Title
Statistical Reconstruction of 20th Century Antarctic Sea Ice

Author
Maierhofer, Thomas Johannes

Publication Date
2023

UNIVERSITY OF CALIFORNIA
Los Angeles

Statistical Reconstruction of 20th Century Antarctic Sea Ice

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Statistics

by

Thomas Johannes Maierhofer

2023

ABSTRACT OF THE DISSERTATION

Statistical Reconstruction of 20th Century Antarctic Sea Ice

by

Thomas Johannes Maierhofer
Doctor of Philosophy in Statistics
University of California, Los Angeles, 2023
Professor Mark Stephen Handcock, Chair

The short satellite-observed period from 1979 to 2023 has seen the Antarctic sea ice change dramatically. The sea ice generally increased until 2014 then precipitously decreased from 2014 to 2017. Record lows were then observed in February 2022 and February 2023. To evaluate these recent changes in the context of anthropogenic climate change requires information on Antarctic sea ice variability over the full 20th century. However, only temporally and spatially sparse data are available before 1979, creating a need for statistical reconstructions. We create a stochastic ensemble reconstruction of monthly Antarctic sea ice extent from 1900-1979 using a novel Bayesian spatio-temporal model. This model produces a set of 2500 plausible reconstructions of sea ice extent by sector. These reconstructions improve on prior approaches with realistic month-to-month changes and interdecadal trends as well as plausible interactions between the sectors. These unique features allow the direct computation of extreme event probabilities for the pre-satellite period of the 20th century. For example, we compute a 0.44% probability of reconstructing a decline in total Antarctic sea ice extent as extreme or more extreme than the 2014-2017 decline. We compute a 16% probability of observing a sea ice minimum as low or lower than the February 2022 minimum in total sea ice extent of 2.22 mio. km^2, and a probability of 4% for the February 2023 minimum of 2.04 mio. km^2.

We also propose a novel approach to modeling the Antarctic sea ice edge using a functional regression model. The sea ice edge on a given day is treated as a continuous set of points circling the South Pole and the latitudes of these points are modeled as a function of their longitude. This enables the estimation and visualization of a circumpolar annual cycle and interannual development of the sea ice, providing novel insight into the regionality of sea ice variability.

We develop statistical techniques for reconstructing 1966-1978 Antarctic sea ice extent using early satellite images in the visible and infrared spectrum. These images have not previously been analyzed due to severe missingness and cloud cover occluding the sea ice. Combining extensive preprocessing with modern statistical methods, we produce monthly sea ice concentration maps from which we compute historic sea ice extent.

The dissertation of Thomas Johannes Maierhofer is approved.

Karen McKinnon

Marilyn N. Raphael

Paik Frederic R. Schoenberg

Ying Nian Wu

Mark Stephen Handcock, Committee Chair

University of California, Los Angeles

2023

To my future students …

do as I say, not as I do.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank the members of my committee for all their help and support. In particular I wish to thank Professor Raphael who has been an incredible mentor, always made time to explain the science behind the data and kept me and my projects on track.

Parts of this work were supported by the National Science Foundation (NSF) under the Office of Polar Programs under grant NSF-OPP-1745089. Parts of Chapter 2 have previously been submitted as part of my written qualifying exam for advancing to candidacy. Parts of Section 3.2 have previously been submitted as a final project for the class BIOSTAT M234 in Winter 2020 with Professor Rob Weiss. Parts of Chapter 6 have previously been submitted as a class project for GEOG 214 in Winter 2020 with Professor Yongwei Sheng.

VITA

2019, 2023	Lecturer, Department of Statistics, UCLA. Taught Introduction to Statistical Reasoning two times.
2018–2023	Teaching Assistant, Department of Statistics, UCLA. TAed Python and other Technologies for Data Analysis, Introduction to Statistical Reasoning, Introduction to Mathematical Statistics, Introduction to Computational Statistics with R, High-dimensional Statistics.
2017–2018	Senior Statistician, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
2017	M.S. (Statistics), Thesis: Classification of Functional Data, University of Munich.
2016–2017	Teaching Assistant, Department of Statistics, University of Munich. Introduction to Probability Theory and Inductive Statistics, Practical course for Statisticians, Practical course for Beginners.
2014–2015	Study Abroad as Erasmus Scholar, Mathematics, Operations Research and Statistics, Cardiff University.
2015	Teaching Assistant, Cardiff University, Preliminary Mathematics and Statistics
2013–2017	Statistical Consultant, University of Munich.
2016	B.S. (Statistics), Thesis: Estimation of Entropies, University of Munich.

PUBLICATIONS

B. Summer, D. Lill, K. Remmel, A. Schraml, C. Schopf, I. J. Banke, H. Kuechenhoff, T. Maierhofer, S. Endres, and P. Thomas. “An interleukin-1 polymorphism additionally intensified by atopy as prognostic factor for aseptic non-mechanical complications in metal knee and hip arthroplasty.” *Frontiers in Immunology*, **13**, 2022

Christian Boekels, Thomas Maierhofer, and Fabian Schroeter, “Method for Checking Plug Connections”, *US Patent US20200083644A1*, Bayerische Motoren Werke AG, 2020

Thomas Maierhofer, Markus Iseli, and Eric Savitsky. “Functional Regression for Analyzing Human Ratings of Ultrasound Probe Alignment.” Poster presented at: *CRESSTCON18*; Los Angeles, CA, 2018

Anna Lee and Thomas Maierhofer. “Latent Ability Estimation in Serious Games Through Gameplay Data Visualization.” Poster presented at: *CRESSTCON18*; Los Angeles, CA, 2018

T. Maierhofer, F. Pfisterer, A. Bender, H. Kuchenhoff, O. Moerer, H. Burchardi, and W. Hartl (2017). “Cost analysis as a tool for assessing the efficacy of intensive care units.” *Medizinische Klinik-Intensivmedizin und Notfallmedizin*, **1-7**, 2017

CHAPTER 1

Introduction

Antarctic sea ice is an important part of the global climate system. As such, it is an important indicator for climate change, but has only been observed during the short satellite observed period starting with the launch of NIMBUS-7 in October 1978. Due to its vastness and remoteness it is basically unobserved before then. Every year the sea ice in the Southern Ocean grows from under 5×10^6 km^2 in February to more than 17×10^6 km^2 in September (Par19), a difference greater than the entire land area of Canada which is about 10×10^6 km^2. This dramatic annual cycle impacts all aspects of the Antarctic climate and ecosystem. In contrast, the interannual change in Antarctic sea ice extent is on a much smaller scale, but nonetheless significant for the global climate. Over the relatively short satellite-observed period from 1979 to 2014 the sea ice then increased (as opposed to a major decrease in Arctic sea ice) and then dramatically decreased between 2014 and 2017, reaching a record low in February 2022 and an even smaller low in February 2023. To fully understand these observed changes and extreme events, it is essential to extend the record of observed sea ice comprises only 42 years of data. Anthropogenic climate change started to have a major impact on the global climate with the industrial revolution at the turn of the 20th century (FGA13), but we have little to no data on Antarctic sea ice before 1979. In this dissertation I explore statistical methods to reconstruct Antarctic sea ice from 1900 to 1978. My results provide a better understanding of recent developments in Antarctic sea ice by analyzing them in the wider context of 20th century Antarctic sea ice. For this purpose, I develop novel statistical methods and apply state-of-the-art statistical methods to understand, model, and interpret changes in Antarctic sea ice.

The National Snow and Ice Data Center (NSIDC) provides satellite-based Antarctic sea ice observations as daily images of Antarctic sea ice concentration on a 25 km × 25 km grid. This data is analyzed on three levels of aggregation, the sea ice concentration (0-100%) per grid cell, the sea ice edge which is computed as the 15% level curve of the sea ice concentration, and the sea ice extent which is computed as the area within the sea ice edge. Developing statistical models for Antarctic sea ice on all three levels of aggregation is an important task. In this dissertation, I introduce a statistical model for Antarctic sea ice extent reconstruction, develop a statistical model for the Antarctic sea ice edge, and create a statistical model to extract sea ice concentration from early satellite images.

The remainder of this dissertation is structured as follows:

Chapter 2 is a literature review on Antarctic sea ice estimation. It summarizes methods of estimating Antarctic sea ice concentration from 1979-today using satellite images, discusses observed trends in the satellite observed sea ice, and summarizes available data on Antarctic before 1979. It further reviews current statistical models for Antarctic sea ice on its three levels of aggregation - sea ice concentration, sea ice edge, and sea ice extent.

Chapter 3 assesses existing approaches to Antarctic sea ice extent reconstruction. It focuses on existing approaches using ice core proxies, their weaknesses, and introduces a Bayesian time series model to improve upon them. This chapter also discusses the weather station based approach used by (FSR22). In particular, the variable selection problem and model selection problem are discussed and how these concerns impact the resulting sea ice reconstructions.

Chapter 4 introduces a Bayesian time series model for 20th century Antarctic sea ice extent reconstruction using temperature and sea level pressure measurements from weather stations in the Southern hemisphere. We reconstruct monthly Antarctic sea ice extent by sector, creating an ensemble reconstruction of 2500 reconstructions where each reconstruction contains plausible month-to-months progressions and interactions between the sectors, as well as interannual and decadal trends. Using this ensemble reconstruction, we analyze extreme events like the dramatic decline in sea ice extent from 2014-2017 as well as the record low in sea ice extent in February 2022. Our reconstruction gives probabilistic answers to the question how likely it is that events of similar magnitude have occurred in the 20th century before the beginning of the satellite record.

Chapter 5 applies a functional regression model to modeling the Antarctic sea ice edge, a framework that has to date not been used in this context. Modeling the Antarctic sea ice edge as opposed to extent gives a detailed insight into the regional distribution of the sea ice. The sea ice extent can easily be computed for any sector by integrating over the sea ice edge and subtracting the area of the Antarctic continent. The latitude (distance to South Pole) of the Antarctic sea ice edge is modeled as a function of longitude (angle around South Pole) simultaneously across all longitudes. Using functional regression to model the Antarctic sea ice edge results in interpretable coefficient maps that give novel insight into the regionality of its annual cycle and interannual development. We perform a graphical evaluation of our model and compute several internal performance measures. We compute our predictions' average distance to the true sea ice edge, the error in sea ice extent our modeled sea ice edge implies and the coverage of 95% prediction intervals.

Chapter 6 analyzes early satellite images taken between 1966 and 1978 that were digitized and made publicly available by (Cam19). These early satellites took images in the visible and infrared spectrum. This means that observations of sea ice are occluded by the cloud cover, which is 60-90% depending on latitude, and looks virtually identical to the sea ice underneath in these spectra. Additionally, during the long Antarctic winter, these satellites could not record any data due to the darkness. These two sources of missingness as well as poor image quality are the main reasons why these images have not yet been evaluated. This dissertation offers a first attempt at extracting information from these images by digitally cleaning and preprocessing the images, estimating monthly sea ice concentration from the resulting images, identifying the sea ice outline, computing the sea ice extent, and interpolating missing months. We use a Bayesian framework for this analysis that enables us to use the modern satellite record as prior information, informing our Bayesian latent discriminant analysis to distinguish between sea ice and open water.

This dissertation concludes with a summary and discussion.

CHAPTER 2

Literature Review of Antarctic Sea Ice Estimation

2.1 Introduction

This is a literature review of Antarctic sea ice estimation. Antarctic sea ice is an important topic of research as it is a driver and indicator of the global climate. This chapter is structured as follows: Section 2.2 introduces the most important algorithms for detecting sea ice using remote sensing, summarizes trends that have been observed since the start of the modern satellite era in 1978, and reviews reconstruction methods for Antarctic sea ice before 1978. Section 2.3 proposes three aggregation levels for Antarctic sea ice, the sea ice concentration (0-100%) per grid cell, the sea ice edge which is computed as the 15% level curve of the sea ice concentration, and the sea ice extent which is computed as the area within the sea ice edge. Statistical models for all three aggregation levels are reviewed. Section 2.4 discusses methods for detecting clouds and creating cloud-free observations and their applicability around the South Pole.

2.2 Antarctic Sea Ice

Estimation of the sea ice concentration on a continental scale in Antarctica is only feasible using remote sensing (i.e., satellites) due to its vastness and remoteness. As a background, this section starts with an introduction to electromagnetic radiation and its application in remote sensing. The modern-day satellite record starts in 1979 with the launch of NIMBUS-7 in October 1978. Section 2.2.1 discusses the most relevant algorithms for the estimation of sea ice concentration using remote sensing. Section 2.2.2 reviews important papers analyzing the development of the Antarctic sea ice extent since the beginning of the modern satellite era in 1979. Section 2.2.3 gives an overview of the literature reconstructing Antarctic sea ice before 1979.

Electromagnetic Radiation in Remote Sensing

We begin by reviewing the fundamental aspects of the electromagnetic spectrum and its role in remote sensing. This provides important background knowledge to understand the differences between different generations of remote sensing instruments that were used to obtain sea ice concentrations in Antarctica. Photons travel at different wavelengths with different energy. Different wavelengths are perceived very differently in our everyday life, and are thus divided into frequency bands with distinct names. From low frequency (long wavelength, low energy) to high frequency (short wavelength, high energy), these bands are: radio waves (<300 GHz), microwaves (300MHz - 300 GHz), infrared (300 GHz - 30THz), visible light (430750 THz), ultraviolet (800PHz to 30EHz), X-rays (300PHz - 30 EHz), and gamma rays (around 300 EHz) at the high frequency (Source: Wikipedia).

The trade-off for using longwave verses shortwave radiation in remote sensing, is that shortwave radiation (i.e., visible or infrared light) creates images with a higher resolution, but longwave radiation (i.e., microwave) can penetrate clouds and observe the surface of the planet directly. Due to these limitations, modern satellites record images at wide range of spectra which can be combined in subsequent analyses.

2.2.1 Sea Ice Concentration 1978 to the Present

The modern-day satellite record starts in 1979 with the launch of NIMBUS-7 in October 1978. Since then, images that can be used to estimate the Antarctic sea ice concentration are available on a one to two day resolution. One of the most commonly used algorithm to obtain the sea ice concentration from the satellite imagery is the NASA Team algorithm (CGC84). The algorithm was originally developed for the Scanning Multichannel Microwave Radiometer (SMMR) as mounted on NIMBUS-7 to estimate the surface brightness temperature. SMMR is a multispectral (using wavelengths of 0.81, 1.4, 1.7, 2.8, and 4.6 cm) and dual-polarization (vertical, horizontal) instrument. Open water, first-year ice and multi-year ice show distinct profiles across the spectra and polarizations. The original algorithm was later adapted for Special Sensor Microwave / Imager(SSM/I) by (CCD91). SSM/I has seven channels at four frequencies. The 19-, 37-, and 85-GHz frequencies are dual polarized (horizontal and vertical) and the 22-GHz frequency has only a single vertically polarized channel. The sea ice concentration is estimated as a linear function of the polarization ratio (horizontal brightness vs. vertical brightness at a certain wavelength λ) and the gradient ratio (slope of vertically polarized brightness across different wavelengths). Ratios are useful to avoid issues in the calibration. The NASA Team sea ice algorithm is intrinsically based on an instrument that measures different channels (polarization and wavelength) and is thus not applicable or transferable to a one-channel instrument.

It was noticed that the NASA Team algorithm was affected by weather (e.g., wind-roughening of the ocean surface, cloud liquid water, rainfall) and would incorrectly estimate the presence of low ice concentrations (e.g. GCC93). This was fixed with so called weather filters, which are implemented by (GC86) and later improved by (Cav94) as a separate thresholding condition that, if met, will set the sea ice concentration to 0. The filters on the other hand are prone to underestimating the presence sea ice concentrations below 8-12%, which can be problematic for nautical traffic. The estimation of the overall ice extent is robust this bias in the weather filters, as a a 15% threshold is be applied to the sea ice concentration before computing the sea ice extent.

The most important update to the NASA Team algorithm is the Bootstrap algorithm introduced in (Com86). While the algorithm is structurally similar to the NASA Team algorithm, it improves on it particularly in areas of low sea ice concentration. It uses brightness ratios in the frequency channels of 19 Ghz (wavelength = 1.58 cm, previously not available) and 37 Ghz (wavelength = 0.81 cm) to estimate the sea ice concentration. The Bootstrap algorithm also uses a weather filter and an ocean mask, as described in (CS86).

The Cal/Val algorithm was developed in (Hol91) as a first calibration/validation algorithm of the SSM/I sensor. It uses a linear combination of the 19V and 37V channels within the ice pack, and 37V and 37H near the ice edge. It is calibrated to be very sensitive to the presence of any ice and thus tends to overestimate the sea ice concentration.

The NASA Team 2 algorithm is an enhancement of the original NASA Team algorithm developed by (MC00), with its main update being the use of the 85GHz channel. The 85H channel complements the 19- and 37-GHz channels and provides additional information for estimating the sea ice concentration, as it is less sensitive to inhomogeneities in the snow or ice cover. Table 2.1 summarizes the four main sea ice algorithms (NASA Team, Bootstrap, Cal/Val, NASA Team 2) presented above. The next generation of remote sensing instruments uses the NOAA Advanced Very High Resolution Radiometer (AVHRR), a polar-orbiting sensor with channels in the visible and infrared spectrum. (CS01) introduce an algorithm to obtain the sea ice concentration from these images. In summer, the surface temperature of both ice and water is close to freezing and they are best distinguished by their surface albedo (AVHRR Channel 2, 420-270THz). A linear interpolation between tie points with 100% ice and 100% water is used to obtain the sea ice concentration. In winter the surface temperature of the ice is a lot lower than for open water and they can be distinguished using thermal infrared (AVHRR Channel 4, 29-27THz). This allows a simple threshold of 271 Kelvin ($\approx 28^\circ$ Fahrenheit) to separate between ice and open water. These initial predictions are then aggregated to increase consistency.

2.2.2 Trends in Sea Ice Extent 1978 to the Present

A series of papers has looked into the trends of the Antarctic sea ice over the decades since modern-day satellite images are available. (GCC93) describe the annual cycle of the polar sea

Table 2.1: Algorithms for sea ice concentration, taken from (Mei05). Primary and secondary channel are electromagnetic wave frequency in GHz using horizontal (H) or vertical (V) polarization. Tie points are empirically determined brightness temperature values for the extreme surface types, i.e., 100% ice or 100% open water.

Algorithm	**Primary Channels**	**Secondary Channels**	**Tie points**
NASA Team	19V, 19H, 37V	37V, 37H (near edge)	winter, summer
Bootstrap	37V, 37H (multiyear ice)	19V, 37V (first-year ice)	winter, multiple summer
Cal/Val	19V, 37V (within pack)	37V, 37H (near edge)	winter, summer
NASA Team 2	19V, 19H, 37V, 85V, 85H	37V, 37H (near edge)	winter, summer

ice extent (and area) using its shape, amplitude, minima, and maxima. They are using the 9 year satellite record available from 1978-1987. (CPV03) use a 30 year satellite record (1972-2002) to detect decadal trends in the Arctic and Antarctic sea ice. They notice the opposing trends of Arctic and Antarctic sea ice that they attribute to an anomaly of the Antarctic sea ice in the late 1960's and early 1970's. Linear regression models are used to describe the decadal trends. (CP08) analyze a 28 year record (1979-2006) of the Antarctic sea ice. They create monthly, yearly, and seasonal averages and describe their trends and anomalies using linear models and averages. They describe contrasting trends for different sectors of the Antarctic sea ice. (PC12) describe similar results for an additional 4 years (1979-2010, 31 years total). (Sim15) describe the sea ice extent for the Arctic, the Antarctic, and globally

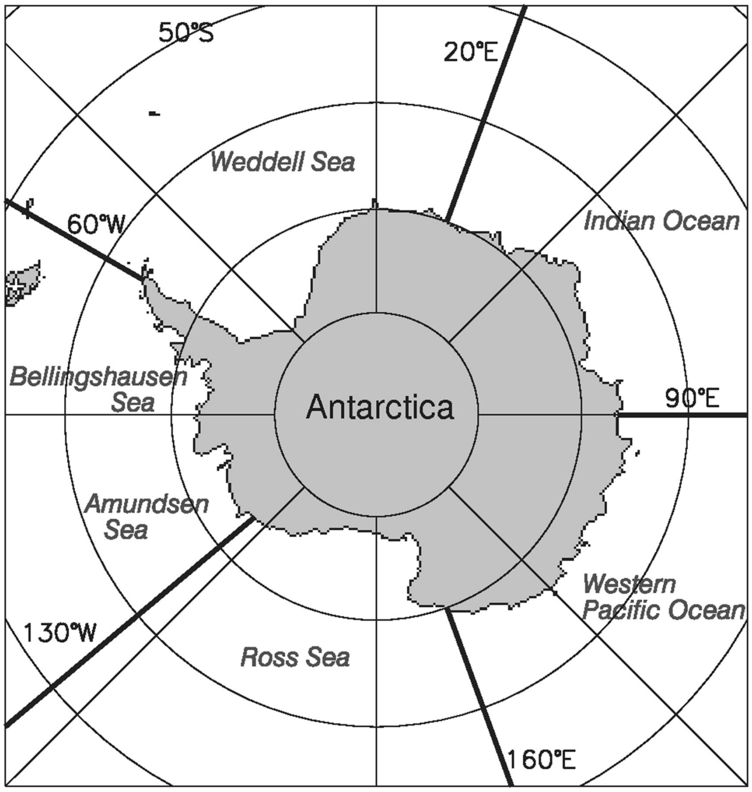


Figure 2.1: Sectors of the Antarctic Ocean (Source: Par19).

(Arctic plus Antarctic sea ice) from 1978-2013. They describe a decrease in Arctic sea ice that is roughly 3 times the size of the increase of the Antarctic sea ice. (CGS17) identify a strong negative correlation (-0.94 during the growth season, -0.86 during the melt season) of the sea ice extent with the surface ice temperature (record starting in 1981). The correlation is even larger when using a 1-month lag on the time series. The Antarctic sea ice has declined dramatically from a record high in 2014 to a record low in 2017 (Par19). (HR20) explain the low average sea ice extent in 2016 by an unusually early decline in sea ice.

The Antarctic Ocean is split into 5 sectors, the Weddell Sea (60° W - 20° E), Indian Ocean (20° E - 90° E), Western Pacific Ocean (90° E - 160° E), Ross Sea (160° E - 130° W), and Bellingshausen & Amundsen Sea (130° W - 60° W), see Figure 2.1. The development of the sea ice concentration within these sectors over the course of the satellite record is analyzed for example by (Par19), who find that all sectors except the Bellingshausen & Amundsen Sea have increased at a rate of 1.7% to 4.3% per decade. The Bellingshausen & Amundsen Sea has decreased at a rate of 2.0% per decade. The intraannual trend is shaped very similarly across all sectors.

2.2.3 Antarctic Sea Ice before 1979

There has been limited research into the Antarctic sea ice before 1979 due to a lack of available data. A timeline of available data sources is compiled in Figure 2.2. Available data is mostly created by in-situ observation and thus spatially and temporally extremely sparse. These observations are also chosen for convenience rather than sampled representatively. A lot more research has been conducted into the Arctic sea ice during the 20th century, due to its proximity to, and accessibility from Europe and Asia and its geopolitical importance. Many scientific tools and procedures developed for the Antarctic have been applied successfully to the Antarctic.

(ED16) compile information about the Antarctic sea ice edge for the years 1897-1917 from ship logbooks. They identify 191 logbook entries of an observed ice edge.

An estimate of the monthly average ice edge based on ship logbooks from before 1939 is compiled by (MH40) and updated in (Mac72). I was unable to find either of the original publications, but their estimated sea ice edge can be found for example in (AWC03).

(RPH03) provide a continuous record of the Antarctic sea ice from 1871 (publicly available from NSIDC `https://nsidc.org/data/G00799/versions/1`). A German monthly sea ice extent climatology for 1929-1939 (Bud50). This is based on maps created by the German Antarctic expedition led by Alfred Ritscher in 1938/1939, see Wikipedia. is cited to estimate the Antarctic sea ice from 1871-1939. The period from 1939-1947 is listed as missing and interpolated. For the years between 1947 and 1962, a Russian publication (Tol66) is cited. This is based on research conducted by the State Oceanographic Institute GUGMS of the

Soviet Union, starting in 1947 with the Slava whaling flotilla. The Antarctic sea ice record for the time period from 1962-1973 is reported as missing and interpolated.

The only available records for this time period are Nimbus I, II, and III images. (MGC13) manually detect ice edges in imagery from Nimbus I (operation 30 August to 19 September 1964) and create an estimate for the sea ice extent in September 1964. They use the idea of combining individual images into one composite image using a minimum brightness filter to remove clouds, as open water has the darkest pixel values and ice is locally more stable than clouds. This provides an initial, albeit biased, indication of the sea ice edge. For the actual extent, all images were individually examined, and the ice edge marked manually. These edges are then overlaid to create an estimate for the true sea ice edge. An "inner outline" and an "outer outline" are created to estimate the uncertainty of the procedure. The labeled data is publicly available online at NSIDC (NSIDC). (GCM14) use a similar manual procedure to obtain estimates for the sea ice extent for May-August 1966 based on Nimbus II images. Estimates for the sea ice extent in 1969 based on Nimbus III images are available on the NSIDC website as well, but I could not find a corresponding paper.

(Cam19) publish a data set of recently digitized satellite images in the visible (1966-1978) and infrared (1973-1978) spectrum. It provides direct observations of the Antarctic sea ice in the currently unobserved period. This data has not been analyzed before due to poor image quality; both instruments suffer from missing values, weather contamination, and the cloud cover. Analyzing this data set could provide a valuable extension of the modern sea ice record.

The AVCS instrument (1966-1971) as equipped on the ESSA satellites, ITOS, and NOAA 1, recorded a single channel in the visible wavelength band from 0.45 - 0.65 μm with a resolution of 2.2 km - 3.0 km and a swath width of 2300 km. The SR instrument (1970-1978) as equipped on ITOS 1 and NOAA 1-5 captured two channels, one in the visible (wavelength 0.50 - 0.70 μm, 3.2 km resolution, 2900 km swath width) and one in the infrared (wavelength 10.50 - 12.50 μm, 8km resolution, 2900 km swath width) spectrum.

(GWC74) create maps of the polar ice from December 1972 to February 1973 in monthly intervals based on data from the Electrically Scanning Microwave Radiometer (ESMR) mounted on NIMBUS 5. The brightness temperature map is converted to a sea ice concentration map by a linear interpolation algorithm between tie points. (CPV03) extend the estimation of the Antarctic sea ice extent from 1973-1977 using NIMBUS 5. The sea ice is computed in a similar fashion to (GWC74), but a 7-day moving average is applied to reduce weather effects near the ice edge.

(GGF15) model the available satellite record from after 1979 and extrapolate to the 1964 estimates of (CPV03).

Estimates based on indirect observations of the Antarctic sea ice are available for the entire 20th century. (KPM13) use temperatures measured on the Antarctic peninsula to reconstruct the monthly mean latitude of the nearby sea ice edge at 70° W for the years 1945 to 1954.

(Mar97) use whaling records to obtain information about the Antarctic sea ice edge from 1931-1987. Whales are usually caught in open water near the sea ice edge (and definitely not in the ice pack), so the southern-most latitude of all whale catches in a longitude window and date window can be used as a proxy for the sea ice edge. The latitude of the southern-most catch is estimated to be on average $0.11° - 0.14°$ north of the ice edge based on sea ice charts available of the Joint Ice Center. After correcting for this, a reasonable estimate of the true sea ice extent is obtained. The methods and results of this paper are heavily doubted by (AWC03), who argue that a consistent offset between satellite and ship observed ice edge, regional sea ice perturbations, and a lack of one-to-one correspondence between the whaling location and the ice edge can explain the developments observed by (Mar97). A transition in the main whaling locations and hunted species casts further doubt on the results. (De 09) reassert the original claims by validating the whaling-based estimate of the sea ice edge with direct observations.

(MCS95) provide a continuous record of fast ice on the Orkney Islands, a set of islands bordering the Weddell Sea close to the Antarctic peninsula, from 1903 to 1994 based on direct observation. This fast ice record has proven to correlate well with the sea ice in the Weddell Sea. Fast ice is defined as ice which forms and remains fast along the coast where it remains attached to shore. A record of start date, end date, and duration of the fast ice is provided for Laurie Island from 1903 to 1975, and Signy Island from 1947 to 1994. The authors update the record in (MCA14) to include the years up until 2008.

A series of ice cores has been used to reconstruct the Antarctic sea ice, see (AWC13) and (TAE19) for a review. Ice cores can predict the regional sea ice better than the overall Antarctic sea ice extent. Salt concentration in the ice core has proven useful for long-term sea ice reconstruction, but for reconstructing sea ice extent from the last century methanesulphonic acid (MSA), a derivative of dimethylsulphide emissions of marine algae and phytoplankton, is the most used proxy. An overview of other possible biogenic and ice core sea ice proxies is summarized in (DGG13) and (TAE19). As a sea ice proxy, the concentration of MSA is tracked in a dated ice core and compared to the known sea ice extent during the satellite era.

(MMK02) use the MSA concentration of an ice core collected from the South Pole covering the time period from 1487 until 1992. (COM03) use ice cores collected in 1995 from Law Dome (adjacent to the Western Pacific sector) to estimate the yearly average and maximum Antarctic sea ice extent since 1841. They identify large cyclical variations, with periods of about 11 years and estimate a 20% decline in SIE between 1950 and 1994. (BCC09) use MSA concentration in an ice core collected on Talos Dome in 1995 adjacent to the Ross Sea to predict Antarctic sea ice extent. Another ice core using MSA concentration collected in 2006 on land adjacent to the Ross Sea supports the mid-century decline in Antarctic sea ice (SBB14). (ATM10) use the MSA concentration in 4 ice cores collected from the Antarctic peninsula in 1988, 1991, 1997, and 2004, to estimate the overall sea ice extent and specifically the Bellingshausen sea ice extent for the 20th century. (AMA11) use the accumulation, isotopic, and chemical signals (including MSA) of an ice core collected in 2008 on Ross Island to estimate the Antarctic sea ice extent from 1967-2008. (PPM16) use an ice core collected in 2010 from the Antarctic Peninsula (66.03°S, 64.07° W) to predict the sea ice extent in the nearby Bellingshausen Sea. (TA16) use an ice core collected in 2010 in Ellsworth Land, West Antarctica (74.57°S, 86.90°W) to predict the sea ice extent in the nearby Amundsen and Ross Sea. (HCA16) reconstruct sea ice extent in different sectors based on one (aggregated) proxy per sector. They find that the correlation between the sea ice extent of the Ross Sea and the MSA levels in the ice core used by (TA16) as well as the ice core used by (SBB14) is not stable during the modern satellite era. This indicates that these ice cores should not be used to extrapolate the sea ice extent for the Ross Sea.

The Coupled Model Intercomparison Project version 5 (CMIP5) is a collection of global climate models that include gridded sea ice concentration (TSM12). The collection tends to have a large variability, but can provide useful information on large scale trends.

The Antarctic sea ice has been found to be associated a number of other climate variables that have not been used for sea ice reconstruction to date. These include ocean-driven large scale climate indicators in the Southern Hemisphere like the El Niño - Southern Oscillation (ENSO) (e.g. SMS08) or the interdecadal Pacific oscillation (IPO) (e.g. PEC16). Rich data exists for sea level pressure observations from the Southern Hemishere midlatitudes in the 20th century that have to date not been used for sea ice reconstruction. (FJG16; FCN18) provide comprehensive collections of these data.

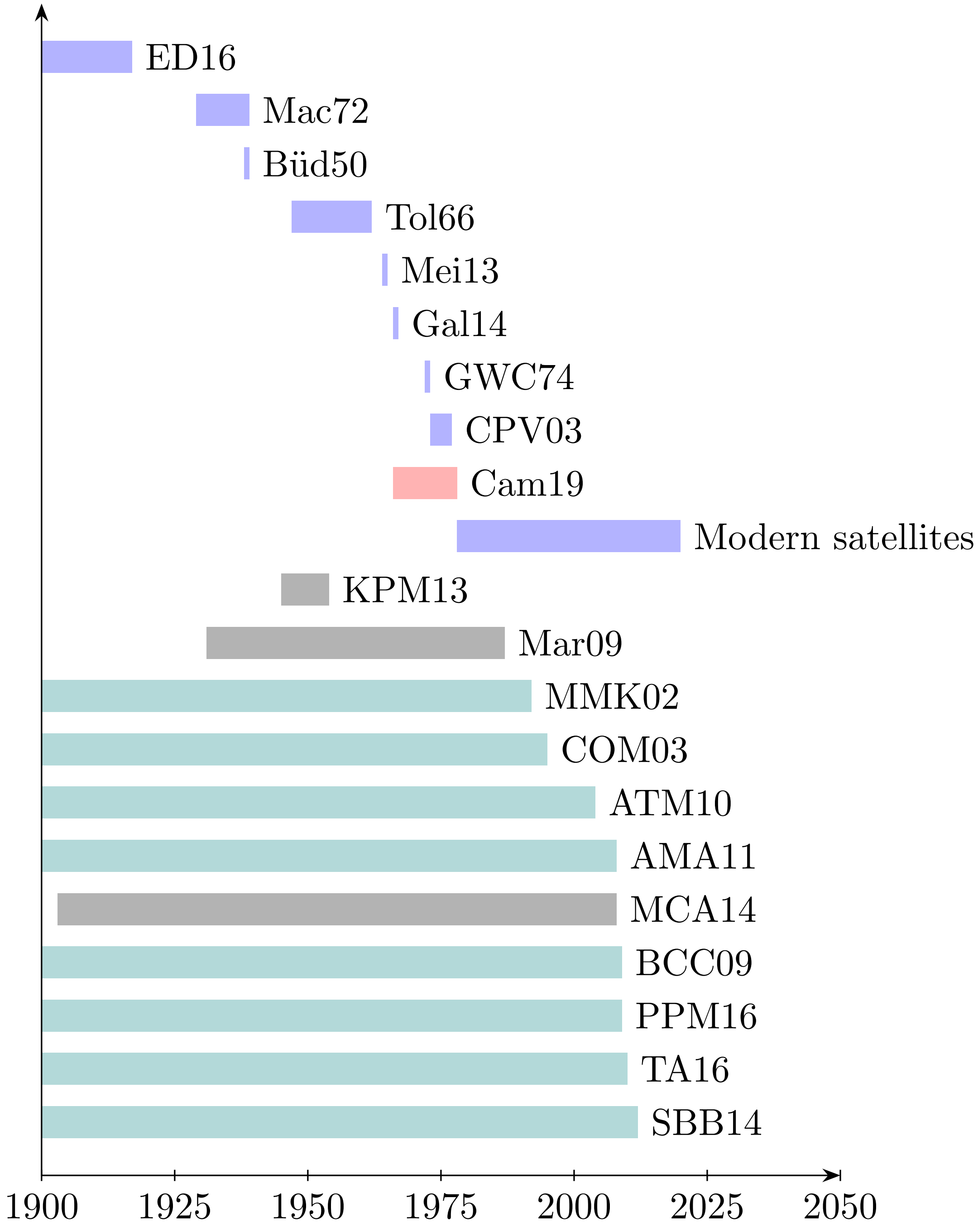


Figure 2.2: Timeline of estimates available for 20th century Antarctic sea ice extent. Blue: Estimates based on direct observation of the ice; grey: estimates based on proxies; green: estimates based on ice cores; red: raw satellite imagery.

2.3 Statistical Models for Sea Ice

The Antarctic sea ice can be modeled on three levels of aggregation, its concentration (Section 2.3.1), outline (Section 2.3.2), or extent (Section 2.3.3). The main product is the sea ice extent, but the concentration and outline map directly onto the extent.

2.3.1 Modeling Sea Ice Concentration

The sea ice concentration on a given day is a 2d map with entries between 0% (open water) and 100% (solid ice). Every entry corresponds to one pixel between 0 and 100% correspond to the percentage of ice coverage of this pixel. As consecutive days are highly correlated, introducing time as a third dimension in the model will be essential.

Ising models can be applied under the simplification that sea ice only takes the discrete values 0 or 100%. This simplification should not be too influential as only a rather narrow boundary zone between open water and the ice pack takes values between 0 and 100% ice concentration. For the subsequent sea ice edge and sea ice extent computation a binary threshold is applied to the concentration. Continuous extensions of the Ising model exists, see e.g., (BK79) and can be applied if necessary.

Another promising approach for modeling sea ice is the use of Gaussian Markov random fields making use of the 2d lattice structure implied by the spatial neighborhood of pixels (Bes74). An application in Bayesian image analysis is shown in (BGH95) who apply random fields to gamma camera imaging. Similar models also find application in the segmentation of medical ultrasound images (e.g. JT04). A pairwise difference formulation of Intrinsic Conditional Auto-Regressive (ICAR) models should be sufficient for modeling sea ice extent, as large parts of the grid are known to have ice concentrations of 0 or 100% (BK95).

(ZC19) create a hierarchical spatio-temporal model for the Arctic sea ice concentration with the purpose of modeling the pixel-by-pixel transition probabilities of ice and water between consecutive years. These transition probabilities give insight into the process of the

Arctic sea ice decline. Whether the ice concentration $z_t(s)$ at time t and location s is above or below the threshold is modeled as conditionally independent Bernoulli random variables conditional on a latent process $y_t(s)$,

$$z_t(s)|y_t(s) \sim \text{Bernoulli}\big(p_t(s)\big),$$

where $y_t(s) = g\big(p_t(s)\big)$ where the logit function is used as the link function $g(\cdot)$. The latent process is modeled as an auto-regressive Gaussian random variable that depends on location s, time t, and covariates $x_t(s)$,

$$y_t(s) = x_t(s)^T\beta_t + S_t(s)^T\eta_t + \zeta_t(s),$$

with

- $x_t(s)$ a p dimensional covariate vector at time t and location s (known),
- β_t a p dimensional vector of corresponding regression coefficients (to be estimated),
- $S_t(s)$ an r dimensional basis function evaluated at t and s (known),
- η_t a corresponding r dimensional mean-zero Gaussian random vector (to be estimated),
- and $\zeta_t(s)$ a univariate Gaussian random variable with mean 0 (known) and fixed variance (to be estimated).

The term $x_t(s)^T\beta_t$ captures effects of known covariates, the term $S_t(s)^T\eta_t$ models temporal and spatial correlation, and $\zeta_t(s)$ models local variation. Spatially sparse bisquare basis functions with two different resolutions are used as basis vectors $S_t(s)$ to allow small scale and large scale correlation while keeping the computational burden manageable.

2.3.2 Modeling Sea Ice Outlines

The sea ice outline is the first aggregation step of the sea ice concentration. Given the sea ice concentration, a 15% threshold is applied to separate the sea ice from open water. This 15% level curve on the sea ice concentration is used as the outline of the sea ice in a subsequent estimation of the sea ice extent. It is also possible to model the sea ice outline directly without estimating the sea ice concentration first.

The ice / water edge is identified manually by (MGC13) and composited across images to estimate the true outline. Enough points were identified on the ice / water edge across multiple days to create one continuous outline. The edge detection can be automatized using edge detection techniques from the computer vision literature, e.g., canny edge detection (DG01).

(DRB19) propose a probabilistic model for the edge of the Arctic sea ice outline using a Bayesian general contour model. The Arctic is segmented into 7 different areas that are modeled separately. The contour is modeled as an ordered set of points, that are referenced by their distance to a fix point (i.e., a point in the center of a cyclical ice edge) or a fixed line that is roughly parallel to the sea ice edge. This model does not allow the outline to curve backwards, but the contortion this introduces in the modeled sea ice outline is reduced by selecting the 7 areas and respective fixed reference line (or point) accordingly.

Methodology from shape analysis can be used to create statistical models for the ice edge (e.g. Pav78; ZL04). For estimating the sea ice outline, a generative model is required, i.e., a model that is can create actual outlines from estimated parameters. A centroid profile, i.e., representing the sea ice outline as a function of the distance of every point on its outline to its centroid, is a possible model for the outline of an object (e.g. Dav04). Every point on the outline is referenced by its angle and distance to the centroid. For the Antarctic sea ice, the South Pole is an obvious choice as the centroid, and the latitude (angle) of every point on the outline is given as a function of its longitude (distance). The problem with the outline of the Antarctic sea ice is, that the outline is curving backwards, i.e., the longitudes do not map onto a unique point on the sea ice outline. This cannot be modeled using a centroid profile.

Instead, the latitude and longitude of every point on the outline of the model can be modeled as a function of its boundary distance. The boundary distance of a point on the outline is the cyclical distance on the outline to an arbitrary but fixed reference point (e.g. Dav04). This method suffers from instability when the length of the outline changes dramatically between images, as is the case when polynyas connect to the open sea.

2.3.3 Modeling Sea Ice Extent

Given an outline of the sea ice, its extent can be computed by summing over the area of all pixels within. The resulting sea ice extent is a univariate time series. Depending on the data source, a daily, monthly, or yearly sea ice extent is reported. The annual cycle of the Antarctic sea ice and its variability are depicted in Figure 2.3 (top). The development of the Antarctic sea ice is shown in the bottom of Figure 2.3 Most papers in the literature use daily, monthly, and yearly averages to report the cycle of the Antarctic sea ice, and linear regression models to describe its trends (e.g. GCC93; CPV03; CP08; PC12; Sim15; Par19). (HR20) propose for daily sea ice extent that explicitly allows variations in amplitude and phase of the annual ice cycle. This is a more sophisticated representation of the amplitude and phase, as simple daily averages across multiple years ignore the variation in amplitude and phase.

2.4 Modeling Clouds

Clouds cover up to 90% of the study area, see Figure 2.4 (Lac10). The ice edge moves between 80° and 55° degrees south, where the cloud cover tends to be consistently above 50%. Clouds need to be explicitly handled if the instrument is sensitive to them. In the visible spectrum, clouds occlude the underlying sea ice and appear as a similar brightness. For the purpose of analyzing sea ice, clouds can either be removed or explicitly modeled as a separate latent layer. Creating cloud-free observations simplifies all subsequent analysis dramatically. (LTL13) remove clouds in modern satellite images in a two-step procedure. First, a binary cloud mask is created that identifies all clouds. Then, all pixels that were marked as clouds are replaced with temporally close observations of the same pixel that were not flagged as clouds. The resulting images are estimates of “cloud free” observations. (HR05) create a 6 step sequence where pixels identified as clouds are replaced with increasingly sophisticated algorithms that use pixels in temporal and local proximity. Running minima and medians are used in (GCD13) to identify and remove clouds.

The cloud cover in satellite images can be modeled similarly to the sea ice concentration, see Section 2.3.1. The same model classes can be applied to clouds as well, as they can be modeled as a binary mask or a continuous cloud fraction (0-100%). This is a promising area of future research.

Identifying the sea ice underneath moving clouds can also be viewed as a foreground/-background separation problem. This problem is described in the computer vision literature, e.g., (SG99) who use a Gaussian mixture to model each pixel in a stream of images. Based on the mean and variance of each pixel, slowly adapting Gaussian distributions for the background are estimated. Pixels are classified as foreground, if they do not fit the slowly adapting background with sufficiently high probability. This can theoretically be used to separate clouds (fast moving, foreground) from the sea ice (slow moving, background). In practice, ice and clouds are so similar in brightness, so that only open water and clouds can be distinguished. Another problem is that most areas (i.e., most pixels) are cloudy more than 50% of the time, especially around the ice edge. This will make the clouds appear as the pixel’s background and the open water as its foreground.

2.5 Discussion

The estimation of Antarctic sea ice is of growing interest to the scientific community. The estimation of Antarctic sea ice concentration based on modern satellite images has been studied with satisfactory results; the Antarctic sea ice concentration, edge, and extent, are considered to be known for every day since 1979.

A range of data sources have been used to estimate the Antarctic sea ice before 1979. Direct observations of the Antarctic sea ice in this time were mainly convenience samples and are temporally as well as spatially extremely sparse. (Cam19) have recently provided a data set of satellite images from the years 1966-1978. To date, these images have not been used to reconstruct the Antarctic sea ice extent of this time period, due to poor image quality and considerable occlusion of the sea ice by clouds. Analyzing this data set could extend the 40 year modern satellite record by an additional 12 years (24%).

A number of different proxies have been proposed to estimate the Antarctic sea ice extent locally, for example ice cores and temperature measurements. These proxy models are based on simple linear regressions, in the sense that the target variable is the one-dimensional sea ice extent in a certain area at a certain time, and the predictor variable is a spatially close one-dimensional proxy variable at the same time. A high-yield improvement would be the application of statistical models that reconstruct the sea ice extent in multiple areas simultaneously using multiple proxy variables as predictors while making use of the temporal and spatial correlation of the Antarctic sea ice.

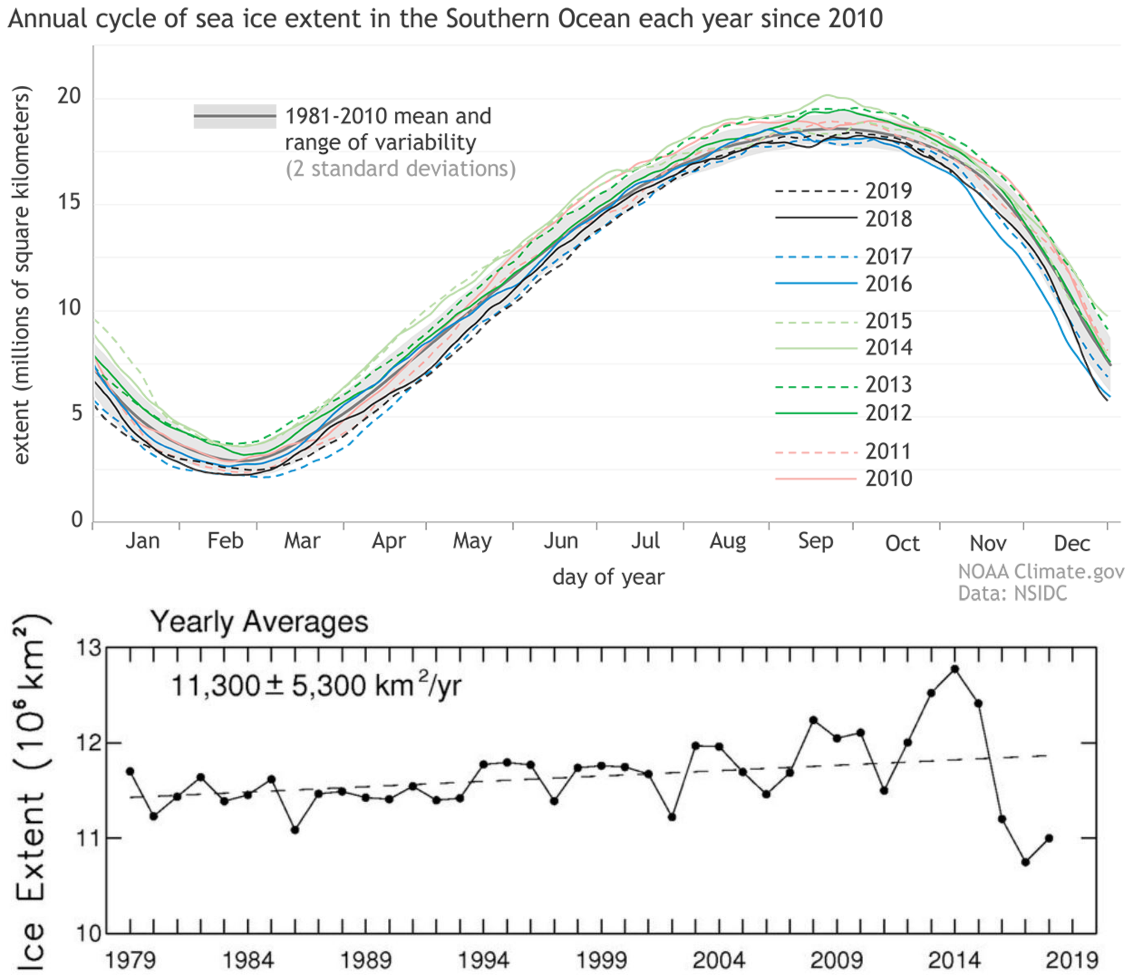


Figure 2.3: Top: Annual cycle of the Antarctic sea ice (Source: nsidc.org). Bottom: Development of the yearly average Antarctic sea ice extent since the beginning of the satellite record with linear trend line (Source: Par19).

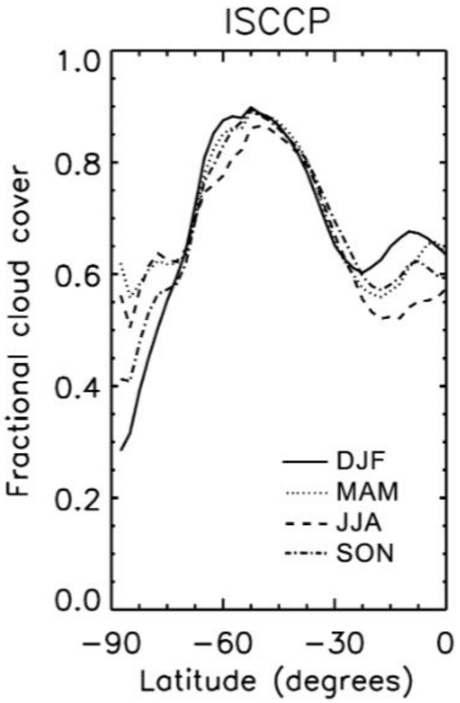


Figure 2.4: Average cloud fraction per latitude and season in the Southern hemisphere (South Pole = −90° South, equator = 0° South) as reported by the International Satellite Cloud Climatology Project (ISCCP) (Image source: Lac10).

CHAPTER 3

Assessing Existing Antarctic Sea Ice Extent Reconstructions

3.1 Introduction

Antarctic sea ice reconstruction is a key component of understanding the effects of climate change. Due the relative brevity of the observed sea ice record in comparison to the anthropogenic influence on the global climate, Antarctic sea ice reconstruction is an important task. A growing body of research is focusing on reconstructing pre-satellite Antarctic sea ice extent, including (Mar97; MMK02; COM03; CG07; De 09; BCC09; ATM10; AMA11; SBB14; PPM16; TA16; HCA16; TAE19; YXL21; DGR21; FSR22). Most of these reconstructions focus on using ice core data or weather data as predictor variables for (local) sea ice extent. These reconstructions concur that overall, there was more sea ice in the Southern hemisphere at the beginning of the 20th century but different reconstructions vary on the exact scale of this sea ice decline. We focus on the Ferrigno ice core analyzed by (TA16) as a typical example for an ice core based reconstruction in Section 3.2, and (FSR22) as a typical weather station based reconstruction in Section 3.3.

3.2 Ice Core Based Reconstruction

A series of ice cores (MMK02; COM03; BCC09; ATM10; AMA11; SBB14; PPM16; HCA16; TA16) have been used for local sea ice reconstruction, see e.g., (AWC13) and (TAE19) for a summary, but these have not been used to compile total and sectoral Antarctic sea ice extent reconstructions, and it is not straight-forward how to do so. Ice cores provide useful local reconstructions but tend to be geographically restricted in their predictive range. Another drawback of these papers is that only an average sea ice extent is reconstructed with a lack of rigorous uncertainty estimation, lack of measures of external and internal validation, and disregard for the temporal autocorrelation structure of Antarctic sea ice.

A completely different modeling approach for ice core based sea ice reconstruction is chosen by (DGR21) who reconstruct regional Antarctic sea ice on an annual resolution based on ice cores and tree ring widths using a Bayesian framework. They use the Bayesian framework to incorporate multiple simulated runs from a climate simulation model (isotope-enabled Community Earth System Model version 1, iCESM1; BSB19) as a prior distribution for Antarctic sea ice during the reconstruction period. Using a climate model trained on satellite observations to create a prior distribution means that information from the data (the observed satellite record) is used to create the prior distribution, which is problematic as the prior distribution should represent the state of knowledge before observing the data. Their approach preserves spatial autocorrelation in sea ice, but neglects temporal autocorrelation. This means that their individual reconstructions are not (and should not be) reported, but only the reconstruction average with (slightly overconfident) uncertainty bands.

Even with all these shortcomings in modeling approaches ice cores have been successfully used for sea ice reconstruction. Particularly, methanesulphonic acid (MSA), a derivative of dimethylsulphide emissions of marine algae and phytoplankton, has been linked to local sea ice extent. MSA flux, the MSA concentration adjusted for the atmospheric concentration, has proven to be a better predictor than unadjusted MSA concentration and is therefore used as "MSA" in this chapter. We re-analyze the 136m Ferrigno ice core that was collected in 2010 in Ellsworth Land, West Antarctica (74.57°S, 86.90°W, 1354m above sea level) and dated to within ± 3 months for the years 1810-2010. This ice core that has been used by (TA16) to predict the sea ice extent in the nearby Amundsen and Ross Sea at 146°W. The

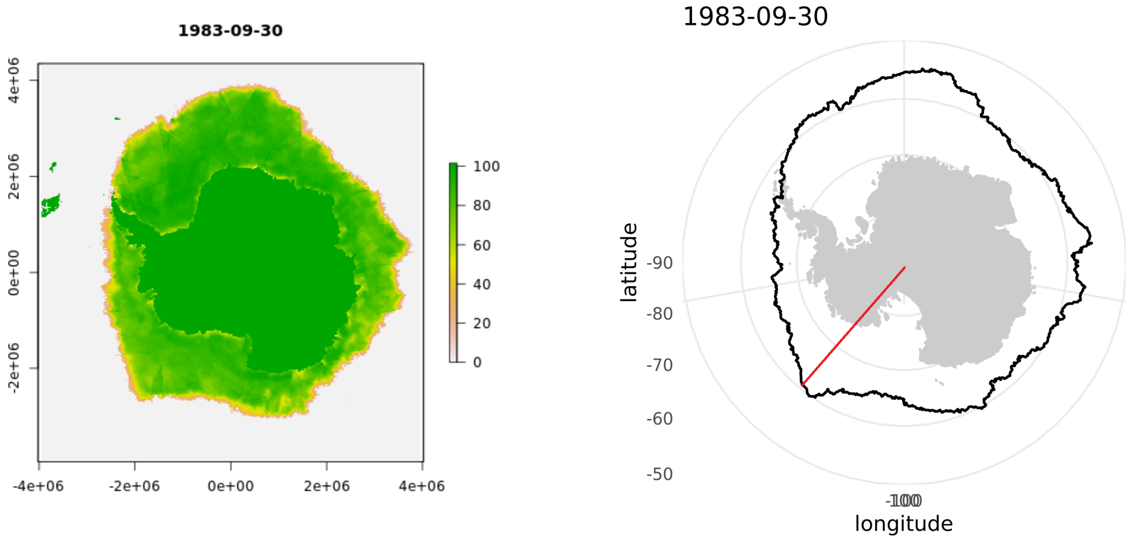


Figure 3.1: Sea ice concentration (left) and edge (right) on September 30th, 1983. The sea ice extent at 146°W is the latitude of the intersection between the 146°W longitude line (drawn in red) and the sea ice outline. For reference, the South Pole is 90°S $= -90°$ latitude, Cape Horn in South America at 56°S $= -56°$ latitude.

local sea ice extent at a certain longitude, here 146°W, can be quantified as the latitude of the sea ice edge at this longitude which is the distance of the ice edge to the South Pole, see Figure 3.1. This simplifies all modeling dramatically, as it simplifies an area of a sector or a complex outline to one scalar measurement.

3.2.1 Concerns with Existing Ice Core Based Reconstruction Models

We showcase the advantage of using Bayesian models for analyzing ice cores and discuss theoretical as well as practical improvements over the currently used methodology and will introduce a new state of the art for the analysis of ice cores. In the literature, the sea ice extent SIE_t in year t is estimated as a linear function of the MSA flux in year t denoted by MSA_t,

$$\hat{\mathrm{SIE}}_t = \beta_0 + \beta_1 \mathrm{MSA}_t + \varepsilon_t, \tag{3.1}$$

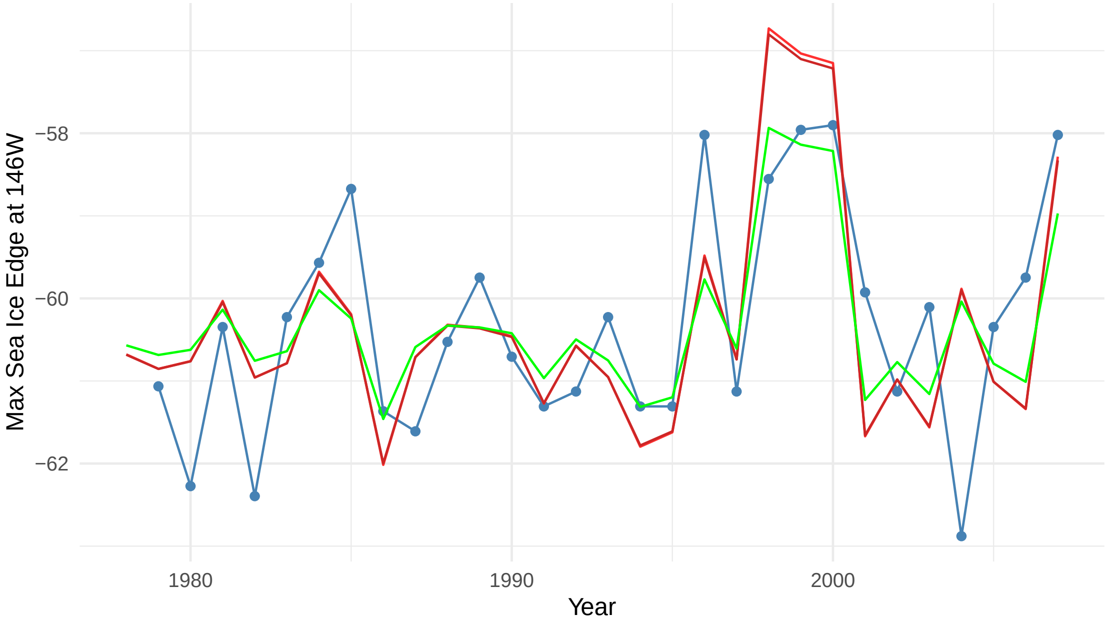


Figure 3.2: Linear models used to predict annual maximal sea ice extent at 146°W (blue) based on average annual MSA concentration. The OLS regression is plotted in green, the original MAR prediction from (TA16) in light red, the recreated MAR prediction in dark red. Higher degrees of latitude imply more sea ice, more negative latitudes imply less sea ice.

see Figure 3.2. While some authors (e.g. COM03) use ordinary least squares regression (OLS), i.e.,

$$\beta_0, \beta_1 = \underset{\beta_0,\beta_1}{\arg\min} \sum_t (\beta_0 + \beta_1 \mathrm{MSA}_t - \mathrm{SIE}_t)^2,$$

more and more authors are using major axis regression (MAR) (a.k.a. geometric mean regression), i.e.,

$$\beta_0, \beta_1 = \underset{\beta_0,\beta_1}{\arg\min} \sum_t (\beta_0 + \beta_1 \mathrm{MSA}_t - \mathrm{SIE}_t)^2/(1+\beta_1^2).$$

The geometric interpretation is, that OLS minimizes the squared errors of the regression line to all points in y-direction, whereas MAR minimizes the squared errors perpendicular between each point and the regression line. Proponents of MAR regression claim that it is more robust towards errors in the independent variable (here MSA measurements in the ice core), as it minimizes the error in both directions. The use of MAR regression is problematic, as it always estimates a larger slope than OLS. This is particularly problematic in this case where correlation between sea ice and ice core proxies should be, if anything, shrunk towards 0. The proxies need to convince us that they actually contain information about the sea ice and we should make sure to not inflate their association with the local sea ice extent.

3.2.2 Bayesian Time Series Modeling for Ice Core Based Reconstructions

Bayesian time series models have been used successfully in a wide range of applications (Har90). The proposed model is

$$\hat{\mathrm{SIE}}_t = \beta_0 + \beta_1 \mathrm{MSA}_t + \beta_2 \varepsilon_{t-1} + \varepsilon_t, \tag{3.2}$$

where $\hat{\mathrm{SIE}}_t$ is the estimated latitude of the sea ice edge at 146°W at time t, MSA_t is the standardized MSA Flux at time t, ε_{t-1} is the error term at time $t-1$, and ε_t is the error term at time t. We use prior distributions

$$\begin{aligned}
&\beta_0 \sim N\left(\overline{\mathrm{SIE}} = -60.33, \frac{\mathrm{Var(SIE)}}{N_{\mathrm{eff}}(\mathrm{SIE})} = \frac{1.87}{20.70} = 0.3^2\right), \\
&\beta_1 \sim N\left(0, 1/0.1^2 = 10^2\right), \\
&\beta_2 \sim N\left(0, 0.5^2\right), \\
&\varepsilon_t \sim N(0, 1/0.1^2 = 10^2)\forall t,
\end{aligned}$$

where $\overline{\mathrm{SIE}}$ is the average latitude of the sea ice edge at 146°W, Var(SIE) is the variance in the latitude of the sea ice edge at 146°W, and $N_{\mathrm{eff}}(\mathrm{SIE})$ is the effective sample size we want to use for the strength of our prior distribution. We use weak priors on the regression parameters β_0, β_1 and error terms ε that are centered around zero.

This model is implemented using the R package `bsts` (Sco19). The estimated posterior distributions for the parameters are summarized in Table 3.1. The model can estimate the average sea ice edge precisely, see small standard deviation (sd) or narrow CI of β_0. MSA flux has a strong positive effect on the ice edge, see the positive value of β_1. For an increase of MSA Flux by one standard deviation we expect the ice edge at 146°W to be 0.74°further South. The lag 1 autocorrelation is found to be small $\beta_2 \approx 0$. The distribution of the error terms ε is symmetric and approximately centered around 0. The predictions for the 20th century sea ice extent based on the model proposed in Equation (3.2) are shown in Figure 3.3. The 95% prediction interval is almost instantly at its full width due to the small autocorrelation of the error terms. Figure 3.4 compares the different reconstructions for the maximal annual sea ice extent at 146°W.

Table 3.1: Mean, standard deviation, 2.5% quantile, 97.5% quantile, and probability of coefficient being greater than 0, of the posterior distribution of the parameters for the BSTS model specified in Equation (3.2). The effect of $\beta_1 = 0.74$ means that an increase of 1sd in MSA flux is associated with a shift of the sea ice edge by 1 degree further North.

	mean	sd	2.50%	97.5%	P(>0|Y)
β_0	-60.31	1.06	-60.86	-59.83	0.00
β_1	0.74	0.25	0.29	1.15	0.99
β_2	0.02	0.28	-0.55	0.73	0.37
ε	0.06	0.43	-0.59	0.69	0.53

3.2.3 Summary and Possible Extensions

A Bayesian time series framework is a promising approach for Antarctic sea ice reconstruction. A Bayesian reconstruction creates an ensemble of plausible reconstructions that can be used readily in further analyses, a much better product than the mean reconstruction of currently used linear regression models. Furthermore, we can employ our knowledge about Antarctic sea to create informative priors for some model coefficients. Another important aspect of Antarctic sea ice is its autocorrelation structure that is ignored by current modeling approaches.

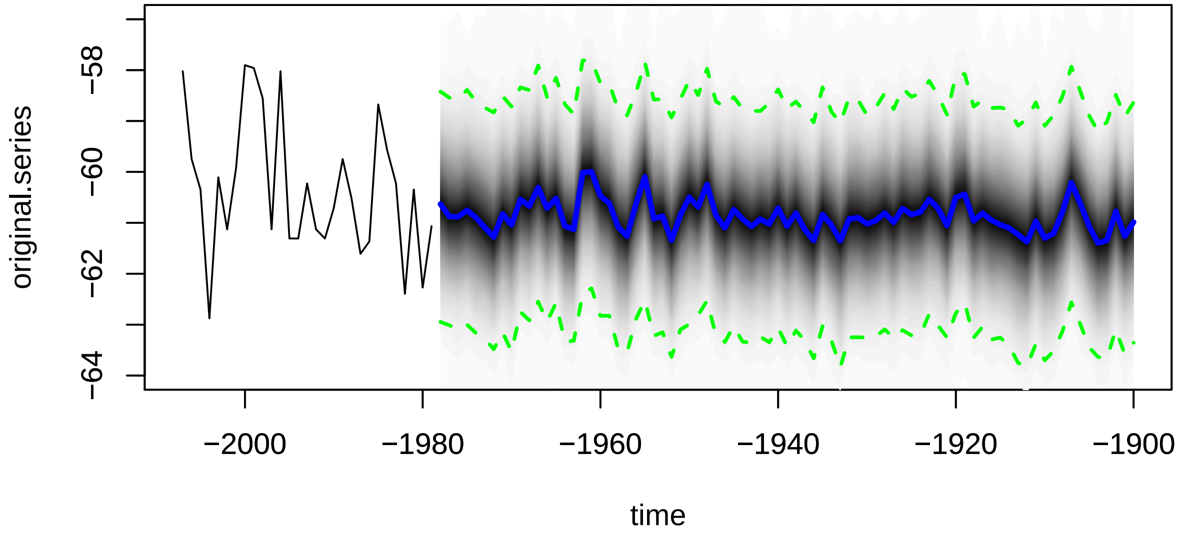


Figure 3.3: Prediction of maximal sea ice extent at 146°W using the BSTS model proposed in Equation (3.2). The posterior mean is plotted in solid blue with shaded posterior density and 95% prediction interval in dashed green.

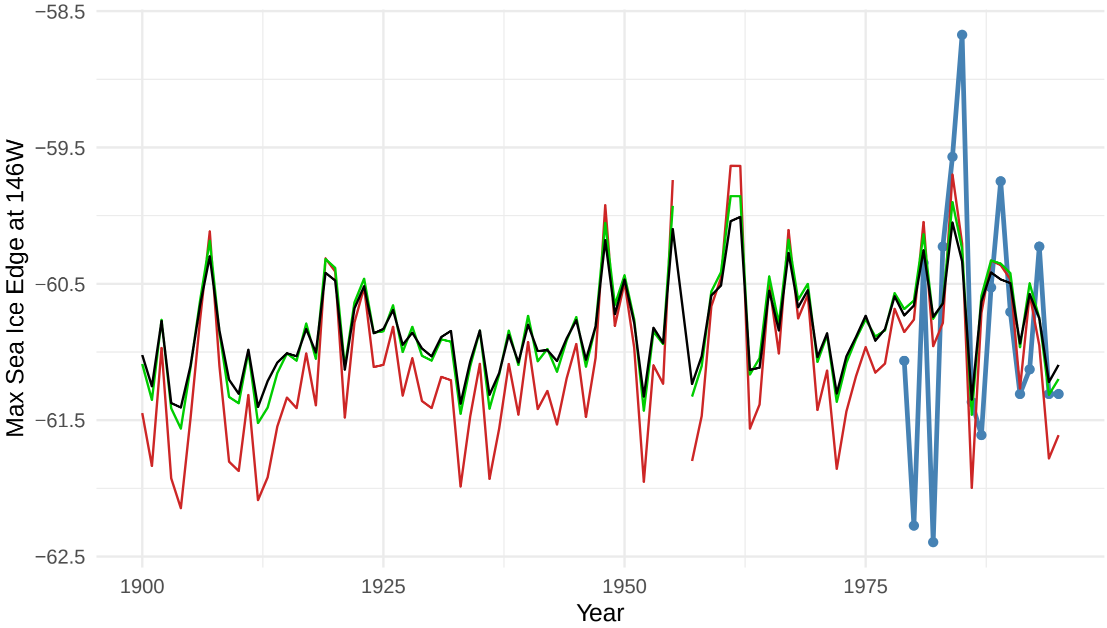


Figure 3.4: Reconstruction of maximal annual sea ice extent at 146°W for the 20th century using annual MSA concentration. The true maximal ice extent as directly measured using remote sensing is plotted in blue, the OLS prediction is plotted in green, the MAR prediction in red, and the BSTS prediction in black.

To improve the stability of a reconstruction model, a target variable with smaller variability could be used. The average sea ice extent in the month of September (the month of maximal sea ice extent in Antarctica) could be used as a target variable instead of the maximum. Choosing the sea ice extent of a narrow sector rather than one longitude would also increase stability.

There are a number of ways to extend the proposed BSTS model for sea ice reconstruction based on ice core proxies. Time varying intercepts and coefficients might be important for other ice cores, where the strength of the association between the ice core proxy and the sea ice extent wanes over the course of the satellite era, see e.g., (HCA16) who exclude ice cores from their analysis where they detect time varying effects. For the reconstruction presented in this paper a time varying coefficient for MSA Flux was tested but found to be constant.

It is also possible to combine multiple ice cores or multiple proxies from the same ice core to create a more accurate and stable reconstruction of the local sea ice extent. Novel models could be developed to reconstruct sections or even the entire sea ice edge based on multiple ice cores.

3.3 Weather Station Based Reconstruction

(FSR22) reconstruct Antarctic sea ice by sector and season for the 20th century using temperature and sea level pressure recorded by weather stations in the Southern Hemisphere as compiled in (FJG16). Each season and sector are modeled and reconstructed completely independently, ignoring their spatio-temporal correlation. (FSR22) provide ensemble reconstructions by replicating their reconstruction model for different subsets of covariates, but implicit biases in the selection of covariates lead to an underestimation of reconstruction uncertainty. Their approach incorporates some reconstruction uncertainty in average reconstructions but neglects the reconstruction uncertainty in the unexplained error terms. The completely separate reconstruction for each season is a weak point, leading to disjoint reconstructions between neighboring seasons (i.e., Spring and Summer of 1962) with no matching or continuity between reconstructions. Even within a single reconstructed ensemble member there is no temporal continuity, in the sense that the same season in neighboring years should be correlated. The lack of explicit modeling of temporal autocorrelation between neighboring seasons and years makes it impossible to assemble complete reconstructed time series ensembles from these seasonal ensemble reconstructions.

(FSR22) create a seasonal reconstruction of Antarctic sea ice by sector. For each of the five sea ice sectors (King Haakon VII, Ross Sea, East Antarctica, Weddell Sea, Bellingshausen Amundsen Sea) and each of the four seasons (DJF, MAM, JJA, SON), a set of linear models are created. As predictors, we use monthly temperature and sea level pressure measurements from weather stations in the Southern hemisphere and 0 to 9 climate indices (IPO, AMO, PDO, SOI, SSTsNino 1 & 2, SSTsNino 3 & 4, SSTsNino 3, SSTsNino 4, SAM). This data set is first augmented by computing lags of one, two, and three months for each variable, and then aggregated to seasonal averages, resulting in seasonal variables and seasonal lagged variables X, with X_i denoting the $d = 167$ dimensional covariate vector for year i. Each linear model predicts the sea ice for one season and one sector denoted Y, with Y_i being the ith year in the data, based on 0 to 35 principal components $\tilde{T}_i$ computed from these temperature and sea level pressure measurements and climate indices that are significantly correlated to the sea ice in this sector and season over the entire 42 observed years at decreasing p-value thresholds $p = \{0.1, 0.05, 0.025, 0.01\}$. One model is created including all temperature and sea level pressure measurements and climate indices. This means that for every sector and season a total of

$$(4 \text{ significance thresholds} \times 10 \text{ climate ind.} + 1 \text{ all variables}) \times 36 \text{ PC retained} = 1476$$

models are created. Choosing the optimal number of principal components to retain using highest coefficient of efficiency (CE; e.g. CMS99), 41 models remain per season and sector.

The CE is a R^2 type measure quantifying the variance explained by the model in question,

$$\mathrm{CE} = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \tag{3.3}$$

with Y_i the sea ice in the ith year, $\hat{Y}_i$ the model's out of sample estimate for the ith year, and $\bar{Y}$ the average sea ice across all years (here $\bar{Y} = 0$ as we work on sea ice anomalies). The CE takes values in $[-\infty, 1]$, and only values above zero indicate predictive power with one corresponding to a perfect prediction. Each model uses a set of temperature measurements, sea level pressure measurements and climate indices, that are orthogonalized using principal component analysis (PCA; e.g. AW10), to create a set of principal components (PC). Principal component analysis is a data projection of the original d-dimensional covariates X into an orthogonal space,

$$T = XW, \tag{3.4}$$

where T is the d dimensional data in the transformed space and W is a $d \times d$ dimensional projection matrix chosen such that it maximizes the variability of X captured in the first k columns of T for all $k = 1, \ldots, d$. This can be used to achieve a dimensionally reduced data set $\tilde{T}$ of dimensionality $\tilde{d} \ll d$,

$$\tilde{T} = X\tilde{W}, \tag{3.5}$$

where $\tilde{W}$ are the first $\tilde{d}$ columns of W. In order of magnitude of their correlation with the sea ice in the respective season and sector, d^* of them (up to 35 PCs) are retained as predictor variables T^* in a regression model predicting sea ice. Using vector notation, this principal component regression model predicting sea ice in a sector and season as target variable Y using PCs $\tilde{T}$ is of the form

$$Y_i = \tilde{\beta}_0 + \tilde{\beta}\tilde{T}_i + \varepsilon_i, \tag{3.6}$$

for every observation $i = 1, \ldots, n$, with intercept $\tilde{\beta}_0$ and $\tilde{d}$-dimensional regression coefficient $\tilde{\beta}$. This is the same model as regression model on the original covariates with

$$Y_i = \beta_0 + \beta X_i + \varepsilon_i \tag{3.7}$$

if $\beta_0 = \tilde{\beta}_0$ and $\beta = \tilde{W}\tilde{\beta}$. This model will result in identical error terms ε_i and predictions for the target variable Y_i for all i. As we have multiple measurement types (sea level pressure, temperature) and lags (1, 2, 3) from each weather station, we compute the importance of each weather station by summing the absolute values of all regression coefficients corresponding to any of its variables. The aggregated variable importance for each sector and season are shown in the left column of Figures 3.6 to 3.10.

The individual PC regression models' performance is evaluated using leave-one-out cross validation (LOOCV) which is akin to the jackknife, see e.g., Chapter 11 of (TE93) for an introduction. This not really an ideal performance measure, as the actual reconstruction of interest is a prediction outside of the data's time frame, as opposed to new observations within. Thus, the LOOCV is not an ideal validation for reconstruction accuracy but it measures the credibility of our model. The LOOCV creates an out of sample prediction for every year in the data set that we call the validation time series. All other years except the two neighboring ones are used to form a training data set; we leave out the additional two-year window around the "left-out year" to reduce contamination of the training data due to autocorrelation. In this training data set, the previously computed PCs are used to estimate a linear regression model for the relationship between the PCs and the sea ice in this season and sector, which is then used to predict the sea ice in the year that was originally left out, thus creating a time series of predicted sea ice for every year which we call the validation time series, which can be used to evaluate the method's predictive performance. A full algorithm for our LOOCV to compute the validation time series is:

```
1. INPUT numeric [42] vector $Y$; numeric [42, $d^*$] matrix $T^*$
2. INIT numeric [42] vector $\hat{Y}$; integer test_index; integer [37-39] vector training_indices
3. FOR $i$ in {1, ..., 42} {
    3.1. SET test_index $= i$
    3.2. SET training_indices $= \{1, \ldots, 42\} \setminus \{i-2, i-1, i, i+1, i+2\}$
    3.3. INIT numeric $\beta_0$, numeric [$d^*$] vector $\beta^*$
    3.4. SET $\beta_0, \beta^* = \underset{\beta_0, \beta^*}{\operatorname{argmin}} \sum_{j \in \text{ training\_indices}} \left(Y_j - (\beta_0 + \beta^* T_j^* + \varepsilon_j)\right)^2$
    3.5. SET $\hat{Y}_i = \beta_0 + \beta^* T_i^*$, for test_index $i$
    ENDFOR
4. OUTPUT $\hat{Y}$
```

taking inputs Y the 42 dimensional vector of sea ice anomalies in a specific season and sector and T^* the $42 \times d^*$ dimensional matrix containing the selected principal components and returning the 42 dimensional vector $\hat{Y}$ containing the validation time series of only predicted values. We use the notations $\{\ldots\}$ to denote a set and $\setminus$ the "set difference" operator $A \setminus B := \{x \in A : x \notin B\}$. The validation time series and the known true sea ice extent are used to compute the validation correlation and validation CE. The validation CE is the defined as the CE between the true sea ice Y and the validation time series $\hat{Y}$, see Equation (3.3), the validation correlation is defined their Pearson correlation coefficient.

3.3.1 Concerns

The main concern is that the measures of performance from the LOOCV reported by (FSR22) are overly confident due to variable selection and model selection effects. The predictor and sea ice data from all years including the left-out year are used for variable selection and model selection, making the LOOCV optimistic in its assessment of model performance as it does not account for a lack of separation between training and test data.

3.3.1.1 Variable selection problem

The first problem is variable selection, which is a type of multiple testing problem. In statistics, a multiple testing problem occurs when a set of many statistical inferences are analyzed simultaneously (e.g. Mil81), here the original method tests for 647 variables whether or not they are significantly correlated with the sea ice in a specific season and sector. In general, the more inferences are made, the more likely it is for one or more erroneous inferences to occur. In our case it means that when testing for one covariate whether it it significantly correlated with the sea ice in a season and sector, the significance threshold does limit the probability of erroneously claiming a correlation, but when testing all 647 covariates it is exceedingly likely that at least one of the covariates that are truly uncorrelated to the sea ice will have a p-value below the threshold. Statistical techniques developed to address this issue usually limit some sort of error rate using a stricter significance threshold for individual comparisons to adjust for the number of inferences being made.

Models are created using predictor variables chosen based on their correlation with the sea ice over the entire data set. These models are then evaluated in a LOOCV with the values from the two years next to the left-out year being left out as well, but the covariates and sea ice from the left-out years have already been used for variable selection in the previous step. The correlations to select the temperature and sea level pressure variables as well as the climate indices to go into the principal component analysis have been computed on the entire data set. This violates the separation of training and test data underlying the LOOCV. Furthermore, choosing PCs highly correlated with the sea ice over the entire time period violates the separation of training and test data in the LOOCV again. Not separating training and test data in the LOOCV is the main concern which can lead to dramatic overconfidence in the model performance on new data. To showcase this issue in an extreme case, consider a thought experiment where we have many possible predictor variables to choose from, but none of them have any predictive power whatsoever, e.g. all predictor variables are draws from a random number generator. In this scenario we expect $\alpha \times 100\%$ of the predictor variables to be correlated with the sea ice at a significance threshold α. When we use these significantly correlated variables in a LOOCV they will have some predictive power for the left-out year, because we know that they are correlated significantly with the entire time series. In an alternative thought experiment, where we have many predictor variables that are not correlated with the sea ice at all, but now we only use the correlation between the predictor variables and the sea ice in all years but the one we try to predict for variable selection, the correlation will not hold up for the left-out year.

3.3.1.2 Model selection problem

A second problem is model selection, which is a multiple testing problem as well. In the original method a total of 1476 (or 41 if not considering the number of PCs retained) models are being evaluated in the LOOCV for every season and sector. The LOOCV allows the identification of the best model from those being tested, but the model performance of the best model in the LOOCV cannot be taken at face value for new data. Ideally, it's model performance on new data would be evaluated using a holdout data set, to see how it performs on entirely unseen data. Intuitively, the more models we compare in a LOOCV, the more likely we are to observe models who just happen to be doing well without any actual predictive power on outside data. To showcase this problem, consider a thought experiment where we have one model without any predictive power, for example a random number generator. It is likely that it won't do well in a LOOCV. If instead we try a large number of such models it is likely that some of them will give results that happen to be close to the true sea ice. If we then use the seemingly best of these models to predict new data outside of the training data, it will not have any predictive power whatsoever.

3.3.2 Impact on Results

We propose two ways to analyze the impact of the variable selection problem and the model selection problem. The first approach adapts the method slightly to do all variable selection and PC selection independently for each year LOOCV using only the respective training data, thus getting rid of the variable selection problem but not addressing the model selection problem. The second approach is the creation of a true holdout data set, which addresses both the variable selection problem and model selection problem simultaneously but suffers from high statistical uncertainty due to small sample sizes.

3.3.2.1 Complete LOOCV

The variable selection problem can be solved in a straight-forward fashion. The original method presented in Section 3.3 can be adapted slightly to not suffer from a variable selection problem by doing the variable and PC selection process independently for each left-out year of the LOOCV instead of once for all years. This means that the LOOCV is complete, it not only encompasses the uncertainty from the data but also the variable selection process. For every left-out year in the cross validation, different predictor variables might be selected which in turn leads to different principal components and a different regression model. This is not a problem; it just changes the LOOCV's interpretation slightly from "This specific regression model has validation CE x" to "A regression model created according to these rules has validation CE x". By doing the variable selection and PC selection inside of the LOOCV, we make sure that the information of the left-out years is not being used in their prediction. The results are a lot worse, the average validation correlation across all sectors and seasons drops from 0.843 in the original method to 0.370 using complete LOOCV, and the average validation CE drops from 0.700 to 0.027. A breakdown of the model performance by season and sector is shown in Figure 3.5. The original method reports higher, i.e., better, validation correlation and validation CE across all seasons and sectors. Note that the purported increase

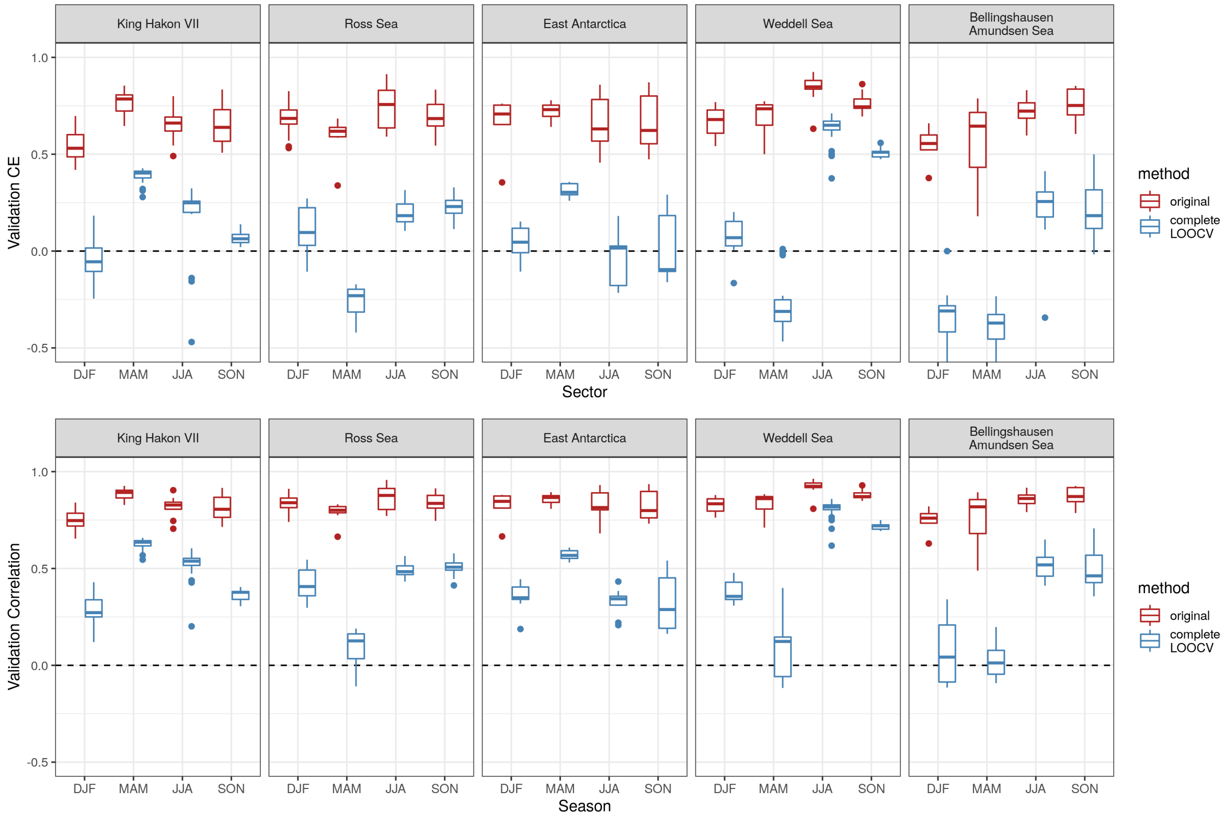
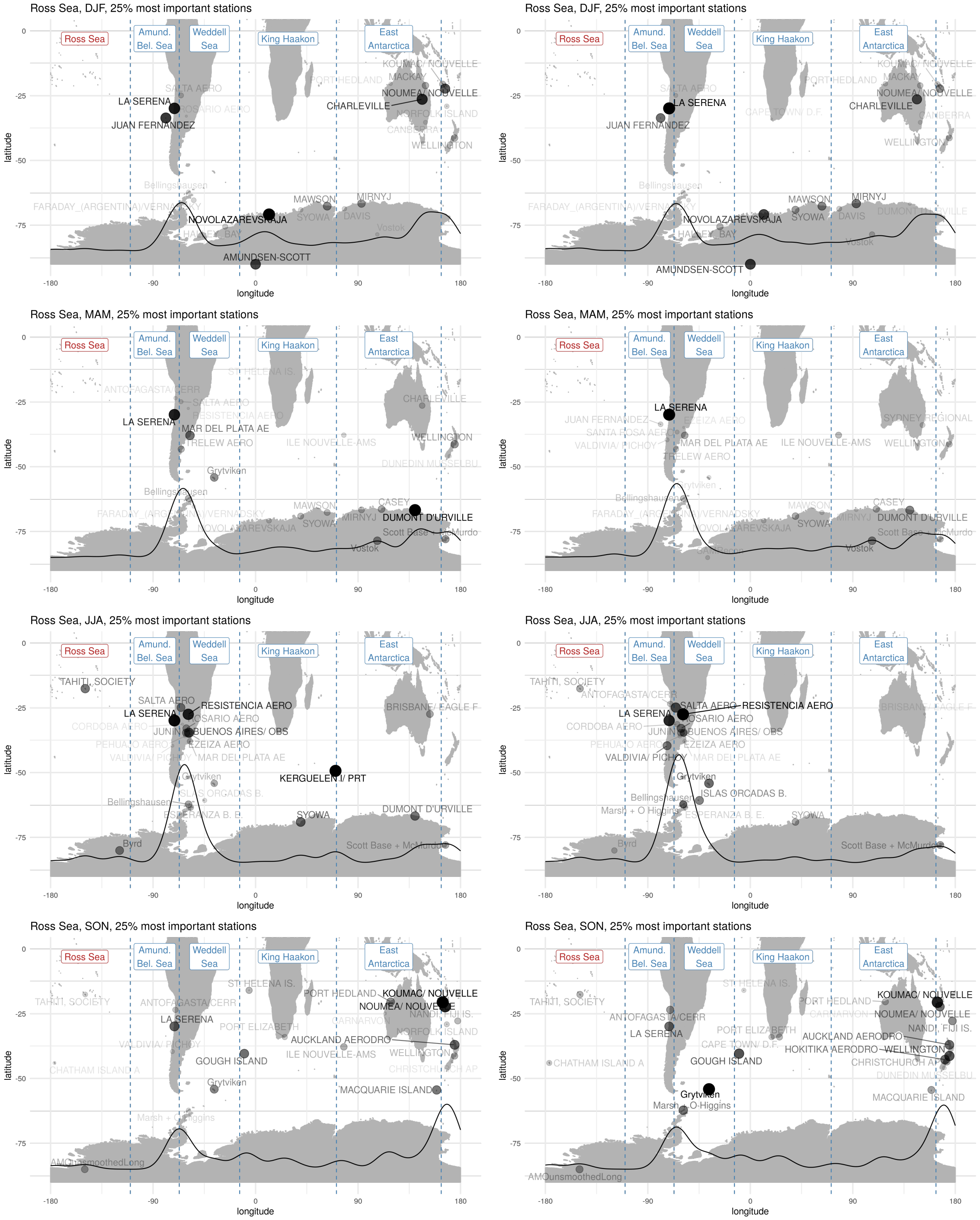


Figure 3.5: Performance in LOOCV by season (boxplots) and sector (panels) of original method (red) and complete LOOCV method (blue) using validation CE (top) and validation correlation (bottom).

in performance of the original method is exclusively due to the use of data from the years that were supposed to be left out in the LOOCV in the variable selection process. All other aspects of the original method are replicated faithfully. In analogy with the original method, we can compute a variable importance for each weather station by summing the absolute values of all regression coefficients corresponding to any of its variables (SLP and TMP with their lags) and the climate indices. The aggregated variable importance for each sector and season are shown in the right column Figures 3.6 to 3.10. We note that the figures shown here are virtually identical between the original method (left column) and the complete LOOCV (right column). This is true even for the individual models when averaging across all left-out years of the cross validation. This is not surprising, as the complete LOOCV averaged over all left-out years of the LOOCV should result in a similar variable selection and weighting. The differences between the two approaches can thus be reasoned to be found in individual models for each left-out year, where the complete LOOCV selects different variables and fits different models. On average the original method and the complete LOOCV select similar variables with similar weights though, which makes sense as the original method fit the "same" model on the "same" data.

3.3.2.2 Performance assessment on holdout data

The joint effect of the variable and model selection problem can be assessed using a true holdout data set as a means of estimating their compounded effect on the predictive performance of new data. Again, this is done to estimate the method's predictive performance when reconstructing sea ice of the entire 20th century. We create a validation scenario where we leave out the first 6 years and reconstruct them using the remaining 37 years, which is a small-scale version of what we are trying to do in the 20th century. For this purpose, we set aside the first 6 years of the data as a holdout data set, and pretend it doesn't exist until we have identified the one very best model using the remaining data. This one model gets to predict the held-out data, and its performance on this data set is a true estimate of how well it will perform on new data. Using the original method on the $42 - 6 = 36$ remaining years of data we create a set of linear regression models for each season and sector. When applying these models to the holdout data set, their performance decreases dramatically, see Figure 3.11 or Figure 3.12 for a breakdown by season and sector. The average correlation decreases from 0.843 in the original method to 0.539 on the 6 year reconstruction period, the average CE drops from 0.700 to a median of -0.733 (the mean was not meaningful due to extreme outliers) on the hold out data. In an actual study we would only test the performance of the one best performing model on our holdout data set, but for the sake of illustration we test every model created by the original method on our holdout data. Almost all of them


Ross Sea, DJF, 25% most important stations

Ross Sea, DJF, 25% most important stations

Ross Sea, MAM, 25% most important stations

Ross Sea, MAM, 25% most important stations

Ross Sea, JJA, 25% most important stations

Ross Sea, JJA, 25% most important stations

Ross Sea, SON, 25% most important stations

Ross Sea, SON, 25% most important stations

Figure 3.6: Average variable importance for Ross Sea sector per season.

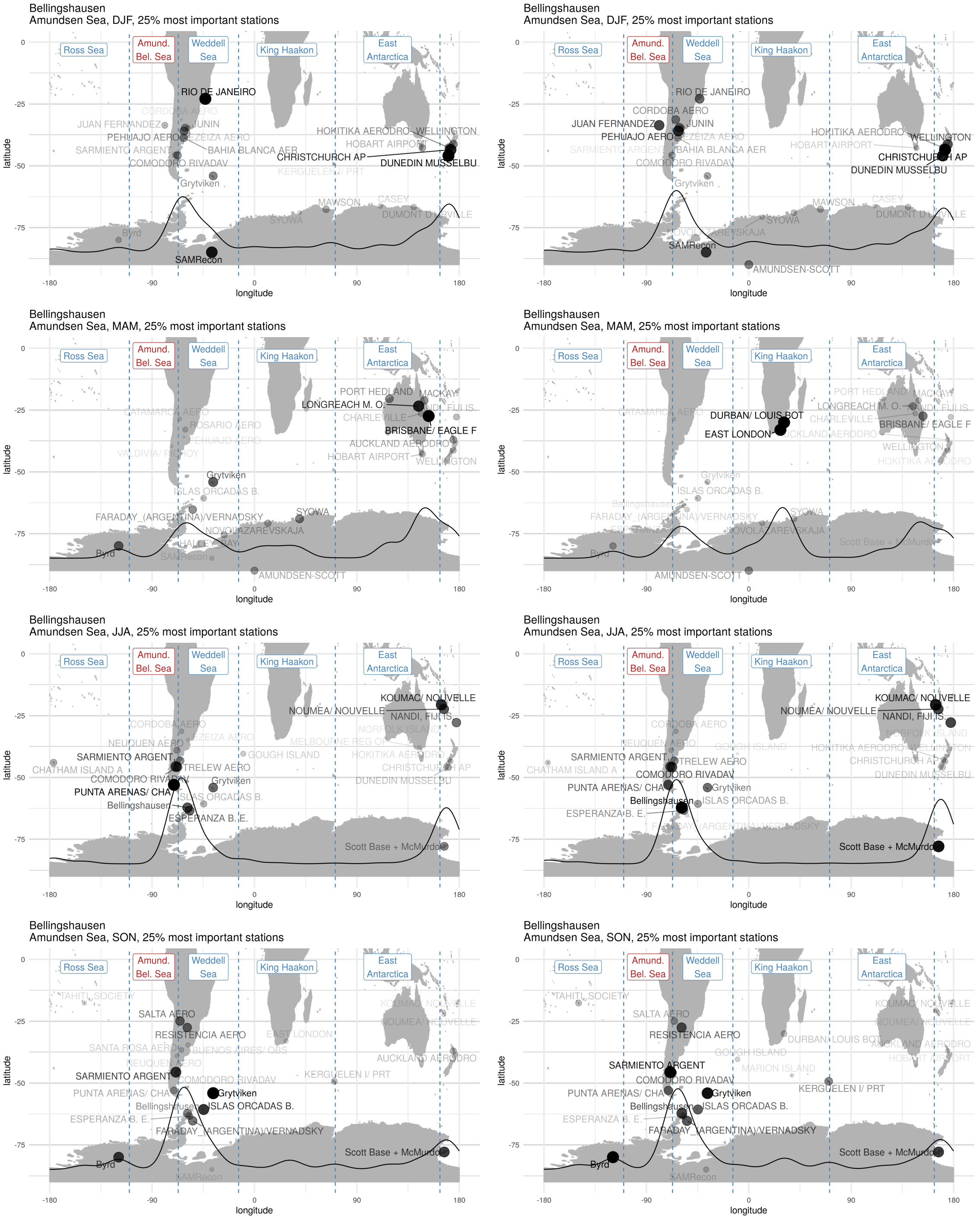


Figure 3.7: Average variable importance for Bellingshausen Amundsen Sea sector per season.

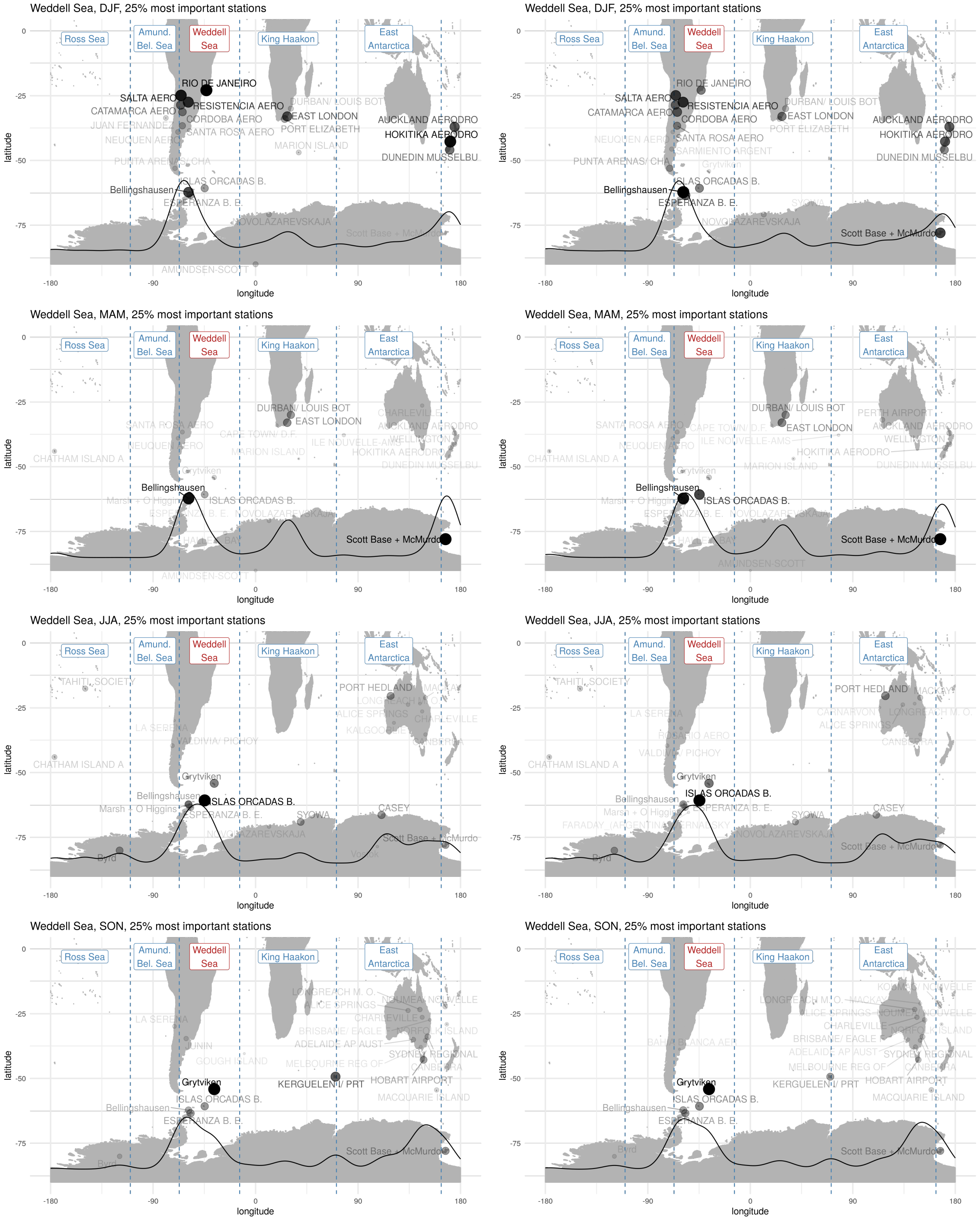
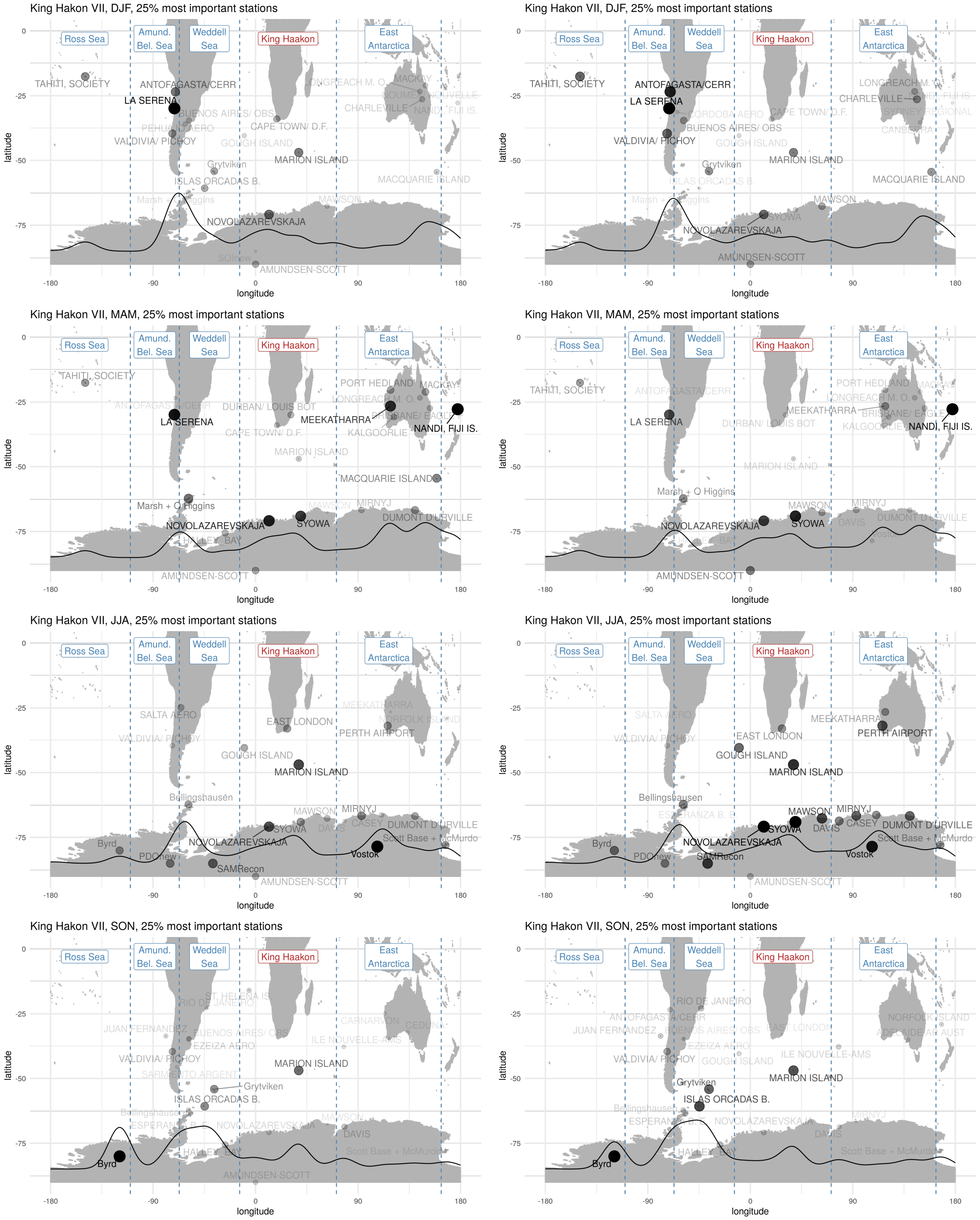


Figure 3.8: Average variable importance for Weddell Sea sector per season.



King Hakon VII, DJF, 25% most important stations

King Hakon VII, DJF, 25% most important stations

King Hakon VII, MAM, 25% most important stations

King Hakon VII, MAM, 25% most important stations

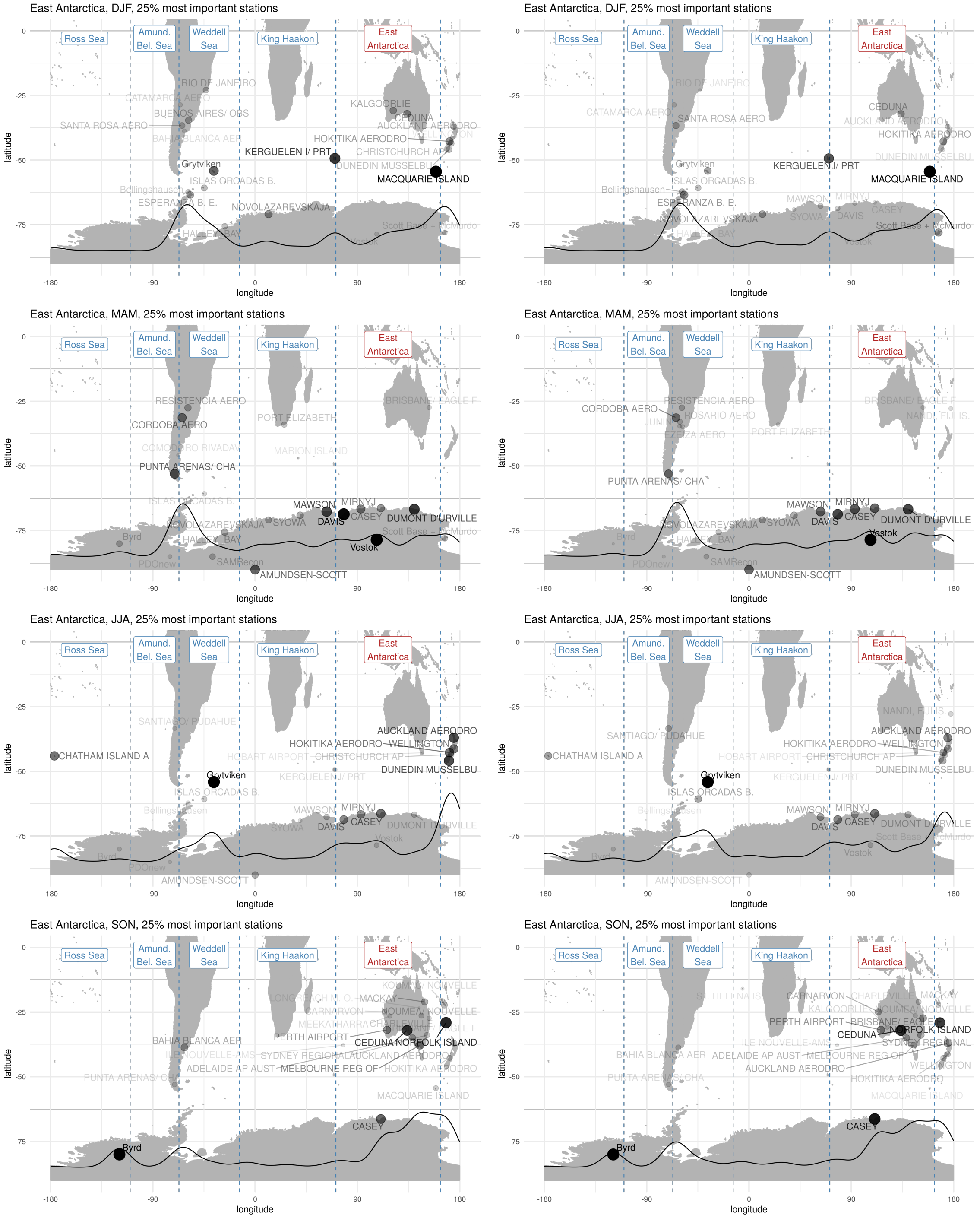
King Hakon VII, JJA, 25% most important stations

King Hakon VII, JJA, 25% most important stations

King Hakon VII, SON, 25% most important stations

King Hakon VII, SON, 25% most important stations

Figure 3.9: Average variable importance for King Haakon VII Sea sector per season.


East Antarctica, DJF, 25% most important stations

East Antarctica, DJF, 25% most important stations

East Antarctica, MAM, 25% most important stations

East Antarctica, MAM, 25% most important stations

East Antarctica, JJA, 25% most important stations

East Antarctica, JJA, 25% most important stations

East Antarctica, SON, 25% most important stations

East Antarctica, SON, 25% most important stations

Figure 3.10: Average variable importance for East Antarctica sector per season.

perform worse on the holdout data than in the LOOCV and there is no clear association between the performance measures in the LOOCV validation and holdout data. We would expect that models which perform exceptionally well in the LOOCV should also perform exceptionally well on the holdout data. This is not the case.

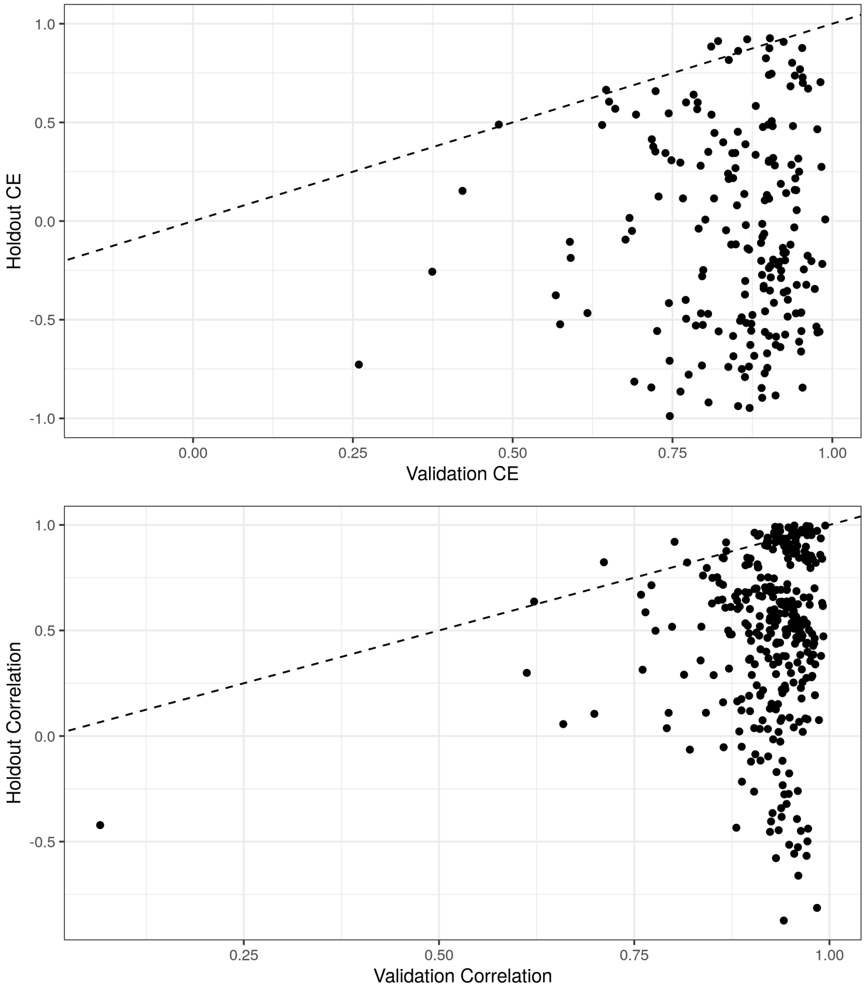


Figure 3.11: Validation vs. holdout CE (top) and validation vs. holdout correlation (bottom). The dashed line graphs identity. This is a combination of all plots in Figure 3.12.

3.3.3 Stability of Variable Selection

Analyzing the difference between the original method and LOOCV on the level of individual left-out years of the LOOCV is not humanly possible for 42 left-out years of 41 models in 5 sectors and 4 seasons (that's 34440 models to look at). Instead, we focus on how stable the variable selection is across the left-out years of the LOOCV by analyzing the inclusion probability of each covariate in the models for each season and sector. As the variable selection is based on the p-values associated with the correlation coefficients between the sea ice and each covariate, we analyze the stability of these p-values first. To this end, we analyze the distribution of p-values for the correlation coefficient between the sea ice and all 647 covariates (including lags) across the 42 left-out years of the LOOCV for every season and sector. If these p-values are stable, we expect little variation in them from one left-out year to another and a standard deviation close to zero, and the subsequent variable selection is stable as well. If the p-values are unstable, their standard deviation approaches its upper bound which is the standard deviation of a Uniform random variable on $[0, 1]$, which is $\sqrt{1/12} \approx 0.29$, and the variable selection will be unstable as well. The distribution of p-values of the correlation coefficient for the Ross Sea sea ice anomaly in MAM and the Kerguelen 1/Prt station over all left-out years in the LOOCV is shown in Figure 3.13. Some correlations appear to be stable (e.g., Kerguelen SLP at lag 1 and lag 2), but others display notable change in the correlation coefficients p-value. This includes covariates that are significantly correlated in some left-out years but not others, e.g., Kerguelen temperature at lag 2 and lag 3. The stability of the correlation coefficients' p-values over all left-out years in every season and sector is depicted in Figure 3.14, showing the distribution of the p-values' standard deviations across the left-out years in the LOOCV for all covariates by season and sector. This lack of stability in correlation coefficients p-values leads to a lack of stability in variable selection of the complete LOOCV method. This in turn explains a difference in performance. The complete LOOCV method only selects variables that are correlated in the training data, and the high degree of variability observed in the correlation

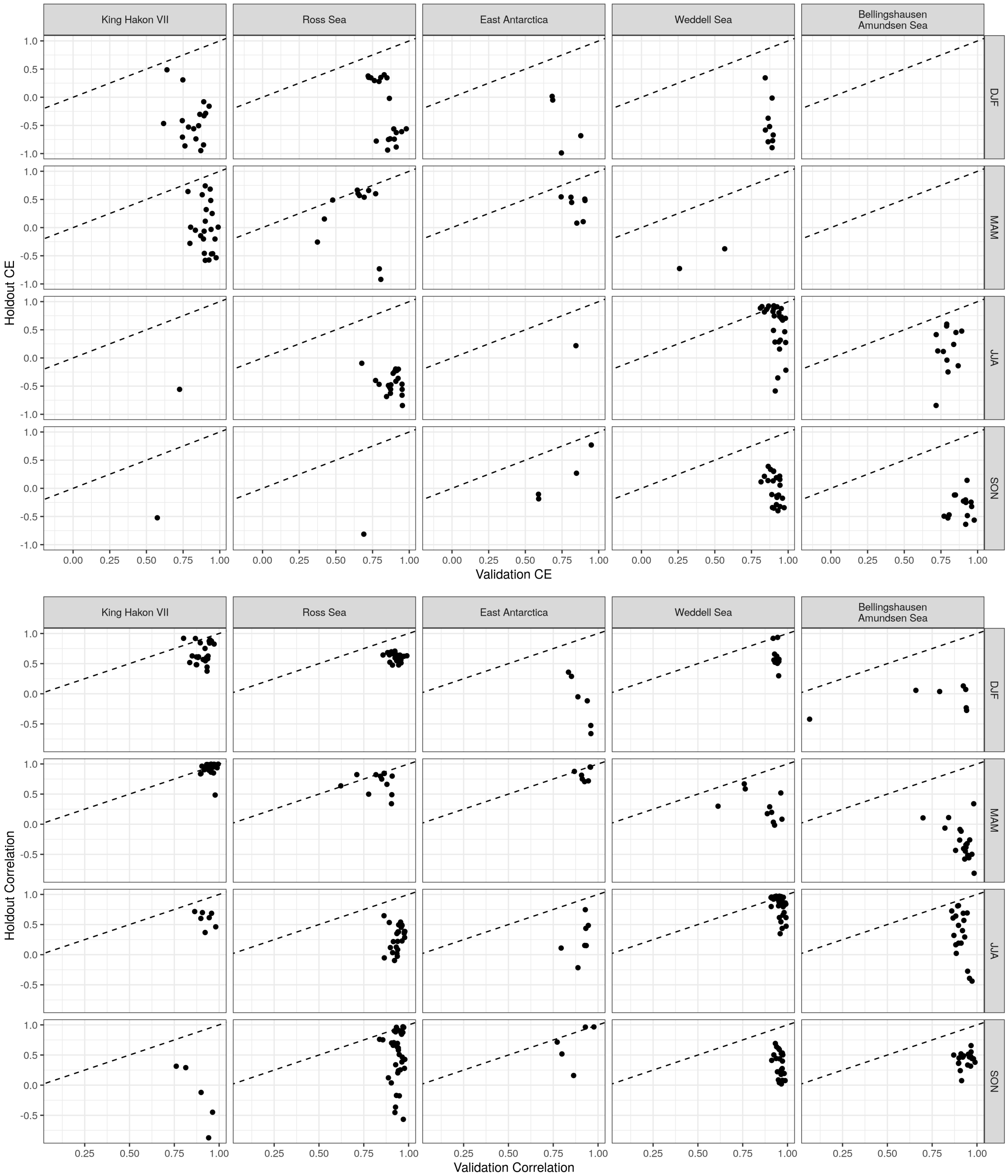


Figure 3.12: Validation vs. holdout CE (top) and validation vs. holdout correlation (bottom) by season and sector. The dashed line graphs identity.

p-values computed in every left-out year of the LOOCV indicates that these correlations do not necessarily hold up in the left-out test set.

As a direct measure of model stability, we analyze the stability of the variable selection across the 42 left-out years of the LOOCV. For every season and sector, the correlation coefficient and its p-value between every variable and the target sea ice are computed, see Figures 3.13 & 3.14. Then, for every variable, we tally up in what percentage of the 42 left-out years it would have been used at a given correlation threshold, we chose 0.05 for the purpose of illustration. The distribution over the proportions of how many times each variable is used are depicted in Figure 3.15. A stable variable selection means that a variable is either never used (the proportion is zero), or always used (the proportion is one). Proportions in between indicate an unstable variable selection. As the variable selection process is independent for each covariate, highly correlated covariates will not be chosen in substitution of one another as they might be in a joint variable selection process using for example p-values of partial correlation coefficients. This means that highly correlated covariates will be jointly selected (or not), but do not bar one another from being selected.

3.3.4 Permutation Test

To assess the effect of the variable selection problem we perform a permutation test (Fis37) to test the null hypothesis “The model performance obtained in (FSR22) was obtained by random chance” with the alternative hypothesis “The model performance obtained in (FSR22) was **not** obtained by random chance”. The idea of a permutation test is to permute the data at random in order to scramble the association between one or multiple variables. When randomly reordering one of two variables, we know for a fact that any association between them is purely due to chance, and when repeating the permutation process multiple times we can obtain a distribution of any statistic measuring association between them under the null hypothesis “There is no association between the variables”, which can be used to compute a p-value for the observed value of this statistic in the original data. Here, we

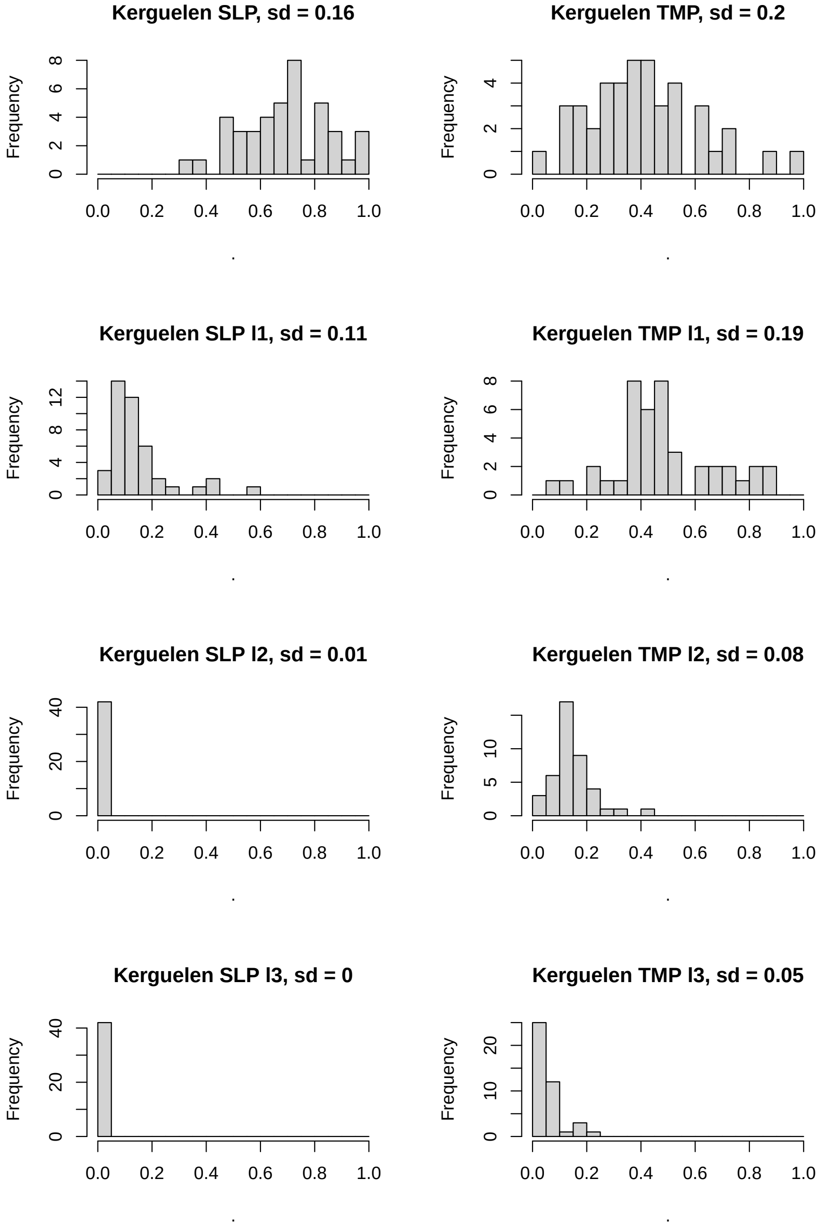


Figure 3.13: Distribution of p-values of the correlation coefficient for the Ross Sea sea ice anomaly in MAM and the Kerguelen 1/Prt station over all left-out years in the LOOCV.

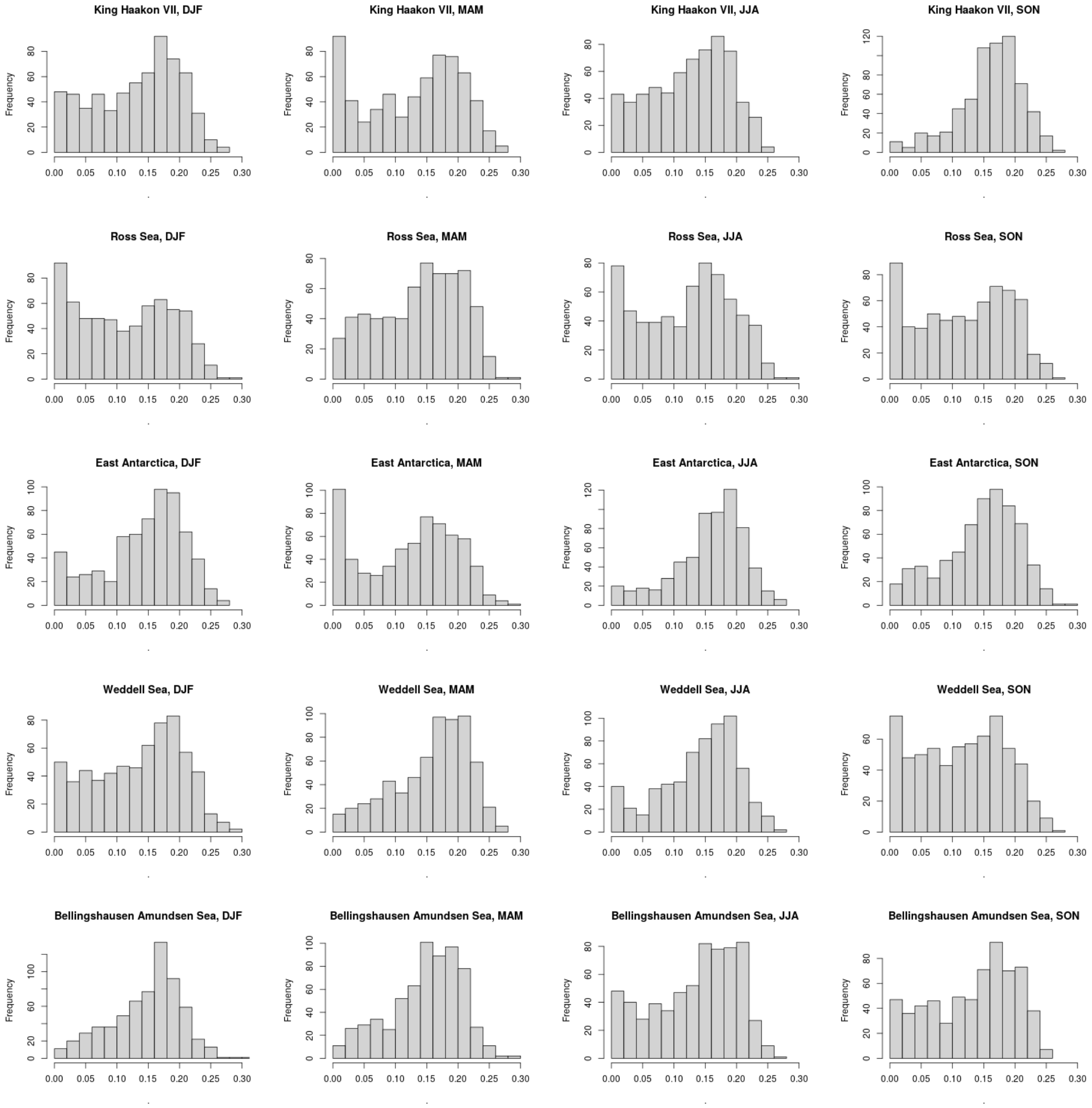


Figure 3.14: Distribution of standard deviations of p-values over all left-out years in the LOOCV per covariate in every season and sector.

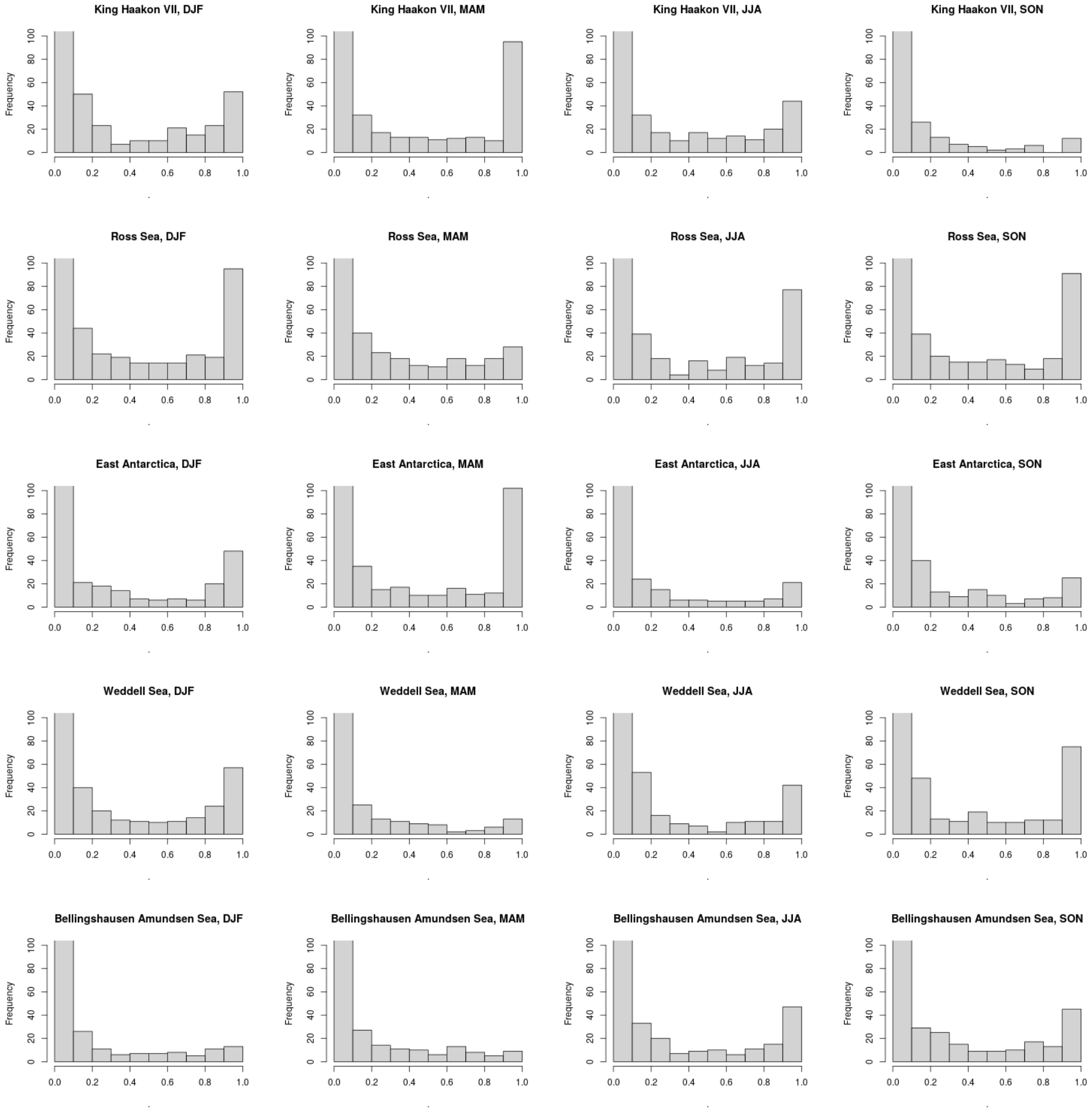


Figure 3.15: Distribution of all 647 covariates' selection probabilities over all left-out years in the LOOCV per season and sector. The y-axes are truncated at a count of 100 to allow focus on variables that were actually selected.

randomly permute the seasonal sea ice anomalies within their respective seasons, but keep the ordering of the covariates in chronological order to preserve their autocorrelation and principal component structure. The model performance of the original method on permuted data is shown in Figure 3.16, showing the model performance on permuted data in the LOOCV by sector and season, with the original model performance on chronological data for reference. The original method on the chronological data notably outperforms the original models on permuted data, indicating that the original method does extract real information from the data. We can safely reject the null hypothesis "The model performance obtained in (FSR22) was obtained by random chance" in favor of the alternative hypothesis. The original model on permuted data performs a lot better than random chance would suggest (CE $\approx$ 0, Corr $\approx$ 0), indicating a strong impact of the variable selection problem. The observed difference between "no predictive power" and the observed "predictive power" of the permuted models is due to the variable selection problem (plus some random noise due to the permutations drawn). The difference between the predictive power of the models on permuted data and chronological data is due to the actual predictive capabilities of the covariates captured by the model. There is some predictive power, but the original method is overconfident in its predictive power.

3.3.5 Summary

We replicate the Antarctic sea ice reconstruction method introduced by (FSR22), a principal component regression model fit independently for every season and sea ice sector in a frequentist framework, resulting in models fit on 42 observation (one per year in the observation period). The principal components are computed on a subset of the 647 predictor variables that are significantly correlated with the sea ice in the respective season and sector, and up to 35 PCs are used in the final models. Predictive performance is assessed using the CE and correlation coefficient in a LOOCV. Due to the variables being selected on the full 42 years of observed data before analyzing the models' predictive power in a LOOCV, information

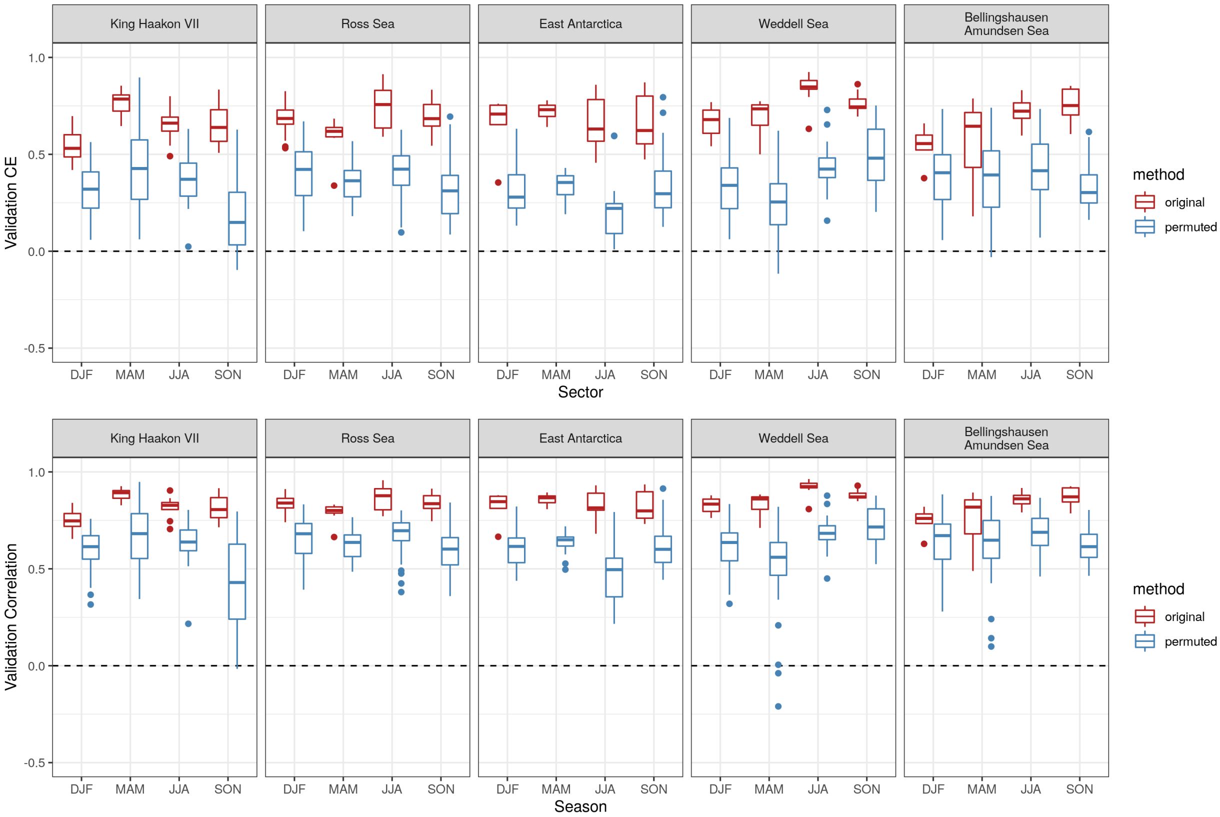


Figure 3.16: Performance in LOOCV by season (boxplots) and sector (panels) of original method on chronologically ordered data (red) and original method on permuted data (blue) using validation CE (top) and validation correlation (bottom).

from the left-out year is incorporated into the models predicting them, causing the LOOCV to be overly optimistic in the original method's predictive performance. We estimate this skew in the methodology using a complete LOOCV where we perform the variable selection process independently for every left-out year in the LOOCV. This method shows a marked decrease in predictive performance, indicating a large effect of the original method's variable selection problem. To pinpoint where the original method and complete LOOCV differ, we first compare the average station importance in both methods showing only marginal differences between different models for the same season and sector, indicating that differences occur on the level of individual left-out years of the LOOCV. This theory is validated by testing the stability of the variable selection process for every left-out year. The variable selection is based on the p-value of the correlation coefficient between each covariate and the sea ice in the respective season and sector, which are found to be unstable. In a second step, the actual inclusion probability for every covariate can be computed based on the p-value associated with its correlation, which are also found to be unstable. The large effect of the variable selection problem is corroborated by a permutation test which, when performed on the original method, still shows strong positive predictive power in the LOOCV, even though the relationship between sea ice and covariates has been artificially eliminated. The original method showing predictive power in the permutation test is further evidence of a profound methodological skew in the original method. Furthermore, for every season and sector 41 models of this type are fit, but only the performance of the best model is reported which adds an additional optimistic skew in the original method's reported predictive performance.

3.4 Discussion

The BSTS model proposed in Section 3.2.2 for sea ice reconstruction using ice cores allows the specification of a statistically rigorous and plausible covariance structure of the observations and error terms. The BSTS model explicitly accounts for the time series structure of the sea ice by including an auto-regressive error term. It also shrinks the effect of the ice core proxies towards 0 in comparison with the OLS model. This is in stark contrast with the MAR model that inflates the effect of the ice core proxies in comparison with the OLS model, which is highly problematic. Effect shrinkage is more in line with a conservative approach to science where effects have the burden of proof to be justified and should not be over-reported.

The method introduced in (FSR22) suffers from a number of drawbacks, most notably a variable selection problem. This can be addressed using rigorous variable selection in a frequentist framework or, more elegantly, using sparsity priors in a Bayesian framework. Variable selection / model sparsity is so essential to the reconstruction problem at hand, that the Bayesian framework's natural ability to incorporate the uncertainty of the model sparsity into the sea ice reconstruction makes it a better choice. Additionally, the Bayesian framework automatically creates a ensemble of reconstructions in the form of posterior draws for the sea ice in the reconstruction period. Another major struggle of the (FSR22) method is the relative brevity of the 42 year satellite observed sea ice record, there are only 42 observations for sea ice in every season and sector. This problem can be mitigated by "borrowing power" across seasons as we can expect the same covariates to have a similar effect on the sea ice in one sector year-round. This relation is even more pronounced when fitting reconstruction models to monthly instead of seasonal data, as the sea ice in neighboring months of year can be assumed to have similar associations with Southern Hemisphere weather. This motivates the method presented in Chapter 4.

CHAPTER 4

A Bayesian Model for 20th Century Antarctic Sea Ice Extent Reconstruction

4.1 Introduction

Antarctic sea ice is a key component of the Southern Hemisphere climate system, but has largely been unobserved before the start of the modern satellite record with the launch of NIMBUS-7 in October 1978. Estimation of the sea ice concentration on a continental scale in Antarctica is only feasible using remote sensing (i.e. satellites) due to its vastness and remoteness. Since 1978, images that can be used to estimate the Antarctic sea ice concentration are available on a one to two day resolution (MFW16). Detailed analysis of Antarctic sea ice variability over the satellite-observed period shows that it is mainly driven by its strong annual cycle. Total Antarctic sea ice extent ranges from about 3 million km^2 in February to about 19 million km^2 in September, a difference of about twice the land area of the contiguous United States. After accounting for this strong annual cycle, a small long-term positive linear trend has been identified in the Antarctic sea ice extent from 1980-2014, contrasting with a decline in Arctic sea ice of a much larger magnitude. This positive trend is subject to strong regional trends of opposing nature, in particular a sea ice decrease in the Bellingshausen-Amundsen Sea, emphasizing the importance of analyzing the Antarctic sea ice extent by sector. On a local scale, Antarctic sea ice variability is driven by changes in the atmosphere and the ocean (HMS16). Antarctic sea ice extent has declined dramatically since a record high in 2014 to a record low in 2017 (Par19). (HR20) relate the low average sea ice extent in 2016 by an unusually early decline in sea ice. The 2014-2017 decline in Antarctic sea ice is unprecedented in the satellite record. In congruence with (FSR22) we show that even in the wider context of the entire 20th century sea ice variability this was an unusual decline that has likely not happened before. This recent sea ice decline has continued to a further record minimum in February 2022 (RH22) and even lower in February 2023, extreme values that we show to be possible but not likely to have previously occurred in the 20th century. To understand the full merit of our approach it is helpful to place the task of Antarctic sea ice reconstruction into the field of statistics. Antarctic sea ice extent is a time series, a multivariate time series if analyzed by sector as is the case here. This is relevant, as multivariate time series modeling allows the estimation and incorporation of temporal and spatial autocorrelation, here sea ice extent of neighboring months and different sectors are correlated. The task of reconstruction can be viewed as a missing data problem, specifically an imputation, see (RL19) for a comprehensive review of the field of missing data. It makes sense to treat reconstruction as a missing data problem as opposed to a time series prediction or forecasting problem, as we have observations of our predictor variables throughout the reconstruction period. Additionally, there are a few missing months during the satellite-observed period as well that need to be imputed. The Antarctic sea ice extent is missing at random (MAR) with the missingness exclusively depending on time (missing before the start of the satellite record, observed afterwards), but not on any other observed or unobserved variables or the sea ice itself, which provides the theoretical basis for asymptotically unbiased estimation. There are many strategies to imputing missing values, with using regression predictions to perform deterministic imputation being the current state-of-the-art in the Antarctic sea ice literature. This approach dramatically underestimates the variability in reconstructions, as the predicted values from the regression tend to be much less variable than the original data. As a showcase, consider a regression model with an R^2 of 30%, a good coefficient of determination in the context of Antarctic sea ice reconstruction. This value means that this model explains 30% of the variance in Antarctic sea ice (or 54% of its standard deviation), which means that its predictions will have a variance of only about 30% of the original variance. Using this model's predictions will severely underestimate the variability in reconstructed sea ice and have a strong tendency to predict values close to the climatology. This is the kind of modeling approach used in most papers in the literature. A next step would be so called random regression imputation, where the unexplained variability gets simply added on top of the regression prediction's deterministic imputations. Ideally, this step is repeated multiple times, resulting in multiple imputations with different random errors added to the deterministic predictions. In the context of time series reconstructions, we can and have to do better, as the error terms in our regression model are not independent between temporally and spatially close observations. Our choice of fitting a regression model in a Bayesian framework naturally creates multiple imputations of missing values (all values in the reconstruction period as well as the missing values in the satellite observed period) by explicitly modeling reconstruction uncertainty in average reconstructions as well as spatio-temporally autocorrelated random errors.

Pre-satellite-era Antarctic sea ice has previously been reconstructed in a number of papers (Mar97; MMK02; COM03; CG07; De 09; BCC09; ATM10; AMA11; SBB14; PPM16; TA16; HCA16; TAE19; YXL21; DGR21; FSR22), but we present the first monthly ensemble reconstruction for Antarctic sea ice extent by sector and in total for the entire 20th century. Previous reconstructions generally agree that Antarctic sea ice extent was higher in the first half of the 20th century than in the second half, but different reconstructions vary on the exact scale of this sea ice decline. This general decline is captured in our reconstructions as well as local trends on a smaller timescale. (Mar97; CG07; De 09) model discrete sections of 10° longitude and 10 day periods without taking into account the considerable spatio-temporal correlations of nearby longitudes and days of year. Our model improves upon this by explicitly modeling the covariance between sectors as well as the temporal autocorrelation structure of Antarctic sea ice. This allows us to create an ensemble of realistic reconstructions of Antarctic sea ice, as opposed to a mere of an average reconstruction with an uncertainty estimate. A series of ice cores (MMK02; COM03; BCC09; ATM10; AMA11; SBB14; PPM16; HCA16; TA16) have been used for local sea ice reconstruction, see e.g., (AWC13) and (TAE19) for a summary, but these have not been used to compile total and sectoral Antarctic sea ice extent reconstructions, and it is not straightforward how to do so. Ice cores provide useful local reconstructions but tend to be geographically restricted in their predictive range. Another drawback of these papers is that only an average sea ice extent is reconstructed with a lack of rigorous uncertainty estimation, lack of measures of external and internal validation, and disregard for the temporal autocorrelation structure of Antarctic sea ice. (DGR21) reconstruct regional Antarctic sea ice on an annual resolution based on ice cores and tree ring widths using a Bayesian framework. They use the Bayesian framework to incorporate multiple simulated runs from a climate simulation model (isotope-enabled Community Earth System Model version 1, iCESM1; BSB19) as a prior distribution for Antarctic sea ice during the reconstruction period. Using a climate model trained on satellite observations to create a prior distribution means that information from the data (the observed satellite record) is used to create the prior distribution, which is problematic as the prior distribution should represent the state of knowledge before observing the data. Their approach preserves spatial autocorrelation in sea ice, but neglects temporal autocorrelation. This means that their individual reconstructions are not (and should not be) reported, but only the reconstruction average with (slightly overconfident) uncertainty bands. (FSR22) reconstruct Antarctic sea ice by sector and season for the 20th century using temperature and sea level pressure recorded by weather stations in the Southern Hemisphere as compiled in (FJG16). Each season and sector are modeled and reconstructed completely independently, ignoring their spatio-temporal correlation. (FSR22) provide ensemble reconstructions by replicating their reconstruction model for different subsets of covariates, but implicit biases in the selection of covariates lead to an underestimation of reconstruction uncertainty. Their approach incorporates some reconstruction uncertainty in average reconstructions but neglects the reconstruction uncertainty in the unexplained error terms. The completely separate reconstruction for each season is a weak point, leading to disjoint reconstructions between neighboring seasons (i.e., Spring and Summer of 1962) with no matching or continuity between reconstructions. Even within a single reconstructed ensemble member there is no temporal continuity, in the sense that the same season in neighboring years should be correlated. The lack of explicit modeling of temporal autocorrelation between neighboring seasons and years it makes impossible to assemble complete reconstructed time series ensembles from these seasonal ensemble reconstructions.

A particular strength of our reconstruction is its implementation in a Bayesian statistical framework, see (GCS04) for an introduction, allowing us the sampling of arbitrary numbers of complete time series of plausible reconstructions of Antarctic sea ice, its overall evolution, and in particular its month-to-month progression in each reconstruction. The more commonly used frequentist framework only creates an average reconstruction and its uncertainty, without ever creating actual time series of complete reconstructions. This is good enough to describe average trends, but cannot answer research questions pertaining to individual reconstructions and their variability. The possibility to compare observed events to individual reconstructions is lost. Our reconstructions incorporate the entire information provided by weather stations in the Southern Hemisphere compiled by (FJG16), show realistic spatial correlation patterns between sectors as well as realistic temporal autocorrelation patterns for neighboring months within sectors, while faithfully including reconstruction uncertainty. A set of these reconstructions as well as the reconstruction mean are shown in Figures 4.10 & 4.11. Our ensemble of 2500 reconstructions for each sector and in total are publicly available (Mai23a).

4.2 Data

Our sea ice reconstruction involves a large set of climatological data across the Southern Hemisphere. We use monthly sea-level-pressure and temperature records from weather stations in the Southern hemisphere to trains our model to predict satellite-observed Antarctic sea ice from 1979-2020. The longer historic record of weather stations informs our model's reconstruction of sea ice for the entire 20th century.

4.2.1 Antarctic Sea Ice

These satellite sea ice data form the core of our reconstruction procedure, as they are the calibration data set used for modeling the relationship between sea ice and the climatological data in the satellite era. The monthly sea ice extent is computed from the Nimbus-7 Scanning Multichannel Microwave Radiometer and the Defense Meteorological Satellite Program Special Sensor Microwave Imager-Special Sensor Microwave Imager/Sounder (SSM/I-SSMIS). We used the climate data record (CDR) daily concentration fields from the NOAA/NSIDC CDR of Passive Microwave Sea Ice Concentration, Version 4 (https://nsidc.org/data/g02202; MFW16). This data set provides sea ice concentrations on the SSM/I-SSMIS polar stereographic grid (25 km × 25 km) from 25 October 1978 to 31 December 2020 and are daily except before July 1987, when they are given every other day. We use the 15% sea ice concentration isoline as the sea ice edge and compute the sea ice extent as the sum of the area of each grid cell to its South, which may include polynyas and other enclosed areas of lower sea ice concentration. The sea ice extent for any missing days, there is a gap in the observations from 3 December 1987 to 12 January 1988 and every other day is missing from 1978 to 1987, is stochastically imputed using a Bayesian Auto-Regressive Integrated Moving Average (ARIMA) model, a model very similar to the one used for reconstruction in our work. These imputations can be assumed to have little to no impact on our analysis due to small variation in the imputed values. Based on the completed daily sea ice extent record, we compute monthly sea ice extent as the average of all days within each month. Sea ice anomalies are computed by subtracting an invariant annual cycle computed analogously to (HR20). The sea ice extent by sector is computed as the sum of the area of each grid cell South of the sea ice edge within each sector, using the sea ice based sectors introduced in (RH14). These sectors shown in Figure 4.2 are identified by their longitudinal range as follows: Ross Sea (162°E - 250°E), AmundsenBellingshausen Sea (250°E - 290°E), Weddell Sea (250°E - 34°E), King Haakon VII (346°E - 71°E), and East Antarctica (71°E - 162°E). Note that this definition ensures the total sea ice extent is exactly equal to the sum of the sea ice extent in all five sectors.

The sea ice extent anomalies by sector and in total are depicted in Figure 4.1. The magnitude of the extrema in total sea ice anomaly (blue) tends to be larger than in any individual sector (gray). The covariance matrix for the sea ice extent anomalies in million km^2 are reported in the top half of Table 4.1. We report a covariance matrix as opposed to a correlation matrix to preserve the scale of the variances and covariances, as the magnitude of sea ice variability varies notably among sectors. The variances on the diagonal measure the magnitude of the variability of the sea ice anomaly in each sector, with larger numbers implying higher variability. Even though the total sea ice anomaly is equal to the sum of the sea ice anomalies by sector, its variance is smaller than the sum of the variances by sector. Instead, the total sea ice anomaly's covariance with the sectors can be interpreted as their contribution to the total sea ice variability, in fact the sum of the covariances of each sector with the total is equal to the variance of the total. This holds for an arbitrary number of sectors $Y_1, \ldots, Y_{N_Y}$ concatenated into an N_Y dimensional vector Y and their total sum $Y_{\sum} = \sum_s Y_s = \mathbf{1}^T Y$, as

$$\begin{aligned} \mathrm{Var}(Y_{\sum}) &= \mathrm{Var}(\mathbf{1}^T Y) \\ &= \mathrm{Cov}(\mathbf{1}^T Y, \mathbf{1}^T Y) \\ &= \mathbf{1}^T \mathrm{Cov}(Y, \mathbf{1}^T Y) \\ &= \sum_s \mathrm{Cov}(Y_s, \mathbf{1}^T Y) \\ &= \sum_s \mathrm{Cov}(Y_s, Y_{\sum}), \end{aligned} \tag{4.1}$$

where $\mathbf{1}$ denotes the vector of all ones of matching length.

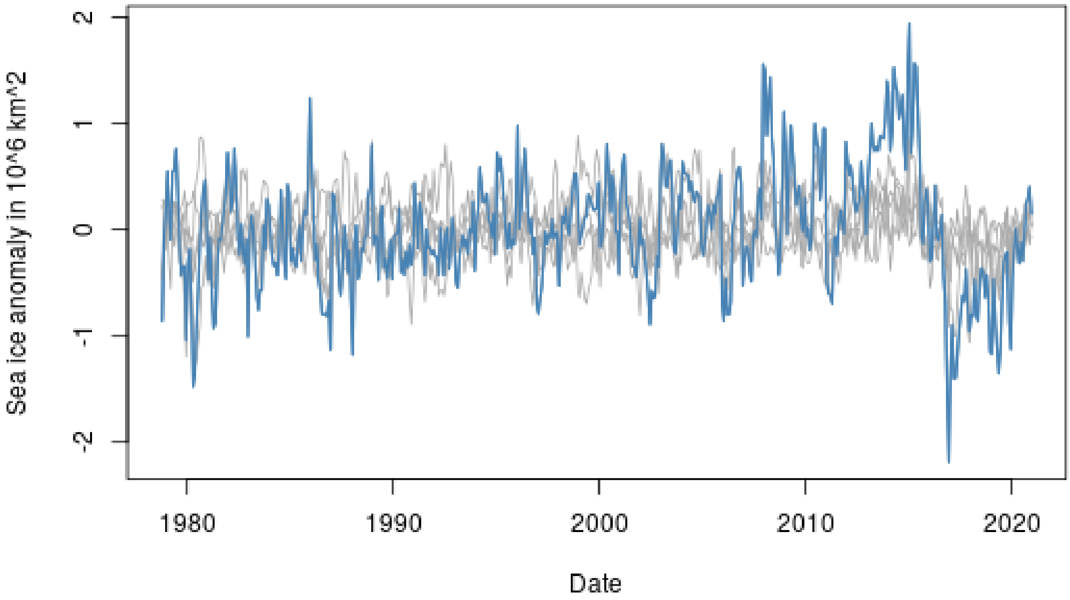


Figure 4.1: Sea ice extent anomaly in million km^2 by sector (gray) and in total (blue) over the course of the observation period 1979-2020. Note the larger magnitude of total sea ice variability due to positively correlated sectoral sea ice anomalies.

4.2.2 Predictor Variables

As predictor variables in our model for Antarctic sea ice extent we use the data compiled by (FJG16) from the University Corporation for Atmospheric Research data archive dataset ds570.0 (`https://rda.ucar.edu/datasets/ds570.0/#!description`) containing observations from 78 weather stations in the Southern hemisphere (see Figure 4.2) with a long measurement record for temperature (TMP) and sea level pressure (SLP). These stations cluster heavily on Australia, South America, South Africa, and the Antarctic peninsula, with some island stations and stations on the Antarctic mainland.

Additionally, nine climate index reconstructions are collected in this data set. Note that during our reconstructions period these indices are reconstructions themselves, based on

Table 4.1: Covariance matrix of sea ice extent anomaly in mio. km^2 by sector and in total (top) and covariance matrix of error in predicted sea ice extent in mio. km^2 by sector and in total (bottom). Reported numbers are multiplied by a factor 10 to increase readability.

	Ross Sea	B.A. Sea	Weddell Sea	King H. VII	East Ant.	Total
Ross Sea	1.18	-0.11	-0.22	0.21	0.09	1.15
B.A. Sea	-0.11	0.29	-0.08	-0.13	-0.04	-0.08
Weddell Sea	-0.22	-0.08	0.93	-0.02	0.08	0.68
King H. VII	0.21	-0.13	-0.02	0.85	0.00	0.92
East Ant.	0.09	-0.04	0.08	0.00	0.32	0.44
Total	1.15	-0.08	0.68	0.92	0.44	3.12

	Ross Sea	B.A. Sea	Weddell Sea	King H. VII	East Ant.	Total
Ross Sea	1.06	-0.09	0.05	0.09	0.11	1.23
B.A. Sea	-0.09	0.23	-0.06	-0.06	-0.03	-0.01
Weddell Sea	0.05	-0.06	0.56	0.12	0.05	0.73
King H. VII	0.09	-0.06	0.12	0.76	0.03	0.94
East Ant.	0.11	-0.03	0.05	0.03	0.28	0.45
Total	1.23	-0.01	0.73	0.94	0.45	3.33

measurements of the exact same weather stations already included as predictor variables. This means, that these indices are (non-)linear transformations of the other covariates. Our model handles this potential collinearity effortlessly through harsh penalization of highly correlated covariates in the form of horseshoe priors. A “double-counting” of stations does occur if a station is included as a predictor, and then again as part of a climate index based on its measurements, but this is not a problem for our model. A more severe form of “double-counting” occurs through geographical clustering of weather stations, in particular in Australia and South America. Geographically dense clusters of weather stations mean that the weather in this area is offered to the model multiple times without any added

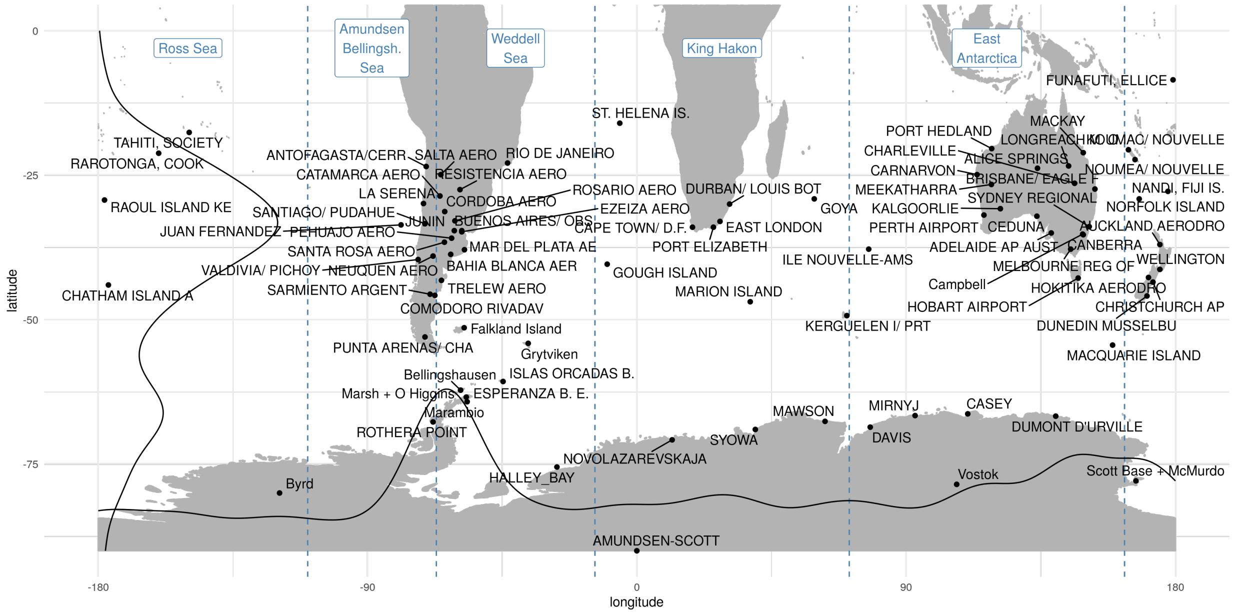


Figure 4.2: Location of weather stations collected by (FJG16) and sea ice sectors defined in (RH14). A density estimate of the station's distribution over longitude and latitude is shown along the axes.

predictive value. To mitigate these effects we tried the use of (spatially weighted) empirical orthogonal functions (EOF) of the weather stations as predictors instead of the raw data, but found that our horseshoe priors are more effective at identifying predictive covariates and filtering out duplicate information.

The nine climate indices included are the northwest and southwest Pacific from the NOAA Extended Reconstructed Sea Surface Temperature dataset, Version 5 (ERSST v5; HBF15), the Pacific Decadal Oscillation (PDO; MH02), four indices calculated from Niño SST from ERSSTv5 namely, the Niño 1 + 2 SSTs (averaged over 0° - 10° S, 270°E - 280°E), the Niño 3 SSTs (5°N - 5°S, 210°E - 270°E), the Niño 3.4 SSTs (5°N - 5°S, 190°E - 240°E) and the Niño 4 SSTs (5°N - 5°S, 160°E - 210°E), the Southern Oscillation Index (SOI; Aus22), the Atlantic Multidecadal Oscillation (AMO; EMT01), the Southern Annular Mode index (SAM; FM20) merging the "Fogt" seasonal SAM index reconstructions (JFW09; FPM09) which extend back to at least 1905 (`http://polarmet.osu.edu/ACD/sam/sam_recon.html`) with the observation-based SAM index (Mar03) after 1957 (`http://www.nerc-bas.ac.uk/icd/gjma/sam.html`). All indices are available on a monthly scale, except for the SAM index which is only available seasonally and downsampled to monthly intervals by smooth interpolation. These indices describe large scale atmospheric and oceanic circulation patterns in the Southern Hemisphere that are known to influence Antarctic sea ice. Seven out of nine indices describe ocean variability, as the weather stations providing our other predictor data are land-based and geographically sparse.

4.3 Creating Spatially Representative Station-level Weather Information

4.3.1 Introduction

We use monthly averages of temperature and sea level pressure measurements from weather stations in the Southern hemisphere to predict Antarctic sea ice extent. Every single stations measures the weather at its location, but for sea ice prediction large scale weather pattern in the Southern hemisphere are more useful than local weather. During the model exploration phase we tried using (spatially representative) principal components instead as predictor variables instead of the original station-level data. This approach led to no significant improvements in sea ice extent prediction in comparison to using station-level data once horseshoe priors were used for variable selection and was thus abandoned. It is still included in this dissertation because the spatially weighted principal component analysis of the weather station data has merit on its own.

Principal component analysis (PCA) is a standard method of identifying and quantifying the major modes of variability which corresponds to geographically clustered weather patterns in our data set, see Section 4.3.2. A drawback of the principal component analysis is that large clusters of stations, for example in Australia and South America, dominate the first principal components, to the point where principal components explaining unstructured patterns in Australia are perceived as more important (i.e., more variance explained) than actual patterns in isolated stations on islands in the Southern Ocean. The higher the number of measurement variables we have in one region, the higher the contribution of these variables in the total variance of the data set. PCA will then try to explain these variables as much as possible, while neglecting individual isolated stations with completely different patterns, as these stations only contribute a small fraction of the total variance in the data set. In order to create a spatially representative PCA, geographically clustered stations need to be downweighted and geographically isolated stations upweighted. Computing the PCA on spatio-temporally interpolated data (Section 4.3.3) achieves this reweighting of the variables to create spatially representative principal components (Section 4.3.4). The predictive power of the spatio-temporally interpolated data and its principal components is comparable to the original data and its principal components, see Section 4.3.5.

4.3.2 PCA of Station-level Weather Data

We use the monthly temperature (TMP) and sea level pressure (SLP) compiled by (FSR22), containing data from $p = 87$ weather stations in the southern hemisphere for $n = 1464$ months (January 1899 to December 2020). The observations are spatially sparse, i.e., every station only measures at its specific locations and all points between stations are unobserved. The resulting data set of measurements

$$X = (X_{\text{TMP}}, X_{\text{SLP}}); X \in \mathbb{R}^{n \times 2p}, X_{\text{TMP}}, X_{\text{SLP}} \in \mathbb{R}^{n \times p}$$

is projected into orthogonal space using principal component analysis (PCA)

$$XW = T; W \in \mathbb{R}^{2p \times 2p}, T \in \mathbb{R}^{n \times 2p}$$

where the projection matrix W contains the weights of the principal components and T the orthogonalized data with columns corresponding to the loadings of the principal components

(Cre15). The variance explained by each principal components are the diagonal elements of the covariance matrix,

$$\text{diag}(\text{Var}(T)) \in \mathbb{R}^{2p \times 1},$$

and the proportion of variance explained by each principal component its variance divdied by the total variance in the data set

$$\frac{\text{diag}(\text{Var}(T))}{\sum \text{diag}(\text{Var}(T))} \in \mathbb{R}^{2p \times 1}.$$

A dimensionally reduced version $\tilde{T} \in \mathbb{R}^{n \times l}$ of the original data $X \in \mathbb{R}^{n \times 2p}$ can be achieved by only using the first l principal components,

$$\tilde{T} = X\tilde{W}; \tilde{W} \in \mathbb{R}^{2p \times l}, T \in \mathbb{R}^{n \times l}, l \ll 2p.$$

As expected, we find that principal components load on spatial clusters of stations, i.e., stations on the same continents or coasts of continents. Such clusters of weather stations appear in South America and Australia, see Figure 4.2.

4.3.3 Spatio-temporal Interpolation of Weather Data

In order to create a spatially representative data set spatio-temporal kriging (Cre15) is used to interpolate the temperature data and sea level pressure data separately on a complete spatio-temporal grid using the R-package `gstat` (GPH16). The empirical variogram and fitted spatio-temporal variogram model for temperature and sea level pressure are shown in Figure 4.3, where both variables show strong spatial autocorrelation and weaker temporal autocorrelation. A separable spatio-temporal variogram model, a model fitting a separate variogram for space and time each and then combining them, is employed to model the spatio-temporal dependence in the data, allowing the prediction of unobserved locations. The station-level data is interpolated on a spatial grid of 20° longitude by 15° latitude on the southern hemisphere of size $m = 180/20 \cdot 90/15 = 54$, with the original n months being used to create a spatio-temporal grid of dimensionality $n \times m$. Interpolation creates the spatially

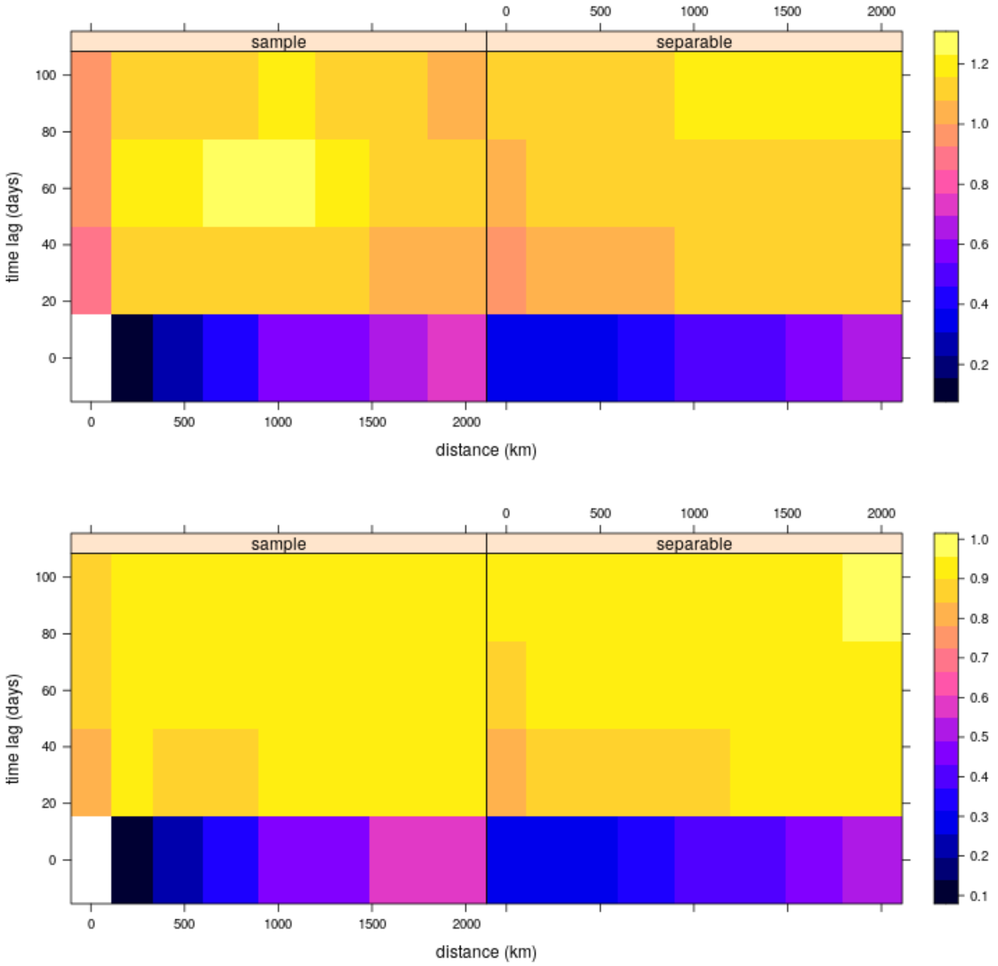


Figure 4.3: Empirical variogram (left) and fitted variogram model (right) for temperature (top) and sea level pressure (bottom).

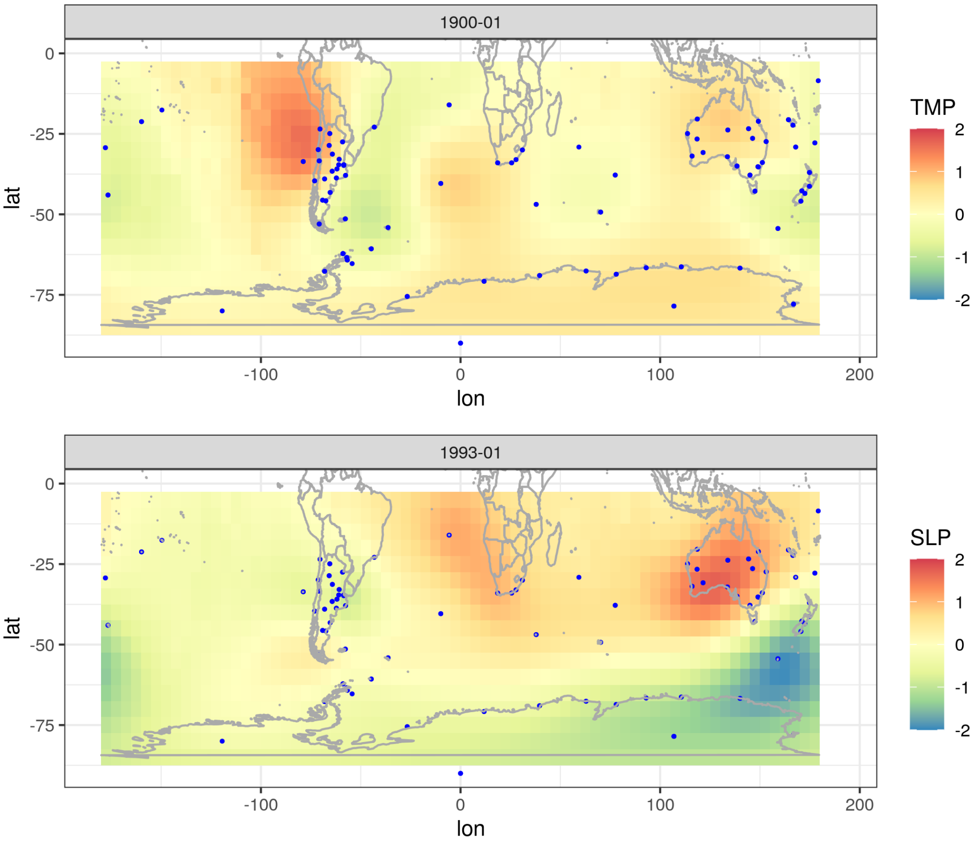


Figure 4.4: Interpolated temperature (top) and sea level pressure (bottom) anomaly for January 1993. Station locations are marked by blue dots, continent outlines are drawn in gray for reference.

representative data Y containing two separate data sets Y_{TMP} and Y_{SLP} for temperature and sea level pressure respectively that are interpolated to the same spatio-temporal grid,

$$Y = (Y_{\text{TMP}}, Y_{\text{SLP}}); Y \in \mathbb{R}^{n\times 2m}, Y_{\text{TMP}}, Y_{\text{SLP}} \in \mathbb{R}^{n\times m},$$

see Figure 4.4 for the interpolated temperature and sea level pressure anomalies in January 1993.

4.3.4 PCA of Spatio-temporally Interpolated Weather Data

The spatially interpolated data Y can also be decomposed in a PCA analogously to the original station level data X, either in two separate PCAs for temperature and sea level pressure or in one combined PCA. These spatial principal components are often called empirical orthogonal functions (EOF) in the literature.

4.3.4.1 PC Loadings on interpolated weather data

The loadings of the first ten principal components of temperature anomaly and sea level pressure anomaly are shown in Figures 4.5 and 4.6. The loadings tend to cluster geographically, i.e., there are large neighboring areas of high positive or high negative loadings. An interesting finding is that the major principal components tend to focus on a contrast between two geographic areas by having a strong positive loading on one area and an equally strong negative loading on a different area. This means that these principal components will be of large magnitude if one of the areas has a large positive anomaly and the other area a large negative anomaly. The loadings of a joint PCA cluster geographically as well and tend to have opposite loadings on sea level pressure and temperature in the same geographical area, see Figure 4.7, due to the negative correlation of temperature and sea level pressure.

4.3.4.2 Reconstruction of interpolated weather data from station-level PCA

In general, the proportion of variance explained by a principal component in the data set it was computed from is used as an interpretable measure of its importance. The first principal components is computed as to maximize its variance explained in the data set it is computed from, with subsequent principal components explaining decreasing proportions of the total variance. In order to assess the difference between the PCA on the station-level PCA and the PCA on the interpolated data, the proportion of variance explained by the original principal components on the interpolated data is assessed. The variance explained in the interpolated

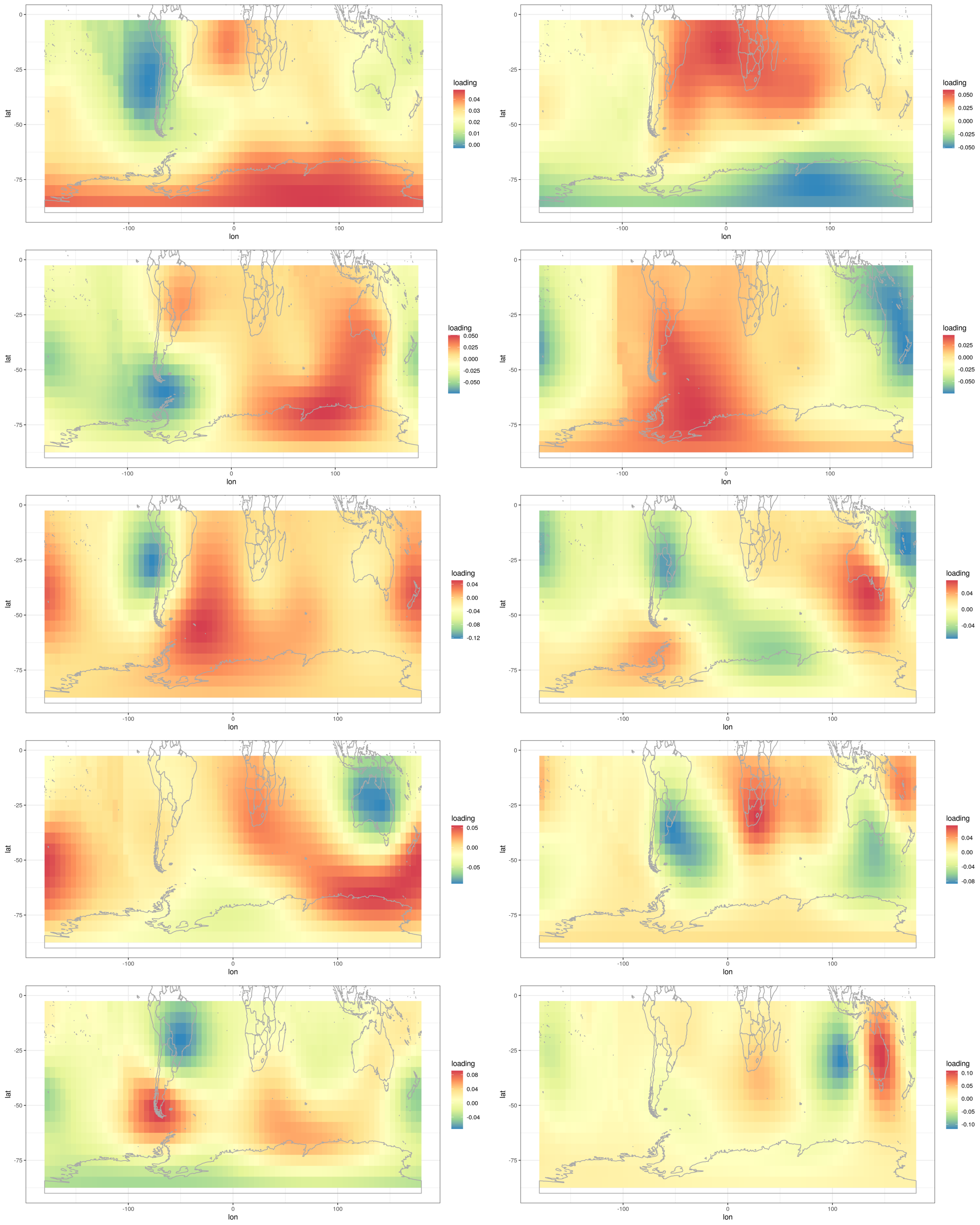


Figure 4.5: Loadings of first ten principal components (top left to bottom right row-wise) of interpolated temperature data.

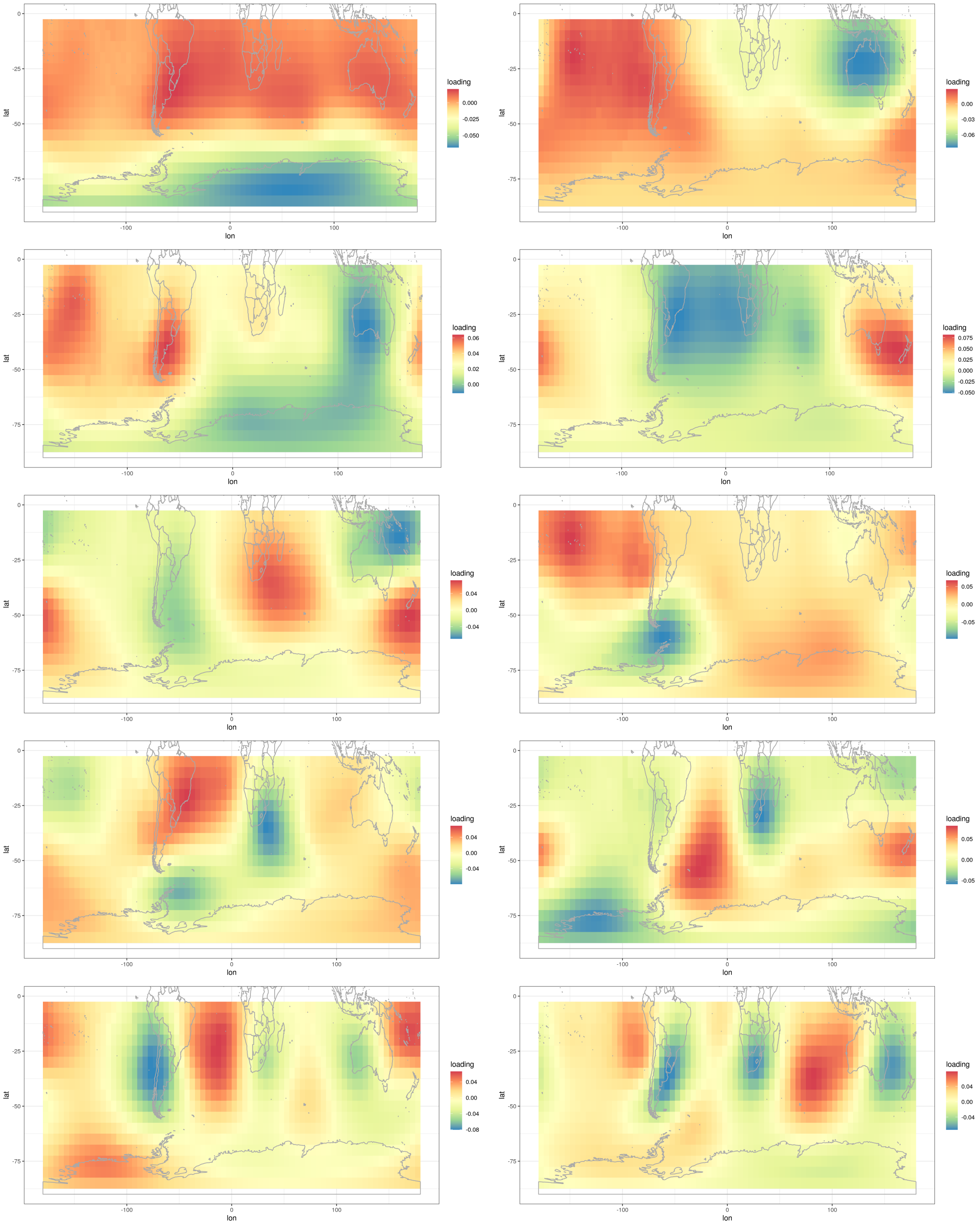


Figure 4.6: Loadings of first ten principal components (top left to bottom right row-wise) of interpolated sea level pressure data.

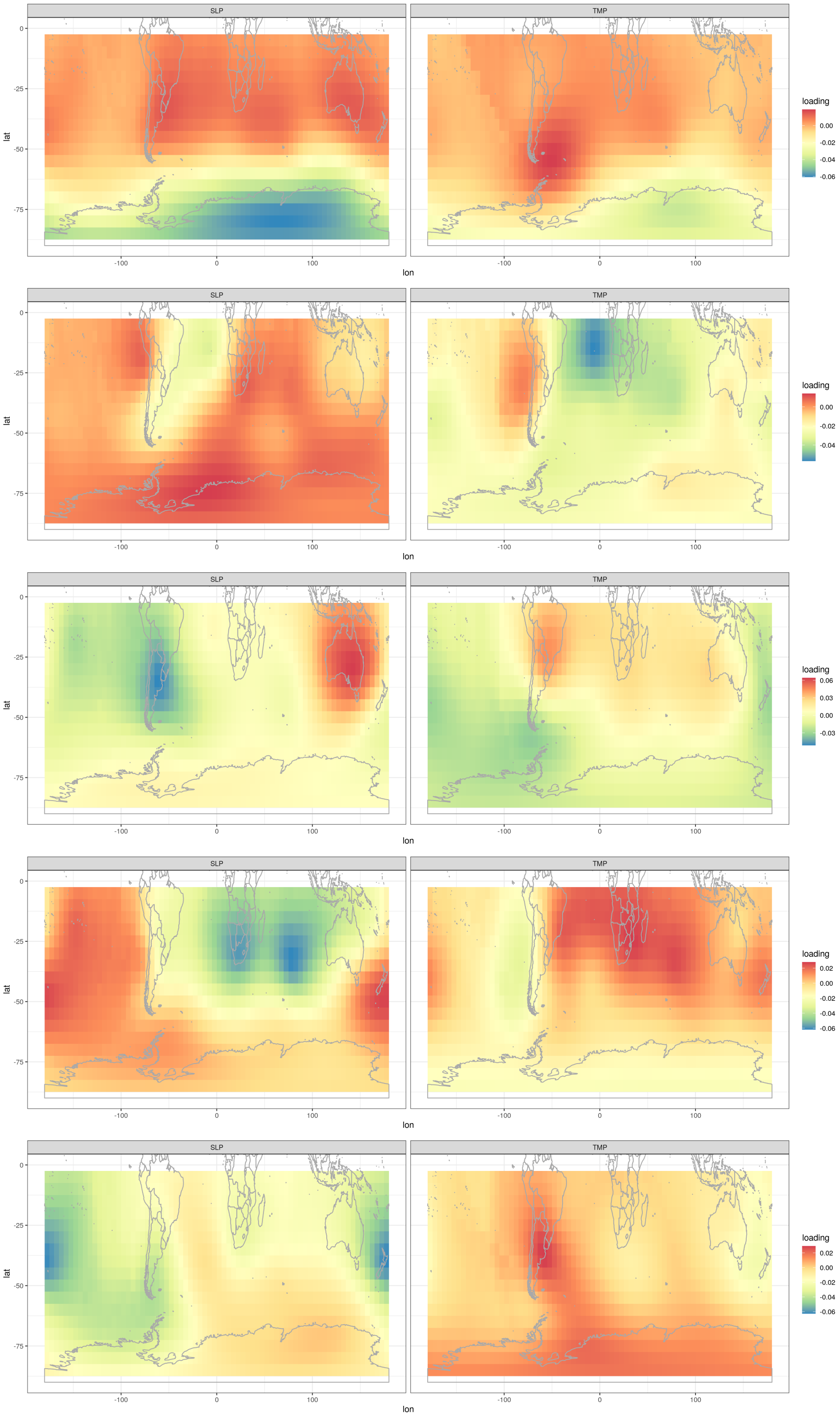


Figure 4.7: Loadings of first five principal components (top to bottom) of interpolated temperature (left) and sea level pressure (right) data.

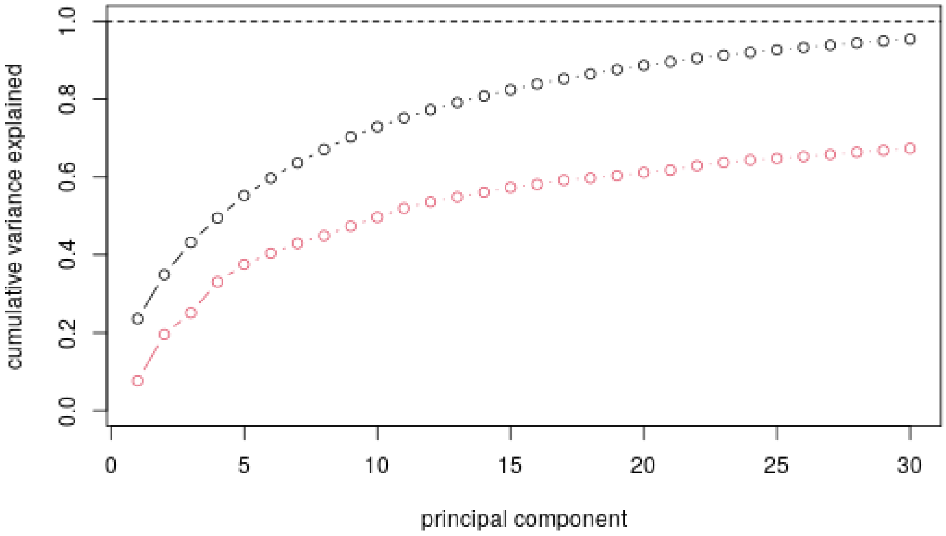


Figure 4.8: Cumulative variance explained of interpolated data by principal components of interpolated data (black) and station-level data (red). As the number of principal components nears the number of covariates (but usually much earlier) they explain 100% of the variance in the data set, see dashed line at 1.

data Y by the ith principal component of the station-level data X can be computed based on the variance explained in the multivariate regression model

$$\underbrace{T_{\cdot i}}_{n\times 1}\underbrace{\beta}_{1\times d}+\underbrace{\varepsilon}_{n\times d}=\underbrace{Y}_{n\times d},$$

where the variance explained by the ith principal component is $\mathrm{diag}(\mathrm{Var}(T_{\cdot i}\beta))$, which also equates to the total variance minus the unexplained variance $\mathrm{diag}(\mathrm{Var}(Y)) - \mathrm{diag}(\mathrm{Var}(\varepsilon))$. The cumulative variance explained by the first i principal components of the station-level data in the interpolated data is shown in Figure 4.8 and compared to the cumulative variance explained by the principal components of the interpolated data.

4.3.5 Predictive performance of interpolated data

The spatially representative interpolated data performs similar in comparison with the station level data in an exploratory analysis predicting sea ice extent by sector. We use a random forest model (Bre01) in an 8 fold cross validation to predict sea ice by sector. As a performance measure, the proportion of improvement in root mean squared error (RMSE) of the predicted values over predicting the overall mean (zero) is used, which can be interpreted as proportion of explained variability. A value of 0.1 indicates a 10% improvement over predicting the global mean. Figure 4.9 shows the performance of the station-level data (top), interpolated data (center), and principal components of interpolated temperature and sea level pressure (bottom). Note that the boxplots shown suffer from small sample artifacts; an 8 fold cross validation implies that only 8 points of data are summarized in every boxplot, leading to some points being classified as outliers or minima / maxima somewhat randomly. Lags of one, two, three, four, and twelve months were included for all variables in all models, leading to a major improvement in predictive performance across the board. Model performance without lagged variables is not reported. The station-level data outperforms the interpolated data by a small but consistent margin across all sectors. A principal component transformation of the interpolated data has a negligible effect on predictive performance of the random forest model, while dramatically reducing the dimensionality of the data. Note that using the joint principal component transformation of temperature and sea level pressure as shown in Figure 4.7 resulted in dramatically reduced predictive power (results not shown). This might be due to the combined PCA loading primarily on sea level pressure patterns, but temperature being a similarly important predictor of sea ice.

4.4 Modeling Framework

Predicting Antarctic sea ice extent in five sectors using temperature and sea level pressure measurements from weather stations in the Southern Hemisphere is a multivariate time

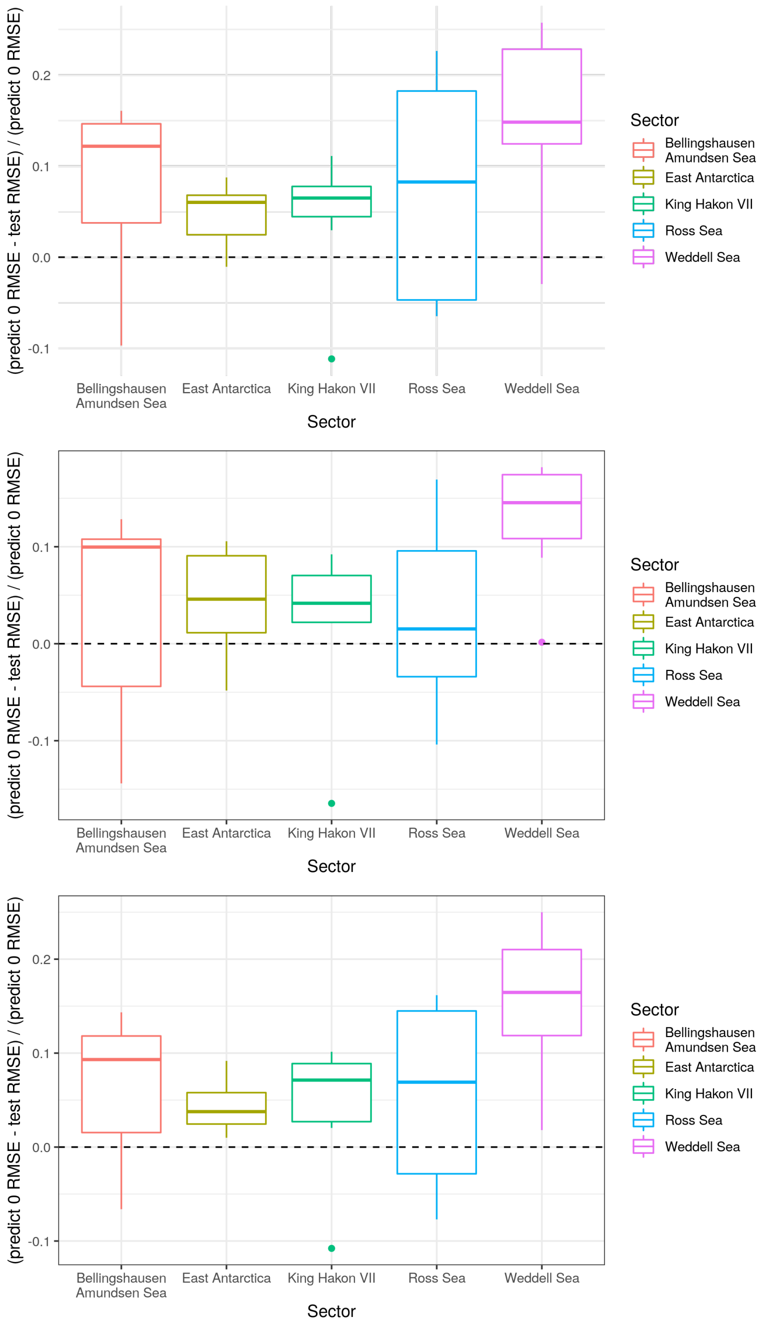


Figure 4.9: Performance of station-level data (top), interpolated data (center), and principal components of interpolated data (bottom), in an 8 fold cross validation of 5 year blocks. Note the different scale of the y-axes.

series modeling task with multiple predictor variables. As a statistical framework for this task consider the following setup. Let $Y_t(s)$ be the value of a spatio-temporal process at location $s \in \{s_1, \dots, s_{N_Y}\}$ and time $t \in 1, \dots, T$, with a N_X dimensional observation of a spatio-temporal covariate process $X_t(s)$. Here, $Y_t(s)$ is the monthly sea ice extent at time $t \in \{$January 1900, February 1900, ..., December 2020$\}$ for sector $s \in \{$King Haakon VII, Ross Sea, East Antarctica, Weddell Sea, Bellingshausen Amundsen Sea$\}$. The sea ice is thus a 5 dimensional time series and we adopt the notation $Y_{s,t}$ for sector s at time t or Y_t for short. We fit our model on monthly standardized anomalies due to the strong annual cycle in Antarctic sea ice. For each month across all years we compute the average sea ice extent and its standard deviation, which we use to standardize the observations in our data set for the purpose of modeling. For the purposes of reporting results and evaluating prediction accuracy we re-transform our reconstructions to sea ice extent anomaly in million km^2 by multiplying the model predictions with the monthly standard deviations. Our model is seasonally varying but does not allow coefficients to change over the years. This implies the common assumption that the relationship between the atmosphere as measured by temperature and sea level pressure and the Antarctic sea ice extent is stable over time.

Fitting our model in a Bayesian framework is an elegant choice for our application. For estimating the model parameters, it naturally allows us to incorporate prior knowledge of model parameters in the form of prior distributions. Most prominently, we know that the sea ice only has a weak relationship with measurements from each individual station in the Southern hemisphere and that the sea ice in many sectors is totally unrelated to weather measured by geographically distant stations in other sectors. Additionally, a Bayesian framework naturally incorporates model uncertainty, parameter uncertainty, and unexplained variability in its predictions, in our case the reconstructed sea ice. A last but maybe most important argument for the Bayesian framework is that it automatically provides an ensemble of reconstructions, i.e., many plausible reconstructions for all sectors over the entire time period. These ensembles incorporate all sources of uncertainty and can readily be used in further analyses. Each individual reconstruction has a plausible month-to-month progression and overall trend for the sea ice in each sector and incorporates credible relationships between the reconstructions in each sector.

The model we propose for Antarctic sea ice reconstruction is a seasonal VARMA (vector auto regressive moving average) model fit in a Bayesian framework using regularized horseshoe priors on the regression coefficients. The model needs to be a **Vector** ARMA model due to the target variable, the sea ice in all five sectors, being a multivariate vector. Modeling the five sectors independently would neglect the correlation between them, leading to implausible joint behavior of the reconstructions and inaccurate variability in the reconstructed total sea ice extent as the sum of all sectors. We chose a Vector **Auto-Regressive Moving-Average** model because the target variable is a time series. Time series in general, and sea ice extent is no exception, tend to be autocorrelated, meaning consecutive measurements in the time series are correlated with each other, and consecutive residuals in time series models are correlated as well. Explicitly modeling this self-dependency will result in reconstructions with plausible month-to-month progressions and plausible deviations from overall trends. Putting a regularized horseshoe prior (PV17) on regression coefficients is a state of the art method to ensure identifiability and sparsity in Bayesian regression models. The regularized horseshoe prior shrinks most coefficients either to zero (no effect) or leaves them almost unpenalized (almost full effect). This is plausible in this scenario where we expect measurements of most weather stations to have no association whatsoever with the sea ice extent in a specific sector, but some stations might have reasonably strong predictive power.

Before introducing our model in Section 4.4.5, we first review its individual modeling components, the multivariate regression model (Section 4.4.1), the VARMA model (Section 4.4.2), the seasonal regression model (Section 4.4.3), and the regularized horseshoe prior (Section 4.4.4).

4.4.1 Multivariate Regression Model

Recall the multivariate regression model

$$Y_t = HX_t + \varepsilon_t \tag{4.2}$$

at time t with multivariate target variable is $Y_t \in \mathbb{R}^{N_Y \times 1}$, multivariate predictor variables $X_t \in \mathbb{R}^{N_X \times 1}$, coefficient matrix $H \in \mathbb{R}^{N_Y \times N_X}$ and error terms ε_t which follow a multivariate normal distribution centered around zero

$$\varepsilon_t \sim N_{N_Y}(0, \Sigma), \forall t \tag{4.3}$$

with positive semidefinite covariance matrix $\Sigma \in \mathbb{R}^{N_Y \times N_Y}$.

4.4.2 Vector Autoregressive Moving Average Model

Building upon the multivariate regression model, recall the VARMA(p, q) model with p autoregressive terms and q moving-average terms of the form

$$Y_t = c + \sum_{i=1}^{p} A_{i,\cdot} Y_{t-i} + \sum_{j=1}^{q} B_{j,\cdot} \varepsilon_{t-j} + \varepsilon_t, \tag{4.4}$$

with $A_{i\cdot}$ denoting the ith row of the autoregressive coefficient matrix $A \in \mathbb{R}^{p \times N_Y}$ and $B_{i\cdot}$ denoting the ith row of the moving average coefficient matrix $B \in \mathbb{R}^{q \times N_y}$.

4.4.3 Seasonally Varying Model

Recall the univariate seasonally varying regression model for monthly data

$$Y_t = f\left(\text{month}(t)\right) X_t + \varepsilon_t, \tag{4.5}$$

where $f : [1, 12] \rightarrow \mathbb{R}$ is a smooth cyclical function describing the regression coefficient of $X_t \in \mathbb{R}$ on $Y_t \in \mathbb{R}$ for every month of year month$(t) \in [1, 12]$.

Due to the strong seasonality of the Antarctic climate system we extend the VARMA(p, q) model to allow seasonally varying autoregressive and moving-average terms. For every month

$k = \text{month}(t) \in [1, 12]$ and lag $i = 1, \dots, p$ and sector s we estimate the autoregressive coefficient $A_{l,s,k}$. Using the same generalisation on the moving average coefficients, we can formulate a seasonally varying VARMA model as

$$Y_t = c + \sum_{i=1}^{p} A_{i,\cdot,\text{month}(t)} Y_{t-i} + \sum_{j=1}^{q} B_{j,\cdot,\text{month}(t)} \varepsilon_{t-j} + \varepsilon_t. \tag{4.6}$$

4.4.4 Regularized Horseshoe Prior

The regularized horseshoe prior (PV17) is a useful tool to enforce sparsity in the coefficients β of a regression model of the form

$$Y_t = \beta X_t + \varepsilon_t, t = 1, \dots, T, \tag{4.7}$$

where $Y_t \in \mathbb{R}, \beta \in \mathbb{R}^{N_X}, X_t \in \mathbb{R}^{N_X}$, and $\varepsilon_t \in \mathbb{R}$. Model sparsity is particularly important in the case when we have fewer observations than covariates $T < N_X$, or when many covariates can be assumed to have no impact ($\beta_m = 0$) on the target variable. Both conditions are met in our case. The regularized horseshoe is an extension of the horseshoe prior (CPS10) that puts a prior on the regression coefficients $\beta_m, m = 1, \dots, N_X$, shrinking all β towards zero and setting some of them to be exactly equal to zero. The unregularized horseshoe is of the form

$$\beta_m \sim N(0, \tau^2 \lambda_m^2), \tag{4.8}$$

$$\lambda_m \sim \text{Half-Cauchy}(0, 1) = \text{Half-Student } \mathrm{t}_1(0, 1), \tag{4.9}$$

with local horseshoe parameter λ_m controlling the local shrinkage and global horseshoe parameter τ controlling the overall shrinkage. While the global horseshoe parameter τ pulls all coefficients towards zero for small values of τ, the thick Cauchy-tails of the local horseshoe parameters λ_m allow some coefficients β_m to be almost unaffected. The regularized horseshoe is an extension of the horseshoe prior containing an additional regularization parameter c controlling the shrinkage of large parameters that would be unaffected by the unregularized horseshoe prior. It is of the form

$$\begin{aligned}
\beta_m &\sim N(0, \tau^2 \tilde{\lambda}_m^2), && (4.10)\\
\tilde{\lambda}_m^2 &= \frac{c^2 \lambda_m^2}{c^2 + \tau^2 \lambda_m^2}, && (4.11)\\
\lambda_m &\sim \text{Half-Cauchy}(0,1) = \text{Half-Student } \mathrm{t}_1(0,1), && (4.12)
\end{aligned}$$

with regularized horseshoe parameter $\tilde{\lambda}_m$ controlling the shrinkage via the local horseshoe parameter λ_m controlling the local shrinkage, global horseshoe parameter τ controlling the overall shrinkage, and regularization parameter c controlling the shrinkage for $\lambda_m \to 0$ when the unregularized horseshoe would not shrink the parameter at all. For small coefficients β_m with strong shrinkage $\tau^2 \lambda_m^2 \ll c^2$, the regularized horseshoe parameter converges to the unregularized horseshoe parameter $\tilde{\lambda}_m \to \lambda_m$. For large coefficients β_m with weak shrinkage $\tau^2 \lambda_m^2 \gg c^2$, the regularized horseshoe parameter $\tilde{\lambda}_m^2 \to c^2/\tau^2$ and the resulting prior distribution converges to

$$\beta_m \sim N(0, c^2), , \qquad (4.13)$$

making c^2 the shrinkage parameter of an otherwise unregularized coefficient. Our implementation follows the alternative parametrization proposed in (PHS14). A Half-Cauchy prior is put on the global shrinkage parameter τ

$$\tau \sim \text{Half-Cauchy}(0, \tau_0), \qquad (4.14)$$

where the scale parameter τ_0 determines the overall shrinkage, i.e., what proportion of coefficients β will be set to zero. The regularization parameter c^2 is chosen according to an inverse gamma distribution

$$c^2 \sim \text{Inv. Gamma}(0.5\nu, 0.5\nu). \qquad (4.15)$$

which translates to a Student $t_\nu(0,1)$ distribution with ν degrees of freedom for coefficients β_m far from zero, because the inverse gamma distribution as a prior on an unknown variance (Equation 4.15) in a Normal distribution (Equation 4.13) results in a Student's t distribution.

4.4.5 Complete Model

In this section we assemble our full model, a seasonal VARMA model in a Bayesian framework using regularized horseshoe priors on the regression coefficients, from its parts presented in the previous Sections 4.4.1 to 4.4.4. For sea ice $Y_{s,t}$ in sector s at time t our model takes the form

$$Y_{s,t} = \sum_{j=1}^{N_X} f_k(\text{month}(t)) X_{k,t} + \sum_{i=1}^{p} A_{i,s,\text{month}(t)} Y_{s,t-i} + \sum_{j=1}^{q} B_{j,s,\text{month}(t)} \varepsilon_{s,t-j} + \varepsilon_{s,t}, \tag{4.16}$$

where the error terms ε_t at time t which follow a multivariate normal distribution centered around zero

$$\varepsilon_t \sim N_{N_Y}(0, \Sigma), \forall t, \tag{4.17}$$

with positive semidefinite covariance matrix $\Sigma \in \mathbb{R}^{N_Y \times N_Y}$.

The seasonally varying regression coefficients f_k are implemented as cyclical splines with four basis functions. A horseshoe prior is set on these splines coefficients that creates sparsity in the coefficients. The model used to create the reconstructions reported here does not have seasonally varying regression coefficients. Even though there is strong evidence for the relationship between atmosphere and Antarctic sea ice being seasonal, (e.g. HBR15), we find that seasonally varying regression coefficients are not warranted with the short observed record and the huge number of predictor variables to choose from. After accounting for seasonal variability through monthly standardization of the Antarctic sea ice extent, we found only insubstantial changes in our results when using seasonally varying regression varying coefficients at a dramatic computational burden.

Even though seasonally varying regression coefficients are not warranted in this application, we find that seasonally varying autoregressive coefficients are required. To ease the burden of computation we use strong priors to shrink the seasonally varying autoregressive terms towards their overall mean per lag and season $\tilde{A}_{l,s}$. For each lag l we specify a multivariate normal prior for the mean autoregressive coefficient across all sectors,

$$\tilde{A}_{l,\cdot} \sim N_{N_Y}\left(0, \begin{bmatrix} v_s & & 0 \\ & ... & \\ 0 & & v_s \end{bmatrix}\right), \tag{4.18}$$

with prior variance parameter v_s regulating the shrinkage. This specification allows us to borrow power across sectors, which is useful, as the autoregressive coefficients for each lag tend to be similar in all sectors. For the seasonally varying autoregressive terms we specify a multivariate normal prior of the form

$$A_{l,s,\cdot} \sim N_{12}\left(\tilde{A}_{l,s}, \begin{bmatrix} v & c & 0 & ... & 0 & c \\ c & v & c & 0 & ... & 0 \\ 0 & c & v & c & 0 & ... \\ ... & 0 & c & v & ... & 0 \\ 0 & ... & ... & ... & ... & c \\ c & 0 & ... & 0 & c & v \end{bmatrix}\right). \tag{4.19}$$

This prior specification implies shrinkage of the seasonal coefficients toward their overall mean $\tilde{A}_{l,s}$. The covariance matrix has the shrinkage parameter v on the diagonal regulating how much the seasonal coefficients can vary from their overall mean. On the off-diagonal and in the bottom left and top-right corner is the covariance parameter c which is used to enforce smoothness across neighboring months. Note that the first month January and the last month December are neighboring.

The final model has a huge number of parameters but the strong priors imposed on them result in stable parameter estimates and reconstructions. All model parameters are sampled using the Hamiltonian Monte Carlo algorithm implemented in the software package Stan (Sta22). We draw 2500 sets of parameters giving a good estimate of their joint posterior distribution. These 2500 parameter sets correspond to 2500 models and 2500 reconstructions, which are interpreted in the following section.

4.5 Results

This section reports our model's reconstructions of 20th century Antarctic sea ice and describes their trends and variability. Our reconstructions display the previously established higher total sea ice extent in the first half of the 20th century than the second half, and match earlier Antarctic ice extent estimates obtained from direct satellite observations before 1980 well. We analyze our model by examining which weather stations it selects for reconstructing sea ice in which sector. Reassuringly, the model tends to select weather stations within or close to each sector and tends to focus on weather stations to the far South. We validate our model's predictive power internally using cross validation. The section ends with a display of one of the key strengths of our ensemble of complete time series reconstruction, a systematic analysis of the 2014-2017 sea ice decline and the February 2022 minimum in the context of 20th century sea ice reconstructions.

4.5.1 Trends in 20th Century Antarctic Sea Ice Extent Reconstructions

Here we give a brief overview of the trends and variability in Antarctic sea ice extent revealed by the reconstructed data. Our ensemble of reconstructions is displayed in Figure 4.10 showing overplotted individual reconstructions (in black) and the overall reconstruction mean (in red) for all sectors and the total sea ice. The plots exhibit black vertical streaks in a pinstripe pattern due to the large difference in reconstruction uncertainty by month of the year. The reconstruction uncertainty is highest during the sea ice melt season, lower during the sea ice growth season, even lower close to the maximal sea ice, and lowest during the low sea ice season. This can be seen in detail in Figure 4.12, showing the ensemble reconstruction for February 1977-February 1978. The Ross Sea sector is the only sector exhibiting lower reconstructed sea ice levels for the beginning of the 20th century than the satellite observed period, but the reconstruction uncertainty is fairly large in this sector and almost always encompasses zero, the climatological mean. For the Bellingshausen Amundsen

Sea sector our reconstructions exhibit generally higher sea ice extent than in the satellite period, with a notable increase in sea ice from 1900 into the 1970s, preceding the decrease in sea ice observed in the satellite period. The reconstructed sea ice in the Weddell Sea sector continuously declines over the course of the 20th century, with comparatively minor subdecadal fluctuations. In the King Haakon VII and East Antarctica sectors there are not major trends in the reconstruction averages, but the reconstruction uncertainty is fairly high, making it impossible to rule out a positive or negative trend, or no trend at all. The trend in reconstructed total sea ice is the sum of all trends in all sectors, as the total sea ice is computed as the sum of sea ice in all sectors. The large trends in the Weddell Sea and Bellingshausen Amundsen Sea, and the smaller contrasting trend in the Ross Sea sum out to an overall higher level of reconstructed total Antarctic sea ice in comparison with the satellite observed period, and a sharp decline in Antarctic sea ice in the 1970s. It is apparent that the positive trend in satellite-observed SIE began around the time that satellites began recording sea ice observations and according to our reconstructions this was preceded by a negative trend. This outcome at the monthly level is consistent with the seasonal reconstruction reported on by (FSR22) who suggested that there was a regime shift in Antarctic sea ice in the 1960s and 1970s. To show the variability within and between individual reconstructions a set of six randomly picked individual reconstructions of the total sea ice are showcased in Figure 4.11. Each of these six individual reconstructions is an equally likely depiction of a possible evolution of Antarctic sea ice extent in the 20th century. Their variability matches the satellite record on small as well as large time scales so accurately, that the reconstruction period is indistinguishable from the satellite observed period by eye.

4.5.2 Comparison with Early Satellite Observations

For a few months in our reconstruction period direct satellite observations of the Southern Hemisphere exist that have been analyzed manually by (MGC13) and (GCM14) to obtain estimates of Antarctic sea ice. This is a difficult and labor intensive task, as these images are shot in the visible spectrum and the sea ice edge is obscured by a 40-90% cloud cover depending on season and location. In every satellite image, observed fractions of the sea ice edge are manually identified, grouped by month and then manually compiled into a complete sea ice edge. This process accrues multiple sources of uncertainty but nonetheless gives a reasonably good estimate of historic sea ice extent. Table 4.2 shows the mean and standard error of the estimated total sea ice anomaly for the satellite observed sea ice and our reconstruction. For September 1964 and May through July 1966 we find good agreement between our reconstruction and the satellite observed sea ice, with our reconstructed average always being closer to zero. This damper effect can be explained by the regularization scheme of our method which brings all regression coefficients and thus average predictions closer to zero in order to increase stability. Note that the variability of individual reconstructions is not dampened by our regularization scheme as error terms are unpenalized. This means that extreme events at the monthly level are created true to the internal variability. The reliability of the reconstruction for August 1966 created by (GCM14) has been questioned by the authors due to missing values in that month. A decrease in sea ice anomaly was deemed plausible in this month, but probably not at that magnitude.

As a quantitative analysis of the agreement between the satellite observed reconstruction and our reconstruction, we compute the Kullback-Leibler divergence (KL divergence; KL51), a distance metric between distributions p_1 and p_2, given as

$$\begin{aligned} D_{\mathrm{KL}}(p_1||p_2) &= \mathbb{E}_{p_1}\left[\log\frac{p_1(x)}{p_2(x)}\right] \\ &= \int_x \log\frac{p_1(x)}{p_2(x)} p_1(x)\partial x \end{aligned} \tag{4.20}$$

between the sea ice anomaly distributions produced by the satellite observed reconstruction and our reconstruction. For practical reasons, we assume both the satellite observed reconstruction and our reconstruction for a given month follow a normal distribution using the mean and standard deviation given in Table 4.2; not an unreasonable assumption since the distribution of the satellite observed reconstructions is unknown and our reconstructions follow the typical bell-shaped distribution seen in normal distributions. This distribution assumption of normality works in our favor by allowing the use of the closed form solution for the KL divergence for normally distributed random variables $X_1 \sim p_1 = N(\mu_1, \sigma_1^2)$, $X_2 \sim p_2 = N(\mu_2, \sigma_2^2)$ which can be derived to be

$$\mathrm{KL}(p_1||p_2) = \log \frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} - \frac{1}{2}. \tag{4.21}$$

Large values in KL divergence imply a larger difference in distributions, but in order to interpret the scale of these divergences, we use the statistical relation that $2 \times n_0$ times the KL divergence between a distribution with m_0 estimated parameters and the same distribution with fixed reference parameters approximately follows a Chi-square distribution with m_0 degrees of freedom (e.g., BA11),

$$2n_0 D_{\mathrm{KL}}(p_1||p_2) \overset{\text{approx.}}{\sim} \chi^2_{m_0}, \tag{4.22}$$

where n_0 is the number of samples on which the parameters were estimated. In our case, we compare a normal distribution with estimated mean and standard deviation ($m_0 = 2$) for one month ($n_0 = 1$) to a normal distribution with parameters given by (MGC13; GCM14), so 2×1 times the computed KL divergence approximately follows a Chi-square distribution with 2 degrees of freedom. This distribution approximation allows us the computation of p-values testing the null hypothesis “Our reconstruction follows the same distribution as the satellite observed reconstruction” vs. the alternative hypothesis “Our reconstruction does **not** follow the same distribution as the satellite observed reconstruction”. This confirms that only for August 1966 our reconstruction differs significantly from the satellite observed reconstructions created by (MGC13) and (GCM14).

4.5.3 Importance of Individual Weather Stations

Figure 4.13 shows the importance of individual stations in the sea ice reconstruction for each sector. For each sector, the importance of a station is measured as the sum of the

Table 4.2: Average (mean) and standard error (se) of estimated sea ice anomaly from direct satellite observation (sat.) and our reconstruction (rec.). We also report two times the Kullback Leibler divergence (2×KL) between the two distributions and its p-value (p val) in a Chi-square distribution with 2 degrees of freedom. *We compute the standard error by treating the range given in (MGC13) as a 95% confidence interval in a Normal distribution. **This estimate is seriously doubted by the authors due to missing observations.

Month	sat. mean	sat. se	rec. mean	rec. se	2×KL	p-val
September 1964	0.90	0.40*	0.75	0.56	0.41	0.82
May 1966	0.40	0.20	0.50	0.52	4.09	0.13
June 1966	1.00	0.40	0.50	0.53	1.76	0.41
July 1966	0.60	0.50	0.47	0.51	0.07	0.97
August 1966	-1.60 **	0.40 **	0.55	0.51	28.89	0.00

magnitudes of all regression coefficients relating to measurements of this station, i.e., the sum of the absolute values of the regression coefficients for the station's temperature and sea level pressure measurements and lags thereof. We notice that our model heavily relies on measurements from stations within the same sector for sea ice reconstruction, in particular for the Weddell Sea, King Haakon VII Sea, and East Antarctica. Note the lack of weather stations in the Ross Sea itself predicting Ross Sea sea ice extent due to the nonexistence of stations in this sector South of the Chatham Island station at 44° South, see Figure 4.2 showing the location of all stations in our data. This probably contributes to relatively low predictive power of our model in this sector. There is also a lack of stations in the Amundsen Bellingshausen Sea, with stations in this sector mainly on the South American mainland, see Figure 4.2, but our model still achieves a good reconstruction using nearby stations.

4.5.4 Predictive Performance

We perform a leave-one-out cross validation (LOOCV, e.g., Cel14) to estimate the predictive performance of our model. In a traditional LOOCV one observation is left out of the data set and predicted using a model fit on the remainder of the data. This process is repeated for every observation in the data resulting in a prediction associated with every observation. These predictions are then compared to the known true observation to obtain measures of model performance. Following (FSR22), we do not only leave out the data of the one month we are trying to predict, but also all data from its year and the two years before and after, resulting in 42-5=37 years of data for each model. This is necessary due to the autocorrelated nature of the data set and our model. This process creates a set of validation reconstructions, whose average prediction per month and sector is compared to the climatology. We also perform the same cross validation on the climatology, computing the climatological mean using only those years not left out.

For our average validation reconstruction and our climatology we compute the correlation, coefficient of efficiency (CE; CMS99), and root mean squared error (RMSE) with the true observed sea ice by sector, see Table 4.3. Correlations are bound between $[-1, 1]$ and **higher** values imply better performance, the CE is bound between $]-\inf, 1]$ and **higher** values imply better performance, the average error (AE) is bound between $[0, \inf[$ and **lower** values imply better performance, and the RMSE is bound between $[0, \inf[$ and **lower** values imply better performance. Note that "almost by design" the climatology has a negative correlation and CE with the ground truth. When computing the climatological mean on detrended data, the climatological mean is always exactly zero. When computing the climatological mean on detrended data where a few years are left out, it will shift in the opposite direction of the anomaly in the left out years. Our model's sea ice predictions for the Ross Sea, King Haakon VII Sea, and East Antarctica have a moderate correlation with the observed sea ice values, strong correlation for the Bellingshausen Amundsen Sea and Weddell Sea, but only a weak positive correlation with the total sea ice. The lack of a stronger correlation for the total sea ice is surprising, as the predicted total sea ice is computed as the sum of the predictions for each sector. Similarly, we find that our model outperforms the climatology in terms of RMSE in every sector, but not in total, see Table 4.3.

Table 4.3: Performance of our model and the climatology in the LOOCV. For every sector we report the correlation coefficient (cor), coefficient of efficiency (ce), mean absolute error (mae), and root mean squared error (rmse) for the cross validated climatology (cli) and our model's average reconstruction (mod).

Sector	cor_cli	cor_mod	ce_cli	ce_mod	mae_cli	mae_mod	rmse_cli	rmse_mod
Ross Sea	-0.51	0.34	-0.05	0.08	0.28	0.25	0.35	0.33
B.A. Sea	-0.26	0.57	-0.04	0.32	0.14	0.11	0.17	0.14
W. Sea	-0.17	0.65	-0.04	0.42	0.25	0.18	0.31	0.23
K.H. VII	-0.31	0.38	-0.05	0.13	0.22	0.21	0.30	0.27
E. Ant.	-0.06	0.30	-0.03	0.03	0.14	0.14	0.18	0.17
Total	-0.45	0.08	-0.05	-0.13	0.44	0.44	0.57	0.59

Table 4.4: Performance of our model and the climatology in the LOOCV from January 1979 to December 2013. This excludes the 2014-2017 decline in Antarctic sea ice.

Sector	cor_cli	cor_mod	ce_cli	ce_mod	ae_cli	ae_mod	rmse_cli	rmse_mod
Ross Sea	-0.50	0.50	-0.05	0.24	0.28	0.23	0.35	0.30
B.A. Sea	-0.27	0.55	-0.05	0.30	0.14	0.11	0.17	0.14
W. Sea	-0.15	0.68	-0.04	0.45	0.25	0.18	0.31	0.22
K.H. VII	-0.30	0.46	-0.05	0.20	0.22	0.20	0.29	0.26
E. Ant.	-0.06	0.36	-0.03	0.08	0.14	0.13	0.18	0.17
Total	-0.37	0.31	-0.05	0.05	0.38	0.36	0.48	0.46

To understand how our model can perform worse in total than in any of the sector consider the following. First, note that the RMSE is the root mean **squared** error, and that the errors our model makes in each sector tend to be positively correlated, see bottom of Tabke 4.1. As the sum of squared errors is smaller than the sum of errors squared for positively correlated errors, it is indeed possible for our model to outperform the climatology in each sector but not in total. Evidence for this theory can be found in Tables 4.1, showing the covariance matrix of the sea ice extent anomalies by sector and the covariance matrix of cross-validated prediction errors. The bottom of Tables 4.1 shows the covariance among the cross-validated prediction errors of all sectors to be positive. The only exception is the Bellingshausen Amundsen Sea, where the errors are negative correlation with errors in all other sectors. This means, that except for the Bellingshausen Amundsen Sea sector, our model tends to make the same kind of error simultaneously in all sectors, i.e., we tend to predict too much or too little sea ice in all sectors at the same time. This in turn implies that the total sea ice will be predicted even more incorrectly, as the errors in the sectors tend to sum up, instead of canceling each other out. This makes prediction errors in the total larger than in any individual sector. With the same statistical reasoning as for the covariance matrix reported in the top of Table 4.1, we can interpret the covariance between the errors by sector with the error in total sea ice (last row and last column of the bottom half of Table 4.1) as their contribution to the total error. Furthermore, as the scale of the sea ice anomalies underlying the top half of Table 4.1 and sea ice prediction error underlying the bottom half of Table 4.1 is the same, their numerical values can be compared directly. This reveals, that while the variances in prediction errors are smaller than the variances in sea ice anomalies in every sector, their contributions to (covariances with) the total prediction error are larger. This leads to a larger variance in prediction error for the total sea ice than variance in total sea ice anomaly.

Another explanation why our model performs worse in total than in any of the sectors can be concluded from Figure 4.14, showing our model's cross-validated performance over the satellite observed period 1979-2020. The predictions on holdout data in our LOOCV for every month in our data set are plotted, with the average reconstruction on holdout data in red, and the true observed sea ice in blue. The good numerical performance of our reconstruction model in the Bellingshausen-Amundsen Sea and Weddell Sea manifests itself in the good visual agreement between the average reconstruction (in red) and the true sea ice (in blue) in these sectors. Our model fails to replicate the 2014-2017 decline in total Antarctic sea ice and relevant sectors, with the observed total sea ice at the margin of our reconstruction uncertainty. Two main explanations come to mind, either the relationship between Antarctic sea ice and Southern Hemisphere weather has changed, or an unobserved driver not captured in our covariates has caused the dramatic decline. Our model's failure to replicate this dramatic sea ice decline highlights the fact that this occurrence is highly unusual. This sea ice decline from 2014-2017 is mainly driven by a decline in each sector, which manifests itself in a much more dramatic decline in total sea ice than in any sector. The years after 2014 contribute to our model's underperformance in comparison with the climatology, in particular in terms of squared errors as measured by the RMSE. Unsurprisingly, our model's predictive performance for the years 1979-2013, see Table 4.4, is substantially improved in comparison to the entire time period, and our model outperforms the climatology in terms of RMSE of total sea ice extent.

To check if the variability in our reconstructions is appropriate, we compute 95% credible intervals for our reconstructions of sea ice by sector and in total, see Figure 4.15. Note that we compute credible intervals for the actual sea ice extent anomaly, not the **average** sea ice extent anomaly. This corresponds to a 95% prediction interval in a frequentist framework, as opposed to a 95% confidence interval for the reconstructed mean. Over the satellite observed period, our 95% credible interval as estimated in the LOOCV cover 91% for the Ross Sea and 91% East Antarctica, 92% for the Weddell Sea, 93% for the Bellingshausen Amundsen sea, 87% of the observations for the total Antarctic sea ice extent. This means that our reconstruction uncertainty undercovers the true variability. When analyzing the temporal distribution of our undercoverage, we find that in particular in the 2010s are undercovered in all sectors and the total.

4.5.5 Extreme Events in Historic Context

Using a Bayesian framework for fitting our reconstruction model allows answering subtle questions in a straightforward fashion that might be difficult or impossible to answer in the more commonly used frequentist setting. The Bayesian framework creates a sample of complete Antarctic sea ice time series reconstructions for the entire 20th century. These can subsequently be analyzed to compute summaries like the average predicted sea ice extent and its uncertainty. The frequentist framework only creates an average reconstruction and its uncertainty, without ever creating actual time series of complete reconstructions, which is good enough to describe average trends, but cannot answer research questions pertaining to individual reconstructions and their variability. In particular the possibility to compare observed events to individual reconstructions is lost. One of the questions that can only be answered using an ensemble of realistic time series reconstructions is how extreme events in the satellite record like the 2014-2017 decline or the February 2022 and 2023 Antarctic sea ice minima compare to the unobserved sea ice in the early 20th century. Our framework allows the estimation of how likely a similar event might have happened, by checking in each of our 2500 reconstructions whether or not the event occurred in the reconstruction period, whereas a reconstruction of an average and its uncertainty cannot be compared to these events. The reconstructed average will never get close to any observed extreme event like the February 2022 and 2023 minima, and statements like "the February 2022 sea ice minimum is within 2 standard errors of our average reconstructed sea ice extent the following years in the reconstruction period ..." are unsatisfying. Extreme events like the 2014-2017 sea ice decline are impossible to detect in an average reconstruction with uncertainty estimates as the month-to-month progression is lost. Our Bayesian framework applies the same logic as the attribution of extreme weather and climate related events, see e.g., (SCO16) for an overview, using multiple runs of climate simulation models to estimate the probability of extreme events occurring. We define the event associated with the 2014-2017 decline in Antarctic sea ice as a difference of 4.13 million km^2 or more in a three year window. This is the difference between the largest positive anomaly of +1.94 million km^2 in January 2015 and the largest negative monthly anomaly of -2.19 million km^2 in December 2016 observed in this time period. We compute for each of our 2500 sea ice reconstructions the maximal difference in sea ice anomaly in a three year sliding window, see Figure 4.16, and find that only 11 out of 2500 reconstructions have a difference greater than the observed three year difference of 4.13 million km^2. We can think of this as a p-value of sorts: How likely was it for a three year difference in Antarctic sea ice anomalies of 4.13 million km^2 or more to have occurred between 1900 and 1979? About $11/2500 = 0.44\%$. Note that this answer is mainly driven by our estimates for the autocorrelation structure of the Antarctic sea ice time series, as the variability in the error terms is much larger than the variability in estimated mean, see Figure 4.10. The difference in reconstructed sea ice averages across all reconstructions are relatively small for any given time window of three year in comparison with the differences in one sea ice reconstruction within a three year time window. This shows that the correct estimation of temporal autocorrelation structures and their continuation in realistic individual reconstructions is key to answering these research questions.

For the second extreme event, the February 2022 Antarctic sea ice minimum of 2.22 million km^2, we check each of our reconstructions to see whether or not a sea ice extent of or below that value was predicted. We identify the lowest sea ice minimum for each reconstruction, see Figure 4.18, and find that $400/2500 = 16\%$ of all reconstructions have a lowest sea ice extent of 2.22 million km^2 or less, which always occurred in the month of February. Therefore, while uncommon, the reconstructions suggest that the SIE might have previously reached extremely low values of such a magnitude in the 20th century. We further analyze in which year the lowest February anomaly occured, and find the late 1970s in general and in particular the year 1977 to have unusually low February reconstructions, see Figure 4.17. The year 1977 has independently been identified as a year of low sea ice in the literature (e.g. ZPC83) Of the reconstructions where a February as extreme or more extreme than the February 2022 minim occurred, about 26.3% (105/400) are predicted to have happened in February of 1977. Such extrema were also found with slightly increased probabilities for February 1975 and 1976, but not nearly as likely as 1977. Other probable years are more spread out over the course of the reconstruction period and do not occur in noteworthy clusters.

The third extreme event we analyze is the recent February 2023 minimum in total sea ice extent of 2.04 mio. km^2. A February anomaly this low or lower occurs only in 108 / 2500 (4.4%) of our reconstructions. Again, the year 1977 is the most likely year where this anomaly might have happened, with 35 / 108 (32.4%) of our reconstructions with such a February anomaly predict it to have happened in this year.

4.6 Discussion

We present the first complete monthly time series reconstruction of Antarctic sea ice for the 20th century. We provide a set of 2500 reconstructions of Antarctic sea ice by sector and in total, with realistic month-to-month progression, sector-to-sector covariability, and total sea ice extent variability. This is achieved by the introduction of a seasonal VARMA model that we fit in a Bayesian framework. Our model draws its realism from formulating a joint time series model for all sectors (VARMA) which creates plausible month-to-month progression and sector-to-sector covariability resulting in plausible total sea ice extent variability. The Bayesian framework naturally creates and ensemble of reconstructions as posterior draws from the joint distribution of the unobserved sea ice extent from 1900-1978 for all sectors.

We validate our method internally using a LOOCV, showing predictive power particularly in the Bellingshausen-Amundsen Sea, Weddell Sea, and the Ross Sea, King Haakon VII Sea, and East Antarctica to a lesser degree. The overall sea ice is not reconstructed better than by a climatology, but our reconstructions contain realistic variability as well as spatio-temporal autocorrelation patterns not contained in the climatology. For many applications the use of a climatology will result in dramatic underestimation of uncertainty due to lack of variability in sea ice anomalies. An further reassurance of the validity of our model is the geographic distribution of weather stations deemed important for sea ice reconstruction by sector. Stations with high importance tend to cluster within or nearby the reconstructed sector. Our method is externally validated by high agreement with the sea ice reconstructions created for September 1964 (MGC13) and May-August 1966 (GCM14) based on direct satellite images. Our reconstructions are uniquely powerful by allowing a comparison of extreme events in the satellite record like the 2014-2017 decline or the February 2022 Antarctic sea ice minimum to the unobserved sea ice in the earlier 20th century. We can state that the 2014-2017 decline is likely unprecedented with only 11 out of 2500 reconstructions exhibiting a difference as great or greater than the observed three year difference of 4.13 million km^2 in three consecutive year. For the February 2022 sea ice extent minimum we find that $400/2500 = 16\%$ of all reconstructions have a lowest sea ice extent of 2.15 million km^2 or less, showing that that this was indeed an unusual event but one that might have happened previously in the 20th century. For the February 2023 sea ice minimum we find that only $105/2500 = 4.4\%$ of our reconstructions contain a February anomaly this low or lower.

The analyses of (De 09) indicate an average shift in circumpolar ice edge latitude around 2.41° (range $1.89° - 2.80°$) southwards between their early period (Austral Summer 1931/1932 to 1955/1956) and their late period (Austral Summer 1971/1972 to 1986/1987), a major decline in sea ice. Consistent with our reconstructions, their regional analyses show that the largest changes occurred in the South Atlantic (mostly covered by our Weddell Sea sector), but they also detect change across the Indian Ocean (our King Haakon VII Sea & East Antarctica sectors). Their smaller scale sea ice decline between their early and late season in the Ross Sea is found in our reconstructions. Our reconstructions d not show their substantial (but not significant with -value 0.856) sea ice decline in their Southern Pacific Ocean sector (corresponding to our Bellingshausen Amundsen Sea sector). Our reconstructions agree well with the ice core based reconstructions summarized in (TAE19), including their reported linear trends from 1900-1990 for. They also find a positive trend in the Ross Sea over the course of this time period, as well as a negative trend in the Weddell Sea. The find no (or a small negative) trend in the Indian Ocean sector, where we find not strong trends in our reconstruction for our corresponding King Haakon VII sector. Their finding of a negative trend for the Western Pacific Ocean is found in our small negative trend in our corresponding East Antarctica sector. To compare our reconstructions with (FSR22) we create seasonal reconstructions by averaging the reconstructions for each set of three months within each season (not shown). We compare our reconstructions for Austral winter (June, July and August) to the respective reconstructions in (FSR22) with overall high consistency. In the Ross Sea reconstructions, (FSR22) do not detect our overall small positive trend in sea ice, but sub-decadal increases and decreases match our reconstructions well. For the Weddell Sea, a sector where both our reconstruction and (FSR22) show a high degree of confidence, our reconstructions match exceptionally well. In the Bellingshausen-Amundsen Sea, King Haakon VII, and East Antarctica sectors our reconstruction and/or those reported in (FSR22) are of small magnitudes and well within each other's reconstruction uncertainty.

This work invites a wide array of new research opportunities. Dirichlet auto regressive moving average model as introduced in (KBW23) are another promising approach for modeling Antarctic sea ice extent in the framework of Bayesian time series analysis. Instead of creating a multivariate model for sea ice extent in all sectors and the total sea ice extent reconstructed as the sum of all sectors, a Dirichlet model reconstructs the total sea ice extent and then models its distribution across the sectors. We chose our multivariate approach over the Dirichlet approach because it is more intuitive and somewhat simpler in its implementation. Including a temporal and spatial autoregressive component as well as sparsity inducing priors, our model's statistical building blocks are probably close to optimal, more work can be done on hyper-parameter tuning and model selection. Our choice of which order of autoregressive parameters to include as well as the strength of our sparsity prior are chosen based on expert knowledge but might be improved using more rigorous statistical methods. Collecting more covariates, and more powerful covariates is certainly fruitful endeavor for future research that will probably have the biggest impact on reconstruction accuracy.

An exciting extension of our work made possible by its Bayesian formulation is the possibility of using prior information in the form of direct sea ice observations to improve reconstruction accuracy. In this application, we can think of prior distributions as containing outside information that can be used to increase the model precision. If we were to use the early satellite observations analyzed by (MGC13) and (GCM14) as prior information for sea ice extent, our model posterior reconstructions would be something like a weighted average with reduced standard error between the satellite based reconstruction and our current reconstruction. In addition to the early satellite observations analyzed by (MGC13) and (GCM14), a vast amount of information about early 20th century Antarctic sea ice remains hidden in ship logbooks (e.g., ED16). These contain notes of direct sea ice observations at a certain date and location. This information can be related to a (local) sea ice edge or (local) sea ice extent using a to-date nonexistent statistical model, which will result in prior information for (local) sea ice extent reconstruction. This method should obtain useful priors for the sea ice all the way back to the beginning of the 20th century. These priors "anker" the reconstructions and are expected to lead to dramatic reductions in reconstruction uncertainty. Preliminary analysis using the direct satellite observations reported in (MGC13) and (GCM14) show promising results in reducing reconstruction uncertainty in months with available prior information as well as neighboring months.

Our reconstructions can also be used to calibrate climate models; using our reconstructions as opposed to the climatology will ensure a plausible trend and variability in historic levels of Antarctic sea ice extent.

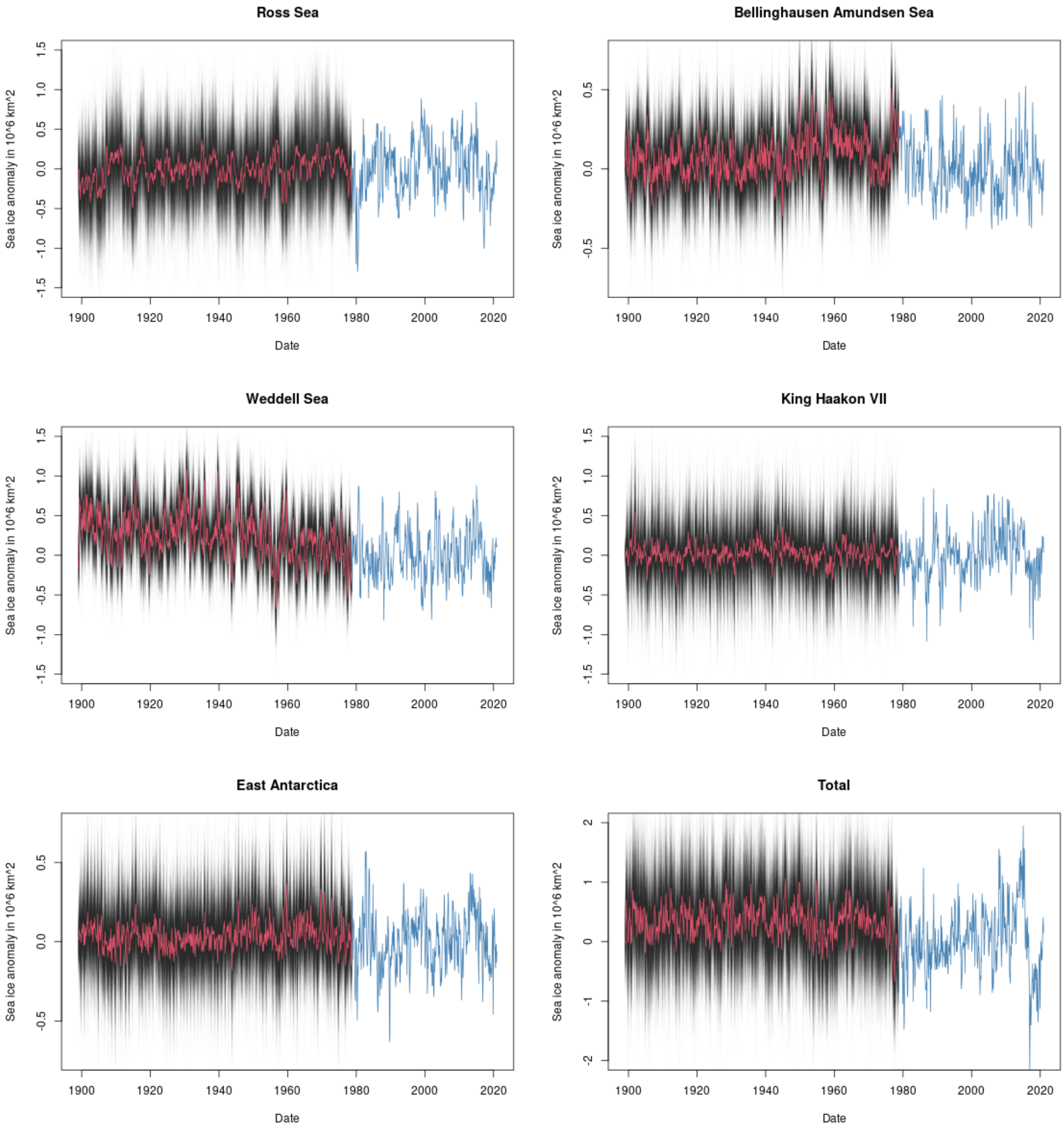
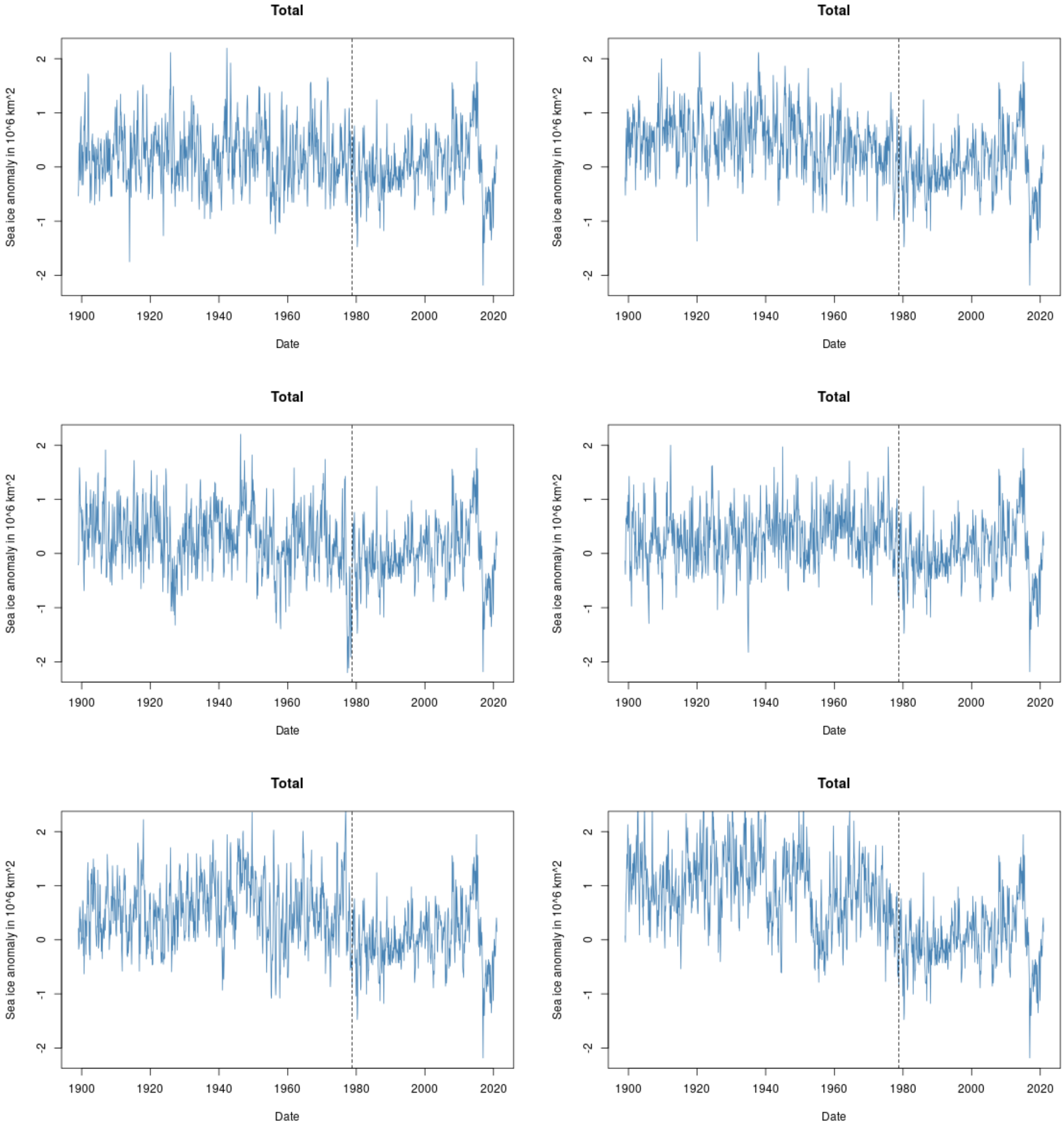


Figure 4.10: Ensemble reconstruction of 20th century Antarctic sea ice. Individual reconstructions are drawn in light gray, the average reconstruction in red, and observed sea ice in blue.


Total

Total

Total

Total

Total

Total

Figure 4.11: Individual reconstructions of 20th century total Antarctic sea ice. The dashed line separates reconstruction from observed data.

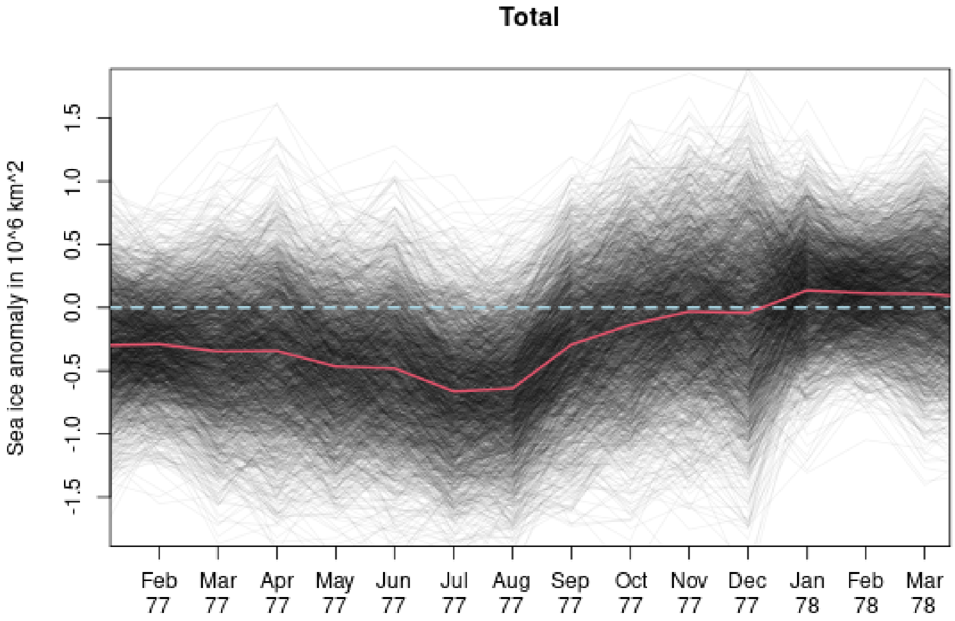


Figure 4.12: Ensemble reconstruction of total Antarctic sea ice anomaly from February 1977 to February 1978. Individual reconstructions are drawn in light gray and the average reconstruction in red.

Ross Sea, 25% most important stations

Bellingshausen Amundsen Sea, 25% most important stations

Weddell Sea, 25% most important stations

King Haakon VII, 25% most important stations

East Antarctica, 25% most important stations

Figure 4.13: Importance of weather station in sea ice reconstruction by sector. A station's importance is measured as the sum of the absolute values of all regression coefficients pertaining to its data.

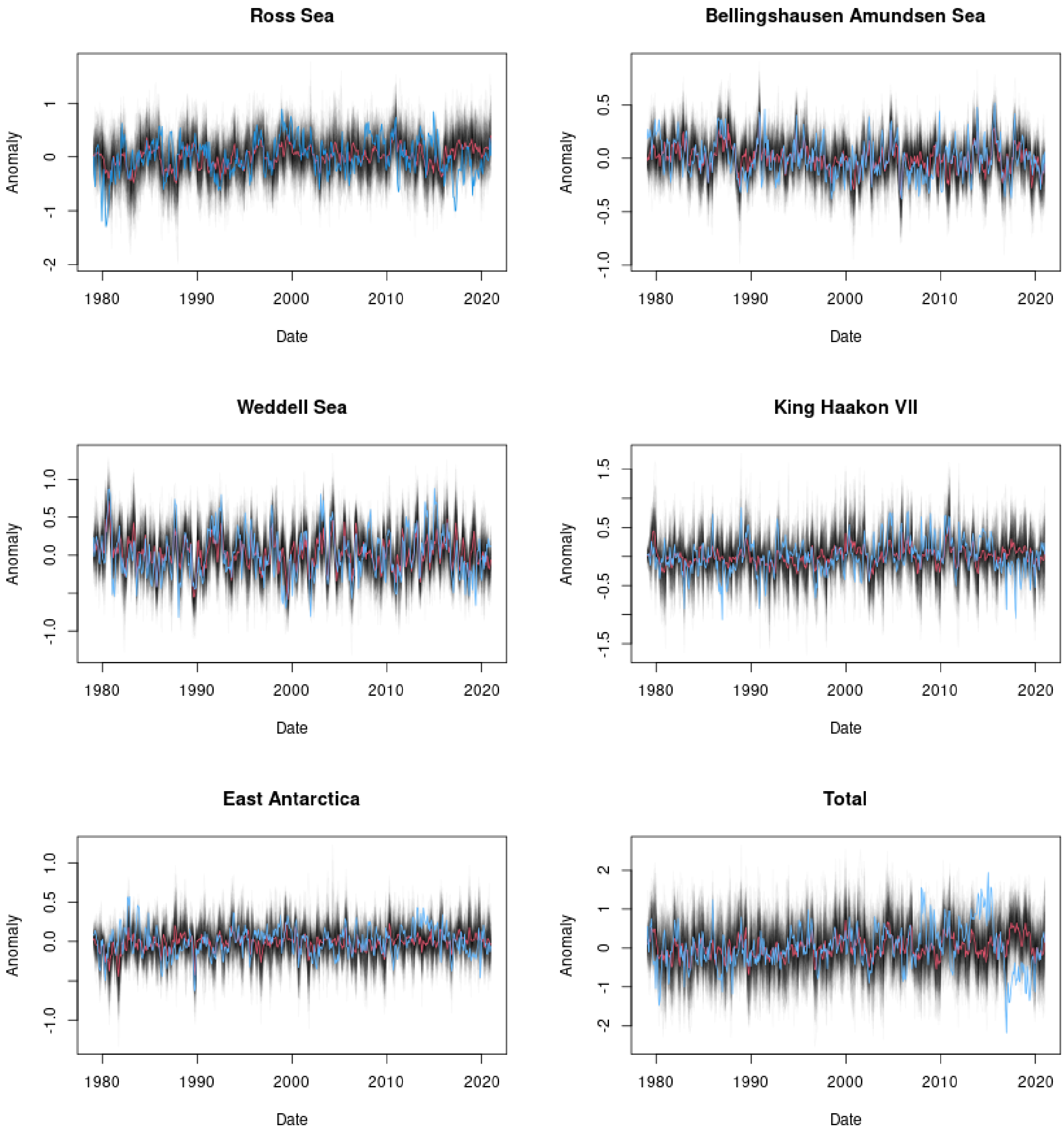


Figure 4.14: Out of sample reconstruction in the LOOCV. Individual reconstructions are drawn in light gray, the average reconstruction in red, the true sea ice in blue.

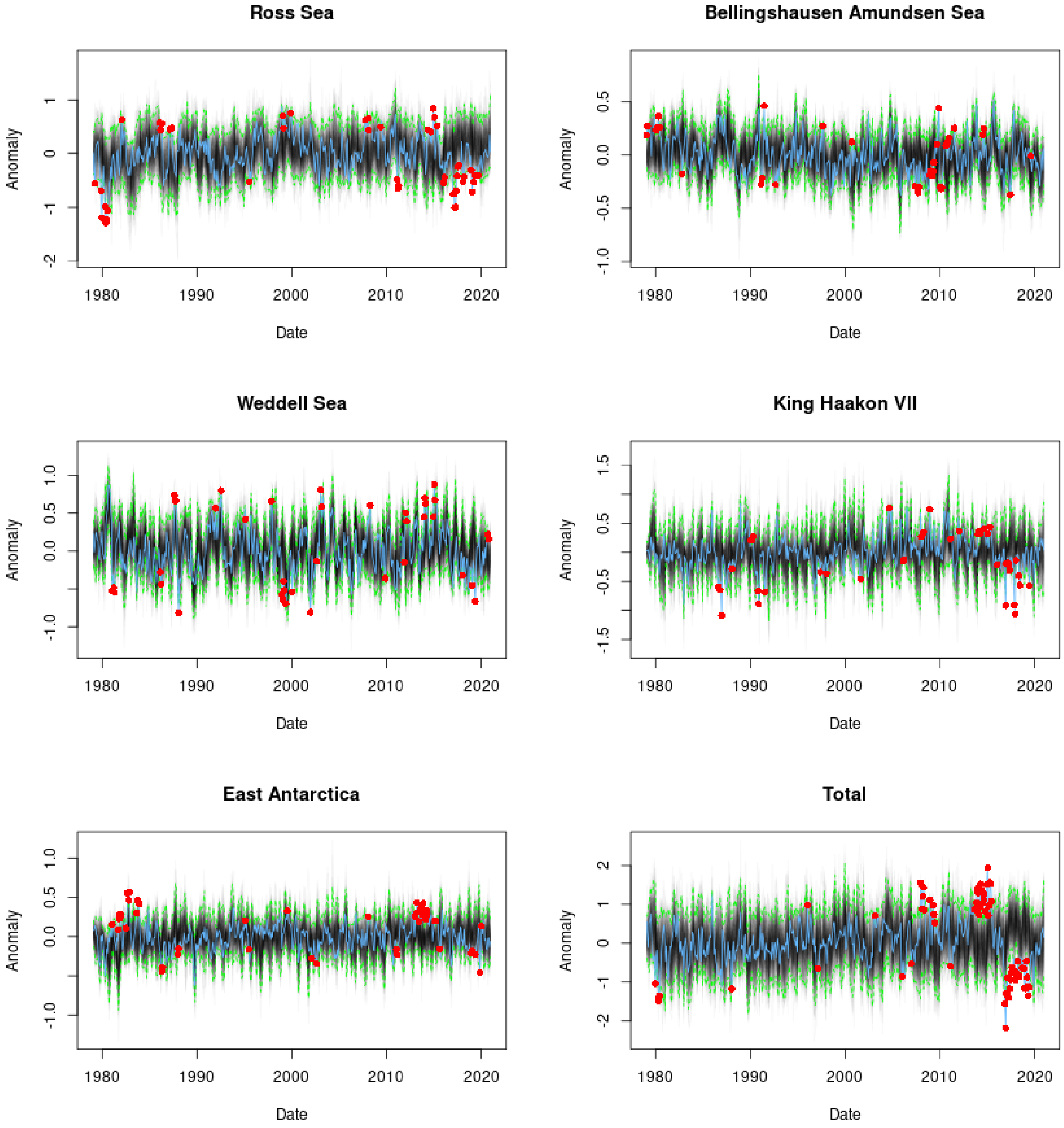


Figure 4.15: Coverage of 95% credible intervals for the sea ice anomaly (Bayesian equivalent of a 95% prediction interval) in the LOOCV. Individual reconstructions are drawn in light gray, bounds of the 95% CI in green, and the true sea ice in blue. Months in which the observed sea ice extent is outside its credible interval are marked with red dots.

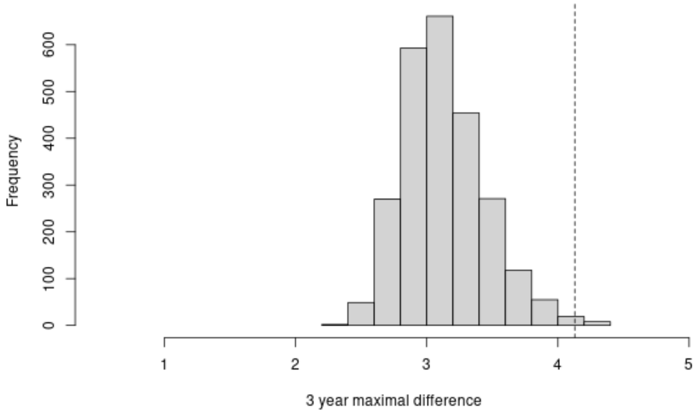


Figure 4.16: Distribution of three year sliding windows' maximal difference over the reconstruction period 1900-1979. The dashed line marks the observed three year difference of 4.13 million km^2 between the anomalies of January 2015 and December 2016.

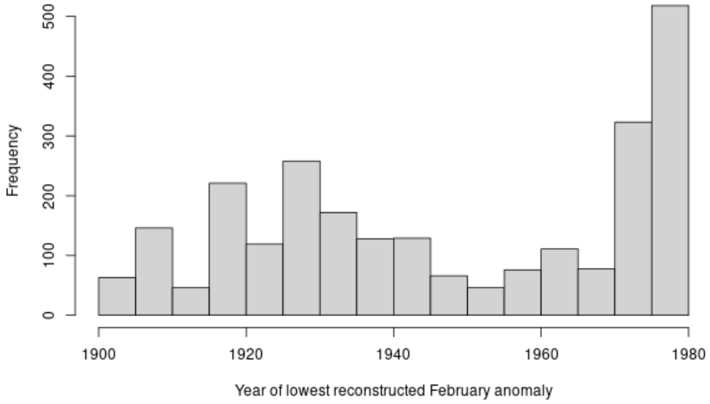


Figure 4.17: Distribution over the year in which each of the 2500 reconstructions had the lowest February anomaly.

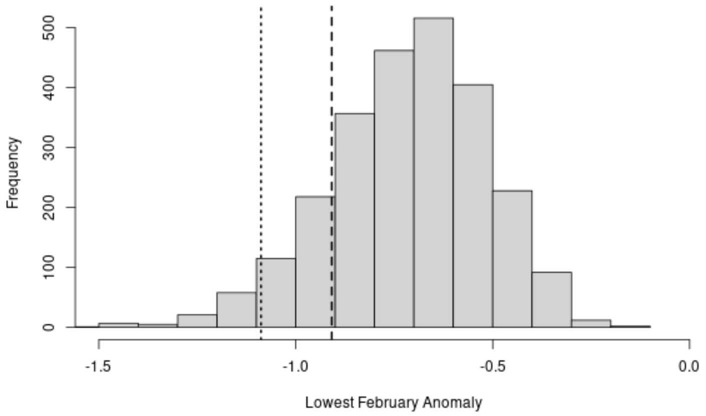


Figure 4.18: Distribution over the reconstructed lowest February anomaly per reconstruction over the reconstruction period 1900-1979. The dashed line marks the observed February 2022 anomaly, the dotted line the observed February 2023 anomaly.

CHAPTER 5

Functional Regression Model for the Antarctic Sea Ice Edge

5.1 Introduction

Antarctic sea ice is a key component of the Southern Hemisphere's complex climate system. It is one of the key components of the Antarctic climate system with extreme seasonality and remarkable interannual change. Every year the sea ice in the Southern Ocean grows from under 5×10^6 km^2 in February to more than 17×10^6 km^2 in September (Par19), a difference greater than the entire land area of Canada which is about 10×10^6 km^2, impacting all aspects of the Antarctic climate and ecosystem. The sea ice edge is the region between open water and dense pack ice. Its location is mainly driven by upper-layer ocean circulation and surface level winds moving the ice, and local thawing and freezing (HMS16). The ice edge zone is a nutrient rich zone of high biomass and primary productivity sustaining a complex ecosystem that is also home to many fishes, birds, and mammals (Kno06). The following models with different purposes have been proposed in the literature for the Antarctic (or Arctic) sea ice edge. (Mar97; CG07; De 09) use whaling records to obtain information about the development of the Antarctic sea ice edge from 1931-1987. Whales are usually caught in open water, usually near the sea ice edge and certainly not in the ice pack, so the southern-most latitude of a whale catch per longitude and date can be used as a proxy for the sea ice edge. This is still a biased estimate, but the bias tends to be small and correctable. The sea ice edge's latitude is modeled as step function using a linear model with main effects for longitude discretized into 10° of longitude blocks, time of year discretized into 10 day intervals, and whaling season (i.e., year), and interactions between longitude blocks and time of year blocks. A (Bayesian) general contour model has been proposed for the Arctic sea ice edge by (DRB17; DRB19) with the purpose of bias correcting ensemble forecasts from numerical prediction systems, in their specific case a 25-member ensemble provided by the European Centre for Medium-Range Weather Forecasts (ECMWF, available from Cop19). This means that they do not train their model to predict the actual sea ice at any location or time but instead focus on estimating the bias in other models' sea ice predictions. Their sea ice edge prediction is computed in a second step by subtracting the bias estimated by their model from the sea ice edge predicted by a numerical prediction system. Their model does not contain an annual cycle but only an interannual development of the bias estimated. Nonetheless, they do offer a theoretical model for the Arctic sea ice edge. They segment the Arctic into seven different regions that are modeled separately due to the fragmented sea ice area in the Northern hemisphere. The sea ice edge in each region is modeled as an ordered set of points that are referenced by their distance to a fixpoint (i.e., a point in the center of a cyclical ice edge) or a fixed line that is roughly parallel to the sea ice edge. In the case of the central Arctic, the region that is most like the Antarctic, a fixpoint is used, and the sea ice edge is operationalized as a set of points that are indexed by their angle and their distance to the fixpoint. For the Antarctic, an obvious choice for the central fixpoint is the South Pole implying that the points on the edge are indexed by their longitude (angle around fixpoint) and latitude (distance to fixpoint). (ZC19) model the Arctic sea ice as binary indicators on a gridded map, implying sea ice edges of arbitrary shapes. For computational feasibility only a snapshot of the month of September is considered in every year. Pixel by pixel ice/water transition probabilities are interpreted but interannual trends are only reported aggregated over the entire study area, whereas spatial trends are only reported as snapshots from one year to the next. The temporal and spatial development of the model is too high dimensional to be plotted in two dimensions.

The remainder of this chapter is structured as follows: Section 5.2 describes the data acquisition process for the sea ice concentration using remote sensing, how the sea ice edge is computed, and challenges when working with this data. Section 5.3 briefly reviews functional data analysis (FDA), discusses how the Antarctic sea ice edge fits into this framework, and introduces the functional regression model used to model the Antarctic sea ice edge. It also reports and interprets the model's estimated coefficients as the Antarctic sea ice's annual cycle, interannual trend, and variability. Section 5.4 reports the model's performance based on visual agreement with the true sea ice edge, average distance to the true sea ice edge, error in implicit sea ice extent by sector, and the coverage of 95% prediction intervals. Section 5.5 discusses the contributions of this chapter to the literature and its limitations. We conclude with an outlook on potential future research.

5.2 Data

Estimation of the sea ice on a continental scale in Antarctica is only feasible using remote sensing (i.e., satellites) due to its vastness and remoteness. Polar-orbiting sun-synchronous satellites are launched into orbits such that each successive orbital pass occurs at the same local time of day, resulting in images with roughly uniform lighting and weather patterns for that time and place. The modern-day satellite record for sea ice concentration, edge, and extent, starts in October 1978 with the launch of the NIMBUS-7 Scanning Multichannel Microwave Radiometer (SMMR). Since then, estimates of the Antarctic sea ice are available on a two day resolution and since the launch of a series of Special Sensor Microwave Imager (SSM/I) and Special Sensor Microwave Imager/Sounder (SSMIS) instruments on Defense Meteorological Satellite Program (DMSP) satellites in August 1987 on a daily resolution. This data is collected in the climate data record (CDR) daily concentration fields from the NOAA/NSIDC CDR of Passive Microwave Sea Ice Concentration, Version 4 (https://nsidc.org/data/g02202 MFW16), providing sea ice concentrations on the SSM/I-

SSMIS polar stereographic grid (25 km × 25 km) from 25 October 1978 to 31 December 2020. We use the sea ice based sectors introduced by (RH14) to summarize and reference the location of the sub-sector level trends described here.

There is a unique set of challenges when using this data. The data is troublesomely large and small at the same time. The sea ice concentration is measured on a 332×316 grid where each grid cell is $25\text{km} \times 25\text{km}$, about ten times the size of Manhattan. During the 42 year period from 1978-2019 we have images for 13,391 days. With each image consisting of 104,912 pixels, we observe a total of $1.4 \cdot 10^9$ pixels, which is an overwhelmingly large data set for computationally expensive models. At the same time, we only have 42 observations of the complete annual sea ice cycle, which is a small number for estimating its variability and large-scale development. This motivates the creation of an ice model focusing only on the relevant part in these sea ice concentration maps, the sea ice edge, as vast areas are almost fully covered in ice or virtually ice free. An efficient method for reducing the size of our data set is to only store and model the sea ice edge instead of the entire map of sea ice concentrations. Using the daily gridded sea ice concentrations provided by NSIDC as ground truth, we operationalize the sea ice edge as the most Northern circumpolar 15% sea ice concentration isoline, which may enclose polynyas and other encircled areas of lower sea ice concentration. This means that the sea ice edge on any given day is one circular set of points around the South Pole. The 15% sea ice concentration threshold is somewhat arbitrary and can lead to volatile edges and “jumps” in the sea ice extent between consecutive observations. Particularly during the melting season, large chunks of sea ice are breaking away from the ice pack, creating a vastly different sea ice edge between neighboring days. Storing only the pixels on the sea ice edge instead of the entire gridded map reduces the size of the data set to about $1.6 \cdot 10^7$ pixels, which is about 1.2% of its original size.

5.3 Functional Regression Model for Antarctic Sea Ice Edge

To model the sea ice edge, we propose a functional regression model. Functional data analysis (FDA) is the field of statistics that analyzes data that are curves, trajectories, surfaces, or any other functions varying over a continuous support (RS06). Functional data are data for which every observation is not just univariate or multivariate, but an entire smooth function defined on a continuous domain, often time. Examples of functional data include growth processes over time, e.g., the height of children as a function of their age, temperature/precipitation curves as a function of day of year, mass spectra as a function of the mass-to-charge ratio, image outlines as a function distance traveled on its outline. The key difference to multivariate data is that the function can (theoretically) be evaluated at an infinite number of arbitrary points in the domain and measurements are obviously strongly correlated when they are nearby. The difference between functional data and time series analysis is more subtle: time series usually assume discrete time steps and the eponymous time as its support, whereas functional data assume a general continuous support that can be of any nature. In many applications of time series, the support is continuing past the observed range, i.e., time is expected to continue past the present into the future and the future development of a time series is of interest. In functional data, the support is usually known and fixed, and functional observations do not extend beyond it. The Antarctic sea ice edge on any given day is known to be a closed loop around the South Pole. It is a reasonably smooth function on known, bounded support, making a functional regression model a good choice for analyzing its behavior. The true sea ice edge at time t in $[1978.00, 2019.00]$ is an ordered set of points on the globe with longitude lon and latitude lat,

$$\mathrm{edge}(t) = \{\mathrm{lon}, \mathrm{lat}\}(t), \tag{5.1}$$

which is a general representation allowing for arbitrary shapes, in particular folds, loops, and intersections. A simplification with precedence in the literature (e.g. Mar97; CG07; De 09; RH14; DRB17; DRB19) is to model the latitude of the points on the edge at time t as a

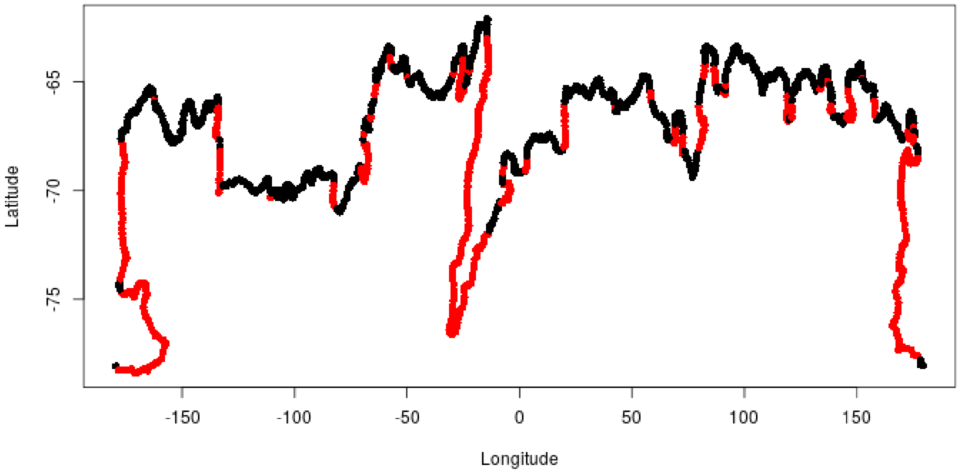


Figure 5.1: Geographic projection of sea ice edge on 06-01-1990 based on a 15% sea ice concentration threshold. Where multiple values of latitude are observed at the same longitude only the northernmost point is colored in black, all others are colored in red. The points below the northernmost observation make up 36% of the edge on this day.

function of their longitude and time. This model simplification is problematic, as the edge's latitude cannot truly be expressed a function of its longitude, as a function in the mathematical sense maps every element in the domain (here longitude) to exactly one element in the codomain (here latitude). This is not the case because we observe multiple latitudes for the same longitude as some edges "curve backwards" or "fold", see Figure 5.1. Note that the sea ice edge's latitude is not a function of longitude, not even a not-bijective function, not-injective-function, not-surjective function, or any other function. It is truly no function at all. In our simplified model, the ordered points on the edge are treated as an unordered point cloud and a smooth function is fit through them. (DRB17; DRB19) model the sea ice edge's convex hull instead. In any case, the simplified model has the great advantage that the target variable is a one-dimensional function, which makes the interpretation of the model and its coefficients more straight-forward and intuitive, in particular the estimated annual cycle as well as interannual trend by longitude. This simplified model where the edge's latitude lat is modeled as a function of its longitude lon, time t, and day of year $\mathrm{doy}(t)$, can be represented as a functional regression model of the form

$$\mathrm{lat}(\mathrm{lon}, t) = f(\mathrm{lon}) + g(\mathrm{lon}, \mathrm{doy}(t)) + h(\mathrm{lon}, t) + \varepsilon(\mathrm{lon}, t), \tag{5.2}$$

with intercept f, annual cycle g, interannual trend h, and functional residual ε. All coefficients are estimated as smooth functions across all longitudes simultaneously. The residual ε at time t is a function of longitude with pointwise expectation zero and pointwise variance $\sigma(\mathrm{lon}, \mathrm{doy}(t))$ depending on longitude and day of year. All coefficients are cyclical in longitude, as the bounding longitudes −180°E (=180°W) and 180°E coincide. The 1d coefficient function $f(\mathrm{lon})$ is the intercept per longitude, representing the average latitude of the sea ice edge at a given longitude across all days of year. The 2d coefficient function $g(\mathrm{lon}, \mathrm{doy}(t))$ is a smooth function describing the annual cycle for every longitude. It is cyclical in day of year as well, as the bounding days December 31st and January 1st are neighboring dates. For modeling purposes, $\mathrm{doy}(t)$ is implemented as the fractional day in the tropical year $\in [0, 365.2422]$ to avoid the irregularity of leap years. The 2d coefficient function $h(\mathrm{lon}, t)$ describes a smooth trend for all longitudes over the years. The residual at time t is a smooth function of longitude with pointwise expectation zero and pointwise variance $\sigma(\mathrm{lon}, \mathrm{doy}(t))$ depending on longitude and day of year. The variance estimate $\sigma(\mathrm{lon}, \mathrm{doy}(t))$ can be interpreted as the variability of the sea ice edge for a given longitude and day of year. This model parameters are estimated using the R package **refund** (GSH19). To speed up computation the model is trained on the observed sea ice edge of every 8th day. The estimates based on this subsampled data theoretically have a larger variance but are still unbiased. As a test of robustness, we computed the model based on the observed sea ice edge of every 16th day, resulting in virtually indistinguishable coefficient estimates and performance measures, assuring that the model converges on the subsampled data.

A strength of the proposed functional regression model is its interpretability. The annual cycle and interannual development of the sea ice at all longitudes are explicitly modeled using coefficient maps that can be visualized and interpreted directly. The coefficient function $f(\mathrm{lon})$ is the intercept per longitude, see top left of Figure 5.2. The intercept is the average latitude of the sea ice edge per longitude which is mainly driven by the shape of the Antarctic continent and not particularly meaningful by itself. The sea ice edge tends to be further north in the longitudes from 50°W to 150°W. These longitudes correspond to the Weddell Sea, King Haakon VII, and East Antarctica sectors. The coefficient function $g(\mathrm{lon}, \mathrm{doy}(t))$ is a smooth function describing the annual cycle for every longitude, see bottom left of Figure 5.2. This coefficient function is a two-dimensional function that maps day of year and longitude onto a difference in latitude. It is identifiable due to a sum-to-zero constraint for every longitude. This coefficient map can be interpreted as the annual cycle around the intercept per longitude. One can imagine a vertical slice through the map at a specific longitude as the annual cycle at this longitude. We can see that the annual cycle is a lot stronger in the Weddell Sea and Ross Sea than in the Bellingshausen-Amundsen Sea and East Antarctica. The average latitude of the sea ice edge at longitude lon^* and day of year $t^* \in [0, 365.2422]$ can be computed by adding the annual cycle at that longitude and day of year $g(\mathrm{lon}*, t^*)$ to the intercept $h(\mathrm{lon}^*)$, that is the expected sea ice edge for a given day of year is the sum of intercept and annual cycle. The coefficient function $h(\mathrm{lon}, t)$ describes a smooth trend for all longitudes over the years, see top right of Figure 5.2. This two-dimensional coefficient map can be interpreted as the inter-annual trend of the sea ice edge per longitude. It is identifiable due to a sum-to-zero constraint for every longitude. The strong incline around 2010 and drop-off around 2014 identified by (e.g. Par19) are captured by our model. The “untypical” development of the sea ice in the Bellingshausen & Amundsen Sea in comparison with other sectors is also found by the model. The variance $\sigma(\mathrm{lon}, t)$ of residual $\varepsilon(\mathrm{lon}, t)$ is modeled as a function of longitude and day of year, see bottom right of Figure 5.2. This two-dimensional coefficient map identifies strong trends in residual variability between sectors and seasons, with high variability during the sea ice melt and freeze, particularly in the Ross and Weddell Sea.

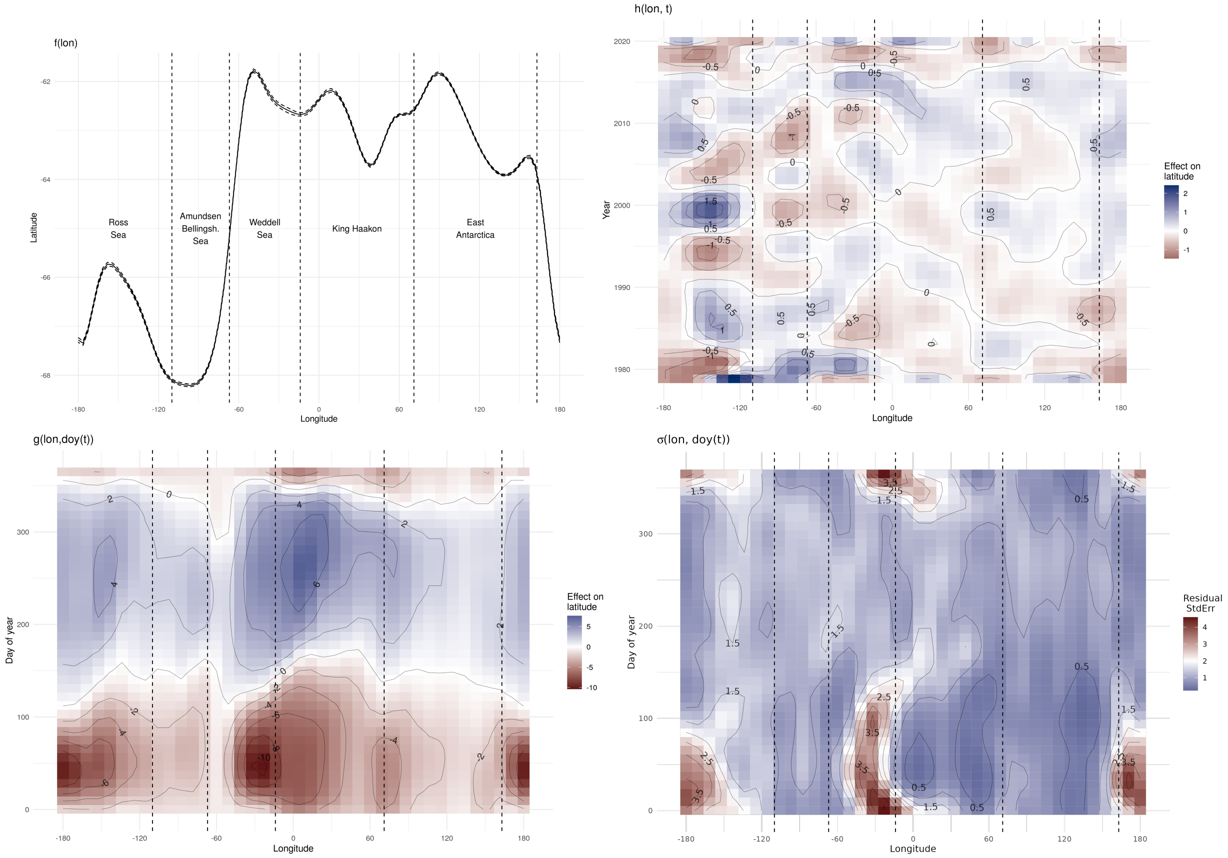


Figure 5.2: Coefficient functions for functional regression model specified in Equation (5.2). Note the different scales in the coloration. The dashed lines separate the sea ice based Antarctic sectors defined by (RH14).

5.4 Model Evaluation

In this section we report measures of discrepancy between the model and the true sea ice and evaluate them. The model performance is evaluated on visual inspection, average distance of modeled sea ice edge to true sea ice edge, coverage of 95% prediction intervals (95% PI), and implicit error in sea ice extent. Even though the model is trained on data from 8 day intervals, it can predict the sea ice edge for any day in our study period. The true sea ice edge is available on a two day resolution from October 1978 to July 1987, and a one day resolution afterwards. To even weight to all time periods in our quantitative performance measures, we choose to predict the sea ice edge and compare it to the true sea ice edge on a two day resolution across the entire study period.

5.4.1 Graphical Evaluation via Visual Fit

The first step of the model evaluation is a visualization of the predicted and true sea ice edge over the course of the study period. Snapshots of selected days are shown in Figure 5.3 and an animation over the entire study period is publicly provided (Mai23c). This shows the functional regression model fits the data well over the entire time period. The two major inlets in the Antarctic continent, the Weddell Sea and the Ross Sea, have a high variability in the timing and speed of the melt, leading to higher discrepancies between estimated and true sea ice edge in January, February, and March.

5.4.2 Numerical Performance Measure: Average Distance to True Sea Ice Edge

Numerical performance measures are important for an objective evaluation of the model that allows comparisons with existing and future models in the literature. The distance of the modeled sea ice edge to the actual sea ice edge is computed per 1 degree of longitude and defined as the distance along the latitude lines (as opposed to the closest true sea ice edge

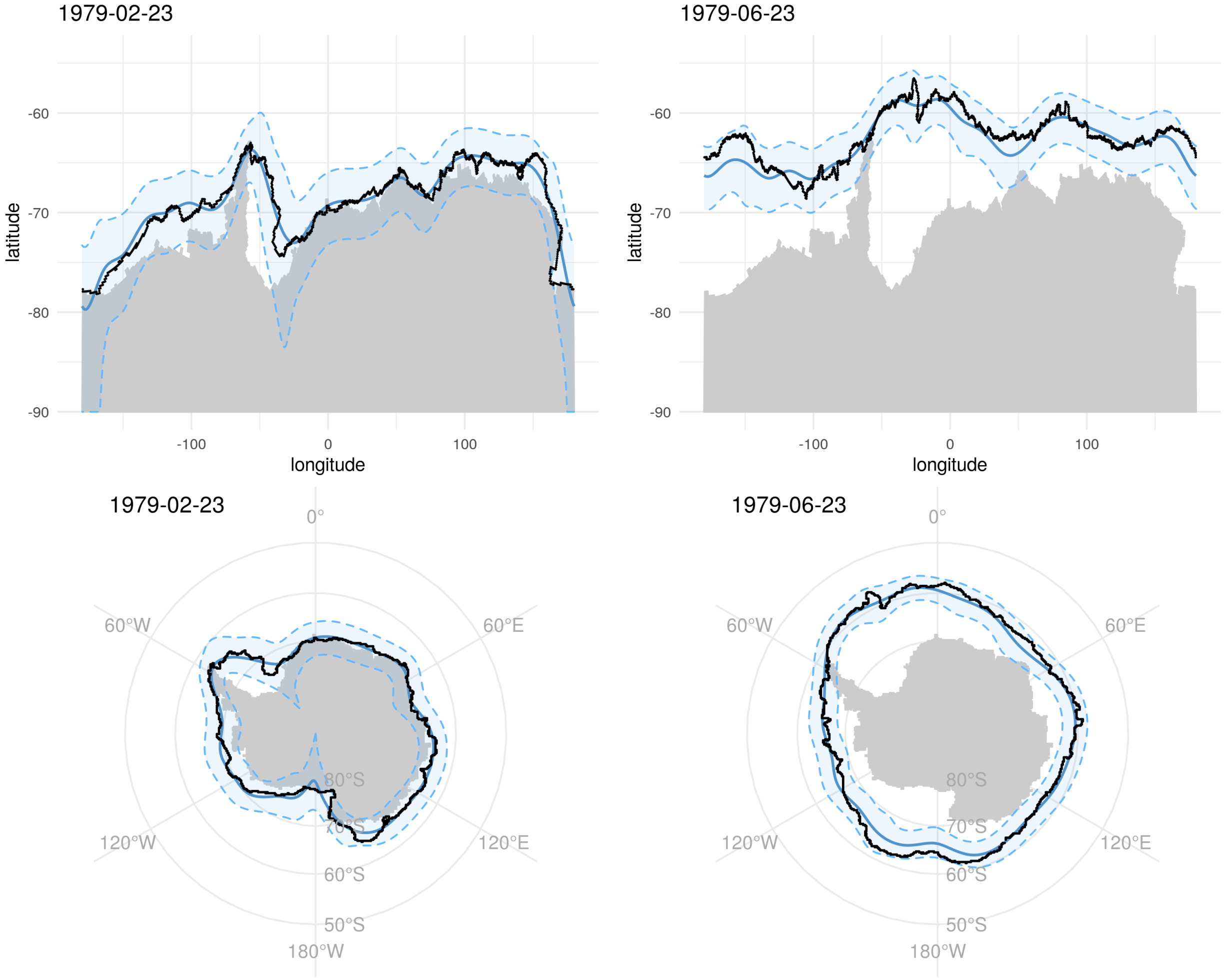


Figure 5.3: True sea ice edge based on 15% sea ice concentration threshold (black) and modeled sea ice edge (blue) with 95% prediction interval (shaded area between dashed blue lines) for February 23, 1979 (left) and June 23, 1979 (right). Top and bottom show the same day in geographic and polar stereographic projection.

in any direction). We can think of this measure as the expected distance one would have to walk to get to the true sea ice edge if one was dropped at a random point on the estimated sea ice edge in the time period and sector of interest. The average distance of the estimated sea ice outline $\widehat{\mathrm{lat}}(\mathrm{lon}, t)$ to the true sea ice outline $\mathrm{lat}(\mathrm{lon}, t)$ for a time certain sector and time period is computed by averaging their distance over the desired longitudes and time, i.e.,

$$\frac{1}{\#(\mathrm{period})} \sum_{t \in \mathrm{period}} \left(\frac{1}{\#(\mathrm{sector})} \int_{\mathrm{lon} \in \mathrm{sector}} |\widehat{\mathrm{lat}}(\mathrm{lon}, t) - \mathrm{lat}(\mathrm{lon}, t)| \, d\mathrm{lon} \right), \tag{5.3}$$

where $|\cdot|$ denotes the absolute value and $\#(\cdot)$ denotes the cardinality of the discrete or continous set respectively, i.e.,

$$\#(\mathrm{period}) := \sum_{t \in \mathrm{period}} 1,$$
$$\#(\mathrm{sector}) := \int_{\mathrm{lon} \in \mathrm{sector}} 1 \, d\mathrm{lon}.$$

This means that the cardinality of a time period is defined as the number of days t within it, and the cardinality of a sector is defined as the fractional count of longitudes within it, which is the difference between the bounding longitudes for contiguous sectors. Across all longitudes and years, i.e., sector = [-180, 180] and period = [1979, 2019] in the equation above, the modeled sea ice outline is on average 0.98° or about 109km away from the true sea ice edge. The average distance of modeled sea ice edge to the true sea ice edge by sector and month is shown in Figure 5.4. The distance is consistently small ($< 1°$), except for the Ross Sea and Weddell Sea from December to April. These two sectors coincide with the two largest embayments in the Antarctic continent and the time of quick decline and minimal sea ice. The average distance of modeled sea ice edge to the true sea ice edge by 1 degree of longitude and month is shown in Figure 5.5. This figure corroborates Figure 5.4, with the largest deviations occurring in the same sectors and months, and further validates the use of the sea ice based sectors introduced in (RH14). The longitudes of the large peaks in deviation match the sectors extraordinarily well, showing that these sectors provide a valid

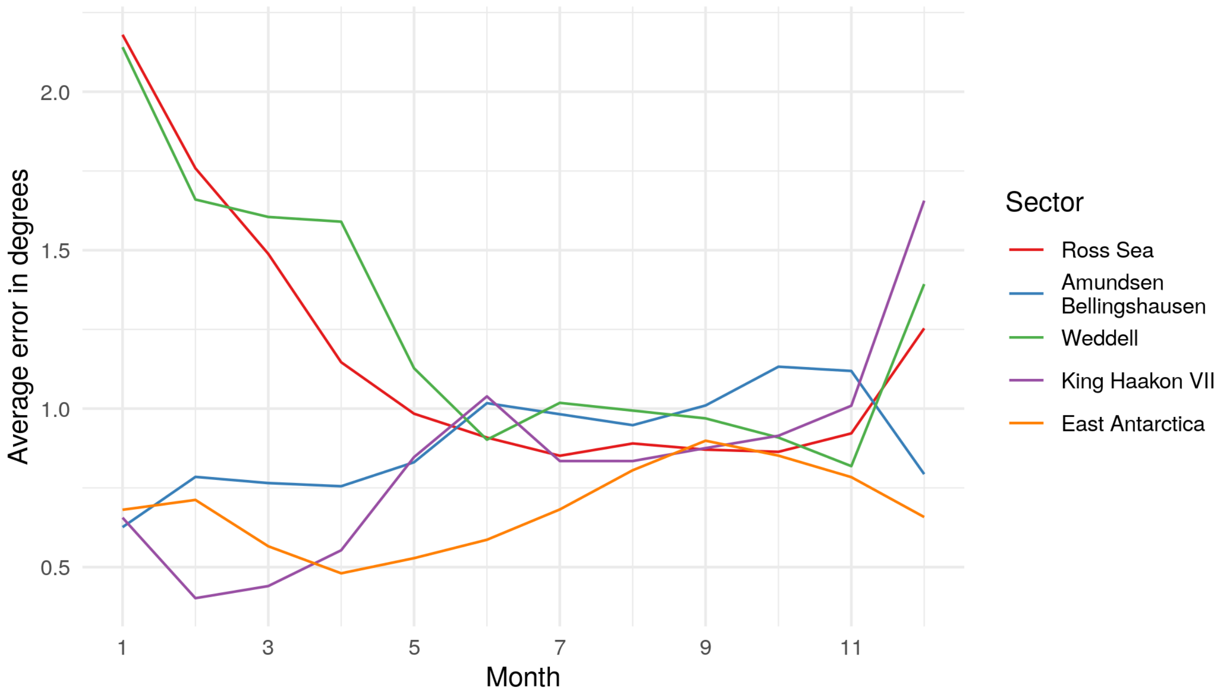


Figure 5.4: Average distance in degrees of modeled sea ice edge to true sea ice edge per sector and month.

aggregation of the underlying longitudes. Note that errors in degrees can be converted to kilometers by multiplying by 111 as latitude lines are approximately 111 km apart.

5.4.3 Numerical Performance Measure: Error in Implicit Sea Ice Extent

The modeled sea ice edge implies a sea ice extent overall and for any sectors of choice. This section quantifies the sea ice area misclassified by the model. A sector's sea ice extent is the area of ocean within it that is covered in ice, which can be computed as the total area of ice minus the area of ice-covered land. To compute the error in implicit sea ice extent for a time period and sector, the true sea ice extent and estimated sea ice extent are computed separately for every time point and sector. The performance metrics are then computed within each sector across all time points in the respective time period. Table 5.1 reports a summary of the mean absolute deviation (MAD) in mio. km^2, mean relative deviation

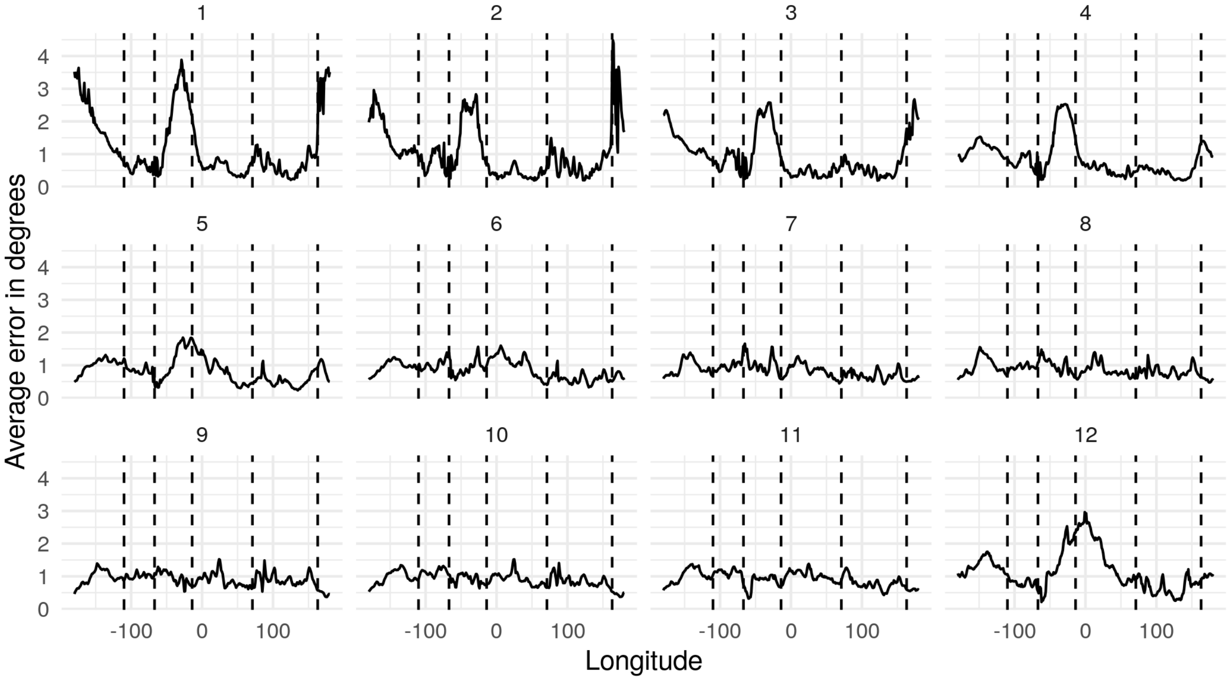


Figure 5.5: Average distance in degrees of modeled sea ice edge to true sea ice edge per 1 degree of longitude and month. Dotted lines separate the sea ice based sectors introduced by (RH14).

(MRD), and root mean square error (RMSE) in mio. km^2 using the sea ice based sectors defined in (RH14). The model presented in Equation (5.2) is optimized to minimize the squared error between the sea ice edge's observed and predicted latitude, i.e., the integral over the squared error in latitude over all longitudes. This does not minimize the (squared) error in sea ice extent, as predicting the sea ice edge to be 1 degree too far North is a much larger error in sea ice area than predicting it to be 1 degree too far South. This is more of a feature than a bug in the model specification that can be adjusted by either optimizing a different loss function or transforming the data set. In this case, transforming the latitude of the outlines in our data set to represent "area to South Pole" instead of "angle to South Pole" will achieve the desired result. For this purpose, any equal-area projection like the Lambert cylindrical equal-area projection (Lam70) or Gall-Peters projection (Gal85) will suffice. Using Lambert's cylindrical equal-area projection of the latitudes to "areatudes"

Table 5.1: Model accuracy of sea ice extent by sector measured by mean absolute deviation (MAD) in mio. km^2, mean relative deviation (MRD), and root mean square error (RMSE) in mio. km^2. Mean, 2.5% quantile, and 97.5% quantile of the sea ice extent per sector are included for reference.

Sector	Mean	q0.025	q0.975	MAD	MRD	RMSE
Total	10.63	2.69	16.94	0.40	0.06	0.51
East Antarctica	1.83	0.51	3.13	0.12	0.09	0.16
Ross Sea	3.10	0.51	4.77	0.25	0.16	0.31
Amundsen Bellingshausen	0.75	0.16	1.44	0.13	0.24	0.16
Weddell Sea	2.54	0.88	3.75	0.21	0.11	0.27
King Haakon	2.41	0.09	5.05	0.19	0.21	0.26

$$\text{art} = \sin\left(\frac{2\pi}{360}\text{lat}\right), \tag{5.4}$$

see Section 5.6 for a derivation, and specifying the functional model as

$$\text{art}(\text{lon}, t) = f(\text{lon}) + g(\text{lon}, \text{doy}(t)) + h(\text{lon}, t) + \varepsilon(\text{lon}, t), \tag{5.5}$$

will optimize the squared error in predicted sea ice extent. To increase human readability of the transformed model, the latitude of the predicted sea ice edge can be obtained applying the inverse function of Equation (5.4), i.e.,

$$\mathrm{lat} = \frac{360}{2\pi} \arcsin(\mathrm{art}). \tag{5.6}$$

We found that in our case a model trained on areatudes (5.5) as opposed to latitudes (5.2) resulted in a negligible improvement in predicting sea ice extent. This is presumably due to demonstrated weak nonlinearity of the transformation in our region of interest, see the Taylor series approximation of degree two quantifying the transformation's nonlinearity in Section 5.6, and we only report model performance based on the latitude based model.

To point out the difference between the error in implicit sea ice extent and average distance to true sea ice edge, assume that the true sea ice edge was truly a function of longitude and that there was no land mass in the Southern Hemisphere. Then the sea ice extent for a sector can be computed by integrating over the sea ice edge in a cylindrical equal-area projection. The mean absolute deviation (MAD) for a sector and time period, as reported in Table 5.1 and Figure 5.6, can then be computed as

$$\frac{1}{\#(\mathrm{period})} \sum_{t \in \mathrm{period}} \left| \int_{\mathrm{lon} \in \mathrm{sector}} \widehat{\mathrm{art}}(\mathrm{lon}, t)\, \partial \mathrm{lon} - \int_{\mathrm{lon} \in \mathrm{sector}} \mathrm{art}(\mathrm{lon}, t)\, \partial \mathrm{lon} \right| . \tag{5.7}$$

Note the different order of the integration and computation of the absolute deviation in comparison to Equation (5.3).

5.4.4 Numerical Performance Measure: Coverage of 95% Prediction Intervals

The expected sea ice edge as plotted in solid blue in Figure 5.3 is usually close to the true sea ice edge but notably smoother. 95% pointwise prediction intervals are computed for the sea ice edge, see shaded are between dashed blue lines in Figure 5.3 as

$$95\%\mathrm{PI}(\mathrm{lon}, t) = \widehat{\mathrm{lat}}(\mathrm{lon}, t) \pm 2 \cdot \Big(\widehat{\mathrm{se}}_{\mathrm{lat}}(\mathrm{lon}, t) + \hat{\sigma}\big(\mathrm{lon}, \mathrm{doy}(t)\big) \Big) \tag{5.8}$$

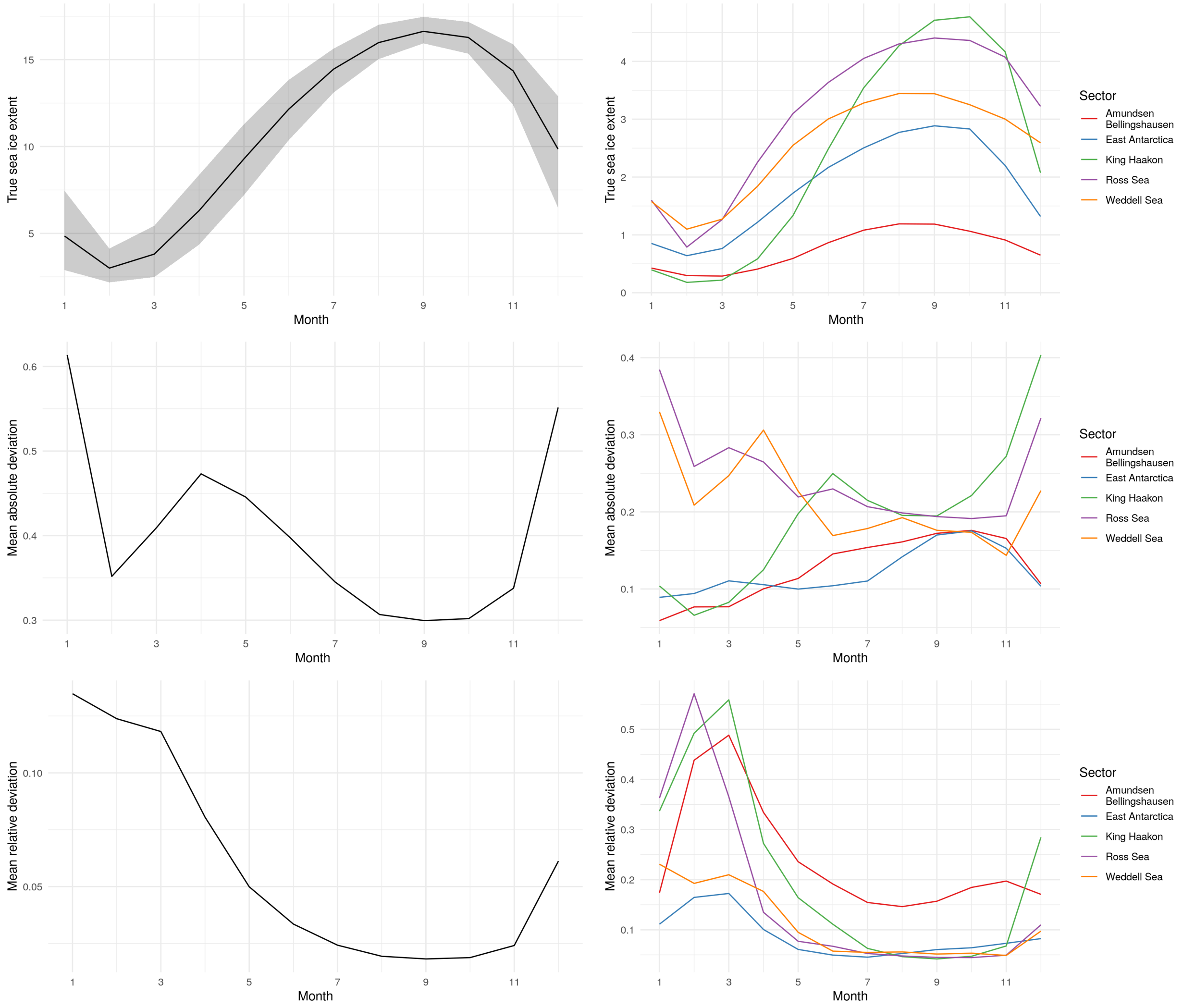


Figure 5.6: Average sea ice extent in million square kilometers (top) for reference, mean absolute deviation (MAD) in million square kilometers per sector (center), and mean relative deviation (bottom) without unit, split up by total sea ice extent on the left and by sector on the right.

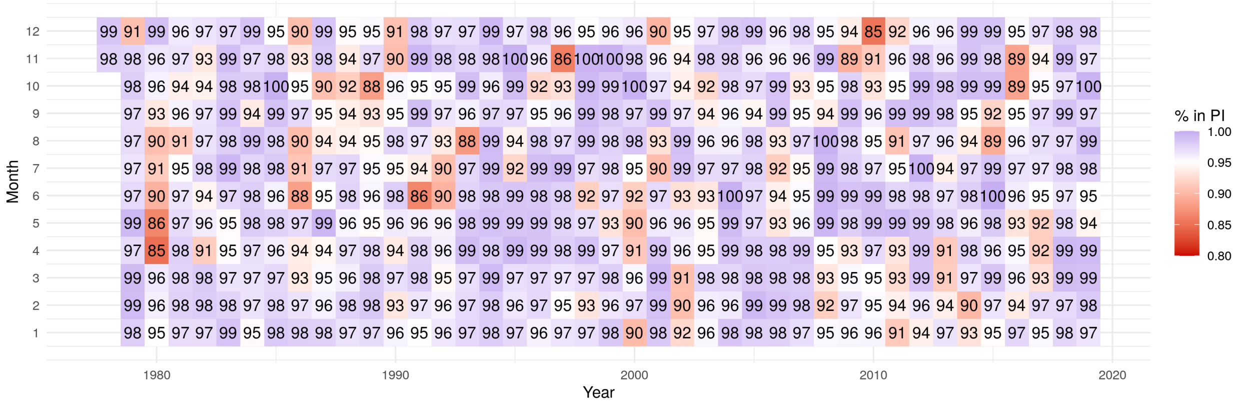


Figure 5.7: Percentage of true sea ice edge covered by 95% prediction interval by year and month.

where $\widehat{\mathrm{se}}_{\mathrm{lat}}(\mathrm{lon}, t)$ is the estimated standard error in the mean function and $\hat{\sigma}(\mathrm{lon}, \mathrm{doy}(t))$ is the estimated standard deviation in the residuals $\varepsilon(\mathrm{lon}, t)$. We choose to report prediction intervals (2× standard error in estimated mean plus 2× standard deviation of residuals) instead of confidence intervals (2× standard error in estimated mean) as the uncertainty in actual sea ice edge is scientifically of greater interest than the uncertainty in its estimated mean location. Poor coverage of our prediction intervals would indicate an underestimation of the sea ice variability or model uncertainty. This is not the case, as the percentage of the true sea ice outline covered by our 95% prediction intervals across the entire study period is 96%. Figure 5.7 breaks this number down by month in year. Vertical "streaks" of poor coverage found for sets of consecutive months, most notably April to September 1980, indicate that the model struggles to capture the sea ice behavior in anomalous years. The lack of vertical streaks of poor performance indicates that the annual cycle is captured well.

5.5 Discussion

We create the first statistical model for the Antarctic sea ice edge using a functional regression model. This framework allows us to explicitly model sea ice edge's latitude at all longitudes simultaneously, which has previously been unattempted, as well as the novel estimation of an annual cycle and interannual development of the Antarctic sea ice for every degree of longitude. Previous efforts have either focused on the Northern hemisphere or been temporally and spatially sparse. Our model improves upon (Mar97; CG07; De 09) by modeling sea ice edge's latitude at all longitudes and days of year simultaneously as opposed to modeling discrete sections of 10° longitude and 10 day periods, while taking into account the spatio-temporal connection of nearby longitudes and days of year.

Our model suffers from a number of limitations, with the most concerning one being that the sea ice edge's latitude is not a function of longitude, even though we chose to model it as such. The sea ice edge does fold backwards onto itself, resulting in multiple observations of latitude for the same longitude and the edge not being a function of its longitude. (Mar97; CG07; De 09) evade this problem by modeling the Antarctic sea ice edge discretely at every ten degrees of longitude, a much larger scale than the backward folds in the sea ice edge. (DRB17; DRB19) evade this issue in their model for the Arctic sea ice edge by modeling the convex hull of the 15% sea ice concentration isoline as the sea ice edge. Modeling these backward folds explicitly is possible, but comes at the cost of an increased dimensionality, making the model less interpretable and computationally more expensive, see (ZC19), who model the Arctic sea ice as a binary indicator on a two-dimensional map which is capable of creating arbitrarily shaped sea ice edges. Over all seasons and years, the treatment of the backward folds only has a marginal impact on the estimated sea ice, major improvements are gained from explicitly modeling an annual cycle, spatial autocorrelation, and interannual trends.

Thus, we deliberately choose the simplification of modeling the sea ice edge's latitude as a function of longitude to capitalize on a wide range of subsequent benefits. For one, the computational burden lightens enormously, allowing a daily model for the sea ice edge. Also, simplifying the sea ice edge to a one-dimensional function of longitude allows the interpretation of the 2d coefficient maps of the annual cycle and interannual development by longitude. We estimate and visualize a previously unobtained annual cycle and interannual development of the Antarctic sea ice across all longitudes simultaneously. The sea ice extent, a climatological variable of great interest, can be evaluated for any desired sector from the estimated sea ice edge. Our model shows a high predictive power except for local areas of high variability during the melt in December and January.

Our model provides further evidence that the sea ice based sectors proposed in (RH14) provide a fitting subdivision of the Antarctic ocean with respect to the location, annual cycle, interannual development, and variability of the sea ice edge. We find that our longitude-continuous analysis of the Antarctic sea ice edge shows similar patterns within these sectors, and more dissimilar patterns between them.

Our model is based purely on observations of the Antarctic sea ice edge, and employs temporal and spatial repetitiveness of the sea ice to create a statistical model. It is explicitly agnostic towards all other climatological variables or any physical relations and processes. This makes it distinct from the class of numerical climate models like the Community Earth System Model version 2 (DLB20) that are commonly used in the field, providing a novel and independent point of view.

This chapter clears the way for several future research opportunities by providing an explicit model for the Antarctic sea ice edge. A possible extension to our model is to explicitly model the covariance of the residuals as a process with expectation zero and covariance depending on longitude and time of year. This would allow us the creation of plausible sea ice edges based on locally and temporally sparse observations, which is entirely unnecessary when using the temporally and spatially complete modern satellite data but is crucial for attempting to reconstruct Antarctic sea ice for the time period before the satellite record.

Such a model can be used to extract information from temporally and locally sparse observations like ship logs (e.g., ED16; MH40; Mac72) about the unobserved circumpolar location of the sea ice edge. Creating a model for the circumpolar and sectoral sea ice edge can also be used to enhance local proxy reconstruction methods like ice core based reconstructions, see e.g., (AWC13) and (TAE19) for a summary, by providing a framework to compose a circumpolar sea ice edge based on local indirect observations.

As is, our model is not designed for short-term forecasting of the sea ice edge. For a good one-step prediction, the model would have to explicitly model the correlation between the residuals of neighboring days, which is assumed to be zero in our model. A good short-term forecast based on days with a backward curving outline would also predict the outline to be curving backwards in a similar fashion, which our model is not capable of.

5.6 Latitude to Areatude Transformation

This section derives the latitude to "areatude" transformation that can be used to optimize the error in sea ice extent as opposed to the error in the sea ice edge's latitude. To optimize the edge's implicit sea ice extent instead of its angular location, we propose a variable transformation in Equation (5.4). Measuring the sea ice edge's location in degrees latitude, an "angle to the South Pole" unit, causes the model's fitting algorithm to optimize the error in angular location, i.e., degrees longitude. When we want to optimize the sea ice edge's implied sea ice extent instead, as straight forward solution is to measure the sea ice edge's location in "area to the South Pole" units we call "areatudes". Note that the effect of this transformation of variable could also be achieved by explicitly optimizing a different loss function. Without proof, note that the area between a line of latitude and the South Pole, i.e., the area of a spherical cap A_{cap}, is given as

$$A_{\text{cap}} = 2\pi R^2 \left(1 + \sin\left(\frac{2\pi}{360}\text{lat}\right)\right), \tag{5.9}$$

with $R = 6357km$ the radius of earth between the poles. Consider a sea ice edge comprising of points $\{\text{lon}, \text{lat}\}_i$, with longitudes evenly spaced with distance

$$\Delta\text{lon} = \text{lon}_i - \text{lon}_{i-1} = \text{lon}_{i+1} - \text{lon}_i = \text{const.} \tag{5.10}$$

Then, for small values of Δlon, $\text{lat}_i \approx \text{lat}_{i+1}$, and the area of the triangle between neighboring points on the edge $\{\text{lon}, \text{lat}\}_i$ and $\{\text{lon}, \text{lat}\}_{i+1}$ and the South Pole is given as

$$\begin{aligned} \text{art} &= \frac{\Delta\text{lon}}{360} 2\pi R^2 \left(1 + \sin\left(\frac{2\pi}{360}\text{lat}\right)\right) \\ &\propto 1 + \sin\left(\frac{2\pi}{360}\text{lat}\right) \end{aligned} \tag{5.11}$$

completing the derivation of Equation (5.4) which is the Lambert cylindrical equal-area projection. Note that evaluating the area up to linearity is sufficient for our model, as it minimizes the squared error in the residuals. Applying any linear transformation, i.e., a transformation of the form $f(x) = ax + b$ with constants a and b, to our data will result in the same model. The inverse function of Equation (5.4) can be deduced as

$$\begin{aligned} \text{art} &= 1 + \sin\left(\frac{2\pi}{360}\text{lat}\right) \\ \text{art} - 1 &= \sin\left(\frac{2\pi}{360}\text{lat}\right) \\ \arcsin(\text{art} - 1) &= \frac{2\pi}{360}\text{lat} \\ \text{lat} &= \frac{360}{2\pi}\arcsin(\text{art} - 1), \end{aligned} \tag{5.12}$$

finishing the derivation of Equation (5.6). To assess the impact of the latitude-to-areatude transformation, we are using a Taylor series approximation of degree two to quantify the nonlinearity of Lambert's cylindrical equal-area projection around $a = -64°$ latitude,

$$\begin{aligned}
\text{art} &=: \varphi(\text{lat}) \\
&\overset{\text{Taylor exp.}}{=} \sum_{n=0}^{\infty} \frac{\varphi^{(n)}(a)}{n!}(\text{lat} - a)^n \\
&\overset{\text{2nd deg. approx.}}{\approx} \varphi(a) + \varphi'(a)(\text{lat} - a) + \frac{1}{2}\varphi''(a)(\text{lat} - a)^2 \\
&= \sin\left(\frac{2\pi}{360}(-64)\right) + \frac{2\pi}{360}\cos\left(\frac{2\pi}{360}(-64)\right)(\text{lat} + 64) \\
&- \frac{1}{2}\left(\frac{2\pi}{360}\right)^2 \sin\left(\frac{2\pi}{360}(-64)\right)(\text{lat} + 64)^2 \\
&\approx -0.89879 + 0.00765(\text{lat} + 64) + 0.00014(\text{lat} + 64)^2,
\end{aligned} \tag{5.13}$$

showing a weak nonlinearity of the transformation in our region of interest. In our region of interest between 55° South and 76° South where most of the sea ice edge lies, the distortion between Lambert's cylindrical equal-area projection and the geographic projection is less than 4° or 450km.

CHAPTER 6

Reconstructing 1966-1978 Antarctic Sea Ice Extent using early Satellite Images

6.1 Introduction

The National Snow and Ice Data Center (NSIDC) currently lists the Antarctic sea ice record for the years 1962-1973 as missing. This chapter fills a large part of this gap by reconstructing the Antarctic sea ice extent for the years 1966-1978 based on satellite images in the visible spectrum (1 band, wavelength 0.50 - 0.70 μm) that were recently digitized and made available by (Cam19). We show that information about the Antarctic sea ice can be extracted from historic satellite records in the visible spectrum as published by (Cam19) and that these reconstructions match previous estimates of the Antarctic sea ice for this time period. A major hurdle of using this data for sea ice reconstruction is that most of the surface is occluded by clouds (60-90% depending on the latitude). To make matters worse, clouds and ice are virtually indistinguishable in these images. To circumvent this problem, a minimum brightness filter is applied to all images available in each month, resulting in one cloud free image per month. These monthly images are then classified pixel by pixel into the classes open water versus ice using a Bayesian latent discriminant analysis. This type of classification model allows the use of the 40 year (1979-2019) sea ice concentration record provided by modern satellites as an informative prior. From these ice-versus-water probability maps, a 30% concentration threshold is used to find the sea ice edge. The estimated sea ice extent for this month is given as the area within the sea ice edge minus the area of the Antarctic continent. After adjusting for a constant offset, the reconstructed sea ice extent agrees well with the sea ice extent estimated for the years 1973-1977 by (CPV03) based on NIMBUS 5 data in the microwave spectrum. These corrected sea ice extent estimates agree well with previous estimates of sea ice extent in this time period. The method proposed in this paper is by no means specific to data set provided by (Cam19) but can be applied to other satellite images that suffer from cloud occlusion.

(Disclaimer: This chapter only analyzes the images from 1974-1978 due to unlabeled missingness in the data that could not be handled automatically. Given the limited time frame, the author did not manage to manually outline the unlabeled missing areas in more than five years worth of images.)

6.2 Methods

This paper focuses on a set of satellite images recently published by (Cam19), who provide an early series of satellite images in the visible (1966-1978) and infrared (1973-1978) spectrum. Table 6.1 lists the satellites with their operational dates and instrumentation. The AVCS instrument as equipped on the ESSA satellites, ITOS, and NOAA 1, recorded in the visible wavelength band from 0.45 - 0.65 μm with a resolution of 2.2 km - 3.0 km and a swath width of 2300 km. The SR instrument as equipped on ITOS 1 and NOAA 1-5 captured images in the visible (wavelength 0.50 - 0.70 μm, 3.2 km resolution, 2900 km swath width) and infrared (wavelength 10.50 - 12.50 μm, 8km resolution, 2900 km swath width) spectrum. This data set has not been analyzed yet but provides direct satellite observation for the time period from 1966-1973 that the NSIDC Antarctic sea ice record currently lists as missing.

The analysis is created in a series of consecutive steps:

- Data cleaning and preprocessing (Section 6.2.1)
- Estimation of sea ice concentration (Section 6.2.2)

Table 6.1: Overview of operational dates and Instrumentation of satellites in data set. It is unclear whether the data recorded for ITOS 1 and NOAA 1 stems from the AVCS or the SR instrument (Source: Cam19). The data available to the author start in December 1966 and end in March 1978.

Satellite	Operational Dates	Instrument
ESSA 3	02 October 1966 - 02 December 1968	AVCS
ESSA 5	20 April 1967 - 20 February 1970	AVCS
ESSA 7	16 August 1968 - 10 March 1970	AVCS
ESSA 9	26 February 1969 - 12 November 1972	AVCS
ITOS 1	23 January 1970 - 18 June 1971	AVCS or SR (uncertain)
NOAA 1	11 December 1970 - 19 August 1971	AVCS or SR (uncertain)
NOAA 2	15 October 1972 - 30 January 1975	SR
NOAA 3	06 November 1973 - 31 August 1976	SR
NOAA 4	15 November 1974 - 18 November 1978	SR
NOAA 5	29 July 1976 - 16 July 1979	SR

- Estimation of sea ice outline and extent (Section 6.2.3)
- Interpolation of missing days (Section 6.2.4)
- Adjustment of estimated sea ice extent (Section 6.2.5)

6.2.1 Data Cleaning and Preprocessing

Videos of the the data before and after our preprocessing step are publicly available (Mai23b). The satellite images in the database suffer dramatically from missingness. Many images are partially unobserved due to the satellite's imperfect orbit or technical malfunctions. These missing values occur spatially and temporally in clustered, but their missingness can be assumed to be independent of the variable of interest, i.e., satellite malfunction can be assumed to be independent of the Antarctic sea ice. This superficially trivial independence implies that missing observations are missing completely at random, which will simplify its treatment in subsequent analyses.

For some months in the study period there are no images available. Due to total darkness during the Antarctic winter, satellite images in the visible spectrum would be completely dark and no records were created. This leads to a total lack of observations around the South Pole between April and September. The sea ice extent for these months is interpolated using a time series model, see Section 6.2.4.

6.2.1.1 Identification of Poor Image Quality

Sections of poor image quality (i.e., missingness) are already identified in the images. After a closer inspection, these sections are not comprehensive, to the point where a majority of the available images need additional areas to be labeled as missing, see Figure 6.1. This process is very labor intensive and cannot reasonably be automatized. Due to this unexpected impediment in the data cleaning, only the years 1974-1978 are cleaned sufficiently for further analysis. The rest of the images will be cleaned and analyzed in future work.

6.2.1.2 Coordinate Registration

The satellite images contain circles for every 10th degree of latitude and straight lines for every 10th degree of longitude, see Figure 6.1. The latitude lines are not perfectly circular, the longitudinal lines are not perfectly straight. The data contains the information of the latitude and longitude of every pixel, but it does not perfectly match the lines drawn on the physical medium. This information was added semi-automatically to the original data, where humans identified the coordinates of four points on the scanned image which was used to reference the rest of the image. Potential sources of error are the physical image (image could be incorrect or physically warped over time), the scanning process, or the hand labeled

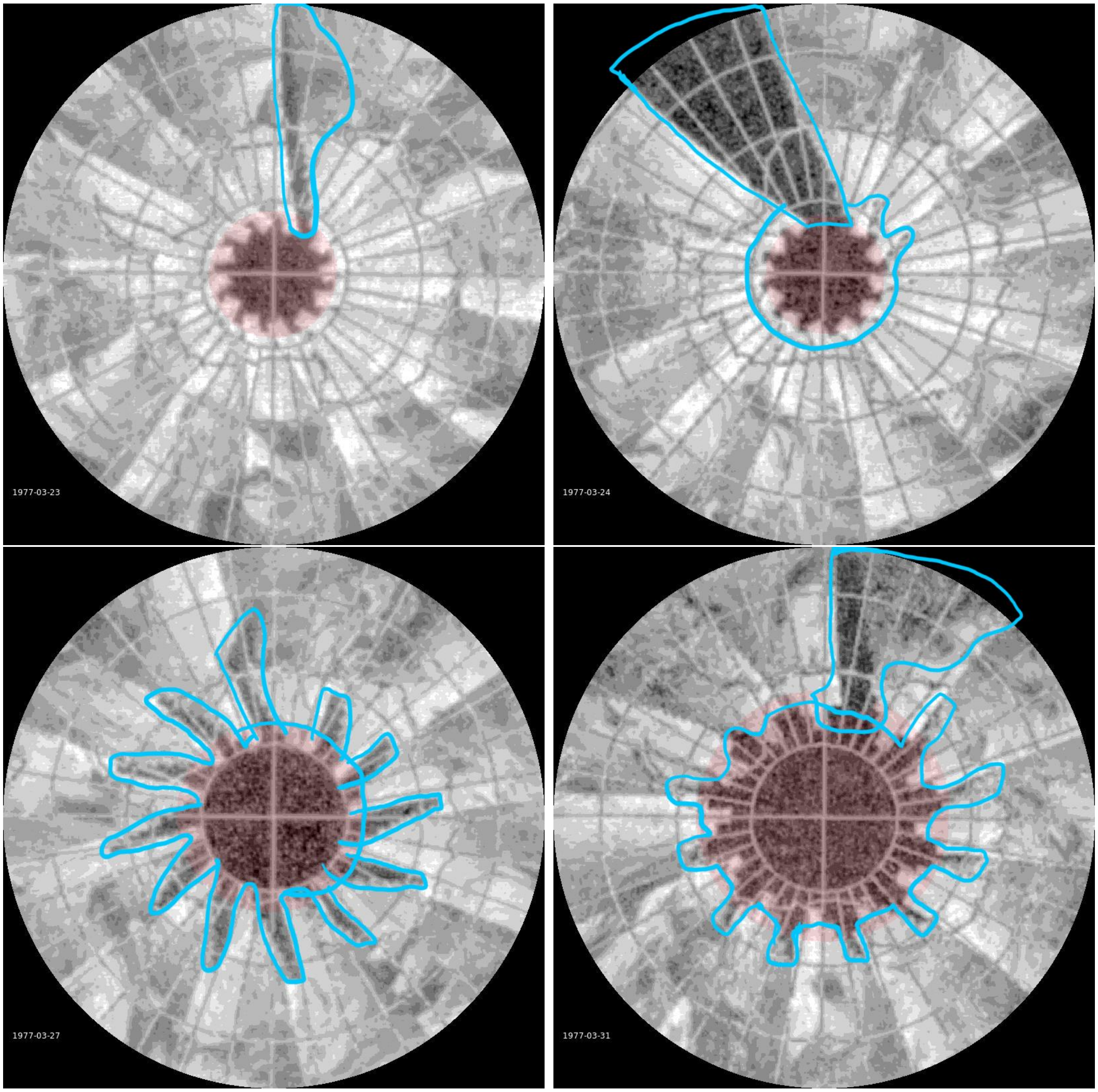


Figure 6.1: Selection of raw images from March 1977. Red areas are marked as missing in the original data set, areas outlined in blue are manually labeled by the author as additionally missing.

reference points. In this analysis, all images are registered (a.k.a. warped) automatically to line up with one reference coordinate grid. The reference grid is created by drawing latitude and longitude lines for every 10 degrees on a white background, see top right of Figure 6.2. The registered images inherit the latitude and longitude of the reference grid. (MCD14) implement global image registration and provide an R interface in the R package `RNiftyReg` (CMP19). The image registration was allowed 6 degrees of freedom (translation + rotation + scale + aspect ratio + shear) to match the coordinate grid. The steps of this matching process are visualized in Figure 6.2. The top left of Figure 6.2 shows the target image containing only the target coordinate grid. The top right of this figure shows the original image with the target coordinate grid overlaid in white. The overlaid coordinate grid (white) and the original coordinate grid (dark) are mismatched, showing that the image was not correctly mapped. The bottom left shows the registered image overlaid with the target coordinate grid. The original coordinate grid is invisible because it is exactly underneath the overlaid target coordinate grid. The bottom right shows the difference between the original image and the registered image. The original coordinate grid is visible as light-colored traces, the coordinate grid after registration as dark traces. The outline of Antarctica is also drawn in the original satellite images.

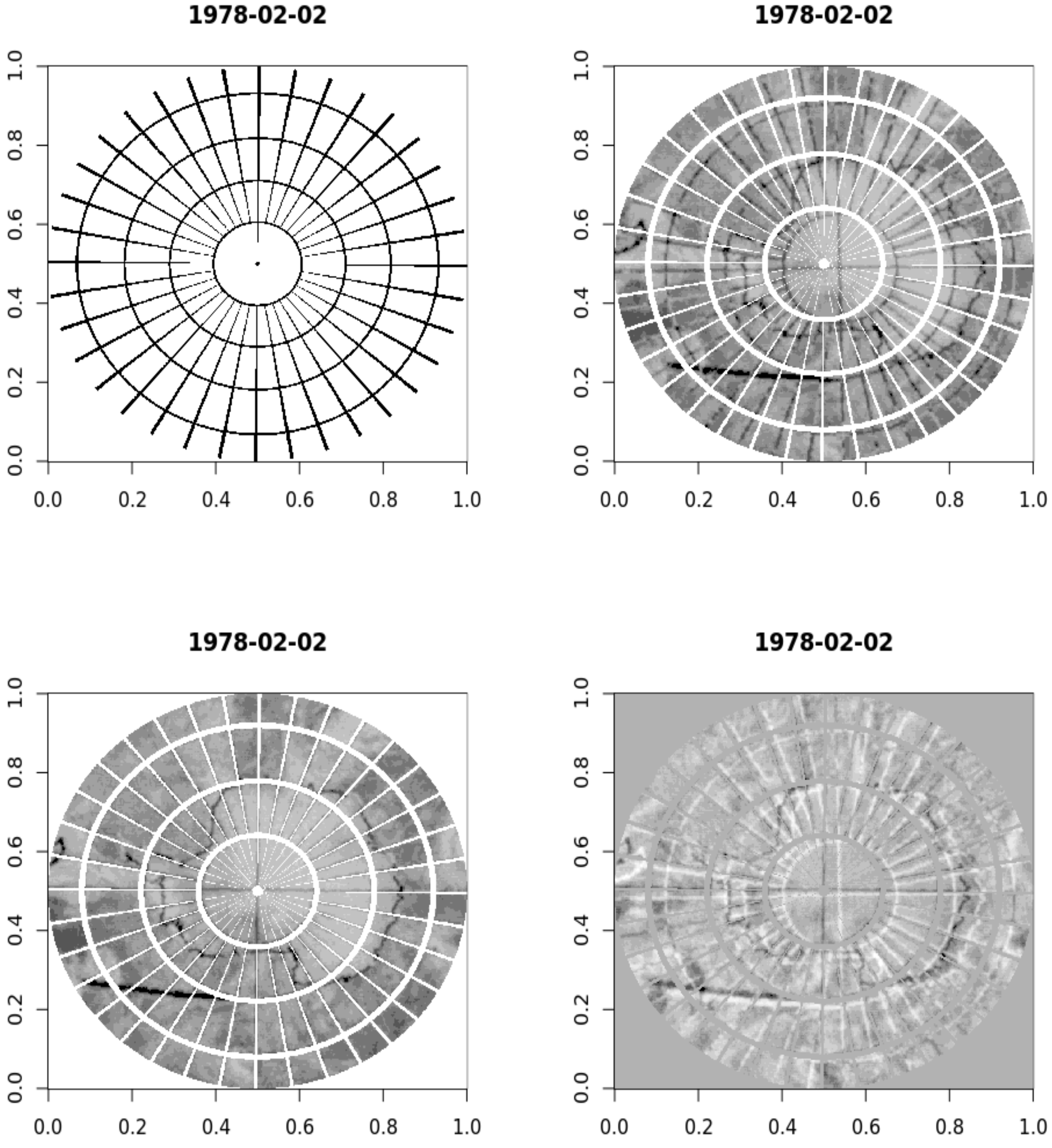


Figure 6.2: Registration process. Top left: Target coordinate grid; Top right: Original image with overlaid coordinate grid; Bottom left: Registered image with overlaid coordinate grid; Bottom right: Difference between original and registered image.

6.2.1.3 Application of Cloud Filter

A large part of the image is occluded by clouds. For this analysis a minimum brightness filter is applied to all images available in each month to remove most of the clouds, see Figure 6.3. This filter also reduces the percentage of missing values in the images. This procedure systematically under-reports the sea ice concentration but majorly increases the stability of any sea ice estimator applied to the data. The increased stability can be seen in the comparison between Figures 6.1 and 6.3.

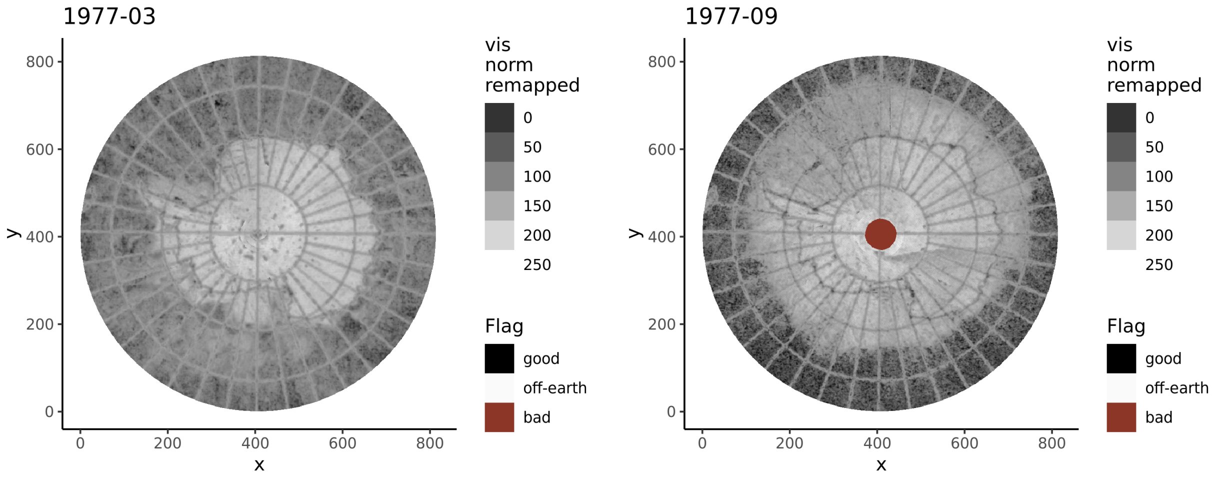


Figure 6.3: Minimum brightness observed per pixel in March and September 1977. This image on the left is aggregated from 30 images available for March 1977, including those shown in Figure 6.1. The image on the right is created from 21 available files for September 1977. March is a month of low sea ice extent, September is a month of high sea ice extent.

6.2.2 Estimation of Sea Ice Concentration

The posterior probability of the sea ice concentration is estimated using a two class Bayesian latent discriminatory analysis (Bayesian LDA). The monthly satellite images are thought of as a matrix of observed brightness values image_{ij} indexed at location ij (ith row and jth colum, $i, j = 1, 2, \ldots, 813$) taking values in $[0, 255]$. The sea ice concentration (or probability of observing ice, we do not differentiate between them) is thought of as an underlying latent matrix with the same dimensionality, with entries ice_{ij} the unobserved sea ice concentration at pixel ij (pixel in i'th row and j'th column) taking values in $[0, 100\%]$.

The quantity of interest in the model is the probability of pixel ij being ice given the observed brightness of that pixel, denoted by

$$P(\text{ice}_{ij}|\text{image}_{ij}).$$

Using the framework of Bayesian LDA, the probability of pixel ij being ice is modeled as

$$\begin{aligned} P(\text{ice}_{ij}|\text{image}_{ij}) &= \frac{P(\text{ice}_{ij})P(\text{image}_{ij}|\text{ice}_{ij})}{P(\text{image}_{ij})} \qquad (6.1) \\ &= \frac{P(\text{ice}_{ij})P(\text{image}_{ij}|\text{ice}_{ij})}{P(\text{ice}_{ij})P(\text{image}_{ij}|\text{ice}_{ij}) + (1 - P(\text{ice}_{ij}))P(\text{image}_{ij}|1 - \text{ice}_{ij})}, \end{aligned}$$

where the prior probability of ice in location ij, denoted by $P(\text{ice}_{ij})$, is estimated from the modern satellite record, and the likelihood of observing a specific brightness at location ij given the underlying sea ice concentration, denoted by $P(\text{image}_{ij}|\text{ice}_{ij})$, is modeled as a Gaussian mixture.

The prior distribution $P(\text{ice}_{ij})$ for the underlying ice concentration for every pixel ij is estimated from the known sea ice concentrations in the years 1978-2019. The prior mean of the sea ice concentration from 1966-1978 is estimated as the average sea ice concentration in the years 1979-2019, see Figure 6.4 showing the average monthly (January = "01", ..., December = "12") sea ice concentration per pixel from 1979-2019. The standard deviation of the prior distribution is estimated in the same fashion (no figure shown).

The sea ice concentration estimated from the data is obtained using a Gaussian mixture model. The distribution of an observed pixel value given the underlying sea ice concentration follows a Normal distribution, i.e.,

$$\text{image}_{ij}|\text{ice}_{ij} = 100\% \sim N(\mu_{\text{ice}}, \sigma_{\text{ice}}), \qquad (6.2)$$

$$\text{image}_{ij}|\text{ice}_{ij} = \quad 0\% \sim N(\mu_{\text{water}}, \sigma_{\text{water}}), \qquad (6.3)$$

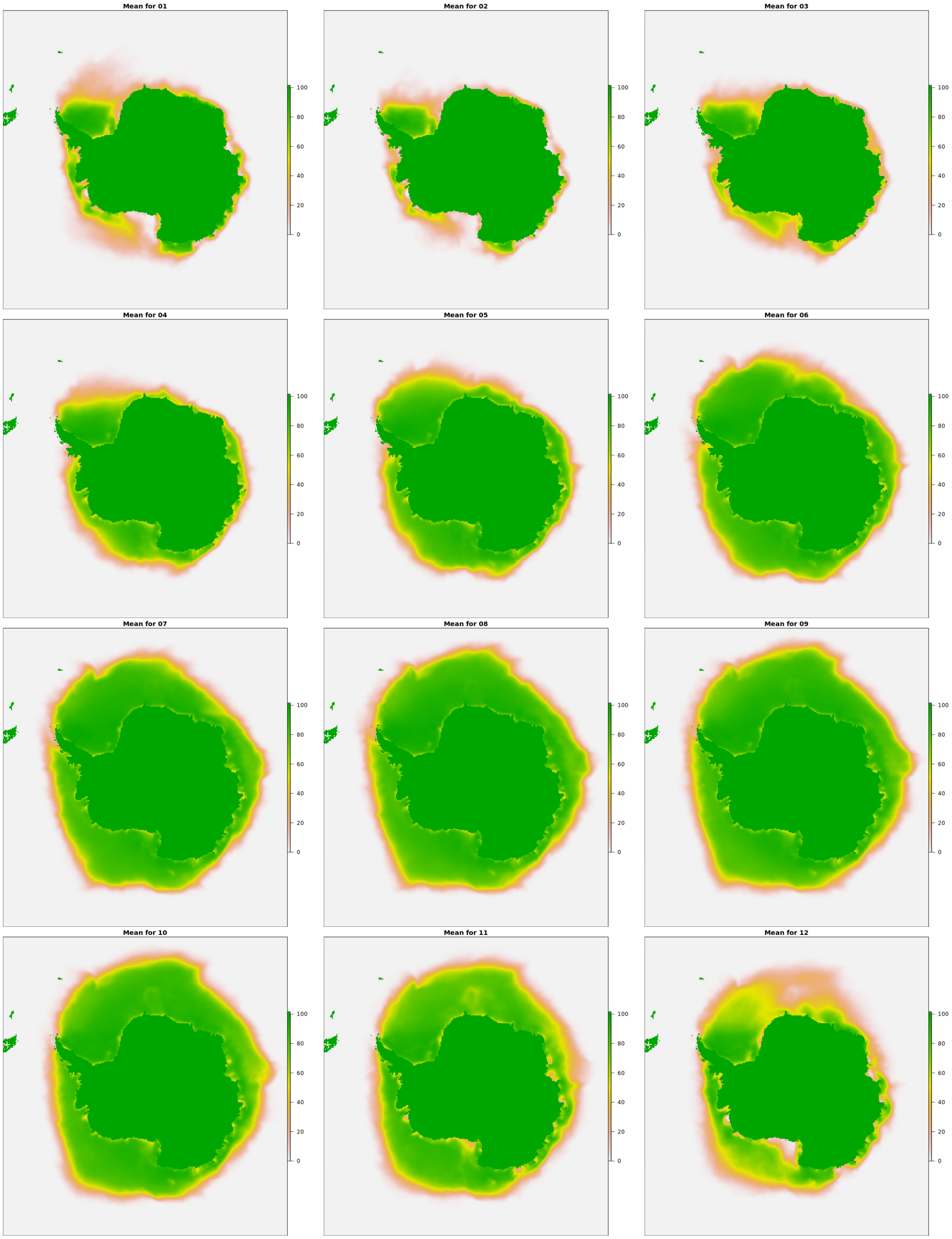


Figure 6.4: Average monthly sea ice concentration per pixel from 1979-2019.

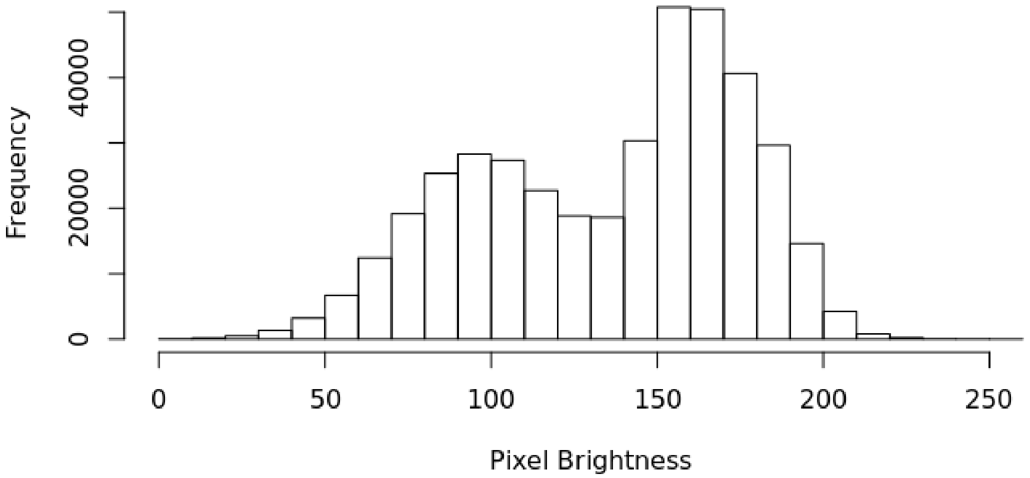


Figure 6.5: Histogram of observed minimum brightness per pixel for September 1977, see Figure 6.3. Two overlapping normal distributions are identified by the Gaussian mixture model proposed in Equations (6.2) and (6.3) with estimated parameters $\mu_{\text{ice}} = 165$, $\sigma_{\text{ice}} = 21$, $\mu_{\text{water}} = 97$, and $\sigma_{\text{water}} = 27$.

where ice is expected to be a lot brighter than water, i.e., $\mu_{\text{ice}} \gg \mu_{\text{water}}$, see Figure 6.5 for a justification.

6.2.3 Estimation of Sea Ice Outline and Extent

The outline of the sea ice extent is found based on a 30% sea ice concentration threshold, see Figure 6.6. The sea ice extent is estimated as the area within this outline minus the area of the continent. The raw sea ice extent is shown in Figure 6.7.

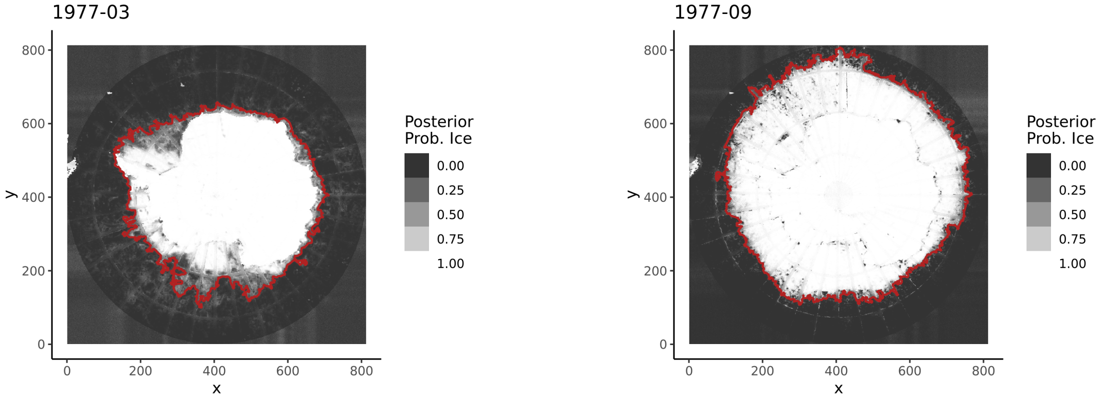


Figure 6.6: Estimated sea ice concentration (color intensity) and outlines (red) for March 1977 and September 1977.

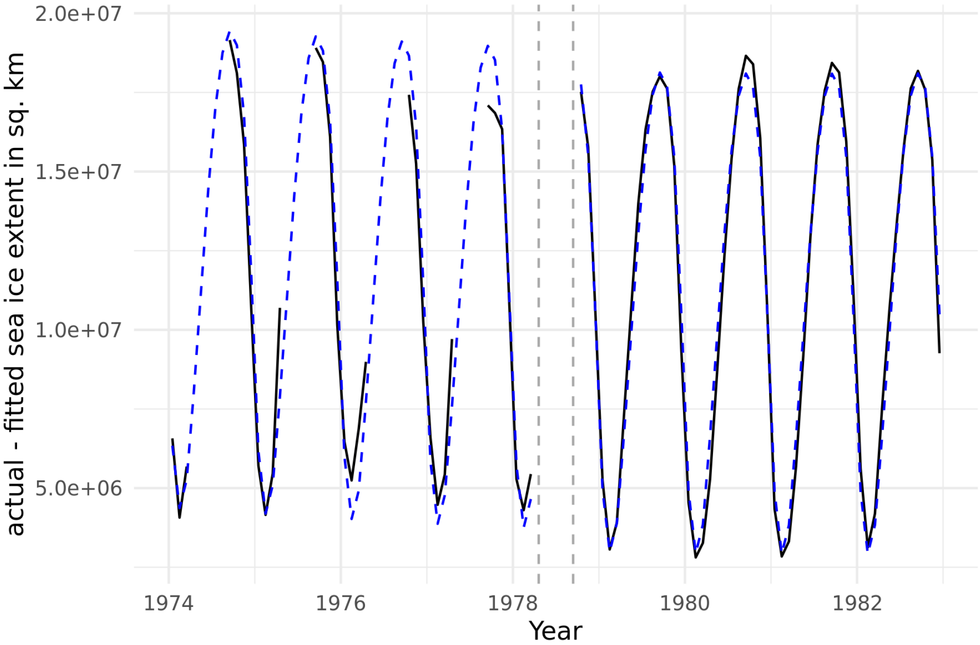


Figure 6.7: Unadjusted estimated Antarctic sea ice extent (solid black) and prediction of the model proposed in Section 6.2.4 (dashed blue). The vertical dashed lines mark the gap in between the data from (Cam19) and the modern satellite data.

6.2.4 Interpolation of Missing Observations

In a final analytical step, a time series model is created taking into account the cyclical nature of the annual trend of the Antarctic sea ice. This will be used to interpolate the sea ice extent of missing months. A time series model with a cyclical annual trend and smooth interannual trend is proposed for the overall sea ice extent SIE(month, year) in each month and year,

$$\text{SIE(month, year)} = \text{cyc(month)} + s(\text{year}) + \varepsilon_{\text{month, year}}, \tag{6.4}$$

where cyc(month) is a cyclical trend over the 12 months in a year (i.e., January and December are as close as any other neighboring months even though they are coded as 'month=1' and 'month=12'), and s(year) is a smooth trend over the years. For the modern satellite record from 1978-2019 this model explains over 99% of the variance in the monthly sea ice extents. The estimated coefficients are depicted in Figure 6.8.

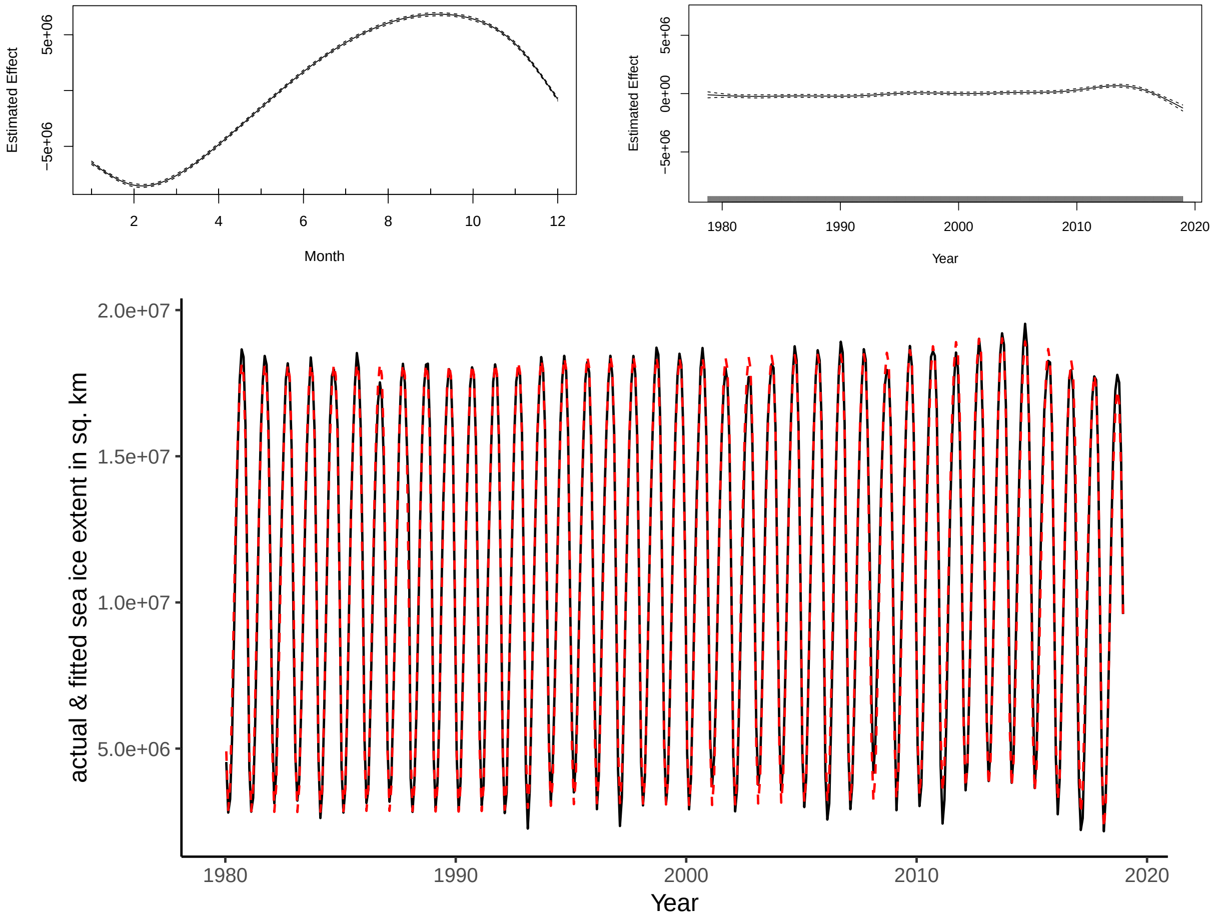


Figure 6.8: Time series model for monthly sea ice extent. Top left: Estimated annual cycle per month (cyc(month)). The characteristic slow growth of Antarctic sea ice March to September and fast melt from October to February is detected; top right, smooth inter-annual trend over the modern-day satellite era (s(year)); bottom: True monthly sea ice extent (black) and sea ice extent fitted by the model (dashed red).

6.2.5 Adjustment of Estimated Sea Ice Extent

In order to include the estimated 1966-1978 sea ice extent, the model proposed in 6.4 is extended to include a constant offset β for the old data. The adjusted estimated sea ice extent is shown in Figure 6.9. This offset accounts for a shift in sea ice extent estimated from old satellite images and the modern satellite images. There are multiple steps in the estimation process that could create such a bias, for example the use of the monthly minimum brightness filter, the choice of threshold for computing the sea ice extent, and the weight the prior probability is given. The sea ice extent as estimated from the old data can be adjusted to the modern satellite data by subtracting this constant.

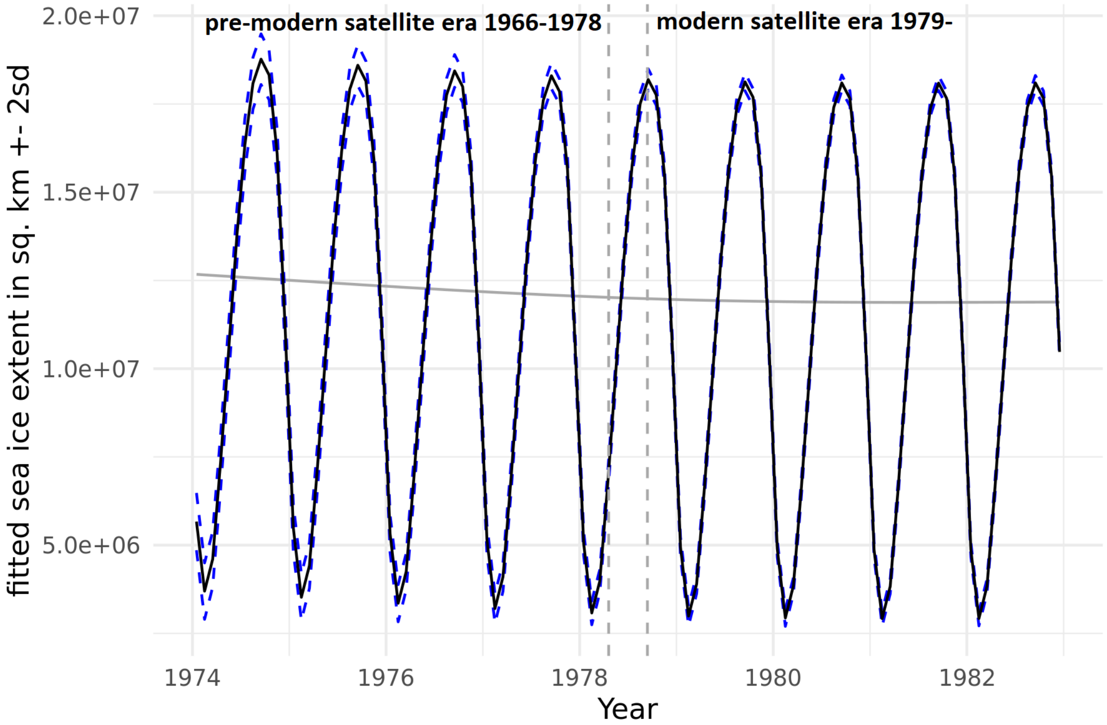


Figure 6.9: Adjusted estimated Antarctic sea ice extent (solid black) with 95% prediction interval (dashed blue) for the years 1974-1983. The grey horizontal line marks the estimated annual mean. The vertical dashed lines mark the gap in between the data from (Cam19) and the modern satellite data.

6.3 Results & Discussion

The Antarctic sea ice is a important driver and gage of the global climate, but a solid record has only been available since 1979. Extending this record back to 1966 increases the available record by 12 years, increasing the available data by 12 / 40 = 24%.

The main products created in this study are the estimates for the sea ice concentration, outline, and extent. Estimates for the monthly sea ice concentration and outline are available from the authors and exemplary shown in Figure 6.6. Estimates of the Antarctic sea ice extent with a 95% confidence interval are shown in Figure 6.9. As expected, the estimated sea ice follows a strong annual cycle very similar to the one observed from modern data. A strong but somewhat uncertain decrease in sea ice extent is picked up by the model in the last years of the Campbell data (1974-1978).

These findings are consistent with previous reconstructions. (CPV03) also find a large increase in sea ice extent in the 1970s, see Figure 6.10. (MGC13) estimate the sea ice extent in September 1966 to be $19.7 \cdot 10^6\text{km}^2$ with a possible range of $[18.9 \cdot 10^6\text{km}^2, 20.4 \cdot 10^6\text{km}^2]$. This high value is consistent with estimates of (CPV03) and also indicates an increase in Antarctic sea ice in the 1970s.

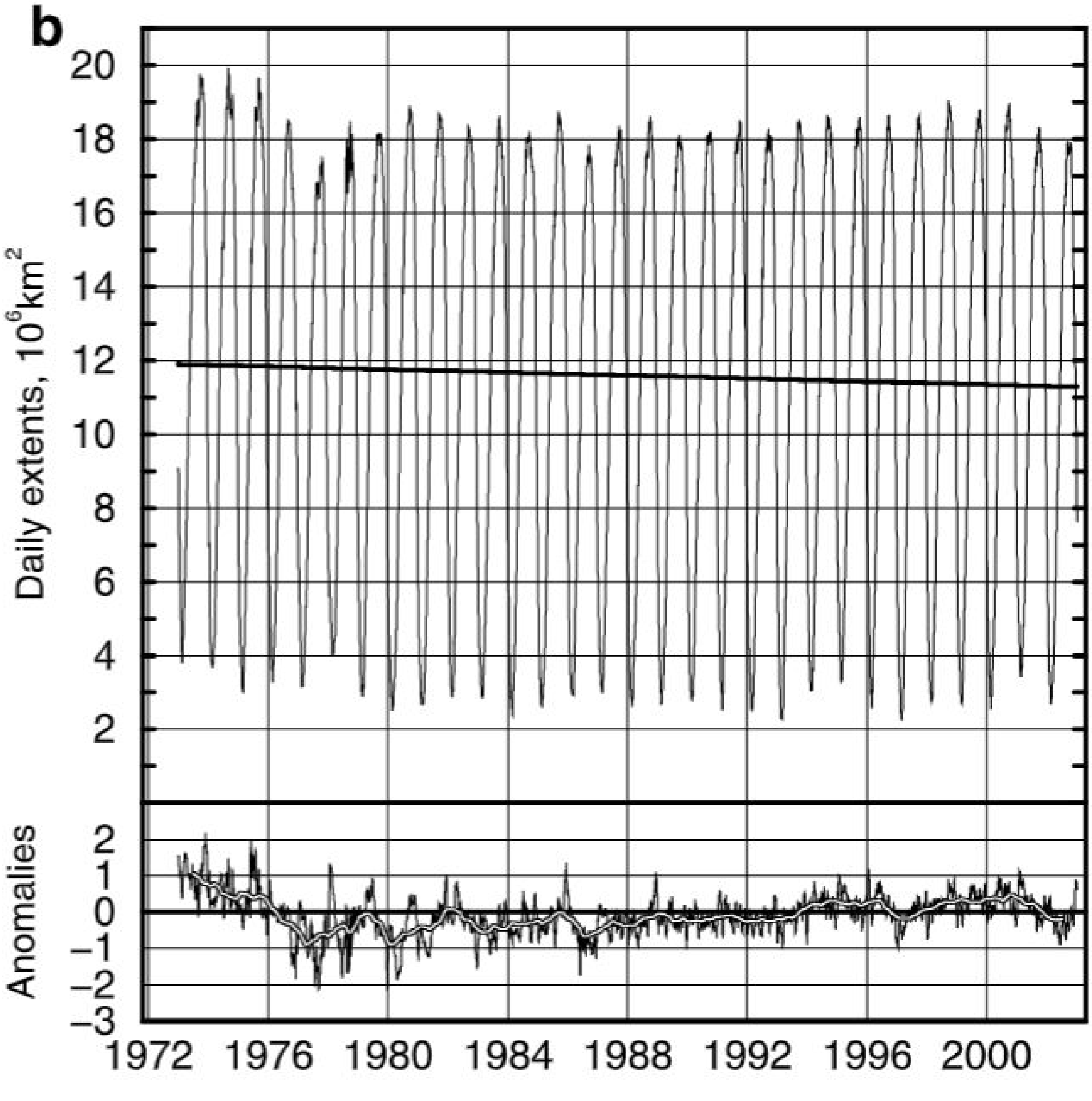


Figure 6.10: Estimates of Antarctic sea ice extent 1972-2002 and anomalies of annual cycle and linear trend. The anomalies show an unusually large increase in sea ice extent between 1972 and 1978 (Source: CPV03).

CHAPTER 7

Conclusions

Antarctic sea ice is an important area of research that can benefit greatly from modern statistical methods. Developing statistical methods for modeling the Antarctic sea ice concentration, sea ice edge, and sea ice extent are essential to bettering our understanding of the complex climate system of the Southern hemisphere. The task of sea ice reconstruction has become increasingly important with the recent dramatic changes in Antarctic sea ice, a dramatic decline in Antarctic sea ice from 2014-2017, a record low in February 2022, and another record low in February 2023. Extending the 42 year satellite record back to the beginning of the 20th century is indispensable to understanding these changes in the bigger context of natural variability and anthropogenic climate change. This dissertation contributes to these goals by introducing a statistical model for Antarctic sea ice extent reconstruction in a Bayesian framework. These ensemble reconstructions of plausible time series reconstructions are superior to current mean reconstructions, by incorporating realistic spatio-temporal autocorrelation and carefully accounting for modeling uncertainty. Future work can expand on this model, chiefly by having a longer satellite observed record available. This will also lead to more predictor variables with a longer period of overlap. Most weather stations used in our reconstruction are still actively collecting data and ice cores drilled now or in the future will have a longer period of overlap with the satellite observed sea ice extent as well. This will allow a more accurate estimation of the association between predictor variables and sea ice extent. Additional data will also support more complex statistical models, including models with more non-zero coefficients, seasonally varying coefficients, and interactions between predictors. A satellite record of 42 years contains sea ice extent observations of 16,000 days. This sounds like a huge sample size, but the relevant number in this case is 42, we only observed the annual cycle and deviations from the annual cycle 42 times. Using the rule of thumb that we need ten data points to estimate one coefficient, models based on current data need to be sparse and can benefit hugely from informative priors.

We also apply functional data analysis to the task of modeling the Antarctic sea ice edge. This novel idea has great potential for future applications like a circumpolar reconstruction of the 20th century Antarctic sea ice edge.

Additionally, we create a statistical methodology for extracting information from early satellite images taken before the beginning of the modern satellite record in 1979. These direct observations of Antarctic sea ice extend the current satellite record substantially, and can readily be applied to other sources of early satellite images as well.

Bibliography

[AMA11] Nerilie J. Abram, Robert Mulvaney, and Carol Arrowsmith. "Environmental signals in a highly resolved ice core from James Ross Island, Antarctica." *Journal of Geophysical Research Atmospheres*, **116**(20):1–15, 2011.

[ATM10] Nerilie J. Abram, Elizabeth R. Thomas, Joseph R. McConnell, Robert Mulvaney, Thomas J. Bracegirdle, Louise C. Sime, and Alberto J. Aristarain. "Ice core evidence for a 20th century decline of sea ice in the Bellingshausen Sea, Antarctica." *Journal of Geophysical Research Atmospheres*, **115**(23):1–9, 2010.

[Aus22] Australian Bureau of Meteorology. "Southern Oscillation Index (SOI) since 1876.", 2022.

[AW10] Hervé Abdi and Lynne J. Williams. "Principal component analysis." *Wiley interdisciplinary reviews: computational statistics*, **2**(4):433–459, 2010.

[AWC03] Stephen Ackley, Peter Wadhams, Josefino C Comiso, and Anthony P Worby. "Decadal decrease of Antarctic sea ice extent inferred from whaling records revisited on the basis of historical and modern sea ice records." *Polar Research*, **22**(1):19–25, 2003.

[AWC13] Nerilie J. Abram, Eric W. Wolff, and Mark A.J. Curran. "A review of sea ice proxy information from polar ice cores." *Quaternary Science Reviews*, **79**:168–183, 11 2013.

[BA11] Dmitry I. Belov and Ronald D. Armstrong. "Distributions of the Kullback–Leibler divergence with applications." *British Journal of Mathematical and Statistical Psychology*, **64**(2):291–309, 2011.

[BCC09] Silvia Becagli, Emiliano Castellano, Omar Cerri, Mark Curran, Massimo Frezzotti, Federica Marino, Andrea Morganti, Marco Proposito, Mirko Severi, Rita Traversi, and Roberto Udisti. "Methanesulphonic acid (MSA) stratigraphy from a Talos Dome ice core as a tool in depicting sea ice changes and southern atmospheric circulation over the previous 140 years." *Atmospheric Environment*, **43**(5):1051–1058, 2009.

[Bes74] Julian Besag. "Spatial Interaction and the Statistical Analysis of Lattice Systems." *Source: Journal of the Royal Statistical Society. Series B (Methodological)*, **36**(2):192–236, 1974.

[BGH95] Julian Besag, Peter Green, David Higdon, and Kerrie Mengersen. "Bayesian Computation and Stochastic Systems." *Source: Statistical Science*, **10**(1):3–41, 1995.

[BK79] George A. Baker Jr and John M Kincaid. "Continuous-Spin Ising Model and λ: $\varphi^4 :_d$ Field Theory." *Physical Review Letters*, **42**(22):1431, 1979.

[BK95] Julian Besag and Charles Kooperberg. "Biometrika Trust On Conditional and Intrinsic Autoregression." *Source: Biometrika*, **82**(4):733–746, 1995.

[Bre01] Leo Breiman. "Random forests." *Machine learning*, **45**(1):5–32, 2001.

[BSB19] E. Brady, S. Stevenson, D. Bailey, Z. Liu, D. Noone, J. Nusbaumer, B. L. Otto-Bliesner, C. Tabor, R. Tomas, T. Wong, et al. "The connected isotopic water cycle in the Community Earth System Model version 1." *Journal of Advances in Modeling Earth Systems*, **11**(8):2547–2566, 2019.

[Bud50] Julius Büdel. *Atlas der Eisverhältnisse des Nordatlantischen Ozeans und Übersichtskarte der Eisverhältnisse des Nord-und Südpolargebietes*. Deutsches Hydrographisches Institut, Hamburg, 1950.

[Cam19] Garrett Campbell. "NOAA Polar Orbiting Satellite (POES) Global Visible and Infrared Band Data from ESSA (1966-1972) and NOAA (1972-1978) Satellites." Technical report, NOAA National Centers for Environmental Information, 2019.

[Cav94] Donald J. Cavalieri. “A microwave technique for mapping thin sea ice.” *Journal of Geophysical Research: Oceans*, **99**(C6):12561–12572, 1994.

[CCD91] D. J. Cavalieri, J. P. Crawford, M. R. Drinkwater, D. T. Eppler, L. D. Farmer, R. R. Jentz, and C. C. Wackerman. “Aircraft active and passive microwave validation of sea ice concentration from the Defense Meteorological Satellite Program special sensor microwave imager.” *Journal of Geophysical Research*, **96**(C12):989–22, 1991.

[Cel14] Alain Celisse. “Optimal cross-validation in density estimation with the L2-loss.” *The Annals of Statistics*, **42**(5):1879–1910, 2014.

[CG07] Cédric Cotté and Christophe Guinet. “Historical whaling records reveal major regional retreat of Antarctic sea ice.” *Deep-Sea Research Part I: Oceanographic Research Papers*, **54**(2):243–252, 2 2007.

[CGC84] D J Cavalieri, P Gloersen, and W J Campbell. “Determination of Sea Ice Parameters With the NIMBUS 7 SMMR.” *Journal of Geophysical Research*, **89**(D4):5355–5369, 1984.

[CGS17] Josefino C. Comiso, Robert A. Gersten, Larry V. Stock, John Turner, Gay J. Perez, and Kohei Cho. “Positive trend in the Antarctic sea ice cover and associated changes in surface temperature.” *Journal of Climate*, **30**(6):2251–2267, 2017.

[CMP19] Jon Clayden, Marc Modat, Benoit Presles, Thanasis Anthopoulos, and Pankaj Daga. *RNiftyReg: Image Registration Using the 'NiftyReg' Library*, 2019. R package version 2.6.7.

[CMS99] Edward R Cook, David M Meko, David W Stahle, and Malcolm K Cleaveland. “Drought Reconstructions for the Continental United States*.” *Journal of Climate*, **12**(4):1145–1162, 1999.

[Com86] J. C. Comiso. “Characteristics of Arctic winter sea ice from satellite multispectral microwave observations.” *Journal of Geophysical Research*, **91**(C1):975–994, 1986.

[COM03] Mark AJ Curran, Tas D van Ommen, Vin I Morgan, Katrina L Phillips, and Anne S Palmer. "Ice Core Evidence for Antarctic Sea Ice Decline Since the 1950s." *Science*, **302**(5648):1203–1206, 2003.

[Cop19] Copernicus Climate Change Service. "Copernicus climate change service climate datastore.", 2019.

[CP08] D. J. Cavalieri and C. L. Parkinson. "Antarctic sea ice variability and trends, 19792006." *Journal of Geophysical Research*, **113**(C7), 7 2008.

[CPS10] Carlos M. Carvalho, Nicholas G. Polson, and James G. Scott. "The horseshoe estimator for sparse signals." *Biometrika*, **97**(2):465–480, 6 2010.

[CPV03] D. J. Cavalieri, C. L. Parkinson, and K. Y. Vinnikov. "30-year satellite record reveals contrasting Arctic and Antarctic decadal sea ice variability." *Geophysical Research Letters*, **30**(18), 9 2003.

[Cre15] Noel Cressie. *Statistics for Spatial Data*. John Wiley & Sons, 2015.

[CS86] J. C. Comiso and C. W. Sullivan. "Satellite microwave and in situ observations of the Weddell Sea ice cover and its marginal ice zone." *Journal of geophysical research: Oceans*, **91**(C8):9663–9681, 1986.

[CS01] Josefino C. Comiso and Konrad Steffen. "Studies of Antarctic sea ice concentrations from satellite data and their applications." *Journal of Geophysical Research: Oceans*, **106**(C12):31361–31385, 12 2001.

[Dav04] E. Roy Davies. *Machine vision: theory, algorithms, practicalities*. Elsevier, 2004.

[De 09] William K. De La Mare. "Changes in Antarctic sea-ice extent from direct historical observations and whaling records." *Climatic Change*, **92**(3-4):461–493, 2 2009.

[DG01] Lijun Ding and Ardeshir Goshtasby. "On the Canny edge detector." *Pattern Recognition*, **34**(3):721–725, 2001.

[DGG13] Anne De Vernal, Rainer Gersonde, Hugues Goosse, Marit-Solveig Seidenkrantz, and Eric W. Wolff. "Sea ice in the paleoclimate system: the challenge of reconstructing sea ice from proxies–an introduction." *Quaternary Science Reviews*, **79**:1–8, 2013.

[DGR21] Quentin Dalaiden, Hugues Goosse, Jeanne Rezsöhazy, and Elizabeth R. Thomas. "Reconstructing atmospheric circulation and sea-ice extent in the West Antarctic over the past 200 years using data assimilation." *Climate Dynamics*, **57**(11-12):3479–3503, 12 2021.

[DLB20] Gokhan Danabasoglu, J.-F. Lamarque, J. Bacmeister, D. A. Bailey, A. K. DuVivier, Jim Edwards, L. K. Emmons, John Fasullo, R. Garcia, Andrew Gettelman, and W. G. Strand. "The Community Earth System Model version 2 (CESM2)." *Journal of Advances in Modeling Earth Systems*, **12**(2):e2019MS001916, 2020.

[DRB17] Hannah M. Director, Adrian E. Raftery, and Cecilia M. Bitz. "Improved sea ice forecasting through spatiotemporal bias correction." *Journal of Climate*, **30**(23):9493–9510, 12 2017.

[DRB19] Hannah M. Director, Adrian E. Raftery, and Cecilia M. Bitz. "Probabilistic Forecasting of the Arctic Sea Ice Edge with Contour Modeling." *arXiv preprint arXiv:1908.09377*, 8 2019.

[ED16] Tom Edinburgh and Jonathan J Day. "Estimating the extent of Antarctic summer sea ice during the Heroic Age of Antarctic Exploration." *The Cryosphere*, **10**(6):2721–2730, 2016.

[EMT01] David B. Enfield, Alberto M. Mestas-Nuñez, and Paul J. Trimble. "The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental US." *Geophysical Research Letters*, **28**(10):2077–2080, 2001.

[FCN18] Ryan L Fogt, Logan N Clark, and Julien P Nicolas. "A New Monthly Pressure Dataset Poleward of 60°S since 1957." *Journal of Climate*, **31**(10):3865–3874, 2018.

[FGA13] Stephen F. Foley, Detlef Gronenborn, Meinrat O. Andreae, Joachim W. Kadereit, Jan Esper, Denis Scholz, Ulrich Pöschl, Dorrit E. Jacob, Bernd R. Schöne, Rainer Schreg, et al. "The Palaeoanthropocene–The beginnings of anthropogenic environmental change." *Anthropocene*, **3**:83–88, 2013.

[Fis37] Ronald Aylmer Fisher. *The design of experiments.* Oliver & Boyd, Edinburgh & London., 1937.

[FJG16] Ryan L. Fogt, Julie M. Jones, Chad A. Goergens, Megan E. Jones, Grant A. Witte, and Ming Yueng Lee. "Antarctic station-based seasonal pressure reconstructions since 1905: 2. Variability and trends during the twentieth century." *Journal of Geophysical Research*, **121**(6):2836–2856, 3 2016.

[FM20] Ryan L. Fogt and Gareth J. Marshall. "The Southern Annular Mode: variability, trends, and climate impacts across the Southern Hemisphere." *Wiley Interdisciplinary Reviews: Climate Change*, **11**(4):e652, 2020.

[FPM09] Ryan L. Fogt, Judith Perlwitz, Andrew J. Monaghan, David H. Bromwich, Julie M. Jones, and Gareth J. Marshall. "Historical SAM variability. Part II: Twentieth-century variability and trends from reconstructions, observations, and the IPCC AR4 models." *Journal of Climate*, **22**(20):5346–5365, 2009.

[FSR22] Ryan L. Fogt, Amanda M. Sleinkofer, Marilyn N. Raphael, and Mark S. Handcock. "A regime shift in seasonal total Antarctic sea ice extent in the twentieth century." *Nature Climate Change*, **12**(1):54–62, 1 2022.

[Gal85] James Gall. "Use of cylindrical projections for geographical astronomical, and scientific purposes." *Scottish Geographical Magazine*, **1**(4):119–123, 1885.

[GC86] Per Gloersen and Donald J. Cavalieri. "Reduction of weather effects in the calculation of sea ice concentration from microwave radiances." *Journal of Geophysical Research: Oceans*, **91**(C3):3913–3919, 1986.

[GCC93] Per Gloersen, William J Campbell, Donald J Cavalieri, Josefino C Comiso, Claire L Parkinson, and H Jay Zwally. “Satellite passive microwave observations and analysis of Arctic and Antarctic sea ice, 1978-1987.” *Annals of Glaciology*, **17**, 1993.

[GCD13] Nicholas R. Goodwin, Lisa J. Collett, Robert J. Denham, Neil Flood, and Daniel Tindall. “Cloud and cloud shadow screening across Queensland, Australia: An automated method for Landsat TM/ETM+ time series.” *Remote Sensing of Environment*, **134**:50–65, 2013.

[GCM14] David W. Gallaher, G. Garrett Campbell, and Walter N. Meier. “Anomalous variability in Antarctic sea ice extents during the 1960s with the use of Nimbus data.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **7**(3):1116–1122, 2014.

[GCS04] Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 2004.

[GGF15] M. Gagné, N. P. Gillett, and J. C. Fyfe. “Observed and simulated changes in Antarctic sea ice extent over the past 50 years.” *Geophysical Research Letters*, **42**(1):90–95, 1 2015.

[GPH16] Benedikt Gräler, Edzer Pebesma, and Gerard Heuvelink. “Spatio-Temporal Interpolation using gstat.” *The R Journal*, **8**:204–218, 2016.

[GSH19] Jeff Goldsmith, Fabian Scheipl, Lei Huang, Julia Wrobel, Chongzhi Di, Jonathan Gellar, Jaroslaw Harezlak, Mathew W. McLean, Bruce Swihart, Luo Xiao, Ciprian Crainiceanu, and Philip T. Reiss. *refund: Regression with Functional Data*, 2019. R package version 0.1-21.

[GWC74] P Gloersen, T T Wilheit, T C Chang, W Nordberg, and W J Campbell. “Microwave Maps ol the Polar Ice of the Earth.” *Bulletin American Meteorological Society*, **55**(12):1442–1448, 1974.

[Har90] Andrew C. Harvey. *Forecasting, structural time series models and the Kalman filter*. Cambridge university press, 1990.

[HBF15] Boyin Huang, Viva F. Banzon, Eric Freeman, Jay Lawrimore, Wei Liu, Thomas C. Peterson, Thomas M. Smith, Peter W. Thorne, Scott D. Woodruff, and Huai-Min Zhang. "Extended reconstructed sea surface temperature version 4 (ERSST v4). Part I: Upgrades and intercomparisons." *Journal of climate*, **28**(3):911–930, 2015.

[HBR15] William Richard Hobbs, Nathaniel L. Bindoff, and Marilyn N. Raphael. "New perspectives on observed and simulated Antarctic sea ice extent trends using optimal fingerprinting techniques." *Journal of Climate*, **28**(4):1543–1560, 2015.

[HCA16] Will Hobbs, Mark Curran, Nerilie Abram, and Elizabeth R. Thomas. "Century-scale perspectives on observed and simulated Southern Ocean sea ice trends from proxy reconstructions." *Journal of Geophysical Research: Oceans*, **121**(10):7804–7818, 10 2016.

[HMS16] William R. Hobbs, Rob Massom, Sharon Stammerjohn, Phillip Reid, Guy Williams, and Walter Meier. "A review of recent changes in Southern Ocean sea ice, their drivers and forcings.", 8 2016.

[Hol91] J. P. Hollinger. "DMSP special sensor microwave/imager calibration/validation." Technical report, Naval Research Lab Washington DC, 1991.

[HR05] Eileen H. Helmer and Bonnie Ruefenacht. "Cloud-free satellite image mosaics with regression trees and histogram matching." *Photogrammetric Engineering & Remote Sensing*, **71**(9):1079–1089, 2005.

[HR20] Mark S. Handcock and Marilyn N. Raphael. "Modeling the annual cycle of daily Antarctic sea ice extent." *Cryosphere*, **14**(7):2159–2172, 7 2020.

[JFW09] Julie M. Jones, Ryan L. Fogt, Martin Widmann, Gareth J. Marshall, Phil D. Jones, and Martin Visbeck. "Historical SAM variability. Part I: Century-length seasonal reconstructions." *Journal of Climate*, **22**(20):5319–5345, 2009.

[JT04] Ameet K. Jain and Russell H. Taylor. "Understanding bone responses in B-mode ultrasound images and automatic bone surface extraction using a Bayesian probabilistic framework." In *Medical Imaging 2004: Ultrasonic Imaging and Signal Processing*, volume 5373, p. 131. SPIE, 4 2004.

[KBW23] Harrison Katz, Kai Brusch, and Robert E Weiss. "A Bayesian Dirichlet Auto-Regressive Moving Average Model for Forecasting Lead Times." *in submission*, 2023.

[KL51] Solomon Kullback and Richard A. Leibler. "On information and sufficiency." *The annals of mathematical statistics*, **22**(1):79–86, 1951.

[Kno06] George A. Knox. *Biology of the Southern Ocean.* CRC Press, 2006.

[KPM13] Michael D. King, Steven Platnick, W. Paul Menzel, Steven A. Ackerman, and Paul A. Hubanks. "Spatial and temporal distribution of clouds observed by MODIS onboard the terra and aqua satellites." *IEEE Transactions on Geoscience and Remote Sensing*, **51**(7):3826–3852, 2013.

[Lac10] Tom Lachlan-Cope. "Antarctic clouds." *Polar Research*, **29**(2):150–158, 8 2010.

[Lam70] Johann Heinrich Lambert. *Beyträge zum Gebrauche der Mathematik und deren Anwendung*, volume 2. Verlag des Buchladens der Realschule, 1770.

[LTL13] Chao Hung Lin, Po Hung Tsai, Kang Hua Lai, and Jyun Yuan Chen. "Cloud removal from multitemporal satellite images using information cloning." *IEEE Transactions on Geoscience and Remote Sensing*, **51**(1):232–241, 2013.

[Mac72] Neil Alison Mackintosh. "Life cycle of Antarctic krill in relation to ice and water conditions." *Discovery Rep.*, **36**:1–94, 1972.

[Mai23a] Thomas J Maierhofer. "20th Century Antarctic sea ice extent anomaly reconstruction by sector.", 05 2023.

[Mai23b] Thomas J Maierhofer. "Videos of 1966-1978 Satellite Images over Antarctica.", 05 2023.

[Mai23c] Thomas J Maierhofer. “Videos of Functional Regression Model for the Antarctic Sea Ice Edge.”, 05 2023.

[Mar97] William K. De la Mare. “Abrupt mid-twentieth-century decline in Antarctic sea-ice extent from whaling records.” *Nature*, **389**(6646):57, 1997.

[Mar03] Gareth J. Marshall. “Trends in the Southern Annular Mode from observations and reanalyses.” *Journal of climate*, **16**(24):4134–4143, 2003.

[MC00] Thorsten Markus and Donald J Cavalieri. “An Enhancement of the NASA Team Sea Ice Algorithm.” *IEEE Transactions on Geoscience and Remote Sensing*, **38**(3):1387–1398, 2000.

[MCA14] E. J. Murphy, A. Clarke, N. J. Abram, and J. Turner. “Variability of sea-ice in the northern Weddell Sea during the 20th century.” *Journal of Geophysical Research: Oceans*, **119**(7):4549–4572, 2014.

[MCD14] Marc Modat, David M. Cash, Pankaj Daga, Gavin P. Winston, John S. Duncan, and Sbastien Ourselin. “Global image registration using a symmetric block-matching approach.” *Journal of Medical Imaging*, **1**(2):024003, 9 2014.

[MCS95] Eugene J Murphy, Andrew Clarke, Carolyn Symon, and Julian Priddle. “Temporal variation in Antarctic sea-ice: analysis of a long term fast-ice record from the South Orkney Islands.” *Deep-Sea Research I*, **42**(7):1045–1062, 1995.

[Mei05] Walter N. Meier. “Comparison of passive microwave ice concentration algorithm retrievals with AVHRR imagery in arctic peripheral seas.” *IEEE Transactions on Geoscience and Remote Sensing*, **43**(6):1324–1337, 6 2005.

[MFW16] W. N. Meier, F. Fetterer, A. K. Windnagel, and J. S. Stewart. “NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration, Version 4.”, 2016.

[MGC13] W. N. Meier, D. Gallaher, and G. G. Campbell. “New estimates of Arctic and Antarctic sea ice extent during September 1964 from recovered Nimbus I satellite imagery.” *The Cryosphere*, **7**(2):699–705, 4 2013.

[MH40] Neil Alison Mackintosh and HFP Herdman. *Distribution of the pack-ice in the Southern Ocean.* University Press, 1940.

[MH02] Nathan J. Mantua and Steven R. Hare. “The Pacific decadal oscillation.” *Journal of Oceanography*, **58**(1):35–44, 2002.

[Mil81] Robert G. Miller. *Simultaneous Statistical Inference.* Springer New York, NY, 1981.

[MMK02] Eric A Meyerson, Paul A Mayewski, Karl J Kreutz, L David Meeker, Sallie I Whitlow, and Mark S Twickler. “The polar expression of ENSO and sea-ice variability as recorded in a South Pole ice core.” *Annals of Glaciology*, **35**:430–436, 2002.

[Par19] Claire L. Parkinson. “A 40-y record reveals gradual Antarctic sea ice increases followed by decreases at rates far exceeding the rates seen in the Arctic.” *Proceedings of the National Academy of Sciences of the United States of America*, **116**(29):14414–14423, 2019.

[Pav78] Theodosios Pavlidis. “SURVEY A Review of Algorithms for Shape Analysis.” *Computer graphics and image processing*, **7**(2):243–258, 1978.

[PC12] C L Parkinson and D J Cavalieri. “Antarctic Sea Ice Variability and Trends.” Technical report, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA, 2012.

[PEC16] Ariaan Purich, Matthew H. England, Wenju Cai, Yoshimitsu Chikamoto, Axel Timmermann, John C. Fyfe, Leela Frankcombe, Gerald A. Meehl, and Julie M Arblaster. “Tropical Pacific SST drivers of recent Antarctic sea ice trends.” *Journal of Climate*, **29**(24):8931–8948, 2016.

[PHS14] Tomi Peltola, Aki S. Havulinna, Veikko Salomaa, and Aki Vehtari. “Hierarchical Bayesian Survival Analysis and Projective Covariate Selection in Cardiovascular Event Risk Prediction.” *BMA@ UAI*, **27**:79–88, 2014.

[PPM16] Stacy E. Porter, Claire L. Parkinson, and Ellen Mosley-Thompson. "Bellingshausen sea ice extent recorded in an Antarctic peninsula ice core." *Journal of Geophysical Research*, **121**(23):886–13, 12 2016.

[PV17] Juho Piironen and Aki Vehtari. "Sparsity information and regularization in the horseshoe and other shrinkage priors." *Electronic Journal of Statistics*, **11**(2):5018–5051, 2017.

[RH14] Marilyn N. Raphael and Will Hobbs. "The influence of the large-scale atmospheric circulation on Antarctic sea ice during ice advance and retreat seasons." *Geophysical Research Letters*, **41**(14):5037–5045, 7 2014.

[RH22] Marilyn N. Raphael and Mark S. Handcock. "A new record minimum for Antarctic sea ice." *Nature Reviews Earth & Environment*, **3**(4):215–216, 2022.

[RL19] Donald B. Rubin and Roderick J. A. Little. *Statistical analysis with missing data.* John Wiley & Sons, 2019.

[RPH03] N. A. Rayner, De E Parker, EB Horton, Chris K Folland, Lisa V Alexander, DP Rowell, EC Kent, and A Kaplan. "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, 2003.

[RS06] James O. Ramsay and B. W. Silverman. *Functional Data Analysis, 2nd ed.* Springer, New York, 2006.

[SBB14] Kate E. Sinclair, Nancy A.N. Bertler, Melissa M. Bowen, and Kevin R. Arrigo. "Twentieth century sea-ice trends in the Ross Sea from a high-resolution, coastal ice-core record." *Geophysical Research Letters*, **41**(10):3510–3516, 5 2014.

[SCO16] Peter A. Stott, Nikolaos Christidis, Friederike E. L. Otto, Ying Sun, Jean-Paul Vanderlinden, Geert Jan van Oldenborgh, Robert Vautard, Hans von Storch, Peter Walton, Pascal Yiou, et al. "Attribution of extreme weather and climate-related events." *Wiley Interdisciplinary Reviews: Climate Change*, **7**(1):23–41, 2016.

[Sco19] Steven L. Scott. *bsts: Bayesian Structural Time Series*, 2019. R package version 0.9.2.

[SG99] Chris Stauffer and W E L Grimson. "Adaptive background mixture models for real-time tracking." *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, **2**:246–252, 1999.

[Sim15] Ian Simmonds. "Comparing and contrasting the behaviour of Arctic and Antarctic sea ice over the 35 year period 1979-2013." *Annals of Glaciology*, **56**(69):18–28, 2015.

[SMS08] S. E. Stammerjohn, D. G. Martinson, R. C. Smith, X. Yuan, and D. Rind. "Trends in Antarctic annual sea ice retreat and advance and their relation to El NiñoSouthern Oscillation and Southern Annular Mode variability." *Journal of Geophysical Research*, **113**(C3), 3 2008.

[Sta22] Stan Development Team. "The Stan Core Library.", 2022. Version 2.31.

[TA16] Elizabeth R. Thomas and Nerilie J. Abram. "Ice core reconstruction of sea ice change in the Amundsen-Ross Seas since 1702 A.D." *Geophysical Research Letters*, **43**(10):5309–5317, 2016.

[TAE19] Elizabeth R. Thomas, Claire S. Allen, Johan Etourneau, Amy C.F. King, Mirko Severi, V. Holly L. Winton, Juliane Mueller, Xavier Crosta, and Victoria L. Peck. "Antarctic sea ice proxies from marine and ice core archives suitable for reconstructing sea ice over the past 2000 years.", 12 2019.

[TE93] Robert J. Tibshirani and Bradley Efron. "An introduction to the bootstrap." *Monographs on statistics and applied probability*, **57**:1–436, 1993.

[Tol66] E. I. Tolstikov. "Atlas of the Antarctic, vol. 1, 126 pp." *Glav. Uprav. Geod. Kart. MGSSSR, Moscow*, 1966.

[TSM12] Karl E. Taylor, Ronald J. Stouffer, and Gerald A. Meehl. "An overview of CMIP5 and the experiment design." *Bulletin of the American Meteorological Society*, **93**(4):485–498, 2012.

[YXL21] Jiao Yang, Cunde Xiao, Jiping Liu, Shutong Li, and Dahe Qin. "Variability of Antarctic sea ice extent over the past 200 years." *Science Bulletin*, **66**(23):2394–2404, 12 2021.

[ZC19] Bohai Zhang and Noel Cressie. "Estimating Spatial Changes Over Time of Arctic Sea Ice using Hidden 2x2 Tables." *Journal of Time Series Analysis*, **40**(3):288–311, 5 2019.

[ZL04] Dengsheng Zhang and Guojun Lu. "Review of shape representation and description techniques." *Pattern Recognition*, **37**(1):1–19, 1 2004.

[ZPC83] H Jay Zwally, C L Parkinson, and J C Comiso. "Variability of Antarctic Sea Ice and Changes in Carbon Dioxide." *Science*, **220**(4601):1005–1012, 1983.