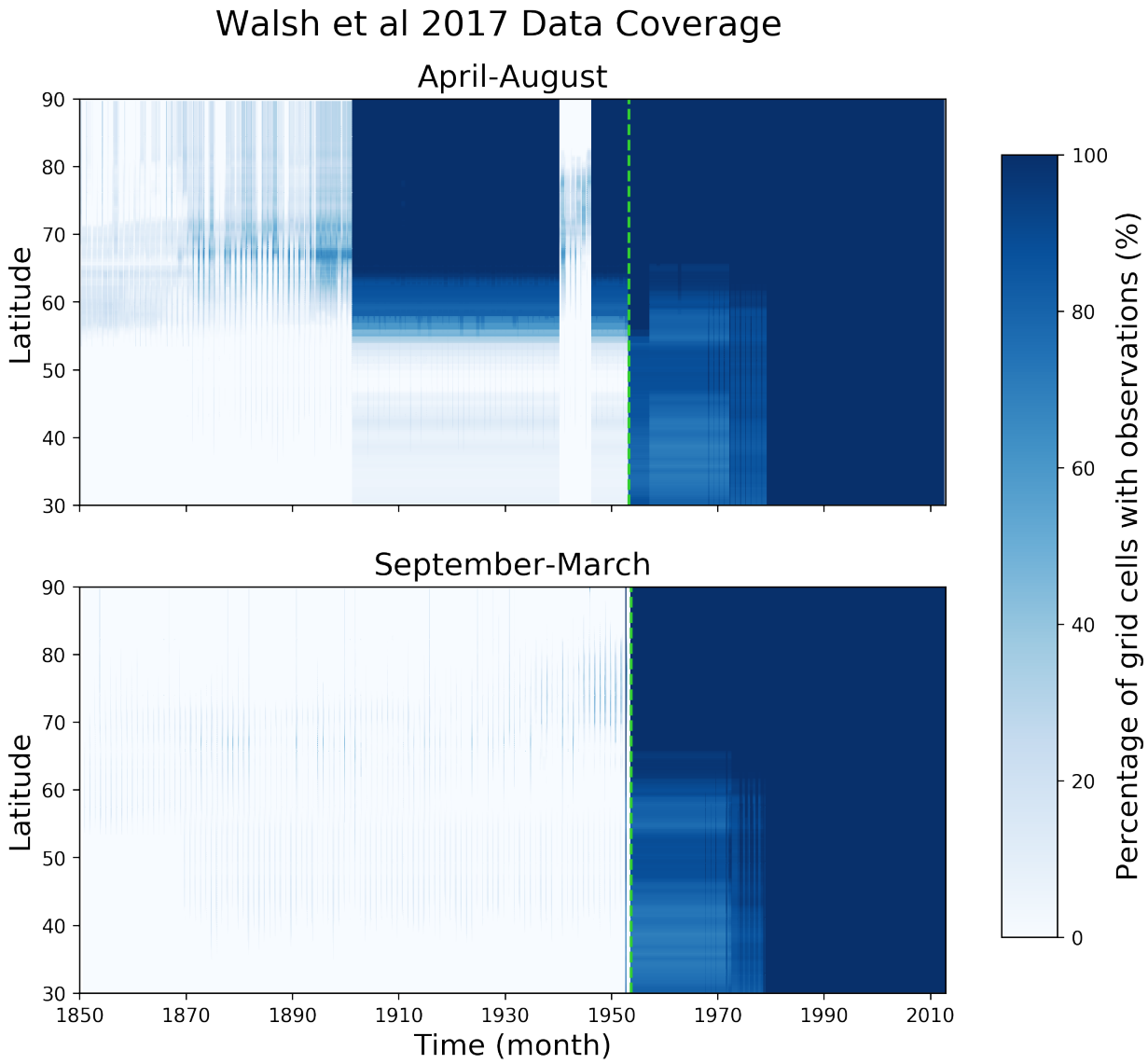

Figure 2.1: Shown is the data availability incorporated into the (Walsh et al., 2017) Arctic sea ice record separated by two seasons over time. The color indicates the percentage of ocean longitude grid cells with an observation available at each latitude for each month. The vertical green lines indicate April 1953 and September 1953 respectively.

months (April–August) is also very low (<40% on average) before May of 1901 and then returns to full coverage intermittently between 1902–1953.

Although direct observations of sea ice are limited in space and time, instrumental observations of TAS are much more abundant. Polar, hemispheric and global mean TAS, both in observations and climate models, are known to be tightly coupled to sea ice variability on annual and longer timescales (e.g. Gregory et al. (2002); Armour et al. (2011); Mahlstein and Knutti (2012); Olonscheck et al. (2019)). Observations show that global-mean TAS was relatively stable between 1850–1900 (Morice et al., 2012), which may explain low decadal sea ice variability in W17SIE during this time period. However, during the early 20th century (1900–1940), an anomalous warming event is well documented across Northern hemisphere high latitudes (e.g. Hegerl et al. (2018)). The magnitude of this early 20th century warming (ETCW) was largest during winter months (Overland et al., 2004; Semenov, 2007) and similar in spatial structure to that observed in the late 20th century.

Despite the TAS anomalies being comparable between the late 20th century and the ETCW, W17SIE shows much less decline during the ETCW than during late 20th century, with one period of decline of $\sim$0.5 x 10^6 km^2 between 1920–1945 (see below). The peak loss in W17SIE also lags the period of largest ETCW temperature anomalies seen in observations, which together with the modest decline in SIE would suggest a weak relationship between temperature and sea ice during the ETCW. In this paper we first investigate the relationship between temperature and sea ice during the Instrumental Era using satellite observations, reanalysis, and W17SIE. Then we use a DA framework to construct a new independent Arctic sea ice reconstruction using more abundant TAS observations. We then explore the decline of sea ice during the ETCW in the new reconstruction, and compare the ETCW decline to that observed and reconstructed in the satellite era.

2.2 Temperature and sea ice in the Instrumental Era

We analyze the relationship between Arctic TAS derived from the HadCRUT version 4.6.0.0 dataset (HadCRUT, Morice et al. (2012)) which uses no infill or interpolation and SIE from

W17SIE. Because the observation availability used in W17SIE depends strongly on both season and time period (Figure 2.2), we partition the analysis into summer and winter seasons (April–August and September–March respectively) and by time period (pre- and post-1953 for W17SIE, plus the satellite era (1979–2017)). Results show that the relationship between TAS and W17SIE and satellite observations generally agree during 1953–present, but differ greatly before 1953 (Figure 2.2).

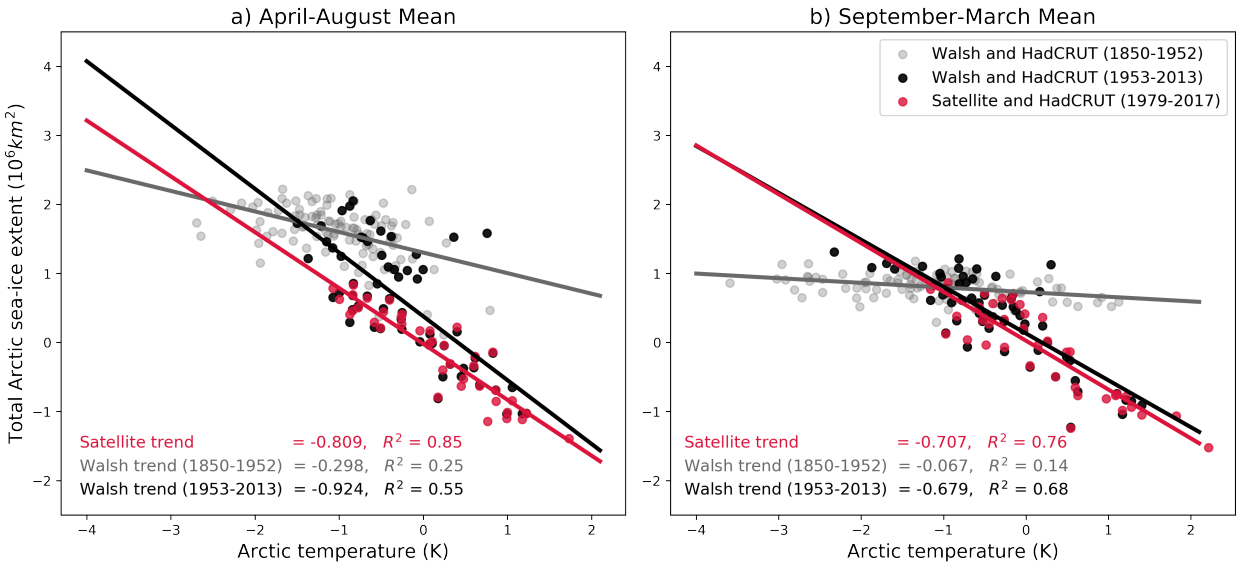

Figure 2.2: Arctic TAS (averaged north of 65°N, derived from HadCRUT) and total SIE in both the satellite data between 1979–2017 (in red) and the (Walsh et al., 2017) dataset between 1850–1952 (in gray) and 1953–2013 (in black). Anomalies are relative to 1979–2013.

Specifically there is a strong linear relationship between TAS and SIE in the satellite record in both seasons as indicated by an R^2-value of 0.76 (September–March) and 0.85 (April–August) in Figure 2.2. Satellite observations also show slightly stronger sensitivity of SIE–TAS in summer months. W17SIE shows a linear relationship for both seasons between 1953–2013, with a similar sensitivity to satellite observations during winter (Figure 2.2b) and a slightly steeper sensitivity compared to satellite observations during summer (Figure 2.2a), but the difference is not statistically different at the 95% confidence level. In contrast,

the earlier part of W17SIE, between 1850–1952, exhibits a much lower SIE–TAS sensitivity in summer relative to the satellite era (Figure 2.2a) and almost no SIE–TAS sensitivity in winter (Figure 2.2b).

There are at least two possible hypothesis for the reduced sensitivity of SIE–TAS in W17SIE before 1953. First, the sensitivity of sea ice may be mean-state dependent, such that in colder, thicker, sea-ice regimes (which may have existed in the Arctic during the late 19th and early 20th century) SIE may be less sensitive to changes in temperature. However, model simulations of sea-ice sensitivity to temperature for different mean states (Armour et al., 2011; Mahlstein and Knutti, 2012) do not support this hypothesis. A second hypothesis is that the apparent reduced sensitivity of sea ice to temperature may simply be due to the fact that there are significantly fewer observations available to W17SIE prior to 1953 (Figure 2.1). When observations are unavailable, W17SIE relies on spatial and temporal infilling, which could dampen variability. Here we explore the second hypothesis, which motivates new methods for reconstructing sea ice.

The fidelity of sea-ice reconstructions has broader implications, as they are used for boundary conditions in reanalysis products. The widely-used sea surface temperature and sea ice concentration HadISST2 product (Titchner and Rayner, 2014) incorporates an earlier version of W17SIE (Walsh and Chapman, 2001), which is based largely on temporal and spatial infilling for the first half of the 20th century. Atmospheric reanalysis during the 20th century such as ERA-20C (Poli et al., 2016) commonly use HadISST2 as a boundary condition. Figure 2.3 shows temperature trends in ERA-20C and the station-based HadCRUT. While HadCRUT shows large magnitude and spatial extent of TAS trends during the ETCW, in some locations comparable to that during the recent satellite-era warming, ERA-20C shows small trends across the Arctic. The NOAA/CIRES 20th Century reanalysis (Compo et al., 2011) shows even larger biases than ERA-20C (see Appendix A Figure A.1).

We postulate that these atmospheric reanalysis biases during the ETCW are strongly influenced by the small inter-annual sea-ice variability in the Walsh and Chapman (2001) sea ice product that serve as boundary conditions. This hypothesis is consistent with Semenov

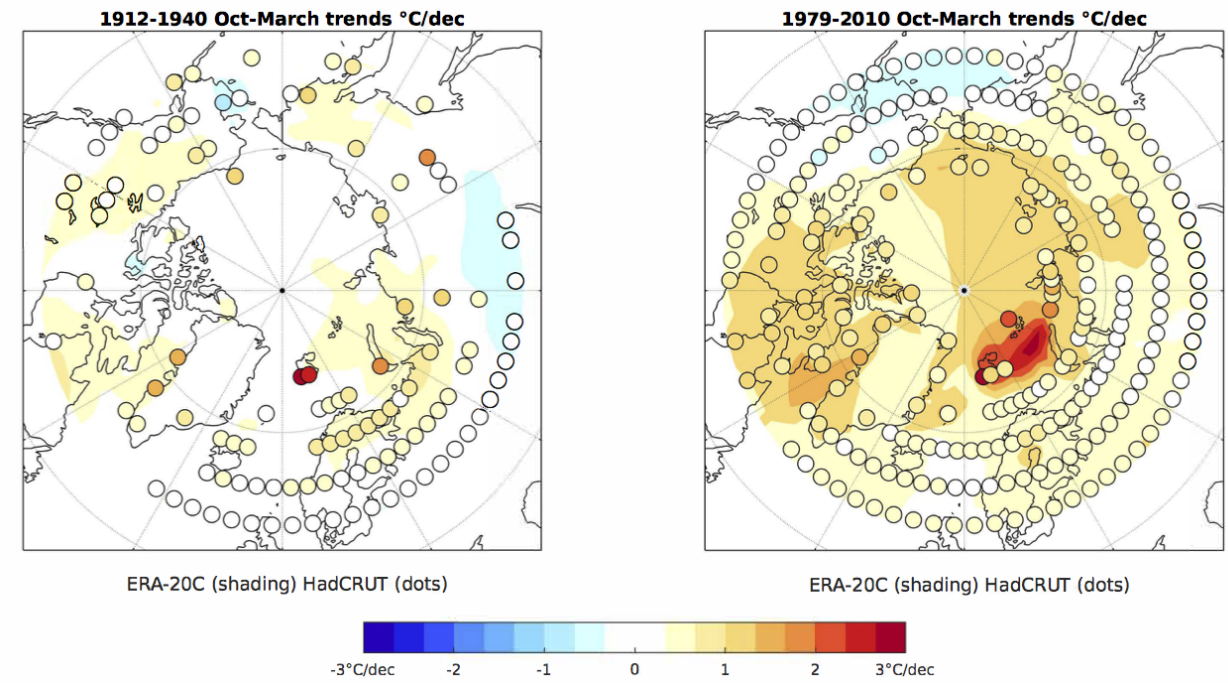

Figure 2.3: Temperature trends from ERA-20C are shown in shading and that from HadCRUT overlaid as shaded dots for both the early 20th century (1912–1940, left) and satellite era (1979–2010, right).

and Latif (2012), who find that the ETCW cannot be simulated with an atmospheric model forced with HadISST1.1 Rayner (2003) boundary conditions. Thus improving sea-ice reconstructions has broad implications, especially for studying high latitude climate variability.

To this end, we exploit the approximately linear relationship between TAS and Arctic SIE evident in Figure 2.2 and in the literature (e.g. Mahlstein and Knutti (2012)) and the fact that TAS is the primary driver of sea ice variability (Olonscheck et al., 2019) to reconstruct sea ice using a DA approach. Other approaches have exploited this relationship to reconstruct sea ice in the 20th century using linear regression models. For example, Connolly et al. (2017) use pre-satellite temperature trends to re-calibrate sea-ice data sources from three regions in the Arctic and find that sea ice retreated after the 1910s and advanced after the mid 1940s, though the magnitude of these changes are small relative to the satellite era. Alekseev et al.

(2016) use the relationship between summer TAS and SIE to reconstruct total summer Arctic SIE with a linear regression model, finding a decline of about half that of the recent event (up to 2012) with a minimum extent in 1936, followed by a recovery that peaked around 1970. The main benefit of the DA approach described here is the use of high quality 2 m air temperature observations with a robust framework for uncertainty quantification (see Section 3). Moreover, the results provide fully-gridded, spatially consistent climate fields that can be used as boundary conditions for models and to probe the dynamics associated with sea-ice variability.

2.3 A new sea-ice reconstruction using data assimilation

2.3.1 DA approach

DA combines information from models with noisy and sparse observations, resulting in a better estimate of climate fields when compared to using models and observations individually. Generally, DA updates a prior estimate, an initial 'best guess', of the climate state with new information from observations. DA allows point-wise observations of temperature to influence broader spatial regions of other climate variables, like sea ice, based on the covariance relationships derived from the prior. The weights given to the prior estimate and observations are inversely proportional to their relative uncertainty.

To reconstruct Arctic sea ice we use an offline (Oke et al., 2002) ensemble Kalman Filter approach to combine Last Millennium climate model simulations (Schmidt et al., 2011; Taylor et al., 2012) with temperature observations. The prior, $\mathbf{x}_b$, is an ensemble of 200 randomly chosen years from these Last Millennium simulations (more details are provided in Section 3.2). The update to this prior estimate derives from the "innovation," which is the difference between annually averaged temperature observations, $\mathbf{y}$, and the prior estimate of these observations $\mathbf{H}\mathbf{x}_b$:

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_b). \tag{2.1}$$

The innovation weight that results in the analysis, $\mathbf{x}_a$, is given by the Kalman gain,

$$\mathbf{K} = \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}, \tag{2.2}$$

where $\mathbf{B}$ is the error covariance matrix of the prior, $\mathbf{R}$ is the error covariance matrix of the observations, and T is the transpose operator. Matrix $\mathbf{H}$ maps the prior to the observations by selecting grid-point data in the prior nearest to the observations. The Kalman gain spreads the new information from temperature observations both spatially and to other climate variables, weighted by the relative uncertainty of each. We sample temperature observations from instrumental data every 10° latitude and longitude, chosen to ensure that the observation errors are uncorrelated and therefore $\mathbf{R}$ is diagonal. This assumption allows us to use serial observation processing, which assimilates observations one at a time, simplifying implementation of spatial covariance localization as described below.

To solve (2.1), we employ a square-root ensemble Kalman Filter (Whitaker and Hamill, 2002), which updates the ensemble mean and the perturbations from the ensemble mean separately. The Kalman gain in the update equation for the ensemble perturbations ($\tilde{\mathbf{K}}$) is adjusted by a constant α to yield the correct posterior covariance matrix. Therefore, $\tilde{\mathbf{K}} = \alpha\mathbf{K}$, where, for a single observation, i,

$$\alpha = \left(1 + \sqrt{\frac{\mathbf{R}_{ii}}{\mathbf{H}\mathbf{B}\mathbf{H}^{T}{}_{ii} + \mathbf{R}_{ii}}}\right)^{-1}, \tag{2.3}$$

where ii denotes the matrix diagonal entry in the ith row and column.

As is standard practice in ensemble DA, we reduce the effect of spurious long-distance covariances using covariance localization (e.g. Hamill et al. (2001)), applying the Gaspari-Cohn fifth order polynomial function (Gaspari and Cohn, 1999) with a localization radius (the distance from observations set to zero influence) of 15,000 km.

Kalman Filter methods rely on the covariance in the prior ensemble between temperature and the variables of interest (here, sea ice concentration, SIC). Climate models tend to underestimate the sensitivity of Arctic sea-ice loss to temperature (Stroeve et al., 2007; Winton, 2011; Rosenblum and Eisenman, 2017). To address this low-sensitivity bias, we

inflate the sea-ice perturbations from the prior ensemble-means for the simulations used here, MPI and CCSM4 Last Millennium simulations (see Section 3.2), by a factor of 1.8 and 2.6, respectively. The inflation factors are determined empirically by goodness of fit to the observed sea-ice trend during the satellite era. Sensitivity of the results to the localization radius and inflation factor is explored in Section 2.4.2.

Since the Kalman Filter method assumes Gaussian distributions, and SIC has a range of 0–100%, unphysical values below and above this range are adjusted to 0% and 100%, respectively.

2.3.2 Data Sources

A 200-member prior ensemble of both TAS and SIC fields are randomly drawn from fully forced Last Millennium model simulations spanning the years 850–1849 CE (Schmidt et al., 2011; Taylor et al., 2012). This ensemble size was chosen because it was computationally inexpensive and revealed small differences in Arctic SIE reconstructions, R^2 >0.97, from tests performed with a 1000-member prior ensemble. Results using the Community Climate System Model version 4 (CCSM4, Last Millennium simulation (Landrum et al., 2013)) and Max Planck Institute for Meteorology (MPI-ESM-P, Last Millennium simulation Taylor et al. (2012)) models are used to determine the sensitivity of the sea-ice reconstructions to climate-model prior, and thus the sensitivity of the results to model physics and sea-ice–temperature covariance structure. All model output is regridded to a $\sim 2°$x$2°$ grid.

Sensitivity to the choice of instrumental temperature record is tested using three different products: HadCRUT, Berkeley Earth (BE, Rohde et al. (2013a)), and NASA Goddard Institute for Space Studies (GISTEMP, Hansen et al. (2010)). An estimate of the uncertainty in these observations is required when using an ensemble Kalman Filter approach (i.e., $\mathbf{R}$ in Equations 2 and 3), and HadCRUT is the only product that provides uncertainty estimates. Various ways of calculating $\mathbf{R}$ were tested, (see Section 2.4.2), but in order to use all three products, and for simplicity, we use an uncertainty estimate of 0.4 K^2, which is the area-weighted mean error variance provided by HadCRUT.

2.4 Arctic sea-ice reconstructions

We first reconstruct annual Arctic SIC from 1850–2018 by assimilating HadCRUT TAS with a prior ensemble drawn from the MPI Last Millennium simulation. Figure 2.4 shows annual Arctic SIE (total area with SIC greater than or equal to 15%) derived from the gridded SIC reconstructions, satellite observations (Fetterer et al., 2017) and W17SIE. The DA derived timeseries is the mean of 5 independent iterations that each use a different 200 member prior ensemble, in order to take into account the uncertainty due to sampling error. Our reconstruction compares well with satellite observations (Figure 2.4) with an R^2 value of 0.89, detrended R^2 value of 0.43, and coefficient of efficiency of 0.89 between 1979–2017. W17SIE also agrees well with satellite observations, as expected, since it uses satellite data during this time period. The trend during this period is well captured in the reconstructions with a value of -0.052 ± 0.012 x 10^6 km^2/year compared to -0.055 x 10^6 km^2/year in satellite observations. Inter-annual variability is overestimated, with a detrended standard deviation of 0.21 x 10^6 km^2 in the satellite observations and 0.28 x 10^6 km^2 in the reconstruction during 1979–2017.

The most notable feature of our reconstruction before the satellite era is the SIE decline during the ETCW, with a total loss of about 1.25 x 10^6 km^2 between 1910–1940 compared to $\sim$2.0 x 10^6 km^2 lost between 1979–2017 in satellite observations. Between 1930–1950, our reconstruction also shows $\sim$0.5 x 10^6 km^2 less SIE than in W17SIE (see Figure 2.4), and the ETCW minimum occurs approximately eight years earlier than in W17SIE. Between 1850–1900 our reconstruction shows a slow increase in SIE, reaching a maximum just after 1900, as opposed to W17SIE, which shows maximum SIE in the 1960s. Overall, prior to the satellite era our reconstruction shows greater decadal variability compared to W17SIE, which has relatively constant Arctic SIE between 1850–1970. Prior to the satellite era (1850–1979) our reconstruction has a time-series standard deviation of 0.31 x 10^6 km^2 whereas W17SIE has standard deviation of 0.22 x 10^6 km^2. Furthermore, the sensitivity between SIE and TAS in our reconstructions agree well with that in satellite observations throughout the recon-

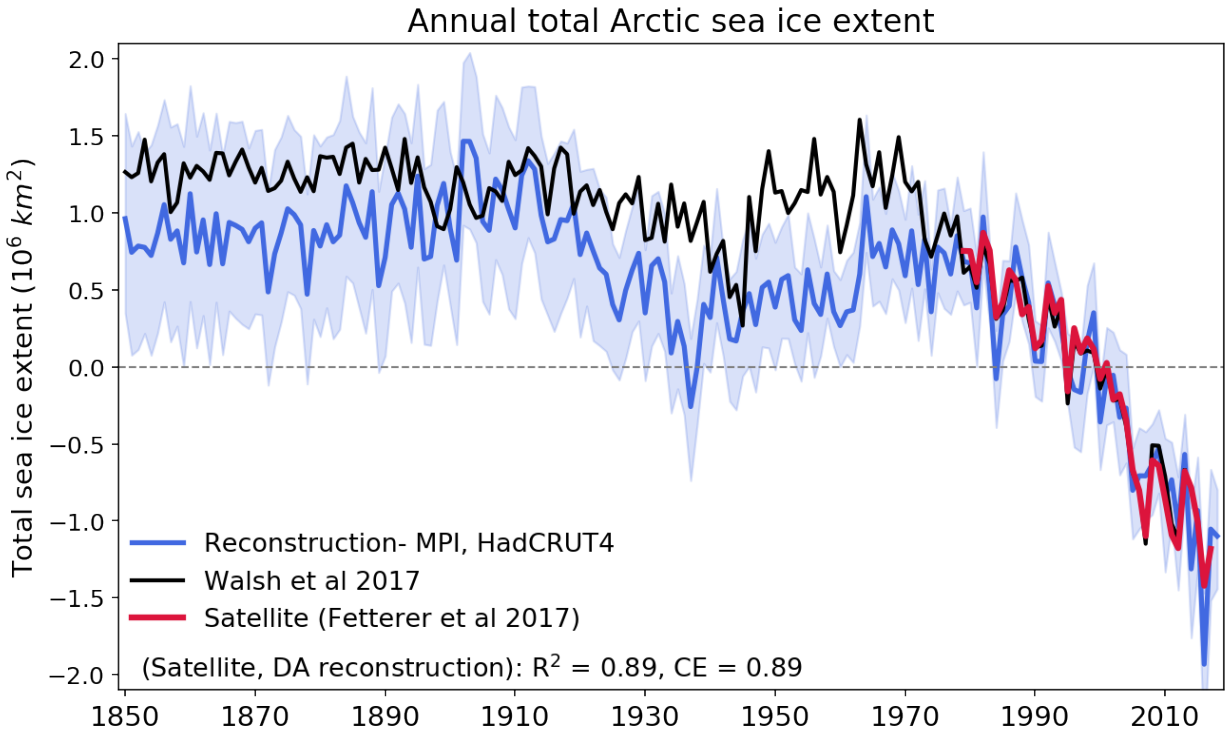

Figure 2.4: Reconstructed Arctic SIE from DA (blue), Walsh et al. (2017) (black), and satellite observations (red). For our reconstructions annually averaged HadCRUT temperature data was assimilated with a prior ensemble drawn from the MPI Last Millennium simulation. Anomalies are centered about 1979–2013.

structed period, whereas W17SIE shows a shift in this sensitivity in the mid 20th century (see Figure 2.5). We hypothesize that a key source of the discrepancy between our reconstruction and W17SIE generally arises from the limited sea ice observations available to W17SIE during the earlier portion of the record. When observations are not available, W17SIE relies on spatial and temporal infilling, which could explain the damped variability observed in their reconstructions between 1850–1953. Our annual-mean reconstructions generally agree with the summer reconstructions of Arctic SIE in Alekseev et al. (2016), except that their reconstruction shows much larger interannual variability in the early 20th century, and a loss of $\sim$2.0 x 10^6 km^2 during the 1970s that is not found in our reconstruction.

2.4.1 Trends and variability

Here we compare the magnitude and significance of Arctic SIE trends during the ETCW relative with the satellite era. In our reconstructions, the SIE decline in the ETCW is shorter lived (∼25–30 years) than that in the satellite era (∼40 years), so we compare the distribution of 25-year trends for both the satellite era and ETCW.

Figure 2.6 shows the distribution of trends calculated for each ensemble member (from reconstructions using both MPI and CCSM4 model priors) for all possible 25-year segments during the satellite era (1979–2017) and the ETCW (1910–1940). The distribution of all 25-year trends between 1979–2017 for both W17SIE and satellite observations are also shown as boxplots below the distributions. For W17SIE, ETCW trends were calculated between 1918–1948 and are also shown as a boxplot (we use a later window for a fair comparison since the minimum SIE occurred 8 years later in W17SIE). The median 25-year trend found in W17SIE during the ETCW of -0.18 x 10^6 km^2/year falls at the 98th and 99th percentiles of our reconstructions based on the MPI and CCSM4 model priors, respectively. Our reconstructions slightly underestimate the mean 25-year trend in the satellite era. Comparing these two time periods in our reconstructions, the satellite era trends are ∼33–38% greater than the ETCW trends.

2.4.2 Sensitivity of results

Our reconstruction of SIE depends on the gridded temperature product (and associated errors), the climate model prior, and sample-error mediation in the DA solution (localization length scale and ensemble variance inflation factor). A range of choices for these aspects has been tested, with details provided below. Overall we find that the choice of observational dataset and model prior make little difference to pan-Arctic indices, but that variance inflation and spatial localization have a larger affect.

Sensitivity to the observations

To test the sensitivity to the choice of assimilated observations, we assimilate three temperature products: HadCRUT, GISTEMP and Berkeley Earth (BE). The original temperature measurements used to create these products are mostly the same, and the main difference is the amount of interpolation (or infill) from grid cells with observations to grid cells with no observations for GISTEMP and BE. Reconstructions using all three temperature products are shown in Figure 2.7. The skill of the reconstruction during the satellite period is slightly higher when HadCRUT is assimilated, as measured by the R^2 values and coefficient of efficiency. Overall the source of the temperature observations has little effect on the overall variability of the reconstructions, with R^2 values with satellite data ranging from 0.82–0.89 for the MPI prior and 0.79–0.89 for the CCSM4 prior (described below). This is expected given the overall agreement among temperature products (e.g. (Rohde et al., 2013a)).

As described above, an observed uncertainty of ($\mathbf{R}$ in Equation 2) 0.4 K^2 is used for all three products for all of these experiments. Other uncertainty estimates tested include: (1) using the annually averaged diagonal elements of the error covariance matrix provided with HadCRUT, and (2) using the variance across all three datasets at each point. Method (1) is ideal, but can only be applied to HadCRUT which has fewer data points than GISTEMP and BE because it does not use interpolation. Further, applying Method (1) shows similar results to using a constant value at all observation locations. For Method (2) the variance across these datasets is very small, given that they often use the same original temperature observations. This led to an over-weighting of observations in the Kalman gain and a SIE reconstruction with an inter-annual variability much larger than the satellite record.

Sensitivity to the prior

We use the MPI and CCSM4 Last Millennium simulations to test the sensitivity of the results to the choice of model prior. Figure 2.7 shows Arctic SIE from these two experiments (note that we use different inflation factors of 1.8 and 2.6 for the MPI and CCSM4 priors

respectively, see below). Results show differences in inter-annual variability, but overall the decadal variability and the timing and magnitude of the ETCW are in close agreement (Figure 2.7). MPI-based reconstructions show slightly higher correlation with satellite data, with R^2=0.82–0.89, as compared to CCSM4-based reconstruction, with R^2=0.79–0.89.

Sensitivity to sample error: prior inflation and localization

The prior ensemble-perturbation inflation factor and prior spatial localization length scale are both determined empirically based on correlations with the trend in Arctic SIE in satellite observations, and correlation and coefficient of efficiency with satellite observations between 1979–2017. A series of experiments are performed with inflation factors ranging from 1.6–2.0 (incremented by 0.1) for the MPI prior and 2.3–2.7 (incremented by 0.1) for the CCSM4 prior. For each inflation factor, reconstructions are performed for localization radii of 5,000, 7,500, 10,000, 15,000, 20,000, and 25,000 km. As the basis of comparison, the trend, detrended variance, correlation and coefficient of efficiency with respect to satellite observations between 1979–2017 are determined across all iterations and ensemble members for each of the 30 parameter combinations (see Figure 2.8 and 2.9).

Increasing the localization length scale and the ensemble inflation of sea ice relative to temperature, both increase the temporal variance and trend in the reconstructions of SIE (Figures 2.8, 2.9). The results indicate that there is not only a trade-off between capturing the trend versus the inter-annual variability, but that there are various parameter combinations that show similar performance. Overall, all experiments described above using HadCRUT observations resulted in R^2 values greater than 0.86 and CE value greater than 0.77 for MPI and R^2 values greater than 0.76 and CE value greater than 0.62 for CCSM4.

With an offline DA approach, all temporal variability in the reconstruction comes from the observations. Increasing the localization radius and the ensemble variance inflation of sea ice relative to temperature both increase the influence of temperature observations, which is realized as larger temporal variability (Figure 2.8 and 2.9). The results show a trade-off between capturing decadal variability versus inter-annual variability. For experiments using

the MPI prior and HadCRUT temperature observations, a localization length scale of 15,000 km leads to the best reconstructed trend for nearly all inflation factors, so we chose to use this localization length scale. Given a localization length scale of 15,000 km, the skill metrics are best for an inflation factor of 1.8. For a prior drawn from the CCSM4 Last Millennium simulation and HadCRUT observations, the same localization length scale of 15,000 km was used and an inflation factor of 2.6 gave the best skill scores. Overall, these experiments show a range of compatible values of localization and ensemble inflation that result in skillful reconstructions relative to satellite observations, and similar reconstructions of sea ice for earlier time periods.

2.5 Conclusions

The relationship between SIE and TAS is approximately linear during the satellite era in observations, but this relationship is much weaker or even absent in W17SIE prior to the 1950s. We speculate that this lower sensitivity of SIE to TAS in W17SIE is due to a lack of high quality sea-ice observations, especially during fall and winter prior to 1953. We also show that 20th century atmospheric reanalyses underestimate the magnitude of the ETCW (1910–1940) in the Arctic. Since previous versions of W17SIE are used as boundary conditions for 20th-century atmospheric reanalysis, we speculate that the low variability of W17SIE could be a reason atmospheric reanalysis do not fully capture the ETCW, but leave exploration of this hypothesis to future work.

We exploit the relationship between TAS and SIC using an ensemble Kalman Filter data assimilation approach to produce a new sea-ice reconstruction. This method combines instrumental temperature observations and model data from Last Millennium simulations to yield skillful Arctic sea-ice reconstructions with annual resolution. Validation against satellite observations yields an R^2-value of 0.89 and coefficient of efficiency of 0.89. Prior to the satellite era, our reconstructions show Arctic SIE loss of $\sim$1.25 x 10^6 km^2 during the ETCW, which is greater than the ETCW loss of $\sim$0.75–1.0 x 10^6 km^2 estimated in W17SIE, yet smaller than the SIE loss during the satellite era of $\sim$2.0 x 10^6 km^2. The

average reconstructed 25-year trends of SIE indicate that the rate of sea-ice loss during the ETCW was about ∼33–38% smaller than the 25-year trends observed during the satellite era.

Overall, these reconstructions show more inter-annual variability in W17SIE during the Instrumental Era with standard deviation ∼40% (∼90,000 km^2) greater between 1850–1979, a significant part due to the ETCW. The ETCW has been ascribed to a combination of anthropogenic forcing and strong natural variability (Delworth, 2000; Wood and Overland, 2009; Fyfe et al., 2013; Beitsch et al., 2014). Here we find that during the satellite era, Arctic sea-ice loss was larger and longer lasting than during the ETCW, which implies that the current decline likely necessitate external anthropogenic forcing, as previous results have shown (Kay et al., 2011; Notz and Marotzke, 2012; Ding et al., 2017). Future work will extend our data assimilation approach to reconstructing seasonal variability and sea-ice thickness to improve our understanding of sea ice during the Instrumental Era.

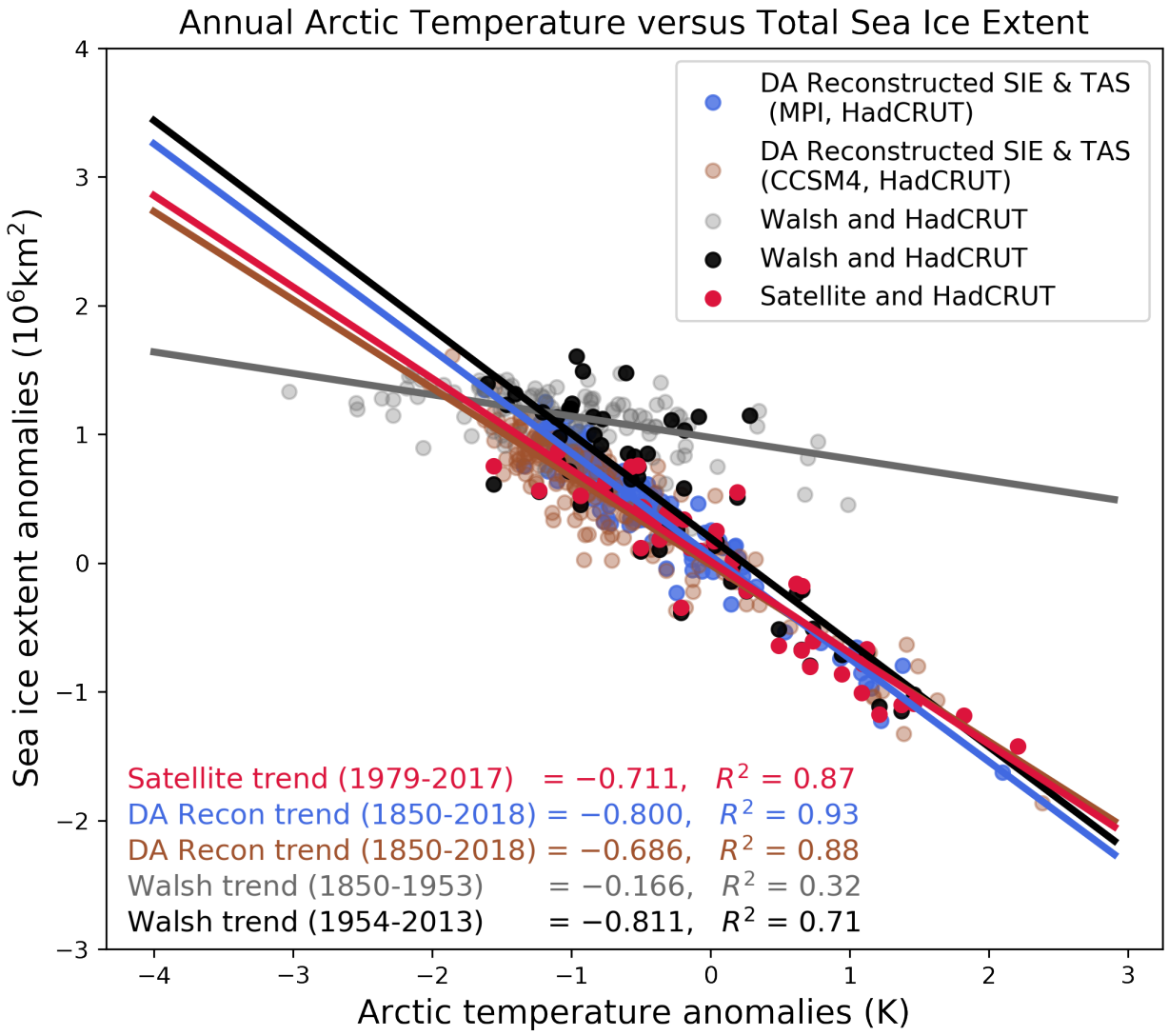

Figure 2.5: Arctic 2 m air temperature (averaged north of 65°N, derived from HadCRUT) and total Arctic sea ice extent in both the satellite data between 1979–2017 (in red) and Walsh et al. (2017) data set between 1850–1952 (in gray) and 1953–2013 (in black). The same relationship from reconstructed SAT and SIE between 1850–2018 are also shown for two reconstructions: one using MPI as the prior and HadCRUT temperature observations (in blue) and one using CCSM4 as the prior and HadCRUT temperature observations (in brown). Anomalies are relative to 1979–2013.

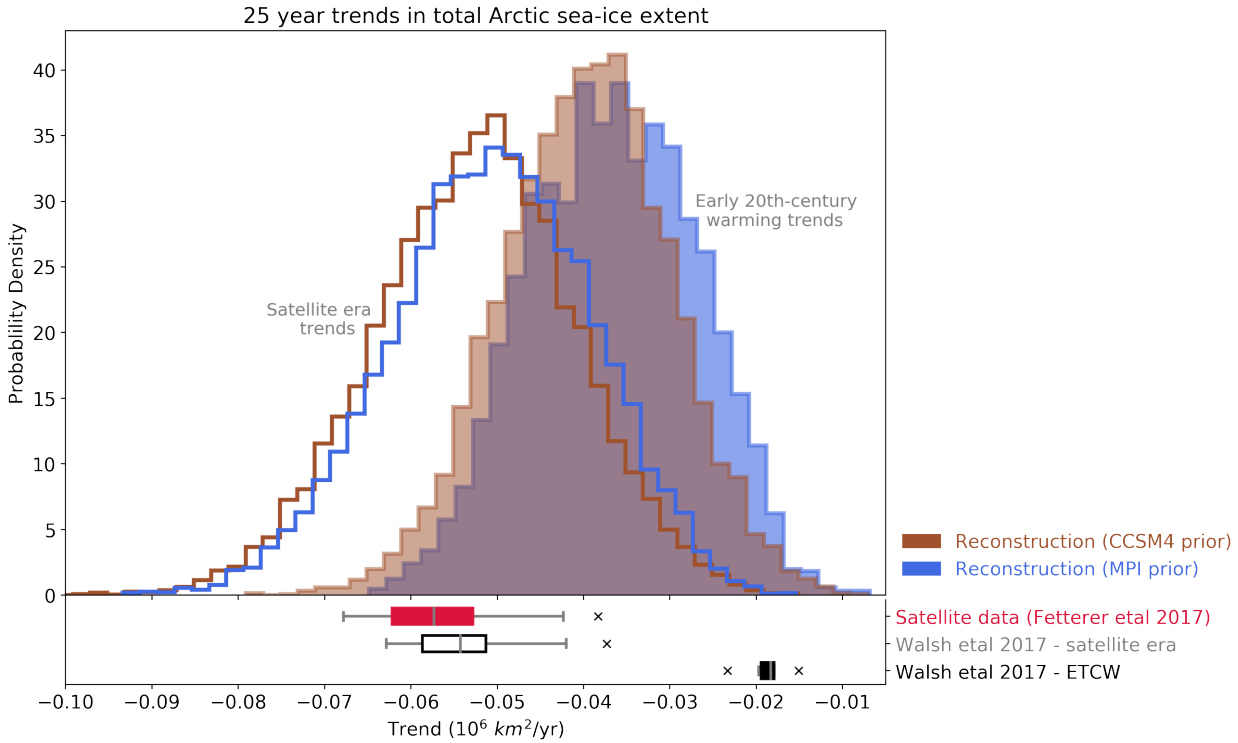

Figure 2.6: The distribution of all possible 25-year trends in Arctic SIE during the satellite era (1979–2017) and ETCW (1910–1940) for 5 prior iterations, each containing 200 ensemble members. The probability density functions show reconstructed SIE trends using MPI as the model prior (blue) and CCSM4 (brown). Below the histograms, the spread of trends calculated in W17SIE (black) and satellite observations (red) are displayed as box plots. The ETCW for W17SIE was calculated between 1918–1948.

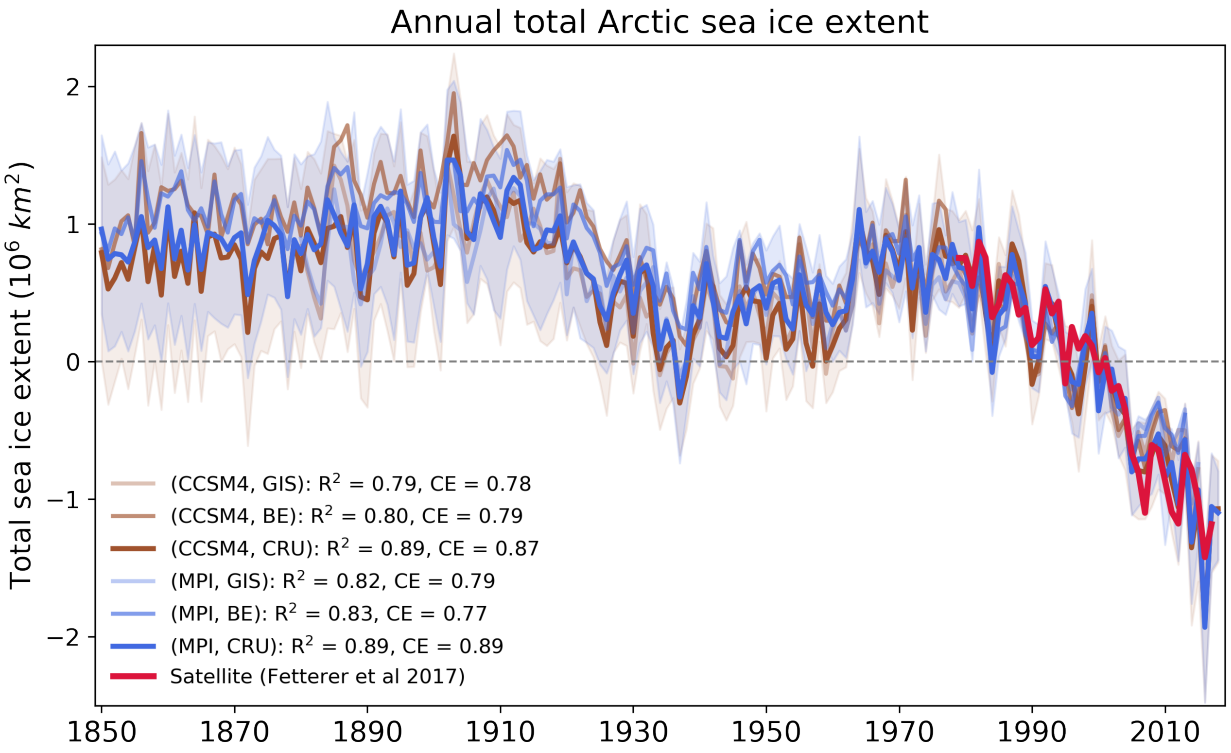

Figure 2.7: Total Arctic SIE reconstructed using priors drawn from two models (MPI and CCSM4 Last Millennium simulations) and three temperature datasets (HadCRUT, GISTEMP, and BE). For all experiments a localization length scale of 15,000 km is used and an inflation factor of 1.8 for MPI and 2.6 for CCSM4. The 97.5 and 2.5 percentiles of the ensemble spread are shown in blue and brown shading.

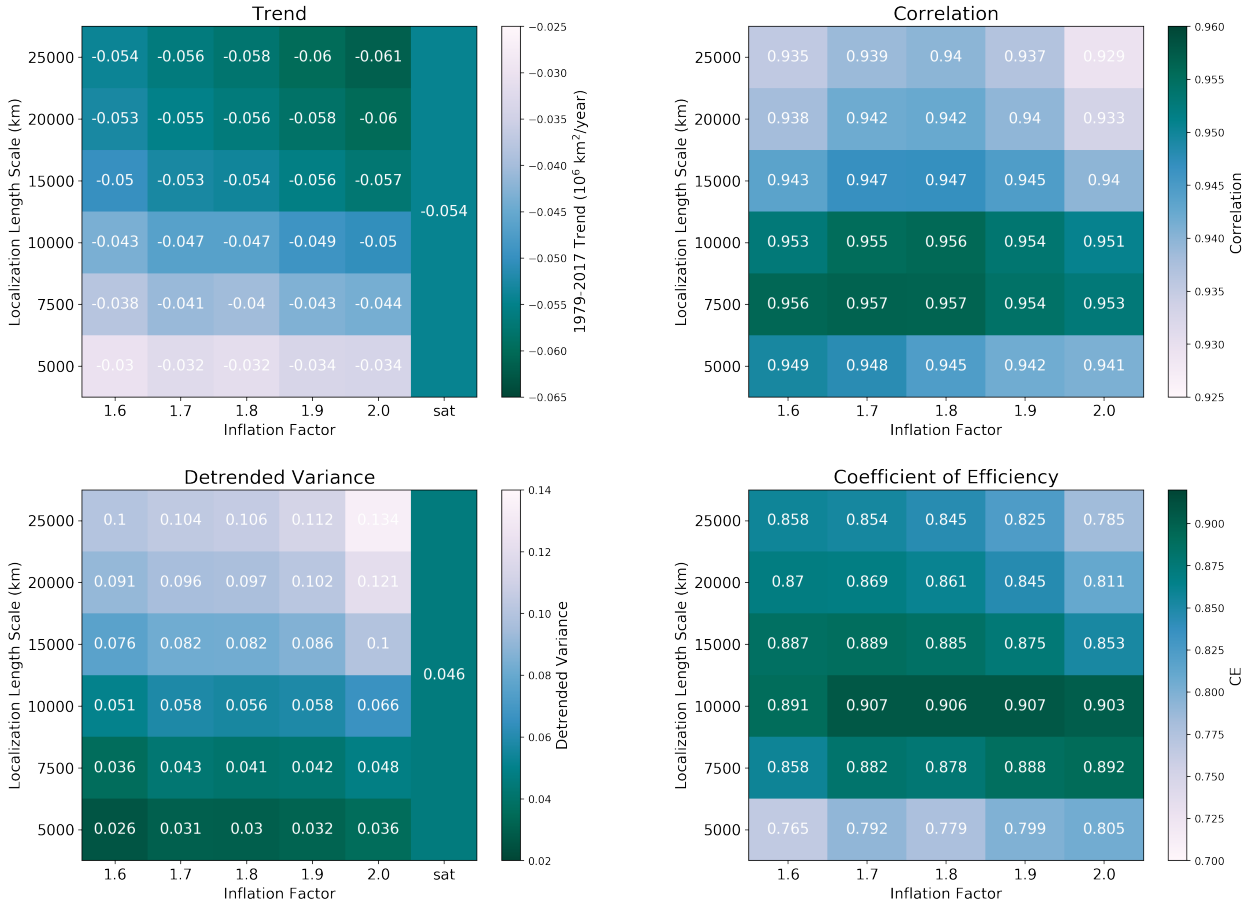

Figure 2.8: Verification statistics for 30 reconstructions performed using MPI as a model prior, HadCRUT observations, and different combinations of localization length scales (y-axis) and inflation factors (x-axis) are shown. Trends and detrended variances during the satellite era are shown in the two boxes on the left and the values observed in the satellite record (Fetterer et al., 2017) are shown in the column on the right. The correlation and coefficient of efficiency of these reconstructions when compared with (Fetterer et al., 2017) are shown in the two boxes on the right.

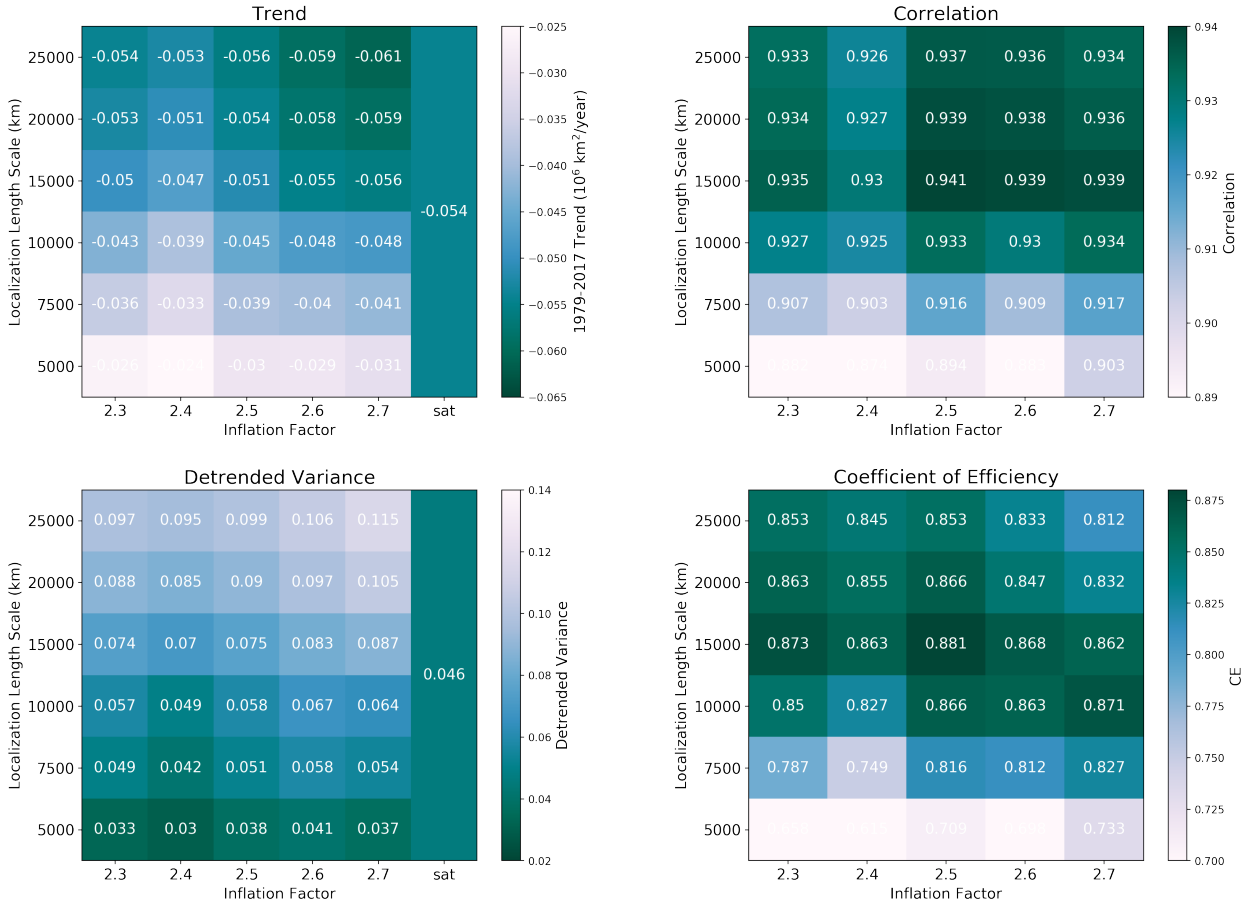

Figure 2.9: Verification statistics for 30 reconstructions performed using CCSM4 as a model prior, HadCRUT observations, and different combinations of localization length scales (y-axis) and inflation factors (x-axis) are shown. Trends and detrended variances during the satellite era are shown in the two boxes on the left and the values observed in the satellite record (Fetterer et al., 2017) are shown in the column on the right. The correlation and coefficient of efficiency of these reconstructions when compared with (Fetterer et al., 2017) are shown in the two boxes on the right.

Chapter 3

MONTHLY ARCTIC SEA-ICE PREDICTION WITH A LINEAR INVERSE MODEL

3.1 Introduction

Large climatic changes observed in the Arctic in recent decades have made prediction of Arctic sea ice on monthly to seasonal timescales increasingly relevant (Jung et al., 2016). Younger, thinner ice in the Arctic (e.g. Maslanik et al. (2007); Lindsay and Schweiger (2015)) has allowed for new shipping routes (e.g. Smith and Stephenson (2013)) and a growing prevalence of industrial and tourist (Hall and Saarinen, 2010) activities in the region. Further, fast changing sea ice conditions present hazards for local communities as conditions becomes increasingly challenging to anticipate and predict (Eicken, 2013).

Arctic sea ice has been shown to contain inherent predictability with high persistence of sea ice concentration (SIC) up to five months (Lemke et al., 1980; Lindsay et al., 2008; Blanchard-Wrigglesworth et al., 2011a), and strong area–thickness coupling provide predictability for even longer timescales (Blanchard-Wrigglesworth et al., 2011b). Blanchard-Wrigglesworth and Bushuk (2019) find that predictability in models is likely robust within a constant climate mean state, but these relationships might not be stationary in a warming climate (Holland and Stroeve, 2011; Holland et al., 2019; Bonan and Blanchard-Wrigglesworth, 2020). Beyond persistence, other major contributors to Arctic sea ice predictability include dynamical advection of sea ice anomalies by mean Arctic circulation patterns (Guemas et al., 2016) atmospheric temperature variability (Olonscheck et al., 2019), and ocean heat flux (Bitz et al., 2005).

Given such promise, there has been a growing effort to predict Arctic sea ice on subseasonal to seasonal timescales, and the Sea Ice Outlook (SIO, Stroeve et al. (2014)) represents

a substantial sea-ice-research community effort to develop and improve Arctic sea ice prediction. Starting in 2008, the SIO has accepted forecasts of September sea-ice conditions from groups around the world, and though much progress has been made, there remains considerable room for improvement. Stroeve et al. (2014) evaluated the skill of SIO forecasts from 2008–2013 and found that overall the skill, regardless of method, was generally poor when sea ice conditions depart from the long-term trend. Further statistical models tend to outperform dynamical models (Stroeve et al., 2014) likely due in part to uncertainty in model physics (Blanchard-Wrigglesworth et al., 2015) and initial conditions (Blanchard-Wrigglesworth et al., 2017) .

Most statistical, data intensive approaches for predicting Arctic sea ice to-date involve predicting pan-Arctic or regional quantities using linear regression techniques (e.g. Drobot et al. (2006); Lindsay et al. (2008); Tivy et al. (2007); Petty et al. (2017). Lindsay et al. (2008) combine the use of observed atmospheric indices and sea ice coverage with ocean and ice thickness fields derived from PIOMAS in order to develop an empirical linear model and find substantially more skill than previous linear regression approaches that include only SIC and atmospheric variables (Drobot et al., 2006). More recently, other data-driven techniques have been applied to Arctic sea ice prediction on seasonal timescales. Andersson et al. (2021) uses convolutional neural networks trained on both climate model and observational data to predict 6 months of SIC fields and are able to outperform a dynamical model in season forecasts. (Hogg et al., 2020) apply a Koopman Mode Decomposition (KMD) to satellite observations of SIC in order to make future predictions of SIC in both hemispheres. KMD is a linear decomposition of the state that is able to capture nonstationary, growing or decaying trends in data.

Here we employ a Linear Inverse Model (LIM) to predict Arctic sea ice on monthly timescales. A LIM is ideal for this application as it uses linearized dynamical modes and persistence estimates from a training dataset, harnessing large contributors to Arctic predictability, to estimate the evolution of a climate state. The LIM framework assumes a dynamical system exhibits stationary statistics and can be modeled as a linear combination

of linearly predictable dynamics plus stochastic noise. Yuan et al. (2016) used a Linear Markov Model trained on sea ice, oceanic and atmospheric variables from reanalysis data to predict Arctic sea-ice concentration on monthly timescales. The Linear Markov Model described in Yuan et al. (2016) is similar approach to the LIM methods, however they did not de-trend the data and only use 35 years (1979–2013) of data to train and validate their model on. A LIM has become a common tool for predicting and probing the dynamics of tropical atmospheric and oceanic variables on seasonal timescales (e.g. Penland and Sardeshmukh (1995); Penland (1996); Penland and Matrosova (1998); Winkler et al. (2001); Alexander et al. (2008); Newman et al. (2009); Cavanaugh et al. (2015); Huddart et al. (2017); Dias et al. (2019); Henderson et al. (2020); Shin et al. (2021)). Perkins and Hakim (2020) build a multivariate LIM for forecasting climate states on annual timescales. We build on their work here with a focus on predicting Arctic sea-ice conditions and other climate fields in the Arctic region on monthly timescales.

In this study we focus on answering two main questions: (1) Can a LIM be used to predict Arctic sea ice and climate conditions on monthly timescales? (2) If so, under what conditions does the LIM work well and under what conditions does it fail to make skillful predictions? To address the first question, in Section 3.1 we use last millennium simulations to train a LIM given these simulations provide a long time series of relatively stationary statistics. Given the absence of a strong external forcing during the last millennium and large availability of data, we are able to optimize the parameters of the LIM for predicting Arctic sea ice coverage and assess its skill. We will focus on 1–3 month predictions, though the LIM framework allows for a prediction of an arbitrary number of months. We then address the second question in Section 3.2 by initializing the LIM with historical simulations and future scenarios from different models and present a metric for predicting when a LIM will fail.

3.2 Methods

3.2.1 Linear Inverse Modelling

A LIM is an empirically determined linearization of a dynamical system about the mean state. LIMs were first applied to geophysical applications for predicting and understanding El Niño Southern Oscillation (Penland and Matrosova, 1994; Penland and Sardeshmukh, 1995; Penland, 1996).

Generally, the tendency of a multivariate state vector $\mathbf{x}$ can be represented as

$$\frac{d\mathbf{x}}{dt} = \mathbf{L}\mathbf{x} + \zeta \tag{3.1}$$

(equation 2 in Penland and Sardeshmukh (1995)). Here, $\mathbf{L}$ is the deterministic dynamical operator which propagates the predictable linear dynamics in time and ζ represents the unpredictable dynamics as white noise forcing in time; correlated in space. Integrating equation (3.1) with respect to time gives an equation to propagate the state from time t to $t + \tau$,

$$\mathbf{x}(t + \tau) = exp(\mathbf{L}\tau)\mathbf{x}(t) + \mathbf{n}. \tag{3.2}$$

Here, $\mathbf{n}$ is a random error vector which arises from the integration of white noise (ζ). Since $\mathbf{L}$ is a matrix, it has an eigenvalue decomposition such that $\mathbf{L}\mathbf{u}_m = \lambda_m\mathbf{u}_m$, where $\mathbf{u}_m$ are the eigenmodes of $\mathbf{L}$ and λ_m are the corresponding eigenvalues. The dynamical system is assumed to be stable, meaning that the forecast modes of $\mathbf{L}$ ($\mathbf{u}_m$) are all damped (exhibiting eigenvalues with negative real parts); however they can also have a complex part which represents oscillating modes. Since the forecast modes are not orthogonal, there can be constructive interference among them resulting in transient anomaly growth despite the individual decay of each mode (e.g. Farrell (1982)).

Next, we use inverse methods to determine $\mathbf{L}$ empirically as described in Penland (1989) by defining a matrix $\mathbf{G}_\tau = exp(\mathbf{L}\tau)$, which is determined through τ-lag covariances,

$$\mathbf{G}_\tau = \mathbf{C}_\tau \mathbf{C}_o^{-1}. \tag{3.3}$$

Here $\mathbf{C}_\tau = \langle \mathbf{x}(t+\tau)\mathbf{x}^T(t) \rangle$ represents the sample covariance of $\mathbf{x}$ for a time lag of τ and $\mathbf{C}_o = \langle \mathbf{x}(0)\mathbf{x}^T(0) \rangle$. Assuming that the state and the error are uncorrelated, we can use equation (3.2) to determine the propagation of the covariance of the state in time

$$\mathrm{cov}(\mathbf{x}(t+\tau), \mathbf{x}(t+\tau)) = \mathbf{G}_\tau \mathrm{cov}(\mathbf{x}(t), \mathbf{x}(t)) \mathbf{G}_\tau^T + \mathbf{N}_\tau. \tag{3.4}$$

Therefore, we can use equation (3.3) to solve for $\mathbf{G}_\tau$ and equation (3.4) to solve for $\mathbf{N}_\tau$ at time zero. Therefore,

$$\mathbf{N}_\tau = \mathbf{C}_o - \mathbf{G}_\tau \mathbf{C}_o \mathbf{G}_\tau^T. \tag{3.5}$$

Given these two quantities ($\mathbf{G}_\tau$ and $\mathbf{N}_\tau$) we are able to propagate the anomalies of the climate state $\mathbf{x}$ in time for a given τ using equation (3.2) and the covariance using equation (3.4). For validation of the LIM here, we consider forecasts for which we take the expectation of equation (3.2) and the noise term becomes zero.

3.2.2 LIM Training Procedure

As mentioned in Section 2.1, an assumption of the LIM is that the state being predicted is stable over long time periods. Thus, to prepare the training data, we first select the training period, linearly de-trend the data and remove the climatological mean for each variable. Given that the trend in time is different across the seasonal cycle (particularly for Arctic sea ice variables, e.g. Serreze et al. (2007)), we determine and remove the linear trend and mean for each month individually at each grid point. We limit our domain to include only data north of 40°N in order to optimize for Arctic prediction.

Given the large number of degrees of freedom in our state vector, it is helpful to reduce the dimensionality of the state using an empirical orthogonal function (EOF) decomposition before calculating the lagged covariance. Each variable is area weighted before the EOFs are

calculated. Then, the leading 50 modes for each variable are retained and normalized by the square-root of the sum of the variance in time such that the variance across the truncated modes of each individual variable sum to 1. Once truncated and normalized, all variables included in the LIM are stacked on top of one another in an array such that

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_{TAS} \\ \mathbf{x}_{PSL} \\ \cdots \\ \mathbf{x}_{SIC} \end{bmatrix}. \tag{3.6}$$

With this truncated state, we determine $\mathbf{G}_\tau$ using Equation 3.3 for $\tau = 1$ month.

Given the large seasonal cycle of Arctic sea ice coverage, we anticipate the need to potentially build a LIM for different months separately. To investigate this, we built a LIM on a single month transition (e.g. train a LIM on January and February alone with monthly means and trends removed) as well as building a LIM using all months (with seasonal cycle and trends removed). However, we found that using only two months during training led to a prevalence of positive eigenvalues in $\mathbf{L}$ indicating that the assumptions of the LIM (e.g. stable state dynamics) were not met. We believe this is because on one month time scales the location of variability during one month is different than that during the next month for sea ice (the location of the sea ice edge changes). In order for the LIM to produce anomalies in locations where there isn't much variability in the training dataset, positive eigenvalues result.

We perform both *in-* and *out-of-sample* validation, which refer to the time period used for validation relative to the training period. For in-sample validation the time period used for validation is also used to train the LIM and for out-of-sample validation the validation time period is not used in training. We will use *intra-model* validation to indicate when data originating from the same model run is used for both training and validation of the LIM and *cross-model* validation to indicate when the LIM is trained and initialized using data originating from different model simulations.

3.2.3 Data Sources

For training our LIM, we use monthly averaged data from the Community Earth System Model version 1 (CESM1) Last Millennium Ensemble (LME, Otto-Bliesner et al. (2016)). We train and validate our LIM on a single ensemble member spanning years 850–1850 CE. Model fields that we include to test their affect on LIM forecast skill include: 2-m air temperature (TAS), sea level pressure (PSL), 500hPa geopotential height (ZG500), sea surface temperature (SST), sea ice thickness (SIT), and SIC. These variables were selected because they have all been associated with sea ice variability and we hypothesize that including each of these variables may contribute positively to the LIM's ability to predict sea ice concentration. Specifically, Arctic sea ice has been shown to exhibit strong coupling with surface air temperature, particularly on longer timescales (e.g., Gregory et al. (2002); Armour et al. (2011); Mahlstein and Knutti (2012); Olonscheck et al. (2019)). PSL is associated with surface winds which can drive sea ice motion and variability (e.g., Rigor et al. (2002)) and similarly ZG500 is associated with the large scale dynamical state of the atmosphere which can also contribute to sea ice variability. Further, sea ice concentration persistence has been shown to be strongly influenced by both SSTs and SIT (e.g. Blanchard-Wrigglesworth et al. (2011a); Bushuk et al. (2015)).

We also initialize and train the LIM using various other models and data products simulating various time periods in order to test the sensitivity of the LIM to different mean states and covariance structures across variables. During the last millennium we use monthly averaged data from both the Community Climate System Model version 4 (CCSM4, Landrum et al. (2013)) and Max Planck Institute for Meteorology Last Millennium simulations (MPI, Jungclaus et al. (2012)), which were run from 850–1850 CE. Both these simulations were part of the the Coupled Model Intercomparison Project, phase 5 (Taylor et al., 2012), Paleoclimate Modelling Intercomparison Project phase 3. During the historical period we use monthly averaged data from the Coupled Model Intercomparison Project, phase 6 (CMIP6, Eyring et al. (2016)) Community Earth System Model version 2 (CESM2), MPI Earth Sys-

tem Model version 1.2 (low resolution, (Mauritsen et al., 2019)), and Geophysical Fluid Dynamics Laboratory Earth System Model version 4.1 (GFDL, Dunne et al. (2020)) simulations which were run from 1850–2014 CE. We also validate the LIM using the European Center for Medium-Range Weather Forecasts Re-Analysis product (ERA5, Hersbach et al. (2020)), which spans 1979–2020 CE (note that ERA5 does not include a sea ice thickness variable). For future simulations we use one ensemble member from the CESM1 Large Ensemble (Kay et al., 2015) which simulates 1920–2100 CE. All variables and from all simulations are regridded onto the native ocean and atmosphere grids from the CESM1 LME simulations.

3.2.4 Skill metrics

To validate our predictions, we use the R^2 and CE value as defined in Section 1.3. To assess the skill at individual grid cells we use Arctic mean root mean squared error (RMSE). The RMSE is defined as

$$RMSE = \sqrt{\frac{\sum_i^n (x_i - v_i)^2}{(n-1)}}, \tag{3.7}$$

where i represents a given forecast, v is the verification data and x is the forecast being evaluated. The RMSE is calculated for each grid cell and then an area-weighted mean is taken across the ocean domain (north of 40°N, land not included). When evaluating the performance of the LIM the forecasts are projected out of EOF space and compared with the verification data (not truncated) in full latitude-longitude space.

3.3 Results

Given that the LIM assumes stable statistics, we expect the LIM to produce skillful forecasts during the Last Millennium when there is less anthropogenic forcing than the Instrumental Era or future projections. We anticipate that the LIM will also perform well when performing intra-model validation given that the covariances in the LIM will match that in the target state. In Section 3.3.1 we will use intra-model experiments to optimize the LIM parameters

and in Section 3.3.2 we will use cross-model validation to quantify under what time periods the LIM produces skillful forecasts. Throughout the results auto regressive model of order one (AR1) forecasts will be used as a baseline the LIM forecast needs to exceed in order to be considered skillful.

3.3.1 LIM optimization and Intra-model performance

To start, we train and validate a LIM during last millennium conditions using intra-model validation with a CESM1 LME simulation, which provides ample data to test the sensitivity of the LIM to various parameters during a period of relatively stable dynamical conditions. Here we outline the experiments used to test the sensitivity of LIM performance to the following parameters: number of training and validation years, number of EOFs included in the truncation, and variables included in training.

To investigate the number of training years and validation years necessary to get consistent and reliable performance metrics of the LIM, we perform two experiments. For the first, we fix the training period from 850–1650 CE (800 years), and vary the number of validation years from 10 to 200 year segments between 1651–1850 CE. All possible non-overlapping validation segments of each length are used. Similarly, we next fix the validation period from 1751–1850 CE (100 years), and then vary the number of training years ranging from 100–900 years for all possible non-overlapping segments between 850–1750 CE. We perform these same experiments using an AR1 and both sets of results for 1 month forecasts are shown in Figure 3.1. For both an AR1 model and a LIM the skill asymptotes to 0.75 and 0.60 respectively as the number of validation and training years increases, which represents the best estimate of the skill. This indicates that a fairly large dataset is necessary in order to train and validate a LIM optimized for monthly prediction. Ideally more than 500 years of training data and 100 years of validation data are needed to get reliable skill metrics.

Next we vary the number of EOFs included in the truncation during the training procedure (Section 2.4) from 5 to 250 for each variable and perform both in-sample (851-1050 CE) and out-of-sample (1651–1850 CE) validation. The percent of variance explained relative to

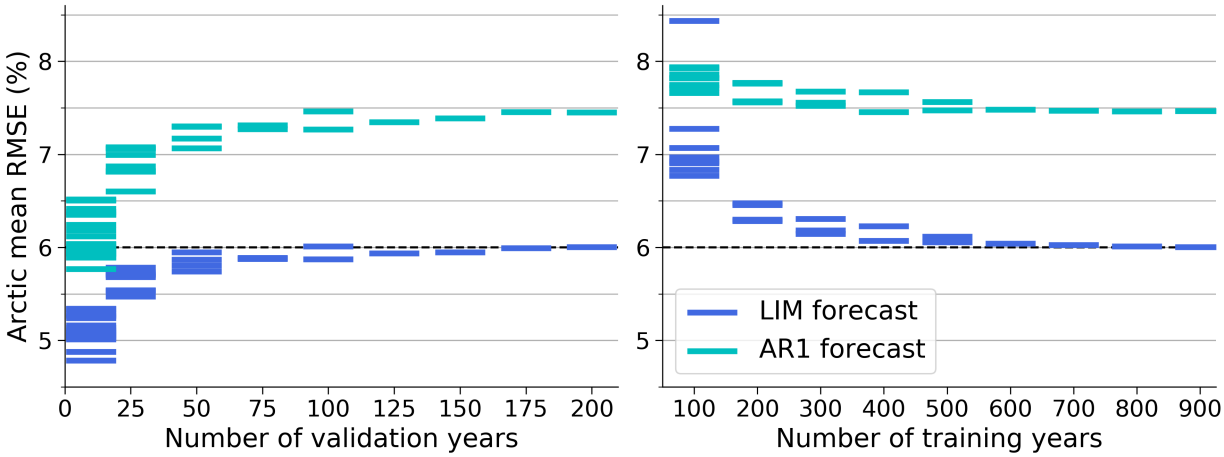

Figure 3.1: Arctic mean root mean squared error (RMSE) for 1 month lag forecasts of sea ice concentration using various number of validation years (left) and training years (right) of monthly data for both LIM forecasts (blue) and an autoregressive model of order one forecasts (turquoise). For both models, a CESM1 LME simulation was used for both training and validation. For the figure on the left, all models were trained using 800 years of monthly data between 850–1650 CE, and the different validation periods of various lengths fell between 1651–1850 CE. For the figure on the right, all models were validated using monthly data between 1751–1850 CE, and the different training periods of various lengths fell between 850–1750 CE. Each line segment represents a different non-overlapping period of validation (left) and training (right) years of a given length indicated on the x-axis.

the number of EOFs retained for each of the six variables in the CESM1 LME simulation data is shown in Figure 3.2. Generally the addition of more EOFs retained for each variable leads to increased forecast skill. For all experiments retaining 10 EOFs for each variable or more, the out-of-sample forecasts outperform an AR1 forecast for lead times up to around 5–6 months. The results for both in- and out-of-sample validation for lead times of one to three months are shown in Figure 3.3. Up to approximately 40 EOFs retained, the in- and out-of-sample forecast skill is nearly identical, after which they deviate. Deviation of in- and out-of-sample forecast skill is one indication of overfitting. As the number of EOFs retained in the training dataset increases, the LIM is able to better capture the training data (in-sample validation), but eventually it becomes overfit and is unable to forecast conditions

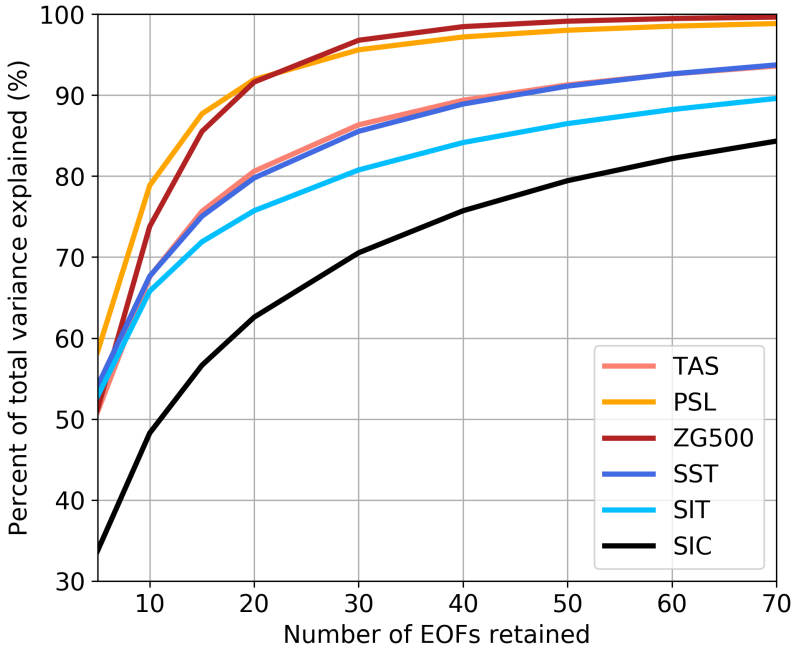

Figure 3.2: Percent of the variance explained for a given number of empirical orthogonal functions (EOFs) retained in the truncation for six variables from a CESM1 LME simulation: 2m air temperature (TAS), sea level pressure (PSL), 500hPa geopotential height (ZG500), sea surface temperature (SST), and sea ice thickness (SIT), and sea ice concentration (SIC).

not present in training (out-of-sample validation). The forecast skill continues to decrease by including more than 50 EOFs per variable (minimum Arctic mean RMSE for out-of-sample validation corresponding with 200 EOFs retained for one month lead times); however, to avoid overfitting, we limit our truncation to 50 EOFs per variable.

Finally, we investigate how including different variables contribute the the predictability of Arctic SIC at different lead times. We consider the role of atmospheric (TAS, PSL, and ZG500), oceanic (SST), and another sea ice variable (SIT) on SIC predictions. Seven different LIMs are trained using: only SIC, SIC plus TAS, SIC plus PSL, SIC plus ZG500, SIC plus SST, SIC plus SIT, and a LIM trained using all six variables. For each LIM we use 800 training years (851-1650 CE), 200 validation years (1651-1850 CE) and 50 retained

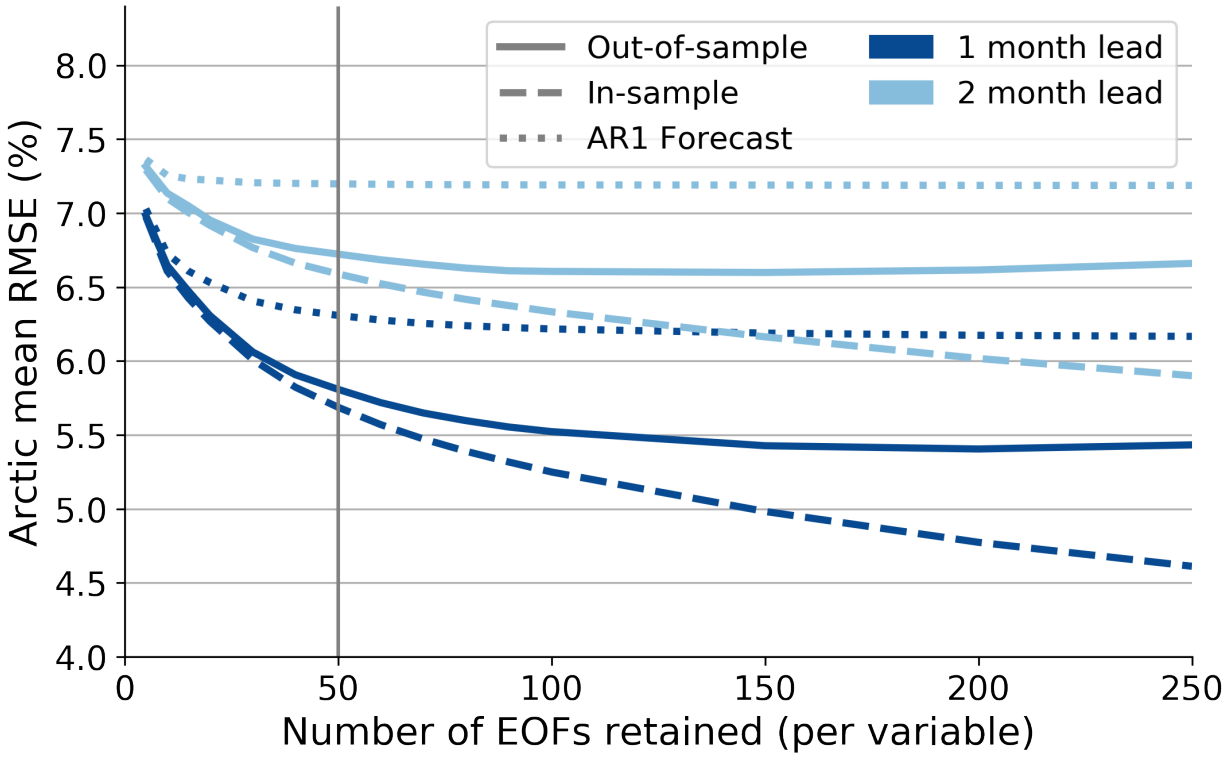

Figure 3.3: Skill of a LIM forecast for one and two month forecasts of sea ice concentration. Arctic mean root mean squared error (RMSE) is shown as a function of the number of empirical orthogonal functions (EOFs) retained in the truncation of all six variables included in the LIM. The LIM is trained on a monthly CESM1 LME simulation between 850–1650 CE and validated between 1651–1850 CE.

EOFs per variable, which were determined to be sufficient at the beginning of this section. Figure 3.4 shows the performance of each LIM at different lead times relative to the SIC-only LIM. In general, all variables contribute faborably to forecast skill of SIC. TAS and SST contribute the most to SIC forecast skill on 1–3 month lead times, while SIT along with TAS and SST contribute the largest increase in skill on 4–6 month lead times. PSL and ZG500 show very similar contributions to the forecast skill in SIC at all lead times, though adding both variables increases forecast performance more than adding either variable individually (not shown). Given these results, we include all six variables in LIM training except when validating on reanalysis data, which does not include SIT.

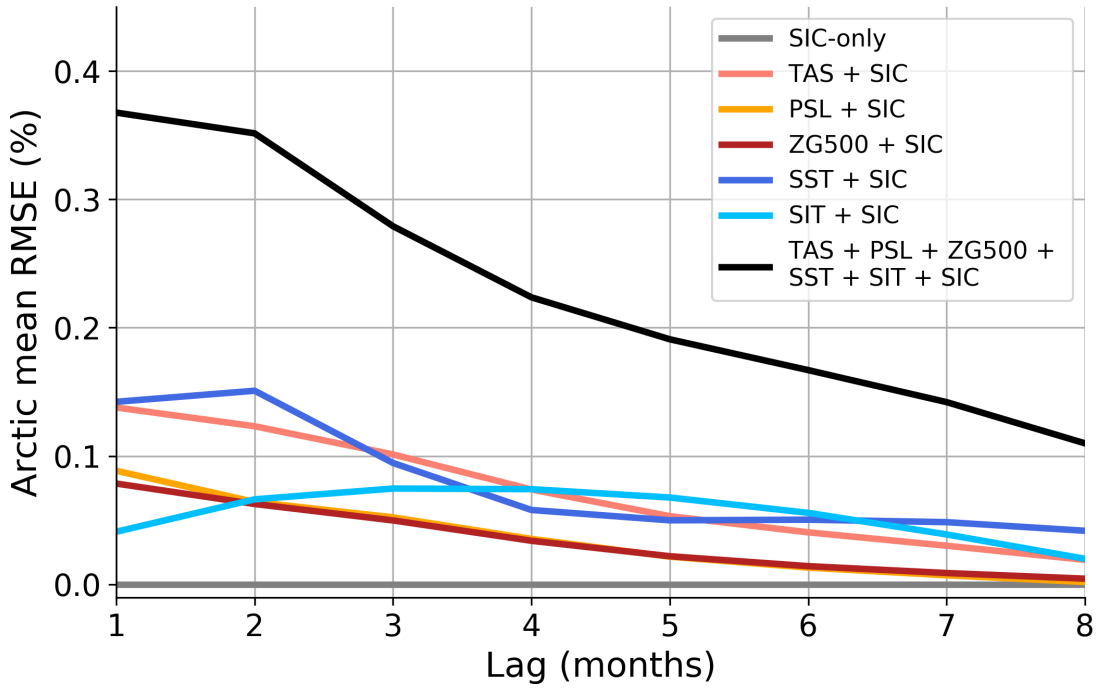

Figure 3.4: Improvement in forecast skill for Arctic sea ice concentration (SIC) relative to a SIC-only LIM forecast as a function of lead time for six different LIMs containing: 2-m air temperature (TAS), sea level pressure (PSL), 500hPa geopotential height (ZG500), sea surface temperature (SST), and sea ice thickness (SIT) each plus SIC; and all variables. Arctic mean root mean squared error (RMSE) is plotted relative to a LIM containing only SIC. Each LIM was trained using a CESM1 LME simulation from 850–1650 CE and validated between 1651–1850 CE.

Using the number of training years, validation years and number of EOFs retained for each variable determined above, we perform out-of-sample validation using a LIM trained on a CESM1 LME simulation (850–1650 CE) during the last millennium conditions (1651–1850 CE). The LIM shows skill above an AR1 forecast for all variables at one month lead times (see Figure 3.5). For pressure fields (PSL and ZG500), such skill is short-lived and disappears for two month or greater lead times. For sea ice variables (SIT and SIC) and TAS the LIM outperforms an AR1 forecast through eight month lead times and seven month lead times

for SST. Generally the shorter-lived predictability of PSL and ZG500 makes physical sense given the relatively short-lived nature of pressure anomalies in the atmosphere relative to SST anomalies, which tend to be more persistent. The longer lived predictability of SIT is also consistent with previous work that has found SIT anomalies persist for up to a year (Blanchard-Wrigglesworth et al., 2011a).

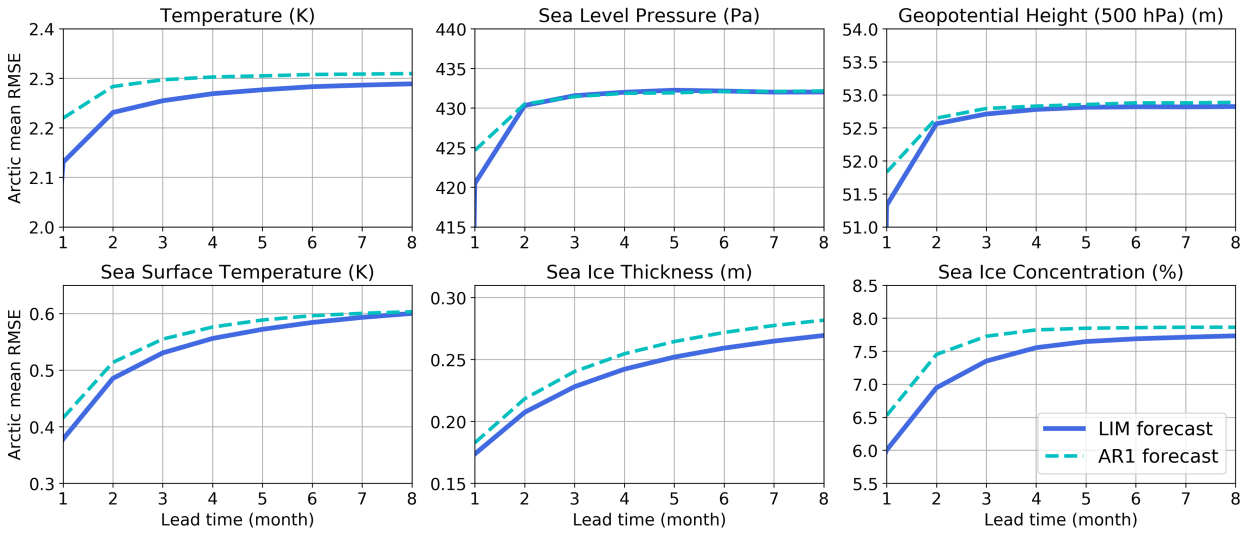

Figure 3.5: Arctic mean root mean squared error (RMSE) as a function of lead time for each variable contained in the LIM (blue line) compared to that from an autoregressive model of order one (AR1) forecast (turquoise dashed line). Both models were trained using a monthly CESM1 LME simulation from 850–1650 CE and validated between 1651–1850 CE.

The spatial skill of the LIM is shown in Figure 3.6. In general we find positive CE values everywhere except for small regions near the ice edge where sea ice variability is only present during a small number of months in the validation dataset. We see a maximum RMSE of approximately 20% and the RMSE is generally greatest starting near the Fram Strait and following south along the east coast of Greenland along the sea ice edge. Figure 3.7 shows the difference in skill of forecasts performed with a LIM minus an AR1 model forecast. The LIM outperforms an AR1 forecast in nearly all regions, increasing the correlation and CE value by a maximum of approximately 0.23 and 0.24 respectively. The RMSE is also reduced or

remains the same everywhere with a maximum reduction of approximately 3.8%. In general skill is highest where variability of SIC is greatest. Further we find that the LIM continues to outperformed an AR1 model forecast when initialized on CESM1 LME data outside of the last millennium from 1851–2004 CE.

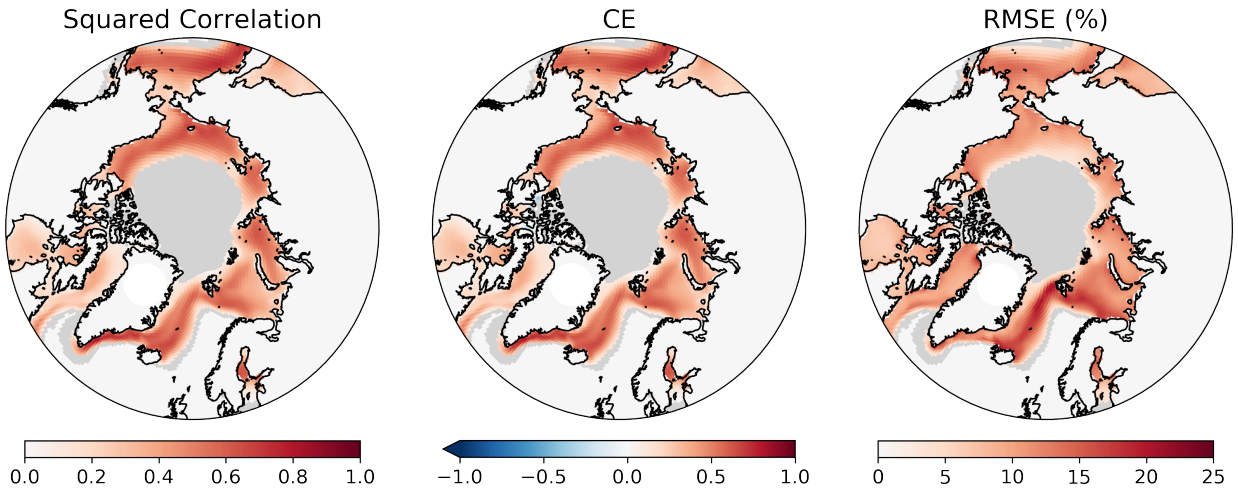

Figure 3.6: Squared correlation coefficient, coefficient of efficiency (CE value) and root mean squared error (RMSE) at each grid cell for 1 month forecasts of sea ice concentration using a LIM trained on a monthly CESM1 LME simulation between 850–1650 CE and validated using 1651–1850 CE. Regions shown in gray are grid cells where the variance in sea ice concentration across all validation time steps is less than 0.001.

These intra-model, out-of-sample results indicate that the LIM can not only predict SIC well, but other important oceanic and atmospheric variables that give a broader picture of the Arctic climate state. We have optimized the LIM to predict SIC, but it maintains skill in other variables for long lead times (up to 8 months). Further, the LIM is able to predict SIC on one month lead times throughout the Arctic region, showing better skill than an AR1 forecast everywhere. These results indicates that the LIM is useful tool for emulating coupled-model simulations.

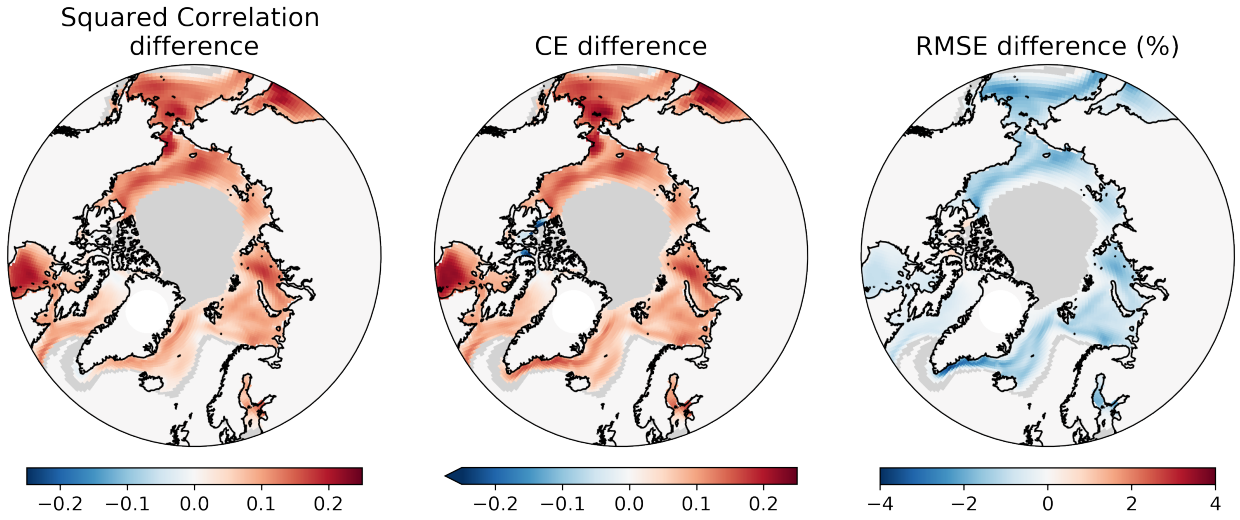

Figure 3.7: Squared correlation coefficient, coefficient of efficiency (CE value) and root mean squared error (RMSE) difference between one month forecasts of Arctic sea ice concentration from a LIM and autoregressive model of order one at each grid cell. Both models were trained using a monthly CESM1 LME simulation from 850–1650 CE and validated between 1651–1850 CE. Regions shown in gray are grid cells where the variance in sea ice concentration across all validation time steps is less than 0.001.

3.3.2 Cross-model sea ice prediction

To investigate our second research question, and determine under what conditions the LIM works well for predicting Arctic sea ice conditions, we perform cross-model validation. For these experiments we initialize the LIM using data that originates from a different model than the LIM training data (CESM1 LME simulation from 850–1650 CE). Generally we find that the LIM fails to outperform an AR1 forecast (trained on the same model data as the LIM) when initialized with models outside CESM1 and outside the mean state conditions of the Last Millennium (Figure 3.8).

When validated on other last millennium simulations the LIM is able to outperform and AR1 model forecast when initialized with CCSM4 last millennium simulations, but not with MPI last millennium simulations. Similarly we initialized the LIM with CMIP6 historical simulations (1851–2014 CE) from MPI, GFDL and CESM2 as well as a future scenario

from the CESM2 LE (2007–2057 CE) and none of these predictions were able to outperform an AR1 forecast. Finally we tried to validate the LIM using ERA5 reanalysis data from 1980–2020 CE and this LIM also failed to outperform an AR1 forecast (SIT not included).

To better understand why the LIM is unable to make skillful predictions when initialized on these other datasets, we examine the location and magnitude of the variability in the validation data relative the the training dataset on a grid cell by grid cell basis. To do this we calculate the variability in time at each grid cell for the training data as well as for each of the model simulations used for cross-model validation. We then calculate the spatial correlation between the training and validation patterns of variability to represent spatial differences. Figure 3.8 shows this spatial correlation versus the Arctic mean RMSE for different LIM forecasts relative to an AR1 forecast of sea ice concentration. Overall this spatial correlation is a strong predictor of LIM performance. Thus the location of the variability in the training and validation datasets have to be sufficiently correlated in order for the LIM to outperform an AR1 forecast. Applying least-squares regression to the data shown in Figure 3.8 yields an R^2-value of 0.77. The x-intercept of this best fit line is 0.66 indicating that a spatial correlation between the training and validation data variance must exceed that value in order to likely outperform an AR1 forecast. A similar analysis was also done for sea ice thickness (see Figure B.1) and the x-intercept of the best fit line is 0.52 indicating that high spatial correlation between the training and validation data is required for skillfully forecasting thickness as well. These results confirm that LIMs are most useful when applied to time periods exhibiting stationary statistics and that the LIM inherits model biases from the training dataset.

3.4 Conclusions and Discussion

Overall we find that a LIM trained using a CESM1 LME simulation performs well at predicting Arctic sea ice and other climate variables when validated on the same model output and mean state, even when out-of-sample in time. We find that the 500 years or more is necessary for training a LIM for monthly prediction and ideally 100 or more years of monthly

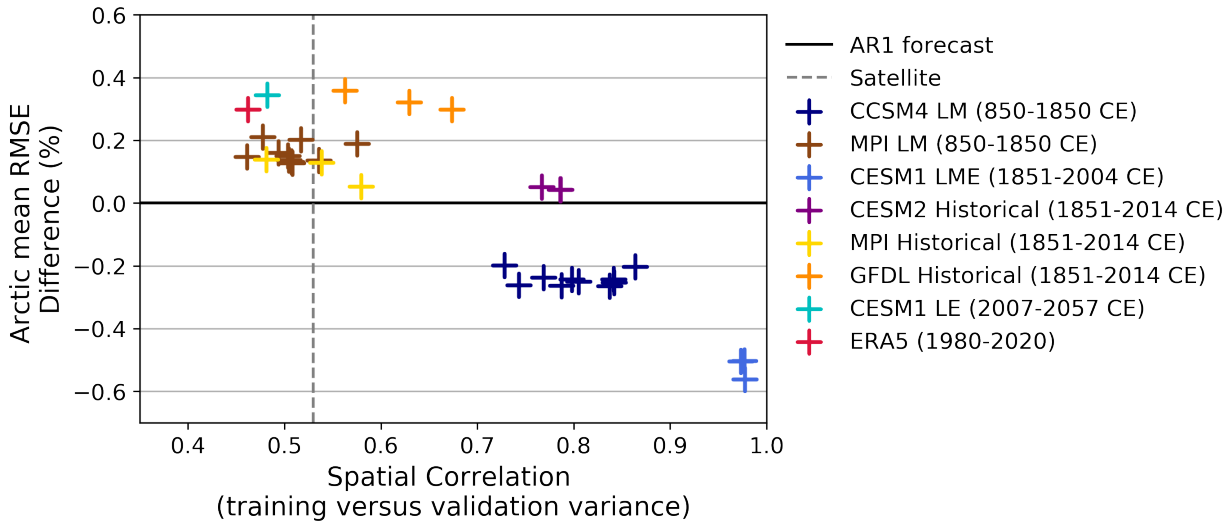

Figure 3.8: Arctic mean root mean squared error (RMSE) for one month LIM forecasts of sea ice concentration validated on various different model simulations (different colors) versus the spatial correlation between the variance (across time) of each validation dataset and the LIM training dataset. The Arctic mean RMSE difference is the RMSE value of the LIM forecast minus the RMSE of an autoregressive model of order one (AR1), thus negative values indicate when the LIM forecast outperforms an AR1 forecast. For all forecasts, the LIM was trained using a CESM1 LME simulation from 850–1650 CE. Values were calculated for 100 year non-overlapping segments for last millennium simulations (LM) and all 50 year non-overlapping segments for historical simulations.

validation data is needed to properly quantify the skill. We also find that TAS and SST contribute the most to Arctic SIC prediction on one to three month timescales, while SIT becomes more important on four to six month timescales.

When validated on out-of-sample intra-model data, the LIM performs well not only for for SIC prediction but for all of the six variables included. The predictions outperform an AR1 forecast for all variables except for PSL and ZG500 for lead times of 2 months or longer. This highlights the ability of the LIM to capture a broader picture of the Arctic system despite being optimized for predicting sea ice coverage. The LIM outperforms an AR1 forecast throughout the Arctic, and shows the most skill in regions near the ice edge with the most variability in SIC.

We test the ability of the LIM to predict conditions when validated with different model output and in general the LIM fails to make strong predictions. We do this by initializing the LIM with other last millennium simulations, historical simulations, reanalysis data and a future scenario and none outperform an AR1 forecast (except CCSM4 last millennium simulation). This is likely due in part to model biases and in part to mean state changes depending on the scenario of interest. Previous work has also found a large range of persistence estimates across models as well as a general overestimation of persistence in models compared to observations (Blanchard-Wrigglesworth and Bushuk, 2019; Giesse et al., 2021), which could also contribute the lack of skill we see during cross-model validation. We find that the spatial correlation between the variability (in time) of the training versus validation data can be used as a good predictor (R^2-value of 0.77) of whether the LIM will outperform an AR1 forecast. According to a least-squares regression of spacial correlation versus Arctic mean RMSE of LIM forecasts of SIC, we find that a spatial correlation greater than 0.66 between the training and validation data could yield a LIM forecast which outperforms an AR1 forecast.

Given this metric, one can pursue training and validation datasets that best fit a period of interest. If a LIM were to be used for present day forecasting, ample training data is needed, ruling out the use of satellite and reanalysis products. Furthermore, a training dataset's variability would need to correlate well with satellite observations. The training dataset we use here, CESM1 LME, has a spatial correlation of 0.57 with satellite observations between 1980–2015 CE which does not pass the threshold determined here. This value drops to 0.42 between 2000–2015 CE.

As described here, the LIM is best used as a model emulator as it seems to inherit model biases from the training dataset. However, given the LIM can predict a diverse set of variables across the climate system at minimal computational cost, it could be a useful tool for probing model dynamics in the Arctic.

Chapter 4

MONTHLY ARCTIC SEA-ICE RECONSTRUCTIONS USING ONLINE DATA ASSIMILATION

4.1 Introduction

Arctic sea ice has declined dramatically in recent decades (e.g.Cavalieri and Parkinson (2012); Stroeve et al. (2012)). These large declines have been observed in all months of the year; however, not all months have been declining at the same rate, leading to shifts in seasonality. Observations indicate that the melt season has lengthened, with earlier melt onset and later freeze-up dates (e.g. Markus et al. (2009); Bliss and Anderson (2014)). The decline of sea ice coverage has also been accompanied by a thinning of the ice (Lindsay and Schweiger, 2015; Kwok, 2018; Schweiger et al., 2019). Further, some argue that the decline in total Arctic sea ice extent has not been constant, the second half of the satellite record showing declines 2 times faster that the first (Stroeve et al., 2012; Serreze and Stroeve, 2015).

In order to better understand these changes across seasons and to put them in context, a longer record than satellite observations (1979–present) is needed. Further, reconstructing a broader set of climate fields that span different components of the Arctic climate system is needed. There are two reconstructions of Instrumental Era sea ice conditions that produce the fully gridded fields with monthly resolution. In the first (Walsh et al., 2017, 2019), the authors combine a ranked list of synthesized historical records along with analog and linear interpolation methods to spatially infill regions and times when observations are not available resulting in monthly sea ice coverage from 1850–2017 CE. The Walsh et al. (2017) reconstructions show little change in the difference between Arctic sea ice extent in March and September until ∼1980 when the September extent begin to decline much more rapidly than the March values. However, as shown in Brennan et al. (2020), the observations used

for these reconstructions are particularly sparse prior to 1953 during winter months, making seasonal investigations with this dataset challenging.

The second reconstruction of Instrumental Era sea ice conditions was published in Schweiger et al. (2019) (PIOMAS-20C) . For this reconstruction, the authors focus on reconstructing Arctic sea ice volume using a sea-ice–ocean model (PIOMAS, (Schweiger et al., 2011; Zhang and Rothrock, 2003)) forced with an atmospheric reanalysis product. They assimilate the sea ice edge using HadISST2 (Titchner and Rayner, 2014) data which relies heavily on Walsh and Chapman (2001) prior to the year 1971. According to this reconstruction, the satellite era declines in total Arctic sea ice volume are 6 times greater than that observed during the ETCW (approximately 1901–1940 CE). Also according to PIOMAS-20C, total Arctic sea ice volume values in the early 1900s are comparable to that in the 1990s, despite Arctic mean temperatures (north of 60°N) having warmed by roughly 1°C between those two periods in temperature observations (Morice et al., 2020).

Here we aim to produce a new independent reconstruction of Arctic SIC and thickness along with atmosphere and ocean fields on monthly timescales throughout the Instrumental Era. As outlined in Chapter 3, major contributors of Arctic sea ice variability on monthly timescales include persistence of sea ice anomalies (Lemke et al., 1980; Lindsay et al., 2008; Blanchard-Wrigglesworth et al., 2011a), dynamical advection of sea ice by atmospheric circulation patterns (Walsh and Johnson, 1979; Fang and Wallace, 1994; Guemas et al., 2016), atmospheric temperature variability (Olonscheck et al., 2019), and ocean heat flux (Bitz et al., 2005). In order to skillfully reconstruct sea ice at this higher temporal resolution, we anticipate the need to incorporate some of these dynamical processes into the DA framework. To do this we will employ online DA, which combines information from temperature observations with a prior and then forecasts the assimilated state in time thereby incorporating dynamical memory into the reconstructions. Forecasting the coupled ocean-ice-atmosphere state is computationally expensive, so we will use a LIM to perform the forecast step in our online DA scheme.

Online DA has long been ubiquitous in weather forecasting for many years, but was

first implemented on longer timescales for paleoclimate field reconstructions only recently (Perkins and Hakim, 2017, 2021). Perkins and Hakim (2017) use an online DA framework to reconstruct temperature throughout the Instrumental Era on annual timescales, and find that incorporating a deterministic LIM in the forecast step increases the skill in global mean temperature reconstructions over offline reconstructions. Overall, adding a LIM into the DA scheme showed improvement in skill in the North Atlantic and Barents Sea regions near the sea ice edge which shows promise for applying these methods to Arctic sea ice reconstructions. Perkins and Hakim (2021) builds on their prior work to reconstruct coupled atmosphere–ocean climate states throughout the last millennium (1000–2000 CE) on annual timescales. Perkins and Hakim (2021) find that the addition of online data assimilation significantly increases the variability of the ensemble mean reconstructions for global mean TAS, SST and 0-700m ocean heat content particularly prior to 1850 CE.

Here we build on the work of Perkins and Hakim (2017, 2021) to reconstruct atmosphere–ocean–sea ice climate fields on monthly timescales throughout the Instrumental Era. Specifically, we will use a Kalman Filter approach to assimilate temperature observations with climate output and then forecast the assimilated state using a LIM that was developed in Chapter 3.

4.2 Methods

4.2.1 Online Data Assimilation

In chapter 2 we used an ensemble Kalman Filter technique to reconstruct Arctic sea ice in an offline (no temporal memory in the prior) framework. Here we employ an online framework for which two steps are carried out at each time step: (1) the update step and (2) the forecast step. For the update step, we assimilate temperature observations using a Kalman Filter, and for the forecast step we use a LIM to project the mean state and covariance in time. The forecasted state is then used as the prior for the update step at the next time step and the process repeats. These methods follow closely to those found in Hakim et al. (2022) where

this approach was applied to sub-seasonal forecasting.

Update Step: Kalman Filter

The update step uses Equation 2.1, except here $\mathbf{x}$ is no longer an ensemble and represents the mean state of the system. For a given set of observations ($\mathbf{y}$), the analysis ($\mathbf{x}_a$) is

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_b) \tag{4.1}$$

with covariance ($\mathbf{B}_a$)

$$\mathbf{B}_a = (\mathbf{I} - \mathbf{H}\mathbf{K})\mathbf{B}_b. \tag{4.2}$$

Here $\mathbf{x}_b$ is the prior estimate, or the initial 'best guess' of the mean climate state. $\mathbf{H}$ is the forward operator from the model prior to observation space and $\mathbf{I}$ is the identity matrix. The Kalman Gain matrix is given as

$$\mathbf{K} = \mathbf{B}_b\mathbf{H}^T(\mathbf{H}\mathbf{B}_b\mathbf{H}^T + \mathbf{R})^{-1} \tag{4.3}$$

where $\mathbf{R}$ is the error covariance of the observations. We will refer to the update as $\mathbf{d} = (\mathbf{y} - \mathbf{H}\mathbf{x}_b)$.

Forecast Step: Linear Inverse Modeling

To forecast the mean state and covariance produced by the update step, we employ a LIM (developed in Chapter 3) with τ equal to 1 month. We forecast the mean state of our analysis ($\mathbf{x}_a$) in time using Equation 3.2, and given the expectation of the noise term is zero, we are left with a deterministic forecast

$$\mathbf{x}_a(t + \tau) = \mathbf{G}_\tau \mathbf{x}_a(t). \tag{4.4}$$

Given a fundamental assumption when using a LIM that the state and the errors are uncorrelated, the error covariance can be propagated in time using

$$\mathbf{B}_a(t+\tau) = \mathbf{G}_\tau \mathbf{B}_a(t) {\mathbf{G}_\tau}^T + \mathbf{N}_\tau \tag{4.5}$$

where $\mathbf{N}_\tau$ is the stochastic error covariance specific to time lag τ, and $\mathbf{B}_a(t) = \mathrm{cov}(\mathbf{x}_a(t), \mathbf{x}_a(t))$. Since Equation 4.5 is valid for all time, solving this equation at $t = 0$ yields an equation for $\mathbf{N}_\tau$ (Equation 11 in (Penland, 1989)),

$$\mathbf{N}_\tau = \mathbf{C}_0 - \mathbf{G}_\tau \mathbf{C}_0 {\mathbf{G}_\tau}^T. \tag{4.6}$$

These propagated states can then be used as the prior mean state ($\mathbf{x}_b = \mathbf{x}_a(t+\tau)$) and covariance ($\mathbf{B}_b = \mathbf{B}_a(t+\tau)$) in the update step (Equations 4.1-4.3) at the next time step $(t+\tau)$. This process is repeated for all time steps reconstructed.

Uncertainty calibration:

All validation and error quantification is done in full latitude-longitude space. If the errors in the observations and prior are well calibrated, the covariance of the update ($\mathbf{d}$), when averaged over all validation times, should be approximately equal to the denominator of the Kalman Gain from Equation 4.3 (e.g. Houtekamer and Mitchell (1998). Therefore, for each observation (i) we have,

$$\frac{\mathrm{cov}(\mathbf{d}, \mathbf{d})_{ii}}{[\mathbf{H}\mathbf{B}_b\mathbf{H}^T + \mathbf{R}]_{ii}} \approx 1. \tag{4.7}$$

We will refer to this ratio as the calibration ratio, and use it to assess how well the uncertainty in the observations and prior is tuned.

Observations

Observations are drawn for every month from gridded datasets described in Section 4.2.2. Given the LIM forecast is performed in truncated EOF space, the observations also need to be projected into truncated EOF space using

$$\mathbf{y} = \hat{\mathbf{H}}\hat{\mathbf{x}} = \mathbf{H}\mathbf{x} + \epsilon. \tag{4.8}$$

Here $\hat{\mathbf{x}}$ is the full lat–lon gridded state and $\mathbf{x}$ is the truncation of $\hat{\mathbf{x}}$ in EOF space. The forward operator in EOF space ($\mathbf{H}$) is related to the forward operator in lat–lon space ($\hat{\mathbf{H}}$) by $\mathbf{U}$ such that $\mathbf{H} = \hat{\mathbf{H}}\mathbf{U}$. The rows and columns of $\mathbf{U}$ represent the the number of lat–lon gridcells and number of EOFs respectively.

The relative size of the observation and prior uncertainty determines the weight given the each (Equation 4.3) and thus must be properly calibrated (Equation 4.7). In general, the observation uncertainty is a combination of measurement error, representativeness error of the HadCRUT grid cell, and representativeness error from the EOF truncation. Here, the observation uncertainty is approximated as a constant at all observation locations for all time. Specifically, the diagonal elements of $\mathbf{R}$ (equation 4.3) are set to 1.0. Constant values of 0.01, 0.2, 0.4, 0.6, 1, 2, and 10 were all tested. Given relatively larger uncertainty in the prior, observational uncertainty between 0.2–1.0 all resulted in roughly the same performance (as measured by R^2 and CE values) for northern hemisphere (NH) mean values for many variables reconstructed using imperfect pseudo observation experiments (see Figure C.1). Uncertainty of 1.0 was chosen because it led to the best performance in total Arctic sea ice thickness and nearly the best for total Arctic sea ice area. Other ways of quantifying the observation uncertainty were tested including estimating the observation uncertainty as truncation error as outlined in Hakim et al. (2022), which not only allows for the uncertainty to change at different locations, but also in time. However, using the EOF truncation error led to worse performance in imperfect pseudo observation experiments as measured by NH mean values for all variables reconstructed. This is likely because there are many off-diagonal terms in the EOF truncation estimation of $\mathbf{R}$, leading to an overestimation of the uncertainty which are poorly constrained (spurious correlations) and reduce the effective number of observations.

Next, we find that assimilating only a subset of available observations at each time step leads to the best performance in pseudo observation experiments. For each time step

reconstructed, 100 observation locations are randomly drawn from the full set of observations available. In many cases, 20 iterations are then performed to estimate the uncertainty due to the observation locations assimilated and the average across all 20 iterations is reported unless stated otherwise. The fact that assimilating only a subset of observations at each time step leads to the better performance likely indicates that the uncertainty in the observations is not perfectly calibrated. Further work could be done to explore other ways to quantify observation uncertainty so that more observations could be utilized in each reconstruction.

4.2.2 Data Sources

Climate fields reconstructed throughout this chapter include: 2-m air temperature (TAS), sea level pressure (PSL), 500hPa geopotential height (ZG500), sea surface temperature (SST), sea ice thickness (SIT), and sea ice concentration (SIC). As described in Chapter 3, these variables were selected because they have all been associated with sea ice variability and we hypothesize that including each of these variables may contribute positively to the performance of reconstructing sea ice variables. Specifically, Arctic sea ice has been shown to exhibit strong coupling with surface air temperature, particularly on longer timescales (e.g., (Gregory et al., 2002; Armour et al., 2011; Mahlstein and Knutti, 2012; Olonscheck et al., 2019)). PSL is associated with surface winds which can drive sea ice motion and variability (e.g., (Rigor et al., 2002)) and similarly ZG500 is associated with the large scale dynamical state of the atmosphere which can also contribute to sea ice variability. Further, sea ice concentration persistence has been shown to be strongly influenced by both SSTs and SIT (e.g. (Blanchard-Wrigglesworth et al., 2011a; Bushuk et al., 2015)). We therefore include all of these variables in our reconstructions.

Data Used in the Update Step:

For the update step we assimilate observations from the Met Office Hadley Centre/Climatic Research Unit global surface temperature data set (HadCRUT5, Morice et al. (2020)) between 1850–2020 CE. The uncertainty in HadCRUT5 observations (diagonal elements of $\mathbf{R}$)

is estimated as a constant value of 1.0 as described in section 4.2.1. The initial mean state ($\mathbf{x}_b$ for the first time step), December of 1850 CE is used from a single ensemble member from the Community Earth System Model version 1 (CESM1), Last Millennium Ensemble (LME, Otto-Bliesner et al. (2016)) . The initial covariance ($\mathbf{B}_b$ for the first time step) is calculated as a time average from the CESM1 LME between 1651–1850 CE.

Data Used in the Forecast Step: LIM training

For performing our reconstructions, we use a LIM trained on monthly averaged data from a single ensemble member from the CESM1 LME. Training was performed using years 850–1650 CE (800 years), which was determined to be sufficient in Section 3.3.1. The linear trend and mean for each month is determined and removed at each grid point as was described in Section 3.2.2. The domain is also limited to the NH for all ocean and atmosphere variables and north of 40°N for sea ice variables.

As was described in Chapter 3, we reduce the dimensionality of the state by performing an EOF decomposition of each variable individually before training each LIM. Each variable in the training data is area weighted and then the EOFs are calculated. Then the first leading 50 EOFs are selected and normalized by the square-root of the sum of the variance in time so that the variance across the truncated modes of each individual variable sums to 1. Once truncated and normalized, all the variables are stacked as shown in Equation 3.5. With this truncated state vector, we then determine $\mathbf{G}_\tau$ using Equation 3.3 for $\tau = 1$ month.

4.2.3 Pseudo-observation experiments

Idealized experiments are performed where a climate model simulation is taken to be the target state. A sparse network of *pseudo* TAS observations are drawn from the target, and assimilated by the DA scheme (e.g., Steiger et al., 2014) and the resulting reconstructions of Arctic climate state are compared with the target state in order to assess the skill of the framework. Both perfect and imperfect model pseudo observation experiments are performed. For perfect model experiments, the pseudo observations, prior and LIM are drawn

from and trained on the same model output, which provides a useful upper bound on performance of a reconstruction for a given observation network. For imperfect model experiments, pseudo observations are drawn from one model simulation and initial prior state and the LIM are trained using a different model simulation. Imperfect model experiments provides a more challenging, but realistic test since the prior and forecasted covariance do not match that in the target state.

For perfect model experiments, a CESM1 LME simulation is taken to be the target state. Pseudo observations of TAS are drawn from the grid cells in the CESM1 LME simulation closest to HadCRUT5 observation locations for each month of the simulation between 1850–2005 CE. For each time step, 100 of the pseudo observations are randomly drawn from the full set of observations so that a constant number of observations is maintained throughout the reconstruction. The prior mean state is initialized using December of 1850 CE from the CESM1 LME simulation. The LIM used to forecast the mean state and covariance at each time step is trained on CESM1 LME monthly data between 850–1650 CE and the initial prior covariance is also estimated from the LIM training data. As mentioned in Section 4.2.1 the uncertainty in the observations is set to be a constant value of 1.0 and random error with variance 1.0 is also added to the pseudo observations. For the perfect pseudo experiments, the posterior reconstructions are validated against the target state (CESM1 LME simulation from 1851–2005 CE). In this setup, the prior and LIM covariance is the same as that from the target state, and thus the covariance in the online DA step is perfect within sample error.

For imperfect model experiments, the setup is the same as perfect pseudo observation experiments except the target state is taken to be the CMIP6 Max Planck Institute for Meteorology (MPI) Earth System Model version 1.2 historical simulations (low resolution, (Mauritsen et al., 2019)). Pseudo observations are drawn from the MPI historical simulation at HadCRUT5 observation locations between 1850–2005 CE in the same way as in perfect model experiments and the posterior reconstruction can again be validated against the target state (MPI historical simulation). The LIM and prior covariance are both trained and calculated using a CESM1 LME simulation between 850–1650 CE and the prior mean state

is initialized using December 1850 CE from the same simulation. This is a much stricter test of the method, and an important one, given that it mimics real reconstructions where the prior and LIM covariance is estimated from a model that is an imperfect representation of the target covariance. Both of these tests were performed by reconstructing all months between 1851–2005 CE, and 20 iterations were performed for each to estimate the uncertainty due to the observation locations assimilated. An average across all 20 iterations is reported unless stated otherwise.

4.2.4 Online versus offline data assimilation:

Online DA refers to when both the update and forecast step are performed at each time step. In order to test whether performing online DA produces more skillful reconstructions we perform four baseline reconstructions: *offline*, *fixed mean*, *fixed covariance* and *online* DA. Offline DA is when only the update step is performed and thus the prior is the same at each time step. In the forecast step, both the mean and covariance are forecasted separately. For fixed mean DA the covariance is forecasted in time and the mean is held constant, while for fixed covariance DA the mean is forecasted in time and the covariance is held constant. For online DA, both the mean and covariance are forecasted in time. These experiments will allow us to determine whether forecasting the mean, covariance or both are contributing to the skill in the reconstruction.

4.3 Results

4.3.1 Online Pseudo-observation experiments

For online perfect model experiments (both the mean and covariance are forecasted in time), we validate our reconstructions against the target state, which is taken to be the linearly de-trended CESM1 LME simulation from 1851–2005 CE. First, in Figure 4.1, we evaluate the skill of NH mean anomalies for each variable reconstructed throughout this time period. As expected, the reconstructions for TAS perform the best when compared to the target

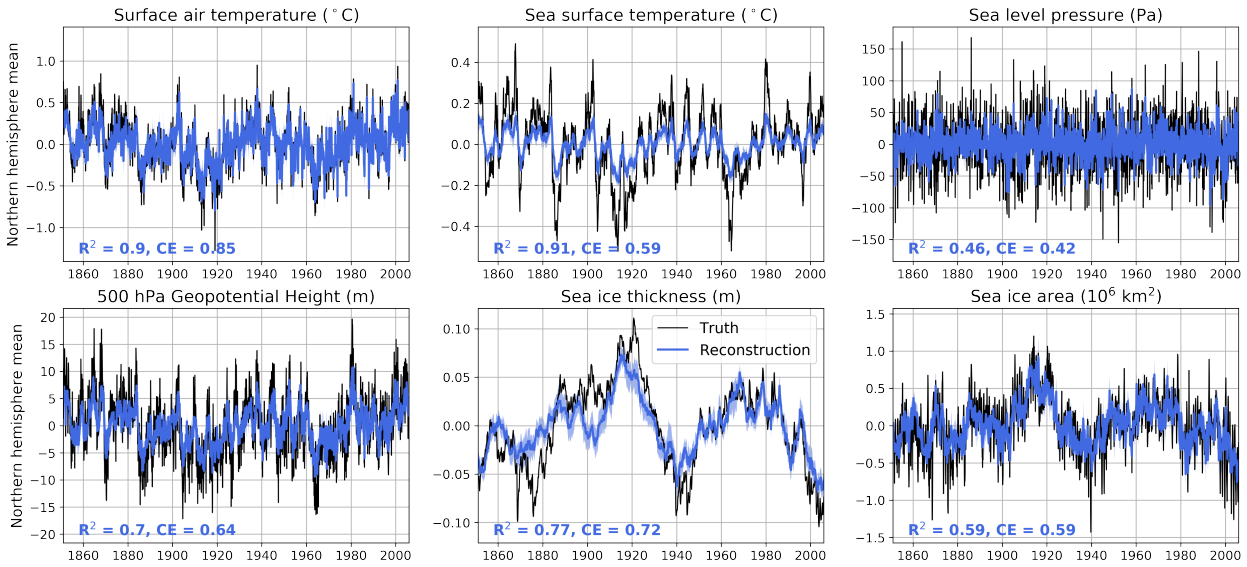

Figure 4.1: Northern hemisphere mean quantities for six target (black) and reconstructed (blue) fields. The target values are taken from CESM1 LME simulation and reconstructed fields produced using a perfect pseudo observation online data assimilation scheme. The initial prior mean state is taken as December 1850 from a CESM1 LME simulation, the LIM used in the forecast step was trained using CESM1 LME from 1650–1850 CE, and the initial prior covariance was calculated using the same training data as the LIM. Pseudo observations are taken from CESM1 LME between 1851–2005 CE at HadCRUT5 observation locations (with added random error drawn from a normal distribution with variance $(1°C)^2$). At each assimilation time step, 100 pseudo observations are drawn randomly. 20 Monte Carlo iterations of the reconstruction are performed subsampling observations to assimilate and the mean across all the iterations is shown in blue and the 2.5th–97.5th spread across these iterations is shown by the blue shaded region (the blue shaded region is not a an estimate of the total error). Anomalies are relative to the entire period shown (1851–2005 CE) and the squared correlation coefficient and coefficient of efficiency are shown from comparing reconstructed and target values for each variable.

simulation with an R^2-value of 0.90 and CE value of 0.85. Sea surface temperature reconstructions also perform well with an R^2-value of 0.91, yet show damped variability relative to the target simulation as indicated by a lower CE value of 0.59. The reconstructions of sea level pressure perform the worst with an R^2-value and CE value of 0.46 and 0.42 respectively. For sea ice variables, the reconstructions of Arctic mean sea ice thickness show an R^2-value

of 0.77 and CE value of 0.72 and the reconstructions of total Arctic sea ice area show an R^2-value of 0.59 and CE value of 0.59 when compared to the target simulation.

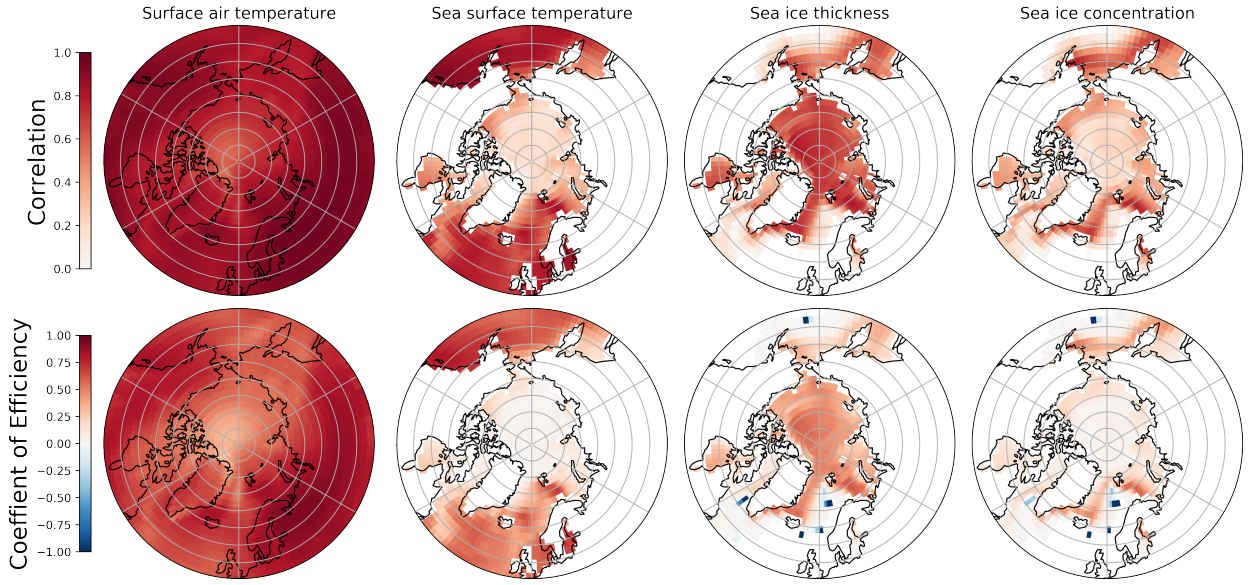

Figure 4.2: The spatial performance for four reconstructed variables produced using perfect pseudo observation online data assimilation experiments. Correlation (top) and coefficient of efficiency (bottom) were calculated at each grid cell between 1851–2005 CE between the reconstructed and target states. The target state and pseudo observations are taken from a CESM1 LME simulation. Regions shown in gray are grid cells where the variance in the target state across all time steps is less than 0.001.

The spatial performance of all reconstructed variables shows good performance throughout the Arctic. The correlation and CE value at each grid cell for TAS, SST, SIT and SIC are shown in Figure 4.2. The spatial skill of PSL and ZG500 is positive and uniform throughout the Arctic for both correlation and CE value and is thus not shown here. High correlation values are observed everywhere throughout the Arctic, particularly for TAS and SST variables. Positive CE values are observed throughout the Arctic, though the magnitude of the CE values are smaller than the correlation values everywhere. These results indicate that the online DA can skillfully combine sparse temperature observations together with a prior and LIM trained on last millennium simulations to skillfully reconstruct Arctic climate fields

from multiple components of the climate system in this perfect model framework.

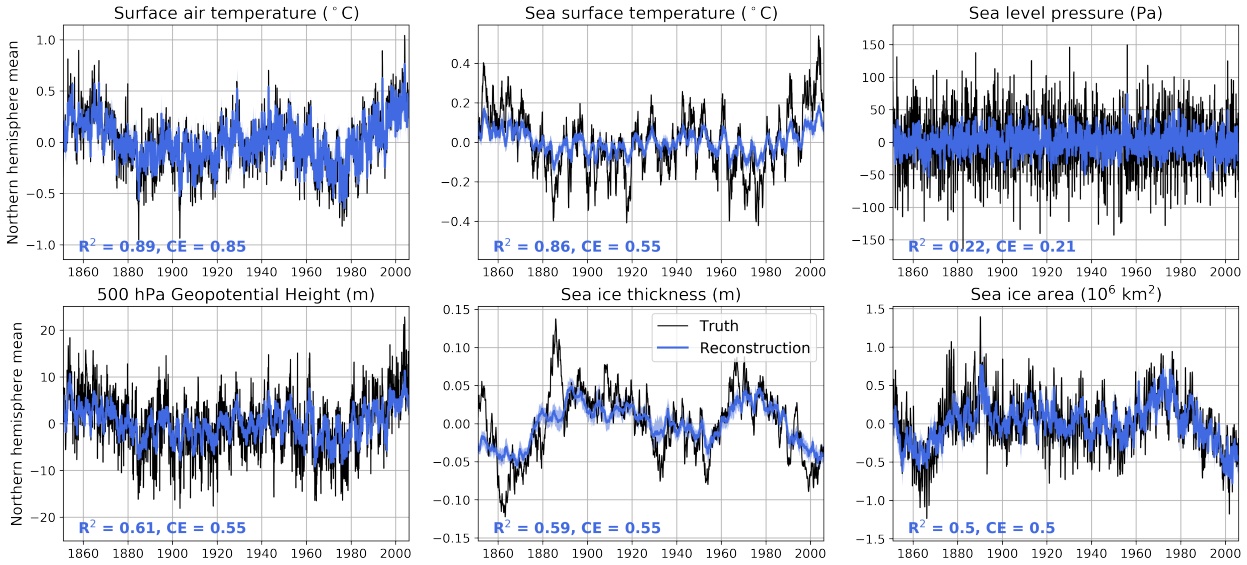

Figure 4.3: Northern hemisphere mean quantities for six target (black) and reconstructed (blue) fields. The target values are taken from the CMIP6 MPI Historical simulation and reconstructed fields produced using a imperfect pseudo observation online data assimilation scheme between 1851–2005 CE. The initial prior mean state is taken as December 1850 from a CESM1 LME simulation, the LIM used in the forecast step was trained using CESM1 LME from 1650–1850 CE, and the initial prior covariance was calculated using the same training data as the LIM. Pseudo observations are taken from a CMIP6 MPI Historical simulation between 1851–2005 CE at HadCRUT5 observation locations (with added random error drawn from a normal distribution with variance $(1°C)^2$). At each assimilation time step, 100 pseudo observations are drawn randomly. 20 Monte Carlo iterations of the reconstruction are performed subsampling observations to assimilate and the mean across all the iterations is shown in blue and the 2.5th–97.5th spread across these iterations is shown by the blue shaded region (the blue shaded region is not a an estimate of the total error). Anomalies are relative to the entire period shown (1851–2005 CE) and the squared correlation coefficient and coefficient of efficiency are shown from comparing reconstructed and target values for each variable.

For online imperfect model experiments (both the mean and covariance are forecasted in time), we draw observations from and validate our reconstructions against the target state, which is taken to be the linearly de-trended CMIP6 MPI historical simulation from 1851–2005 CE. As expected, by all metrics, the imperfect model experiments reveal weaker performance

than those seen in the perfect model framework. The skill in NH mean quantities is shown in Figure 4.3 and again, the reconstruction of NH mean TAS performs the best when compared to the target simulation with an R^2-value of 0.89 and CE value of 0.85. The reconstructions of NH mean sea surface temperature also perform well with an R^2-value of 0.86, yet have a CE value of 0.55 revealing more damped variability than the perfect model results. The reconstructions of sea level pressure again, perform the worst with an R^2-value and CE value of 0.22 and 0.21 respectively. For sea ice variables, the reconstructions of Arctic mean sea ice thickness show an R^2-value of 0.59 and CE value of 0.55 and the reconstructions of total Arctic sea ice area show an R^2-value of 0.50 and CE of 0.50 when compared to the target simulation.

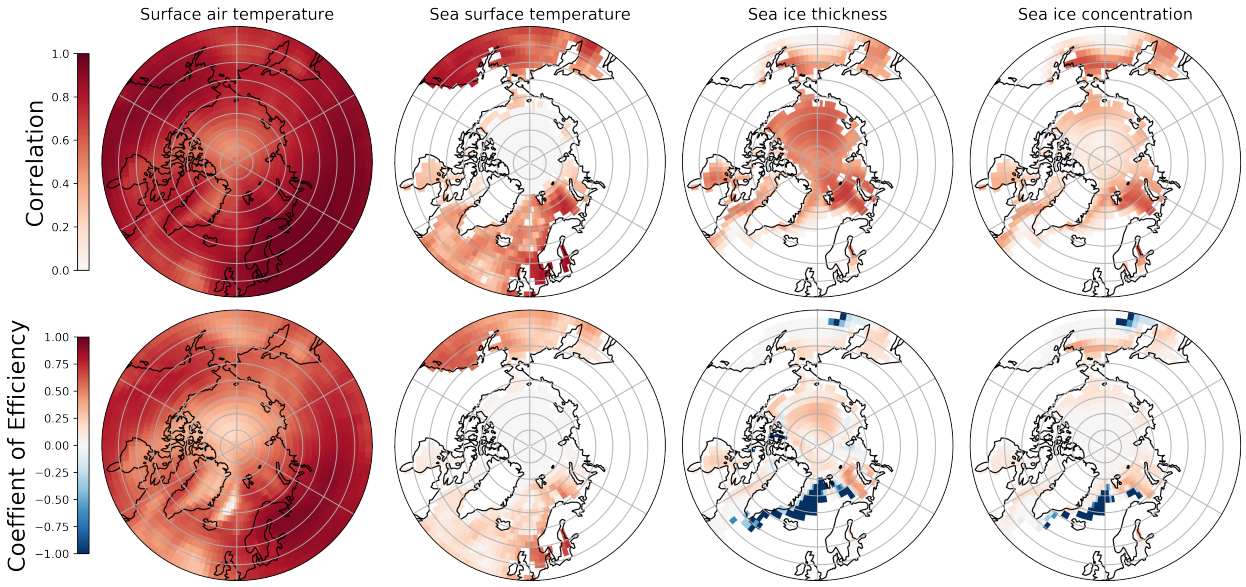

Figure 4.4: The spatial performance for four reconstructed variables produced using imperfect pseudo observation online data assimilation experiments. Correlation (top) and coefficient of efficiency (bottom) were calculated at each grid cell between 1851–2005 CE between the reconstructed and target states. The target state and pseudo observations are taken from the CMIP6 MPI Historical simulation. The initial prior mean state is taken as December 1850 from a CESM1 LME simulation, the LIM used in the forecast step is trained using CESM1 LME from 1650–1850 CE, and the initial prior covariance was calculated using the same training data as the LIM.

For the imperfect model experiments, the spatial performance shows positive correlation everywhere in the Arctic (see Figure 4.4). Again, the spatial skill of PSL and ZG500 is positive and uniform throughout the Arctic for both correlation and CE values and is thus not shown. The results for the four other reconstructed fields is shown in Figure 4.4. In general, the performance is lower everywhere in the imperfect model pseudo observation experiments when compared to the perfect model experiments. The CE values are positive everywhere for all variables except near the sea ice edge along the east coast of Greenland for both SIT and SIC. These regions where negative CE values occur generally are where there is sporadic sea ice present in the target state, but consistent sea ice coverage and variability in the training dataset. The sea ice edge in the LIM training and prior data (CESM1 LME simulation) extends further east of Greenland then in the target simulation (MPI historical simulation). This results in mean biases and variance differences throughout this region between the posterior and target state leading to negative CE values.

Overall, these results indicate that this online DA approach can skillfully reconstruct various Arctic climate fields on monthly timescales in both perfect and imperfect model frameworks. We see high squared correlation values between reconstructed and target values of NH quantities, which indicates that much of the variability in the target state is explained by the reconstruction. For all fields, the CE values are lower than one, indicating that the reconstructions tend to under-estimate the magnitude of the variability in each field.

In order to quantify if the use of the forecast step in online DA is improving the reconstruction, we perform a single iteration of a fixed covariance (only the mean is forecasted in time), fixed mean (only the covariance is forecasted in time) and offline DA reconstruction for comparison. The NH mean quantities for each variable reconstructed using online, offline and fixed covariance DA are compared in Figure 4.5. In general, the addition of the forecast step improves the skill of the reconstructed NH mean quantities for all variables relative to the offline DA results. The largest boost in performance occurs for both mean SIT and total Arctic sea ice area, where the percent of the target variance explained by reconstruction (R^2-value) nearly doubles when comparing the offline and online reconstructions. Online

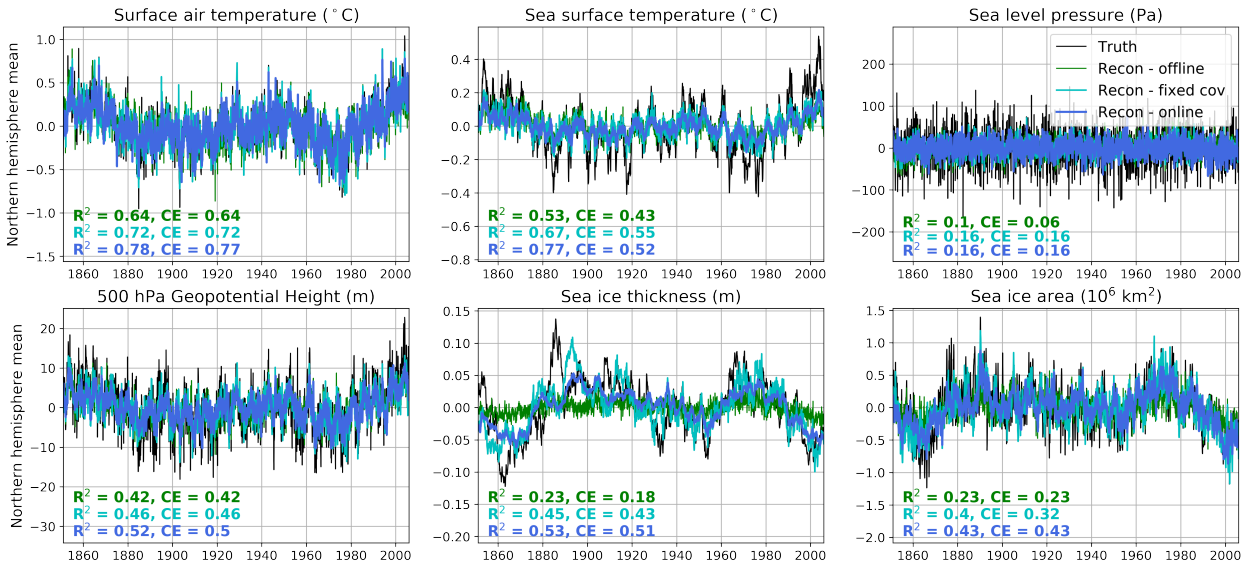

Figure 4.5: Northern hemisphere mean quantities for six target (black) and reconstructed (color) fields. The results from three different reconstructions are shown: offline DA (green), fixed covariance DA (turquoise), and online DA (blue). The target values are taken from the CMIP6 MPI Historical simulation and reconstructed fields were all produced using an imperfect pseudo observation scheme between 1851–2005 CE. The initial prior mean state is taken as December 1850 from a CESM1 LME simulation, the LIM used in the forecast step for the fixed covariance and online DA reconstructions was trained using CESM1 LME from 1650–1850 CE. The initial prior covariance was calculated using the same training data as the LIM. Pseudo observations are taken from a CMIP6 MPI Historical simulation between 1851–2005 CE at HadCRUT5 observation locations for each reconstruction with added random error drawn from a normal distribution with variance $(1°\text{C})^2$. At each assimilation time step, 100 pseudo observations are drawn randomly and the results shown represent one Monte Carlo iteration of each reconstruction. Anomalies are relative to the entire period shown (1851–2005 CE) and the squared correlation coefficient and coefficient of efficiency are shown from comparing the reconstructed and target values for each variable.

reconstructions also outperform the fixed covariance reconstructions for all variables, though the improvement is marginal compared to the improvement achieved between the offline and fully online results. Fixed mean (only the covariance is forecasted in time) reconstructions show similar or slightly better performance to the offline reconstructions of NH mean quantities for all variables (not shown), which indicates that most of the skill coming from moving

to an online DA framework is coming from forecasting the mean. Overall, these results confirm that implementing the online DA frameworks does improve the reconstruction skill on monthly timescales mainly through forecasting the mean.

4.3.2 Online Reconstruction Validation

Given the pseudo observation experiments show skillful results, we next assimilate real HadCRUT5 temperature observations between 1851–2020 CE using the fully online DA approach. We again use December 1850 from a CESM1 LME simulation as our initial prior mean state, the LIM used in the forecast step was trained using CESM1 LME from 1650–1850 CE, and the initial prior covariance was calculated using the same training data as the LIM. 100 HadCRUT5 observations are randomly drawn at each time step and assimilated using the update step. We reconstruct all months between 1851–2020 CE and 20 Monte Carlo iterations are performed to estimate the uncertainty due to different random draws of observations assimilated.

To begin, we first validate our reconstructions of temperature with other records available during the Instrumental Era. Figure 4.6 shows the comparison of the reconstructed NH mean temperature with that from HadCRUT5, Berkeley Earth (BE, Rohde et al. (2013a)), and NASA Goddard Institute for Space Studies (GISTEMP, Hansen et al. (2010)). Our reconstruction agrees strongly with all three datasets with an R^2-value of 0.89 when compared with HadCRUT5, 0.95 with GISTEMP, and 0.95 with BE. The CE values are also high, with values of 0.85, 0.91 and 0.89 when comparing the reconstruction with HadCRUT5, GISTEMP, and BE respectively. Between 2000–2020 CE, HadCRUT5 anomalies are slightly smaller in magnitude than anomalies in GISTEMP and BE. Our reconstructed anomalies follow those from HadCRUT5 given those are the observations assimilated in the online DA scheme.

Next, we compare our reconstructions of NH mean SST to other observationally based datasets. Reconstructed NH monthly mean sea surface temperatures from 1851–2020 CE along with that from the Hadley EN4 v4.2.1 dataset from 1940–2010 CE (Good et al.,

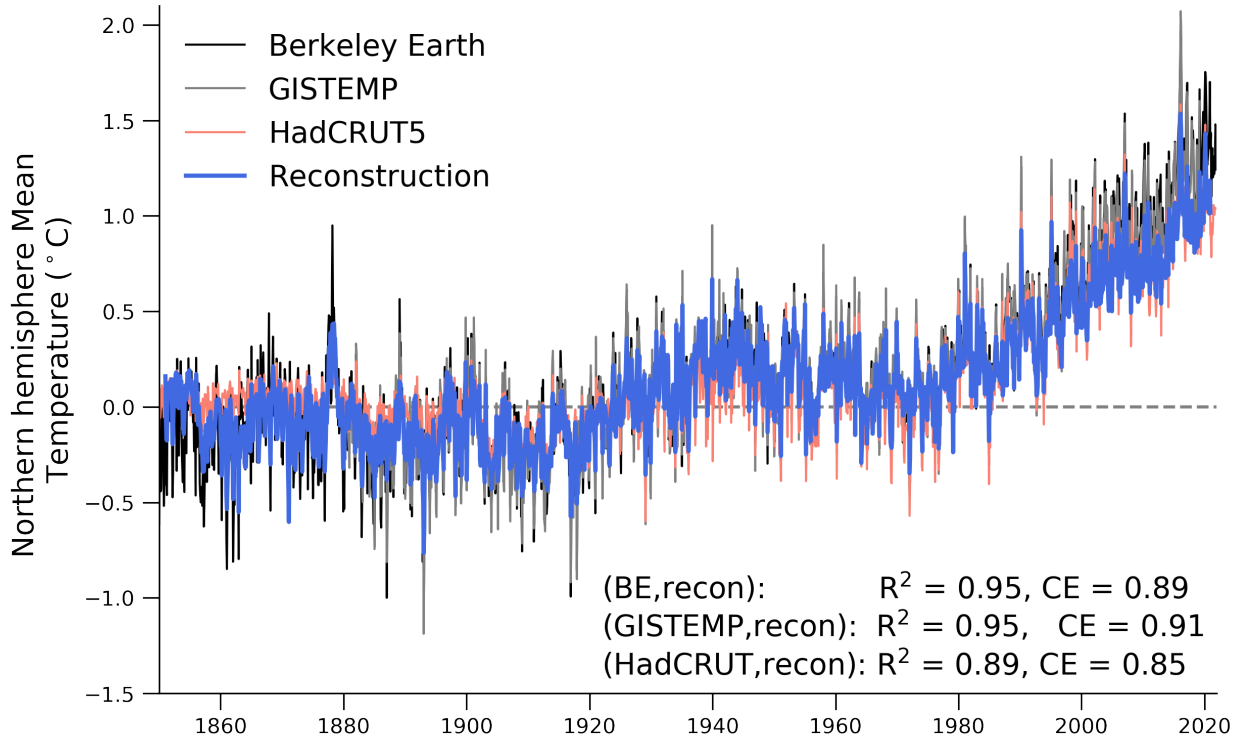

Figure 4.6: A reconstruction of monthly northern hemisphere mean surface air temperature anomalies from 1851–2020 CE (blue). The northern hemisphere mean air temperature from Berkeley Earth is shown in black (Rohde et al., 2013a), GISTEMP in gray (Hansen et al., 2010) and HadCRUT5 in pink Morice et al. (2020). The monthly mean over the entire period of each dataset as been removed and then anomalies are all relative to the mean calculated between 1880–1970 CE.

2013), the Geophysical Fluid Dynamics Laboratory Ensemble Coupled Data Assimilation dataset from 1961–2010 CE (GFDLECDA, Chang et al. (2013)), and the European Center for Medium-range Weather Forecasting ORA-20C dataset from 1900–2009 CE (de Boisséson et al., 2018) are shown in Figure 4.7. The reconstructions show the strongest comparison with both ORA-20C (R^2-value of 0.88 and CE value of 0.78) and GFDLECDA (R^2-value of 0.9 and CE value of 0.73) and the weakest comparison with Hadley EN4 (R^2-value of 0.64 and CE value of 0.44). The variability in the reconstruction is damped relative to these other datasets as indicated by the lower CE values. SST values were not directly assimilated,

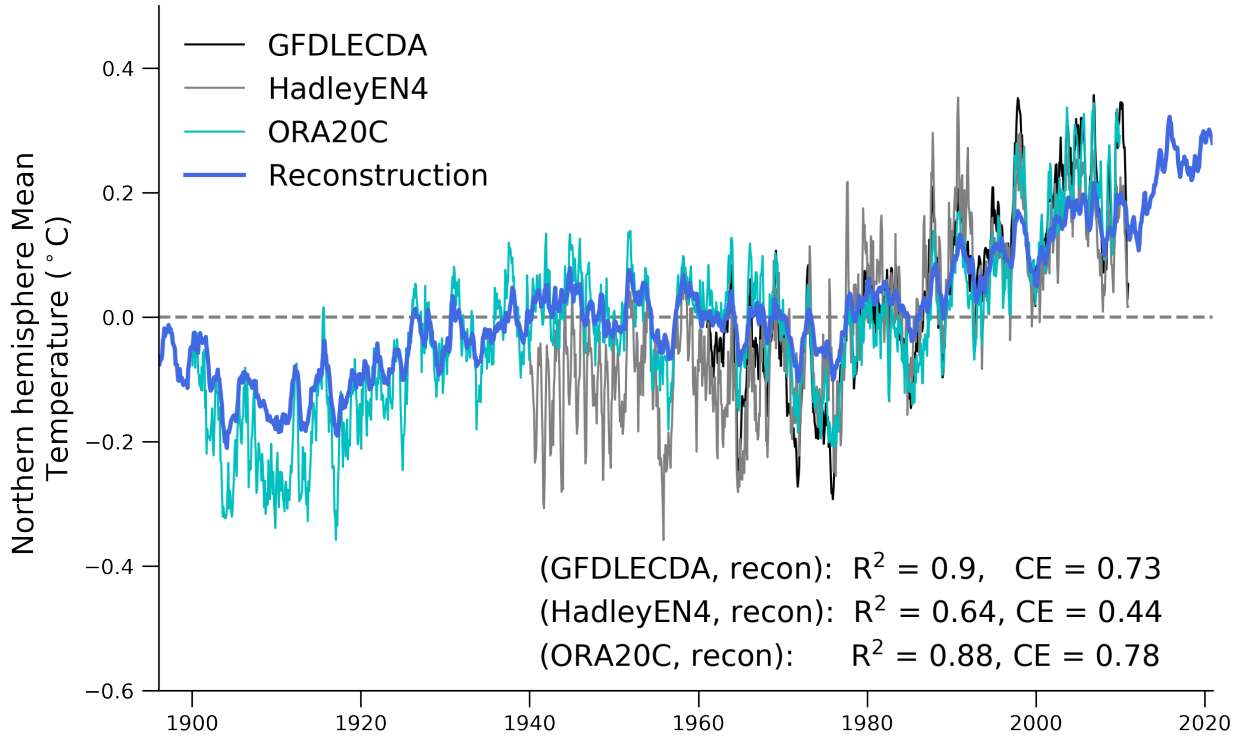

Figure 4.7: A reconstruction of monthly NH mean sea surface temperature anomalies from 1851–2020 CE is shown in blue. The NH mean sea surface temperatures from the Hadley EN4 v4.2.1 dataset (Good et al., 2013) is shown in gray, the Geophysical Fluid Dynamics Laboratory Ensemble Coupled Data Assimilation dataset (GFDLECDA, Chang et al. (2013)) is shown in black, and the European Center for Medium-range Weather Forecasting ORA-20C dataset (de Boisséson et al., 2018) is shown in turquoise. The mean of each month as been removed from all datasets shown. Anomalies are all relative to the mean calculated between 1961–2000 CE.

thus the variability is coming through the updated and forecasted covariance between TAS and SST. The high R^2-values indicate that the online DA scheme is capturing monthly fluctuations well, but the magnitude of such fluctuations is perhaps underestimated.

Figure 4.8 shows a comparison of monthly total Arctic sea ice area from our reconstructions with that from satellite observations (Fetterer et al., 2017) and (Walsh et al., 2019). Compared to satellite observations of total Arctic sea ice area, our reconstructions show high correlation on monthly timescales (R^2-value of 0.75) but damped variability (CE value

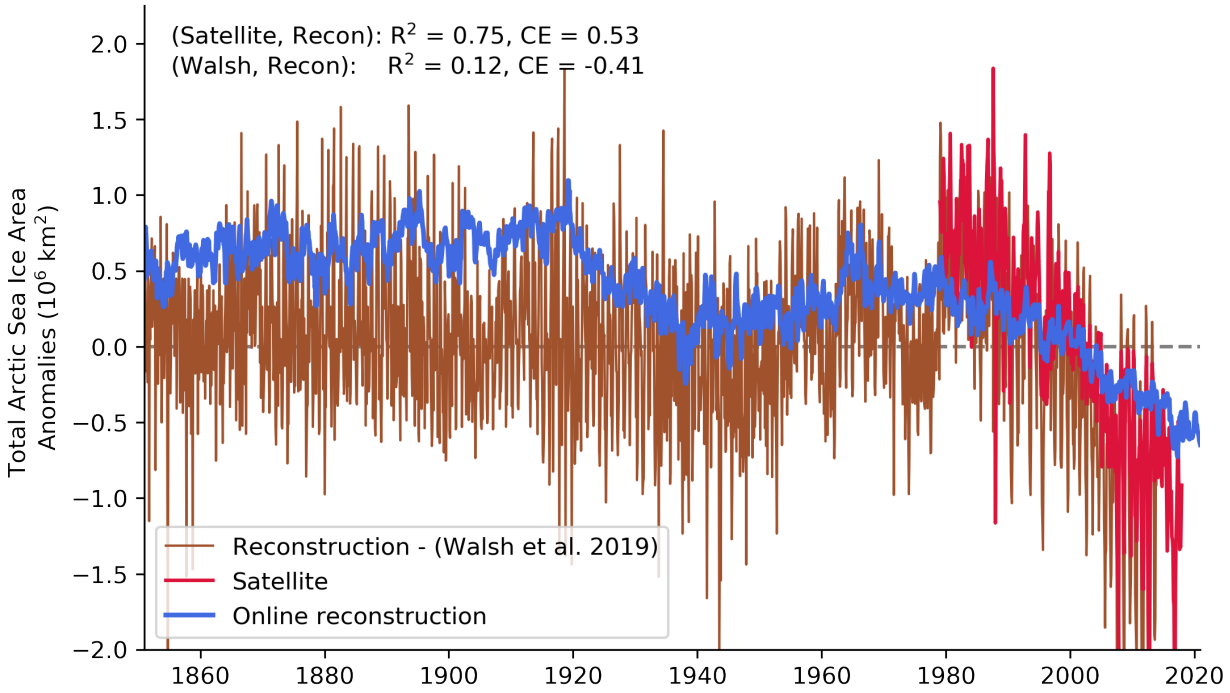

Figure 4.8: A reconstructions of monthly total Arctic sea ice area anomalies from 1851–2020 CE shown in blue produced using an online DA scheme. The total Arctic sea ice area from Walsh et al. (2019) is shown in brown and satellite observations (Fetterer et al., 2017) are shown in red. Anomalies are all relative to the mean calculated between 1979–2005 CE.

of 0.53), which indicates that further tuning of the prior and observation uncertainty or prior covariance inflation (as was used in Chapter 2 above) may lead to improved performance. Comparing the online reconstructions with Walsh et al. (2019) on monthly timescales throughout the Instrumental Era results in a weak agreement with an R^2-value of 0.12 and CE value of -0.41. The reconstructions show a stronger comparisons with other reconstructions on annual timescales (R^2 value of 0.94 and CE of 0.28 with satellite observations) as shown in Figure C.2, which also includes a comparison with Brennan and Hakim (2021).

Individual monthly anomalies in the online reconstruction and satellite observations across the satellite era are shown in Figure 4.9. We first note that the magnitude of monthly anomalies in the online reconstructions are smaller than those in satellite observations (note

difference in color bars in Figure 4.9). Also, the reconstructed anomalies vary throughout the year as seen in the top panel of Figure 4.9. Overall, the reconstructions tend to underestimate anomalies in the summer and fall (June–October). This is likely because the sensitivity of temperature and sea ice coverage in models is underestimated (Stroeve et al., 2007; Winton, 2011; Rosenblum and Eisenman, 2017) particularly in the summer and fall (Figure C.3), leading to a weaker covariance and update in the online DA scheme during these seasons.

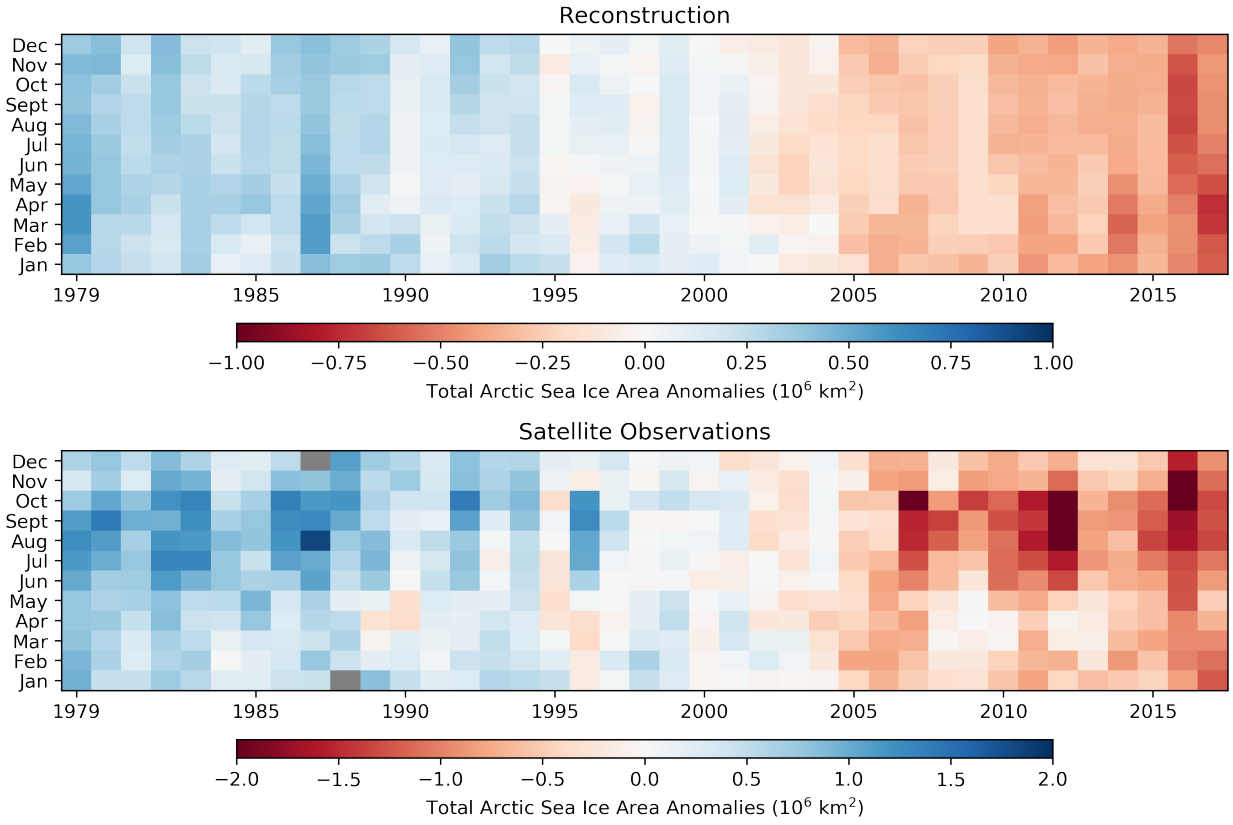

Figure 4.9: Monthly total Arctic sea ice area anomalies across all months throughout the satellite era from both an online reconstruction (top) and satellite observations (Fetterer et al., 2017). Monthly anomalies are calculated relative to the mean between 1979–2017 CE.

Finally, we compare our reconstructions of total Arctic sea ice volume to reconstructions produced using PIOMAS-20C from Schweiger et al. (2019) in Figure 4.10. We first note that

the domain configuration in the PIOMAS-20C dataset only includes regions north of 49°N so we therefore use this same domain for our calculations of total Arctic sea ice volume in our reconstructions.

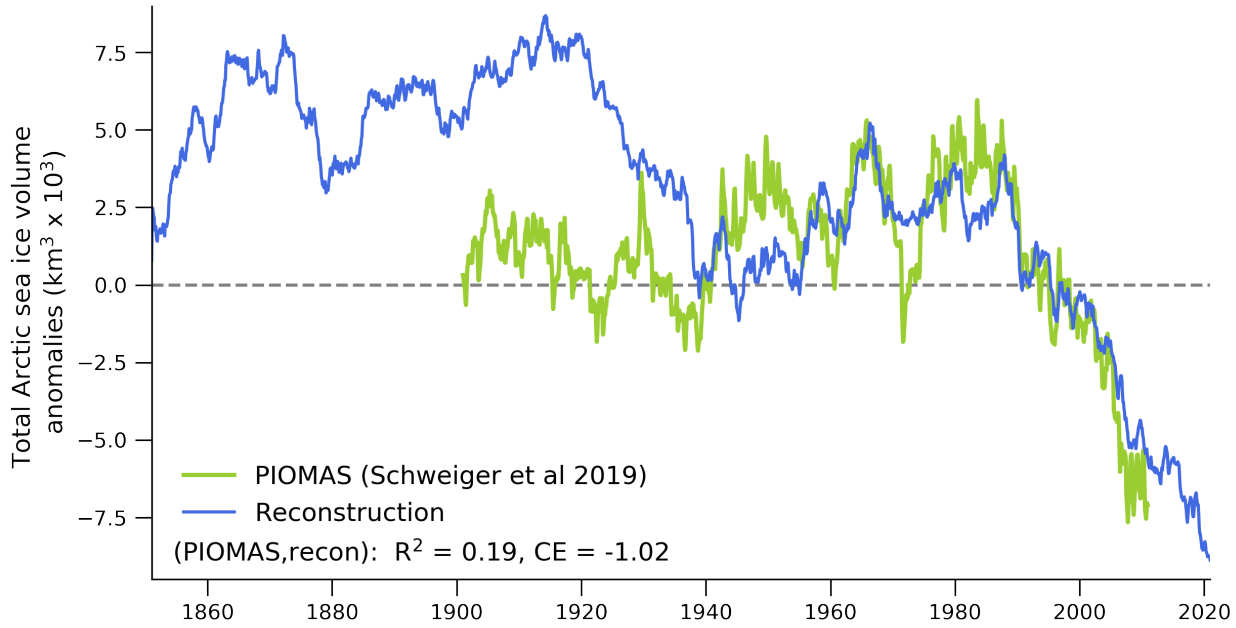

Figure 4.10: A reconstruction of monthly total Arctic sea ice volume anomalies (north of 49°N) from 1851–2020 CE shown in blue produced using an online DA scheme. The total Arctic sea ice volume from Schweiger et al. (2019) is shown in green. Anomalies are all relative to the mean calculated between 1979–2010 CE.

Our reconstruction agrees strongly with PIOMAS-20C during the satellite era (1979–2010 CE) with an R^2-value of 0.94 and CE value of 0.89. Between 2000–2010 CE, the anomalies in our reconstruction are smaller in magnitude than those in PIOMAS-20C, which may indicate that covariance inflation (as was used in Chapter 2 for SIC) is necessary for SIT in the LIM training and prior datasets, and this is left for future work. When comparing the full period of overlap between our reconstruction and PIOMAS-20C (1900–2010 CE) the R^2-value and CE values drop to 0.19 and -1.0 respectively. The magnitude of the anomalies in each reconstruction deviate from one another prior to around the year 1960. In general,

our reconstructions show significantly more sea ice volume between 1900–1940 CE and reveal a period of steady and rapid decline between 1920–1940 CE corresponding with the ETCW.

Overall these comparisons reveal high correlation with the validation datasets, indicating the the timing of monthly variability is well-captured in our reconstructions. The variability is however underestimated in fields other than temperature when compared to other records, which indicates that further tuning of the reconstructions is needed.

4.3.3 Online Reconstruction Trends

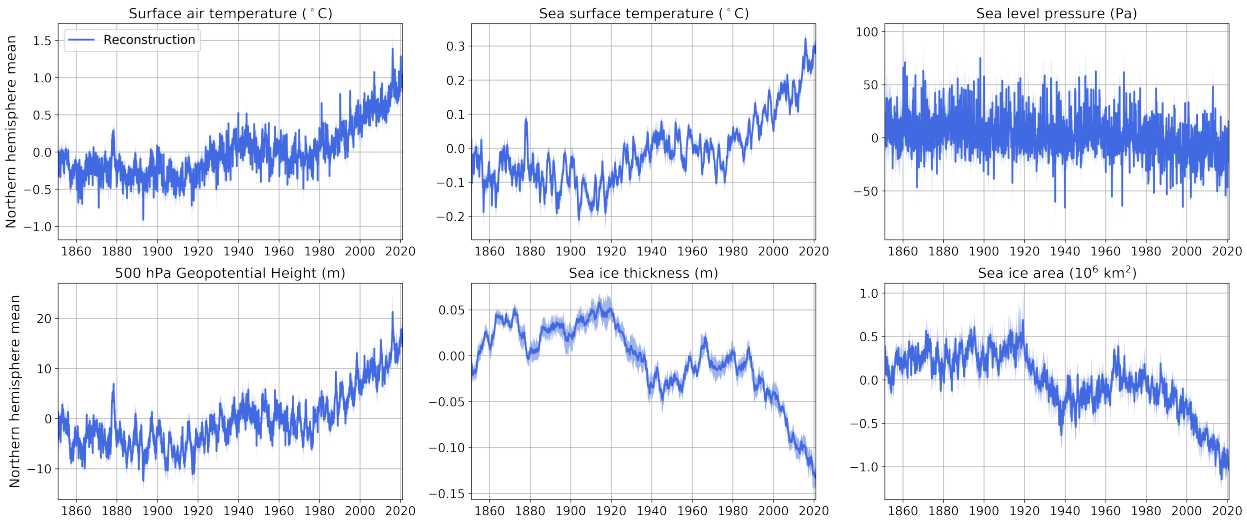

Figure 4.11: Northern hemisphere mean quantities for all six fields reconstructed for all months from 1851–2020 CE produced by assimilating HadCRUT5 temperature observations using online data assimilation. The initial prior mean state is taken as December 1850 from a CESM1 LME simulation, the LIM used in the forecast step was trained using CESM1 LME from 1650–1850 CE, and the initial prior covariance was calculated using the same training data as the LIM. At each time step, 100 HadCRUT5 observations are drawn randomly and assimilated with the prior state. 20 Monte Carlo iterations of the reconstruction are performed and the mean across all the iterations is shown in the blue line and the 2.5th–97.5th spread across these iterations is shown by the blue shaded region.

Here we present the NH mean quantities resulting from assimilating HadCRUT5 observations in the online DA scheme (Figure 4.11). Again, an initial prior mean state taken as

December of 1850 from a CESM1 LME simulation, the LIM used in the forecasting step was trained using a CESM1 LME simulation from 850–1650 CE and the initial prior covariance was also calculated from the same LIM training data. As indicated in the previous section, the magnitude of the anomalies in this reconstructions in fields other than TAS are likely damped therefore we refrain from using these reconstructions to estimate the magnitude of changes in time. We can however, compare different periods of decline within our reconstruction.

We first note that the online reconstructions show a large decline in both Arctic sea ice area as well as Arctic mean thickness during the ETCW (approximately 1915–1940 CE) that are comparable to that during the satellite era. To compare these two periods, we calculate the magnitude of trends in total Arctic sea ice area for 10–40 year windows during both the ETCW (starting in 1917) and the satellite era (ending in 2020). We find that the magnitude of trends in total Arctic sea ice area calculated during the ETCW exceed that during the satellite era for windows ranging from 10–26 years long. Similarly, the magnitude of trends in Arctic mean sea ice thickness during the ETCW exceed those calculated in the satellite era for time windows ranging from 10–16 years in length calculated in the same manner. Further work is needed to properly calibrate the magnitude of the variability in the online reconstructions, these preliminary results suggest that the ETCW was associated with sea ice declines comparable to those observed during the satellite era.

4.4 Conclusions

In this chapter, we developed an online DA framework that assimilates temperature observations with climate model data in order to reconstruct various Arctic climate fields on monthly timescales. We use a Kalman Filter to assimilate temperature observations and a LIM to forecast the assimilated state in time. Using this framework, we reconstruct six variables spanning the atmospheric, oceanic and sea ice components of the Arctic climate system on monthly time scales throughout the Instrumental Era. Full spatial fields of each variable are produced and overall we find that the online DA scheme produces skillful reconstructions.

First, both perfect and imperfect pseudo observation experiments are performed to test the skill of the online DA framework. For both sets of experiments the initial mean state, LIM training data and initial prior covariance where acquired from a CESM1 LME simulation. For perfect model experiments, pseudo observations were also drawn from a CESM1 LME at 100 random HadCRUT5 observations locations at each timestep. The perfect model reconstructions show high skill in NH mean temperature with an R^2-values of 0.9 and CE value of 0.85 when compared with the target state. For sea ice variables, the reconstructions of both Arctic mean thickness and total Arctic sea ice area compare well with the target state with R^2-values of 0.77 and 0.59 respectively. Further, positive correlation and CE values are observed at all grid cells throughout the Arctic for all six variables reconstructed.

For imperfect model experiments, pseudo observations are drawn from a CMIP6 MPI Historical simulation at 100 random HadCRUT5 observation locations at each timestep, while the covariance used to update the prior were derived from a CESM1 LME simulation. The covariance used in the DA scheme to spread the information from the observations is therefore known to differ from the target state, making this a much more challenging, but realistic test of the framework. As expected, we see decreased skill in our reconstructions for all variables when compared to the perfect model experiments. When compared with the target state, the skill in reconstructed NH mean temperature remains high with an R^2-values of 0.89 and CE value of 0.85. For sea ice variables, the reconstructions of both Arctic mean thickness and total Arctic sea ice area have an R^2-values of 0.59 and 0.50 respectively when compared with the target state. These comparisons also reveal damped variability in reconstructions of sea ice variables with CE values of 0.55 and 0.50 when comparing Arctic mean thickness and total Arctic sea ice area with their target states. When calculated at each grid cell, the correlation across time is positive almost everywhere for all variables reconstructed. Some negative CE values are observed for sea ice variables near the sea ice edge along the east coast of Greenland where two models used in the DA scheme disagree on the location of the sea ice edge. This results in sporadic sea ice coverage in the target state, but consistent sea ice coverage and variability in the training dataset leading to mean

state biases, difference in variability and thus negative CE values. Overall, these results lend confidence that the online DA scheme works (even with an imperfect forecast model) and can produce skillful reconstructions of Arctic climate conditions spanning various components of the climate system.

In order to establish whether the use of online DA improves the reconstruction skill of each variable we perform an offline and fixed covariance reconstruction using the same imperfect pseudo observation scheme described above. We found that moving from an offline to online framework nearly doubles the R^2 and CE values when comparing the reconstructed with the target state NH mean quantities. The online reconstructions perform marginally better than fixed covariance experiments by the same metrics indicating that most of the increased skill is coming from forecasting the mean. Overall, these results show that implementing the online DA framework improves the reconstruction skill on monthly timescales.

Next, we assimilate HadCRUT5 temperature observations with the initial mean state, LIM training data and initial prior covariance taken from a CESM1 LME simulation outside the period being reconstructed. This results in monthly reconstructions of six climate variables from 1851–2020 CE. The online reconstructions of NH mean temperature compare strongly with BE (R^2 and CE value of 0.95 and 0.89), GISTEMP (R^2 and CE value of 0.95 and 0.91) and HadCRUT5 (R^2 and CE value of 0.89 and 0.85) datasets. We also compare the online reconstruction of NH mean SST with other records and find weaker comparisons than that for temperature comparisons. Further the variability of the reconstructed SSTs are damped when compared to the other records (CE values ranging between 0.44–0.78). The same is true for the online reconstruction of total Arctic sea ice area. When comparing monthly reconstructions with satellite observations, the R^2-value is 0.75, however the CE value of 0.53 reveals large differences in variability. In general we find that the magnitude of the anomalies is mainly damped in the summer and fall months (June–October) in the reconstructions relative the the satellite data. We also compare our reconstructions of total Arctic sea ice volume with that from PIOMAS-20C and find strong agreement between these two datasets during the satellite era (R^2 and CE values of 0.94 and 0.89) but little agree-

ment is seen prior to the year 1960. Our reconstructions show much higher total Arctic sea ice volume anomalies between 1900–1920 CE followed by a steady and rapid decline from 1920–1940 CE corresponding with the ETCW.

Further work is needed in order to investigate ways to amplify the signal from the observations in the reconstructions. This could be done by either exploring covariance inflation techniques of the LIM training and prior datasets or through further tuning the observation and prior uncertainties in order to weight the observations more in the update step. Despite these variability issues with the reconstructions, we investigate the relative changes in Arctic sea ice during the ETCW and satellite era. These preliminary results, show that the trends calculated on annual timescales during the ETCW exceed those calculated during period in the satellite era (ending in 2000 CE). Specifically, trends in total Arctic sea ice area calculated over windows ranging from 10–40 years, starting in 1917, exceed those calculated for periods that end in 2000. The same is found for trends calculated in Arctic mean thickness during the same periods. This indicates that trends during the ETCW may have been as large as trends observed in the very recent past.

Chapter 5

CONCLUSIONS AND FUTURE WORK

5.1 Conclusions

Throughout this dissertation we have shown that assimilating temperature observations with climate model simulations is a useful tool for reconstructing Arctic climate conditions throughout the Instrumental Era. We begin our investigation by employing offline DA using an ensemble Kalman Filter to reconstruct Arctic sea ice conditions on annual time scales. Then, in order to reconstruct higher temporal resolution, we apply an online DA framework using a Kalman Filter and LIM to reconstruct monthly Arctic climate conditions throughout the Instrumental Era.

Prior to this work, the longest spatially complete reconstruction of sea ice was produced using a collection of digitized historical records that were ranked based on their reliability in all regions of the Arctic (Walsh et al., 2017, 2019). When observations were not available, Walsh et al. (2017) relied on climatology and infilling. Being the first of its kind, this reconstruction has been used extensively to force climate model simulations and as boundary conditions for 20th century reanalysis products (e.g. Rayner (2003)). However, in Chapter 2 we show that this reconstruction differs from observations, particularly prior to the 1950s. Specifically, we show that the relationship between SIE and TAS is approximately linear according to satellite observations, but the relationship is much weaker or even absent in the Walsh et al. (2017) SIE reconstruction prior to the 1950s. Prior to the 1950s, the Walsh et al. (2017) SIE reconstruction shows very little decadal to centennial scale variability, however the lack of sensitivity between SIE and TAS calls into question whether the sea ice reconstruction is reliable during this period. This further motivates the need for a sea ice reconstructions that is consistent with more well-constrained temperature and SST data.

By assimilating TAS observations with climate model simulations, we exploit the relationship between SIC and TAS in these climate model simulations to reconstruct sea ice. In Chapter 2 we use and offline ensemble Kalman Filter approach to combine HadCRUT4.6 observations with last millennium climate model simulations from CMIP5, PMIP4. These reconstructions validate strongly with satellite observations (R^2 of 0.89 and CE value of 0.89) and exhibit a large Arctic SIE loss of $\sim$1.25 x 10^6 km^2 during the ETCW. This ETCW loss in our reconstructions is greater than that shown in the Walsh et al. (2017) SIE reconstruction of $\sim$0.75–1.0 x 10^6 km^2 during the same time period, but less than the SIE loss observed during the satellite era in our reconstructions of $\sim$2.0 x 10^6 km^2. To put the satellite trends in context with the rest of the Instrumental Era, we investigated 25-year trends of SIE and find that the rate of sea-ice loss during the ETCW was about $\sim$33–38% smaller than that observed during the satellite era. Overall, our annual reconstructions show more inter-annual variability than in the Walsh et al. (2017) SIE reconstruction with standard deviation $\sim$40% ($\sim$90,000 km^2) greater between 1850–1979 due largely to difference in SIE loss during the ETCW in each reconstruction.

Next, we explored the use of a LIM to forecast sea ice and other Arctic climate variables on monthly timescales. The initial motivation for this work was to develop a model emulator that could be used for making computationally efficient forecasts in an online DA scheme, though we explore a LIM's ability to make forecasts in the Arctic across various mean state conditions. Overall we found that a LIM works well as a model emulator and is able to make skillful forecasts of Arctic climate on monthly timescales when initialized on the same model it was trained on.

We found that 500 years or more of monthly training data and 100 years or more of monthly validation data are needed in order to properly quantify the forecast skill. This large number of training years needed limits the use of satellite observations, reanalysis data, and historical simulations, thus we began by evaluating a LIM trained on a CESM1 LME simulation between 850–1650 CE. We explored how much including various atmospheric, oceanic and sea ice variables in the LIM training contributes to SIC forecast skill and found

that TAS and SST contributed the most on one to three month lead times, while SIT becames more important on four to six month lead times. However, all six variables tested contributed positively to SIC forecast skill. When validated on out-of-sample (in time) intra-model data (from CESM1 LME), the LIM forecasts outperformed an AR1 forecast for all variables for up to 8 month lead times. At a one month lead time, the LIM forecast was skillful for all variables, however at a two month lead time the LIM is no longer skillful at predicting PSL and ZG500 which we speculate is due to the faster timescales of anomaly propagation for these variables. Above an AR1 forecast, the most skill for SIC is found near the sea ice edge. Overall, the LIM is able to capture broad climate conditions that span the atmosphere, ocean and sea ice climate states particularly under pre-industrial conditions.

In order to evaluate the LIM's ability to predict modern sea ice conditions, we next validated the LIM on various different models and a reanalysis product during the Instrumental Era and found that in general the LIM fails to outperform and AR1 forecast when validated using models it was not trained on. This is likely because the LIM inherits model biases from the training datasets. In particular, for SIC, models vary greatly in their spatial pattern of variability near the sea ice edge. To quantify this we calculated the spatial correlation between the variability in time of SIC in the training data versus the validation data, and found that this quantity serves as a strong predictor of LIM forecast skill (R^2-value of 0.77). Unfortunately spatial correlation between the training data (CESM1 LME) and satellite observations and ERA5 data is low, however this metric could be useful for determining what training data would produce skillful predictions when validated using these observationally based products. Overall, we find that the LIM is a useful tool as a model emulator that requires minimal computational cost to train and make a forecast of Arctic climate conditions.

Using the LIM developed in Chapter 3, we develop and online DA framework that assimilates temperature observations with climate model data in order to reconstruct various Arctic climate fields on monthly timescales throughout the Instrumental Era. In this framework the initial mean state, LIM training data and initial prior covariance where acquired

from a CESM1 LME simulation. We first test this framework using perfect and imperfect model pseudo observations experiments. Overall, we find the framework is able to produce skillful reconstructions of all six variables included in the framework. As expected, the imperfect model results reveal lower skill for all variables when compared to the perfect model experiments. Further, we find that the variability in the reconstructions is damped for all variables, but particularly for variables other than TAS. We also find that offline reconstructions outperform online reconstructions in an imperfect pseudo observation framework, indicating that the use of online DA does improve reconstruction skill in the Arctic on monthly timescales.

Next, we assimilate real HadCRUT5 observations in the online framework resulting in monthly reconstructions of six climate variables between 1851–2020 CE. The online reconstructions of NH mean temperature compare strongly with other temperature records (R^2 values ranging from 0.89–0.95). Comparing the online reconstructions of NH mean SST with other records reveals weaker performance with R^2 values ranging from 0.64–0.9. Further the variability of the reconstructed SSTs are damped when compared to the other records (CE values ranging between 0.44–0.78).

We compare the online reconstruction of total Arctic sea ice area with that from satellite observations and find that the the R^2-value is high (0.75 for monthly data), however the CE value (0.53 for monthly data) reveals damped variability in the reconstructions mainly in the summer and fall months (June–October). This indicates that further work is needed in order to investigate ways to amplify the signal from the observations. This could be done by either exploring prior covariance inflation techniques or through further tuning the observation and prior uncertainties in order to weight the observations more in the update step. We also compare our reconstructions of total Arctic sea ice volume to those from PIOMAS-20C and find strong agreement during the satellite period (R^2 of 0.94 and CE value of 0.89) yet large discrepancies prior to the year 1960. In particular, our reconstructions show much higher positive anomalies between 1900–1920 CE followed by large steady declines during the ETCW (approximately 1920–1940 CE).

In our reconstructions of both total Arctic sea ice area and volume, there are large declines during the ETCW. Noting the damped variability in the reconstructions of area, we investigate the relative changes in Arctic sea ice during the ETCW and satellite era in our reconstructions. These preliminary results indicate that the trends during the ETCW may have been as larger than the trends observed during satellite era prior to the year 2000 CE, though further work is needed.

5.2 Future Work

Further work is needed in order to optimize the online DA framework for reconstructing Arctic sea ice. Particularly, different methods of prior covariance inflation could be explored in order to boost the variability of Arctic SIC. Inflation was used for annual offline DA reconstructions in Chapter 2 in order to replicate the trend and variability of SIE observed in satellite observations. The monthly reconstructions produced by online DA underestimate both the trend and variability of Arctic sea ice area by nearly a factor of two indicating that either prior covariance inflation or further tuning of the observation uncertainty is needed. We explored simply inflating the covariance of SIC relative to TAS at each time step after the update step in the online DA scheme and though this did not substantially boost the variability in the reconstructions. Other methods that could be explored, include inflating the magnitude of the variance in SIC relative to TAS in the LIM training datasets before training the LIM. We speculate that training the LIM on these inflated states may lead to more skillful forecasts and reconstructions.

Next, these results include reconstructions of SIT, however, we do not do any detailed validation of these fields with satellite observations that are available. There are fewer satellite products that measure SIT available and they contain a higher amount of uncertainty relative to SIC satellite measurements, though there still are products that could be used for comparison. Comparison with satellite observations of SIT could be used to optimize the variability of the thickness reconstructions via prior covariance inflation or further tuning the uncertainty in the observations.

Other ways to improve the monthly reconstructions include exploring ways to improve the monthly forecast in the online DA scheme. One way to do this could be explore ways to make the LIM less tied to one particular model. Ideas of how to decrease the influence of model bias include training a LIM on an ensemble of different model simulations, which we refer to as a multi-model LIM. Preliminary work in this area was done using four CMIP6 historical simulations linearly detrended individually and then concatenated in time. This LIM still did not outperform an AR1 forecast unless validated on model data that was included in the LIM training, however we hypothesize that using a larger ensemble of model simulations could improve a multi-model LIM's ability to produce skillful cross-model forecasts. We could also try to use a different, non-linear models to perform the forecast step in the online DA scheme. Machine learning techniques are an area left unexplored, but have shown promising results in their ability to predict Arctic sea ice (Andersson et al., 2021). Comparing the performance of machine learning techniques to a LIM could be explored and if machine learning outperforms a LIM, embedding it in the online DA scheme could lead to better reconstructions.

Beyond improving the forecast step, the online DA reconstructions could potentially be improved by assimilating a more diverse set of observations beyond just TAS. Here, we assimilatde the HadCRUT observations which include both air temperature and SST measurements to create a joint product. However, the UK Met Office also produces separate TAS (CRUTEM) and SST (HadSST) products, which we could explore assimilating separately. Furthermore, there are PSL observations available throughout the historical period as well as some quantitative records of SIC that were included in the Walsh et al. (2017) reconstruction which we could potentially assimilate directly with our DA framework.

Finally, once the online DA monthly reconstructions are further optimized, there are many scientific questions worth exploring. Future work could include further investigating the dynamical conditions that led to the ETCW and the response of Arctic sea ice to this event. Many hypotheses of what initiated the ETCW have been proposed. Bengtsson et al. (2004) found that the ETCW was associated with increased westerly winds north of Norway, which increased ocean and atmospheric heat transport into the Barents Sea, driving sea

ice loss in the region which then sparked even more warming. Wood and Overland (2009) also argued that the atmospheric circulation drove sea surface temperature anomalies in the Atlantic. Tokinaga et al. (2017) proposed the ETCW was driven by the Pacific through shifts in the phase of the Atlantic and Pacific interdecadal modes. We could investigate what our reconstructions indicate drove the ETCW and compare it to these other mechanisms that have been proposed.

BIBLIOGRAPHY

Alekseev, G., Glok, N., and Smirnov, A. (2016). On assessment of the relationship between changes of sea ice extent and climate in the Arctic. *International Journal of Climatology*, 36(9):3407–3412.

Alexander, M. A., Matrosova, L., Penland, C., Scott, J. D., and Chang, P. (2008). Forecasting Pacific SSTs: Linear Inverse Model Predictions of the PDO. *Journal of Climate*, 21(2):385–402.

Andersson, T. R., Hosking, J. S., Pérez-Ortiz, M., Paige, B., Elliott, A., Russell, C., Law, S., Jones, D. C., Wilkinson, J., Phillips, T., Byrne, J., Tietsche, S., Sarojini, B. B., Blanchard-Wrigglesworth, E., Aksenov, Y., Downie, R., and Shuckburgh, E. (2021). Seasonal Arctic sea ice forecasting with probabilistic deep learning. *Nature Communications*, 12(1):5124.

Andrews, M. B., Ridley, J. K., Wood, R. A., Andrews, T., Blockley, E. W., Booth, B., Burke, E., Dittus, A. J., Florek, P., Gray, L. J., Haddad, S., Hardiman, S. C., Hermanson, L., Hodson, D., Hogan, E., Jones, G. S., Knight, J. R., Kuhlbrodt, T., Misios, S., Mizielinski, M. S., Ringer, M. A., Robson, J., and Sutton, R. T. (2020). Historical Simulations With HadGEM3GC3.1 for CMIP6. *Journal of Advances in Modeling Earth Systems*, 12(6):1–34.

Annan, J., Hargreaves, J., Edwards, N., and Marsh, R. (2005). Parameter estimation in an intermediate complexity earth system model using an ensemble Kalman filter. *Ocean Modelling*, 8(1-2):135–154.

Armour, K. C., Eisenman, I., Blanchard-Wrigglesworth, E., McCusker, K. E., and Bitz, C. M. (2011). The reversibility of sea ice loss in a state-of-the-art climate model. *Geophysical Research Letters*, 38(16):n/a–n/a.

Atwood, A. R., Wu, E., Frierson, D. M. W., Battisti, D. S., and Sachs, J. P. (2016). Quantifying Climate Forcings and Feedbacks over the Last Millennium in the CMIP5PMIP3 Models*. *Journal of Climate*, 29(3):1161–1178.

Beitsch, A., Jungclaus, J. H., and Zanchettin, D. (2014). Patterns of decadal-scale Arctic warming events in simulated climate. *Climate Dynamics*, 43(7-8):1773–1789.

Belt, S. T., Massé, G., Rowland, S. J., Poulin, M., Michel, C., and LeBlanc, B. (2007). A novel chemical fossil of palaeo sea ice: IP25. *Organic Geochemistry*, 38(1):16–27.

Bengtsson, L., Semenov, V. A., and Johannessen, O. M. (2004). The Early Twentieth-Century Warming in the ArcticA Possible Mechanism. *Journal of Climate*, 17(20):4045–4057.

Bhend, J., Franke, J., Folini, D., Wild, M., and Brönnimann, S. (2012). An ensemble-based approach to climate reconstructions. *Climate of the Past*, 8(3):963–976.

Bitz, C. M., Holland, M. M., Hunke, E. C., and Moritz, R. E. (2005). Maintenance of the Sea-Ice Edge. *Journal of Climate*, 18(15):2903–2921.

Blanchard-Wrigglesworth, E., Armour, K. C., Bitz, C. M., and DeWeaver, E. (2011a). Persistence and Inherent Predictability of Arctic Sea Ice in a GCM Ensemble and Observations. *Journal of Climate*, 24(1):231–250.

Blanchard-Wrigglesworth, E., Barthélemy, A., Chevallier, M., Cullather, R., Fučkar, N., Massonnet, F., Posey, P., Wang, W., Zhang, J., Ardilouze, C., Bitz, C. M., Vernieres, G., Wallcraft, A., and Wang, M. (2017). Multi-model seasonal forecast of Arctic sea-ice: forecast uncertainty at pan-Arctic and regional scales. *Climate Dynamics*, 49(4):1399–1410.

Blanchard-Wrigglesworth, E. and Bitz, C. M. (2014). Characteristics of Arctic sea-ice thickness variability in GCMs. *Journal of Climate*, 27(21):8244–8258.

Blanchard-Wrigglesworth, E., Bitz, C. M., and Holland, M. M. (2011b). 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. (2019). Robustness of Arctic sea-ice predictability in GCMs. *Climate Dynamics*, 52(9-10):5555–5566.

Blanchard-Wrigglesworth, E., Cullather, R. I., Wang, W., Zhang, J., and Bitz, C. M. (2015). Model forecast skill and sensitivity to initial conditions in the seasonal Sea Ice Outlook. *Geophysical Research Letters*, 42(19):8042–8048.

Blanchard-Wrigglesworth, E. and Ding, Q. (2019). Tropical and Midlatitude Impact on Seasonal Polar Predictability in the Community Earth System Model. *Journal of Climate*, 32(18):5997–6014.

Bliss, A. C. and Anderson, M. R. (2014). Snowmelt onset over Arctic sea ice from passive microwave satellite data: 19792012. *The Cryosphere*, 8(6):2089–2100.

Bonan, D. B. and Blanchard-Wrigglesworth, E. (2020). Nonstationary Teleconnection Between the Pacific Ocean and Arctic Sea Ice. *Geophysical Research Letters*, 47(2):1–12.

Brennan, M. K. and Hakim, G. J. (2021). Reconstructing Arctic Sea Ice over the Common Era Using Data Assimilation. *Journal of Climate*, 35(4):1231–1247.

Brennan, M. K., Hakim, G. J., and BlanchardWrigglesworth, E. (2020). Arctic SeaIce Variability During the Instrumental Era. *Geophysical Research Letters*, 47(7).

Bushuk, M., Giannakis, D., and Majda, A. J. (2015). Arctic Sea Ice Reemergence: The Role of Large-Scale Oceanic and Atmospheric Variability*. *Journal of Climate*, 28(14):5477–5509.

Cavalieri, D. J. and Parkinson, C. L. (2012). Arctic sea ice variability and trends, 19792010. *The Cryosphere*, 6(4):881–889.

Cavanaugh, N. R., Allen, T., Subramanian, A., Mapes, B., Seo, H., and Miller, A. J. (2015). The skill of atmospheric linear inverse models in hindcasting the MaddenJulian Oscillation. *Climate Dynamics*, 44(3-4):897–906.

Chang, Y.-S., Zhang, S., Rosati, A., Delworth, T. L., and Stern, W. F. (2013). An assessment of oceanic variability for 19602010 from the GFDL ensemble coupled data assimilation. *Climate Dynamics*, 40(3-4):775–803.

Chevallier, M., Smith, G. C., Dupont, F., Lemieux, J. F., Forget, G., Fujii, Y., Hernandez, F., Msadek, R., Peterson, K. A., Storto, A., Toyoda, T., Valdivieso, M., Vernieres, G., Zuo, H., Balmaseda, M., Chang, Y. S., Ferry, N., Garric, G., Haines, K., Keeley, S., Kovach, R. M., Kuragano, T., Masina, S., Tang, Y., Tsujino, H., and Wang, X. (2016). Intercomparison of the Arctic sea ice cover in global oceansea ice reanalyses from the ORA-IP project. *Climate Dynamics*, 49(3):1–30.

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, Ø., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D., and Worley, S. J. (2011). The Twentieth Century Reanalysis Project. *Quarterly Journal of the Royal Meteorological Society*, 137(654):1–28.

Connolly, R., Connolly, M., and Soon, W. (2017). Re-calibration of Arctic sea ice extent datasets using Arctic surface air temperature records. *Hydrological Sciences Journal*, 62(8):1317–1340.

de Boisséson, E., Balmaseda, M. A., and Mayer, M. (2018). Ocean heat content variability in an ensemble of twentieth century ocean reanalyses. *Climate Dynamics*, 50(9-10):3783–3798.

de Vernal, A., Gersonde, R., Goosse, H., Seidenkrantz, M.-S., and Wolff, E. W. (2013). Sea

ice in the paleoclimate system: the challenge of reconstructing sea ice from proxies an introduction. *Quaternary Science Reviews*, 79:1–8.

Delworth, T. L. (2000). Simulation of Early 20th Century Global Warming. *Science*, 287(5461):2246–2250.

Dias, D. F., Subramanian, A., Zanna, L., and Miller, A. J. (2019). Remote and local influences in forecasting Pacific SST: a linear inverse model and a multimodel ensemble study. *Climate Dynamics*, 52(5-6):3183–3201.

Ding, Q., Schweiger, A., L'Heureux, M., Battisti, D. S., Po-Chedley, S., Johnson, N. C., Blanchard-Wrigglesworth, E., Harnos, K., Zhang, Q., Eastman, R., and Steig, E. J. (2017). Influence of high-latitude atmospheric circulation changes on summertime Arctic sea ice. *Nature Climate Change*, 7(4):289–295.

Dirren, S. and Hakim, G. J. (2005). Toward the assimilation of time-averaged observations. *Geophysical Research Letters*, 32(4):n/a–n/a.

Drobot, S. D., Maslanik, J. A., and Fowler, C. (2006). A long-range forecast of Arctic summer sea-ice minimum extent. *Geophysical Research Letters*, 33(10):n/a–n/a.

Dunne, J. P., Horowitz, L. W., Adcroft, A. J., Ginoux, P., Held, I. M., John, J. G., Krasting, J. P., Malyshev, S., Naik, V., Paulot, F., Shevliakova, E., Stock, C. A., Zadeh, N., Balaji, V., Blanton, C., Dunne, K. A., Dupuis, C., Durachta, J., Dussin, R., Gauthier, P. P. G., Griffies, S. M., Guo, H., Hallberg, R. W., Harrison, M., He, J., Hurlin, W., McHugh, C., Menzel, R., Milly, P. C. D., Nikonov, S., Paynter, D. J., Ploshay, J., Radhakrishnan, A., Rand, K., Reichl, B. G., Robinson, T., Schwarzkopf, D. M., Sentman, L. T., Underwood, S., Vahlenkamp, H., Winton, M., Wittenberg, A. T., Wyman, B., Zeng, Y., and Zhao, M. (2020). The GFDL Earth System Model Version 4.1 (GFDLESM 4.1): Overall Coupled Model Description and Simulation Characteristics. *Journal of Advances in Modeling Earth Systems*, 12(11):1–56.

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

England, M., Jahn, A., and Polvani, L. (2019). Nonuniform Contribution of Internal Variability to Recent Arctic Sea Ice Loss. *Journal of Climate*, 32(13):4039–4053.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9(5):1937–1958.

Fang, Z. and Wallace, J. M. (1994). Arctic Sea Ice Variability on a Timescale of Weeks and Its Relation to Atmospheric Forcing. *Journal of Climate*, 7(12):1897–1914.

Farrell, B. F. (1982). The Initial Growth of Disturbances in a Baroclinic Flow. *Journal of the Atmospheric Sciences*, 39(8):1663–1686.

Fetterer, F., Knowles, K., Meier, W. N., Savoie, M., and Windnagel, A. K. (2017). Sea Ice Index, Version 3. *NSIDC: National Snow and Ice Data Center, Boulder, Colorado USA*.

Fyfe, J. C., von Salzen, K., Gillett, N. P., Arora, V. K., Flato, G. M., and McConnell, J. R. (2013). One hundred years of Arctic surface temperature variation due to anthropogenic influence. *Scientific Reports*, 3(1):2645.

Gaspari, G. and Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125(554):723–757.

Giesse, C., Notz, D., and Baehr, J. (2021). On the Origin of Discrepancies Between Observed and Simulated Memory of Arctic Sea Ice. *Geophysical Research Letters*, 48(11).

Good, S. A., Martin, M. J., and Rayner, N. A. (2013). EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, 118(12):6704–6716.

Goosse, H., Renssen, H., Timmermann, A., Bradley, R. S., and Mann, M. E. (2006). Using paleoclimate proxy-data to select optimal realisations in an ensemble of simulations of the climate of the past millennium. *Climate Dynamics*, 27(2-3):165–184.

Goosse, H., Roche, D., Mairesse, A., and Berger, M. (2013). Modelling past sea ice changes. *Quaternary Science Reviews*, 79:191–206.

Gregory, J. M., Stott, P. A., Cresswell, D. J., Rayner, N. A., Gordon, C., and Sexton, D. M. H. (2002). Recent and future changes in Arctic sea ice simulated by the HadCM3 AOGCM. *Geophysical Research Letters*, 29(24):28–1–28–4.

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., Koenigk, T., Salas y Mélia, D., and Tietsche, S. (2016). A review on Arctic sea-ice predictability and prediction on seasonal to decadal time-scales. *Quarterly Journal of the Royal Meteorological Society*, 142(695):546–561.

Hakim, G. J., Emile-Geay, J., Steig, E. J., Noone, D., Anderson, D. M., Tardif, R., Steiger, N., and Perkins, W. A. (2016). The last millennium climate reanalysis project: Framework and first results. *Journal of Geophysical Research*, 121(12):6745–6764.

Hakim, G. J., Snyder, C., Penny, S. G., and Newman, M. (2022). Subseasonal Forecast Skill Improvement from Strongly Coupled Data Assimilation with a Linear Inverse Model. *Geophys. Res. Lett.*, pages 1–10.

Hall, C. M. and Saarinen, J. (2010). Polar Tourism: Definitions and Dimensions. *Scandinavian Journal of Hospitality and Tourism*, 10(4):448–467.

Hamill, T. M., Whitaker, J. S., and Snyder, C. (2001). Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter. *Monthly Weather Review*, 129(11):2776–2790.

Hansen, J., Ruedy, R., Sato, M., and Lo, K. (2010). Global Surface Temperature Change. *Reviews of Geophysics*, 48(4):RG4004.

Hegerl, G. C., Brönnimann, S., Schurer, A., and Cowan, T. (2018). The early 20th century warming: Anomalies, causes, and consequences. *Wiley Interdisciplinary Reviews: Climate Change*, 9(4):e522.

Henderson, S. A., Vimont, D. J., and Newman, M. (2020). The Critical Role of Non-Normality in Partitioning Tropical and Extratropical Contributions to PNA Growth. *Journal of Climate*, 33(14):6273–6295.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., MuñozSabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049.

Hogg, J., Fonoberova, M., and Mezić, I. (2020). Exponentially decaying modes and long-term prediction of sea ice concentration using Koopman mode decomposition. *Scientific Reports*, 10(1):16313.

Holland, M. M., Landrum, L., Bailey, D., and Vavrus, S. (2019). Changing Seasonal Predictability of Arctic Summer Sea Ice Area in a Warming Climate. *Journal of Climate*, 32(16):4963–4979.

Holland, M. M. and Stroeve, J. (2011). Changing seasonal sea ice predictor relationships in a changing Arctic climate. *Geophysical Research Letters*, 38(18):n/a–n/a.

Houtekamer, P. L. and Mitchell, H. L. (1998). Data Assimilation Using an Ensemble Kalman Filter Technique. *Monthly Weather Review*, 127(6):1374–1377.

Huddart, B., Subramanian, A., Zanna, L., and Palmer, T. (2017). Seasonal and decadal forecasts of Atlantic Sea surface temperatures using a linear inverse model. *Climate Dynamics*, 49(5-6):1833–1845.

Jahn, A., Kay, J. E., Holland, M. M., and Hall, D. M. (2016). Geophysical Research Letters How predictable is the timing of a summer ice-free Arctic? *Geophys. Res. Lett.*, 43:9113–9120.

Jahn, A., Sterling, K., Holland, M. M., Kay, J. E., Maslanik, J. A., Bitz, C. M., Bailey, D. A., Stroeve, J., Hunke, E. C., Lipscomb, W. H., and Pollak, D. A. (2012). Late-twentieth-century simulation of arctic sea ice and ocean properties in the CCSM4. *Journal of Climate*, 25(5):1431–1452.

Johannessen, O. M., Bengtsson, L., Miles, M. W., Kuzmina, S. I., Semenov, V. A., Alekseev, G. V., Nagurnyi, A. P., Zakharov, V. F., Bobylev, L. P., Pettersson, L. H., Hasselmann, K., and Cattle, H. P. (2004). Arctic climate change: observed and modelled temperature and sea-ice variability. *Tellus A*, 56(4):328–341.

Jung, T., Gordon, N. D., Bauer, P., Bromwich, D. H., Chevallier, M., Day, J. J., Dawson, J., Doblas-Reyes, F., Fairall, C., Goessling, H. F., Holland, M., Inoue, J., Iversen, T., Klebe, S., Lemke, P., Losch, M., Makshtas, A., Mills, B., Nurmi, P., Perovich, D., Reid, P., Renfrew, I. A., Smith, G., Svensson, G., Tolstykh, M., and Yang, Q. (2016). Advancing Polar Prediction Capabilities on Daily to Seasonal Time Scales. *Bulletin of the American Meteorological Society*, 97(9):1631–1647.

Jungclaus, J., Giorgetta, M., Reick, C., Legutke, S., Brovkin, V., Crueger, T., Esch, M., Fieg, K., Fischer, N., Glushak, K., Gayler, V., Haak, H., Hollweg, H.-D., Kinne, S., Kornblueh, L., Matei, D., Mauritsen, T., Mikolajewicz, U., Müller, W., Notz, D., Pohlmann, T.,

Raddatz, T., Rast, S., Roeckner, E., Salzmann, M., Schmidt, H., Schnur, R., Segschneider, J., Six, K., Stockhause, M., Wegner, J., Widmann, H., Wieners, K.-H., Claussen, M., Marotzke, J., and Stevens, B. (2012). CMIP5 simulations of the Max Planck Institute for Meteorology (MPI-M) based on the MPI-ESM-P model: The past1000 experiment, served by ESGF. *World Data Center for Climate (WDCC) at DKRZ*.

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., Arblaster, J. M., Bates, S. C., Danabasoglu, G., Edwards, J., Holland, M., Kushner, P., Lamarque, J.-F., Lawrence, D., Lindsay, K., Middleton, A., Munoz, E., Neale, R., Oleson, K., Polvani, L., and Vertenstein, M. (2015). The Community Earth System Model (CESM) Large Ensemble Project: A Community Resource for Studying Climate Change in the Presence of Internal Climate Variability. *Bulletin of the American Meteorological Society*, 96(8):1333–1349.

Kay, J. E., Holland, M. M., and Jahn, A. (2011). Inter-annual to multi-decadal Arctic sea ice extent trends in a warming world. *Geophysical Research Letters*, 38(15):2–7.

Kjeldsen, K. K., Korsgaard, N. J., Bjørk, A. A., Khan, S. A., Box, J. E., Funder, S., Larsen, N. K., Bamber, J. L., Colgan, W., Broeke, M. V. D., Nuth, C., Schomacker, A., Andresen, C. S., Willerslev, E., and Kjær, K. H. (2015). the Greenland Ice Sheet since ad 1900. *Nature*.

Klein, F., Goosse, H., Mairesse, A., and de Vernal, A. (2014). Modeldata comparison and data assimilation of mid-Holocene Arctic sea ice concentration. *Climate of the Past*, 10(3):1145–1163.

Kwok, R. (2018). Arctic sea ice thickness, volume, and multiyear ice coverage: losses and coupled variability (19582018). *Environmental Research Letters*, 13(10):105005.

Landrum, L., Otto-Bliesner, B. L., Wahl, E. R., Conley, A., Lawrence, P. J., Rosenbloom,

N., and Teng, H. (2013). Last Millennium Climate and Its Variability in CCSM4. *Journal of Climate*, 26(4):1085–1111.

Lemke, P., Trinkl, E. W., and Hasselmann, K. (1980). Stochastic Dynamic Analysis of Polar Sea Ice Variability. *Journal of Physical Oceanography*, 10(12):2100–2120.

Lin, G., Penner, J. E., Flanner, M. G., Sillman, S., Xu, L., and Zhou, C. (2014). Radiative forcing of organic aerosol in the atmosphere and on snow: Effects of SOA and brown carbon. *Journal of Geophysical Research: Atmospheres*, 119(12):7453–7476.

Lindsay, R. and Schweiger, A. (2015). Arctic sea ice thickness loss determined using subsurface, aircraft, and satellite observations. *The Cryosphere*, 9(1):269–283.

Lindsay, R., Wensnahan, M., Schweiger, A., and Zhang, J. (2014). Evaluation of Seven Different Atmospheric Reanalysis Products in the Arctic*. *Journal of Climate*, 27(7):2588–2606.

Lindsay, R. W., Zhang, J., Schweiger, A. J., and Steele, M. A. (2008). Seasonal predictions of ice extent in the Arctic Ocean. *Journal of Geophysical Research*, 113(C2):C02023.

Mahlstein, I. and Knutti, R. (2012). September Arctic sea ice predicted to disappear near 2C global warming above present. *Journal of Geophysical Research: Atmospheres*, 117(D6):n/a–n/a.

Markus, T., Stroeve, J. C., and Miller, J. (2009). Recent changes in Arctic sea ice melt onset, freezeup, and melt season length. *Journal of Geophysical Research: Oceans*, 114(12):1–14.

Maslanik, J. A., Fowler, C., Stroeve, J., Drobot, S., Zwally, J., Yi, D., and Emery, W. (2007). A younger, thinner Arctic ice cover: Increased potential for rapid, extensive sea-ice loss. *Geophysical Research Letters*, 34(24):L24501.

Massonnet, F., Fichefet, T., Goosse, H., Bitz, C. M., Philippon-Berthier, G., Holland,

M. M., and Barriat, P.-Y. (2012). Constraining projections of summer Arctic sea ice. *The Cryosphere*, 6(6):1383–1394.

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., Brovkin, V., Claussen, M., Crueger, T., Esch, M., Fast, I., Fiedler, S., Fläschner, D., Gayler, V., Giorgetta, M., Goll, D. S., Haak, H., Hagemann, S., Hedemann, C., Hohenegger, C., Ilyina, T., Jahns, T., JimenézdelaCuesta, D., Jungclaus, J., Kleinen, T., Kloster, S., Kracher, D., Kinne, S., Kleberg, D., Lasslop, G., Kornblueh, L., Marotzke, J., Matei, D., Meraner, K., Mikolajewicz, U., Modali, K., Möbis, B., Müller, W. A., Nabel, J. E. M. S., Nam, C. C. W., Notz, D., Nyawira, S., Paulsen, H., Peters, K., Pincus, R., Pohlmann, H., Pongratz, J., Popp, M., Raddatz, T. J., Rast, S., Redler, R., Reick, C. H., Rohrschneider, T., Schemann, V., Schmidt, H., Schnur, R., Schulzweida, U., Six, K. D., Stein, L., Stemmler, I., Stevens, B., Storch, J., Tian, F., Voigt, A., Vrese, P., Wieners, K., Wilkenskjeld, S., Winkler, A., and Roeckner, E. (2019). Developments in the MPIM Earth System Model version 1.2 (MPIESM1.2) and Its Response to Increasing CO 2. *Journal of Advances in Modeling Earth Systems*, 11(4):998–1038.

Meier, W. N., Hovelsrud, G. K., van Oort, B. E., Key, J. R., Kovacs, K. M., Michel, C., Haas, C., Granskog, M. A., Gerland, S., Perovich, D. K., Makshtas, A., and Reist, J. D. (2014). Arctic sea ice in transformation: A review of recent observed changes and impacts on biology and human activity. *Reviews of Geophysics*, 52(3):185–217.

Morice, C. P., Kennedy, J. J., Rayner, N. A., and Jones, P. D. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *Journal of Geophysical Research: Atmospheres*, 117(D8):n/a–n/a.

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., Dunn, R. J. H., Osborn, T. J., Jones, P. D., and Simpson, I. R. (2020). An Updated Assessment

of NearSurface Temperature Change From 1850: The HadCRUT5 Data Set. *Journal of Geophysical Research: Atmospheres*, 126(3):1–28.

Nash, J. and Sutcliffe, J. (1970). River flow forecasting through conceptual models part I A discussion of principles. *Journal of Hydrology*, 10(3):282–290.

Newman, M. (2007). Interannual to Decadal Predictability of Tropical and North Pacific Sea Surface Temperatures. *Journal of Climate*, 20(11):2333–2356.

Newman, M., Sardeshmukh, P. D., and Penland, C. (2009). How Important Is AirSea Coupling in ENSO and MJO Evolution? *Journal of Climate*, 22(11):2958–2977.

Notz, D. and Marotzke, J. (2012). Observations reveal external driver for Arctic sea-ice retreat. *Geophysical Research Letters*, 39(8):n/a–n/a.

Oke, P. R., Allen, J., Miller, R. N., Egbert, G. D., and Kosro, P. M. (2002). Assimilation of surface velocity data into a primitive equation coastal ocean model. *Journal of Geophysical Research*, 107(C9):3122.

Olonscheck, D., Mauritsen, T., and Notz, D. (2019). Arctic sea-ice variability is primarily driven by atmospheric temperature fluctuations. *Nature Geoscience*, 12(6):430–434.

Otto-Bliesner, B. L., Brady, E. C., Fasullo, J., Jahn, A., Landrum, L., Stevenson, S., Rosenbloom, N., Mai, A., and Strand, G. (2016). Climate Variability and Change since 850 CE: An Ensemble Approach with the Community Earth System Model. *Bulletin of the American Meteorological Society*, 97(5):735–754.

Overland, J. E., Spillane, M. C., Percival, D. B., Wang, M., and Mofjeld, H. O. (2004). Seasonal and Regional Variation of Pan-Arctic Surface Air Temperature over the Instrumental Record*. *Journal of Climate*, 17(17):3263–3282.

Penland, C. (1989). Random Forcing and Forecasting Using Principal Oscillation Pattern Analysis. *Monthly Weather Review*, 117(10):2165–2185.

Penland, C. (1996). A stochastic model of IndoPacific sea surface temperature anomalies. *Physica D: Nonlinear Phenomena*, 98(2-4):534–558.

Penland, C. and Matrosova, L. (1994). A Balance Condition for Stochastic Numerical Models with Application to the El Niño-Southern Oscillation. *Journal of Climate*, 7(9):1352–1372.

Penland, C. and Matrosova, L. (1998). Prediction of Tropical Atlantic Sea Surface Temperatures Using Linear Inverse Modeling. *Journal of Climate*, 11(3):483–496.

Penland, C. and Sardeshmukh, P. D. (1995). The Optimal Growth of Tropical Sea Surface Temperature Anomalies. *Journal of Climate*, 8(8):1999–2024.

Perkins, W. A. and Hakim, G. (2020). Linear Inverse Modeling for Coupled Atmosphere-Ocean Ensemble Climate Prediction. *Journal of Advances in Modeling Earth Systems*, 12(1).

Perkins, W. A. and Hakim, G. J. (2017). Reconstructing paleoclimate fields using online data assimilation with a linear inverse model. *Climate of the Past*, 13(5):421–436.

Perkins, W. A. and Hakim, G. J. (2021). Coupled AtmosphereOcean Reconstruction of the Last Millennium Using Online Data Assimilation. *Paleoceanography and Paleoclimatology*, 36(5):1–29.

Petty, A. A., Schröder, D., Stroeve, J. C., Markus, T., Miller, J., Kurtz, N. T., Feltham, D. L., and Flocco, D. (2017). Skillful spring forecasts of September Arctic sea ice extent using passive microwave sea ice observations. *Earth's Future*, 5(2):254–263.

Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., Laloyaux, P., Tan, D. G. H., Peubey, C., Thépaut, J.-N., Trémolet, Y., Hólm, E. V., Bonavita, M., Isaksen, L., and Fisher, M. (2016). ERA-20C: An Atmospheric Reanalysis of the Twentieth Century. *Journal of Climate*, 29(11):4083–4097.

Polyak, L., Alley, R. B., Andrews, J. T., Brigham-Grette, J., Cronin, T. M., Darby, D. A., Dyke, A. S., Fitzpatrick, J. J., Funder, S., Holland, M., Jennings, A. E., Miller, G. H., O'Regan, M., Savelle, J., Serreze, M., St. John, K., White, J. W., and Wolff, E. (2010). History of sea ice in the Arctic. *Quaternary Science Reviews*, 29(15-16):1757–1778.

Polyakov, I. V., Alekseev, G. V., Bekryaev, R. V., Bhatt, U. S., Colony, R., Johnson, M. A., Karklin, V. P., Walsh, D., and Yulin, A. V. (2003). Long-Term Ice Variability in Arctic Marginal Seas. *Journal of Climate*, 16(12):2078–2085.

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):4407.

Ridgwell, A., Hargreaves, J. C., Edwards, N. R., Annan, J. D., Lenton, T. M., Marsh, R., Yool, A., and Watson, A. (2007). Marine geochemical data assimilation in an efficient Earth System Model of global biogeochemical cycling. *Biogeosciences*, 4(1):87–104.

Rigor, I. G., Wallace, J. M., and Colony, R. L. (2002). Response of Sea Ice to the Arctic Oscillation. *Journal of Climate*, 15(18):2648–2663.

Rohde, R., Muller, R., Jacobsen, R., Perlmutter, S., and Mosher, S. (2013a). Berkeley Earth Temperature Averaging Process. *Geoinformatics & Geostatistics: An Overview*, 01(02).

Rohde, R., Muller, R. A., Jacobsen, R., Muller, E., Perlmutter, S., Rosenfeld, A., Wurtele, J., Groom, D., and Wickham, C. (2013b). A New Estimate of the Average Earth Surface Land Temperature Spanning 1753 to 2011. *Geoinformatics & Geostatistics: An Overview*, 01(01):1–7.

Rosenblum, E. and Eisenman, I. (2017). Sea Ice Trends in Climate Models Only Accurate in Runs with Biased Global Warming. *Journal of Climate*, 30(16):6265–6278.

Schmidt, G. A., Jungclaus, J. H., Ammann, C. M., Bard, E., Braconnot, P., Crowley, T. J., Delaygue, G., Joos, F., Krivova, N. A., Muscheler, R., Otto-Bliesner, B. L., Pongratz, J.,

Shindell, D. T., Solanki, S. K., Steinhilber, F., and Vieira, L. E. A. (2011). Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.0). *Geoscientific Model Development*, 4(1):33–45.

Schweiger, A., Lindsay, R., Zhang, J., Steele, M., Stern, H., and Kwok, R. (2011). Uncertainty in modeled Arctic sea ice volume. *Journal of Geophysical Research*, 116(9):C00D06.

Schweiger, A. J., Wood, K. R., and Zhang, J. (2019). Arctic Sea Ice Volume Variability over 19012010: A Model-Based Reconstruction. *Journal of Climate*, 32(15):4731–4752.

Semenov, V. A. (2007). Structure of temperature variability in the high latitudes of the Northern Hemisphere. *Izvestiya - Atmospheric and Ocean Physics*, 43(6):687–695.

Semenov, V. A. and Latif, M. (2012). The early twentieth century warming and winter Arctic sea ice. *The Cryosphere*, 6(6):1231–1237.

Serreze, M. C., Holland, M. M., and Stroeve, J. (2007). Perspectives on the Arctic's Shrinking Sea-Ice Cover. *Science*, 315(5818):1533–1536.

Serreze, M. C. and Stroeve, J. (2015). Arctic sea ice trends, variability and implications for seasonal ice forecasting. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 373(2045):20140159.

Shin, S.-i., Sardeshmukh, P. D., Newman, M., Penland, C., and Alexander, M. A. (2021). Impact of Annual Cycle on ENSO Variability and Predictability. *Journal of Climate*, 34(1):171–193.

Singh, H. K. A., Hakim, G. J., Tardif, R., Emile-Geay, J., and Noone, D. C. (2018). Insights into Atlantic multidecadal variability using the Last Millennium Reanalysis framework. *Climate of the Past*, 14(2):157–174.

Smith, L. C. and Stephenson, S. R. (2013). New Trans-Arctic shipping routes navigable by midcentury. *Proceedings of the National Academy of Sciences*, 110(13):E1191–E1195.

Steiger, N. J., Hakim, G. J., Steig, E. J., Battisti, D. S., and Roe, G. H. (2014). Assimilation of time-averaged pseudoproxies for climate reconstruction. *Journal of Climate*, 27(1):426–441.

Stroeve, J., Hamilton, L. C., Bitz, C. M., and Blanchard-Wrigglesworth, E. (2014). Predicting September sea ice: Ensemble skill of the SEARCH Sea Ice Outlook 2008-2013. *Geophysical Research Letters*, 41(7):2411–2418.

Stroeve, J., Holland, M. M., Meier, W., Scambos, T., and Serreze, M. (2007). Arctic sea ice decline: Faster than forecast. *Geophysical Research Letters*, 34(9):1–5.

Stroeve, J. and Notz, D. (2018). Changing state of Arctic sea ice across all seasons. *Environmental Research Letters*, 13(10):103001.

Stroeve, J. C., Serreze, M. C., Holland, M. M., Kay, J. E., Malanik, J., and Barrett, A. P. (2012). The Arctic's rapidly shrinking sea ice cover: a research synthesis. *Climatic Change*, 110(3-4):1005–1027.

Svendsen, L., Keenlyside, N., Bethke, I., Gao, Y., and Omrani, N. E. (2018). Pacific contribution to the early twentieth-century warming in the Arctic. *Nature Climate Change*, 8(9):793–797.

Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Hanna, S., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Sigmond, M., Solheim, L., von Salzen, K., Yang, D., and Winter, B. (2019). The Canadian Earth System Model version 5 (CanESM5.0.3). *Geoscientific Model Development*, 12(11):4823–4873.

Swart, N. C., Fyfe, J. C., Hawkins, E., Kay, J. E., and Jahn, A. (2015). Influence of internal variability on Arctic sea-ice trends. *Nature Climate Change*, 5(February).

Tardif, R., Hakim, G. J., Perkins, W. A., Horlick, K. A., Erb, M. P., Emile-Geay, J., Anderson, D. M., Steig, E. J., and Noone, D. (2019). Last Millennium Reanalysis with an

expanded proxy database and seasonal proxy modeling. *Climate of the Past*, 15(4):1251–1273.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A. (2012). An Overview of CMIP5 and the Experiment Design. *American Meteorological Society*.

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., Alt, B., Howell, S., Wilson, K., and Yackel, J. (2007). Long-Range Prediction of the Shipping Season in Hudson Bay: A Statistical Approach. *Weather and Forecasting*, 22(5):1063–1075.

Tokinaga, H., Xie, S.-P., and Mukougawa, H. (2017). Early 20th-century Arctic warming intensified by Pacific and Atlantic multidecadal variability. *Proceedings of the National Academy of Sciences*, 114(24):6227–6232.

Walsh, J. E. and Chapman, W. L. (2001). 20Th-Century Sea-Ice Variations From Observational Data. *Annals of Glaciology*, 33(1979):444–448.

Walsh, J. E., Chapman, W. L., Fetterer, F., and Stewart., J. S. (2019). Gridded Monthly Sea Ice Extent and Concentration, 1850 Onward, Version 2. *Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center.*

Walsh, J. E., Fetterer, F., Scott stewart, J., and Chapman, W. L. (2017). A database for depicting Arctic sea ice variations back to 1850. *Geographical Review*, 107(1):89–107.

Walsh, J. E. and Johnson, C. M. (1979). Interannual atmospheric variability and associated fluctuations in Arctic Sea ice extent. *Journal of Geophysical Research*, 84(C11):6915.

Wang, L., Yuan, X., and Li, C. (2019). Subseasonal forecast of Arctic sea ice concentration via statistical approaches. *Climate Dynamics*, 52(7):4953–4971.

Wegmann, M., Brönnimann, S., and Compo, G. P. (2017). Tropospheric circulation during the early twentieth century Arctic warming. *Climate Dynamics*, 48(7-8):2405–2418.

Whitaker, J. S. and Hamill, T. M. (2002). Ensemble Data Assimilation without Perturbed Observations. *Monthly Weather Review*, 130(7):1913–1924.

Widmann, M., Goosse, H., van der Schrier, G., Schnur, R., and Barkmeijer, J. (2010). Using data assimilation to study extratropical Northern Hemisphere climate over the last millennium. *Climate of the Past*, 6(5):627–644.

Winkler, C. R., Newman, M., and Sardeshmukh, P. D. (2001). A Linear Model of Wintertime Low-Frequency Variability. Part I: Formulation and Forecast Skill. *Journal of Climate*, 14(24):4474–4494.

Winton, M. (2011). Do climate models underestimate the sensitivity of Northern Hemisphere sea ice cover? *Journal of Climate*, 24(15):3924–3934.

Wolff, E. W., Rankin, A. M., and Röthlisberger, R. (2003). An ice core indicator of Antarctic sea ice production? *Geophysical Research Letters*, 30(22):2–5.

Wood, K. R. and Overland, J. E. (2009). Early 20 th century Arctic warming in retrospect. *International Journal of Climatology*, 30(9):n/a–n/a.

Yamanouchi, T. (2011). Early 20th century warming in the Arctic : A review. *Polar Science*, 5(1):53–71.

Yuan, X., Chen, D., Li, C., Wang, L., and Wang, W. (2016). Arctic Sea Ice Seasonal Prediction by a Linear Markov Model. *Journal of Climate*, 29(22):8151–8173.

Zhang, J. and Rothrock, D. A. (2003). Modeling Global Sea Ice with a Thickness and Enthalpy Distribution Model in Generalized Curvilinear Coordinates. *Monthly Weather Review*, 131(5):845–861.

Appendix A

CHAPTER 2 SUPPORTING MATERIAL

A.1 The early 20th century warming in reanalysis

Figure A.1 shows a comparison between annually averaged Arctic (north of 70N) mean temperature observations from HadCRUT and reanalysis data (NOAA-20C and ERA-20C) during the 20th century. Between 1953–2011 there is good agreement between HadCRUT and ERA-20C with an R^2-value of 0.85 and R^2-value of 0.41 between HadCRUT and NOAA-20C. In contrast before, 1953 these two records diverge with an R^2-value of 0.27 and 0.01 respectively. Particularly from 1900–1953, ERA-20C temperature anomalies hover just below 0°C and NOAA-20C shows very cold anomalous temperatures of around -3°C, while HadCRUT increases from approximately -2°C to 1°C over the same time period. These discrepancies illustrate that neither of these reanalysis products capture the early 20th century warming.

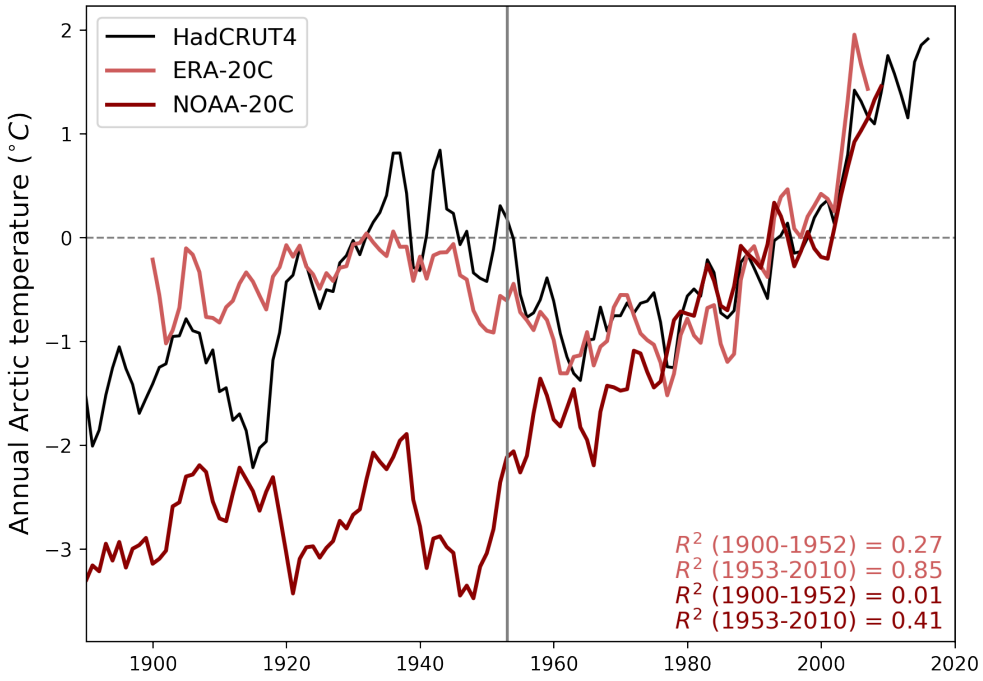

Figure A.1: Arctic (north of 70N) mean surface air temperatures anomalies from HadCRUT, NOAA-20C, and ERA-20C. The vertical gray line indicates the year 1953, when availability of observations of sea ice in the Arctic decline substantially in the (Walsh et al., 2017) record. Anomalies are centered about 1979-2011.

Appendix B

CHAPTER 3 SUPPORTING MATERIAL

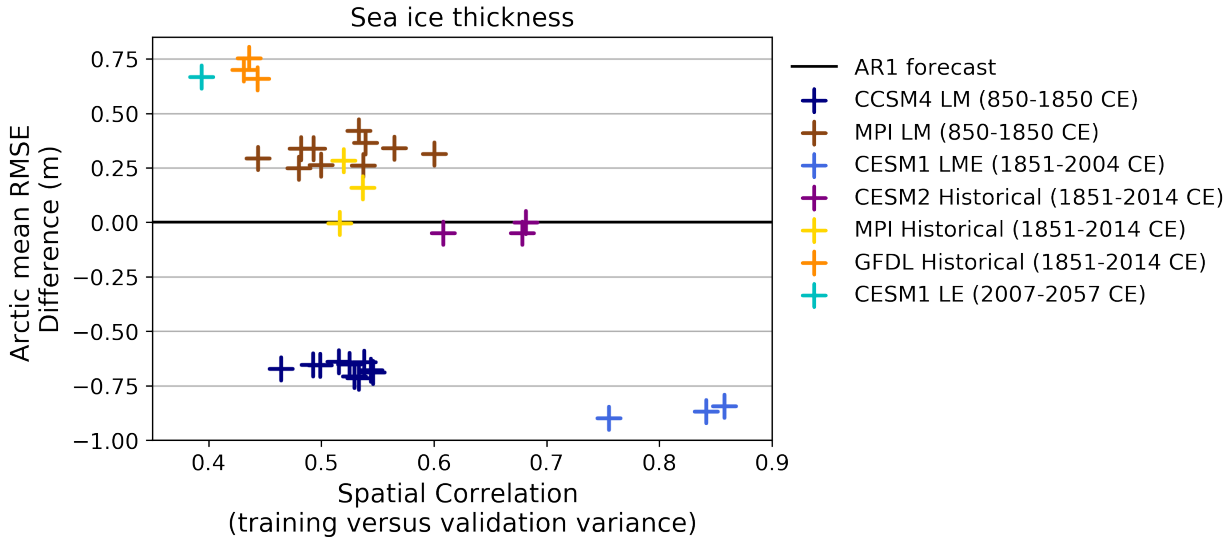

Figure B.1: Arctic mean root mean squared error (RMSE) for one month LIM forecasts of sea ice thickness validated on various different model simulations (different colors) versus the spatial correlation between the variance (across time) of each validation dataset and the LIM training dataset. The Arctic mean RMSE difference is the RMSE value of the LIM forecast minus the RMSE of an autoregressive model of order one (AR1), thus negative values indicate when the LIM forecast outperforms an AR1 forecast. For all forecasts, the LIM was trained using a CESM1 LME simulation from 850–1650 CE. Values were calculated for 100 year non-overlapping segments for last millennium simulations (LM) and all 50 year non-overlapping segments for historical simulations.

Appendix C

CHAPTER 4 SUPPORTING MATERIAL

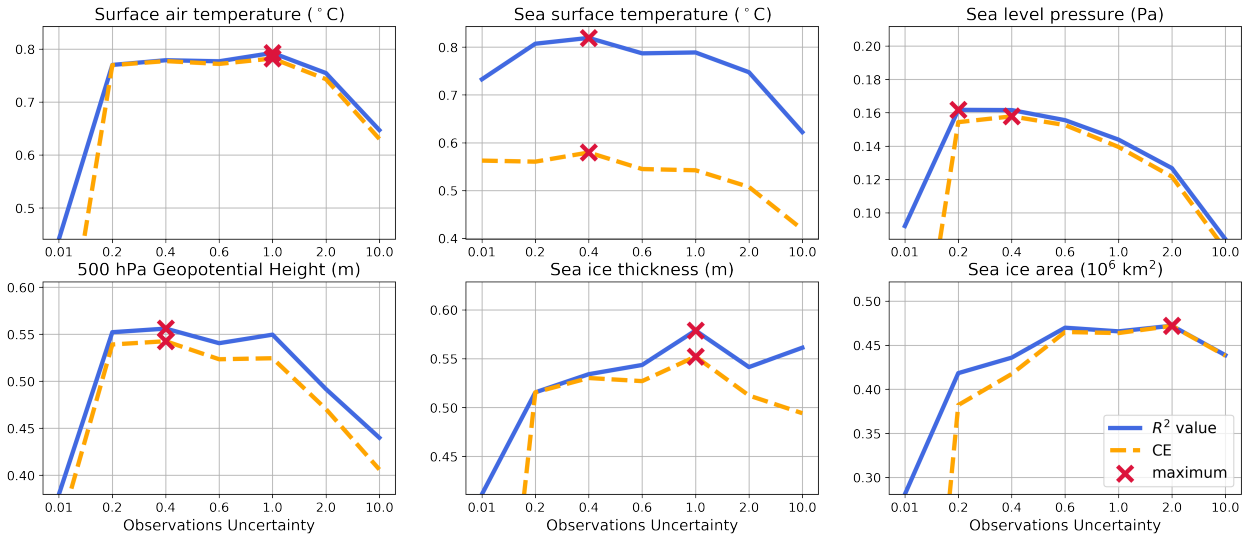

Figure C.1: Squared correlation (R^2, blue) and coefficient of efficiency (CE value, orange) calculated for northern hemisphere mean quantities for six target and reconstructed fields produced using different estimate of observation uncertainty (x-axis) in the DA scheme. The target values are the CMIP6 MPI Historical simulation and reconstructed fields are produced using an imperfect pseudo observation online data assimilation scheme between 1851–2005 CE. 100 observations are drawn randomly at each time step from HadCRUT5 observation locations and a random error drawn from a normal distribution with variance equal to the observation uncertainty is added to each observation. The results shown represent one Monte Carlo iteration for each experiment. The maximum values for both R^2 and CE values are shown as red X's for each field. Each observation uncertainty value considered was applied at all observation locations and held constant in time.

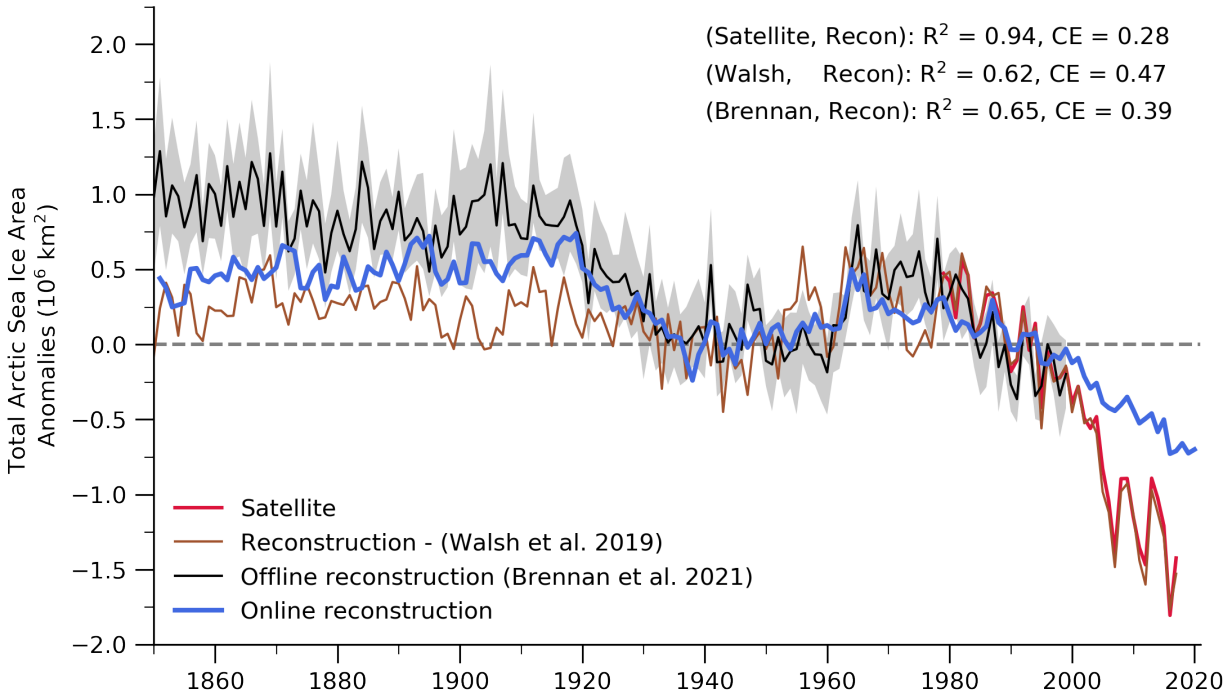

Figure C.2: A reconstructions of total Arctic sea ice area anomalies from 1851–2020 CE shown in blue produced using an online DA scheme. The total Arctic sea ice area from the Brennan and Hakim (2021) produced using offline DA, Walsh et al. (2019), and satellite observations (Fetterer et al., 2017). Annual means of all datasets are presented for comparison. Anomalies are all calculated between 1979–2005 CE.

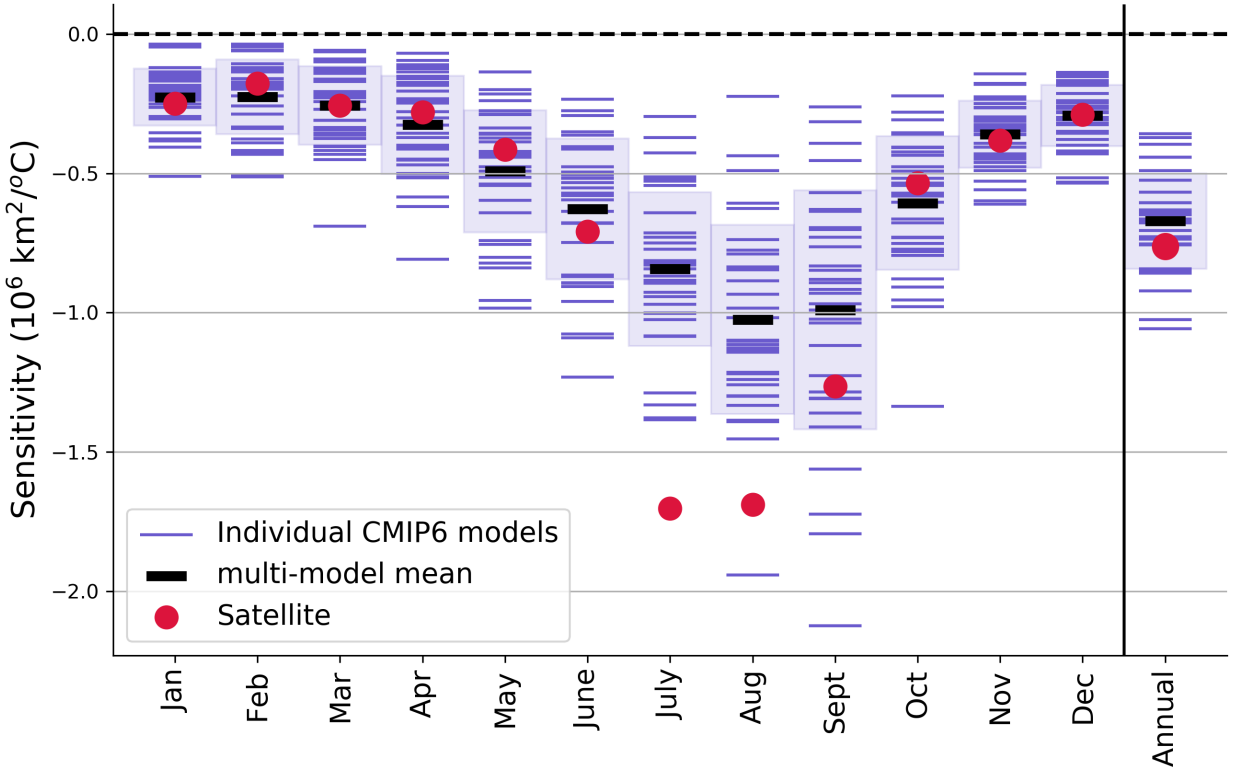

Figure C.3: Sensitivity of total Arctic sea ice extent to Arctic mean temperature (north of 60°) during the satellite period (1979–2014 CE) for both CMIP6 historical simulations and satellite observations (red). Individual CMIP6 model estimates from 31 different models are shown as purple lines and the multi-model mean is shown in black. The purple shaded regions represent one standard deviation about the multi-model mean. This figure is taken from a manuscript in preparation by Madison Smith, Robin Clancy and M. Kathleen Brennan.